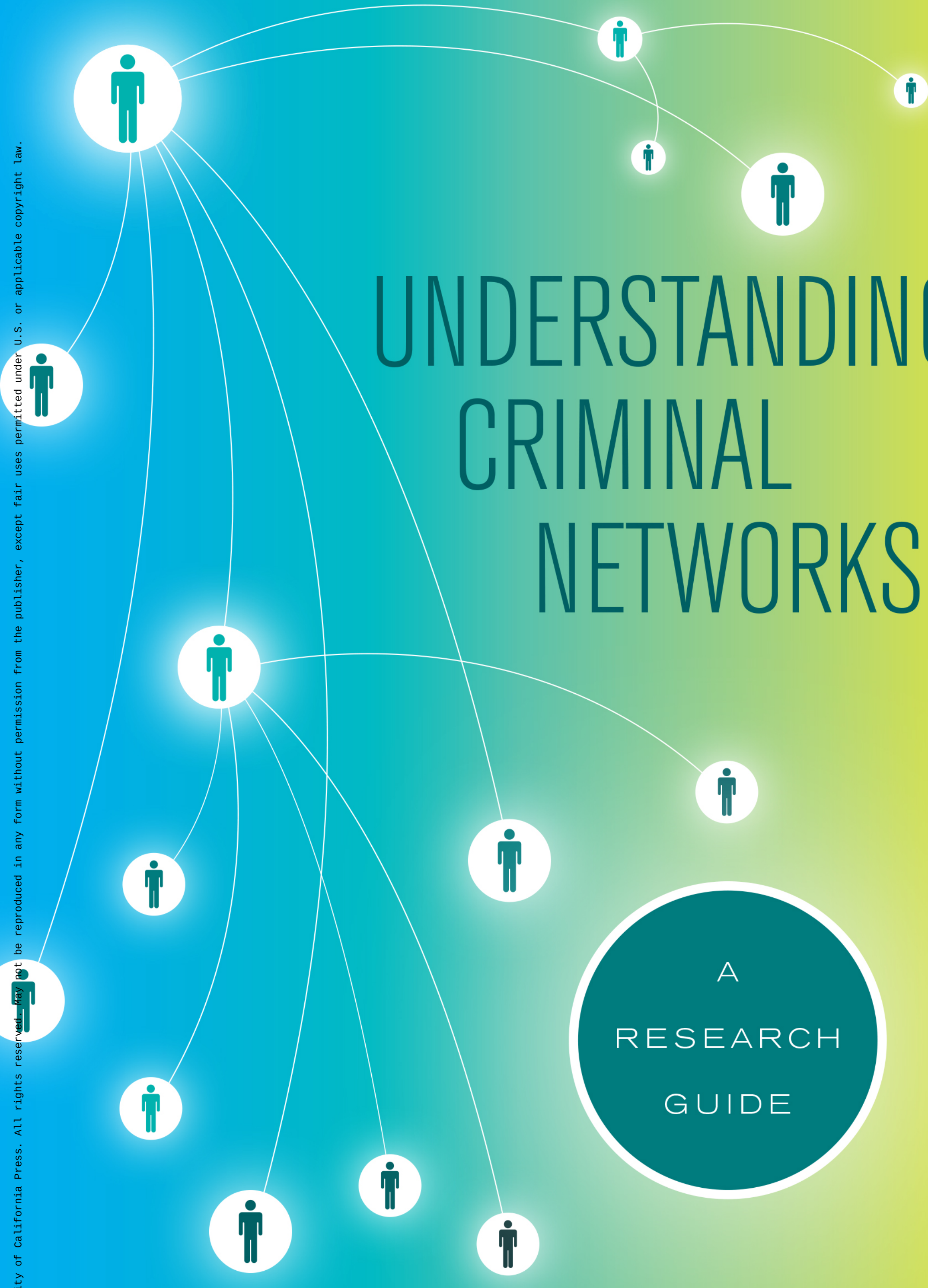


UNDERSTANDING CRIMINAL NETWORKS

A RESEARCH GUIDE

GISELA BICHLER

Understanding Criminal Networks

A Research Guide

GISELA BICHLER

University of California Press

Understanding Criminal Networks

Understanding Criminal Networks

A Research Guide

GISELA BICHLER

University of California Press

University of California Press
Oakland, California

Cataloging-in-Publication Data is on file at the Library of Congress.

ISBN 978–0-520–29704–3 (cloth : alk. paper)
ISBN 978–0-520–29705–0 (pbk. : alk. paper)
ISBN 978–0-520–96956–8 (ebook)
LCCN: 2019943248

28 27 26 25 24 23 22 21 20 19
10 9 8 7 6 5 4 3 2 1

To my partner, Illya, for his unwavering support, readiness
to celebrate every win and every loss, and his ability
to keep himself occupied while I hammer away at the keyboard.

Contents

Illustrations and Tables

FIGURES

TABLES

1. Read First

> The concept of a network emphasizes the fact that each individual has ties to other individuals, each of whom in turn is tied to a few, some, or many others, and so on.
>
> WASSERMAN AND FAUST, *Social Network Analysis*

Social context is important. In fact, some argue that it is the most significant predictor of human behavior: understanding why people act the way they do requires an investigation of the social context within which people are embedded. Your connections with others, in other words, ensnare you in a web of relations. Interactions stemming from relations with friends and associates, colleagues, and family influence your attitudes, opportunities, and activities. While many of these influences have positive effects (e.g., finding a new job, discovering a TV show for the next binge session, getting travel advice on an upcoming vacation), networks also expose people to crime and deviance. It is not just the influence of direct contacts that is of interest but also the pattern of connections involving the people you are indirectly connected to. Since your connectivity with others forms channels through which information may pass, sometimes the information passed to you through a friend will significantly affect your thoughts and behavior. To illustrate how patterns of connectivity create social context, let's consider two scenarios.

Charlie, represented in figure 1.1 by the gray emoji, has an idea. Charlie is fed up with using heroin. Too many of his friends have died or left the neighborhood to escape the situation. He feels terrible all the time, and he is tired of all the things he has to do to feed his habit. Luckily, Charlie lives in Seattle, where he was contacted by a representative of the LEAD program—Seattle's Law Enforcement Assistance Diversion program. (For more information, see Marcella Gaviria's 2016 documentary *Chasing Heroin*.) The objective of LEAD is to reduce the harms associated with drug addiction and to help people engage in effective treatment.

In the first panel, Charlie is part of a close-knit group of five people, including Charlie. The lines connecting emojis indicate active relationships. People represented with sad faces do not support LEAD. Notice that Charlie

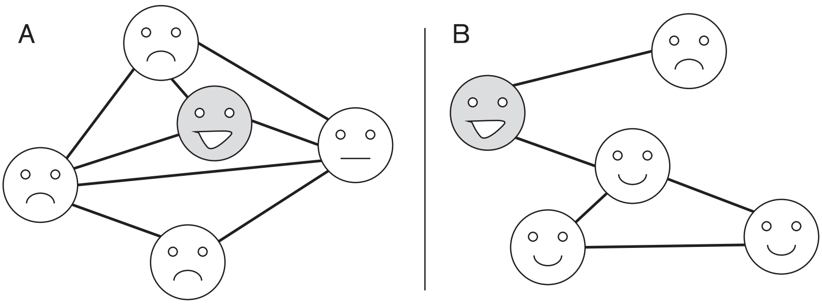


FIGURE 1.1. Two illustrations of the influence of social structure on Charlie. *A*, Charlie's network is close-knit; *B*, Charlie's network is loosely structured.

does not have a direct relationship with the person appearing below him; however, this individual is connected to two of Charlie's closest friends. Even though the third party has an indirect connection, it is still an important part of Charlie's local social world. This person exerts direct influence on two of Charlie's close associates and may cause the LEAD-ambivalent friend to turn against the program.

The social structure depicted in panel A will cause considerable problems for Charlie. If Charlie's closest associates (the ones he would need to rely on for social support to do well in the program) do not trust LEAD outreach workers, or they are ambivalent about the program, then Charlie will face even more obstacles to getting clean. His friends, who make up an insulated group, talk when Charlie is not around, reaffirming their positions. Whether intentional or not, they engage in similar interactions with Charlie, behaving in ways that continue to enable Charlie's addiction. After all, the group might feel that Charlie is much more fun when he is high and that he is really good at supplying drugs to the group, in which case his sobriety would impact their lifestyle. Change is difficult in any circumstance, but it is even harder in this situation because the nature of the change conflicts with the group's social norms.

In the second panel, Charlie is integrated into a very different social structure. In this scenario, his network still includes five people, including two close associates, who are split between two different sets of people. Having unique clusters of friends, one of which supports Charlie's enrolling in LEAD, provides a social structure more conducive to change—the influence of the grumpy friend (sad face emoji) could be countered by the supportive friend (happy face emoji). In addition, while those directly connected to an individual are important for understanding behavior, the influence exerted indirectly must be accounted for—that is, through a friend of a friend. Charlie's supportive friend is connected to others, who also support the efforts of LEAD. These third parties may bolster Charlie's friend, encouraging that person to continue helping Charlie get to drug counseling even when it gets difficult. In other words, indirect influences from people two steps removed may directly influence Charlie's friend and indirectly reinforce Charlie's efforts to do his best with the opportunities the LEAD program affords him.

From this example, we begin to see why social context is important. Interactions with others generate the social structure within which people form ideas, make decisions, encounter opportunities, and engage in behaviors. While geography is important—Charlie might not be able to access a program like LEAD if he lived elsewhere—Charlie's relationships play a critical role in his decision to treat his drug addiction and in his toughing it out when his sobriety is threatened. Whom you know is a vital factor in understanding what you think and do. But social support is not the only phenomenon associated with crime that flows through networks—even crime risks transmit through networks.

For example, after decades of inquiry into the patterns of homicide and gunshot victimization, researchers have found that gun violence is concentrated among a small, interlinking portion of the community, suggesting that risk of being caught up in violence has less to do with who you are and more to do with whom you interact with. Andrew Papachristos (2009, 75) argues that "Gang members do not kill because they are poor, black, or young or live in a socially disadvantaged neighborhood. They kill because they live in a structured set of social relations in which violence works its way through a series of connected individuals." This means that the risk of becoming involved in violence, as a victim or as an offender, is contingent on the structure of social relations—violence acts much like other social phenomena, spreading through a network as individuals react to the behavior of others. Studying gun violence in Chicago, Papachristos (2009, 76) found that social networks place "adversaries in positions where each must attempt to defend, maintain, or repair their reputation." Patterns of networked violence emerge when we aggregate information about individual-level disputes (e.g., rivalries among competing groups, retribution for perceived injustice or harm, and the need to avoid subjugation by others) and internal struggles for control of the group. This line of thought leads him to remark that gang-related homicides are best explained by understanding the way violent conflict works its way through a series of direct and indirect connections.

Unfortunately, conventional research and analytic approaches do not offer many ways to study relationships among people and within groups. For this reason, a growing number of researchers and analysts are turning to the field of social network analysis (SNA). Rather than assuming independence, as many conventional scientific methodologies do, a central axiom underlying the social network perspective is that *dependencies matter.* Working from an interdisciplinary perspective, network scholars are continually developing analytic techniques and metrics specifically crafted to study relations, as conventional approaches do not work. This scientific petri dish has drawn the attention of criminologists, practitioners, and analysts interested in understanding how social structures enable crime and deviance.

A PRACTICAL EXAMPLE

At this point in the opening narrative, I turn to a practical example to illustrate why looking at criminal activity through a network perspective is useful to crime control efforts. The example comes from the Cincinnati Police Department and was executed under the leadership of Police Chief Eliot Isaac. The key personnel on the project were Captain Maris Herold, Lieutenant Matthew Hammer, MS; senior crime analyst Blake Christenson, MA; and research consultant Tamara Madensen, PhD, as well as six other police officers and an analyst. Known as P.I.V.O.T (Place-Based Investigations of Violent Offender Territories), this innovative project earned the distinction of being named the winner of the 2017 Herman Goldstein Award for Problem-Oriented Policing. As an introduction to the project, here is a quick explanation of problem-oriented policing.

Problem-Oriented Policing

Problem-oriented policing (POP) is extolled as an effective process through which to identify and learn about discrete issues of police business that pose problems, investigate new and effective strategies for resolving problems, and document the effectiveness of initiatives, so as to build on the body of policing knowledge (Goldstein 1979). POP efforts hinge on a four-step process—scanning, analysis, response, and assessment (Eck and Spellman 1987). First, the process begins by scanning for problems. Scanning involves tapping into varied information to identify those issues of concern to the community that pose a real threat to public safety, that the public expects law enforcement to intervene in and resolve, and that demonstrate evidence of recurring dangerous behaviors (Clarke and Eck 2003). Second, deep analysis is conducted to fully understand the nature of the problem, focusing on offenders, victims, and the locations (analysis often applies routine activities or crime pattern theory; see chapter 5 if you are not familiar with these theories). Third, once the problem is fully understood, investigative efforts seek to uncover and assess innovative responses. The goal is to develop a package of strategies to modify all circumstances that contribute to the problem. Community partners are key to the development and implementations of effective responses. Fourth, after some time has passed, the POP initiative is subjected to a detailed and rigorous evaluation, preferably by an independent researcher, to assess the impact and sustainability of the initiative as a whole and explore the possibility of crime displacement. Systematic analysis underscores all steps in the process, which means that crime and intelligence analysts are integral to POP initiatives. Analysis is so important, that the Center for Problem-Oriented Policing (http://www.popcenter.org/) offers two training modules and has published several books, guides, and toolkits to support analytic efforts. I highly recommend checking out the website. Now that you are familiar with the structure of POP projects we can turn to a discussion of Cincinnati's P.I.V.O.T. project.

P.I.V.O.T

Scanning crime incidents occurring in the City of Cincinnati in 2015, the City of Cincinnati and its police department made reducing gun-related violence a joint priority (Cincinnati Police Department 2017; Madensen et al. 2017). Why? Because violent crime was highly concentrated in specific locations. Analysis revealed that from 2012 to 2015, 25.7 percent of serious violent crimes (included were those that were compiled by the FBI and were considered part 1 offenses) and 42.6 percent of shootings involving a victim occurred at only twenty-three microlocations (1.4 percent of city areal coverage). Resolving crime problems at microlocations is an efficient use of resources because even though sites can be smaller than an address—that is, a single apartment or an area beside a dumpster—the reduction in crime can be significant. One of the most problematic locations was an open-air drug market. The center of this drug market was the intersection of Baltimore and McHenry Avenues.

Subjected to a 2007 focused-deterrence gang violence initiative (see box 1.1 for a discussion of focused deterrence), the site had a long history of violent, gang-related drug trafficking. The aim of the original project was to constrain the behavior of exceptionally violent gang members. Before community partners, social service providers, and multiple law enforcement agencies could leverage influence on specific individuals, analysts mapped the network of offenders and victims, identifying the most violent drug-involved gang members for targeted intervention. Despite the considerable effort invested in addressing gang violence, the 2007 initiative met with only

BOX 1.1. FOCUSED DETERRENCE

Focused-deterrence initiatives target high-risk offenders (e.g., prolific or violent criminal offenders). Originating as a problem-oriented policing initiative to address youth-gang gun violence in Boston in the late 1990s (Boston Police Department 1998), the focused-deterrence—or "pulling-levers"—strategy seeks to leverage the power of prosocial networks (Braga 2008; Kennedy 1997). Building lines of communication between criminal justice agencies, social services, community members, and prosocial family members, the web of relations surrounding high-risk offenders is tightened in order to constrain their illegal behavior. One step out of line, and a serious criminal case that was placed on hold becomes activated, putting the individual in prison, often for a significant length of time. Underpinning the strategy is the argument that if people believe they will be caught and punished, they will be discouraged from committing crimes. To influence offender perceptions of the risks of crime, individuals are subject to concentrated law enforcement attention—that is, identification and callout in a notification meeting, often in the presence of family or community members, who know what they get up to—and coordinated and strategic prosecution if the individuals fail to refrain from crime. Focused deterrence is premised on the idea that law enforcement action alone is insufficient to change behavior. Intensive police attention is coupled with social services, through direct interaction and careful monitoring, by the people who can reach the offender. Pulling-levers strategies are most frequently applied to problems of gang violence or gun violence, often in connection with drug markets (e.g., Braga, Apel, and Welsh 2013; Braga, Hureau, and Papachristos 2014). For more information see www.popcenter.org/responses/focused_deterrence/.

limited success. While gang-related homicides significantly declined after an effort to target the most prolific offenders, the reduction was not sustained and gun violence returned, spilling out from the initial intervention site, to an intense clustering on two intersecting street segments.

The limited success of the 2007 initiate is not surprising, if we look at it through the lens of social network theory. Targeted removal of central actors does not necessarily destroy a network; if the community of relations among remaining gang members and their support networks continue to function, the network will repair itself. Social networks have resiliency, if the mechanisms that enable interaction remain largely intact after the removal of key members. In this case, the drug market itself had resiliency because new relations formed and the gaps in the network were filled as individuals adjusted to the loss of key players. So in 2015, under the direction of Captain Herold, the department sought to bolster the focused-deterrence efforts by targeting the places that supported gang member interactions—now the place network was being investigated.

Rather than focusing on person-to-person interactions directly, the new network mapping strategy identified all of the places that were used by gang members and their associates. Locations were identified by combing through intelligence, surveillance, and informants' reports. The logic behind the new initiative was that the open-air drug market was only one of the locations frequented by offenders. There was a string of other places—places to stash weapons, hide drugs, or socialize—that were necessary to sustaining drug activities. Analysts mapped offender-activity spaces, and then looked to see where activity spaces overlapped. Places were important if several gang members were observed to frequent the location. (In chapter 5, we will discuss in more detail why mapping activity space is important in crime control efforts.) This exercise revealed that the prior initiative in 2007 failed because, even when specific members of the gang were removed, the remaining members were able to easily meet up and reorganize.

By mapping a network that links places frequented by multiple offenders, the P.I.V.O.T. team was able to better understand how (and where) the community of offenders interacted. Through this exercise, the team found private hangout places, meeting locations, staging locations, and a supply location. Some of the sites appearing in figure 1.2 include the following:

- Specific street-parking locations used for hand-to-car and car-to-car sales
- Three parcels of low-density rental homes used by offenders in various ways
- A commercial/retail property with an unsecured dumpster and an illegal vendor linked to drug activity
- Another drug sales site that generated only a few police calls for service
- A bighted, vacant, and abandoned location where offenders hung out

To generate the place-to-place network, analysts linked pairs of locations because different gang members or associates visited both sites repeatedly. Crime sites, shown in dark gray, could be instrumental in drug sales or violence; staging (white) and supply (light gray) locations were important in sustaining drug activity; and meeting locations (medium gray) contributed to sustaining social relations and criminal enterprise.

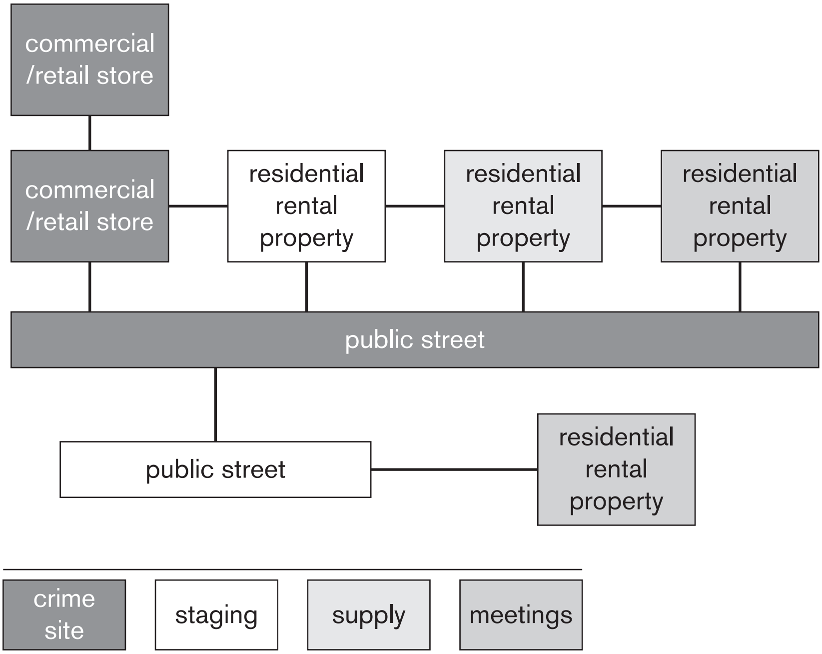


FIGURE 1.2. Place-to-place network of locations instrumental to the drug market. Adapted from Cincinnati Police Department (2017).

Since each site was used by offenders for different purposes, it is not surprising, that the sites had different place-based characteristics—parking spaces were unregulated, absentee owners were ineffective place managers, and vacant properties were unsecured (Madensen et al. 2017). A set of responses was necessary to address each site and its facilitating conditions—for example, permanent on-street parking restrictions to deflect vehicle drug sales; code enforcement against a property to remove illegal vending; and activation of unused public space, improvements in lighting, and demolishment of blighted property to improve natural surveillance and enhance the effect of directed patrols.

Assessing the project's effectiveness, analysts found a significant reduction in gun violence. In 2015, there were eighteen shooting victims. In 2016, after the place-focused efforts, there were three shootings, and by June 2017, there was only one. Moreover, the time between shootings increased from an average of 32 days preintervention to 130 days postintervention. Blight levels declined and overall violence in the area dropped significantly. Project leaders concluded that removing central places that fostered social interaction necessary to sustain the drug market was key to dismantling the violent drug operation and facilitating sustained reductions in violence (Madensen et al. 2017). The drug network was less able to react and rebuild after targeted actions because interaction sites were removed. (For more information about this project, see Cincinnati Police Department 2017; Madensen et al. 2017.)

Utility of Social Network Analysis

As you may have gathered from this innovative POP project, people often get excited about SNA because it provides new ways of thinking about crime problems, which may spark major advances in crime control. SNA-oriented thinkers seek to model the mechanisms through which interaction, and the social context it produces, fosters and constrains human behavior—legal or deviant. Not surprisingly, criminal network research already boasts an impressive array of studies, with the topics ranging from investigating the structure of criminal enterprise, transnational crime, illicit markets, and peer deviance to exploring the vulnerabilities of subversive groups (e.g., terrorism and gangs) and the systems exploited by corporate and cyber criminals. The widespread applicability of SNA in criminology leads some to argue for the coming of a networked criminology (Papachristos 2011).

While there is still some debate as to what to call this emerging field—illicit networks, dark networks, deviant networks, or criminal networks—I agree with Papachristos (2011) in that the term *networked criminology* is more inclusive and propose the following working definition: *networked criminology* refers to an interdisciplinary branch of social science that applies social network perspectives and analytics, in tandem with criminological theory, to study crime events, deviance, and criminality, as well as informal and formal responses to these problems. When we talk about the social structures underlying crime and deviance, there is usually an effort to differentiate our work from conventional SNA, which for the most part investigates legitimate and legal networks. So the debate about what to call this emerging field also includes a sidebar conversation about how to describe the class of networks under investigation. In respectful acknowledgment of Carlo Morselli's argument that referring to these structures as "illicit networks" is a general classification that could describe a range of social structures of interest to criminologists, this book is called *Understanding Criminal Networks,* rather than *Understanding Illicit Networks.* Why? Because if we ignore the age of the perpetrator, much of what preoccupies current applications in this emerging field is related to criminal behavior.[1] So while this discussion continues, I will use the term *criminal networks* in the broadest sense to refer to the range of social structures of interest to network criminologists.

In some ways, the current excitement about criminal network analysis is reminiscent of the revolution in place-based criminology triggered by widespread adoption of advances in geographic information sciences. For this reason, we must learn from prior mistakes and be careful. We cannot import SNA wholesale. In order to use SNA in our investigation of crime and deviance, we must think carefully about the potential differences between legitimate networks and criminal networks. Criminal networks operate in a hostile environment. Members face social censure and punishment if their associations and behavior become known; they are under attack from various criminal justice agencies; crime groups are not immune from internal struggles caused by warring factions; and they operate within a competitive arena, facing hostilities from other criminally minded groups and individuals. And yet, much of their activity links them to legitimate behavior, so group boundaries are fuzzy. Carlo Morselli (2009, 8), one of the foremost network criminologists, encourages us to remember that "criminal networks are not simply social networks operating in a criminal context. The covert settings that surround them call for specific interactions and relational features within and beyond the network." These operational demands shape network structure. And in doing so, the nature of what we aim to investigate, in some ways, restricts which theory and methods we can use to conduct criminal network research.

OBJECTIVE OF THIS BOOK

Crossover between disciplines is complicated. One must quickly get up to speed with a large body of research, and this requires considerable investment in learning theories, methods, and analytic techniques, in addition to becoming familiar with the idiosyncrasies of new software, of funding opportunities, and of the personalities controlling the publication process—that is, the personal quirks and demands of potential reviewers, editors, and

1. Readers should note that this statement is not to be construed as a rejection of the large body of work illustrating that illicit and legitimate activity overlaps. For example, coconspirators involved in methamphetamine production may also be relatives who are investing in a legal home-remodeling business. For this reason, determining the boundaries of dark and legitimate networks is sometimes difficult. For the present, let us just visualize these two types of networks as separate and distinct. I will get into the fuzzy boundary problem and the importance of considering multiplex relationships later.

publishers. And as noted above, it is necessary to consider how the nature of criminal activity differs from legitimate network behavior, so as to apply appropriate methodologies. In short, there is a lot to learn in a short period of time. Without the guidance of a seasoned scholar, a novice may be quickly overwhelmed. SNA researchers specializing in networked criminology, however, are in short supply, and you may not have someone with experience in the field in an office down the hall from you. This brings me to the impetus for this book—to generate a guide that will help interested criminal justice academics, or practitioners, analysts, or students, dive into criminal network analysis.

I kept two types of readers in mind when writing this book. (1) The career scholar or practitioner who is interested in jumping into something new but has little time to invest in a long, labored approach to learning the basics; and (2) the eager new scholar or analyst who wants to get into SNA but faces pressure to produce results quickly in order to establish credibility as a researcher or to satisfy the boss. In both situations, I expect the reader is looking for a straightforward and brief introduction to the field. I kept the methods and statistical chapters to the basics because excellent books already cover these topics in detail. Moreover, this book was written to be read wherever you do your best thinking, whether it is behind a locked door until someone else needs the restroom, sitting beside the grill with a glass of your favorite beverage while you make sure that dinner does not burn, or commuting to work—provided of course that someone else is operating the vehicle, bus, train, or subway. In light of these considerations, the book needs to be short and can't cover any of the topics in great depth. To address this limitation, each chapter suggests resources, including videos when possible.

ORGANIZATION OF THE BOOK

Covering just the essentials, I organized the chapters around three themes: (1) the theoretical basis of networked criminology, (2) methodological issues associated with studying criminal networks plus useful analytic tools, and (3) tips for producing professional products in this new field. Below I describe each chapter within the three sections.

Section 1: Theoretical Ties

In some respects, chapter 2 stands alone. One of the most difficult hurdles to diving into a new subject is to wrap your mind around the field's language and standard methodological concepts. While we get into the mechanics of criminal network research in detail later, chapter 2 works through a fictional example to provide an overview of what you will learn in this book.

Chapter 3 is a brief synopsis of axioms, central theories, and key concepts used in social network research. This discussion lays the foundation for readers' understanding of the material provided in subsequent chapters. I supply a series of graphics to illustrate ideas, and a list of "must reads" points out important reference books and articles by notable SNA scholars.

Organized around tables of congruence, chapters 4 and 5 illustrate the theoretical correspondence between fundamental SNA principles (introduced in the prior chapter) and criminological theory. A discussion of select theories of criminality appears in chapter 4, whereas chapter 5 covers theories of crime. The aim of both chapters is to show readers how SNA can help to develop our explanations of crime and deviance. I hope this conversation will ignite some ideas for future research.

While the application of SNA to crime is relatively new, a number of active and notable scholars have generated important streams of research. Anyone embarking on research in these areas must be aware of this foundational work. Thus, chapter 6 outlines research streams in the emerging field of networked criminology, identifies key scholars, and lists notable publications.

Section 2: Designing Research

Four chapters make up the section on designing research. The first two chapters in this section should be read in succession. Chapter 7 covers the basics of gathering network-oriented data, reviews the strengths of different data collection techniques, and compares the utility of commonly used information sources. Chapter 8 opens with a discussion of data integration and entity resolution, followed by an introduction to assessing the effects of missing data and sensitivity testing to ensure findings are robust. Several graphics illustrate key concepts.

Without getting into too much detail, chapter 9 reviews some basic descriptive statistics that are widely used to report on whole networks, clustering, actor positions, and egocentric networks. This chapter includes a table defining key statistics, with notes about when to use them. In addition, I provide a list of resources to learn more.

Chapter 10 supplies a comprehensive discussion about transitivity, an explanation of techniques used to identify subgroups, and a brief overview of some advanced analytic options designed for testing hypotheses. Pay attention to the boxes linking analytic strategies with software and training resources.

Section 3: Publishing

The final section has just one very important chapter intended to support the readers' productivity. Because SNA is relatively new to crime science, most people do not understand what they are reading. Within this somewhat hostile setting, academics, practitioners, and analysts must generate professional products that describe network research. This chapter provides tips on publishing in peer-review outlets, generating graphics, and making presentations about networks. I hope that this advice will reduce the number of rotten tomatoes aimed at you. Chapter 11 also introduces drivers of this new field—the professional associations and organizations dedicated to advancing SNA scholarship, particularly in relation to criminal networks.

Readers who are new to the field should continue reading. If you have some SNA experience, you can skip chapter 2 and continue with the text at your leisure. Otherwise, it is best to read the second chapter immediately. The crash course will ease your entry into criminal network research by demystifying SNA jargon.

2. Demystifying Social Network Analysis

> Social network analysis reveals what is hidden in plain sight.
>
> KADUSHIN, *Understanding Social Networks*

At this point, you may be wondering, what is social network analysis (SNA) and where did it come from? Rest assured, SNA is a bona fide discipline and not some one-off analytic technique. It has its own theories, methods, and statistical applications for studying relationships, as well as the larger social structure—economic, political, and social environment—emerging from social interactions. While the theoretical roots of SNA date to the mid to late 1800s with the work of Georg Simmel and Emile Durkheim, the field of SNA emerged in the 1930s, with the introduction of analytical methods based on a branch of discrete mathematics called graph theory. In recent years, there has been a significant growth in statistical applications available to study social structures (i.e., exponential random graph modeling, stochastic actor-based models, and multilevel modeling) and a proliferation in software applications (I once saw a list of fifty different programs). Widespread adoption, in fields as divergent as communications and sociology to complex systems and theoretical physics, partially accounts for its developmental trajectory. The drawback to such a rapid and widespread adoption is that the field is still evolving and developing a common lexicon.

To ease your emersion into the field, this chapter explains some SNA jargon. Bold or italic font identifies key terms. We begin with a short set-up for a project, after which I review some data options. Next, the conversation turns to units of analysis and how methodological decisions influence analysis. The chapter concludes with a brief notation about analytic options.

EXAMPLE PROJECT

One of the most fundamental SNA concepts is the notion of central positioning. But centrality does not necessarily mean that someone sits in the middle of a group of people, nor is there only one type of centrality. Rather, a **highly central actor** is an entity that, relative to others, has a positional advantage over others. What that positional advantage is depends on the type of positioning under investigation, the nature of information captured in the network, and the analytic choices made to study the relative positioning of actors.

For instance, let's suppose that you are interested in mapping the activities of a gang involved in illicit drug distribution—perhaps you intend to replicate P.I.V.O.T. or the original focus-deterrence projected implemented by the Cincinnati Police Department. On the basis of a working hypothesis that within your jurisdiction only a few gang members are critical to the flow of drugs, you set out to investigate the structure of illicit drug operations. More specifically, you want to know which offender in the distribution network acts as a gatekeeper, bridging different subsets or clusters of actors, and by doing so, controls how product moves between groups. The reason is that if the drug of interest is heroin, which is not sourced locally, a critical junction in the flow of drugs might be someone who smuggles the drugs across the border or a midlevel supplier providing a purer form of heroin. Since the identity of this theorized person is unknown, you want to map drug operations using the bits of intelligence you have and predict who might act in this capacity.

Identifying central individuals for additional investigation, use as informants, or targeted enforcement is a common practice for criminal network analysts that has implications for crime control (Hashimi and Bouchard 2017). As reported by one analyst interviewed by Burcher and Whelan (2018, 283), "What I'm usually drawn to, is to go, who's the bridging person between this group of offenders and that group of offenders." The analyst continues, suggesting that network applications "make it clearer who are the more powerful players that might not be obvious" when dealing with an assemblage of snippets of intelligence wherein nothing stands out (283). In an effort to strengthen the use of SNA metrics for this purpose, several crime scientists are working to develop rigorous but user friendly analytics that prioritize targets by combining relative position in the network with individual attributes and access to resources (e.g., Bright et al. 2014; Hashimi and Bouchard 2017).

As you will discover in chapters 3 through 6, there is strong support for the desirability of identifying central actors who have bridging relations with distinct subsets of others; identifying bridging relations can be used to weed out the central actors. And an individual positioned in this way could exert an even greater influence on the network if there are no indirect links that could be used to sidestep or work around using the central bridge—that is, they are the only ones with ties between distinct groups of people, so they are the only channels through which product can move from production to retail distribution.

Finding relevant and complete information about a group's clandestine activities is not easy, forcing most practitioners to extract and compile information from many sources. While chapters 7 and 8 cover the logistics of data acquisition in detail, at this stage there are a few points worth noting. First, no matter how comprehensive your information-retrieval process is, people, and the links among them, might be missed. Second, irrelevant individuals and their connections might be included. And third, people often have multiple identifiers (e.g., several nicknames). For these reasons, always treat the mapped network as a sample and use caution when reporting results. Also, test the robustness of your analysis by repeating it on networks generated using different inclusion criteria—the wisdom of this point will become apparent soon. Meanwhile, it is time to review some key methodological concepts relating to units of analysis.

METHODS

Units

The most basic unit of a social network is a **dyad,** composed of two actors and a relation between them. The term *actor* is often replaced with the more neutral term *node.* Nodes can represent any social agent—individuals, groups, organizations, nations, and etcetera. Of note, nodes can also reflect stages in a complex system or any other agent representing the outcome of a social or interactive process. Returning to our hypothetical project, we might study the social structure of gang activity using gang members as nodes connected to each other in dyads by the existence of a coarrest. Since understanding the versatility of what a node can represent is key to learning how to use SNA, a few other illustrations follow.

1. In an investigation mapping activity space such as was done in the P.I.V.O.T. project, nodes could be *addresses* visited by people.
2. Studies of a crime syndicate might look at *members* (nodes) who jointly attend meetings.
3. An examination of terrorist activity might examine how *terror group cells* (nodes) are linked by the transmission of information or materials.

4. Money laundering might be tracked by identifying *accounts* (nodes) involved in transactions.

The connection between nodes, called a ***tie, link, edge*** or ***arc*** (if there is a direction to the relation), can represent a somewhat durable association (e.g., a family relation) or dynamic property (e.g., a phone call passing information). While in our project we were able to record a tie between gang members who were coarrested, in this example, the relation between nodes is not directed, as there is no basis on which to determine who initiated the partnership. If instead we were able to connect people using communications, such as recording both the originators and recipients of telephone calls, then we would describe the relations as arcs.

Returning to the other four examples of nodes described above, we can identify the relations used to form dyads. (1) Addresses are linked when the same person *visits both places* (ties); (2) members of a crime syndicate are connected if they *attended the same meeting* (ties); (3) terror cells are connected through the *transmission of information or materials* (arcs); and (4) accounts are linked through *financial transactions* (arcs). Notice how the connections for the latter networks are referred to as arcs. Since the origin and recipient of transmissions or financial transactions can be determined, it is possible to capture the directionality of the interaction. Our conversation about relations needs to be put on hold for a few minutes while I explain other units of analysis.

Analysis may also focus on a slightly more complex configuration of nodes, that being the **triad,** or sets of three actors and the relations or potential connectivity among them. As we discuss in chapter 9, dyads and triads are the basis for most statistics used to describe a node's position within the network, identify subgroups, or characterize the network as a whole. As you may have guessed, triad-based metrics are higher-order measures, often used in comparison with dyad-level metrics to reveal important details about the network's structure. Since the statistics covered in this book have been calculated from counts of either dyads (relational patterns among pairs of nodes) or triads (relational patterns found among sets of three nodes), we now turn to an image so that you have an opportunity to visualize dyads and triads.

Figure 2.1 illustrates relational information of a fictitious gang. Circles represent gang members, and the lines connecting the circles indicate an association identified by examining all gang intelligence gathered during one year (panel A) or identified from co-offending as revealed by arrest records (panel B). This gang includes seventeen people. Of interest, person 17 is thought to be part of the gang but is not linked directly to other

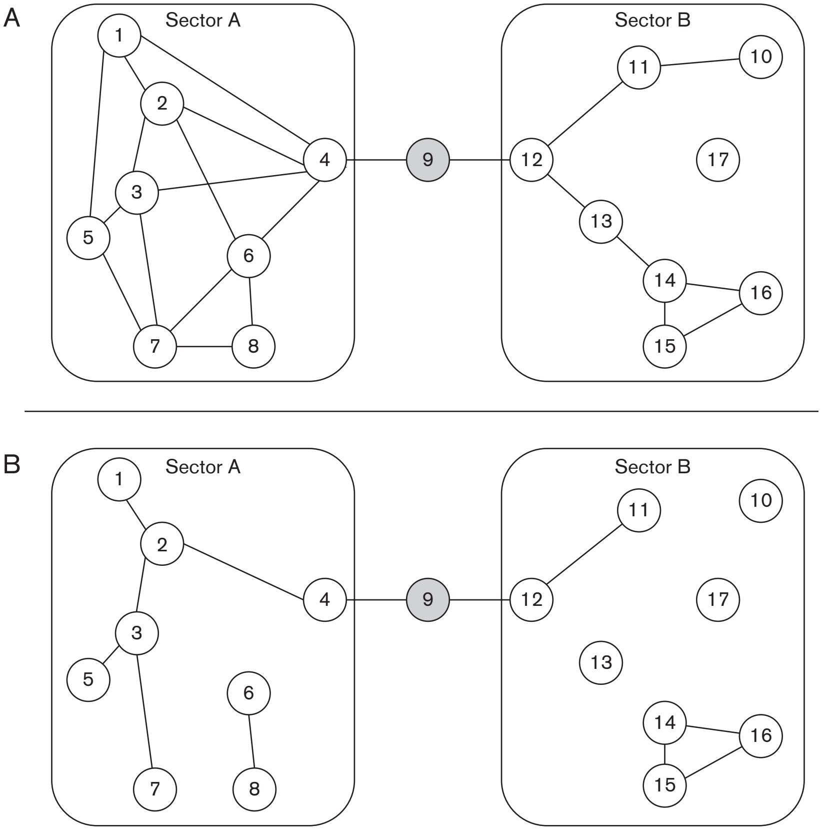


FIGURE 2.1. Illustration of units of analysis. *A*, network of gang associations using all intelligence generated in one year; *B*, network of co-offending for one year based on arrest records.

members in the group by recent information collected for the year under observation. For the moment, we will concentrate on panel A. Please take a good look person 9's connections.

From a dyadic perspective, person 9 is involved in two relations (two dyads), one with person 4 and one with person 12. Turn your attention to person 4. Person 4, who is part of five dyads, has a relationship with persons 1, 2, 3, 6, and 9. We could continue to explore the network individual by individual, looking at patterns of direct connectivity using dyads. But what about indirectly connected pairs? Are persons 9 and 1 connected? Not directly, but this pair is connected indirectly through a third party. This is one of the ways triads are useful. If we consider three people—persons 4, 9, and 12—this set does not form a complete triad because there is no direct connection between persons 4 and 12—a relation does not exist for this pair of actors. However, persons 4 and 12 can influence each other by sending information through their mutual connection. Refocusing on sets of three people, we can examine all of the indirect channels of influence among gang members.

Another way to examine patterns of connectivity is to compare **egocentric networks.** If we consider only person 9's direct relations and the ties among the associates of person 9, we are examining an egocentric network. The egocentric network is the local social neighborhood encapsulating a person—the ego. Egocentric networks are not generally considered in isolation; rather, the analysis is more informative when we compare people. For instance, compare person 12's local social world to that of person 1. While both gang members have only three direct ties, person 1 is part of a more cohesive set of relations, and person 12 has a loosely structured personal network.

It is also common to examine the relations among a full set of people. By including the relationships that each gang member has with others, we generate the **whole network.** Whole networks include everyone designated as part of the group and the relations among them. Whole network research allows us to study the nesting of individuals within the group and the structure of subgroups, as well as the group structure as a whole. It is also possible to extend this meso-level analysis to a macrolevel investigation by including connections that group members have to individuals in other groups. Here we are interested in the *community* of actors. A community is not a single group; instead, it is a set of interconnected groups. For example, it is possible to learn about interconnectivity among crime-involved groups by including all of the intelligence about interpersonal relations for all individuals involved in drug trafficking within an entire jurisdiction. Aggregating intelligence will expose individuals with ties to different groups; these individuals generate the interconnectivity among groups that results in the formation of the larger drug-trafficking community. We will see an example of this in chapter 10.

Returning to the drug-trafficking group depicted in figure 2.1, panel A, we see two sectors, denoted by rectangles. Each sector exhibits distinctly different substructures. Sector A exhibits a high level of interconnectivity, and many triads are visible. This **subgroup** is highly **cohesive** in that most people connect directly to each other. Sector B includes only a small cohesive subgroup involving persons 14, 15, and 16. All others join in **chainlike structures,** and, as noted above, person 17, who has no observed associations, is an **isolate.** Either information about person 17 is missing or that person is no longer an active member. Clearly, the data used to map the network reveal information about group structure.

Focus

When gathering information about social relations, it is important to predetermine whether you are interested in studying the **architecture** of the group or the **flow** of something through the network. In figure 2.1, panels A and B exhibit all known associations using two different strategies. Underlying the construction of panel A is the need to map all potential relations of gang members; it is a *static representation* of relations that may reveal the architecture of the group's structure. Contrast this image with panel B. Co-offending activity is more fluid and may represent how criminal partnerships ebb and flow. People engage in different activities over time with different individuals. If we were to compare co-offending activity for different observation periods, we may discover both stable and temporary associations: in SNA, this type of research is referred to as *dynamic analysis*. Another data source commonly used to examine crime flows is communications—surveillance of meetings, wiretap data, or electronic messaging. If we had flow data for an extended time, we would assume that the full architecture of the group might be revealed (Borgatti and Lopez-Kidwell 2011).

Returning to figure 2.1, contrast panel B with panel A. Using co-offending information for one year provides a different picture of the relations among gang members compared with the network depicting relational information generated by a year of intelligence. Consider person 9. The intelligence data (panel A) may lead to the conclusion that person 9 is a critical gatekeeper bridging subgroups, but recent co-offending activity (panel B) suggests something different—that person 9 is no more important than most others.

Capturing information to map both the architecture of relations and current flow of activity is useful. The graph showing flow tells us something about recent activity, and the mapped architecture of the network may illustrate where the flow would diverge or be displaced if a channel were blocked. A growing body of literature supports the use of SNA for this purpose. For instance, Morselli and Roy (2008) illustrated the flexibility of two automobile theft rings, assessing how targeted removal of key actors reduces the scope of alternatives for continued operations; Duijn, Kashirin, and Sloot (2014) explored the effects of several disruption strategies and recovery mechanisms with simulation models to understand how to better disrupt the Dutch marijuana industry without inadvertently strengthening the criminal network; and, taking a systems perspective, Bichler, Bush, and Malm (2015) demonstrated how network models can help us anticipate the effects of crime control policy on transnational illicit markets.

As noted previously, you should be aware of one of the biggest challenges to using SNA to study criminal networks—*incomplete network data*. Many of the metrics and analytics we apply assume that there is no missing information about nodes or connections among nodes. Descriptions of network structures are biased if there is too much missing data, but there is some good news. First, network criminologists, who have begun to do sensitivity testing, are finding some positive results (we discuss this in chapter 8). Additionally, the inherent instability of some metrics offers tactical advantages. For example, mapping networks observed at six-month intervals, Calderoni (2015) showed how shifting position within a network may identify emerging leaders within mafia-style criminal enterprise groups, and Bichler, Lim, and Larin (2017) demonstrated that monitoring the network position of individuals connected to a serial-murder investigation may help avoid investigative failure and aid in the identification of suspects.

It is also important to keep in mind that mapped networks are representations of social structures. As alluded to above, they are as good as the information we have to map them. Moreover, decisions made at the onset of a project dictate how we interpret the results: nodes and the relations among them can represent different phenomena, which will affect what we can say about the network we are investigating. This returns our conversation to the issue of relations.

Relations

One reason researchers from many disciplines use SNA is the flexibility of what nodes and relations can represent. As we learned above, nodes could denote any social entity. Similarly, relations among actors can also take many forms. A fleeting encounter exemplifies this point. Several years ago I bumped into a biologist while rock climbing and camping in the desert at Joshua Tree National Park, California. As we stood chatting by the campfire, I discovered that she applied the same social network techniques I used to study criminal networks to the study of genetics. In her research, the node is a gene, and dyads indicate heredity.

Within criminological research, relations often represent the following:

- **Co-activity** (e.g., a conversation, a coarrest, attendance at an event, or a violent attack)
- **Social connection** (e.g., friendship, kinship, coemployment, business association, co-ownership, comembership, enemy, or competitor)
- **Emotion** (e.g., trust, like or dislike, fear, or avoidance)
- **Role-based associations** (e.g., chain of command, supervision, or mentorship)

- **Product or information flow** (e.g., movement of illicit goods, money transfers, social media posts and reposts, or travel between places)
- **Intersection** (e.g., hyperlinks among webpages, citations of scholarly work, or linked processes in a complex system)

It is also possible to **quantify relations,** so that they represent the *strength* of an association or a *value*. To return to our fictitious gang, figure 2.2, panel A, illustrates what a valued network would look like. Thicker lines generally reflect high values or stronger relations, and thin lines reflect low values or weak ties. For instance, while person 4 has five connections to others, these relations are not all equal—three are strong and two are weak. Examining quantified relations provides another way of thinking about the structure of the gang.

There are many different ways to quantify a relation. Some relations are inherently more durable—for example, a familial relation is likely to be more durable than an acquaintance. As a general rule, stronger relations are generally harder to disrupt. Also, dyads can share different types of relations—for instance, two members of a gang (group connection) are cousins (familial relation) and they own a business together (legitimate business association). In this scenario, the pair has three distinct relations with each other. This relation is **multiplex.** *Multiplex relations* can represent a stronger bond between the two people than if they were simply part of the same gang. *Multiplex networks* include different types of connections among people. Consider person 4, shaded in gray in figure 2.2, panel A. If this graphic illustrates multiplex relations, we would expect that the connections among persons 4, 2, and 3 would remain intact, even if we were to arrest one of them.

Relations can also be valued in other ways, such as by the frequency of contacts, volume of materials transferred, or the degree of involvement using scales. Similar to bond strength, more frequent contact or relations with greater involvement are harder to disrupt. If figure 2.2, panel A maps cell phone communications, the triad with persons 14, 15, and 16 includes one relation with regular interaction (thickest line)—perhaps daily contact—and two less frequent communication channels, which may be relatively easy to interrupt.

Panel B of figure 2.2 depicts another level of detail about relations that we could add to a network—the **direction** of a relationship. Now the channels between gang members depict a *hierarchy* of sorts. Again, depending on what constitutes a relation, this could represent command structure,

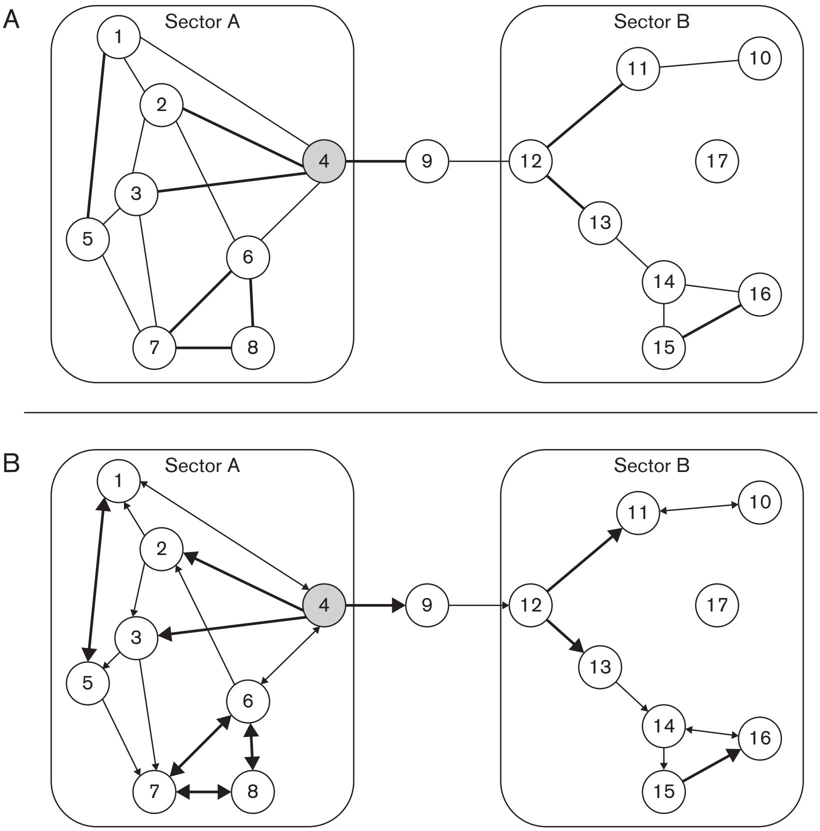


FIGURE 2.2. Illustration of the strength and directionality of relations. *A*, valued network of gang relations; *B*, valued and directed network of gang relations.

trust, or flow of information, among other things. Arrowheads indicate the directionality of relations, such that the arrow originates with the sender and terminates with the receiver. Double-headed arrows specify a mutual exchange. Considering both the strength and directionality of relations within this gang, a different picture emerges from what we saw in figure 2.1—the texture of social relations comes into focus, and this information may better serve drug interdiction efforts. Person 4 (shaded in gray) may control the flow of information to sector B through a strong tie to person 9, and the two weak but mutual relations with persons 1 and 6 provide access to two other strongly connected subsets in sector A. Person 4, it turns out, may be more important than we originally thought.

ANALYTIC OPTIONS

Despite how useful figures 2.1 and 2.2 are, it is important to keep in mind that visualizations of networks can be misleading. Relative positioning and the detail of substructures is difficult to assess and virtually impossible to compare between large networks without using a standardized metric. In most situations, important structural features can be obscured or overlooked. Fortunately, there is an extensive array of analytic options at our disposal. Without giving away too much of chapters 9 and 10, I want to illustrate how calculating a simple descriptive statistic—betweenness centrality—can support an investigation. Betweenness centrality calculates the extent to which each node in a network is likely to sit between pairs of others. High scores can be interpreted as nodes situated in a position, wherein they could act as a gatekeeper bridging subgroups in the network.

For instance, we might wonder if persons 4, 9, and 12 are equals (or potential rivals) in controlling drug activity. Before examining the scores reported in box 2.1, take another look at the position of these people (figure 2.1). What would you guess from looking at the graphic? Now examine the scores reported in bold text. The scores reported are standardized (referred to as being normalized in SNA) to permit comparison between networks of different sizes. Higher values indicate that the person sits directly on the shortest paths between more pairs of people. Scores are reported for the network mapping all intelligence, as well as for the more fluid co-offending network. See how the scores vary for each person depending on the source of information used to graph relations.

Figure 2.1 may have led you to predict that gang members 4, 9, and 12 were similarly positioned, but as you can see, they are not all equal—4 has a substantial advantage over the others (panel A). Reviewing the list below, person 4 has a higher betweenness score, irrespective of the data used to map the network. Looking down the column that reports scores for the co-offending network, we find that even though gang member 2 is the most significant bridge, their potential influence is three times less than what person 4 is projected to have when we consider all intelligence on gang relations. Network statistics reveal differences in the relative positioning of members of the gang, reaffirming something you learned from the quote that opens this chapter: “Social network analysis reveals what is hidden in plain sight” (Kadushin 2012, 6).

Equipped with this basic introduction to the study of networks, along with some key terminology, you can investigate the theoretical basis of social network research. The next chapter reviews what it means to be

BOX 2.1. STANDARDIZED BETWEENNESS SCORES FOR ACTORS IN EACH NETWORK (FIG. 2.1)

Node ID Number	*Intelligence Network*	*Co-offending Network*
	Panel A	*Panel B*
1	1.50	0
2	**1.39**	**10.83**
3	**5.59**	**8.33**
4	**31.32**	**6.25**
5	1.81	0
6	8.75	0
7	3.26	0
8	0	0
9	**26.67**	**5**
10	0	0
11	4.17	0
12	**22.50**	**3.33**
13	12.50	0
14	10	0
15	0.83	0
16	0	0
17	0	0

NOTE: Normalized betweenness centrality scores are reported for each panel.

embedded within a network and how this both constrains and enables social opportunities. Chapter 3 also covers several prominent theories—strength of weak ties, social capital, small worlds and scale-free networks, and hyperdyadic spread. By the end of the next chapter, you will understand the theoretical importance of central positioning within the ego network, subgroup, and whole network.

3. Social Network Theory 101

> Most of us are already aware of the direct effect we have on our friends and family; our actions can make them happy or sad, healthy or sick, even rich or poor. But we rarely consider that everything we think, feel, do, or say can spread far beyond the people we know. Conversely, our friends and family serve as conduits for us to be influenced by hundreds or even thousands of other people. In a kind of social chain reaction, we can be deeply affected by events we do not witness that happen to people we do not know. . . . As part of a social network, we transcend ourselves, for good or ill, and become part of something much larger. We are connected.
>
> CHRISTAKIS AND FOWLER, *Connected*

Social networks generate the living fabric of a society. With each person embedded in a unique constellation of relations, investigating human behavior requires considering how direct and indirect influences—for example, the friend-of-a-friend-of-a-friend—shape decision making, control access to opportunities, and generate belief systems. As with all disciplines, network explanations of social phenomena begin with a common set of assumptions on which theories are built. Chapter 3 provides a general understanding of discipline axioms and essential theories used in social network research. I selected the most commonly discussed ideas that have relevance for crime applications. My intention is to lay a foundation that readers can build on. A series of graphics illustrates concepts, and a list of "must reads" identifies reference books and articles written by notable social network analysis (SNA) scholars.

ASSUMPTIONS OF SOCIAL NETWORK ANALYSIS

SNA is an interdisciplinary, multidisciplinary field of study concerned with exploring the regularities and patterns of social relations and their effects, both immediate and distal, on behavior, perceptions, beliefs, and decisions (Christakis and Fowler 2009; Knoke and Yang 2008; Wasserman and Faust 1994). Social networks are made up of a set of actors and the relations among them. As we learned in chapter 2, *actors* is a general term that refers to any social unit of interest, such as individuals, groups, organizations, websites, or nation-states. (I use actors in this chapter, rather than nodes, as it makes the discussion of theoretical concepts more readable.) Relations can refer to a range of mechanisms, emotions, or behaviors that link social units—for example, coinvolvement in a conspiracy, movement of contraband in a smuggling operation, co-offending activity, or rivalries between gangs. Underpinning the analytics used to map and study social networks is graph theory.

Graph theory is a branch of discrete mathematics used to model many types of processes and relations among pairs of objects in biological, informational, physical, or social systems. In light of graph theory's mathematical roots, readers should not be surprised to learn that social units are often referred to as a set of vertices representing a group of actors and a set of edges (or arcs, if directed) defining the relations among them. Relational information pertaining to a set of actors is recorded in an adjacency or incidence matrix (see box 3.1). The information captured in a matrix format can be illustrated in a sociogram (map of the network) by representing vertices with a nodal symbol, often a circle, and edges as lines connecting the nodes. When early SNA scholars applied graph theory to the study of social systems, several assumptions were made about people that continue to underlie research in this area (see, for example, Wasserman and Faust 1994).

1. Interdependencies among actors explain behavior better than actor attributes

In other words, network forces impel behavior more than internal motivations. The first assumption has implications for how we investigate the shape of a network that influences actors, with **network shape** referring to the structure and topography that emerge from patterns of connections. Wellman (1988, 20) explains, "Analyses focus on the relations between units, instead of trying to sort units into categories defined by the inner attributes (or essences) of these units." Such studies include information about the characteristics of actors as covariates, but these variables are used to understand the generation or termination of relations (or the formation of a larger social structure) that influences actor behavior. For example, a conventional analysis of co-offending might investigate whether co-offending is more prevalent among younger than older offenders. In this example, age is the focal covariate thought to account for the level of co-offending observed. Alternatively, using an SNA perspective, co-offending pairs might be examined to determine whether offenders exhibit a tendency to select co-offenders of the same age group. The focus is on patterns of selection, not whether co-offending occurs. Age is still important to the inquiry, but it aids our understanding of the structure and topography of the network.

BOX 3.1. NETWORK INFORMATION IN A MATRIX

Underlying all visual depictions of networks is a matrix that enables visualization of the network, as well as manipulation of data, the calculation of metrics, and the running of various multivariate analyses. While user friendly software makes it possible to avoid looking at the matrix, do not forget that it is always behind the scenes. To understand analytic techniques, you must consider what functions are being applied to the matrix. Two forms of matrices exist—an adjacency matrix and an affiliation matrix.

An **adjacency matrix** is square, with the actors heading up each column, starting with the second, as well as making up the line captions listed on the left side of each row (table 1). This matrix, used to represent one-mode networks, can capture information for directed or undirected relations. A one-mode network includes only one type of social unit—for example, people. So a network of gang members would record connections among people and not links among people, gangs, and neighborhoods. Cells record whether a connection exists between each possible pair of actors. In the example below, I used a dichotomous coding scheme, where 0 indicates the absence of a tie and 1 indicates the presence of a tie. If the matrix is undirected, it will be symmetric—meaning that if Bruce is connected to Charlie (looking across the row for Bruce), Charlie will also be connected to Bruce (looking across the row for Charlie). A value of 1 will appear in each cell.

Notice that the row and column totals are the same, which is indicative of a symmetric network. The totals report the number of ties each person has. Totals are included for explanatory purposes only. They would not appear in the matrix that the software uses in its calculations. If people are not able to connect to themselves (e.g., you do not count sending email to yourself from a different account when mapping email communications), then the diagonal is empty (or contains 0s), as you see in the example. When self-loops or recursive ties are possible, values appear in the diagonal.

Symmetric adjacency matrix

	Alex	*Bruce*	*Charlie*	*Danielle*	*Erica*	*Row Totals*
Alex	—	1	0	0	0	1
Bruce	1	—	1	0	0	2
Charlie	0	1	—	1	0	2
Danielle	0	0	1	—	1	2
Erica	0	0	0	1	—	1
Column Totals	1	2	2	2	1	

Affiliation or incidence matrix

	CFS1	*CFS2*	*CFS3*	*CFS4*	*CFS5*	*Row Totals*
Alex	1	1	0	0	1	3
Bruce	1	0	1	0	1	3
Charlie	0	0	0	1	1	2
Danielle	0	1	1	0	1	3
Erica	0	1	0	0	1	2
Column Totals	2	3	2	1	5	

Directed networks do not exhibit symmetry in row and column totals. If the relations are directed, they are generally asymmetric—that is, just because you find a relation between Alex and Bruce, it does not mean that Bruce will extend a tie to Alex. Asymmetric networks are unbalanced owing to the lack of reciprocated ties—the row and column totals are not be the same. Since direction is usually recorded from rows to columns, row totals report outgoing connections (focal person or "ego" to their contacts or "alters"), and column totals report incoming connections (alter from ego).

An **affiliation or incidence matrix** is used to capture information representing a two-mode network. This rectangular matrix lists one type of social entity—for example, people in the rows. The second type of social unit is listed across the columns—for example, the events they attend. (The P.I.V.O.T. project used something like a two-mode network when initially connecting people to hangout locations.) While SNA scholars might refer to these tables as affiliation matrices, dark-network researchers tend to investigate phenomena like co-offending or coinvolvement in meetings revealed by surveillance data; thus, the matrix is often described as an incidence matrix. For example, I used this type of matrix to map officer responses to calls for service (CFS). In this network, the rows list officers and the columns list each call for service. Notice in the example matrix that there are different row and column totals. Row totals tell us how many CFSs each officer attended, and column totals tell us how many officers attended each call. Diagonals are not a relevant feature of an incidence matrix.

2. The constellation of relations among actors generates differential influence on each social unit in the network

The pattern of relations with others shapes our behavior, for it is through these connections that we access information; engage in activities or complete tasks; and receive reaffirmation, emotional support, or censoring of our behavior. In short, the pattern of ties provides access to opportunities and, at the same time, potentially constrains our behavior. Because the web of relations exerting influence on each person in the network is likely to vary, so too will aspects of their lives—success, health, and emotional well-being. The network of ties surrounding each person varies for several reasons, one of the most tangible being multiplexity.

Multiplexity refers to the notion that actors are bound by different types of relational ties. The implication is that each type of relational tie could generate a unique network configuration if mapped by itself. Aggregating each unique network will further differentiate the web of ties surrounding each individual. To understand this idea, compare your familial relations with those at work or school. The network-mapping conversations with family members will not necessarily be the same as the network of people you worked with last month. Our personal lives do not necessarily include our professional associations, although it is feasible that some individuals might appear in both networks, if, for example, your sister is your business partner or if you go to school with your cousin. The implication for research is that each person is likely to be part of multiple, potentially overlapping networks. Each type of relation produces a unique network, and when different types of relations are mapped together, a more comprehensive picture of social interactions will emerge.

Interestingly, the more that different networks overlap, the more deeply interwoven our lives are with those we interact with. For instance, imagine that you have only ever worked in a family business, all of your closest relatives live within a couple of blocks, and you all attend the same church and bowling league. Mapping the network of interactions for each facet of your life (work contacts, church associates, etc.) will generate heavily overlapping networks, compared with a situation wherein work, family, religion, and socializing involve unique sets of people. So the value added to network comprehensiveness depends on relational patterns. In the first scenario, in which the family does everything together, multiplexity will add little value, whereas in the latter example, a multiplex mapping protocol will reveal a great deal about social connectivity.

Two other points are important. First, some phenomena are not relationship specific and thus may cut across different types of networks. To illustrate, think about mapping the diffusion of information about crime victimization. If your credit card information were skimmed in a novel way, you might issue identical warnings to your family, friends, classmates, and coworkers. The boundaries of each type of social network are not important. Alternatively, information about a breach of security protocol at work may be relevant only to coworkers. Second, it is also important to recall that relations can be directed, meaning the channels among actors may not open both ways, which further complicates the flow of network influence. For example, you may turn to someone for advice about cyber security, but you do not advise them. Complex patterns can emerge that interweave different networks, and these patterns may not be fully reciprocal. These complex patterns constitute network topography.

Network topography influences the degree of social freedom or constraint imposed on individuals, and the constellation of patterns is likely to be different for different types of ties. For instance, panel A of figure 3.1 illustrates a hypothetical kinship network for actor 8. It is more cohesive because familial customs connect everyone to each other, compared to a professional network (panel B), which exhibits less cohesion. While actor 8 knows everyone in the work unit (8 is connected to E, F, G), only a single link to different subsets of people is maintained—one person in the human resource office (actor C) and one person in the technology support unit (actor K). Topographic differences generate structural variation throughout each network, which affects the nature and amount of opportunities available to, or constraints imposed on, each person. Someone could be well positioned professionally to access information about new opportunities in the company, while also being highly constrained by a tight-knit familial network that does not accept deviations. An example would be that Aunt Mary hosts all holiday meals, attendance is required without exception, and this will not change until her funeral!

Returning to the credit card–skimming example, information about the victimization spreads differently depending on network topography. In a network where everyone is connected to each other, information about your experience may dead-end by recirculating repeatedly among family members. Just think: the credit card incident may become family lore and, to your dismay, could be retold at every holiday meal! Sparsely connected networks, such as the professional network, behave differently, and in this scenario information about your experience may course through the network. For example, the story about your victimization could spread like wildfire, traveling until it dies out. In other words, in sparse networks like the one illustrated in figure 3.1, panel B, information from person 8 is less likely to loop back to its origin, as there are few indirect channels reaching back to 8. If person 8 is situated in the network exhibited in panel A, however, information can loop back through many different sets of people. Notably, the structural constraints (or advantages) that emerge from topography may exist in only a portion of a network rather than being characteristic of the network as a whole.

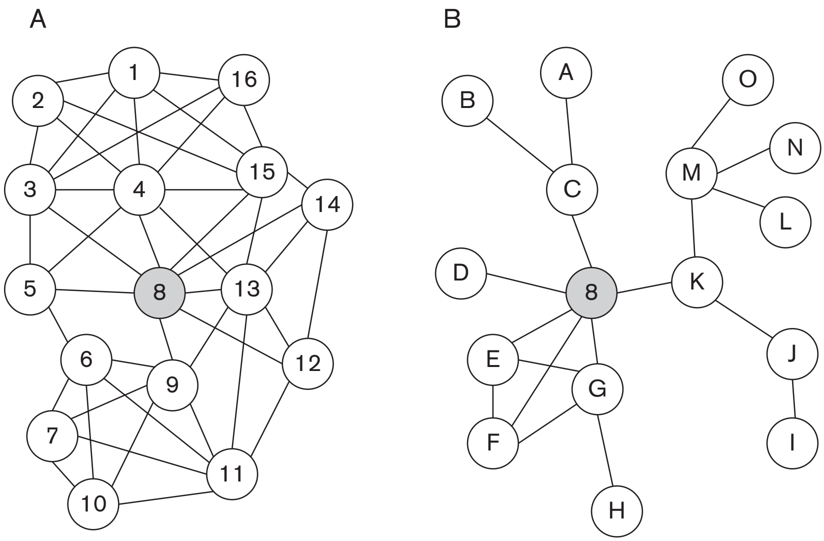


FIGURE 3.1. Idealized cohesive and sparse social network structures. *A*, cohesive group of sixteen actors with forty-three ties; *B*, group of sixteen actors sparsely linked by eighteen ties.

3. Social networks are dynamic, continually evolving as actors interact

Unless a historic network is being investigated where all actors are dead, social networks are assumed to evolve. Actors transform relational structures, intentionally or unintentionally, by forming or discontinuing relations; sharing information and materials; or changing perceptions, beliefs, or feelings. Consequently, the currency of a mapped network depends on the rate of change of the phenomena being studied.

Routinely observing interactions is necessary to capture the dynamic nature of social influence. Unfortunately, I have not encountered much thinking or investigation on how to establish an appropriate observation period. Commonsense suggests three factors. First, when establishing observation periods it is important to consider the *context of the study*—that is, the life cycle of delinquent peer groups may coincide with the academic school year. Second, the *permanence of relations* may vary—the bond that forms between a crime boss and protégé might be somewhat permanent, in that the tie may become dormant for a period of time, but can be easily reactivated. An association between a drug dealer and a tourist, on the other hand, might terminate with the transaction. Third, it is also important to consider the *periodicity of activities*. For example, when monitoring money-laundering transactions that involve wiring funds between various accounts on a daily basis, a month of information may be sufficient to map the entire network. A less active scheme that launders funds through real estate transactions, however, may require several years of data to capture sufficient information about actors and their relations. Be aware that using observation periods that are too short or too long obscures key features of the topography and any substantive changes that may occur with time.

4. Macrolevel social structures and transformations emerge from the combined behavior and preferences of individual actors navigating their social worlds with bounded knowledge of the larger social network within which they are enmeshed

Actors are most acutely aware of their immediate social neighborhood and, to a lesser extent, of the relations of their contacts beyond their egocentric worldview. Each actor will act and react on the basis of what they know. As a consequence, rationality is bounded by the perceived edges of the ego's network. But since influence can extend beyond a friend of a friend, the ego can be influenced indirectly by other actors beyond the periphery of bounded knowledge. Scientists adopt a bird's eye view to understand the macrolevel effects by integrating dyadic or egocentric viewpoints: aggregation reveals the larger community. It follows that behavior and reactions at the microlevel reverberate through a network, resulting in macrolevel phenomena, such as a riot during a sporting event (Christakis and Fowler, 2009).

Since actors are directly influenced by others in their immediate environs, as well as by indirect effects rippling through the community, multilevel influence is always assumed. Models exist to understand how individual-level behavior is influenced, directly by local forces and indirectly by the larger group. Studies investigate an individual's position (egocentric approach) within a local social context, examining the focal actor of interest (ego) and his or her direct connections (alters), along with the associations among alters; cohesive subgroups within the larger network; or the network as a whole. Of importance, results describing analysis conducted at one scale, for instance, an egocentric analysis, must be viewed with the understanding that each level—subgroup and whole network—will exert some unmeasured influence on the ego that is not captured by the reported findings, and the strength of the influence may vary.

5. Social networks persist, even when individual members enter or leave or relations form or dissolve

Routine evolution is inevitable. When actors join or leave a group, the resulting formation and dissolution of ties may change the topography and structure of a network, but the social network itself remains—groups have staying power. Think about the unit or department you work in. Despite changes in personnel, the group will continue to function, albeit differently, owing to the composition of actors and their working relations. With this said, efforts to disrupt activity by strategically removing key persons or interrupting ties extending between subgroups can significantly fracture the network for a period of time, resulting in disconnected graphs. (Recall from chapter 1 the temporary results of the original 2007 focused-deterrence effort used in Cincinnati to deal with a violent drug market.) To understand this phenomenon, a few definitions are needed.

Networks are classed as **connected graphs** if all pairs of actors can be reached—a tie exists linking each person to at least one other person, and the sum of connections generates a single interconnected group. In other words, there are no isolates: each person links to at least one other person. Think of social relations in a kindergarten class, where everyone plays with at least one other person during recess and play groups change on successive days. A strong possibility exists that the play group will constitute a connected graph by the end of the year.

Connected graphs of social relations, however, may not necessarily form a **complete network.** A network is said to be "complete" when all actors can directly reach everyone else. Complete networks may be rare. Even if observers are present for the entire academic year, they may find that not every child plays with everyone else in the class. Instead, they are likely to observe strongly connected subgroups of children, who regularly play together, and a few instances in which individuals branched out to play with someone different.

A social network can also become disconnected. In a **disconnected graph,** separate groups of interconnected actors exist, but they will be unable to reach other groups. Consider an event that triggers a family feud, which splits the extended family network into nuclear cells. In this scenario, even though the social network has fractured, intact components remain. Since networks are in a state of continual change (recall assumption 3), the disconnection might be temporary.[1] For instance, Aunt Mary, who insists

1. Notably, evolutionary trends will be conflated if data are missing, particularly in the presence of variation in which information is missing across successive waves

on bringing the entire clan together for holiday meals, may hold enough sway to reunite the group.

Each of the five theoretical assumptions discussed above is either explicitly or implicitly rooted in formal social network theories. While a thorough discussion of all major social network theories is beyond the scope of this chapter, four of the most essential perspectives are discussed below. Please keep in mind that research testing these ideas is not reviewed here; instead, I report on the key tenets of each perspective. Readers are left to investigate the modern incarnations of each theory and supporting empirical research on their own—some guidance is offered at the end of the chapter.

FOUR ESSENTIAL THEORETICAL PERSPECTIVES

Strength of Weak Ties

In a set of seminal articles, Granovetter (1973, 1983) drew attention to the critical function played by weak ties in the formation of resilient or strong networks. Coining this approach as the strength of weak ties, Granovetter sought to explain how interpersonal ties at the local level were important to macrolevel phenomena as disparate as political organization, diffusion of innovations, and social mobility. He begins his argument by describing tendencies of strong ties. Strong interpersonal relations develop between pairs of people in a nonrandom way. Relational strength is a function of (1) time spent with each other, (2) frequency of contact, (3) the emotional intensity and intimacy of interactions, and (4) the reciprocal nature of exchanges. These factors are the demands of strong ties. When they weaken, so too does the strength of the relation.

Observing the results of landmark research at the time, Granovetter (1973, 1983) noted the tendency of people to form stronger relations with others who are like themselves. **Homophily** (sameness) in relational choice often involves age cohorts but could include political, social, and cultural factors. We often choose to have strong relations with people who resemble us in meaningful ways: the associated cliché is that birds of a feather flock together. Figure 3.2 shows that Bruce is likely to form close relations with people he has a lot in common with (panel A). It is natural that Alex and Chris will have commonalities that will facilitate the formation of a tie at some point in the future (illustrated by the dotted line in panel B). When

---

of data collection. Estimates will be biased because analytic tools typically assume that datasets include the entire set of actors and their relations, for which relational information captures the absence and presence of ties.

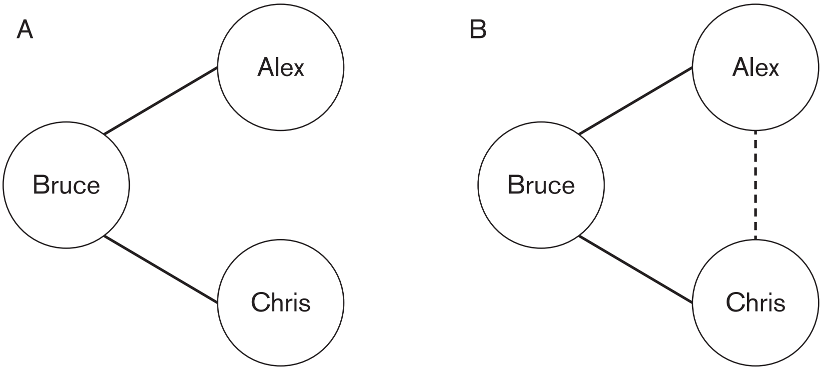


FIGURE 3.2. Cognitive balance leading to transitivity. *A*, Bruce has two strong relations; *B*, a positive relation develops between Alex and Chris.

this additional relation forms, the triad closes and the set of three actors is said to be **transitive.**

As individuals develop relations with others, a problem arises if one's close relations do not like each other. Formulated by Heider (1958), the theory of **cognitive balance** suggests that if a positive tie were not to form between Alex and Chris, a psychological strain would be introduced, generating a forbidden triad (positive associations between Bruce and his friends but a negative relation between Alex and Chris). Forbidden triads are inherently unstable and likely to dissolve. The time Bruce spends with Alex might affect whether Chris sees Bruce as much as she wants. If a positive sentiment were not to develop between Alex and Chris, it would be difficult for Bruce to maintain both relations—he is likely to lose a friend.

Forbidden triads aside, the local network surrounding an individual will thicken over time, in that friends will meet and form ties. As transitivity increases, more associations among sets of three people will develop, and the network will become more resilient. Moreover, the probability that information, or whatever is flowing through the network, will pass between any two people is proportional to the number of common links to others and inverse to the length of those paths (Granovetter 1973). In other words, the more direct connections among members of a group, and the more friends they have in common, the more likely it is that information will travel between Bruce and any other member of his larger social group. If the individuals had unique sets of friends when they first met, their individual networks will over time overlap to a greater and greater extent. This also means that the network will become less flexible or more resistant to change, because ties among group members serve as conduits recirculating information and reinforcing social norms. In the end, individuals will become more constrained and less able to behave in ways that contradict group norms. Yet new information is critical for survival in a social environment that continually evolves, which is where weak ties become important.

Weak ties are so named because they are more apt to form between people who have less in common and are therefore considered "weak." This is because these relations are likely to generate fewer exchanges that involve shorter and less emotional interactions. For these reasons, weak ties are subject to decay. Because the actors party to a weak tie are less apt to have much in common, the weak tie is likely to perform a **local bridging function,** acting as a connection between a pair of people involved in distinct social groups. Figure 3.3 illustrates what this might look like. Here, a solid line shows that Alex has strong ties with Bruce and Chris, and a dashed line indicates a weak bridging tie to Karl. Alex met Karl at the gym and has no other relation with him. Because Alex goes to the gym only sporadically, typically around the holidays, this relation does not strengthen. (We will return to figure 3.3 when the discussion turns to structural holes theory.)

Putting these ideas together shows that when strong ties form among people, there is a greater probability that their social worlds will overlap as a result of the inherent tendency toward homophilous relations and transitivity. As transitivity increases, strong ties are less apt to be bridging ties. Bridging ties are critical to accessing new information and thus play a vital role in maintaining adaptability. Bridging ties are more likely to be weak ties formed with individuals different from ourselves in material ways. While likely to decay, weak ties have a strong influence in binding sections of the network, linking small sets to the larger group.

Structural Holes Theory of Social Capital

Burt's (1992) structural holes theory of social capital draws attention to the importance of having **nonredundant ties** through which to access novel information. A nonredundant tie is a bridging connection that is unique (compare Karl's tie to Alex with his tie to Joshua in fig. 3.3). Used in this context, it means that a person has a relation with someone who is part of another cluster of people, which no one else in the egonet can reach: none of the ego's alters are connected to the other subgroup. Karl has a nonredundant tie to Alex: none of the other people in Karl's local neighborhood (egocentric network) can access the group that Alex belongs to. **Structural holes** are the gaps between nonredundant contacts. If you take a look at the

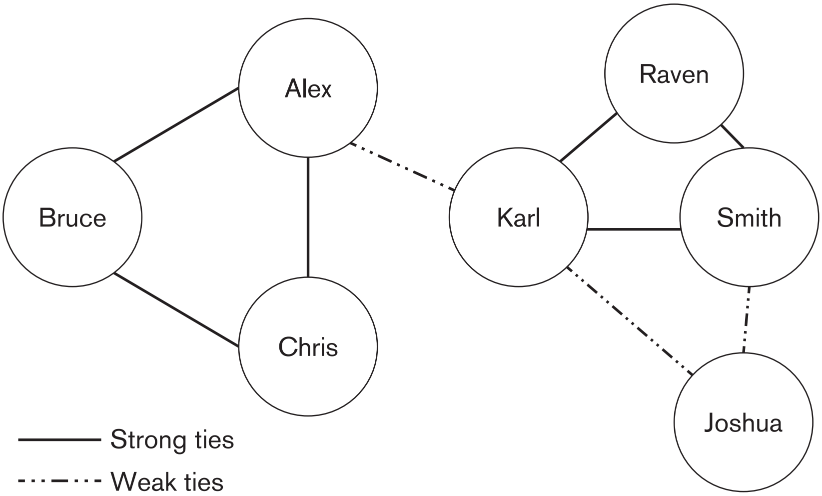


FIGURE 3.3. Bridging function of weak ties.

empty space in figure 3.3, such as the lack of connectivity between Chris and Karl, you will see that this space is a structural hole generating a nonredundant tie between Alex and Karl.

Being the only one who is able to access information flowing in other clusters of people provides benefits, such as being the first to learn about a new way to launder money. Connections that enable someone to access different pools of information within an identifiable group, such as a company or organization, increase one's **social capital** in the group. Social capital is a function of relations—the structural position that an individual occupies in the network as a result of relationships with others, relative to the social position of all others. Individuals with greater social capital can maximize their success, provided they can act on the unique opportunities that the nonredundant ties provide.

Returning to figure 3.3, if the network illustrated officers working different shifts in the same policing district, Karl would be best positioned to know what is going on in the community. His connection to Alex links to the dayshift and provides information about issues reported during business hours, whereas Joshua provides information about the nighttime economy. A structural hole exists between the day shift and the swing shift that only Karl can reach across. For this reason, Karl might be better able to understand the range of issues facing the community than his coworkers are and might thus be better positioned to identify a focal concern. Moreover, his relations with other officers might help him to leverage the resources needed to launch a successful problem-oriented policing (POP) project. If the chief values successful POP efforts, Karl may achieve greater professional success than his fellow officers. Karl has greater social capital as a result of his position in the network, but does this mean he is destined to capitalize on the situation? Not necessarily, as social capital does not in itself predict success.

Burt (1992) argues that two people with the same structural advantage associated with social capital may achieve different outcomes in a competitive setting. Variable outcomes are linked to financial and human capital—two attributes of the individual actor that are typically unrelated to network position. **Financial capital,** such as savings, access to funding streams, or lines of credit, is not evenly distributed among actors. Financial capital is generally an actor attribute, but it is possible that this resource might be associated with one's position in a different network. Karl, for example, might not have the financial means to bankroll a POP project, but his Aunt Mary might be a business owner looking to donate to a local charity; and his wife might be a grant writer working for the metropolitan transit agency, who just received funds to redesign bus shelters. (Aunt Mary strikes again!) If Karl is interested in working with a local nonprofit organization seeking to improve public transportation safety, Karl's personal network might provide the financial capital he needs to initiate the project in his professional network. **Human capital** refers to the natural qualities (e.g., diplomacy and charm) individuals possess, as well as to their acquired skills. Skill acquisition is generally associated with formal education, job experience, and postemployment training. Similar to financial capital, natural qualities and skills are not evenly distributed. In our example, if Karl has more years on the job, attended the recent POP conference, and possesses the necessary negotiating skills to generate agreements among stakeholders, he might be better able to manipulate his contacts into helping with the POP project, thereby ensuring Karl's success and earning him more occupational prestige. In turn, Karl's success makes the chief look good, thereby contributing to the success of the organization as a whole.

Burt (1992) adds that, in addition to enhancing information benefits, structural holes can offer control benefits. **Control benefits** accrue from being positioned between competing individuals who do not have direct access to each other—for example, when the market is saturated, two competing wholesale drug dealers could be played off against each other to drive down the price. So what do control benefits offer in our POP example? Well, if the weak bridging tie between Joshua and Smith were to dissolve, Karl would be in a position to play Joshua off against Alex to take on a greater role in helping to run the POP project.

At this point, you might be wondering why you wouldn't just structure your network so that all of your relations are nonredundant—strategically investing time and energy in cultivating relations with previously unreached clusters of people will maximize your potential for success. There is a limit, however, to the number of relations one can maintain. Moreover, while larger networks are more likely to provide greater diversity of information, and thus greater information benefits, the additional contacts might link to the same secondary people. In other words, the additional contacts might add redundant links that tap information benefits you already enjoy. Simply put, in this case, unbeknownst to you, the person you just met is already a friend of your Aunt Mary. So the advice you just got was actually from Aunt Mary—no wonder it sounded familiar! To optimize the network for effectiveness, Burt (1992) argues that actors should invest in nonredundant primary contacts, who themselves are surrounded by structural holes (referred to as secondary holes). A personal network characterized with many primary and secondary structural holes maximizes the flow of novel information without increasing the effort needed to maintain relationships. In other words, an actor should form nonredundant ties to individuals who themselves have many nonredundant ties. In doing so, the actor will build an efficient network (few primary ties, most of which are nonredundant) that taps into a higher volume of diverse clusters, which in turn provide better-quality information. Unfortunately, this feat is hard to accomplish without having an omnipresent view of the whole network!

To summarize the argument, nonredundant ties to individuals, who themselves have nonredundant ties and are part of unique constellations, produce information benefits. Having structural autonomy that facilitates entrepreneurial motivation provides control benefits. Taking these two things together, this social positioning provides social capital, which when combined with sufficient capacity to act leads to opportunity. The capacity to act on information benefits, and to invoke control benefits, is a function of an individual's financial and human capital: the interaction generates opportunity.

Complicating the situation is the tendency of people to form strong, trusted relations with others like themselves (homophily) and the likelihood of one's contacts to form relations over time. In our example (fig. 3.3), Karl may have bridging ties to different shifts because he worked for a time on the day shift and went through the academy with Joshua (night shift). It is likely that over time, other ties will form among the officers working in this district, particularly if they have something in common. In this instance, Raven and Alex have young children enrolled in the same day care. The children became friends, and now there are joint playdates and birthday parties that strengthen the relation between Raven and Alex. Karl's social capital diminishes with the addition of this bridging tie.

Structural holes theory is somewhat consistent with the ideas proposed by the strength-of-weak-ties argument: Burt states that "weak ties and structural holes seem to describe the same phenomenon" (1992, 27). He claims, however, that the structural holes argument provides a better explanation of the causal mechanisms at work. Bridging ties are not important because they are weak; instead, bridging ties span structural holes, thereby offering information and control benefits. In the next section, the discussion continues to expand on the importance of intermediaries.

Small-World Perspective

> Almost all of us have had the experience of encountering someone far from home, who, to our surprise, turns out to share a mutual acquaintance with us. This kind of experience occurs with sufficient frequency so that our language even provides a cliché to be uttered at the appropriate moment of recognizing mutual acquaintances.
>
> We say, "My it's a small world."
>
> MILGRAM, "The Small World Problem"

Intermediaries play a critical function in generating society. The small-world perspective advances the argument that a large diverse community is bound together by short paths linking intermediaries—chains generally include five to seven people. Chains of acquaintances bind the network, typically through weak ties, because they generate shortcuts. These link people at the ends of the chain, who on the surface would be considered at a great social distance—for example, there is a good chance that you are within five-to-seven steps of being indirectly connected to Halle Berry. Acquaintance chains make up the "threads" of social fabric.

In addition to their low average path length (on average, people are reachable through short chains of intermediaries), small-world networks also exhibit subsections that have high transitivity (most friends are linked to each other). As a consequence, small-world networks are clumpy, with lots of subgroups or cliques containing clusters of interconnected people but with a set of bridging ties able to transmit information between cliques. Networks exhibiting small-world characteristics are more successful, as the mix of diversity and cohesion generates an optimal structure for the rapid diffusion of information among legitimate (e.g., Uzzi 1996, 2005; Watts and Strogatz 1998) and criminal networks (e.g., Medina and Hepner 2008; Natarajan 2006; Xu and Chen 2008). Although this explanation makes small worlds and structural holes sound very similar, recall that Burt (1992) investigated organizations, whereas small-world scholars investigate societal networks (a.k.a. diverse communities).

Two sets of defining characteristics—scale-free properties and self-organization—are associated with mechanisms thought to play pivotal roles in the formation of small worlds. Small worlds are **scale free:** irrespective of the scale of the investigation (e.g., district, city, nation, or global network), charting the number of direct contacts actors have tends to reveal a heavy-tailed distribution that approximates a Pareto distribution or power-law function, according to which a small percentage of actors have many contacts and most actors have relatively few direct links. Actors with a high number of direct contacts are termed *hubs* relative to others in the network. Barabási and Albert (1999) proposed that power-law distributions appear as a result of **preferential attachment.** Applied to the distribution of contacts in a small world, preferential attachment refers to social processes that generate more links for those actors who already have a lot of direct contacts relative to others. In other words, actors with the most friends will gain more associates over time, compared with those having few direct contacts. In addition, major hubs tend to be linked to smaller hubs, and so forth. Hub-to-hub connectivity significantly shortens the distances among all actors in the network.

Social networks with scale-free properties are more resilient to random failure, as the loss of any actor selected at random is likely to be a person with few contacts rather than a hub. If a hub is removed, the network will remain connected owing to the likelihood that, because of preferential attachment, the remaining hubs will inevitably link to each other; however, the group is more apt to fracture from repeated attacks targeting major hubs located at the core of the network than from repeated attacks targeting hubs positioned at the periphery of the group. If this were applied to the typical topography and structure of a drug-supply network, it would be reasonable to expect that removing wholesale drug distributors with high hubness might do more to fracture an illicit drug distribution network than targeting major retailers located at the periphery of the drug-supply chain.

The second factor leading to the formation of small worlds is the **self-organizing principle** of networks. Order will spontaneously emerge out of a disordered social system because of the local interactions among actors.

Triggered by chance encounters and amplified by positive feedback, order will evolve in a group without central planning or intentional influence by external forces. Self-organization tends to generate a decentralized, distributed structure that will regenerate after attack (able to self-repair).

Self-organization is the mechanism generating global structure. Advocates of small-world studies propose that the features of the global structure of a network emerge from the local patterning of relations. This means that connectivity at the societal level is the outcome of local social processes, in that "network ties emerge, persist, and disappear by virtue of actions made locally at the scale of the individual actors in a network (whether they be persons or families or companies or some other social entity). . . . On the basis of their localized view, they form strategies and make decisions that intersect with those of others who are socially proximate. Combinations of these competing or complementary intentions and actions constitute the social processes that make up local patterns of relationships. These local patterns agglomerate to create the global structure" (Robins, Pattison, and Woolcock 2005, 895).

Readers unfamiliar with SNA should be aware that the development of the small-world perspective differs from the other theories discussed. To some extent, the small-world perspective emerged from a debate between mathematicians and experimental social scientists about how best to resolve the small-world problem, framed as what is the probability that any two people, selected arbitrarily from a large population, will know each other? Experimental social scientists led by Stanley Milgram responded by adopting an inductive approach. (Mathematicians used a more deductive process.) As a consequence, no conversation about the small-world argument is complete without mention of the "small-world method." Be advised that we will return to this topic again in chapter 7.

Contagion and Three Degrees of Influence

A funny thing happened when I began to write this section. I brought the book *Connected* by Nicholas Christakis and James Fowler (2009) with me to the salon, in order to look over some notes penciled in the margins about contagion and influence (bibliophiles please excuse my nasty habit). I am not one for small talk and figured it best to keep myself occupied. Having recently moved to the city, I was encountering this stylist for the first time. After a few minutes, the stylist, who had managed to get a look at the book cover, exclaimed in an excited voice, "Oh, I know Christakis. He wrote the book you are reading. My wife used to be his trainer!" Needless to say, I put the book away, and we chatted for the remainder of the appointment. Maria turned out to be a lovely person, and I could not have asked for a better segue from small worlds to contagions.

Christakis and Fowler (2009) observe that social phenomena of any sort, be it disease, violence, information, values, or emotion, spread through networks, moving between actors like pathogens. Their theory of social contagion is outlined in the aforementioned book and highlighted in a TED talk (Christakis 2010a). Terming their idea *hyperdyadic spread,* these authors posit that the phenomena flowing through ties will extend beyond their origins and those directly connected to actors only indirectly associated with a triggering event or idea originator. Moreover, relations among actors influence decision making, beliefs, and behavior, through complex, relatively short pathways.

A return to the example of credit card skimming, and tracing one path your warning took, easily illustrates how influence can spread in a hyperdyadic fashion. Imagine that you passed the warning to a woman at work, who in turn shared it with her spouse, who informed three people from a bridge club. In this example, the bridge players were influenced by your victimization, despite the fact that they are three steps away from you—meaning the information was passed in an indirect, chainlike manner across different networks, starting with your coworker (one step removed from you in a professional network), who passed it to her spouse (two steps removed in a familial network), who conveyed it to bridge players you did not even know existed (three steps removed in a friendship network).

Notably, the ripple effect that extends outward from any origin is unlikely to continue for long. The local neighborhood influencing behavior is three steps. Calling their concept the **three degrees of influence rule,** Christakis and Fowler (2009) argue that influence decays gradually and ceases at a "social frontier" of three links. Consequently, a friend of a friend of a friend may exert some meaningful influence, but someone seven or eight steps removed from you has virtually no impact on your life. The influence of your victimization from credit card skimming will unlikely extend beyond the associates of the bridge players, unless of course someone in this chain amplifies the message as a result of their own experience.

Why social influence typically extends only three steps is still up for debate. Christakis and Fowler (2009) offer three possible answers: (1) the *intrinsic-decay* argument posits that influence fades because the fidelity of information degenerates, causing the wave of influence to dissipate; (2) the *network-instability explanation* holds that all networks are dynamic and that continual evolution renders links beyond three steps less stable; and

(3) the *evolutionary-purpose explanation* proposes that the survival of early hominids depended on people located within three steps. No one existed at four steps, and not enough biological time has passed for evolution to extend influence chains.

To account for hyperdyadic phenomena scaled to the global level, these authors posit a set of rules about how the structure of local social neighborhoods shapes the way social systems behave.

1. People deliberately form and change the shape of their social neighborhood by deciding whom they connect with, as well as how many people to associate with; whether their contacts meet; and how central they will be (e.g., whether they will take a peripheral role on the sidelines or be the host of the party). At the individual level, people are varied. Some have highly transitive networks, in which they are deeply embedded within a single group, whereas others may have contacts with multiple groups.
2. The network, and a person's position within it, shapes behavior and experiential attributes (e.g., health, occupation, and emotional states), and vice versa. While the influence chain extends beyond direct contacts, it is a complex process involving reinforcement from multiple contacts within the context of a dynamic social structure.
3. Social networks have emergent qualities—characteristics and functions that are separate from the actors and subgroups subsumed within them—that are not associated with central organization or control. Crowd effects accrue from the aggregate impact of small contributions and perceptions of individual actors (e.g., spontaneous rioting during a political protest). Networks also reinforce two kinds of social inequalities: *situational inequality* places some people at a socioeconomic advantage over others, and *positional inequality* places some people at a locational advantage within the community network.

Integrating these rules with the prior discussion, Christakis and Fowler (2009) propose that local social neighborhoods aggregate to form societal networks bound together by intermediaries spanning cohesive subgroups. The emergent qualities of the larger societal network generate inequalities among actors and waves of crowd effects, which in turn result in differences in how influence disperses through sectors of the network. To reconsider credit card victimization, if others who are several steps removed from you had suffered the same fate, and had similarly shared knowledge and adopted preventive measures, the collective experience (and response) might have triggered an information wave to ripple across society (crowd effect), which could render that crime technique almost obsolete in a short time. I say *almost* obsolete, because, owing to local social structures, not everyone would receive the information, and some people would remain vulnerable to victimization. At this point, readers should see how hyperdyadic spread expands on tenets of small-world theory.

SUGGESTED READINGS

Diving into the theoretical foundation of SNA is not an easy task. As with any interdisciplinary, multidisciplinary field, wrapping your mind around new concepts is harder when the supporting material is steeped in discipline-specific jargon or buried in arcane sources—who would think to look in a physics or computer science journal for recent scholarship applicable to criminal networks? Now that you have an overview of major streams of social network thinking, reading some original material written by theorists and seeing how these ideas are tested will help to solidify your conceptual understanding. For instance, I highly suggest that you read some of the seminal small-world studies for yourself, in the order listed below. While the language they employ is a bit dated, and may offend our twenty-first century notions of political correctness, many have made breakthroughs that continue to shape our understanding of small worlds that I was unable to cover here. The following list provides guidance to aid in your investigation.

- Strength of Weak Ties: See Granovetter (1973) and Granovetter (1983).
- Structural Holes: See Burt (1992).
- Small Worlds: For theory, see Milgram (1967) and Watts (2003); and for excellent applications, see Travers and Milgram (1969), Dodds, Muhamad, and Watts (2003), and Watts and Strogatz (1998).
- Contagion Processes: For theory, see Christakis and Fowler (2009) and Christakis (2010a); for excellent applications, see Christakis (2010b), Green, Horel, and Papachristos (2017), and Centola (2018); and for good debates, see Aral (2011), Cohen-Cole and Fletcher (2008), and Fowler and Christakis (2008).

In the next chapter, the social network theories explored above will be integrated with criminological theories about criminality. If you are not reading chapters 3 and 4 in succession, it is a good idea to briefly rescan this chapter before moving on. While there will be some review of social network concepts, the discussion assumes you will remember the content covered here.

4. Connected Criminality

> Criminogy's neglect of social network analysis serves as a warning that the discipline is failing to keep up with important developments in scientific inquiry, not to mention the fact that criminology is missing an opportuinty to test and expand upon some of its most treasured theories and concepts. Indeed, criminological theory oozes with network imagery and jargon.
>
> PAPACHRISTOS, "The Coming of a Networked Criminology?"

Most of our ideas about why people offend and how crimes transpire include some reference to an aspect of social networks. To illustrate, I draw your attention to the italicized words in the next sentence. Central concepts of criminological theories often include arguments suggesting that: (1) *interactions* between people influence perceptions, the development of values, exposure to opportunities to commit crime, and ultimately the adoption of criminal and delinquent behavior; (2) the *bonds* formed with individuals, groups, or organizations vary in strength and act as conduits of criminal influence on behavior, values, and norms; and (3) differential group and community-level *organization* impose varying behavioral *controls*, which when absent foster crime problems. If we kept the term *interactions* but replaced the other italicized terms—*bonds* with *relations or ties*, *organization* with *network structure and topography,* and *controls* with *constraint*—most criminological theory could easily find its way into social network thinking, or vice versa.

Reviewing all of the theoretical correspondence between fundamental social network analysis (SNA) principles introduced in the prior chapter and criminological theory is well beyond the scope of the book, let alone a single chapter. Instead, chapter 4 briefly examines the social network implications of select theories on a cursory level and then launches into an illustrative case study about social learning and methamphetamine use. The ideas reviewed here fall under the umbrella of explanations of criminality. In the following chapter, I investigate the social network implications for theories of crime, specifically deterrence, neighborhood ecology, and two opportunity theories—crime pattern theory and routine activity theory (also known as the routine activities approach). Both chapters conclude with a list of suggested readings. Throughout the discussion, tables of congruence link key principles of each theory with SNA implications. I hope this discussion ignites some ideas for future research that will infuse new life into criminological theories. A precautionary note: while many of you might be tempted to skip this chapter, or read intermittently, focusing only on the theories you are interested in, thinking about how SNA ideas can be integrated with each set of theories will help you conceptualize your own integrative efforts. In any event, do not miss the case study integrating social learning theory with SNA to explain methamphetamine use.

OVERVIEW OF THEORETICAL CONGRUENCE

Table 4.1 summarizes some of the theoretical congruence between major criminological theories and SNA, which are explained in the text below. Four theoretical traditions are profiled: biopsychological, control, developmental, and strain theories. While each tradition is rich, having several different and potentially conflicting arguments, only a generalized account—essentially, a central argument—is provided. The table also reviews how SNA concepts might help in either testing the theory or developing theoretical prepositions. These suggestions should be viewed as examples: many other arguments are viable.

Biopsychological Theories

According to proponents of the biopsychological perspective, a great deal of variability in criminal behavior may be linked to individual differences in personality traits—stemming from biological and psychological factors—that cause crime either directly or indirectly through social interactions. Proponents argue that crime involvement varies among people because people are different. Differences arise from **genetic variation** but are also a consequence of genetic predispositions being activated by patterns of social interactions and individual experiences. For example, someone may inherit an aggressiveness trait, but aggressive behavior does not manifest until a series of experiences bring out the behavior. If aggressive individuals associate with other people with the same trait, their behavioral patterns will be reinforced. From this generalized, and perhaps overly simplified characterization of a deep set of theories, it is immediately apparent that integrating tenets of SNA offers at least two direct benefits to biopsychological explanations of criminal behavior: (1) SNA adds a way of conceptualizing social interactions with implications for how to measure this concept, and (2) SNA provides an explanation of how traits influence the selection of associates and the formation of cohesive, like-minded groups

TABLE 4.1. Select theories of criminality and congruence with social network analysis (SNA)

Theory	*Central argument*	*Examples of SNA theoretical congruence*	*Benefits of SNA*
Biopsychological theories (e.g., crime proneness and supertraits)	Variability in criminal behavior linked to individual differences in personality traits that cause crime, directly or indirectly, through social interactions	—Reactions and interactions with others influence behavior in proportion to nonredundancy of the connection and strength of bond between actors, where bond strength refers to frequency and intimacy of contact and reciprocity of exchanges	Adds a measurable conceptualization of social interactions
		—Principle of homophily suggests individuals gravitate toward others with similar characteristics	—Explains how traits influence selection of associates and formation of cohesive, like-minded groups
Control theories (e.g., social bond theory, containment theory, techniques of neutralization, general theory of crime)	Socialization processes instill and reinforce normative values by fostering internal and external controls, curbing urges to act on crime opportunities	—Variation in structure of ego networks explains how individuals experiencing similar parenting behave differently—exposure to one nonredundant, positive influence like a coach may negate effects of poor parenting (strength of weak ties theory)	—Can conceptualize interactions at the egocentric level, with implications for consistent measuring within families to compare different children
		—As network composition evolves, so does information about crime opportunities	—Explains how exposure to opportunity changes

		Being embedded within multiple, overlapping networks deepens constraint imposed on individuals—making it harder to sever ties and engage in counter-normative behavior	—Explains how relationships with conventional organizations can constrain criminal behavior
Developmental theories	Crime results from developmental processes involving interaction between individual traits and social structural conditions; people ensnared in different life trajectories, restricting their ability to engage in conventional social roles or escape criminality	Important life events change composition of ego network	—Argue continuity and change in behavior explained by stability and evolution in structure of networks
Strain theories (e.g., relative deprivation, general strain theory, strain theory)	Failing to achieve socially prescribed expectations, failing to attain aspirations or goals, losing something valued, or experiencing noxious stimuli can produce strain—may lead to criminal behavior	—Relations provide a mechanism through which aspirations form and are reinforced; relations enable comparisons to assess progress toward goals; reactions of others generate strain experienced when people fall short of their desires	—Provides conceptualization of social neighborhood/reference group and explanation for how aspirations and strains are generated through interactions with others
		—Individuals with greater social capital, and ability and resources to use information and control, resulting in benefits, achieve greater success within their group relative to peers; should be less apt to resort to criminal or deviant behavior	—Accounts for variable criminal behavior among individuals experiencing same strain

that reaffirm behavior. The arguments in support of these statements follow.

First, biopsychological theories generally suggest that the variability in criminal behavior is linked to individual differences in personality **traits**. Traits can be developed, or brought out and reinforced, by social interactions between family members and the reactions of parents to a child's behavior. Even if behavior has distinct genetic origins, *social interactions* during psychological development can suppress or bring out the trait. In SNA terms, social interactions during formative years influence the way individuals develop and shape how they will interrelate with others in the future. Networks are dynamic. And it is likely that the social neighborhood of an individual will vary with age. During the early years, it will be made up of immediate caregivers, whereas in later years, the social neighborhood will include a greater mix of people. Reactions and interactions will influence the ego in proportion to the redundancy of the connection and the strength of the bond between actors, where bond strength refers to frequency and intimacy of contact and reciprocity of exchanges. In other words, where redundancy is high, there is likely to be more consistency in how different caregivers react to a child's behavior—that is, if traits related to criminal behavior are suppressed uniformly, the trait will not develop. (This is provided only for groups that are not criminally inclined; otherwise, the influence would be the opposite. A cohesive network among crime-involved family members would reinforce criminal tendencies in the offspring.) Where redundancy is low—little contact or cohesiveness exists among caregivers—there will be little consistency in how people react, which will provide a suitable environment for the criminogenic trait to emerge. A situation like this could occur if caregivers do not share a parenting model.

The second benefit gained by integrating SNA concepts pertains to the biopsychological argument that traits also influence the selection of associates. To further clarify how this occurs, recall the principles of homophily and cognitive balance. The SNA principle of homophily suggests that individuals will gravitate toward others with similar characteristics—that is, aggressive people will be drawn to other aggressive people. Moreover, individuals will form stronger bonds with others like themselves. Strong bonds among homophilous individuals will in time increase transitivity, thereby generating a cohesive network that reinforces group norms and behavior. Cohesion increases because, according to cognitive balance theory, transitive ties form when positive sentiments develop among sets of people connected to a third party. For example, if my two best friends like each other, then I can keep both friends, but if one dislikes the other, I will eventually end up with only one best friend. Cohesive, multigenerational groups will form if these tightly intertwined sets of people remain somewhat isolated from the larger society.

Control Theories

According to control theories, such as social bond theory and the general theory of crime, socialization processes at the local level—for example, within the family—instill and reinforce values by fostering internal or external controls that help individuals refrain from acting on crime opportunities. From the **internal controls** perspective, familial interactions promote behaviors by establishing, or failing to establish, internal controls to curb base urges. While internal controls are thought to be permanent—either you can control your urges or you cannot—criminal behavior will vary throughout a person's life. To explain this phenomenon, theorists argue that the amount of criminal activity someone participates in will fluctuate as a result of shifting exposure to opportunities. **External controls** emerge from involvement in different prosocial community organizations, units, and groups: participation evokes the development of bonds that anchor an individual within conventional behavior. Adopting SNA thinking, provides three potential benefits to control theorists.

First, SNA theories offer a way to conceptualize how internal controls on behavior are developed by interactions with others. For example, variation in the structure of the ego network can account for why individuals whose parenting experiences were similar behave differently—being exposed to one nonredundant, positive influence (e.g., a coach) can negate the effects of poor parenting (extrapolating from the strength of weak ties theory). Moreover, networks are dynamic. Relations form and dissolve. Since relations bring information, they are also the source of opportunities; thus, as the network composition evolves, so too will access to crime opportunities. This means that the second benefit of infusing SNA thinking into control theories is that it provides a measurable argument that accounts for the dynamic nature of crime opportunity.

Third, SNA offers a conceptualization of how involvement in prosocial organizations and groups imposes external controls on behaver. Control theorists focused on external forces argue that behavioral controls, rooted in relationships with conventional organizations and activities, can also constrain criminal behavior. SNA scholars may suggest that this occurs because of multiplexity, which binds individuals in different social networks. An example would be going to the same school (network 1) as other kids from the neighborhood (network 2). Being embedded in multiple, overlapping networks (same school and the same community) deepens the constraint imposed on individuals. Deviating from proscribed norms triggers reactions and potential sanctions in different social groups, magnifying the influence of others and making it harder to sever ties and engage in counter-normative behavior. When individuals have only a single type of relation—they have no siblings at the school and they are bussed to the school from a neighborhood across town—the cohesiveness of their contacts, meaning the degree of interconnections among their contacts, rather than the number of networks someone is enmeshed within, may provide behavioral controls. For instance, schools with a cohesive social structure will deeply embed individuals in a web of transitive relations (i.e., all kids participate in different clubs), whereas schools with minimal interactions among rigidly defined cliques (i.e., kids in band do not associate with jocks) will have low overall cohesion and will thereby impose low levels of constraint on behavior within the school (the community level).

Developmental Theories

According to developmental theorists, crime is the result of a developmental process involving the interaction between individual traits and social conditions throughout the **life course.** People are caught in different life trajectories that restrict, to a varying degree, their ability to engage in conventional social roles, or alternatively their ability to escape criminality by responding to changes that occur at pivotal **turning points** (i.e., graduation, birth of a child, retirement). Delinquent and criminal behavior, in turn, weakens bonds to conventional society in a cumulative fashion.

SNA aids our efforts to explain continuity and change in behavior. Behavioral changes can be explained by evolution in the structure of networks. If individuals experience major shifts in the composition and structure of their social networks at pivotal turning points in their lives, social influences will change, which should alter behavior. Substantive changes in social structure, which may occur as a result of marriage, a new job, becoming a parent, etcetera, can reflect the imposition of a new form of informal social control that triggers desistence from crime, particularly if it is accompanied by cognitive transformations. The mechanism of change is found in the theory of cognitive balance. Cognitive balance theory suggests that cognitive dissonance will occur, resulting in the dissolution of relations, if a favorable sentiment does not exist among people the ego is strongly connected to. For example, if a "good marriage" occurs, and the ego opts to maintain the marriage, the person will need to break off existing, historic relations with antisocial friends. Because interactions between individuals can be characterized by their strength and nature (prosocial being positive and antisocial being negative), network analysis provides a way of assessing the degree of influence individuals and events can have on the life course. Mapping social relations over time provides an opportunity for testing the effects of turning points.

Strain Theories

Relative deprivation, strain theories, and general strain theory share the notion that failing to achieve the same success as one's peers or associates (or failing to attain aspirations or goals), losing something valued, or facing other noxious stimuli produces **strain**, which under certain conditions leads to criminal behavior. I can see two distinct benefits to integrating SNA concepts into strain theories: (1) SNA provides a tangible conceptualization of the social neighborhood/reference group and adds an explanation for how aspirations and strains are generated through interactions with others, and (2) network theories can also account for variable criminal behavior among individuals experiencing the same strain.

As explained in chapter 3, the egocentric network, and indirect influences up to three steps removed that are transmitted to the ego via individuals in close social proximity, constitutes the local social neighborhood within which individuals navigate life. SNA scholars are likely to argue that it is our relations with others that provide the mechanism through which aspirations form and comparisons are made to assess progress toward goals. Further, the reactions of others reinforce aspirations, as well as generate the strain experienced when people fall short of their desires. Consequently, both aspirations and strains are generated through interactions with others.

Integrating the strength of weak ties theory or structural holes theory also provides arguments that may account for variable criminal behavior among individuals experiencing the same strain. For instance, having weak ties to other groups, and potentially other social norms, provides advantages. Individuals socially positioned to be a bridge in the larger community network should be better equipped to deal with strain, if they have access to different perspectives, because information provides more resources to overcome strain. Individuals with less structural advantage will be more constrained by group norms and less able to resolve problems, particularly if they are socially isolated from conventional organizations and resources.

Structural hole theorists might argue that within a working group, and by extrapolation within other goal-oriented criminal groups or high-crime communities, there is an uneven distribution of social capital, human capital, and financial resources. Generally, individuals with greater capital, and the ability and resources to use the information and control the benefits this brings, will achieve greater success within their groups relative to their peers. They should, therefore, be less apt to resort to criminal or deviant behavior. (Alternatively, you could argue that individuals with greater "crime-oriented" capital may be more successful in criminal pursuits compared to others. In this scenario, the objective is not to be successful by conventional means. Instead, actors aspire to "get rich quick.") Either way, individuals experiencing the same strains do not have an equal ability to respond.

There are many other theories of criminality that could be added to this overview of congruence. Omission from this section should not be taken as a suggestion that SNA has little to offer; instead, consider it an invitation to add to the conversation! The material presented in the first part of the chapter was dense, containing many complex ideas. Rest assured, it gets easier from here. In the section that follows, I provide a detailed example to illustrate how social learning theory can be combined with tenets drawn from SNA in order to better understand methamphetamine involvement.

HYPERDYADIC SOCIAL LEARNING IN THE WORLD OF METHAMPHETAMINE

Data Source and Inspiration

If you have not had a chance to read *Methamphetamine: A Love Story* by Rashi Shukla (2016), I highly recommend picking up a copy. Using qualitative methods and providing rich narratives to support thematic analysis, this book is the result of four years of interviews with thirty-three individuals, each with extensive involvement in using, dealing, and distributing methamphetamine. Most notably, all interviewees reported a high level of involvement in the drug trade—twenty-three participants were former manufacturers, and the rest assisted, facilitated, financed, organized, or otherwise supported "cooks." In the course of documenting life histories—a total of 1,238 years of life experience—Rashi Shukla assembled detailed information about how each person got involved, escalated and changed usage patterns, began dealing or distributing, got involved in meth production, and desisted from drug involvement. One of the factors that gripped my attention was how the social networks surrounding each individual evolved with deeper immersion into the drug world. Drug use was learned through social interactions, and, as a consequence of adopting new behavior, the network changed—with each step of emersion and desistence, the interviewees without fail described social learning processes involving hyperdyadic spread of information and behavior.

BOX 4.1. DIFFERENTIAL ASSOCIATION

Four of the nine principles of Sutherland's (1947) theory correspond directly with SNA theory:

2. Criminal behavior is learned in interaction with other persons in a process of communication.
3. The principal part of the learning of criminal behavior occurs within intimate personal groups.
7. Differential associations may vary in frequency, duration, priority, and intensity.
8. The process of learning criminal behavior by association with criminal and anticriminal patterns involves all of the mechanisms that are involved in any other learning.

Learning Theory Refresher

Assuming readers have not thought about learning theories for a while, the network-oriented concepts and tenets of two of them—differential association (box 4.1) and social learning theory (box 4.2)—are reviewed here. Of the two, differential association is one of the theories most often integrated in SNA-oriented research on crime and delinquency. A review of the tenets reported in table 4.1 shows why. In short, the *criminal attitudes* and behavior of an individual are conditioned by the attitudes and behavior of their contacts, as well as the characteristics of their communications (e.g., frequency, duration, priority, and intensity). Mapping the egocentric network and the characteristics of these associations, we can study the process of social learning necessary to participate in crime and delinquency (e.g., Haynie 2001). The second perspective, social learning theory, draws our attention to direct and indirect influences flowing through the local social neighborhood, and to the importance of different types of ties—*primary* and *secondary groups,* as well as to long-lasting and close associations. Moreover, two key aspects of the social structure provide a context for learning—*differential location* in the social structure (attributes shape position or standing within the group) and *differential social location* (membership in and relations to primary, secondary, and reference groups).

Integrating Learning Theory with Hyperdyadic Spread—Methamphetamine Example

Considering the arguments of social learning theories in light of the theory of hyperdyadic spread that Christakis and Fowler (2009) present (see

BOX 4.2. SOCIAL LEARNING THEORIES

Interacting with others involves direct involvement (association and interaction) with others engaging in certain behaviors, as well as indirect association and identification with members of a reference group.

Learning mechanisms operate within a social context (groups) formed by different types of ties.

- Primary groups (family and friends) are most important. Secondary and reference groups (e.g., classes, neighbors) have varying effects.
- Early associations (priority), longer-lasting or more involved contact (duration), more frequent interaction, or closer relations (intensity) have greater influence.
- Most learning results from social exchange. The words, responses, presence, and behavior of others directly reinforces behavior, provides a learning situation, or furnishes a conduit for punishment or rewards.

Two key aspects of the social structure provide learning context—differential location in the social structure and differential social location.

chapter 3 for a review), we find an explanation for why there is fluidity in the composition of, and in the influences exerted by, the social network enveloping each person. Since people intentionally form and change the shape of their social neighborhoods, the network and the ego's relative positions within it changes, and this affects behavior and attributes (e.g., emotional states, financial security, and health). Influence is a complex and highly dynamic process involving chains that extend beyond direct contacts and reinforcement from multiple parties. Narratives about methamphetamine use, initial consumption, and escalation exemplify how social learning occurs within this framework.

Shukla (2016) reports that all first encounters with methamphetamine occurred in small groups, often in intimate settings with a known individual—a friend (sixteen of thirty-two interviewees, or 50 percent), a boyfriend or girlfriend (six, or 18.8 percent), family member (four, or 21.5 percent), or other social acquaintance, including older peers (six, 18.8 percent). What I found more telling was the social context surrounding the initial use. Three examples follow.

Vanessa reports that her introduction came from a family member during a trip when she was sixteen years old. She states (Shukla 2016, 31), "I remember it perfectly, because they were all usin' it and they just came up to me and said, 'Do you wanna get high?' and I was like, 'I'm not snortin' it,' cause it was openly used, it wasn't like it was something that was hidden. And they said, 'Well okay, you can drink it out of a shot glass,' and I was like 'Okay.'" Vanessa remarks that drug knowledge was shared as a rite-of-passage. While her siblings and cousins were users, she had not been invited to participate in drug use until the trip: at that time, she had reached the age when family members thought she might be old enough to use meth. In this example, the influences come from the primary group during a family vacation.

Emme was introduced to meth at sixteen by a boyfriend. Emme states (Shukla 2016, 34),

> I was hanging with the crowd, and I was never into doin' the drug thing, I did the marijuana kinda to fit in. I was more into drinking, and I had lost a couple boyfriends because of it. They thought I was too prude or just wasn't into it. So I said, screw this shit. I'm not doin' this crap anymore. And had found somebody I really, really, really liked, and he introduced me to cocaine first. . . . And then he introduced me to meth, probably about a week later. . . . I was always the one that sat on the outside. I was tired of sitting on the outside. I wanted to do what everybody else did. . . . I was sixteen and he had his own apartment. We always partied there. It was a big thing and a bunch of people had it. Everybody was having a good time.

In her comments, she acknowledges the fluidity of primary contacts (boyfriends) couched within a large social group of secondary associates and how the two sets of influences changed her drug use. She decided to start using after losing *another* boyfriend (unstable primary contact), owing to being labeled a prude (reaction from the larger group). By the time of the interview, Emme was twenty-six and had been involved in dealing and had played a supporting role in manufacturing.

Bryant's story provides a clear picture of the intersection of primary and secondary contacts, and how actors two or three steps removed exert influence. When asked who introduced him to methamphetamine at fourteen years of age, he reported the following:

> I am gonna say the name "Jay," okay? We used to always party out north of town with a guy named [name], [a] trailer house, that's where all the beer, all the drugs, I mean everything happened. And a guy by the name [name], he use to take me out there all the time. He wasn't with me one night and I went out to that trailer house, stupid me but I was sitting in the living room, we done smoked some weed and I kept watching these guys go in and out and I asked Jay what they were doing, and he says, "Hey, why don't you come in here and try this?" I was going on fourteen years old. (Shukla 2016, 33)

In this narrative, the trailer acts as a convergence setting, anchoring partying and drug use for a community. In using the phrase, "these guys go in and out," Bryant suggests that others are present who are part of the larger social network (two or three steps removed) but not necessarily directly connected to him. At the time of the interview, Bryant was forty-one years old and a former longtime dealer and manufacturer.

Similar to initial use, learning how to use meth in different ways accompanied network evolution and often occurred when a greater proportion of primary contacts was also using via a different method. Shukla (2016, 47) describes Ray, a thirty-five-year-old former dealer and manufacturer, who switched to intravenous use when his greater participation in the lifestyle coincided with more interactions with IV users. In an effort to maintain a partying lifestyle, Wes, a twenty-eight-year-old former manufacturer, found his peer group changing:

> When that started happening, when smoking it more started happening, we started to be around people who just wanted to, well, they would leave the room, and just come back totally blitzed. Their eyes, you wouldn't see any color or anything. I'm like, "What the heck is that all about? Did you take some to the bathroom with you?" And they're like "no, I shot up." Then they would just . . . they'd sit there and smoke, just for the taste. But then they'd be off the charts. I'm like, "I'm not getting the same effect; I'm not getting as high. I can still function." These guys are bouncing off the walls. It seemed like they were having more fun. (Shukla 2016, 50)

Transitioning from being a user to other roles in the drug-distribution chain occurred over time. The evidence presented by Shukla (2016) suggests that transitioning from a user to a dealer/distributor or manufacturer accompanies increased use and need for money to pay for drugs and the desire for more sustained access to larger quantities or a better product. But the roles assumed by each of the actors was controlled by their network positions and situations, in light of the larger social system—professions, contacts, and personalities figured prominently. For instance, several of the major distributors interviewed were truck drivers, who tapped into their capacity to transport drugs nationwide; some, including a stripper, were in professions that generated steady access to a large clientele; and others lived in cohesive communities, such as families or neighborhoods, where everyone was a user. In each case, however, interviewees talked about meeting more people in the lifestyle and how personal contacts helped them to move from user, to using and dealing, and then to using, dealing, and producing.

For instance, Troy, a thirty-four-year-old former dealer and cook, got involved with manufacturing when he began acquiring precursor materials for a friend who cooked. He puts it this way:

> Troy: They were making it, and what I'd do is I would go to [state] and buy these pills and stuff, and he'd give me like a certain percentage of dope from this.
>
> Q: From the cook?
>
> Troy: Yeah. And, I was thinkin', "Well I mean if he could do it, why don't I do it." . . . I asked him one day, I said, "Why don't you show me how to do that," and he showed me from start to finish and that's all it took, 'cause I've sit there and observed it, and he went from step one to everything, preparation and all that, and so I said, "Okay, I think I got it, let me watch you one more time." (Shukla 2016, 79)

Conner, a twenty-four-year-old former dealer and manufacturer reports that

> I got involved just because a friend of mine, I was using meth and dealing meth and I was buying from him. . . . Then I started buying this cooked crystal meth, from a local person, locally. And come to find out, he was makin' it right there in his backyard. Or, actually, a shed in the back of his house. And then so I got involved with that, and I started puttin' some money in on it, and he was makin' it for me and I was selling. I was the one selling it. I was kinda like the distributor of the manufacturing meth, and then I learned how the process was and then I started takin' over my own manufacturing. (Shukla 2016, 79)

While some interviewees intentionally changed their involvement and strategically developed contacts, others naturally slid into different roles as they became more involved in the lifestyle and needed money to support their growing habits. To illustrate, Mia, a thirty-nine-year-old former dealer who organized and facilitated cooks, states, "Before, I was just buying it for me. I was buying quite a bit. Of course, I'm sharing and partying with everybody. And then, I don't know what the transition was from me just buying it and partying to me getting rid of some of it. I don't really know why. Well I know why is because I didn't want to keep payin' for it" (Shukla 2016, 58).

People sold through personal contacts first, and some progressed to a wider distribution in time. For instance, stating that she started out selling only to people she knew, Allie, a thirty-nine-year-old former midlevel dealer, who organized and facilitated cooks, explains, "There were some instances where I would sell to strangers because they were friends of the friends. . . . Usually they were males, they were usually at least twenty-one [years old], because it'd be people that went to the bar with me or, and then my friends, family, and then they would have people that their people would call me and I got pretty involved, I was dealing with the KKK [Ku Klux Klan], I was dealing with a lot of people" (Shukla 2016, 60–61). Incidentally, Allie started dealing after reuniting with high school friends who were cooking.

Situationally, individuals were differentially positioned in terms of their contacts, motivation, and resources to identify and act on opportunities, accounting for why some people reported being more successful in changing their participation in the lifestyle—from using and paying for it to making money off the trade to support their own use. Personal contacts spanning different sets of people, either through their profession, partying, jail time, or transiency, provided locational advantages. These patterns are consistent with the hyperdyadic spread argument. At the community level, Christakis and Fowler (2009) argue that local neighborhoods integrate into a larger societal context by intermediaries spanning cohesive subgroups. Individuals are differentially placed within the network that shapes their behavior, as well as within the larger social system. One of the important emergent qualities of the larger social system is that the network reinforces inequalities, placing some people at a socioeconomic advantage over others (situational inequality) or at a locational advantage (positional inequality within the community network). Individuals who were reportedly involved in a higher volume of manufacturing, or who achieved greater profits, were better positioned in their networks compared to others who remained lower-level users and dealers. Situational and positional inequality were both important.

As you may have gathered from this discussion, using a social network perspective to understand criminal behavior does not require mapping exercises. I was able to illustrate the direct and indirect influences on methamphetamine use without drawing a picture. You might be tempted to conclude that qualitative inquires like the one described here are inherently unmappable, because the individuals mentioned in the narratives are not identified. If you came to this conclusion, you are wrong.

Visualizing Anonymous Networks

Qualitative inquiry produces rich narratives that are easily investigated with social network techniques, even when we do not have many specific details about the individuals involved. To map anonymous network information, we can use pseudonyms or roles. For example, when taking notes about lifetime involvement in methamphetamine use, interviewees can be asked to identify up to ten people they were closest at an important life milestone, without using specific names. Start with the people they were

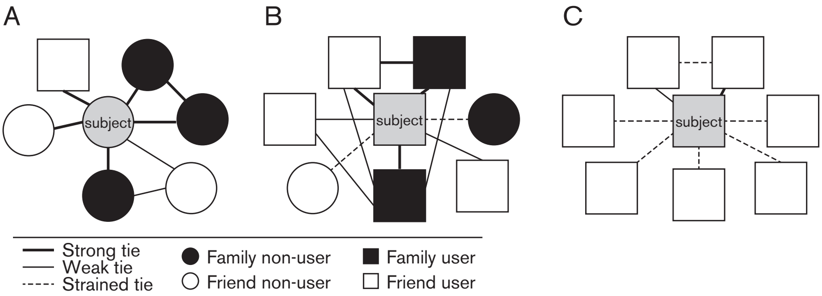


FIGURE 4.1. Dynamic visualization of anonymous networks. *A*, introduced to methamphetamine; *B*, begins intravenous use; *C*, decides to stop using.

closest to at the time they first used. Then follow with a question asking if any of these people knew each other. Figure 4.1, illustrates what this might look like in panel A. Note that black symbols depict relatives and white symbols are friends and associates; bold solid lines are strong relations, thin lines show weak relations, dashed lines represent strained relations, and square symbols indicate methamphetamine use. No specific names are needed to generate this network. The interviewee could nominate ten people but was close to six individuals at the time, so only six people are depicted. The interviewee's close confidants included family and friends.

To explore how the social network changed throughout the life course, repeat the line of questioning for different milestones (e.g., switch to IV use, period of most intensive use, decision to stop using). To make the networks comparable, ask the persons you are interviewing if any of the people they associated with in the prior time period were still present in their lives at the next juncture. Using the same symbolization, we can see how the networks evolve. Notice how different the networks look when intravenous use started (panel B) and when the person decided to stop using methamphetamine (panel C). As drug use deepened, familial relations became strained and ended, and nondrug users disappeared from their lives.

This illustration shows how it is possible to map an anonymous network in a meaningful way without compromising various ethical standards guiding research involving human subjects. As you will learn in chapter 7, plans to map anonymous networks must be developed before interviews start. Consequently, figure 4.1 is a hypothetical example and not a visualization of anyone interviewed by Rashi Shukla. Regardless, I hope this hypothetical exercise satisfies your curiosity.

SUGGESTED READINGS

I am not the first to argue for theoretical integration between theories of criminality and SNA. Several other efforts are exceptional and are arguably better than the cursory discussion provided in this chapter. Before launching your own SNA-inspired investigation of a preferred theory, consider consulting one of the following general discussions of theoretical integration:

- For criminological theories of differential association, subcultures, neutralization, and social bonding, with an eye to gangs and co-offending, see Sarnecki (2001, chapter 1).
- For criminological theories of differential association, social control, Krohn's network theory of delinquency, social disorganization, and differential social organization as they relate to delinquency, gangs, and organized crime, see Carrington (2011).

It is also advisable to consult an empirical study testing a hybrid SNA criminological framework, such as the following:

- For general theory of crime, see Young (2011).
- For differential association—mentoring, see Morselli, Tremblay, and McCarthy (2006).
- For social learning and social control theories, see Ennett et al. (2008) and Goldsmith and Brewer (2015).

In the next chapter, the discussion turns to theories of crime.

5. Connected Events

> Most people do not function as individuals, but have a network of family, friends and acquaintances. These linkages have varying attributes and influence decisions of others in the network. . . . The number of persons in the network varies. The intensity of the relationship between members varies. Readiness to make the decision to commit a crime is not constant: it varies from person to person; and it varies for each individual person across time and space as the backcloth or context varies.
>
> BRANTINGHAM AND BRANTINGHAM, "Crime Pattern Theory"

The compatibility of social network analysis (SNA) and theories about criminality may be self-evident, as criminality is often described within a social context. The detailed discussion of the hyperdyadic process of social learning that appears in the prior chapter underscores this point. What might be less obvious to some readers is the intrinsic connection between theories seeking to explain crime events and tenets of SNA theories. Many of the major principles and practices in environmental criminology and crime analysis are compatible with SNA concepts. For this reason, the material covered in this chapter is likely to appeal to officers, investigators, and analysts of crime and intelligence.

The chapter begins with a general overview of the theoretical congruence between two sets of theories and SNA: (1) deterrence and situational theories and (2) neighborhood ecological theories. While neighborhood ecological theories could have been covered in chapter 4, because of their regular integration with opportunity theories, I thought it best to reserve that material for the present chapter. Before providing an integrated model, I next examine two opportunity theories individually. The reason for this exercise will become apparent shortly. The chapter concludes with suggested readings.

OVERVIEW OF THEORETICAL CONGRUENCE

Theories of crime seek to explain why incidents occur, given that the individuals involved are known to engage in socially acceptable behavior much of the time. Scholars of crime events investigate what factors converge to trigger the event, why events exhibit distinctive patterns across time and space, and what can be done to prevent or suppress crime through opportunity-reduction strategies. Two sets of theories are covered in this overview: (1) deterrence and situational theories, which consider the individual decision making that results in decisions to commit crime; and (2) neighborhood ecological theories, which provide some indication of why spatial patterns emerge from these decisions at the community level. For both sets of ideas, as revealed below, there are grounds for theoretical integration with major tenets of SNA.

Deterrence and Situational Theories

Deterrence and situational theories, such as classic deterrence theory and rational choice perspective, suggest that crime results from a choice, wherein offenders perceive that the **benefits of crime** override the *effort* needed to commit the crime, *costs* associated with committing the crime, or the *risks* of apprehension and punishment. Offenders' decide to commit crimes on the basis of a desire to benefit themselves; they involve a bounded rationality that is constrained by time, the ability to process information, and the availability of information. As shown in table 5.1, the advantage gained by integrating network concepts is that SNA theory can provide an explanation for how rationality and behavior are bounded by the structure of the offender's local social neighborhood.

Perceptions of risk include ideas about the certainty, severity, and celerity (speed) of formal criminal justice system responses and informal sanctions. Individuals routinely gain information through involvement in activities and through information transmitted by associates, friends, and family. While classic deterrence and situational theories reference social sources of information, these perspectives do not fully acknowledge how instrumental local social networks are to perceptions of the risks and benefits of crime. Moreover, there is little discussion of how perceptions form. Integrating SNA provides a mathematical framework to explain how perceptions form and evolve.

Perceptions form through experience, as well as through information based on the experiences of others that is transmitted through the network. For instance, the principle of hyperdyadic spread suggests that influence extends up to three steps removed, with direct contacts having the greatest potential to sway perceptions. With each step away from the focal individual, influence declines significantly, meaning that a friend can shape your perspectives more than a friend of a friend, and so on. To illustrate, the details regarding a theft of supplies that a colleague shares with you will

TABLE 5.1. Select criminological theories of crime and congruence with SNA

Theory	*Central argument*	*Example of SNA theoretical congruence*	*Benefits of SNA*
Deterrence and situational theories (e.g., classic deterrence theory, rational choice perspective)	Commit crime based on desire to benefit oneself; decisions involve a bounded rationality constrained by time, ability, and information about benefits, costs, and risks	—Information gained through activities with others and information passed by others in local social network; changes in peer group membership and structure reflected in altered perceptions and behavior	—Explains how rationality and behavior are bounded by the structure of the egocentric network
		—Variation in criminal success may be attributable to social capital: people with local social neighborhoods having primary and secondary structural holes can access diverse information leading to "better decisions" and greater success	—Accounts for differential success among offenders with access to same information
Neighborhood ecological theories (e.g., social disorganization)	Communities lacking common prosocial values and collective efficacy can't fight crime and disorder because informal social controls break down; consequence is communities may (re)organize around criminal cultures	Communities with dense, strong ties have greater social constraint—stronger social institutions and more effective informal social controls; communities with sparse, weak tie structures exhibit more heterogeneity and change, increasing differences among groups and reducing effectiveness of informal social control	Provides mathematical framework to map and measure changes in multilevel social structures (e.g., family, neighborhood) and variable ineffectiveness of social institutions to exert informal social controls; updates concepts of neighborhood and community to reflect behavioral patterns of social interactions that often transcend spatial proximity

have a greater impact on your behavior than the same information would have on your sibling, or on others to whom your sibling passes the information. Notably, influences on decisions change throughout a criminal career, as a consequence of experience and events that result in changes in lifestyle and peer groups. Since access to opportunities and situations is also controlled by the people offenders interact with, changes in peer-group membership and evolution in the structure of associations within the group should be reflected in altered perceptions about the risks of crime, and ultimately criminal behavior. It follows that members of the local social neighborhood, both directly through the egocentric network and indirectly through the local sector (connections extending up to three steps removed), are the most meaningful actors influencing offender behavior.

SNA theory also offers an argument to account for differential success among a group of offenders, even when they have access to the same information. Variation in criminal success may be attributable to social capital. Recall from chapter 3 that people whose local social neighborhoods are characterized by primary and secondary structural holes have greater access to a diverse pool of information, which should lead to "better decisions" and improve success. When applied to offending decisions, greater success might mean selecting crime opportunities that afford more rewards with fewer consequences. On the other hand, cohesive networks with no bridging ties have less novel information and may result in poor decision making—that is, unsuccessful criminal behavior, such as getting caught and sanctioned for a crime or receiving little benefit from the crime (e.g., a low payout). Social capital theory also suggests that to maximize the benefits attained from maintaining an advantageous position within a social network (referred to as social capital), individuals must also have the human capital—that is the personal skills or abilities to use the benefit—and financial resources to act on the opportunities discovered through network contacts. Consequently, individuals with comparative social capital may not exhibit the same level of criminal success. With respect to conflict within a criminal group, being equipped with the relevant resources to act on an opportunity may be why one person emerges victorious in a battle for control within a criminal enterprise and another loses.

Explanations of crime rooted in situational theories may also benefit from integrating a network perspective. Specifically, social networks appear to be important in connection with rational choices (Cornish and Clarke 2016) and related arguments about the situational precipitators of crime (Wortley 2016). Generally, these perspectives suggest that criminal behavior is influenced by information gained through *routine interactions* with others within a *constrained situational context*. For this reason, changes in peer-group associations and life circumstances, such as marriage, alter knowledge boundaries and consequently affect criminal decision making. Additionally, some of the situational precipitators of crime—for example, provocations, such as crowding or territoriality, or the pressures to conform, obey, and comply with social demands—develop from social processes that can be modeled with SNA.

Theoretical integration is feasible because the fundamental tenets of situation perspectives reflect the five core assumptions of SNA outlined in chapter 3. For example, the situational perspective draws attention to the context of behavior, suggesting that situational interdependencies—the immediate circumstances and interactions with others and the environment—create the opportunities seized when crime occurs. SNA adds to this framework by alerting us to the idea that the actors' relative positions within the social network are part of the situational context of crime. If the five core assumptions of SNA are rephrased to reflect the situational context of crime (see chapter 3), it can be argued that

1. Situational interdependencies that shape the context of crime have greater explanatory power than offender characteristics.
2. Situational interdependencies have differential influence on offenders.
3. Situational interdependences are dynamic, and as a consequence, the opportunities these interdependencies bring continually evolve.
4. Offenders' awareness of situational interdependencies is bounded by their local social neighborhoods, an awareness that is also influenced by decisions and influences of unseen others who shape the local context.
5. Crime opportunities at the local level can persist despite changes in which actors are present, because the larger macrolevel social system within which local opportunities are embedded continues to exist—that is, drug markets remain despite the arrest of a dealer.

Integrating core concepts of SNA with the situational perspective stands to extend situational research in several directions. First, mapping the interactions between people and situational factors provides a standardized set of metrics to identify common elements of opportunity that can be used to compare behavior among different offender networks. For example, using a two-mode network linking people to elements of the modus operandi would make it possible to generate standardized assessments of the importance of situational elements common for many crimes, which could then be compared across jurisdictions. Second, investigating person-to-person relations could reveal how changes in decision making accompany evolving social contacts. To illustrate, it is possible to map co-offending partnerships before and after significant events, such as arrest or incarceration, to map how selection of co-offenders changes. Third, theoretical integration would also benefit event-based explanations for crime that explore the influence of the local social neighborhood. Mapping neighbor interactions (e.g., over the fence and over the road) before and after neighborhood redesign or street modifications would provide a way of assessing how these factors account for changes in fear and reported crime.

Neighborhood Ecological Theories

Deterrence and situational models are not the only explanations for crime that have SNA threads; neighborhood ecological theories also have network elements. Neighborhood ecological theories generally argue that community growth and development shape the social composition of neighborhoods and alter *social cohesion*. In turn, these affect the probability that the group develops or sustains a common set of prosocial values, as well as the *collective efficacy* required to resolve crime and disorder problems. Lack of homogeneity among community members disrupts informal social institutions, causing social controls to break down. As a consequence, criminal cultures emerge, around which the community (re)organizes.

From a network perspective, it is reasonable to suggest that levels of social cohesion will be reflected in the structure of ties among community members. Where communities exhibit a pattern of dense, strong ties, there are greater levels of social embeddedness. Strong ties form among people with much in common. Combining the homophily hypothesis with the tendency toward transitivity suggests that, over time, the networks of strongly connected people will overlap, resulting in greater social constraint (e.g., stronger social institutions and more effective informal social controls). On the other end of the continuum, communities characterized by sparse, weak ties experience greater heterogeneity and higher levels of change. These conditions reduce constraint, facilitate greater differences among groups, and reduce the effectiveness of informal social control. (Admittedly, these types of communities are at a distinct advantage—they are better able to adapt to changing societal conditions, such as an economic crisis.)

I see two advantages to formally integrating SNA tenets with neighborhood ecological theories. First, theoretical integration provides a mathematical framework, which can be used to map and measure changes in community social structure, collective efficacy, exposure to criminogenic influences, and the power of local social institutions, including family units. For example, social disorganization theory suggests that rapid population growth or change, persistent poverty, and heterogeneity at the neighborhood level can disrupt social institutions, causing informal social controls to break down and restricting the communities' ability to censure antisocial behavior. Mapping neighborly ties or other social interactions (e.g., public participation in community events, organizations, and governance) over time provides a window into how changes in social structure accompany population growth, transiency, or variation in population characteristics (e.g., poverty levels and heterogeneity). Using metrics of cohesion, clustering, and the like, it is possible to apply standardized measures to capture change in social structures that can be readily compared across different communities.

The second advantage afforded by adopting an SNA perspective is the possibility of updating the concept of neighborhood and community to reflect behavioral patterns of social interactions that often transcend spatial proximity. Modern life, lived online and in person, is spread out, often across different communities. People live in one neighborhood, work in another, and spend time with family and friends somewhere else. As a consequence, each person becomes embedded in different communities, such as social, familial, and employment. Some of these communities intersect, and if strong bonds form between people, the overlap with and integration of separate groups increase over time. Mapping social ties and applying subgroup identification metrics, it is possible to define neighborhood boundaries of distinct social groups, which may or may not reside in geographic proximity. SNA provides an opportunity to extend neighborhood ecological theory to explain deviance occurring in different kinds of social neighborhoods—cyber- and employment neighborhoods. Crime, however, is not spread evenly throughout neighborhoods. In the next section, the focus shifts to a set of complementary opportunity theories that explain how crime patterns emerge from routine behaviors.

A NETWORKED ACCOUNT OF OPPORTUNITY THEORIES

In this section, I discuss crime pattern theory (CPT) and routine activity theory (RAT), sometimes referred to as the routine activities approach. I decided to provide more coverage of these two opportunity theories for the simple reason that they are widely used, implicitly and explicitly, in practical settings—they directly impact crime control efforts, particularly among law enforcement agencies. Additionally, with the growing use of SNA within the intelligence community, and the integration of intelligence and crime analysis functionality, it seemed prudent to extend the discussion to link opportunity theories and SNA. Readers should note that the rules of CPT were renumbered over a set of successive publications. Thus, the order that appears here may not be consistent with prior versions.

Crime Pattern Theory

The role played by social networks in shaping criminal patterns is clear, as evidenced by two comprehensive explanations of CPT (Brantingham and Brantingham 2008, 2015). Arguing that "most people do not function as individuals, but have a network of family, friends and acquaintances," who influence their behavior, Brantingham and Brantingham (2008, 81) partially rooted their explanation of crime in graph theory (Brantingham and Brantingham 1984). (This is not a surprise if you know that Patricia Brantingham has a degree in mathematics.) CPT stresses that interactions with others shape behavioral patterns, thereby exposing offenders to opportunities and placing potential victims at risk. When the rules that form the basic tenets of CPT (see box 5.1) are paraphrased, it is surprising that the importance of social networks was overlooked for many years. CPT asserts that crime is not random, nor are events uniformly distributed—events cluster across n-dimensions (e.g., time, physical space, and hyperspace). Moreover, because crime event patterns are scale independent, in part because individual patterns can be aggregated, the ten rules of CPT apply to general and detailed analysis.

Historically, efforts to test CPT have focused on the geographic aspect of opportunity. At the heart of spatially contrived opportunity is **spatial awareness**, defined as the knowledge of areas traveled that fall within range of the locations where people engage in different activities (a.k.a. *activity space*). Spatial awareness emerges from routinized travel between central activity anchors (i.e., home, school, and work). Popular places, such as malls, recreation centers, or large institutions, are where the behavioral patterns of many individuals intersect, which means these locations shape crime patterns. Where the routine travel and activity of offenders and victims intersect, crime can occur—it is the intersection that forms opportunity because offenders can commit crimes only at or near the locations that contain targets they know about. It stands to reason that this theoretical framework supports thinking about offender behavioral patterns during the course of an investigation. For instance, Rossmo (2000) argues that we

BOX 5.1. CRIME PATTERN THEORY

The Brantinghams (2008) argue that nonrandom crime can be explained through the combination and interactions of ten rules.

Rule 1: Individuals participate in a set of routine daily activities that evolve in time and occur in different settings (physical and digital).

Rule 2: A dynamic multidimensional backcloth shapes crime occurrence by influencing decisions to commit crime and the routine activities of individuals.

Rule 3: Individuals function within a personal network of family, friends, and acquaintances who influence decisions—linkages have varying attributes. The personal network is connected through various communication mediums and types of interactions.

Rule 4: Participation in activities, occurring in physical and digital forums, shapes decision processes; regularization formulates a decision-making template.

Rule 5: Navigating among activity nodes generates general patterns with nested microactivity and movement regularities. These movements and experiences generate normal activity and awareness space.

Rule 6: Crime is located near normal activity and awareness space.

Rule 7: Crime occurs at the intersection of target and offender activity spaces.

Rule 8: Criminal actors act when they encounter targets that fit a crime template. Experience accumulates and modifies future behavior.

Rule 9: Summative patterns accrue from aggregating individual decision-making and behavioral patterns.

Rule 10: Structural dimensions of the backcloth, such as underlying networks (e.g., transportation, cyber navigation, communications), usage (e.g., land use, communication forums), and features of the built environment, generate regional activity nodes that may be crime attractors (cumulative impact of experience and network communication) or crime generators (cumulative impact of awareness and activity space of many people engaged in daily routines).

NOTE: The rules have been modified from a comprehensive explanation of CPT (Brantingham and Brantingham 2008) to integrate arguments about cyberlife offered by Patricia and Paul Brantingham (2015).

learn something about the offender's activity patterns when spatial clusters emerge in a crime series, and investigators can prioritize suspects on the basis of these spatial tendencies.

Visually, CPT is often represented with a diagram. In the first panel of figure 5.1, appearing on the left-hand side of the image, is a visual depiction

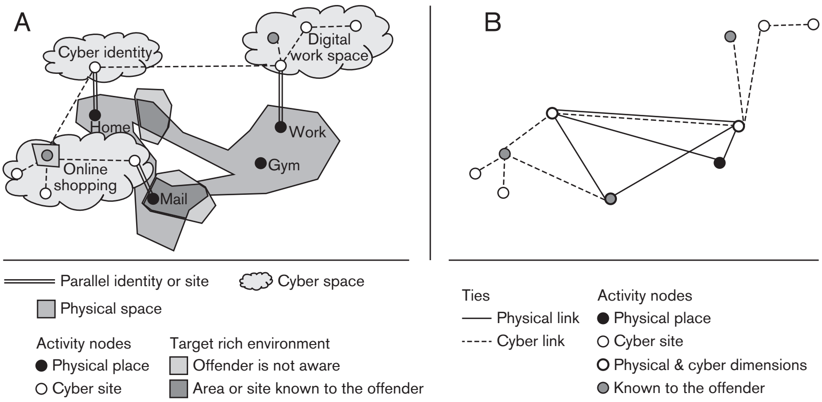


FIGURE 5.1. Connectivity across parallel activity spaces. *A*, activity space illustrated with a conventional diagram; *B*, activity space depicted as a network.

of an offender's **activity space,** defined as their most important places of activity, the area around these locations, and the paths taken to reach them (shaded gray). Within this region are two target-rich areas (shown as dark patches). (Please ignore the clouds for now.) The darkest part of the patch indicates portions of the target-rich environment known to a hypothetical offender—let's call him Nicholas. In this hypothetical scenario, Nicholas makes most of his money burglarizing cars. It is likely that he would select targets from either of these zones, as there are a lot of vehicles to choose from and he has had success in these areas in the past.

In thinking about current crime problems, Brantingham and Brantingham (2008, 2015) extend their notion of activity space to include cyberbehavior. Because modern life involves a considerable amount of digital interactions, the Brantinghams argue that we must consider the totality of activity when investigating crime patterns, assessing risk, or identifying target-rich environments. Updating the notion of activity space in light of modern life, understanding crime patterns requires investigation of physical *and* cyber-activity patterns, because victims and offenders may cross paths in either domain, or activity occurring in one domain, can spill over to the other. Connectivity across parallel-activity spaces is illustrated in figure 5.1.

Two sets of activity space are illustrated for the hypothetical burglar. Nicholas has a unique cyberidentity used for online activities, such as game play and shopping (now look at the clouds). He has another cyberidentity for his office job. Activity patterns associated with each digital identity merge into an *electronic-activity space,* and emailing materials back and forth between his accounts generates digital connectivity among his cyber-activities and two cyberidentities—shipping online purchases to home or work fuses his parallel-activity spaces. Moreover, a gaming store at the mall maintains both a brick-and-mortar location, as well as an ecommerce site. This illustration shows why the Brantinghams (2015) argue that routine activities giving rise to current crime patterns occur on multiple dimensions, which means we must rethink activity space to deal with crime in a digital age.

It does not take much to jump from a map illustrating activity space to a network map of connectivity. In fact, the Brantinghams told us as much years ago (1984!): to better understand crime occurrence, the Brantinghams suggested that we draw on graph theory, stating that while spatial techniques are useful, graph theoretical measures provide the "areal link between the sociological and geographical imagination," which integrates concepts of spatial and social structure (243). CPT rules clearly state that the selection of activities and emerging behavioral patterns are influenced by a personal network involving different types of interactions and use of various mediums of communication. Individual decisions about where to go and what to do are constrained by a social structure—an evolving network of family, friends, and acquaintances influences the daily and weekly activity patterns of offenders and victims. Since this network changes over time, as do primary activity nodes like school or work, modifications in activity and awareness spaces are to be expected (Brantingham and Brantingham 2008, 85). *In short, but for social networks, the behavioral patterns that shape crime occurrence would not exist.*

With the network origins of CPT in mind, take a look at the second panel of figure 5.1. The second panel depicts the **activity network**. First, ties are introduced to connect all activity nodes—dotted lines indicate cyberconnectivity and solid lines indicate physical travel between locations. Node attributes are modified to indicate whether the site is in physical space, cyberspace, or both. While the image only reports Nicholas's activity space, aggregating networks for many victims and offenders provides a way to examine where victimization concentrates. This is similar to the place-to-place network used by the P.I.V.O.T. team (described in chapter 1), the difference being that, in the theoretical example, physical and cyberactivity is combined (and I report on only one offender).

Despite the fact that Brantingham and Brantingham drew attention to networks a long time ago in one of their seminal books, the CPT-SNA link remained an implicit aspect of the theory until recently. Scholars got stuck on the first rule of CPT and did not fully integrate rules 1, 2, and 5. Noting this omission, I began to investigate how social-spatial crime patterns can be investigated with a CPT-SNA approach on two fronts—to account for the formation of intercity crime generators attracting juvenile delinquency (Bichler, Malm, and Christie-Merrall 2011; Bichler, Malm, and Enriquez 2014), and how this approach can be used to investigate serial offending by examining how the networked activity space of suspects and victims overlap (Bichler, Lim, and Larin 2017). If you intend to pursue research in this area, I strongly suggest exploring these articles first.

Routine Activity Theory

A close reading of the set of publications that have advanced routine activity theory since its first articulation in 1979 (see the note for box 5.2) highlights the critical role that social relations play in generating and preventing crime, as well as in sustaining co-offending relations. Synthesizing over thirty years of theoretical developments, I present RAT as a series of ten tenets for two reasons. First, there is a lot of material to integrate, and whether Felson acolytes agree or not, it is time an integrated version were presented. And second, I liked the argument presented by the Brantinghams (2015) that, to facilitate testability and advance computational criminology (another frontier of criminology), we need to formalize our theories by developing testable theoretical rules. To vet these ideas, Marcus Felson, John Eck, and Tamara Madensen reviewed the RAT tenets. No one objected to what appears in box 5.2!

In brief, RAT suggests that the general routines of daily life shape the **convergence** of elements needed for a crime to occur. Convergence and related factors foster crime according to the ten rules appearing in box 5.2. Inherent in Felson's use of the term *convergence*, the most central concept of RAT, is the notion that crime is a consequence of interacting components of a social system. Although not stated explicitly in network terms, each tenet requires an interaction, or lack of effective interaction, between or among actors. For instance, rule 1 necessitates the convergence of central actors. The social network accessible to the central actors (offenders, co-offenders, and victims), however, influences the extent to which convergence occurs devoid of agents of control. To clarify, the network affects the **intimate handlers'** ability to constrain the behavior of would-be offenders. Imagine that a parent enables tracking capabilities for their child's cell phone and has all of the cell numbers of their child's friends and their parents. The parent would be able to monitor movements and reach the juvenile, directly and indirectly, through others with whom the child is likely to interact during precursor activities that may lead to crime. The parent could

BOX 5.2. ROUTINE ACTIVITY THEORY

The modern incarnation of routine activity theory suggests that general routines of daily life shape the convergence of elements needed for a crime to occur. Popularized as a crime triangle, the theory has great appeal as both an explanation for crime occurrence and a template for thinking about crime prevention. At the core of the triangle is the act, which involves the intersection of a motivated offender, suitable target, and the place where the event occurs. The outer triangle contains the relevant social, political, economic, and regulatory agents that influence first-level controllers. What the crime triangle is lacking is that it does not clearly convey *how* the essential elements interact. To better understand interdependencies, consider how the ten rules listed below account for convergence and crime causation.

Rule 1: Crime events require the convergence of a motivated offender or co-offenders with a suitable target at a location, physical or virtual. At the same time, control agents—capable guardians, intimate handlers, place managers, and supercontrollers—must be conspicuously absent or ineffective. In "victimless" crimes, a second set of offenders obfuscates the need for a suitable target.

Rule 2: Convergence occurs within a setting, and settings are situated within communities that are part of a larger social ecology.

Rule 3: Crime events require materials and tools, and success is partially contingent on the availability of and access to these materials, as well as precursor (getting ready for the crime) and postevent (reaping crime benefits) situations.

Rule 4: Each element of the crime—offenders, targets, and settings—is associated with control agents. Only one control agent must be effective to prevent a crime from occurring.

Rule 5: There are at least two levels of control agents: one that is directly linked to the actors or crime site (level 1 controller) and a second that is linked to the level 1 controller (level 2 or supercontroller). Supercontrollers (e.g., regulators) can be situated within the specific setting or larger social ecology and need not be present at a potential crime site to exert influence.

Rule 6: Control agents can be animate (e.g., people, organizations) or inanimate (e.g., policy, firewalls, or cameras).

Rule 7: Control agents can monitor or exert supervision directly or indirectly, and supervision can be temporary (e.g., a bar manager watches a server for a few minutes) or systematic (e.g., algorithms that monitor credit card transactions for fraud).

Rule 8: The situational landscape underlying the interaction of crime elements and control agents involves three dimensions: social proximity (level of responsibility), spatial proximity (degree physical intersection), and temporal proximity (duration or length of interaction between parties).

Rule 9: The suitability of targets varies on many dimensions (e.g., target characteristics and situational attractiveness) that interact with the capabilities and willingness of guardians (e.g., willingness to supervise, ability to detect potential offenders, and willingness to intervene).

Rule 10: The management styles of control agents with sole or joint responsibility for supervising potential crime sites can be designated or self-imposed, and management styles (or interaction of the management styles if jurisdictions overlap) could have a preventative function or they could enable or promote crime.

NOTE: These rules differ from the original formulation in order to integrate many developments that elucidate elements of the crime or extend the theory, including Bichler and Malm (2015b); Clarke (1999); Cohen and Felson (1979); Eck (1995, 2002); Eck and Eck (2012); Farrell and Roman (2006); Felson (1995, 2003, 2006); Felson and Clarke (1998); Felson and Cohen (1980); Madensen (2007); Madensen and Eck (2008); Madensen and Sousa (2008); Reynald (2010); Sampson, Eck, and Dunham (2010); and Trembley (1993).

monitor activity during the event or after, as the group enjoys the spoils of its misbehavior (i.e., Mom has all friends' cell numbers and can confirm whether or not the child is actually in the purported location). By interacting with the child and taking action when suspicious, parents are able to exert influence. Such a situation increases the likelihood that the youth will be prevented from committing the crime. Similarly, the behavior of **suitable targets** is also influenced by their contacts. A target's suitability will shift the moment the person engages with or draws the attention of a **capable guardian**. Moreover, **place managers** can act, or fail to act, in ways that could prevent or promote interactions between offenders and victims. Integrating SNA provides standardized metrics to model the strength, duration, frequency, and timing of the interactions that may have preventative effects or, in their absence, causal ones. Before continuing, please read box 5.2.

The RAT-SNA link really comes to light in rule 3. Rule 3 holds that crime events require materials and tools. Assembling the instruments of crime may involve social interactions or transactions in advance of committing the act. To return to the methamphetamine example discussed in chapter 4, many of the people interviewed reported being involved in rounding up precursor chemicals (Shukla 2016). Before the passage of the Combat Methamphetamine Epidemic Act (CMEA) of 2005, individuals were able to make large purchases or repeated transactions at the same pharmacy. After the act passed, individuals were forced to extend their networks and purchase smaller amounts from different pharmacies, so as to avoid detection. Securing enough materials often required more coconspirators engaging in multiple trips, sometimes out of state. Social networks are also critical in the aftermath of crime. Consider the need to sell the methamphetamine produced during the cook, to use the drugs acquired in a safe location, or to fence stolen property traded for drugs. In each instance, one's contacts are cited as playing a pivotal role.

Rules 2 through 5 invoke the larger community context associated with contacts beyond the ego network. Aggregating the social networks of a set of offenders and victims provides a means to map the setting and social ecology supporting crime occurrence. Understanding how a particular crime and criminal justice response, perhaps the arrest of a local methamphetamine producer, fits within the larger context of drug supply in a region, better positions agents of control to target their resources for maximum disruption. (If this prospect excites you, I am certain that you will enjoy the next chapter.)

Rules 6 through 10 synthesize current thinking about the characteristics and functions of essential ingredients of crime. Reading each rule, you will quickly notice a theme. Element characteristics, whether they refer to the suitability of targets, capabilities of control agents, or situational landscape, rely to some extent on direct or indirect interactions. Network methods and analytics can model all forms of interactions, whether they involve animate actors (e.g., individuals, groups, or organizations) or inanimate mechanisms (e.g., firewalls). Moreover, the proximity, duration, frequency, and intensity of interactions can be calibrated. With a little imagination, it is possible to operationalize theoretical rules with network-oriented analytics. Because explaining the technical aspects of how this can be done is outside the purview of this work, our conversation now turns to a conceptual example that illustrates how network mapping is beneficial for problem diagnosis.

One of the primary benefits of applying a social network framework to RAT is that doing so can focus crime problem diagnosis in a way that zeros in on the opportunity mechanisms that are instrumental for the crime to occur. Efforts to understand how routine behaviors result in crime typically begin with a diagnostic exercise focused on identifying all actors or system elements crucial to crime occurrence by using the crime triangle (see Center for Problem-Oriented Policing, n.d.). While some attempts to classify and assign actors or system elements to the sides of the triangle are easily accomplished, it is often the case that essential elements are missing or are one and the same. For instance, during an investigation of high-value art crime, I came across a tax fraud scheme, wherein a dealer sells art and then acts as the "independent" appraiser, who inflates the value of pieces when his client donates the objects to a charity for a tax credit. The dealer then repurchases donated art from the charity at a discounted price, only to launch the scheme anew by reselling it at another time. In this scenario, the dealer is both the offender and the industry expert appraising value (intimate handler), who is supposed to ensure a system of checks and balances. Categorizing roles in this instance is useful, but it is less informative than mapping out the scheme.

By adding a second step in the diagnostic process that scripts in detail how the crime occurs (we will talk more about crime scripts in an upcoming chapter; for now, think of story boarding), we can map out how various actors and system elements interact in furtherance of a crime. For example, we can represent interactions and directionality of the scheme with edges (lines), connecting individual actors, key controllers, and organizations using different symbols or colors to indicate their roles in the scheme. Although it is a more complex process, mapping the elements of the crime using a network framework helps to pinpoint where interventions could be targeted. Using a standardized coding scheme makes it possible to map and aggregate the schemes used in many case studies, which in turn supports the development of crime control policy on a broader scale. A new concept raised in this discussion is the idea that crime opportunity can be thought of as a network. In the next section, I present an explanation of this concept and offer a set of rules, together with suggested hypotheses, to guide investigation of the structure of crime opportunities.

AN INTEGRATED THEORY OF NETWORKED OPPORTUNITY

As noted previously, I asked Marcus Felson ahead of publication to review the tenets of RAT that I generated for this chapter. My attempt to develop a general set of testable "rules" was the first of its kind, at least to my knowledge, and I thought it was best to check with the master. His response was interesting. While he did not object to what I wrote, Marcus stated that he had come to see the two opportunity theories as one and the same, differing only in their emphasis and points of departure. For instance, RAT identifies the essential elements of a crime and CPT aims to explain how and why the elements of a crime converge. He suggested that I merge the two sets of rules to develop one, unified, testable set of statements. What follows takes the challenge one step further by presenting a unified list of rules that integrates major tenets of crime pattern theory, routine activity theory, rational choice theory, and arguments about situational precipitators with SNA. The proposed integrated theory of networked opportunity does not replace any of the opportunity theories; rather, it is my attempt to think about how social networks shape opportunity.

Table 5.2 presents the eleven testable rules of the integrated theory of networked opportunity. To avoid rehashing the theoretical integration already discussed in this chapter, I note where the ideas fit—see the "origins" column. Notably, I did not specify which social network theory I drew from. To illustrate how to test the rules, I included example hypotheses. Please take a few minutes to look over the table before continuing.

The first thing to notice about the table is that it is divided into three sections—crime events, personal networks, and social systems. The first four rules appearing in table 5.2 account for the patterns of crime. Crime events cluster, irrespective of how behavior is measured. Moreover, clusters are found in networks at all scales of analysis. Why? Because personal networks and social systems interact and generate a highly skewed distribution of exposure to and information about crime opportunities, the capacity to act on perceived opportunities, and the ability to reap crime benefits. To understand how these skewed distributions form, we need to acknowledge that social networks include two interdependent elements—personal networks and the greater social system that accrues from the integration of personal networks. Together, these two interdependent forces generate the situational interdependencies (context) within which the elements necessary for precursor activity, crime events, and postcrime activity converge (recall our discussion of situational conditions from the beginning of this chapter). The RAT roots of this rule are obvious. As we learned from RAT, the necessary elements include suitable targets, motivated offenders, and the conspicuous absence or ineffectiveness of control agents (intimate handlers, place managers, capable guardians, and super controllers). These statements can be tested with network analytics. Studies can also investigate the supposition that crime-related perceptions, attributes, functions, behavior, and the success of individuals vary significantly across the social system. For instance, it is possible to investigate whether significant clustering is present among actors, within subgroups, or social systems.

Rules 5 to 8 clarify the role in crime causation that is played by personal networks. Rule 5 explains that networks form through the routine activities of people. By activities I include actor-to-actor interactions, as well as engagements that might generate potential interactions like participation in events, associations or memberships with groups or organizations, or public discourse (e.g., posting on social media, publishing a report). Routine activities have a locational component because they occur in time and physical or cyberspace.

TABLE 5.2. Integrated theory of networked opportunity

Theoretical Tenet	*Origins*	*Example network hypotheses*
Crime Events		
Rule 1: Crime events cluster.	CPT, SN	Clustering coefficients and contagion models show significant crime clustering.
Rule 2: Clustering is scale invariant.	CPT, SN	Significant clustering is present among actors, within subgroups, or within social systems.
Rule 3: Clustering occurs because social networks generate highly skewed distributions of (a) exposure to information about opportunities, (b) capacity to act on perceived opportunities, and (c) ability to reap crime benefits.	CPT, RAT, SN	Crime-related perceptions, attributes, functions, behavior, and success of individuals vary owing to structural differences in local personal networks.
Rule 4: Social networks are made up of two interdependent elements—personal networks and an emergent social system—that generate situational interdependencies, enabling precursor activity, crime events, and postcrime activity.	CPT, RAT, SP, SN	Personal networks and social systems, facilitating greater convergence of necessary elements, exhibit higher levels of precursor activity, crime, and successful postcrime rewards.
Personal Networks		
Rule 5: Personal networks form through routine activities.	CPT, RAT, SN	When patterns of social interaction change, network structure changes.
Rule 6: A personal network extending up to three steps from the focal individual exhibits the most influence on crime-related perceptions, decision making, and behavior.	CPT, RC, SN	Individuals with more redundant ties to others involved in crime or experiencing victimization also exhibit high levels of criminal behavior (or victimization); networks devoid of potential control agents increase the rate of a victim's or offender's involvement in crime.

Rule 7: Personal networks are dynamic.	RC, SN	Structural changes in the social network result in modified perceptions, decision making, behaviors, and alterations in individual attributes.
Rule 8: Social networks and spatial behavior can correlate.	CPT, RAT, SN	Victims and offenders have overlapping spatialized activity networks, to the degree that they share links to the same social (people) or physical and cyberanchors (places).
Social Systems		
Rule 9: Social systems with emergent crime characteristics accrue from the aggregation of personal networks of community members.	RAT, SN	Removal of randomly selected actors may not affect functionality of the system.
Rule 10: The social system generates crime-related inequalities among individuals.	RAT, SN	Few actors have the highest levels of victimization, exposures, capacity to act, and success.
Rule 11: The environmental backcloth exerts influence on the social system.	CPT, SN	Current and historical political, legal, and economic context influences current social system characteristics and evolutionary change.

NOTE: CPT = crime pattern theory; RAT = routine activity theory; RC = rational choice theory, SP = situational precipitators; SN = social network theory.

Drawing heavily on the concepts of hyperdyadic spread (Christakis and Fowler 2009) and social capital (Burt 1992), rule 6 explains how people are influenced by their personal networks. Central to the argument is that an individual's personal network controls access to information and resources, exposure to criminogenic situations, and ties to potential co-offenders or collaborators, all of which affect perceptions, decisions, and behaviors that are associated with involvement in crime (as victim, offender, or potential control agent). For example, from this point it can be suggested that networks devoid of potential control agents increase the rate of involvement in crime as a victim or offender. Extrapolating from network theory, I argue that individuals with a greater portion of redundant ties to others involved in crime or experiencing victimization also exhibit high levels of criminal behavior (or victimization) and that individuals, both victims and offenders, who are strongly connected have overlapping spatialized activity networks to the degree that they share alters. Overlap increases exposure to crime.

Rule 6 is richer than it first appears, once you consider all of the underlying suppositions about relations. For instance,

A. Personal networks can include direct ties to crime involved and noncriminal family, friends, and acquaintances and indirect ties to others accessible through direct ties.

B. The first level of influence, direct contacts, is greater, than secondary and tertiary contacts (second and third steps removed from the focal individual).

C. Individuals have little knowledge of their personal networks beyond direct contacts—and yet, rationality is bounded by the edges of their local social neighborhood, which channels information from unknown individuals.

D. Spatial proximity and attribute homogeneity (homophily) strengthen associations and potential interactions (i.e., among co-offenders and between offenders and victims).

E. The term *behavior* includes social involvements, travel to activity sites, and participation in crime-related activity, as well as precursor and postcrime activity.

F. Crime influences are amplified by the level of network cohesion and redundancy of ties.

Rule 7 states that personal networks are dynamic, meaning they are in a continual state of purposeful and nonintentional change. When they are examined over time, therefore, we can expect criminal relations to vary in duration and strength—frequency of contact, intensity, intimacy, and reciprocity. Also, relations vary in immediacy and proximity. Since ties form and dissolve, so too do the situational conditions conducive to crime form and dissipate. Notably, the rate of change varies by person and throughout the life course. This leads me to hypothesize that structural or topographic changes in the social network will result in modified perceptions about crime, decision making, and behaviors, as well as in alterations in individual attributes.

Rule 8 notes how the correlation between spatial and personal networks generates convergence among the essential ingredients of a crime. Keep in mind that spatial activity networks include physical and cyberplaces. Overlapping personal networks reveal the convergence of victims and offenders, and the lack of convergence of potential control agents. Individuals, both victims and offenders, have overlapping spatialized activity networks to the degree that they share links to the same social (people) or physical and cyberanchors (places). In keeping with current arguments proposed by the Brantinghams (2015), convergence involves interactions that can be asynchronous or synchronous, reciprocal or not, occur in person or remotely, involve direct or indirect contact (e.g., through the internet or communication media), or include singular or repeated contact.

Rules 9 to 11 bring us to the social system and its influence on crime. This set of rules is critical to understanding how crime opportunity is networked at the community level. First, rule 9 states that social systems accrue from the aggregation of personal networks of community members, and that they have emergent qualities, separate from the people and subgroups within them, which influence crime patterns. In practical terms, criminogenic conditions may emerge and be measurable at the community level, when they are not readily apparent at the individual level. For instance, transnational crime may not be visible when individuals are monitored in a local context, but the aggregation of individual behaviors may reveal crime at the community level. In other words, selling marihuana in California may be legal, but the behavior is criminal if the seller supplies large amounts to someone who, unbeknown to them, transports pot to Mexico. If social systems have crime characteristics and functions that are different from individual-actor networks embedded within the system, removing a specific actor may have no effect on the functionality of the system. (A problem demonstrated over several decades by antidrug policies aimed at retail distributors.)

Rule 10 states that the social system generates inequalities among individuals. Much like the influence of personal networks, social systems produce highly skewed distributions among subgroups of actors, in their ability to avoid victimization or act as a control agents, as well as in their exposure to crime opportunities, capacity to act on opportunities, and potential to reap crime benefits. Since these distributions are highly skewed, few subgroups have the highest levels of victimization, exposure to opportunity, capacity to act, and criminal success.

Rule 11 introduces a key concept from CPT—the *environmental backcloth*. In keeping with the spirit of the Brantinghams' definition, but adding a network orientation, I argue that the environmental backcloth is the byproduct of the historic social system. Manifested as the current legal, political, and economic context, as well as the existing built environment, the backcloth shapes the characteristics of the current social structure and plays a part in its evolution. In other words, the conditions generated by a social system at time 1 will influence the social system structure at time 2—a prior social system generates exogenous circumstances that shape how the current system behaves. For example, if I wanted to investigate how the illicit international trade in small arms and light weapons has evolved over the last ten years, it would be important to control for political alliances, conflict, change in national wealth, and the like.

While I could continue, I think it is best to conclude this chapter. You may be ready for something less abstract and more applied. But, before we move on, I have some recommendations.

SUGGESTED READINGS

Of relevance to this section, Carrington (2011) provides an informative discussion of the network implications of social disorganization theory and differential social organization theory. Empirical examples include the following:

- *Social Ecology Theories*—neighborhood effects and spatial patterns: Hipp and Boessen (2013); Hipp, Faris, and Boessen (2012); Tita and Radil (2011); Papachristos, Hureau, and Braga (2013)
- *Opportunity Theories:* Bichler, Lim, and Larin (2017); Bichler, Malm, and Enriquez (2014)

6. Who Is Who?

> Existing concepts from the discipline of network analysis have been shown to be relevant to the analysis of criminal intelligence. These include several different notions of centrality and of equivalence, and the concept of weak ties. There are many other network analysis concepts which might turn out to be useful also.
>
> SPARROW, "The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects"

> The use of social network analysis in criminology is in its infancy. . . . Nevertheless, a small number of criminologists are knowledgeable in the concepts and methods of social network analysis, and some have shown great ingenuity in finding or generating suitable data. They have produced a number of sophisticated and powerful criminological network analyses over the past decade.
>
> PETER J. CARRINGTON, "Crime and Social Network Analysis"

Thinking about the context underlying the two opening quotations, the situation can be summarized as follows. First, studying criminal networks often necessitates access to criminal intelligence or arrest data. For this reason, practitioner-academic partnerships are crucial. A direct consequence is that crime control and prevention needs are shaping how this field develops. The principal drivers are applied academics, often working in concert with, or to inform, network-disruption efforts. Second, even though networked criminology stands to contribute a great deal to crime control and crime prevention efforts, the lion's share of scholarship remains descriptive. Recent developments, however, suggest that a tipping point is being crossed, and thus the situation may very well have changed by the time you finish reading this chapter. Third, while the application of social network analysis (SNA) to crime is relatively new, a rapidly growing list of active and notable scholars has in a very short time generated important streams of research in the emerging field of networked criminology. Anyone embarking on work in these areas must be aware of this foundational work. Consequently, I have two modest objectives for this chapter.

First, I will document the speed at which the field is growing. Since a list of current research will be outdated by the time this book goes to print, my second objective is to provide examples of notable scholarship shaping inquiry into three dominant streams of research—the nature of co-offending, the structure of criminal groups, and the disruptability of criminal enterprise. The discussion highlights many prominent criminal network scholars, but it should not be mistaken for an exhaustive foray into the current state of criminal network research. Instead, the chapter acts as a guide to get you started. (You can think of this limited literature review as another list of suggested readings.) Keep in mind that the suggested readings noted in prior or subsequent chapters are not repeated here.

RAPID ONSET

Several people have acknowledged the speed at which SNA is being applied to crime-related topics. Two statements stand out. Having been a network-oriented researcher since the 1990s, and a founding member of the Illicit Networks Workshop (described in chapter 11), Carlo Morselli is one of the best people to consult regarding growth in this field. Professor of criminology at the Université de Montréal, Carlo investigates criminal networks and organized crime, most recently illicit markets and collusion within the construction industry. Often his work has practical applications in support of the Service de Police de la Ville de Montréal (Montreal Police Department). Looking back roughly twenty-five years, he comments that it is amazing how things have changed. Citing a presentation by Sean Bergin from the first Illicit Networks Workshop (Wollongong, Australia 2009), Morselli (2014) reports that the annual rate of publications about criminal networks witnessed between 1975 and 2008 tripled in 2001 and doubled again in 2006. These tipping points are interesting: in 2001, the 9/11 attacks occurred, which resulted in a wave of network-oriented studies of terror groups; and in 2006, the first special issue of the journal *Global Crime* that was dedicated to criminal networks was published. Admittedly, though Morselli did not conduct an empirical assessment, his anecdotal experience reviewing manuscripts for journals, editing, writing, and organizing conferences confirms that the growth trend is continuing (2014).

Andrew Papachristos, professor of sociology at Northwestern University, is a prominent scholar of social networks, neighborhoods, street gangs, and interpersonal violence. Involved in the evaluation and implementation of gang and violence reduction strategies (e.g., Project Safe Neighborhoods and the Group Violence Reduction Strategy in Chicago), Papachristos (2011) laments that, despite the rapid uptake of crime-related network research, there is room for improvement. Examining the rate of publications in top sociology, public health, and criminology journals between

1980 and 2010, Papachristos (2011) observes that criminology lags behind other disciplines but exhibits growth with cyclical spikes caused by the productivity of a couple of crime-oriented researchers.

In an effort to update these perspectives, I did an impromptu study. Searching Google Scholar under "social network analysis" combined with three other terms produced the following number of hits: "crime" (17,547), "delinquency" (3,597), and "criminal networks" (1,680). In order to examine the use of SNA theory, for my last search I looked for a specific theoretical construct, combining "social capital" with "criminal networks," which produced 586 hits. It is noteworthy that the search parameters did not include patents or citations, the terms could appear anywhere in the page or searchable document, I included hits from 1991 to 2016, and all searchers occurred on September 9, 2017.

Panel A of figure 6.1 illustrates the distribution of hits over time. Looking at the results of the first search—combining the term "crime" with "social network analysis" —the rapid increase in scholarly material is hard to miss. Because I did not check the nature of each of the 17,547 hits, I am not certain how many of these items pertain to empirical studies; therefore, it is more appropriate to interpret this result to mean that there are simply more materials that mention crime and SNA together. For instance, an author might speculate that it is a crime that SNA is not used more. In this example, the document does not actually refer to a crime-related SNA study. In order to better assess the rise of networked criminology, three more specific searches were conducted.

Panel B offers a better look at the results of searching on the more specific phrases. The trends were striking. The number of delinquency-related materials tripled from 2003 to 2008, and it doubled again within the next five years. From 2005 on, the number of hits including the term "criminal networks" doubled every five years. While it is not feasible to draw direct comparisons with Sean Bergin's findings, it is safe to say that dramatic growth has continued since his report in 2009.

In the final search, I looked for an uptake in social network theory. This exploration is important in light of the Peter J. Carrington quotation that opens this chapter. Carrington, professor of sociology and legal studies at the University of Waterloo, has a long-standing record of research applying SNA to understanding crime and delinquency. In his seminal 2011 book *The SAGE Handbook of Social Network Analysis*, Carrington published a key chapter[1]

1. *The SAGE Handbook of Social Network Analysis* (2011), edited by John Scott and Peter J. Carrington, brings together the best thinking about social

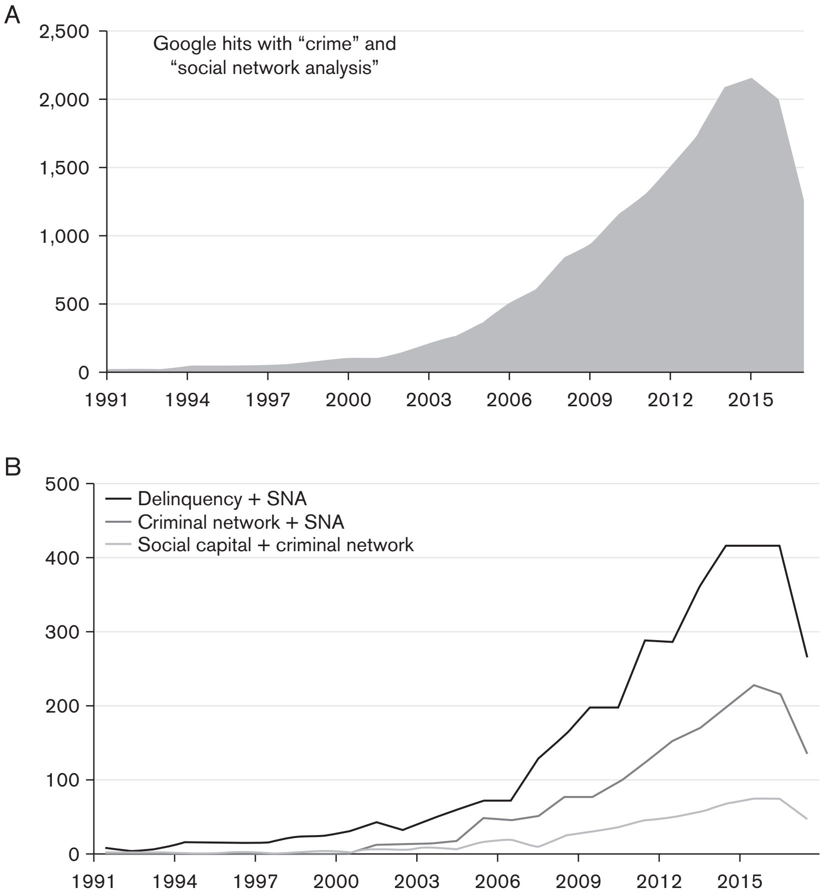


FIGURE 6.1. Frequency of SNA and crime-related search term hits on Google Scholar. *A*, all search terms; *B*, search terms were "delinquency," "criminal network," and "social capital" with "SNA" or "criminal network."

that reviewed the state of networked criminology, exploring the theoretical intersection between criminology and SNA. In doing so, he undertook a review of the network-oriented criminological literature of the day. As indicated in his opening quotation, Carrington (2011) concluded that while most studies had not gone beyond simple descriptive analysis, some powerful and

networks by leading researchers and has been lauded as the first attempt to encapsulate the SNA paradigm. It reviews the discipline and is well worth the price tag. Importantly, proceeds from the sales fund scholarly prizes awarded by the leading professional association, the International Network of Social Network Analysis.

sophisticated modeling had been done and change was afoot. To follow up on this idea, I looked for criminological studies that sought to examine an identifiable theoretical construct—Burt's (1992) notion of social capital. Searching on "social capital" combined with "criminal networks" revealed a modest number of hits and a steady upward trend since 2005. Moreover, since 2011, there have been no less than forty-nine scholarly materials, including both phrases. It is safe to say that at least one SNA theory is being used regularly to understand crime-related phenomena.

What does not appear in figure 6.1 is that the field may have reached a developmental milestone. The year 2017 marks the first publication of a formal effort to synthesize the results of a large set of studies. While methodological variation precluded a meta-analysis, Bichler, Malm, and Cooper (2017) succeeded in completing a systematic review of thirty-four studies describing the structure of fifty-four illicit drug supply networks linked to organized crime groups. This study marked an important milestone in the field, because systematic reviews and meta-analyses are possible only when enough scholarship is available on a specific research question. The field must be mature enough to have reached this threshold. I hope that this is not an isolated case and that more will be published soon! (Incidentally, if I just piqued your interest in the topic of organized-crime involvement in the drug trade, I will explore these findings in detail later in this chapter.)

CURRENT UNDERSTANDING

The rapid uptake of SNA to investigate criminality and crime is encouraging, particularly when studies use an integrated theory. In the next section, I explore criminal network research in three prominent research streams: (1) co-offending, (2) the structure of criminal groups, and (3) the disruptability of criminal enterprise. These broad categories hide the variability of scholarship: several subareas or topics appear in each research stream. Since reviewing all of the relevant literature was not feasible, I chose one study to profile for each topic. These examples constitute a good entry point into the literature.

Co-offending

A substantial portion of networked criminology investigates the prevalence, structure, and nature of co-offending. The six studies described below represent the dominant areas of inquiry with one caveat: studies using co-offending information to investigate specific forms of crime, street gang violence, drug trafficking, and organized crime, are not included. Instead, this selection of studies illustrates what can be gleaned from a focus on understanding co-offending behavior in a general sense. Most of the studies rely on police data, a widely acknowledged limitation in the field.

Prevalence and Structure

If you are interested in studying co-offending, begin with a close reading of Sarnecki (2001). In addition to providing a general overview of the prevalence and structure of co-offending in Stockholm, this book examines patterns in the selection of co-offenders by age, ethnicity, gender, residential location, and offense type, as well as offering a detailed inspection of hooliganism, of the influence of spending time in a secured facility, and of formal gang membership. In short, this book introduces readers to all streams of co-offending research. The study uses police data to map a community of 22,091 people involved as co-offenders (45,676 co-offending ties). The analytic strategies used and the presentation of findings are accessible, and no advanced statistical knowledge is necessary to understand the results. Sarnecki (2001) offers a number of influential findings:

- Most co-offense pairs were of a similar age (76 percent were within two years of age), 89 percent were the same sex, and 67 percent lived in the same postcode ward.
- Twenty percent of the study population (4,130) were part of a large central network that included 83 percent of all co-offense links.
- There was a weak correlation between the number of known co-offenders and number of offenses committed.
- Co-offending relations were unstable and likely to involve weak ties: youths involved in continued delinquency tended to select new co-offenders.

Offense Specialization

The next study looked at offense specialization. McGloin and Piquero (2010) randomly selected 400 individuals (200 lone offenders and 200 people offending in a group) from a list of 60,821 juveniles under eighteen years of age, who were arrested in 1987 in Philadelphia. Offense histories were gathered from January 1976 through December 1994. Selecting a subgroup of offenders arrested for at least two co-offenses during this time, the authors used offense histories to generate egocentric networks for each of 218 persons. Each network included the focal person's co-offending partners and the co-offenses of partners. Then diversity indices were created for each observation (year of age) to rate the ego's offending patterns and the co-offenders' patterns on the basis of three crime categories—violence, property, or drug-related crime.

Controlling for frequency of offending and age of onset, McGloin and Piquero (2010) found that density was a significant predictor of specialized co-offending. When density was observed to increase, meaning that a greater portion of the focal individual's co-offending partners engaged in crime with each other, higher levels of offense specialization were found (less diversity). In other words, juveniles who had more redundant co-offending networks (recall the theory of social capital presented in chapter 3) were more likely to engage in the same kind of offense type when co-offending.

Spatial Proximity

Another topic of interest is spatial proximity. Schaefer (2012) explored the spatial proximity of a random sample of juvenile co-offenders detained in Maricopa County, Arizona. Maricopa County includes Phoenix and several other cities and was home to about four million residents at the time of the study. Information about juvenile detentions was extracted from Arizona's Juvenile Online Tracking System. Arizona's tracking system is a statewide record containing all official contact between youth and juvenile justice personnel. The random sample consisted of 10,629 youth detained in 2000 (about half of that year's juvenile arrests). Collectively, this group committed 7,723 criminal offenses. From this sample, Schaefer identified 3,058 co-offending relationships between youth, of which 72 percent involved co-offending between people living in different census tracts. Home address at time of arrest was used to identify co-offender proximity. Measuring the Euclidean distance between middle points of each census tract, Schaefer (2012) found that, on average, co-offenders lived 4.9 miles from each other (SD, 5.2). A multivariate model controlling for individual and area-level characteristics found that spatial factors, such as attending schools in the same district, were significantly related to co-offending.

Homophily

Homogeneity in offender characteristics is another popular topic. In SNA terms, homogeneity is referred to as homophily. Investigating the issue of homophily, van Mastrigt and Carrington (2014) drew a sample of 18,494 co-offenders (30 percent of offenders) from cases detected during a three-year observation period by a large police force in the United Kingdom. Linking co-offenders resulted in 10,997 groups, the unit of observation for the study. Co-offending groups averaged 2.4 people (group size ranged from 2 to 20). Of the groups, 68 percent were exclusively male, 53 percent were made up of adults (eighteen or over), about 39 percent were juveniles, and the remaining 17 percent included people of multiple age groups. Homophily was found to be significantly higher than expected on all three factors with some notable variations. For example, within-group sex homogeneity was higher among males than females.

Peer Effects

While the study by Haynie, Doogan, and Soller (2014) is not an example of co-offending research per se, because it investigates delinquency, it does provide insight into peer effects. The influence of peers and mentors is an important facet of offending, as well as of the co-offending dynamic. Tapping into the network elements of the AddHealth study, Haynie and colleagues examined the behavior of 1,857 adolescents attending two large schools.[2] Adolescents nominated up to ten of their closest friends during interviews for wave I and again at wave II. Aggregating ego networks for each respondent by school, and then comparing wave I and wave II, provides a way of looking at how the networks changed over the course of a year. The authors used stochastic actor-oriented models to investigate whether girls, compared with boys, are more influenced by, or are more likely to select friends on the basis of, their friends' violent or delinquent behaviors. The results suggest that, while all adolescents tended to select friends according to their friends' behavior, girls exhibited a stronger tendency to do so, leading the authors to conclude that friends' involvement in violence or delinquency is a decisive factor for girls when selecting friends.

Victimization

No discussion of co-offending is complete without mention of work led by Andrew Papachristos. Focusing on gun violence, particularly within communities facing entrenched gang problems, Papachristos applies SNA concepts from Christakis and Fowler's (2009) hyperdyadic spread model to

2. The National Longitudinal Study of Adolescent Health (AddHealth) is a research program designed to investigate health outcomes and behaviors of young people. Designed by J. Richard Udry and Peter Bearman, and funded by a grant from the National Institute of Child Health and Human Development (HD31921), this study involves a nationally representative, saturated sample of ninety thousand respondents drawn from 130 schools. In-school surveys were completed between 1994 and 1995. A subset of in-home interviews generated network data about friends; twenty thousand adolescents were interviewed in wave 1 and almost fifteen thousand completed interviews in wave 2.

investigate risk for homicide and gun-related injuries (see, for example, Papachristos 2009, 2013; Papachristos, Hureau, and Braga 2012). Investigating the contagious properties of violence with logistic regression models, Papachristos and colleagues found that the risk of being involved in gun violence is significantly related to one's position within a social network of risky behavior: "Gangs, as groups, and gang members, as individuals, are not partitioned off from their communities, socially or criminally. The lives of gang members are woven into the larger social fabric of their neighborhoods, social networks, families, and friends (Papachristos, Braga, Piza, and Grossman 2015, 627).

Moreover, while being a gang member significantly increases the odds of being shot, risk is also transmitted to those who are socially proximate to gunshot victims. Using eight years of police reports of fatal and nonfatal shootings to investigate risk transmission, Green, Horel, and Papachristos (2017) mapped a community of co-offenders—the main component included 138,163 people. The results showed that risk of victimization transmits to others located within three steps of a gunshot victim, and victimization of these other people occurred on average about 125 days after exposure (after an associate was shot).

Criminal Groups

It should not be a surprise to learn that a large amount of criminal network research concerns the structure of criminal groups. Practitioner-academic partnerships generate a substantial portion of this work. Admittedly, with so many to choose from, it was difficult to narrow the list to six studies. Two topic areas are popular—organized crime group structure and computer crime/hacking structure. Of note, much of the scholarship investigating criminal group structure concerns groups involved in the illicit drug trade, particularly cocaine, heroin, methamphetamine, and marijuana. I will return to this issue after a quick tour of criminal group structure. The studies reported on below are representative of the variability of data sources used to describe the structure of criminal groups.

Political Corruption and Conspiracies

Baker and Faulkner (1993) were among the first to acknowledge that covert networks do not necessarily behave like normal social networks. Illegal networks need to maximize concealment, suggesting a preference for a decentralized and sparse network structure. And yet, those involved may also need efficient communication to accomplish tasks. These authors theorize that the complexity of the product and market activity will influence unique information-processing requirements, which may facilitate a centralized structure. Illicit operations must strike a balance to remain secret.

Seeking to examine the organizational structure of networks for three price-fixing conspiracies in the heavy electrical equipment industry (switchgear, transformers, and turbines), Baker and Faulkner (1993) used sworn testimony from the Kefauver Committee Report 1961 to map participant involvement. The network included seventy-eight individuals from thirteen companies. Individuals were linked if one directly observed the others participating in price-fixing activities—that is, joint meeting attendance. One of their most important findings concerns actor position within the network. Controlling for network structure, management level, and company size, this study found that central positioning in the network was key—the more people who witnessed an individual participating in price fixing, the more likely the person was to be found guilty. Visibility in the network increased vulnerability to conviction but did not result in more severe sentences or fines. Despite the passage of time, this research remains an important contribution to the study of criminal networks. If this line of research interests you, please review some of the recent work of Robert R. Faulkner and Eric R. Cheney.

Terror Group Structure

Krebs's (2002) study is deceptively simple relative to the other works discussed in this chapter. Do not let this sway your opinion of its importance. What it lacks in scientific rigor, it makes up for many times over in timeliness and ingenuity. This study is a descriptive exploration of a terrorist cell—the 9/11 hijacking network—drawn from news reports published by major news media in 2001: the *New York Times*, *Wall Street Journal*, *Washington Post*, and *Los Angeles Times*. Krebs demonstrated that insightful, criminal network research was feasible without access to confidential intelligence files. In doing so, he single-handedly inspired a wave of research and, arguably, may have facilitated the rise of networked criminology.

Mapping the associations of nineteen hijackers (connected through 66 ties) as initially reported in the media, and comparing this network to a more comprehensive set of thirty-seven individuals (linked by 170 ties) gleaned from subsequent news reports, Krebs demonstrated the importance of trusted prior contacts (e.g., those trained in Afghanistan, old school friends, and kinship ties) in providing the network shortcuts that were instrumental to the function of the cell. These deeply trusted ties wove the network together, though the ties were not readily apparent right away. Moreover, the network of the terror cell was sparse. Most cell members did not know or were not directly connected to many others; however, the network structure shifted with a meeting in Las Vegas. The meeting resulted in six ties (shortcuts), which brought distant actors together temporarily to coordinate activity. Finally, this study showed that the most important actors were not always the most central on various measures of central positioning. While this may be a consequence of strategic positioning, Krebs (2002) cautions that the metrics are sensitive to missing data and should be interpreted with care. (I return to the issue of missing data in chapter 8.) For more information on terror group structure, see Ouellet, Bouchard, and Hart 2017, which is a more recent investigation of Al Qaeda's structural changes pre- and post-9/11.

Street Gang Rivalries

The next study examined the socio-spatial networks of violent gang rivalries in Hollenbeck. Hollenbeck is a neighborhood in Los Angeles with a history of entrenched gang conflict. The research involved a collaboration between George E. Tita and the Los Angeles Police Department, and as a consequence, this study marked a shift in driving forces from academic interest to practitioner-academic partnerships. I opted to showcase Radil, Flint, and Tita (2010) because their research includes summaries of some prior work related to this partnership. Investigating rivalries among twenty-nine gangs, the Hollenbeck study illustrates the complexity of violent gang conflict—the association between social and spatial proximity is not absolute because gangs will attack others that hold gang turf in noncontiguous areas. Additionally, different sets of gangs can be structurally equivalent, and some sets play an important role in facilitating violence. These authors concluded that gang-related violence is shaped by the rivalry network but mediated by the relative location of gang territory, leading them to suggest that crime control efforts would be aided by the use of social and spatial techniques in tandem. Of note, the information used to map gang rivalries was drawn from a survey of police officers and former gang members. Consequently, the findings pertained to known rivalries and may not have been fully representative of emerging conflict.

Organized Crime

Francesco Calderoni (2015) exemplifies some of the work being done to understand the structure and operations of organized crime groups. Calderoni presents a method of identifying instrumental leaders by mapping networks constructed from meeting attendance using pretrial detention records (court documents, which include information from police surveillance activity). A network linking 215 people on the basis of meeting attendance (118 meetings examined) was used to calculate the central positioning of each person. These centrality scores were then added to a logistic regression model to determine whether the scores could predict leaders. Betweenness centrality performed the best. This statistic indicates the tendency for someone to be positioned between other people. As used in this study, high scores indicate that the person attended meetings with different groups of people. If scores are compared for different sets of meetings over time, it may be possible to use surveillance information to reveal emerging leaders or power shifts in the group. Calderoni (2015) demonstrated that law enforcement can surmise operational control of a group from surveillance data.

Collaboration in Organized Crime

Much of the research on organized crime groups aims to illuminate the structure of illicit drug operations. Looking across a distribution chain reveals collaboration between various criminal groups. Using fifteen hundred pages of transcripts of surveillance and electronic records, Tenti and Morselli (2014) mapped collaborative drug trade activity occurring over an eighteen-month period (December 2000 to June 2002). This sample of Italy's illegal drug supply network involved 242 individuals, 187 of whom were known members of a criminal group (9 criminal groups were represented) and 55 of whom were unaffiliated (meaning they were not known to be members of known criminal groups). Demonstrating the importance of examining individual-level connectivity and group-to-group level associations, these authors found that (1) the drug distribution network was more decentralized than centralized; (2) small groups tended to be associated with specific local activities; (3) large groups of importers and suppliers had broader commercial activity and could reach participants across the network; and (4) some groups and individuals functioned as key brokers (two different Italian groups involved in wholesale importation, one Albanian supplier, and an unaffiliated Italian individual acting as a go-between).

Hackers

The final article reviewed introduces a relatively new area of criminal network research—the social structure of hacking communities. A leading example of tactical SNA, the research of Décary-Hétu and Dupont (2012) illustrate that SNA can be used to efficiently examine thousands of communications to identify key players, against whom a tactical strike would have a destabilizing impact on the network. These authors argue that such a strategy would garner the same effects as a broader approach but, because the investigation could focus on a small number of targets, would use less resources. Reallocating the resources saved to the continued monitoring of the network would facilitate even greater gains if, after observing its reaction to the arrests, law enforcement launched new initiatives to limit the network's ability to adapt to the losses. These authors constructed a communications network that mapped interactions between 771 people using the logs of Internet Relay Chats recovered from hard drives seized during an investigation of a network of hackers involved in botnets. Botnets are networks of many infected computers (thousands or even millions) that can be controlled remotely to perform particular actions. The resulting graph mapped 4,714 conversational links among people. I picked this study because it highlights how electronic media can be used to investigate the structural properties of criminal relations. With a little ingenuity, and some data management skills, Internet and phone-based interactions can be mined to understand criminal behavior. If you are interested in this line of research, keep an eye out for David Décary-Hétu, who is very active in this area.

Criminal Enterprise

Perhaps owing to the influence of practitioner-academic partnerships and the links forged between the intelligence community and SNA-oriented researchers, it is not surprising that a substantial portion of criminal network research concerns the structure of criminal enterprise networks and the resiliency of these criminal networks to crime control efforts. More recently, attention is being drawn to using SNA to better understand crime control implementation challenges. Selecting the six studies highlighted in this section was complicated, as some of the most interesting pieces are also technical demonstrations of the mining capacity of new web-crawling and data-mining software, rather than studies of crime phenomena per se. Accordingly, I chose to highlight developmental data-mining work in chapter 11, under the heading practitioner-based drivers.

Structure

Drug Markets. As noted previously in this chapter, our understanding of the structure of criminal enterprise networks reached a milestone with Bichler, Malm, and Cooper (2017). In this systematic review, we examined thirty-four articles, reporting on fifty-four drug-trafficking networks. Since most of the studies mapped networks using law enforcement data

(76.5 percent) or court transcripts from successful prosecutions (17.7 percent) that focused on polydrug trade (48.4 percent) or cocaine (27.3 percent), the five conclusions reported below can be construed only as summarizing what we currently know about these type of operations as a result of criminal investigations (15):

1. Drug-trafficking networks are more apt to be sparse, with central individuals who connect the group and provide links between different groups, suggesting an operational preference for security.
2. Leaders of drug-trafficking networks and those with important roles are identifiable through centrality analysis.
3. A range of metrics and analytic techniques is used to identify central players to target.
4. Degree targeting or degree/human capital strategy performs best, though disruption efforts will vary in effect. Removing well-positioned and well-resourced actors from the trade network should split the network into smaller components and maximize the potential disruption of market activity.
5. Anticrime strategies need to be flexible, as networks continually evolve; attacks on the network originating from conflict within the network, or launched by the criminal justice system, lead to structural evolution.

Given what we learned from the case study of methamphetamine use discussed in chapter 4, readers may be interested to note that a cluster of studies included in the systematic review involved work by David Bright and colleagues, who investigated the structure of methamphetamine production and distribution networks in Australia (e.g., Bright, Hughes, and Chalmers 2012; Bright and Delaney 2013; Bright et al. 2014). Mapping illicit activities from court documents, these studies investigated the centrality of key actors and how networks evolved over time in response to the loss of key members. Readers interested in methamphetamine production should consult these studies.

Money Laundering Money laundering is a critical aspect of criminal enterprise. Despite its importance, few studies have investigated the operational structure of money-laundering activity. Seeking to fill this void, Malm and Bichler (2013) examined the network position of money launderers and their level of specialization in light of global trends toward nonconventional online banking activities. Co-offending activity was extracted from the 2007 "E" Division Provincial Threat Assessment report produced by the Royal Canadian Mounted Police. This report summarizes information about all known criminal enterprise that occurred from 2004 to 2006. Of the 2,197 individuals named in the report, 916 (41.7 percent) were known to be involved in drug market activity. We found that 102 people, about 11 percent of drug-involved individuals, participated in money-laundering activity. Two important findings were associated with this study. First, we found that only a small percentage of people were professional money launders (12 percent); rather, most people laundered their own proceeds from the drug trade (80 percent), while a small portion were opportunistic, laundering for a friend or family member (12 percent). Second, launderers were likely to be positioned between others, and thereby held brokering positions, suggesting a greater ability to control the information or materials flowing through the network.

Cybercrime Another vital element of modern criminal enterprise is the use of the Internet. As Goldsmith and Brewer (2015) argue, the anonymity and access provided by the Internet enable people to engage in criminal and delinquent behaviors in unprecedented ways. Understanding the new variations of crime enabled by the intersection of criminal enterprise and electronic exchange systems is vital to crime control efforts in the digital age. An excellent example of research in this area is Westlake, Bouchard, and Frank (2011). These authors describe a web crawler that they designed to extract information from child exploitation websites. Using data retrieved from the web crawler, they harnessed the power of automated processes to map ten child exploitation networks and suggested a protocol for identifying and prioritizing websites/targets for law enforcement action. Specifically, the study found that targeting sites high on network capital (a combination of severity of content—i.e., images and videos—and connectivity to other websites) can be an efficient way to maximize disruption of the network. We will come back to this study in a subsequent chapter.

Network Resilience to Attack

As illustrated by Westlake, Bouchard, and Frank (2011), a natural progression from understanding the structure of a criminal operation is to think about how this information can be used to disrupt the network. Often the observed network is compared to what it would look like with the removal of key actors (or websites). Metrics calibrate the fragmentation potential or change in actor position for the remaining group members after the potential targets are removed. The studies highlighted here use alternative strategies to calibrate resilience.

Trafficking. Morselli and Roy (2008) introduced a hybrid methodology integrating crime-scripting with SNA in order to map out the flexibility of car theft operations. Crime-script analysis, as explained by Cornish (1994), entails cataloging the procedural aspects and requirements of crime commission. Scripts map criminal procedures by decomposing procedures into logistical steps. For instance, car theft rings involve five scenes or steps—theft, concealment, disguise, marketing to potential new owner, and disposal/delivery of vehicle. Each step can be disaggregated further into the different ways each task can be executed. For instance, a vehicle could be concealed in a warehouse or a private garage. The different options available to accomplish the task are referred to as facets. Morselli and Roy (2008) argue that the combinations of facets across scenes, referred to as permutations within the script, provide a mechanism through which it is possible to assess procedural flexibility. Where a great many permutations are feasible, crime is highly flexible, and therefore difficult to disrupt. Using intelligence generated through two large, multiagency investigations, these authors mapped two ringing operations (car theft for export). Using centrality metrics to identify actors to remove, Morselli and Roy (2008) discovered that removing one person prominent in one facet of activity would significantly reduce flexibility, while removing three would completely disrupt the operation.

Transnational Crime Excited by the method introduced by Morselli and Roy (2008), Bichler, Bush, and Malm (2013) extended the hybrid script–SNA method to investigate an entire industry by mapping the tasks, professional roles, and tools used to produce, value, trade, and collect high-value art and antiquities (collectively referred to as cultural property). Examining professional roles and tools used to advance legal and illicit market objectives, these authors used network statistics to identify the points at which the illicit and legal markets intersected, thereby providing a focus for crime interdiction strategies.

In a subsequent study, Bichler, Bush, and Malm (2015) examined the resilience of the transnational illicit market, in light of the fact that legal and illicit activities often used the same facets—e.g., valuation of a piece and international transport via a specialized international shipping company.

Linking vital aspects of the high-value industry in cultural property by the number of shared facets, we anticipated the effects of two anticrime policy directives—movement control and monitoring financial transactions. By considering the complexity of sectors in the trade system (*complexity*) and the financial incentives to work around crime control efforts (*transformity*), we estimated that maximum decoupling of illicit\legal trade is more likely if crime control efforts target central trade functions associated with valuing and gifting objects. While this study illustrates how network-oriented research could be used to estimate the effects of efforts to dismantle transnational crime, it does not simulate the likely outcomes of crime control activities. I highlight the cultural property studies because we mapped the trade system from the findings of other research, and in doing so showed that meaningful analysis can be applied to novel sources. Tapping into novel sources helps to extend our exploration of the utility of network approaches to investigating crime and brings me to the second reason for showcasing this work. Innovative approaches are needed to increase the feasibility of transnational crime research that goes beyond a case study approach.

Disruption Arguably, Duijn, Kashirin, and Sloot (2014) were among the first to offer a convincing, theoretically derived approach to calibrating which actors are best to target and how the illicit enterprise is likely to rebuild after a strategic attack by law enforcement. Using what appears to be an agent-based simulation model calibrated with observed data, this study compared five disruption strategies (e.g., targeting randomly selected actors, targeting central actors, or targeting individuals with special skills) and three recovery mechanisms (e.g., picking replacements from people who are socially connected). The study found that the Dutch marijuana industry was more cohesive (greater density) and thus more resilient after attacks targeting those with the most human capital (special skills). Networks adjusted to become more efficient, but this reaction decreased security by making individuals playing instrumental roles more visible (e.g., coordinators and international traders) and, consequently, more vulnerable to subsequent law enforcement action. These authors assert that disruption strategies must be long-term efforts requiring repeated targeting. I highly recommend this article. It explains a scientifically rigorous study in a relatively accessible manner without sacrificing vital details. Moreover, Duijn, Kashirin, and Sloot (2014) provide a model process through which others can anticipate the effects of target removal. Please keep this study in mind, because it will be revisited in chapter 7.

For those of you keen to see what has come out since this primer was published, take a look at the current issues of the following journals: *Social Networks, Global Crime, Trends in Organized Crime,* the *Journal of Research in Crime and Delinquency,* and *Criminology.* In the next chapter, the discussion turns to data: specifically, how network information can be gathered, the strengths and limitations of the data collection methods and sources used, defining group boundaries, and the small-world method.

7. Gathering Data

> A frequent confusion about network research has to do with where theory ends and methodology begins. Network analysis is exemplary in the social sciences in basing its theorizing on a foundational construct—the network—that is both empirically meaningful and fully mathematical.
>
> BORGATTI AND LOPEZ-KIDWELL, "Network Theory"

My first question when being introduced to a new method or analytic technique is, "What does the data file look like?" Since a bit of reverse engineering can be useful, I begin this crucial chapter with two examples illustrating the ties existing among five actors. The first is an *edge list* that could first be copied from an Excel file into a text document with a couple of lines of code, and then imported into a social network analysis (SNA) program. We know it depicts the ties in a directed network because not every relation is duplicated—that is, Tori extends a tie to Robert, but Robert does not reciprocate. The first column lists egos, who originate contact, and the second column lists alters, who receive the tie. Each row represents a relation that exists between the pair of actors named. I could add a third column that records the type of relation or a strength or value.

TORI_JONES	ROBERT_WHITE
ROBERT_WHITE	VALENTINA_THOMPSON
ROBERT_WHITE	JASON_KRUGER
VALENTINA_THOMPSON	JASON_KRUGER
JASON_KRUGER	VALENTINA_THOMPSON
JASON_KRUGER	AMY_ROGERS
AMY_ROGERS	

Notice there are no extra spaces or symbols, and I capitalized to minimize typos. Amy Rogers appears in a row with no corresponding contacts because I wanted to record that her relational information is unknown. If Amy were investigated and found not to initiate any contact, she would not receive her own row: the last row depicted here would not exist.

This edge list can also be presented in a *CSV listing*, with egos listed first and with the names that follow as alters.

TORI_JONES, ROBERT_WHITE

ROBERT_WHITE, VALENTINA_THOMPSON, JASON_KRUGER

VALENTINA_THOMPSON, JASON_KRUGER

JASON_KRUGER, VALENTINA_THOMPSON, AMY_ROGERS

AMY_ROGERS

Both listings would support the formation of an actor-to-actor graph. To produce a two-mode graph, you would simply include one node type in the first column (the ego's spot) and list the records relating to the second type of node, where alters appear. When importing the list into SNA software, you would simply note that the data represent a two-mode graph with some simple code or by clicking "two mode" in an import wizard. The software would take care of the rest. As a general rule, all software programs have the capacity to convert lists into the matrix format required for analysis.

Gathering complete information about the members of a clandestine group can be a difficult task. But incomplete data is not a problem unique to the study of criminal networks. Even when we are interested in mapping a network of people involved in legitimate public activities, researchers encounter missing actors, missing links, and missing attribute data—three different types of information could be missing. Recall that the primary unit of analysis is the dyad, made up of two actors and the connection between them. Missing one of the actors or information about the tie could remove the pair from your inquiry. If attribute information is required (e.g., gender), then the gender of *both* actors must also be known. Depending on the aims of the study, the aggregate effects of missing data can be distressing. The good news is that there are ways to anticipate the effects that omitted actors or ties may have on findings. Admittedly, I have yet to encounter vetted tactics for dealing with missing attribute data.

Since data collection shapes all other aspects of analysis, the first chapter in this section introduces readers to the methods used and methodological challenges faced when assembling information about criminal networks. I begin with a short review of some general research design issues and the strategies used to gather network-oriented data. Next is a quick tour of sampling methods and data collection techniques, followed by a review of the strengths of commonly used data sources. The chapter concludes with a detailed description of the small-world method, a network-mapping strategy that highlights a point made in the quotation that opens this chapter. Chapter 8, which should be read in relatively quick succession, first examines the problem of entity resolution, and then describes issues associated with establishing network boundaries, missing data, and sensitivity testing. A forewarning. Despite the breadth of topics covered in this pair of chapters, only a cursory review is possible. Before launching into a project of your own, it is best to consult, or at least have on hand, one of the classic SNA textbooks, such as the ones listed at the end of this chapter.

ASSEMBLING INFORMATION ABOUT CRIMINAL NETWORKS

Research Designs

As with conventional research, SNA studies come in all forms. At one extreme are experiments and time series studies—in the field or in a lab, natural or contrived—designed to test hypotheses about the influence networks have on individuals or social groups. At the other end of the continuum are inductive exercises that describe, with varying levels of methodological complexity, the associations among a group of people. Use of actor-based simulation modeling is also growing and significant developments in large-scale data-mining procedures have taken place. Given the depth and breadth of SNA studies, to simply describe SNA as a "tool" to illustrate connectivity (a.k.a. map a criminal network) is, at the very least, a gross misrepresentation.

While diving into SNA requires a significant commitment to learning new theory, methods, and analytic techniques, all network studies still follow the tenets of the scientific method. This means that most of what you learned about research designs during your formal education or while on the job still applies when engaging in SNA. Investing in SNA simply expands research capacity and extends the kinds of research questions you can answer.

One difference between SNA and conventional research is the focus. Rather than talk about micro-, meso-, or macrolevel research, or individual versus group behavior, SNA-oriented studies seek to investigate connectivity and what that means for emerging structure and the relative position of actors within a personal group (egocentric network), within subgroups of a larger network, or within the whole network. Recall from the first chapter that structure is key: the research goal is often to understand how the structure influences the development of attitudes and perceptions, as well as the access to opportunities and, ultimately, behavior. Often multiple "levels" of analysis appear in the same report. You will recall the discussion in chapter 2 of analytic techniques, which can also adopt different base units—for example, dyadic or triadic metrics.

By far the most common designs used within criminology are cross-sectional in nature, by which researchers and analysts attempt to map connectivity within a suspected group of offenders or a population of actors in order to uncover key players and important subgroups, identify potential correlates that may account for the structures that are observed, and understand how social structures shape criminal opportunity and behavior. The objectives behind cross-sectional research range from mapping the architecture of a group to mapping the structure of something flowing through the network, such as messages or drugs. Less common but growing in prevalence are studies aimed at predicting how criminal networks react to targeted attacks. To illustrate the variability in this emerging field, three examples follow.

Simple correlational study of a purposive sample from an egocentric perspective

The first example is a simple correlational study of egocentric networks to investigate positional importance as revealed by communications flowing among a purposive sample of individuals. Interested in the command structure and survivability of terror-group leaders, Bichler and Bush (2016) applied a business model of competitive advantage to examine how ego network structures correlate with the operational success of command staff. Mapping the communication patterns of central leaders of Al Qaeda and the Islamic State of Iraq (ISI), who were active since 2006 and survived at-large for some period of time until November 2015, we found that less social capital *and* lower constraint improved the likelihood of survival. (For a refresher on Burt's [1992] arguments regarding the importance of social capital gained from maintaining a network with structural holes, see chapter 3.) So where did the information used in this study come from?

The methods used are fairly common in criminal network research. A sample of intelligence documents was scoured for person-to-person linkages. Information about communication among central command staff were extracted from two sets of documents released through the Harmony Program, by the Combating Terrorism Center (CTC) at West Point:

- Thirteen declassified and translated documents (154 pages of text) recording inter- and intraorganizational operations of Al Qaeda dated between September 2006 and April 2011

- Seventeen declassified and translated documents (58 pages of text) recording intraorganizational communications of the Islamic State of Iraq (ISI), and its subsequent and prior incarnations that occurred between 2006 and 2009

Actor-to-actor links were generated by extracting messages embedded within each document. Since messages were often sent via indirect routes, each person mentioned in a message was recorded. The originator, listed as the ego, was linked in a chain to subsequent intermediaries, and so on, until the message chain terminated. An edge list was generated linking each set of dyads. To illustrate, review the passage below, wherein Osama bin Laden writes instructions to his associate Atiyya Abdul Rahman: "Regarding what brother Basir mentioned relating to Anwar al-'Awlaqi, it would be excellent if you inform him, on my behalf in a private message to him, to remain in his position where he is qualified and capable of running the matter in Yemen" (Combatting Terrorism Center 2012).

Two communication chains were extracted from this text. First, bin Laden reveals that he received a message originating from Basir: BASIR → RAHMAN → BIN LADEN. (We assume Rahman delivered the message because of some comments found in a prior paragraph.) In the passage, Bin Laden also dictates a message to be delivered back to Basir: BIN LADEN → RAHMAN → BASIR. By converting these information chains into an edge list, we obtain the following set of ties:

BASIR	RAHMAN
RAHMAN	BIN LADEN
BIN LADEN	RAHMAN
RAHMAN	BASIR

The extraction protocol continues until all identifiable communication chains are harvested and dissected to form edges.

Since several Al Qaeda affiliates were active in ISI, the two groups were reasonably integrated in a network containing 302 actors connected via 541 ties. Ego networks were extracted for a subset of prominent leaders. Notably, the network was based on messages, orders, and documents passing between actors: this means that the network describes how information may flow among a sample of actors and does not necessarily reveal the complete architecture of the respective groups. While this example involves a time component and different variables, it is not like the example of the following dynamic, multivariate study.

Dynamic multivariate study—whole network mapping the architecture of trade relations

Dynamic, multivariate regression models are gaining traction among criminal network scholars as a result of the efforts of network statisticians and methodologists that advance software capabilities while adding user-friendly interfacing. To illustrate what these advanced multivariate models might look like, I turn to a study investigating the international trade in small arms. Working with a former marine, I investigated the common argument that the cessation of armed conflict triggers (pardon the pun) a shift in the flow of weapons (Bichler and Franquez 2014). Specifically, a surplus of secondhand small arms (ranging from small caliber to military-grade, medium-range motor tubes and ammunition) will flow out from the embattled nation into the illicit market through makeshift weapons supermarkets. Reviewing the anecdotal evidence, we found four possible market structures: (1) trade-interchange markets occur when weapon flow reverses course, returning to where it originated; 2) trade mediators exist when conflict nations pass weapons along through different nations acting as intermediaries after the cessation of hostilities; 3) epicenters occur when the conflict nation becomes a supermarket; and 4) trade channels occur when the embattled nation passes a weapons surplus to another nation, which acts as a supermarket. Fortunately, each of these market structures could be represented with specific SNA statistics. Now that we had a focus, we needed to map the gray market in global small arms and light weapons. Fortunately, there is a growing number of publicly accessible secondary data sources that arguably contain information for population of actors (or at the very least, a saturated sample).

Searching the United Nations commercial trade data, we discovered that information about small arms and ammunition transfers was available from 1997 to 2010 for 224 nations. A saturated sample of suspicious and clandestine trade relations was extracted by pulling all transfers that were reported by only one party to the exchange. Then, using dynamic actor-oriented modeling (it is a bit like multivariate logistic regression across networks over time), we investigated how the gray market of gun trade changed after the end of armed conflict, testing the degree to which market activity evolved to reflect each type of weapons supermarket globally, and locally, for two case studies (Egypt and Angola). We found that interchange markets were associated predominantly with conflict cessation on a global level, as recorded by the Uppsala Conflict Data program and controlling for the following:

1. Shared borders (border network constructed from looking at a map of the world);
2. Formal military alliances (alliance network generated from the membership lists of twenty-six different alliances, including the Collective Security Threat Organization, NATO, and so on); and
3. Insurgents operating in pairs of nations (an insurgent network derived from a publicly available terrorism-incident data set produced by the National Consortium for the Study of Terrorism and Responses to Terrorism).

Significant regional variation was also observed—the Egyptian trade network exhibited interchange-market tendencies similar to the global market, whereas the Angolan network evolved to favor an epicenter structure.

While the weapons supermarket study was a step-up in methodological rigor compared with the prior example, it is still a correlational study. The next example illustrates a computational experiment using simulation models to predict how a whole network or population might react to attacks targeting specific individuals.

Computational model testing disruption tactics on a whole network or population

Using computational models to conduct experiments, Duijn, Kashirin, and Sloot (2014) simulated the effects of different tactical strategies on the potential to disrupt the Dutch marijuana industry. Using all intelligence available from January 2008 to January 2012 to map the macrocommunity of organized crime from which criminal cooperation originates (24,284 people with 35,359 ties), in tandem with coarrests (2008–11) representing specific illegal enterprise (6,020 people with 12,073 ties), these authors extracted a subnetwork of 793 individuals (1,388 ties) involved in cannabis cultivation. They then compared five disruption strategies (e.g., targeted individuals were selected at random on the basis of high levels of human capital, a high number of direct contacts, high brokerage, or high scores on both human capital and number of contacts) and three recovery mechanisms (e.g., random, preference by social distance, or number of contacts), the authors tested the likely success law enforcement efforts might have in disrupting illicit trade under different conditions. Overall, the researchers concluded that networks were found to exhibit greater density, and thus more resiliency, after attacks targeting people with the most human capital. Moreover, individuals playing instrumental roles became more visible (e.g., coordinators and international traders), leading the authors to assert that these individuals were more vulnerable to subsequent interdiction efforts after initial attacks on the network. The authors concluded that disruption strategies must be long-term efforts, as networks recover to attack in such a manner that they become more efficient and resilient.

From a methodological standpoint, the study by Duijn, Kashirin, and Sloot (2014) is also notable in its use of script analysis with SNA. Recall from chapter 6 that crime scripting involves dissecting events into all of the smaller tasks (steps) taken in the commission of a crime and then cataloging all of the different ways that each task could be performed. The first to see the potential of developing a hybrid methodology was Morselli and Roy (2008). Extending the hybrid script-SNA method, Bichler, Bush and Malm (2013, 2015) investigated an entire industry. Taking the hybrid approach a step further, Duijn, Kashirin, and Sloot (2014) advanced this line of inquiry by using script analysis, in the computational model described above, to better understand how networks might rebuild after the strategic removal of specific actors. They used scripting to generate a value chain of activities specific to cannabis cultivation, against which they tested disruption tactics. Scripted roles included everything from tending plants, controlling cutters, and disposing waste/leftovers to financing operations and arranging international trade. Of the dozens of roles, two stood out: growshop owner and coordinator were most important for maintaining the flow of drugs and continuing production and distribution.

The examples described above illustrate how diverse SNA research can be in sampling, data sources, and analytic approaches. They also lead us to two important methodological issues that confront researchers at the start of an inquiry—figuring out what type of sample to generate and deciding how to collect the necessary information.

Sampling

Ideally, researchers aim to include all members of a group, and yet it is often not feasible to obtain a current and accurate roster listing all group members. To compensate, researchers trying to identify the participants of criminal networks take one of two strategies—conventional sampling techniques or ad hoc methods to purposively identify populations or communities of interest.

Conventional Sampling Strategies

Three conventional sampling strategies dominate criminal network research: (1) snowball and respondent-driven samples (a.k.a. chain referral method), (2) random samples, and (3) census or saturated samples. In conventional research, **snowball sampling** is so named because of the use of starting bits and a gathering process. When launching a snowball sample, we start by identifying people based on predefined criteria and asking them to participate. Snowballs could also be set into motion by identifying individuals who might know someone who qualifies for the study. Each recruit is asked by the researcher to refer others who might be eligible, and so on. As each contact generates more subjects for the study, referrals build the sample. This method tends to be called **respondent-driven sampling,** because respondents generate additional participants. The difference between snowball sampling and respondent-driven sampling is that, for snowball protocols, individuals nominate others, who may not themselves be part of the sample, whereas for respondent-driven samples, only participants are able to nominate others.

Within an SNA context, the methods are slightly different. The people initially identified as qualifying for the study are referred to as **seeds.** The researcher connects qualifying seeds to the people they nominate (i.e., their contacts, who are referred to as alters). The process repeats, with alters becoming the new focus and with their being linked to all their affiliates, and so on, for a set number of waves until quotas are reached, or the study period expires. Networks generated with the initial seeds, one wave of sampling, are described as involving a *one-step* data collection process. If the connections of alters are added to the network from a subsequent wave of information retrieval, the protocol is described as a *two-step* procedure, and so on with each wave of sampling constituting another step. Criminal network studies routinely extend up to two steps.

Readers should know that seeds can be identified in different ways. For example, I am currently leading a team of researchers investigating the effects that civil gang injunctions have had on the structure of gang violence in Los Angeles over a twenty-year period. This study uses a **researcher-defined** sampling process. The initial set of seeds is made up of the seventy-two gangs listed on the Los Angeles Police Department's website that had active injunctions against them as of October 1, 2017. Searching each seed gang, we identified convictions for serious violence (e.g., homicide, attempted murder, aggravated assault, rape, and robbery). Seed gang members could be victims or offenders. The network was formed by linking the gang of each defendant to the gang of each victim named in the case. In the second wave of sampling, we searched the gangs of alters (gangs involved in violence with seed gangs) and mapped links between defendant and victim gangs. Pilot testing this sampling strategy for twenty-three gangs affiliated with the Bloods or Crips consortiums revealed eighty-eight other gangs. As of this writing, six research assistants are coding hundreds of cases involving the eighty-eight gangs identified in wave two.

Samples of **natural groups** work a little different. The researcher selects a starter group, perhaps self-nominated members, and then each recruit is asked for more candidates until no new names appear—this situation is referred to as nomination redundancy. Recently, I attempted this technique to map networks of people attending electronic dance music (EDM) festivals (a.k.a. Raves). I sent an electronic survey about event attendance and drug use to a group of fourteen self-identified EDM attendees, or people who knew EDM attendees of college age, who frequented events held in the City of San Bernardino, California. Each participant was asked to take the survey, to send the survey link to their friends, and to post the link on their social media pages. In turn, survey respondents in the next wave of were also encouraged to repost the survey link on their media platforms and to encourage their family, friends, and associates to participate, and so on. Our respondent-driven recruitment continued for about two weeks and generated a sample of fifty-five people.

Random samples are conventional probability samples; again, samples can investigate researcher-defined or natural groups/communities. For instance, someone might be interested in social networks that form in juvenile hall. To map relations, the researcher could obtain a list of all youth detailed in juvenile hall on a particular day. Then, since resources are limited, a random sample of people could be generated and these individuals invited to participate in interviews about their associations within the facility. Alternatively, juvenile hall staff could be interviewed to identify social groups within the facility. To generate a stratified sample, people would be randomly selected for interviews from each staff-defined group. If, instead, a particular group is of interest, a set of members from the group as identified by juvenile hall staff could be randomly selected.

To generate a **census or saturated sample,** start by identifying all actors matching predefined selection or inclusion criteria. Generally, this strategy involves tapping into an information source providing details about actors. For example, I am developing an evaluation of a program designed to enhance the employability of at-risk youth—the program is made up of a series of life skills and job skills workshops and a 10-week paid internship involving community service. One of the program's aims is to reorient social support networks away from gang-involved people to positive, prosocial role models. To map changing support networks, all members of the treatment and control groups will be interviewed about the people they trust during an intake orientation. Interviews will be repeated again eleven weeks later when the program finishes (we plan to implement delayed treatment for the control group, meaning that members of the control group will get to participate in the program when data collection is complete). In this example, the intake process identifies all those matching the eligibility criteria. In prior assessments of this program, we found that eligible candidates were in short supply. Since in this case there is a small population of candidates, all eligible referrals will be offered a chance to participate. Arguably, the participants will represent a saturated sample of the eligible population (the group still represents a sample because not every referral is likely to participate).

An alternative strategy to generate a census or saturated sample would be to identify one group member. The individual is asked to name other members. Then, as each nominated person is interviewed, they are shown the list and invited to revise the list. Once data collection is complete, researchers could implement an inclusion rule, i.e., individuals whose membership to the group was confirmed by at least five other people remains in the study.

Sampling Methods to Identify Populations and Communities

Sometimes the focus of a study is not a group; rather, the objective is to investigate a population or community. To sample such a large set of actors, ad hoc methods or scale-down sampling strategies are required.

Ad Hoc Methods **Ad hoc methods** use a purposive process to capture all potential actors using loose criteria regarding group membership: the resulting set constitutes a population. Then, by applying a more restrictive set of rules, it is possible to extract a subset of people with sufficient levels of interaction to suggest they function as a community. Ad hoc sampling methods often begin with the identification of a secondary data source, usually a data management system that is capable of two functions—running relational queries for a specified period of time and filtering by eligibility criteria. For example, as described above, Duijn, Kashirin, and Sloop (2014) extracted all co-offending activity for a designated period of time from a Dutch law enforcement records management system (four years, from January 2008 through January 2012). By querying the records management system for all individuals who were co-arrested for an illegal enterprise (specific criterion), these authors developed the "universe" of individuals known to be co-involved in criminal enterprise. It is widely acknowledged, however, that arrest data are incomplete; not everyone who commits an offense with another person is caught. To fill information gaps (people suspected of collaboration but not yet arrested together), the researchers integrated coarrest data with a broader set of information generated from querying a data system that tracked all intelligence relating to Dutch organized crime activity for the same study period. Readers should be aware that integrating data from disparate sources requires a schema to link database tables that include the information to be matched, algorithms that link individual records, and data fusion processes to merge pairs or groups of records into a clean file (see Christen 2012 for more information). Owing to the importance of this process, I will return to this topic in the next chapter.

From the unified network that maps a more complete universe of actors (people coarrested or suspected of collaborating), the authors extracted a subnetwork that included a "group" of 793 connected individuals (1,388 ties) involved in cannabis cultivation. For this second step, the network boundary was set by the criterion of coinvolvement in cannabis production. The term *group* appears in quotations, because we need to use the word loosely in this context, since the network does not reflect a group in the sociological sense; rather, the subnetwork represents a community of interconnected people. In reality, the community is likely to encompass many clusters of people, who define themselves as a group. For this community of actors, various techniques can be applied to identify subgroups of actors representing social groups. But what happens when the graph is too large to work with? How do you identify communities of actors, let alone identifiable social groups?

Scale-Down Methods Data systems and Web 2.0-enabled forums may have evolved to the point that we are now in the age of too much data. Imagine trying to analyze the connectivity among Facebook's 2 billion users. Identifiable groups exist, but software limitations and computing capacity render analysis problematic. One strategy to study graphs generated by exceptionally large data systems is to deploy a scale-down technique to extract a representative sample from the larger master graph (*subgraph sampling*). Before discussing subgraph sampling, I must note why considerable effort is expended to investigate these huge networks in the first place.

Huge social graphs often result from data-mining processes integrating information from many sources to support intelligence functions or from investigations of online social communities that seek to understand social trends—for example, the spread of extremism, radicalization, and the transmission of criminal innovations. Merging information retrieved from different accounts or platforms offers an opportunity to generate more complete attribute and network information for target individuals. Data integration is feasible because, even though people use multiple platforms or post from different accounts, they often repost the same or similar photos, demographic data, histories, and activities. These shared data points help to identify unique users. Using web crawlers or scrapers to retrieve information, both relational and attributive, and applying an entity-resolution algorithm (discussed in the next chapter) to consolidate information about a single person (thereby removing aliases) generates more complete networks.

But these massive webs of social connectivity include various communities within which well-defined social groups exist—social groups containing target individuals involved in illicit activity. Huge graphs are not manageable, and, more importantly, many social network statistics require that the groups examined have clearly defined boundaries. The presence of excess actors or ties will significantly bias results. Fortunately, many sampling techniques can be used to scale down huge graphs to a manageable size or to focus on a community of actors approximating a definable social group,

Table 7.1 describes some of the processes used to scale down huge graphs. Note the use of the word *some.* The management and analysis of huge graphs is a rapidly developing field, and my intent is simply to introduce sampling strategies. Readers looking to use these strategies will need to invest considerable time and effort into learning how to write and work with the necessary algorithms. The difference between the methods reported in the table and the process used by Duijn, Kashirin, and Sloot (2014) is that the methods described here include a probabilistic process or use of structural attributes to generate a subgraph, whereas Duijn, Kashirin, and Sloot's process invoked a purposive, researcher-defined inclusion criterion to identify a community. Two sets of methods are presented—one that selects at random from the master graph and one that is crawling based.

Randomly selected subgraphs are often generated to represent the original master graph. For this reason, scaling-down techniques can be assessed on their ability to produce subgraphs with the same properties; admittedly, the properties of the original graph may be assumed or estimated if it is not possible to calculate basic descriptive statistics. Two sampling processes are reported in the table—random node selection or random edge selection. As the random methods reported here are self-explanatory, I will focus on the methods that are likely to be less familiar to readers—the crawling-based sampling procedures—rather than recap what is in the table.

TABLE 7.1. Sampling processes used to generate representative subgraphs

Purpose	*Method*	*Example process*	*Potential biases*
Random selection from master graph	Random node selection	At random, select nodes, then incident edges (include alters connected t o selected edges). Node selection options are uniformly random, with probability based on degree centrality, probability based on page rank, and so on.	Master graph properties may not be represented (i.e., scale-free master graph and nonscale free sample, or overly sparse subgroup).
	Random edge selection	Randomly sample edges, then include incident nodes. Options are uniformly random, random node edge, hybrid (perform some uniform edge and random node edge), and so on.	
Crawling	Breadth-first search	Select (explore) least popular nodes and add neighbors (i.e., pick the weblinks least selected from a page and then capture pages linked to them); repeat the process using the least popular nodes from the newly added group.	Oversampling of nodes with lots of connections (high-degree nodes) may occur.
	Depth-first search	Select (explore) most popular nodes and add neighbors (i.e., select people with most contacts and add their neighbors); repeat the process using most popular nodes from the newly added group, and so on.	
	Random walk	Randomly select node, then select (explore) neighbors (i.e., select neighbors uniformly at random according to the most recent activity; if its a webpage, the most recently visited; if a contact, the newest) from a set of newly selected nodes. Repeat the process for a set number of steps (L steps).	
	Forest fire	Randomly select seed(s), then visit (select) neighbors using a predefined process (i.e., a probabilistic breadth-first search with probability set at . 7; repeat the iterations until sample size is achieved or destination (or target) node is reached.	

There are many crawling methods, and several variations on each method. I decided to cover only a few of the simpler procedures, as this book is a primer. Breadth-first and depth-first searches select starting nodes on the basis of a property, such as the nodes' structural position within the full graph. **Breadth-first searches (BFS)** start with the least popular nodes, often those on the edge of the graph, and then gather additional nodes by following their relations for a set number of steps. **Depth-first searches (DFS)** start at the opposite end of the popularity spectrum, selecting the nodes with the most connectivity and then crawling out from these individuals by capturing, for example, their contacts for a set number of steps. Both methods presuppose that it is possible to calculate positional metrics on the original master graph, or that the data source includes metrics about each node. For example, if a subgraph from Facebook is being generated, it is possible to use the number of friends to gauge the popularity of people, from which a breadth-first search or depth-first search could be launched. Once subgraphs reach a set size (predetermined by quota or an assessment of statistical power), the crawl terminates. Since both methods rely on attributes of nodes, these methods can produce biased samples, but this might be acceptable depending on the research aim. For instance, if the objective were to investigate the subgraphs containing the most prolific offenders identified in a master graph of co-offenders, then using a depth-first search method to extract a manageable subgraph would be warranted because the crawl starts with the most active people.

Random walks and **forest fire** sampling introduce random functions, but a higher probability that popular nodes with lots of connections will be oversampled remains. Random walks begin with a random selection of nodes, with the number dependent on the size of the master graph. Stratified sampling is feasible, if strata are available and there are few cases missing the attribute associated with the stratification (e.g., gender, offense history, city of residence). The crawling algorithm will then select incident nodes; that is, nodes connected to the selected node at random, based on probability, or if they satisfy a pre-set node-based criterion, such as recent activity, i.e., most recently added or newest co-offending partner. The process is repeated using the set of newly selected nodes, and so on, for a predefined number of steps. Again, the sample size will be dictated by the research purpose and power calculations.

From what I can surmise, the main differences between random walks and forest fires is that the crawling algorithms of forest fires start with a random selection from a predefined list of seeds, rather than a simple random selection, and the protocol integrates a structural attribute to select nodes in lieu of a probability function (e.g., breadth-first search or least contacts). The process repeats using the set of newly selected nodes until the target sample size or a predefined target node is reached.

To illustrate this type of sampling technique, I draw upon the work of Westlake, Bouchard, and Frank (2011). They developed a web-crawling algorithm that uses a forest fire approach. Interested in disrupting online pornography, the team developed a custom-written crawler, named the Child Exploitation Network Extractor, which when deployed automatically browses the Internet and gathers and assembles information about the pages examined. It also generates the page-to-page network structure of the exploration. Using ten seeds (five websites and five blogs), the Child Exploitation Network Extractor retrieved associated pages, analyzed content, and followed all links out of the page. Repeating the process of capturing links and analyzing content, the crawler added pages by following links if the page included a set of predefined key words (pages were only selected if they had at least seven of the sixty-three key words). The crawl was set to terminate when it reached two quotas: (1) a set number of pages (250,000 pages) had been analyzed and retrieved, and (2) a fixed number of websites or blogs (200 unique sites or blogs) had been examined. Each seed generated a network that was manageable and could support meaningful analysis that identified key pages for targeted removal by law enforcement actions aimed at disrupting access to child pornography. Now that you have a fundamental understanding of sampling, it is time to think about how to gather the necessary information.

Collecting Data

I cannot refrain from stating that network information is everywhere. Contrary to popular sentiment, I am not referring to the abundance of social media platforms and the plethora of tweets that have inundated our waking hours over the last few years. Rather, as we navigate our lives we leave traces about our connectivity. If you do not believe me, look at your phone. If your habits are typical, you will not have cleared recent phone calls and text messages. A network could be mapped showing all of your recent communications, with details about who originated the contact and who received it (a directed network). Using the length of the call (or number of characters in the text) or number of communications, it is possible to value the ties. Date and time become link attributes. If you are not convinced by the list of phone calls and texts, I bet that your email communications stretch back for a considerable amount of time, with some messages dating from several years ago. While examining the frequency and duration of your digital communications will not necessarily reveal the most important people in your life, it is a good start and may in fact do just that.

The method chosen for unveiling information about actors and their connections is driven by conventional issues like research purpose, resources, and feasibility. The most common direct methods of collecting primary source data include surveys and surveillance-based intelligence gathered during the course of an investigation. Indirect methods in widespread use include tapping into archives or extracting relational information from data management systems or web-based electronic media. Each of these methods, along with their pros and cons, is discussed briefly below.

Direct Methods

Surveys Generating a network by asking people directly about their connections is likely to produce viable information, provided you are able to establish sufficient rapport to elicit private information and respondents are convinced of the anonymity of the process. If you are reasonably convinced that one or both conditions are satisfied, then you must make two critical decisions: (1) which format to use and (2) how many contacts to ask about.

Questions can be posed in a close-ended format or an open-ended format. *Close-ended* questions are aided, in that respondents see a list of people with whom they may be connected. Respondents are asked questions about what ties they have; sometimes researchers also ask respondents to place a value on the relations, either in absolute terms (e.g., frequency of contact over a specified time period) or in relative terms (e.g., ranked scale). An example follows. Q1 asks about professional relationships within the police department, and Q2 values that tie. Q3 asks about a different type of relation—trust. Notice how the set of choices repeats with each question; hence the name, *repeated roster.*

Q1. Please indicate whom you would ask for professional advice (e.g., about a promotional exam or a question about policy).

- ☐ John Luther
- ☐ Jimmy McNulty
- ☐ Olivia Benson
- ☐ Danny Reagan
- ☐ Poncherello

Q2. Please indicate how often you have sought advice from each person listed.

	Never	Infrequently	Somewhat Frequently	Very Frequently
John Luther	☐	☐	☐	☐
Jimmy McNulty	☐	☐	☐	☐
Olivia Benson	☐	☐	☐	☐
Danny Reagan	☐	☐	☐	☐
Poncherello	☐	☐	☐	☐

Q3. Please indicate whom you would trust with a personal secret.

- ☐ John Luther
- ☐ Jimmy McNulty
- ☐ Olivia Benson
- ☐ Danny Reagan
- ☐ Poncherello

For convenience, the items can be organized in a multigrid. *Multigrids* can be easier for the respondent to navigate and permit more ready comparison, so that respondents can recheck their answers. In the example that follows, I changed the two repeated rosters described previously into a multigrid.

Q1. Please consider each person listed below and place a check mark on the corresponding box of those from whom you would seek professional advice and those with whom you would share a personal secret.

	Ask for professional advice	Trust with a secret
John Luther	☐	☐
Jimmy McNulty	☐	☐
Olivia Benson	☐	☐
Danny Reagan	☐	☐
Poncherello	☐	☐

Both of these formats are useful if you have a small roster of people who belong to the group, and the boundaries of group membership are clearly defined. Close-ended questioning is generally considered more reliable, in that we are relatively certain that each person has been considered and that they had an equal chance of being selected. The method becomes less feasible for large groups or in situations in which rosters are not possible because group membership is not known—a common circumstance when studying criminal networks. I should mention that there is a potential fix for the last problem.

Digital survey software makes it feasible to pose close-ended questions when researcher-defined rosters are not available or plausible. Recall the electronic dance music (EDM) festival study described previously. This study involved an anonymous survey, and the sample was unknown at the beginning of the study. So how did we use close-ended questions to map the personal networks of respondents? The answer is that we included an initial open-ended question, asking respondents to nominate a set number of friends with whom they attended an EDM.

Q. Using pseudonyms, list up to six people with whom you went to an EDM in the past year. (If you went to events with fewer people, do not enter extra names. If you went to events with more than six people, list the people you went with most often.)

Person 1	enter pseudonym here
Person 2	enter pseudonym here
Person 3	enter pseudonym here
Person 4	enter pseudonym here
Person 5	enter pseudonym here
Person 6	enter pseudonym here

The pseudonyms of individuals would then populate the rosters in subsequent questions. We used repeated roster and multigrid methods to capture information about trust, deviance, and extramural associations occurring away from EDMs.

Open-ended questioning relies on the respondents' ability to recall people and apply the relevant eligibility criteria. While people are generally pretty good at remembering others who are important in their lives, recalling less intimate relations can be more elusive, and studies show that free-listing misses about 25 percent of the eligible contacts (Borgatti, Everett, and Johnson, 2013). Prompts to improve recall, such as mentally walking through places or recalling recent meetings or social events, are best deployed in face-to-face interviews. The other way to ensure that open-ended questions solicit high-quality information is to restrict the number of people nominated. It is widely accepted that asking respondents to nominate no more than six people produces the least number of recall errors, but it does introduce censoring (meaning that there is an artificial cap on how many people the most popular individuals can nominate). Direct questioning is an excellent way to gather network information, but it is often not a feasible method for mapping criminal networks. Sometimes direct observation is necessary. Before moving on to our discussion of surveillance, please reflect on the pros and cons of survey data.

Survey Pros	**Survey Cons**
Enables researcher control over the operationalization of key variables, as well as node and link identification procedures	Requires network boundaries to be apparent (i.e., membership lists can be generated) or identifiable
Captures egocentric networks or small networks	Is resource heavy
Allows for probes to improve data quality	Can have high nonresponse rates

Surveillance Studies of criminal networks often map relations that are revealed through investigations, either from intelligence being gathered for a case or, more frequently, from wiretaps or other forms of surveillance (see, for example, Campana 2011; Natarajan 2000; 2006; Morselli 2009; Varese 2011). Any effort to use information generated through investigations must be used cautiously, as the process used to identify actors and their relations is influenced by the working hypotheses driving the investigation, by the resources deployed, and by time (graphs mapping information from the early stages of the inquiry can look radically different from those at the time the case concludes). With this said, several criminal network scholars have explored validity and reliability issues associated with using surveillance data, specifically wiretaps, and concluded that, so long as certain conditions are met, the data source is viable.

Campana and Varese (2012) argue that using wiretap conversations in criminal network research is valid if the following conditions can be assured through independent assessment or cross validation:

1. The vast majority of individuals under surveillance talk freely, meaning that while messages may be encrypted for general security, the individuals do not purposely avoid contacting others or engage in other forms of self-censorship.

2. The surveillance must include a reasonably wide set of group members: coverage that includes only one or two people when the group numbers in the hundreds is likely to produce a biased representation of the whole network.
3. A large sample of conversations is available. While the sample size will vary according to the nature of the group in question, the research questions driving the study, and the frequency of contact among group members, at least several hundred conversations are needed for any meaningful analysis.

If all three conditions can be assured, then analysis of "conversations may yield valid inferences on the nature and activities of criminal groups and enrich the understanding of the ties within a criminal network" (Campana and Varese 2012, 13).

Investigating the source of wiretap information, and comparing the full set to subsets used in arrest warrants and judicial statements, Berlusconi (2013) found that descriptions of the whole network are best if they are based on the full set of wiretap information generated for a case. At other stages of the prosecutorial process, peripheral actors and links are lost, and, while the key players remain, the full functioning of the criminal group is less apparent.

Surveillance of physical activity can also generate meaningful inferences. For instance, Calderoni (2015) obtained a pretrial detention order issued by an investigative judge of Milan, who was looking into the dealings of members of the 'Ndrangheta mafia. He was able to extract information about 308 individuals who participated in 574 meetings over 4.5 years. Analysis of meeting attendance indicated that important leaders emerged as central to the network, and that an individual's position within the mafia correlated with their relational position in the network. Moreover, meeting attendance provided insight into group structure throughout the investigation. The predictive capacity was stable over time—the structure generated from data gathered at the beginning of the investigation resembled networks from the final stages of the inquiry.

In sum, surveillance sources are good for capturing current group member relations, provided that the data collection activities target a large portion of the group for a reasonable amount of time, and that group members do not excessively self-censor their behavior. The greatest drawbacks include the potential for investigatory factors to bias the sample and the operational challenges to establishing, maintaining, and sharing information across practitioner-academic partnerships.

Indirect Methods

When direct data collection is not feasible, researchers and analysts adopt indirect data collection techniques. Two of the most commonly used indirect data collection methods involve tapping into archives or secondary sources.

Archives Although little is written in most textbooks about what protocol should be established to harvest actor and relational data, archives can be invaluable resources. Their use has at least three distinct advantages: (1) There is no direct subject-researcher interaction, which removes potential threats to study validity (i.e., subject reactivity) and reduces potential harm to participants. Consequently, institutional review board requirements are less onerous, shifting from a full-board review to an exempt or expedited classification. (2) With the exception of navigating access to the documents (e.g., the hours the collection is open), data collection is not time sensitive, and source documents are unlikely to disappear, as some live participants might. Moreover, documents can be reexamined to ensure information was coded in a consistent fashion. (3) The nature and structure of sources can lend themselves well to cross validation: often, information includes official names and other details about actors and their activity that can be compared to other sources to verify the actor, attribute, and relational information extracted.

For example, mapping Al Capone's syndicate, including criminal and noncriminal actors, Papachristos and Smith (2014) showed that using archival sources is useful for investigating criminal groups. Papachristos' team gained access to the Chicago Crime Commission archives and were able to extract actor and relational information from an assortment of documents, such as newspaper clippings, police investigations, personal correspondence, funeral proceedings, and court records. Information was cross-checked, and missing details were retrieved by integrating information from other related archives—the Internal Revue Service, the National Archives of the Great Lakes, Northwestern's History of Homicide in Chicago 1870–1930 database, and Proquest Newspapers. Using a multifaceted relational-coding strategy, Papachristos and Smith (2014) were able to investigate the intersection of organized crime and the legal political, economic, and social life of Chicago. The authors found overlap between criminal and noncriminal worlds among the 1,883 people identified.

There are also disadvantages to drawing on archival sources. Data quality varies significantly in terms of the completeness, depth, and accuracy of the original material and should be assessed on a case-by-case basis. Depending on the focus and resources available to the curators, custodians, and historians responsible for the collection, the information assembled may be biased, omitting critical links and actors. At the other extreme, the archive could be so rich and extensive that it is not feasible to process all of the information. It is fair to say that the nature and scope of an archive influence both the types of research questions that can be investigated and the ways in which the questions can be answered. Along these lines, archives often capture information about events that individuals attended or groups that they belong to, rather than the relations among people. Consequently, it is often necessary to construct a two-mode network and derive actor-to-actor connectivity. But gathering data to map a two-mode network can be a methodological advantage: Kossinets (2006) argues that, when you suspect that the data source is incomplete, using a two-mode data collection protocol can reveal hidden connectivity.

Returning to Papachristos and Smith (2014), coding relations was difficult as the documents examined did not always contain information in a manner that supported the recording of actor-to-actor relations. For instance, the list of pallbearers at a mob boss's funeral should be coded in a two-mode format, whereas information about relatives and kinship is best recorded as actor-to-actor linkages. This inconsistency in data sources prompted the development of a more complicated data-coding protocol. The second issue facing the research team was the expansiveness and depth of the Chicago Crime Commission archives. To deal with this problem, a random seeding method was used to sample a specific type of document—the consolidation files that summarized key figures and events. Finally, the researchers noted that the coverage and quality of information available about each person named was not consistent—omissions, typos, bias, and contradictions were apparent. While the study faced limitations because it relied on archival data, it offers the most comprehensive modern investigation of the structure of organized crime in Chicago during prohibition, and it is worth taking the time to read the original report.

From the preceding discussion we can draw a few factors to remember when deciding if using archival data is right for your project. On the pro side of the argument: (1) the lack of subject-researcher interaction significantly reduces validity threats; (2) data collection is not time sensitive; and (3) cross validation is feasible. On the other hand, data quality and scope vary, which can limit the nature of the study. So investigate an archive fully before deciding to rely on it for your investigation.

Secondary Sources—Official Data Arguably, the most commonly used data collection strategy among criminal network scholars is to solicit a large data dump from a practitioner partner, data warehouse or repository, or private entity, either by permission or through the launch of a web scraper. Having many of the advantages of archival sources, secondary data typically have one additional characteristic that makes it so popular. More often than not, the records management systems warehousing the original data are capable of running targeted queries with user-defined filters and exporting data in a spreadsheet format that makes network generation less resource intensive.

There are also drawbacks to using **secondary sources.** Foremost, it can be difficult to convince agencies to share data. When negotiating a data-sharing agreement, privacy concerns may be restrictive, resulting in aggregated data (e.g., gang-to-gang links instead of actor-to-actor links) or anonymized data (e.g., an analyst with the agency replaces personal identifiers with anonymous research identifiers, with no comments about how the identifications were assigned). And even if agreement can be reached, not all data management systems were designed for pilferage by researchers—the data quality and completeness vary significantly. When collected for other purposes, secondary data may require extensive scrubbing to standardize node identifiers, reclassify ties, and resolve entities (ensuring each actor is uniquely identified).

Despite these issues, secondary data are among the most used sources of information for studying criminal networks. While network information can be extracted from many public sources (e.g., court transcripts, news items, and published threat-assessment documents), a byproduct of the era of big data and transparency is a growing number of publicly available spreadsheets or incident compilations for people interested in thinking about criminal networks. Below, I report on some publicly available relational data sources to illustrate the growing accessibility of secondary data.

Uppsala Conflict Data Program provides information on armed conflict—combatant attributes, conflict information, and dyadic files (http://ucdp.uu.se).

United Nations Office on Drugs and Crime (https://data.unodc.org/) reports drug seizures of a specific size for a range of drugs, incident details, and drug transshipment (country-to-country) linkages. The report on individual drug seizures (IDS Report 2010–2016) is at the following link: https://dataunodc.un.org/ids.

Global Terrorism Database (http://www.start.umd.edu/) collates attack data in several files that can be used as two-mode networks (or derived actor-to-actor networks). Click on the *Data & Tools* tab.

UN Comtrade Database (https://comtrade.un.org/) makes global trade activity accessible, but commodity transfers from territory-to-territory require some data manipulation to extract clandestine activity.

Small Arm Survey Reports (www.smallarmssurvey.org/) contain details about small arms and light weapons that could be entered into relational data files.

Data is Beautiful (www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/) offers a global accounting of data breaches of thirty thousand records or more, and while substantial effort is needed to harvest information, two-mode network construction is feasible.

SMALL-WORLD METHOD—MAPPING ACQUAINTANCE CHAINS

The small-world method is referred to more formally as *contact mapping*. Contact mapping emerged from efforts to figure out how likely it is for two people, selected at random from a defined population, to know each other. One of the most recognizable names associated with launching the experimental method is Stanley Milgram. Made infamous for his obedience to authority studies, Stanley Milgram also made a significant contribution to network thinking by launching a series of studies in response to a mathematical model of the small-world problem proposed by Ithiel de Sola Pool, from the Massachusetts Institute of Technology, and Manfred Kochen of IBM. The Pool-Kochen model assumed that individuals had an average of five hundred contacts (the estimate was drawn from a descriptive study) and estimated that the chance that any two people knew each other was one in two hundred thousand. (The United States had about two hundred million people at the time the model was built.) Milgram found the premise on which the estimate was calculated was not appropriate and set out to solve the small-world problem with an experimental approach that mapped social structure. Arguing for an experimental solution, Milgram (1967, 63) writes, "If you could think of the American population as simply 200 million points, each with 500 random connections, the model would work. But the contours of social structure make this a perilous assumption, for society is not built on random connections among persons but tends to be fragmented into social classes and cliques." What followed in the experimental tradition was a set of seminal studies that charted the characteristics of small worlds.

Partnering with Jeffrey Travers at Harvard University, Stanley Milgram devised an experiment to map contact chains. In the first two implementations, residents from two arbitrarily selected cities—Wichita, Kansas (Kansas Study), and Omaha, Nebraska (Nebraska Study)—were sent letters of invitation to join the contact study. The Kansas sample included 145 participants, and at least 160 were in Nebraska's initial study protocol (subsequent studies increased this sample to 196 and combined it with a Boston subgroup of 100 people). Participants were sent packets containing the information about the target person, rules for reaching the target, and a roster capturing information about the participant and the person to whom they sent the packet. The roster was intended to prevent looping. The protocol was simply for the participant to forward the packet by mail to the target, or to one personal acquaintance known to them on a first-name basis, who might be more likely to know the target. The objective was to get a folder contained in the packet to the intended target—for the Kansas participants, the target was the wife of a divinity school student living in Cambridge, while the participants in the Nebraska study attempted to reach a stockbroker working in Boston but living in Sharon, Massachusetts.

The packet also included tracer cards. Tracer cards captured participant information, as well as details about the chosen intermediary. Participants were instructed to mail the tracer cards to the research team at the same time the packet was forwarded to the intermediary. Tracer cards provided information about the contact chain that was of critical importance, particularly when packet recipients declined to participate in the study. Tracer cards permitted a detailed analysis of study attrition (a.k.a. broken chains that did not reach the intended target) that could be used in reference to completed chains (those reaching the target). The results of the study are not always clearly explained, but I draw from two source documents (Milgram 1967; Travers and Milgram 1969) to summarize some of the highlights below.

Nebraska Study

- Forty four (27.5 percent) of 160 chains were completed, meaning they reached the intended target.
- A median number of five intermediaries linked the originator and the target, suggesting that the path length separating the originator and target contained six steps.

- A substantial portion of chains reached the target through one of three local hubs.

Kansas Study

- Twenty (13.7 percent of 145) of packets reached the target.
- Of the starting sample of 145 people, most sent packets to someone of the same gender (38.6 percent of starting links were female to female, 40 percent were male to male, 12 percent were female to male, and 9 percent were male to female).
- Of the starting links, 85 percent involved participants sending the packet to a friend or an acquaintance, not a close family member.

Following the spirit of the small-world method, Dodds, Muhamad, and Watts (2003) found complementary results by mapping electronic messages on a global scale. The authors recruited 61,168 individuals from 166 countries to participate in an Internet-based contact-mapping experiment. Participants were randomly assigned one of eighteen targets distributed across 13 countries. Targets were drawn from a range of occupations—professors, archival inspectors, technology consultants, law enforcement officers, and veterinarians. Updating the method, the researchers asked participants to forward an email message toward the target by passing the message to a social acquaintance whom they considered to be closer to the target than themselves. Information about participants and intermediaries was captured online. Data capture included the intermediary's name and email address, as well as a description of the relationship between the participant and the intermediary (e.g., duration of the relationship, type of tie, and strength of the relationship). Including participants recruited as intermediaries, the final sample was 98,847. There were 24,163 distinct message chains, of which 384 messages reached their intended targets. Successful messages passed through a median of 5–7 people and favored geographic proximity early in the chains, but after the third step preference switched to other characteristics, such as occupation and education. Unsuccessful searches disproportionately relied on professional contacts. Little funneling appeared, leading authors to conclude that the search process was more egalitarian than Milgram's mail-based study. Notably, attrition was exponential and a function of chain length, with more chains faltering with each additional step.

As noted in the quotation that opened this chapter, newcomers to SNA sometimes find it difficult to distinguish between theorizing and methodology.

A case in point is contact mapping and the small-world perspective—particularly if you try reading Milgram (1967). Contact mapping is an important method of data collection, in part because it helped to shape several of the theoretical arguments reviewed in chapter 3. While criminal network researchers have yet to capitalize on tracing methods, with noted exceptions (Papachristos, Braga, and Hureau 2012; Papachristos et al. 2015; and Westlake, Bouchard, and Frank 2011), such methods are instrumental in tracking how information, resources, and so forth can ripple through a network. I hope that reviewing Milgram's work will spark interest in studies aimed at tracing criminal networks.

SUGGESTED READINGS

The overview of sampling, data collection, and data sources provided in this chapter should make it easier to read beefy methodological textbooks. Now that you are armed with a basic understanding of these topics, I encourage you to pick up three books. First, Borgatti, Everett, and Johnson (2013) have written a thorough and accessible textbook covering essential methods and basic statistics. Once you have digested that material, move on to Wasserman and Faust's (1994) weighty classic. When you need to find something in a pinch, consult Knoke and Yang (2008).

In the next chapter, I get into more technical methodological issues associated with entity resolution, establishing network boundaries, missing data, and sensitivity testing. Warning: chapter 8 marks a transition in writing style. Equations are involved.

8. Mapping Networks

> Collecting network data on entire networks, where information is gathered on all actors, and ties are measured for all pairs of actors, requires a great deal of time and effort, especially when networks are large.
>
> WASSERMAN AND FAUST, *Social Network Analysis*

After spending years in computer labs designing and constructing databases, and cleaning data files until my eyes blurred and my fingers refused to type any more, I made a common faux pas—the type of miscommunication that often occurs when a data geek speaks with policy makers or operations personnel. When chatting with someone from the US Army Research Laboratory, I explained that I built a terrorist network using Al Qaeda documents released to the public through the Combatting Terrorism Center housed at West Point.

"Let me clarify," he declared in alarm, "You did *not build* a terrorist network, *you mapped it.* Is that what you are telling me?"

Momentarily taken aback by the glint of amusement and fear in his eyes, I responded, "Huh? Yes, I mapped communication chains to reveal a network."

Then, it dawned on me. Point taken.

When speaking with policy makers, or people who work with policy makers, be careful how you refer to your work. They operate with an entirely different lexicon. As researchers and analysts, we *map* networks, we *do not build* networks. With this lesson in mind, I now turn to the three most daunting issues related to mapping networks—data integration, missing data estimation, and sensitivity analysis. Then I briefly report on a method of identifying corrupted data.

DATA INTEGRATION

Data integration is an essential and routine exercise when mapping illicit activity because of the inevitability of missing nodes and ties (Sparrow 1991). Sometimes the only way to capture a glimpse of interactions among covert operators is to collapse actor and relational data from multiple sources, often developed by different units or agencies for different purposes. Remember two sets of details must be reconfigured—relations must be recoded into a consistent format, and unique actors must be identified. (Attribute data may be optional depending on the research objective.) Of the two reclassification exercises, identifying unique actors, also known as entity resolution, is the most challenging. So we will start with the easier task, consolidating relational information.

Consolidating Relational Information

Recoding the data about the nature of relations is relatively straightforward, but there are many options. One way to code links is to differentiate sources of information or modes of interaction. For example, when integrating phone, text, and email exchanges, these connections can all be considered digitally enabled communications. Digitally enabled communications are appreciably different from face-to-face interactions. In a multiplex communications network, face-to-face interactions could be coded with a value of 2, given the greater intimacy of the contact; digital communications could be valued at 1. Such a valuation system permits the identification of strictly in-person, mixed method, and strictly digital interactions. I doubt anyone would critique this protocol, providing that the coding scheme is explained.

Relational coding can also reflect different types of associations. In Malm, Bichler, and Van de Wall (2010), relational information was extracted from a consolidated threat assessment and classified into four different types of ties—kinship, co-offending, formal organization membership, and legitimate association.[1] Ties were defined as follows:

- *Kinship:* actors tied through a biological or family-based (in-law) relationship

1. The 2007 " E" Division Provincial Threat Assessment report was jointly produced by Criminal Intelligence Service Canada and the Royal Canadian Mounted Police. Using a triangulated data collection strategy, the 2007 report inventories information about individuals from groups that engaged in criminal enterprise during 2004–2006, within the Pacific region, which includes British Columbia and the Yukon Territory. Integrating information from an array of sources (intelligence reports, crime incidents, interviews with law enforcement personnel and prosecutors, review of wiretap transcripts, and offender interviews), the report provides detailed group-level narratives that include information about all known members of the crime group; co-offending information; demographic characteristics; description of criminal activity; role played in the criminal organization; and all known associates, including legitimate business partners, relatives, friends and co-offenders.

- *Co-offending:* all linked individuals committed a crime together at some time during the observation period
- *Formal organization membership:* individuals linked by involvement with a recognized crime group as reported in the threat assessment
- *Legitimate associations:* individuals tied by an association that involved a legitimate business relationship (most involving co-ownership or investment in real estate), a work relationship, a legitimate trade relationship or other noncriminal association

Since intelligence sources about actors involved in criminal enterprise are likely to reveal information about different types of interactions, it was important to capture all information about the associations among members linked to the criminal activity. Generally, even if the actor is not known to play a role in the criminal enterprise, they might in fact be instrumental, knowingly or unknowingly, to operations. Remember, intelligence information is likely to be incomplete, and thus the criminality of individuals may not be known to law enforcement at the time of the initial analysis. Including all types of relations provided us with an opportunity to better understand the context of criminal enterprise and potential resources an offender could draw on in a moment of need. For example, in a subsequent study using the data extracted about 2,197 people linked to criminal enterprise, we discovered that a substantial number of people involved in drug trafficking laundered their own money, sometimes through the legitimate business activity of family members (Malm and Bichler, 2013).

Coding all relational information into different ties permitted examination of each network defined by type of tie, as well as the full, consolidated network mapping all associations. We found that structural differences were associated with networks mapping various binding mechanisms, suggesting that information on both illicit links (co-offender and formal criminal organization ties) and legal links (business and kinship relationships) among individuals in criminal networks is needed to map the networks (Malm, Bichler, and Van De Walle 2010). Ignoring noncriminal binding mechanisms increases the number of missing actors and ties and could lead investigators to underestimate the cohesion and strength of the criminal network. Also remember that co-offending represents joint activity, which generates a graph of behavior, whereas mapping kinship or other more permanent social ties represents relationships. Sometimes it is informative to examine behavior and relational bonds separately.

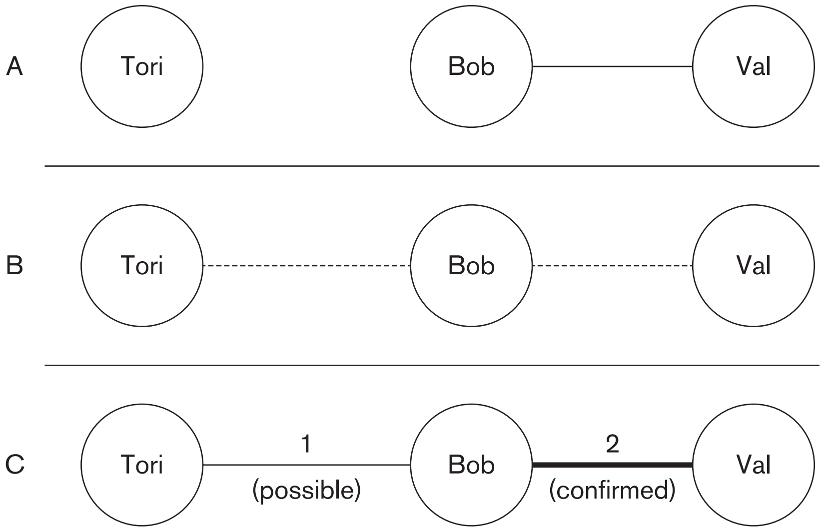


FIGURE 8.1. Relational data integration. *A,* link found in surveillance from a drug trafficking investigation; *B,* co-occurrence at a party as recorded in notes of a gang detective; *C,* consolidated link information.

Commenting on the inherent challenges associated with extracting network data from intelligence sources, Sparrow (1991) argues that it is critical that we remember networks are dynamic. Relations are not static and the context of social interactions is always changing. This means that, instead of looking to see if a connection exists between two individuals, it might be more informative to explore the waxing and waning strength of the tie. Here, tie values could indicate recent contact or coinvolvement, length of association or joint activity, or frequency of interactions. Comparing the level or amount of interactions over time will expose the changing nature of criminal interactions.

A final example illustrates how careful thinking about relations and coding schemes can generate more complete information about group-member interactions, as well as improve the rigor of research. The first panel in figure 8.1 illustrates the linkage between three people that was revealed during a drug-trafficking investigation. An association was found between Bob and Val; perhaps during an interrogation, Bob admitted purchasing methamphetamine from Val. While Tori was seen speaking with Bob before he was picked up, no drug-trafficking interactions were recorded between either Tori and Bob or Tori and Val. Panel A shows the drug transaction link between Bob and Val.

A search for additional information about these three individuals revealed something interesting in the notes of a gang detective (see panel B). A month earlier, during a routine exercise to monitor gang members under a civil injunction, a gang-unit detective noted that Bob and Val were observed exchanging something at the entrance to a parking lot of a gang-controlled apartment complex. A few minutes later, Tori and Bob were observed using methamphetamine at a party occurring near the parking lot. The detective noted this behavior because Bob's gang was under investigation. Tori is an unknown female, and Val has no known gang involvement, but she is associated with a meth cook (the detective used to work in the drug unit).

Panel C integrates all of the information recorded. Because Bob and Val were linked in different sources related to separate occasions, the association suggested in one source (drug arrest interrogation) is confirmed in the other (gang detective's observation). When it comes to Val and Bob's interactions, we have more information to suggest that we captured a real, ongoing association that might be meaningful to our mapping exercise. This confirmed, and potentially more reliable association, is valued by the number of confirmations; in this case, it receives a value of 2. Since the interaction between Tori and Bob was observed once, it receives a value of 1, suggesting a possible relation. It is easy to imagine that as data are assembled, the values associated with each link may increase—it would not take long for some ties to be valued quite highly, while other ties remain valued at 1.

The key to integrating relational information is to be clear about the coding scheme and to code in a manner that provides an opportunity to preform sensitivity analysis. To test the impact of the coding plan used, we could do any number of transformations to see if the positional importance of specific actors or general structure of the network or subgroup changes. For instance, by dichotomizing the network, revaluing all relations worth 3 or more as 1, with relations less than that valued as 0, we can investigate whether networks mapping confirmed associations among actors exhibit structural differences compared to networks mapped with all possible links. In this scenario, at least three independent sources must confirm a relation between actors for the tie between them to be coded as 1, and any tie valued less results in a 0, indicating no substantive link exists. After the analysis is performed, the dichotomization could be set at a different threshold, perhaps 2, and the analysis redone. Comparing results provides insight into how the network is structured when more tenuous associations are included; sensitivity analysis is an important part of testing the robustness of results. We will return to the issue of sensitively analysis shortly.

Entity Resolution

Overriding all methods used to gather data about members of a group and their relations with each other is the problem of entity resolution—that is, applying protocol to ensure that each actor appears in the network under one name or identification number. Because of the peculiarities of software, the spelling and capitalization must be exact. Consequently, if one person is referred to in six ways—Juan Martinez, JUAN Martinez, JMartinez, Juanchi M, JM the Ripper, and Al's brother—then the person will have six different identities in the network. In other words, this one person will be represented by six different symbols, instead of one. Not only does this complicate the network, but the relational position of Juan becomes muddied because his relations are split among six different entities instead of being all attributed to him. Multiply this problem by all the other people who are likely to have aliases and you start to get the idea of why this is so problematic. Figures 8.2 and 8.3 depict what I am talking about.

Figure 8.2 illustrates two networks drawn from unique sources. Each network includes four people. Both networks result from an investigation of Jason, a suspected retail-level drug dealer. (The astute observer might recognize Tori, Bob, and Val from a prior example.) The first network maps surveillance. One evening, Jason was observed spending some time in a car with Val. He stopped by her house, picked her up, and drove to a diner on the west side of town. Val immediately exited the vehicle and walked into the diner. She sat briefly at a table with Bob. Jason remained in the car with a clear view of the table where Bob and Val sat. A few minutes later, Val left the diner and returned to the vehicle where Jason sat waiting. Val used her cell phone. Then, Jason and Val remained talking in the car for about ten minutes. They stopped talking when Tori arrived (on foot). Tori walked into the diner and sat with Bob. She did not look at the car, but, just after she sat down at the table, Jason started the vehicle and drove away.

Meanwhile, one of the detectives on the case mapped wiretap information. Jason was observed to talk with three people on the day of the diner visit. Jason called two people in the morning. He referred to one person as Tina, while the second person appeared to be Amy. Then, around the time he dropped off Val, he received a phone call from Robert, which he did not answer. Figure 8.2 illustrates both networks.

Figure 8.3 provides two visuals illustrating consolidated surveillance and wiretap information. Panel A maps the combined information about which persons Jason interacted with during the day of observations, as well as

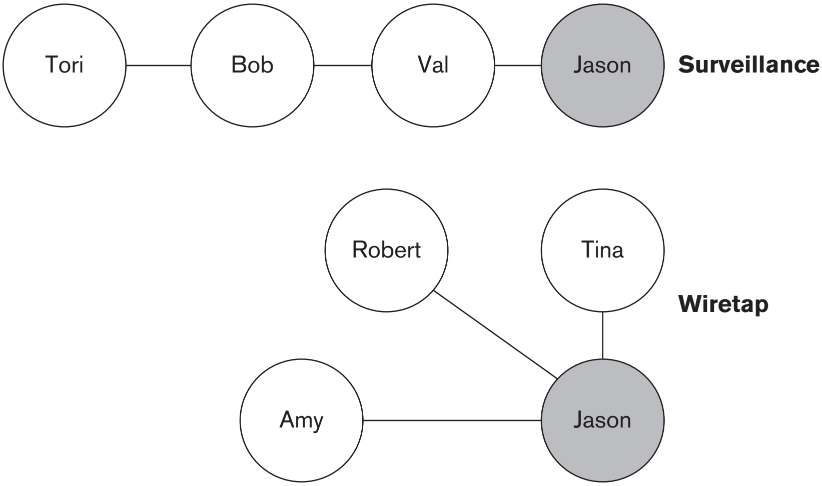


FIGURE 8.2. Unique network structures from two data sources.

whom his contacts interacted with. While Jason did not exchange information with Tori or Bob, they still appear in the network because of the potential links observed during the stakeout. On the surface, the combined information suggests that Jason interacted with four different people. In total, the network includes seven people and exhibits a starlike quality, with most people linked to Jason. Since Jason is the focus of this investigation, the network structure might be an artifact of the data collection process—the network is akin to an egocentric network, with a bit of extra detail about Bob and Tori.

Since people often use nicknames or aliases, the detectives decided to use an entity-matching protocol to ensure that each name actually identified a unique individual. Examining birthdates, location of residence, and occupation (as listed on Facebook), the detectives discover that Bob and Robert were the same person and that Valentina, though not previously named, actually went by two different aliases—Val and Tina. After the identities of the entities were resolved, the network was mapped. Panel B of figure 8.3 now provides a substantively different perspective. There is a cohesive core of three people (Robert, Valentina, and Jason) and two peripheral actors (one woman linked to Jason and one woman linked to Robert). Consolidating information and applying entity resolution protocol may have changed the course of the investigation: the focus of the investigation may now benefit

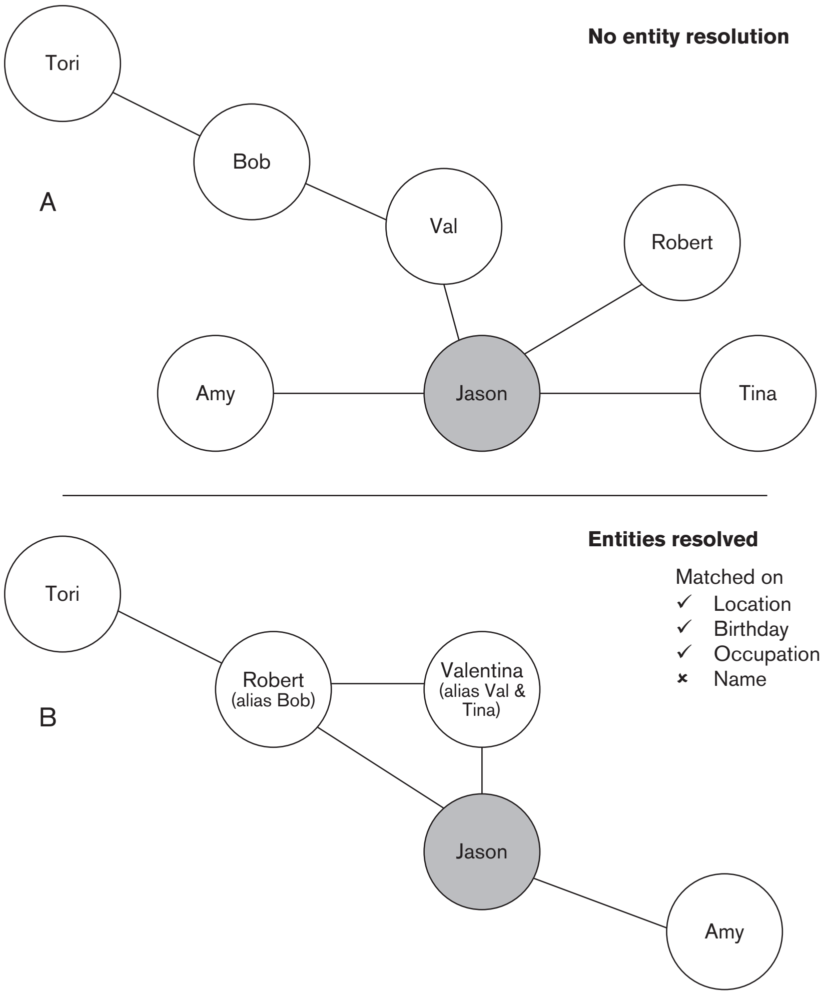


FIGURE 8.3. Illustration of structural change from entity resolution. *A,* network map consolidating information from the investigation; *B,* network map consolidating information after applying a protocol to identify unique individuals.

from an expansion from Jason to his clique. Resolving entities is important to understanding someone's position within the network, as well as network structure.

Entity resolution is important irrespective of the type of network being investigated. In the prior example, I illustrated how the focus of an investigation can change when efforts are made to ensure only unique individuals

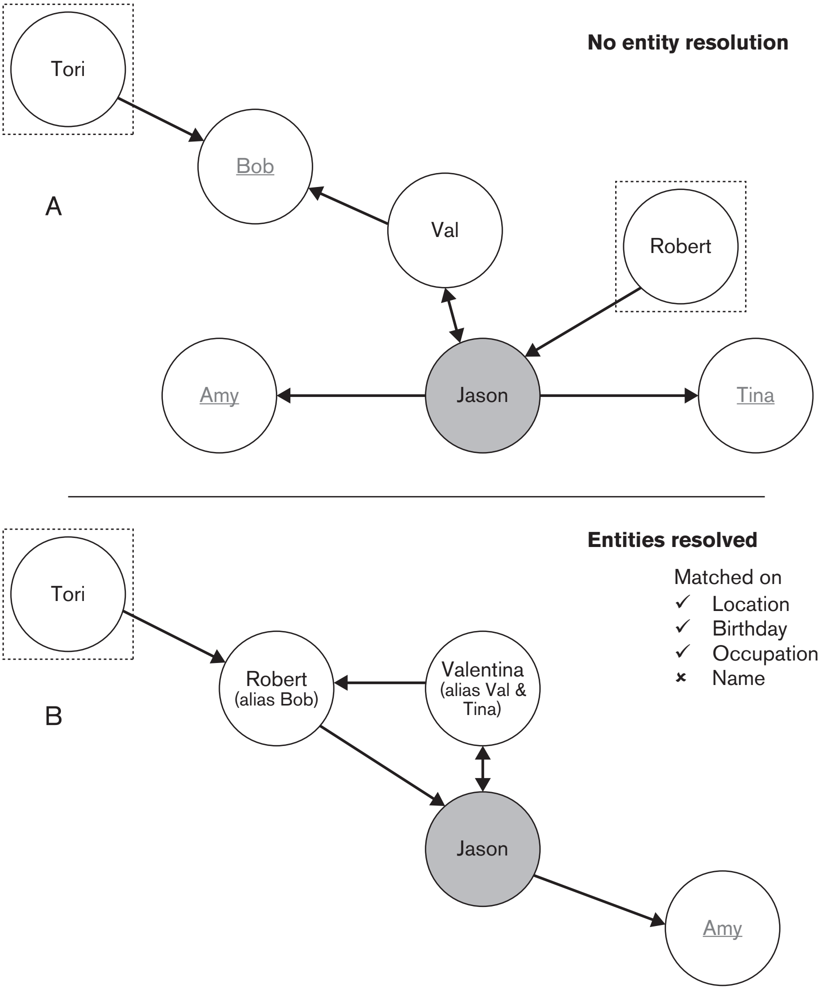


FIGURE 8.4. Mapping directional networks with and without entity resolution. *A*, mapping directional information without entity resolution; *B*, mapping directional relations after applying a protocol to identify unique individuals.

appear in the network; however, there are other considerations. Figure 8.4 adds detail to the networks described above by including the direction of communications. Dashed boxes indicate individuals uniquely positioned to introduce new information into the network, and underlined text identifies people who appear only to receive information—they do not initiate communications.

Consolidating the investigations and adding directionality of contacts, panel A of figure 8.4 illustrates what the network might look like without entity resolution, and panel B depicts the network after entities were resolved and only unique individuals were included. As with the prior example, the differences are striking. The investigatory implication exhibited in panel A is that two people, Tori and Robert, are in the unique position of initiating activity or introducing information to the group. More importantly, Jason is a central relay, as he is positioned to both receive and distribute. Three people, highlighted by underlined text, only received information. Jason appears to be a likely law enforcement target given his central placement in the network. Contrast this supposition with the structure found in panel B.

With entity resolution and directionality added, network flow suggests that there is a central group, not a central person. Tori remains uniquely positioned to introduce new information, which circulates through a clique comprising Robert, Jason, and Valentina. Jason is instrumental but in only two capacities—in completing the feedback loop of the clique and in passing information on to Amy. It may be advisable to continue to target Jason, but the cohesiveness of the clique suggests that it may be relatively simple for contact with Amy to be reestablished by Robert or Valentina soon after Jason's arrest. The network is likely to be more resilient than first thought (referring to panel A).

Entity resolution is just as important when dealing with groups. Recall the gang injunction project described in chapter 7. In this study, we are mapping gang-on-gang conflict over a twenty-year period. Information about violent crime is being pulled from case transcripts posted on Westlaw and LexisNexis. The documents report decisions from appellate court, and, as such, include only a brief summary of the facts and relevant testimony to support judicial decisions. Detailed gang involvement is available, as most cases involve gang-enhancement penalties. This means that gang experts testify about gang affiliations. Even so, entity resolution is a major issue. Los Angeles gangs are large, with many subsets or cliques; groups go by different names, naming conventions evolve over time, and some experts are not clear about which gang individuals belong to. To resolve this problem, we are using an entity-matching protocol that draws on supplemental sources with built-in time stamps. Just as entity resolution for an actor-to-actor network draws on such factors as name similarity and conventional nicknames, demographic characteristics, location, hobbies, and network position, as listed below, equivalent criteria can be developed for groups.

Individuals	**Groups**
Name similarity	Name similarity
Demographic characteristics (i.e., date of birth, gender, ethnicity)	Date founded or group characteristics (i.e., type of criminal activities, symbols, uniform)
Location (home city, url, or address)	Geography (i.e., home turf, safe zone)
Hobbies	Rivalries and alliances
Network position	Organizational hierarchy or subgroup linkages/structure (i.e., clique-gang associations or organizational structure)

(If you are curious, while the final results will not be out for some time, a description of a pilot study offers some preliminary insight; See Bichler, Norris, Dmello, and Randle (2017) for more details about this project.)

When dealing with small networks, it is feasible to engage in manual processes to resolve entities; however, when working with large networks, which are likely to contain hundreds (or thousands) of different people, manual data integration procedures are not feasible. In these situations, algorithms are necessary. Depending on the complexity of the matching process and number of nodes to compare, some machine learning is necessary to train programs. Typically, machine learning uses extracted features of user profiles—name-based, user information, and network topological features—to resolve entities. For instance, pair-wise similarity scores of user profiles can be combined with a network similarity score (overlapping social networks) to integrate files.

Training and testing require several tables of entity information containing two specific subsets: (1) a known set of people who are the same across data files, and (2) a set of people who are not the same but have similar characteristics. The sizeable body of literature on this topic is too varied and deep to review here; suffice to say that the tested performance of different entity-resolution algorithms is generally deemed successful when the algorithms achieve a reasonable level of accuracy (e.g., 80–90 percent). To reach a high level of accuracy, entity resolution algorithms use at least two sets of matching characteristics—similarity in a set of individual characteristics (more than just names), combined with either similarity in posted content or similarity in network positioning (or topography).

One reason why investing considerable resources into merging and cleaning data files is important is to improve the completeness of the data. What most hinders the use of social network analysis (SNA) is excessive levels of missing data. In the next section, I review two techniques for estimating the amount of data that might be missing from a network.

Estimating the Amount of Missing Data

One of the most concise and informative discussions about the effects of missing data can be found in Knoke and Yang (2008). These authors describe several equations that could be used to think about the completeness of data collection efforts. In the interest of brevity, only two calculations will be described here.

Egocentric Networks

Imagine that you are interviewing parolees involved in a reentry program about their social support network. You ask each parolee whom they rely on for support when they are faced with a situation that could lead to a parole violation. Then you ask if any of the people mentioned know each other so well that they would initiate contact with each other, if they were worried about them. Since individuals have limited knowledge about the associations among alters, a potential bias threatens the study—complete information about the links between alters is needed when mapping an egocentric network. You will need to estimate the **relational response rate**, or percentage of relational information that might be missing.

Knoke and Young (2008) recommend using the following equation to estimate the percentage of missing information that you may have upon completion of the interview. The first step is to calculate the number of potential relations that would exist among alters; here, N is the number of alters nominated by the interviewee.

$$C_N^2 = \frac{N!}{2!\,(N-2)!}$$

So if the parolee, the ego, were asked about his or her relations with six people, the interviewer would have to retrieve information on the existence of all potential relations among the six contacts (relations among alters). Using the suggested equation (depicted below), six alters would generate the need for information about fifteen possible nondirected ties among alters.

$$C_N^2 = \frac{6!}{2!\,(6-2)!} = \frac{(6 \times 5 \times 4 \times 3 \times 2 \times 1)}{(2 \times 1) \times (4 \times 3 \times 2 \times 1)} = \frac{720}{48} = 15$$

If the focus of the data collection were to identify directed ties, then you would simply multiply the answer by 2—meaning you would calculate

$2C_N^2$. In this example, your interview would need to enquire about thirty different possible associations among alters: for example, "Would Bob call Charlie?" is not the same as "Would Charlie call Bob?"

The next step is to divide the number of reported relations among alters by the number of relations that might exist. Let us suppose that the respondent confidently expressed knowledge about twelve relations among alters (this could mean that relations definitely exist or perhaps alters do not know each other). Knoke and Yang (2008) suggest that to calculate a relational response rate for an undirected network one simply divides the reported number of ties (present or absent) by the possible number of ties. The relational response rate (R) would be the following if we plug the results from the equation above into the denominator.

$$R = \frac{T}{C_N^2} = \frac{12}{15} = .80$$

If we multiply by a hundred, the response rate for the ego is 80 percent, which means 20 percent of relational data is likely missing—the parolee is unaware if relations exist among some of the alters. Repeat these calculations for each person interviewed. To summarize, average the scores to figure out the estimated mean level of missing data for the study.

Full Networks

Estimating the level of missing data in a full network is more complicated than the egocentric calculation for two reasons. Nodes *or* relations could be missing from the data file. Consequently, when assessing the completeness of network information we need to consider both the *nodal response rate* and the *relational response rate*. Nodal response rates are easy to calculate, if you know the membership of the group. If there are six members of the gang, and you have pulled within-gang, co-offending information on five, then the nodal "response rate" is five of six or 83 percent. I put response rate in quotation marks because criminal justice applications often map networks from secondary data sources (rather than asking people directly about their co-offenders). Obviously, there is a serious limitation to this estimation process. We are assuming that the secondary source captures all co-offending information. This is a very unlikely scenario. To resolve this issue, merge data from multiple files to increase the likelihood that people are not missing from the network (see the prior discussion.)

The calculation of relational response rates depends on the type of network. To estimate the number of relations you should know about—

meaning that you have data on the existence or absence of a tie—simply use one of the following equations:

Nondirected dyadic relations: $2 \times N$; where N is the number of actors.

Directed dyadic relations: $N^2 - N$; where N is the number of actors.

With respect to the gang with six members, a nondirected dyadic network file should contain relational information on twelve different pairs (calculated with the first equation, or 2×6). This means you need to find out if co-offenses occurred involving twelve combinations of people. If you were mapping a directed network, say you wanted to note who initiated co-offending (ego) and who joined in (alter), then you would need to know about $6^2 - 6$, or thirty potential partnerships.

One simple way to calculate the relational response rate is to divide the number of relations you know about by what you should have information on (then multiply by a hundred to generate a percent). Keep in mind that when you sum up the relations you have information on, you need to count the ties (or arcs) that are present and the ones you know are absent. Missing ties (or arcs) are the relations you do not have information about. (Once you dive deeper into SNA, consult Knoke and Yang [2008] for a more sophisticated way of estimating the level of missing data for whole networks.)

There are a couple of additional points to make before moving on to sensitivity analysis. If the network is nondirected and one actor fails to report a relation, you might still pick up the information from the other party. For example, if you were investigating the weapons trade, Israel might fail to report receiving weapons from the United States. If the United States were reporting the export, however, then the relation would not be missing from your network. In essence, you have two opportunities to acquire the trade data—imports reported by Israel and exports reported by the United States. The second point is that nodal response rates are more impactful to directed networks because of the inherent asymmetry in many types of relations. If the network is directed, a missing relation may exist in one direction but not the other, and you may not be able to estimate what is missing.

These missing data calculations are all well and good if you have a clear picture of the group, community, or population you wish to investigate. But what if you are dealing with a hidden population of actors—a criminal group, whose membership is covert and whose relations are intertwined with legitimate ties? Covert groups often have **fuzzy boundaries,** meaning it is difficult to decide which persons should or should not be included in the group—the boundary of the group is blurred (Sparrow 1991). And with crime embedded within routine behaviors (see chapter 5), isolating criminal associations from legitimate ties can be a matter of splitting hairs. In such a situation, it is best to spend some time conducting a sensitivity analysis before going forward with the results of your study.

SENSITIVITY ANALYSIS

As discussed in the beginning of this chapter, when assembling network information, it is possible to use a coding scheme that quantifies how confident you are that a relation exists between a pair of people. Using a filtering function, you can then repeat analysis with different cuts of the data—networks using all information compared to networks made up of a subset of trusted data about actors or relations—to see if the results are stable. For instance, does the structure change remarkably? Are the most central figures still appearing in the top ten?

More formally, sensitivity analysis involves recalculating outcomes under alternative conditions or assumptions. Typically, networked criminologists aim to investigate the robustness of findings, given different expectations of data completeness. Since there are many different analytics available, there is no formal set of specific parameters or analyses to choose from. Instead, most people simply apply different inclusion rules to actors or relations, thereby generating different networks on which they repeat the analysis to see if the results are similar to those produced in the initial investigation. Alternatively, researchers add (or delete) nodes or ties at random or use a specific criterion (e.g., preferential attachment). Compare the original results against those produced in successive runs with the number of actors or ties gradually increasing or decreasing to estimate the robustness of findings. Generally, these investigations increase (or decrease) the sample size with several intervals of 5 or 10 percent—that is 5 percent more, followed by 10 percent more, and so on. Repeating the tests with varying numbers of links shows whether the findings are robust if relational data are missing; repeating the tests with varying numbers of actors shows whether the findings hold in the face of missing actors.

A good example of this process was written up by Xu and Chen (2008). Using three networks—a Global Salafi Jihad network of 366 terrorists mapped by Sageman (2004), a co-offending network of 1,349 people involved in trafficking methamphetamines in Tucson from 1985 to 2002, and a network of 3,917 people involved in gang-related crimes in Tucson from 1985 to 2002—these authors investigated the effects of missing links on the small-world and scale-free properties of the groups. Three link generation functions were tested.

1. *Random effect.* Links were added to randomly selected pairs of actors who were not originally connected.
2. *Common neighbor effect.* Links were added to pairs of unconnected actors who shared common neighbors, determined on the basis of the number of shared connections.
3. *Preferential attachment effect.* Links were added to unconnected pairs depending on the product of their associations—there was a higher probability of being linked if both actors were connected to many others.

They found that adding up to 10 percent more links with either function did not materially change the small-world and scale-free properties of the networks.

Networked criminologists also use sensitivity analyses to simulate how attacking the network might compromise its structure. Xu and Chen (2008) did this as well, and it is well worth reading the study. I would also like to direct your attention to Duijn, Kashirin, and Sloot (2014). The latter study was profiled in chapters 6 and 7, and I refer to it again here for good reason. Their highly rigorous study has a relatively accessible format. (Although, I strongly encourage you to refrain from reading it until you finished this book! Chapters 9 and 10 will help.)

Duijn, Kashirin, and Sloot (2014) tested five different strategies for identifying actors to remove from the network (via arrest), in an effort to destabilize the structure, making it unlikely actors would be able to continue drug-trafficking activities. Generally, destabilization or disruption is assumed to have occurred when (1) connected networks become unconnected (dissolve into unconnected components) as a result of the removal of key actors or functions, (2) the system becomes less efficient, or (3) cohesion levels change significantly. In this study, the authors looked at efficiency and cohesion. The simulation models were tested with zero to thirty actors removed using each attack strategy. Recall from our prior conversation that the five attack strategies were

1. Random strategy, with no preference in the selection of candidates for removal;
2. Target actors with the most contacts (hubs in the trafficking network);
3. Target actors who act as central bridges joining parts of the network that would otherwise be unable to connect;

4. Select individuals playing key roles in the drug production and distribution chain; and
5. Actors are targeted owing to their special skills or knowledge.

Duijn Kashirin, and Sloot (2014) also simulated possible recovery processes that could be deployed as drug traffickers rebuild the system after an attack. They discovered that none of the strategies materially affected the efficiency of the trafficking system, and in fact efficiency actually increased if actors were targeted for their roles in the drug-production chain. Moreover, the network became more resilient after attacks, leading the authors to conclude that "a delicate criminal process, such as cannabis cultivation, is organized in a flexible and adaptive network structure, which is highly resilient against network disruption" (13).

Sensitivity analysis can also be used during the course of an investigation to assess progress and avoid the establishment of cognitive biases, probability errors, and organizational traps that could derail even well-resourced investigations (Bichler, Lim, and Larin 2017). Working with a San Bernardino County deputy sheriff and a second lieutenant in the US Army, I investigated the utility of using SNA metrics to gauge the importance of suspects as they emerged in the Green River serial murder investigation. Using an actor-to-actor network linking eighty-eight key witnesses, victims, and suspects who frequented the same places (1,304 ties), we demonstrated that significant shifts in the centrality of suspects emerged when we tracked the evolution of this case at six-month increments. Most notably, by eighteen months into the investigation, the most central suspect bridging all other clusters of people was Gary Leon Ridgway, the person responsible for killing at least forty-nine women, and not the prime suspect, Melvyn Foster, who had been identified early on in the investigation. We tentatively concluded that the initial working case hypothesis had misled investigators. And that in regard to investigatory policy, testing the robustness of suspect positioning in the network may be useful when tracking the progress of an investigation.

Detecting Corrupted Data

Digit Analysis

We are often confronted with the prospect that our criminal network information may be systematically missing data or corrupted intentionally. While corrupted data are problematic to the external generalizability of the study, detecting it may still be useful to crime control efforts. For instance, you might examine the financial transactions for a set of companies looking for an indication of money laundering. Of interest to those checking for measurement error and data fabrication or those engaged in fraud detection is that a digital analysis using Benford's Law has proved useful in the identification of manipulated and artificially created numbers (e.g., Dickinson 2014; Gunnel and Todter 2009). **Digital analysis** involves detecting abnormal patterns (e.g., recurrences of digits, digit combinations) by examining digits and number patterns in sets of data.

To identify suspicious patterns indicative of corrupted data, you begin with a set of numbers and then extract the first digit for all values in the set of scores. When used in an SNA study, the scores could be a distribution of measures rating each actor's position within the network. Next compare this distribution to *Benford's Law of first significant digits.* Benford's law states that the first digit (the leading nonzero digit) for a large set of naturally occurring numbers follows a predictable pattern. If you create a frequency distribution of values with the first digit—one through nine while ignoring all zeros—ones should occur 30.1 percent of the time, twos should occur 17.6 percent of the time, threes should occur 12.5 percent of the time, and so on (Benford 1938). The first row of table 8.1 reports the expected distribution (in proportions).

When patterns in your observed data significantly deviate from the expected distribution, this can indicate that some kind of data manipulation has occurred or that parts of the data set are missing. Numbers that are randomly selected, sequentially assigned, influenced by human thought, or assigned within a constrained minimum or maximum will not conform to the expected pattern. Comparing the suggested distribution to the observed data will indicate which digits are over- or underrepresented. Another way to compare the distributions is with a Pearson's Correlation Coefficient. When Pearson scores are in the .990 to .999 range, corruption is unlikely to exist in the observed data. An illustrative example follows.

Example Digital Analysis of Global Weapons Trade

To investigate the complex system of small arms and light weapons trade, Bichler and colleagues (Bichler and Birks 2015; Bichler et al. 2015) developed a computational agent-based model to simulate global illicit and legal transfer activity (see box 8.1).[2] Such models permit us to explore trade

2. An agent-based model allows researchers to construct artificial worlds inhabited by autonomous, heterogeneous, goal-directed actors interacting with each other to produce measurable system outcomes. Models are made of a population of agents and a simulated environment. Agents store individual characteristics, as well as behavioral algorithms that allow them to act on those internal conditions and the

BOX. 8.1 GLOBAL SMALL ARMS AND LIGHT WEAPONS TRADE MODEL

The model comprises 224 agents (nodes), each representing a single nation or territory. The edges that connect agents represent weapons transfers. Agents consider their current situations and negotiate trades with other nations to meet their weapons needs. Choosing between suppliers involves evaluating the characteristics of suppliers and the shipment routes through which each can provide firearms.

Trading. Quarterly, nations evaluate how many weapons they need and apply three weighting functions (costs) when considering trade partners: (1) transparency of trade relations, (2) whether embargoes exist, and (3) social and political homophily between nations. A simple depth-first search algorithm ascertains the least costly path between each purchaser-supplier dyad.

Skimming. Firearms stock moves between nations involved in the trade path. If this path is not direct—firearms are transshipped through one or more additional nations—there is a chance that some quantity of firearms may be skimmed from the shipment. Skimmed units are classed as illicit and added to national stockpiles.

As the model runs, it generates a series of outputs—trade paths selected by each nation (used to map the network), quantity shipped (trade value), and current stockpiles. For more information about this model, see Bichler et al. (2015).

dynamics and conduct experiments to assess the likely success of crime control strategies seeking to disrupt illicit trade without incurring political, economic, and social costs. The problem facing us was how to validate the model.

As noted in box 8.1, the agent-based simulation model generated a number of outputs, all of which could be compared to observed data were accurate global estimates accessible. Unfortunately, this was not entirely feasible. We built some indication of global trade but did not know if it was complete.[3] In addition, the model generated quarterly estimates of trade

---

perceived world around them. Behavioral rules, often derived from theory or parameterized from historical data from a target system, dictate how agents perceive, reason, act, and interact within an environment.

3. We began by assembling all the information we could find on military-grade small arms and light weapons transfers among 224 nations and territories. Our primary source of information was the UN Commodity Trade Statistics Database.

TABLE 8.1. FSD analysis of military SALW transfers (2002–13)

Data set	*Probability of first significant digits*									*r*	*N*
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*		
Benford's Law	.301	.176	.125	.097	.079	.067	.058	.051	.046		
Trade structure											
Simulated degree centrality	.304	.192	.133	.093	.087	.071	.051	.038	.030	.995	2,679
Observed degree centrality	.341	.198	.134	.093	.071	.058	.045	.034	.025	.999	2,162
Other model outputs											
Trade value (simulated)	.143	.110	.115	.120	.123	.124	.122	.085	.061	.576	45,530
Stockpiles (simulated)	.365	.053	.272	.086	.014	.021	.085	.059	.045	.781	10,752

NOTE. All reported correlations are significant at the $p < .001$ level.

value and current stockpiles, neither of which are available globally. Consequently, we used a first significant digit (FSD) analysis to assess the completeness of data capture, as well as potential distortions in simulated weapons trade. Table 8.1 reports that while both the observed network and simulated trade activity generate actor positional scores (degree centrality counts the number of direct transfer partners) that appear to follow the expected distribution, the correlation coefficients for trade value (quantity shipped) and stockpiles suggest that the model generates constrained values. The low r values (.576 and .781) tell us it is likely that the rules governing trade behaviors require modification, because the complex system is not functioning well. As you may have guessed, we have gone back to the drawing board to figure out what is wrong with the model!

FINAL WORDS OF ADVICE

As noted by Wasserman and Faust (1994) in the opening quotation of this chapter, gathering relational information for a group of actors is difficult. Summarizing conventional wisdom about the challenges generally associated with gathering network data, Wellman (1988, 40–47) suggests that, when developing methods to generate network data, it is best to remember the following:

1. Ties are usually asymmetrically reciprocal, differing in content and intensity.
2. Ties link network members indirectly as well as directly. Hence, they must be defined within the context of larger network structures.
3. The structuring of social ties creates nonrandom networks—hence clusters, boundaries, and cross-linkages.
4. Cross-linkages connect clusters as well as individuals.
5. Asymmetric ties and complex networks differentially distribute scarce resources.
6. Networks structure collaborative and competitive activities to secure scarce resources.

From the disclosures of 165 to 188 reporting nations (varying by year), we generated trade flow between origin and destination for 224 nations and territories for the study period (2002–13) by integrating information from imports and exports. Next we added details about illegal trade: (1) illicit weapons trade identified in reports of the Small Arms Survey; and (2) news reports about seizures made at border crossings, shipping ports, and airports. The final, directed country-to-country transfer network included reported, suspicious, and illegal trade, where one denotes some type of trade and zero denotes no known trade.

Networked criminologists must also remember that our efforts to assemble complete actor and relational information may require even more resources than we originally anticipated, because membership information is hidden, obscured, or intentionally distorted. So double (or triple) the resources you plan to set aside to generate the network data files. And no matter how much effort you put into to mapping the network, expect that it is incomplete. With these lessons in mind, you are ready to advance to chapter 9—describing networks (with statistics).

9. Describing Networks

> An obvious question to ask is why anyone would want to analyze social network data. The uncontestable answer, of course, is because they want to.
>
> BORGATTI, EVERETT, AND JOHNSON, *Analyzing Social Networks*

What sets scientific enquiry apart from other forms of knowledge building is the adherence to a set of methodological principles designed to ensure that studies are reproducible and findings are credible—meaning there is a low probability that bias has materially distorted conclusions. To accomplish this feat, procedures must be systematic, rigorous, and *public.* Note the emphasis on public. The importance of this term will become clear momentarily.

To achieve the standard of being systematic and rigorous, the same procedures must be followed with every observation (a.k.a. datum gathered), and the inquiry must be as thorough and bias free as possible. Moreover, the data used to develop conclusions must be **collected objectively,** relate to **observable phenomena,** and represent an **aggregation of experience.** Definitions for bolded terms follow.

Collected objectivity: observe things as they are with no attempt to falsify observations to fit with a preconceived idea.

Observable phenomena: study subject matter that can be observed with some tool or instrument by more than one person at a time.

Aggregation of experience (issue of probability): gather data from a wide range of situations to reduce the likelihood of making conclusions on the basis of anomalies.

As noted previously, there is a standard of publicness. All dissemination of research findings—whether shared verbally, in print, through mass media, or via the web—must be accompanied with detailed information about the process or method used to collect the data (covered in chapter 7) and the limitations or problems faced during the research process (covered in chapter 8). Authors must also describe data in summary form. The latter standard is the focus here. We will review basic statistics that are widely used to describe whole networks, actor positions, and egocentric networks. Where possible, I define key statistics and provide notes about when to use them, as well as a list of supplementary resources. A warning: chapter 9 contains equations and is long! Do not read the entire chapter in one sitting.

DESCRIBING THE WHOLE NETWORK

General Descriptive Statistics

Descriptive statistics provide a framework for understanding structure. Often the framework provides a context for understanding the focal analysis driving the study. For example, knowing something about the structure of the whole network provides a reference for a more detailed investigation of identifiable subsections. When describing the whole network, it is generally expected that reports will include the number of nodes and ties, number of components, density, average degree, degree centralization, and average path length for each network under examination. This means that if a subset is drawn, then two sets of values may be necessary—the descriptive statistics for the full network and the descriptive statistics for the subset. To assist your introduction to descriptive statistics, I supplement hypothetical examples with real-world examples drawn from the Al Qaeda and Islamic State of Iraq (ISI) terror-group communication study that was described in chapter 7. The real-world examples illustrate how to report and interpret the statistics. To refresh your memory, see box 9.1 for information about how this network was mapped.

Descriptions of whole networks begin with reporting the number of nodes and ties observed. In prior chapters, we already covered what a node and tie constitute. Combined, these two details equate to providing basic information about the sample size. Box 9.2 reports both the total ties and the unique ties as the network was valued—multiple messages often passed between sets of individuals. The total tie count indicates the volume of communications observed—if a pair of people were observed to communicate fifteen times, the connection between them would contribute fifteen ties to the total tie count. But this information might be misleading, as we have no idea how specific documents were chosen for release to the public. Was this a random selection of documents? Or were the communications of specific people targeted, which means more of their exchanges were captured compared to other actors in the group. Given that this snapshot of interactions may have been biased by the set of documents we examined, investigating

BOX 9.1. MAPPING TERROR-GROUP COMMUNICATION CHAINS

To map a communication network linking central command staff, Stacy Bush and I extracted message chains from two sets of documents released through the Harmony Program by the Combating Terrorism Center at West Point (Bichler and Bush 2016). The first set contained thirteen declassified and translated documents (154 pages of text) recording inter- and intraorganizational operations of Al Qaeda dated from September 2006 to April 2011. The second set contained seventeen declassified and translated documents (58 pages of text) recording intraorganizational communications of the Islamic State of Iraq occurring between 2006 and 2009. Actor-to-actor links were generated by extracting messages embedded within each document. Since messages were often sent via indirect routes, each person mentioned in a message was recorded. The originator, listed as the ego, was linked in a chain to subsequent intermediaries, and so on, until the message chain terminated. An edge list was generated linking each set of dyads. This set of directed, valued communications constituted a community of actors with distinct subsets of information channels among command staff. Joining the two sets of actors were eight individuals who appear in both intelligence sources.

unique ties may be useful. Unique ties may reveal something about the intrinsic architecture of communication channels. The pattern of unique linkages among group members may provide a better context for understanding how information could flow within the community.

In comparing the Al Qaeda and ISI subnetworks, it is clear that ISI intelligence revealed less frequent information exchanges. Most communication channels were unique and likely observed only once in the documents, whereas, with 1,132 messages passing along 362 unique ties, the observed linkages of the Al Qaeda network were likely to support multiple exchanges. On the surface, this information may indicate that the set of intelligence used to map the Al Qaeda subnetwork may be more complete. Mapped communication paths may constitute a more complete set of ties, because on average the source captured multiple exchanges between parties. The ISI source documents, however, did not. Readers should know that one unique tie was found in both intelligence sources; thus, the total number of unique ties for the whole network is 541, not 542. Moreover, eight people are common to both networks; consequently, the total number of actors in the combined dataset does not equate to a sum of the two intelligence sources.

BOX 9.2. REPORTING SAMPLE SIZE

Network	*Total ties*	*Unique ties*	*Actors*	*Components*	*Actors' main component (%)*
Al Qaeda network	1,132	362	195	1	100.0
ISI network	215	180	115	11	75.7
Combined (all data)	1,347	541	302	10	90.7

The second piece of information to report is the number of components. Readers should be familiar with **components**—a set of connected nodes, wherein every node is connected to at least one other in the set. In addition to reporting the number of components, authors must also provide their readers with the percentage of actors appearing in the *main component*. This convention is important. Many standard descriptive statistics and analytics focus on the main component, so readers must be able to reflect on the proportion of the network to which the reported findings pertain.

Reviewing the number of components reported for each network in box 9.2, we discover that the larger number of components found in the ISI network might further support an argument that the set of intelligence used to map communications is incomplete—missing actors or links might explain the lack of connectedness. Knowing, however, that 75.7 percent of all actors are contained in the main component, despite the larger number of ISI components, reassures us that, while missing actors and links are noteworthy, the ten small disconnected subsets are unlikely to derail the analysis. Instead, this information could help focus continued investigation: I would focus subsequent data-gathering efforts to expose how these small subsets link up with the main component.

Turning to the whole network, most, or 90.7 percent, of actors are part of the main component of the combined network. This finding suggests that the eight individuals common to both terror groups unite the two main components of the subnetworks and join the main component with one of the previously disconnected smaller ISI components (otherwise, there would be twelve components). The high percentage of actors in the main component suggests that amalgamating intelligence sources, rather than treating the sources separately, improves the likelihood of generating a more complete snapshot of communication linkages within this community. This is not to say that understanding of the internal structure of specific groups (e.g., Al Qaeda or ISI) has improved; rather the representation of the communications among the larger community within which the groups are embedded improved.

The next step in describing the whole network brings us to the first statistical equation covered in this chapter. Pay close attention. Although basic, this equation sets the stage for what follows. To begin, I will ignore the fact that the original network files are valued and directed; instead, the networks will be treated as if they are **symmetric** and **dichotomous.** A symmetric and dichotomous representation of a network is the most basic mapping of relational structure. As we learned in the prior two chapters, criminal network research is plagued by questions of data completeness. Examining the networks in their simplest form can provide an overview of the organizational structure. In a sense, the transformation is akin to recoding ratio-level data to nominal level data. While we lose precision, we might gain a better understanding of the skeletal structure supporting the social fabric under investigation. The valued terror-group communication network was transformed using two functions—it was symmetrized and dichotomized as explained in box 9.3.

Structural Description of a Whole Network

Two objectives guide how we describe the structural patterns of relations in a whole network: painting a picture of the overall level of cohesion and centralization observed in the network. **Cohesion** is an important concept. When applied to whole networks, it means that a range of statistics is used to investigate patterns of connectivity in order to gain a general understanding of the overall level of constraint that the network might impose on its members. If you are well versed in urban-planning concepts and the work of Jane Jacobs, you might think of this as the tightness in the weave of the social fabric interlinking a group of people or a community. Network investigators are also interested in the **centralization** exhibited, which means within any group of actors we might find someone who stands out—for example, a community leader, popular person, or drug kingpin. Below, I will explain three sets of statistics, which when reported together provide some indication of the overall cohesion and centralization observed in a network.

I begin with density. **Density** is the simplest measure of cohesion. This statistic reports the proportion of ties observed between pairs of nodes

BOX 9.3. TRANSFORMING NETWORKS

The original terror-group communications network was generated as a directed and valued network. To illustrate the process of transforming this network into a nondirected and dichotomous structure, two functions were used. I created a simple hypothetical example to explain the process. A simplified version of a valued, directed original form of the network appears in the first example matrix below (left side of the page). This hypothetical network includes four actors (labeled A through D) with twenty-six ties, which are called arcs because they are directed. To calculate the number of arcs, simply add all the cells. Since communication could not flow through a recursive loop, meaning emails or message were not recorded if someone sent something to themselves, no values appear along the diagonals. We know the network is directed because the values and ties are not all reciprocated—A contacted C three times (light gray cell), but C did not contact A (dark gray cell).

Valued (directed)

	A	B	C	D
A	–	1	3	0
B	9	–	0	5
C	0	1	–	0
D	0	5	2	–

g = four actors
l = twenty-six arcs

Dichotomized (directed)

	A	B	C	D
A	–	1	1	0
B	1	–	0	1
C	0	1	–	0
D	0	1	1	–

g = four actors
l = seven arcs

Symmetrized (not directed)

	A	B	C	D
A	–	1	1	0
B	1	–	1	1
C	1	1	–	1
D	0	1	1	–

g = four actors
l = ten edges

A function to dichotomize the network was run, with all values of 1 or greater assigned a value of 1 and all others remaining 0. The resulting network matrix depicted in the middle has fewer ties but retains all original directed communications among the four actors—seven arcs remain to represent unique ties. Notice the change in the light gray cell. The value of 3 was recoded to 1, and the dark gray cell remained 0. Dichotomizing reduces the number of relations in the network.

The final transformation involves symmetrizing relations. Here we ignore directionality. So if a tie exists between A and C, then it is also assumed to exist between C and A. Compare the matrix on the right side of the page to the others. Symmetrizing often increases the number of unique ties in the network. Now the two halves of the matrix match—the top triangle on the north side of the diagonal of empty cells and the bottom triangle on the south side of the diagonal are mirror images of each other.

relative to what would exist if everyone were connected. A general form of the equation is

$$density = \frac{l}{g(g-1)/2}$$

where l is the number of ties observed and g is the number of actors in the network.

Multiply the value obtained by a hundred to report the percentage of ties observed. Networks with scores nearing 1.0 are considered very cohesive. For instance, a value of .9 (or 90% if the density proportion is multiplied by 100) indicates that nearly everyone in the network is directly connected to each other. This is an unusual situation in a large network but is feasible for a small group. Low scores, such as .12, suggest that only 12 percent of possible connections are observed.

So what does density really tell us when applied to a communications network? Consider a tight-knit group of people to which you have belonged—perhaps the group of friends you grew up with. Everybody knew each other and spoke on a regular basis. This is an example of a very tight network with a density of perhaps 90 percent, meaning that if all telephone calls were observed, 90 percent of the possible exchanges between pairs of people actually occurred. This does not mean that each person spoke with only one other person. Because density is a dyadic metric, we must break down more complex exchanges into the simplest parts. If a group of five people had a teleconference, and everyone spoke to each other, then the five-way conversation would be broken down into ten exchanges or communication ties (ten dyads).

Now think about the people you work with. If you work in a big organization, you might be close to a few people, but it is unlikely that you are in regular contact with everyone. In fact, it is more likely that there are many employees with whom you have never interacted. In this scenario, we would expect a very low-density score. Interpreting what constitutes a low density depends on what the network represents and the length of the observation period. If we found that 10 percent of the possible communication channels were observed among a work unit of five people over the course of four months, I might be inclined to say that density is extremely low. If instead I obtained a score of 10 percent for a working group of twenty people over the course of one hour (with group meetings not captured in the data), I might be inclined to suggest that a moderate level of interactions occurred between pairs of people.

BOX 9.4. DESCRIPTION OF NETWORK STRUCTURES

Network	*Density*	*Average degree (SD)*	*Normed average degree (SD)*	*Degree centralization*
Al Qaeda network	5.37%	3.17 (5.73)	1.63 (2.95)	32.72%
ISI network	2.62%	2.44 (2.57)	2.14 (2.26)	13.00%
Combined (all data)	2.61%	2.96 (4.95)	.98 (1.64)	21.08%

There are a few other things you should know about density. Even if the network is valued, the default calculation always assumes a dichotomous association; valued relations are converted into binary ties before estimation. Also, density is calculated on the largest component. Note that symmetry is not assumed. Asymmetric, directed networks will have a lower density than if you ran this calculation on a symmetrized version of the network. For instance, in its unsymmetrized form, the whole terror communications network has an observed density of 1.48 percent (not reported in the table), whereas when transformed the symmetrized network has a density of 2.61 percent.

Generally, density can be thought of as the probability that a tie exists between any pair of nodes chosen at random. Without even looking at a map of the communications network, it is evident from the very low density reported in box 9.4 that these networks are exceptionally sparse. Few ties exist among nodes, although we found that the Al Qaeda subnetwork exhibits greater density than the ISI subnetwork. Recall our theoretical discussions (chapters 3–5). Certain configurations of sparse networks are thought to offer competitive advantages to some well-positioned individuals, and the group as a whole is more adaptable to changing circumstances.

Before we move on to the next statistic, readers should note the denominator appearing in the density equation. Many statistics used to describe networks, and actor positioning within the network, come in two forms: raw values and standardized metrics (referred to as normed values). Dividing by the size of the network, or some mathematical derivation of size, is regularly done to standardize metrics, so that networks can be compared. We will return to the importance of standardization shortly.

The next statistic appearing in box 9.4 is **average degree centrality.** Average degree centrality tells us about typical connectivity. It tells us how many other actors someone is directly connected to on average—in other words, the average number of contacts each person has. Calculating this value involves two steps. First, we need the degree centrality for each node in the network. Degree centrality counts the number of direct connections each node has and is represented as

$$d(n_i) = \sum_{i=1}^{g} x_{ij}$$

where x_{ij} is each tie connected to node *i*. Note that *i* always represents a focal node and *j* indicates all other nodes. This equation, therefore, tells us to examine each node separately and count the number of others (*j*) connected to it. In the second step we average all of the degree centrality scores.

Notably, this value is still in a raw form, meaning that it has not been normalized. In social network analysis (SNA) terms, we normalize metrics so that we can compare values across networks of different sizes. This process is akin to creating rates, like crime rates, so that we can compare relative scores for different communities. Normalizing individual degree centrality scores using the equation that follows permits individuals from different networks to be compared:

$$C'_D(n_i) = \frac{d(n_i)}{g-1}$$

Do you notice something familiar in the denominator? The denominator is designed to represent the maximum number of people to which a node could be connected. Recall that *g* refers to the number of nodes in the network. If people cannot be connected to themselves, then the maximal number of direct contacts is one person less than the number of people in the group, or $g - 1$. Dividing by this measure of network size enables us to compare scores for different networks, or subsets of a network, that contain a different number of actors.

Normally, the distribution of individual-degree centrality scores, whether raw or normalized, is highly skewed and follows a power law function, such that a few people are directly connected to many others and most actors in the network have few direct ties. This means that irrespective of the type of network under investigation, or the nature of the ties examined, most people will have one or two direct connections and a small number of people will have a lot of connections. Because of the skewed nature of degree centrality, we know that standard deviations calculated on these values will be exaggerated. For this reason, standard deviations are not always reported along with average values; instead, authors often report the minimum and maximum values. Since adding two more columns to report the minimum and maximum values would disrupt the format of the table, I simply provided the standard deviation. As would be expected, the observed standard deviations are rather large, suggesting much variability. If you are curious, the minimum value was 1 for each network, and the maximum values were 66 for the whole network and the Al Qaeda subnetwork and 17 for the ISI subnetwork.

Interpreting the average-degree centrality (or extreme individual scores) is tricky because you must consider the size of the network. This means that what constitutes "a lot" of connections depends on the data. If the network includes a hundred people, theoretically someone might be connected to ninety-nine other people. Depending on the nature of the ties, scores of this magnitude might not be likely. For example, a co-offending network for a group of a hundred gang members, where people are considered to be connected if they were arrested together within a designated time period—let's say one year—is unlikely to exhibit such high-degree centrality scores. I would be shocked if one gang member were arrested with ninety-nine others during the course of one year. If instead we were to investigate a gang that included only five people, then someone with a lot of co-offending activity might be linked to four other people in the course of a year. Score interpretation must be sensitive to the size of the network, as well as the nature of the ties. In the next section of this chapter, where we cover measures of individual positioning, we will explore degree centrality in greater depth. In the meantime, because of the skewed nature of this statistic, it is conventional to report the average score for the whole network in order to give the reader a rough estimate of overall connectivity. The measure can be considered a rough approximation because outliers—a few people with high scores—could inflate the average. Generally, therefore, when it is used in an SNA study, the average-degree centrality tells us the threshold under which most actors fall. Again, sometimes you also see the standard deviation, or the minimum and maximum scores.

A comparison of the two subnetworks showed that individuals named in Al Qaeda intelligence typically communicated with about three people, while ISI affiliates communicated with two people. Note that I rounded these unstandardized values to a whole person. The normalized values are a bit harder to interpret. For instance, controlling for network size reveals that individuals involved in the ISI communications network have, on aver-

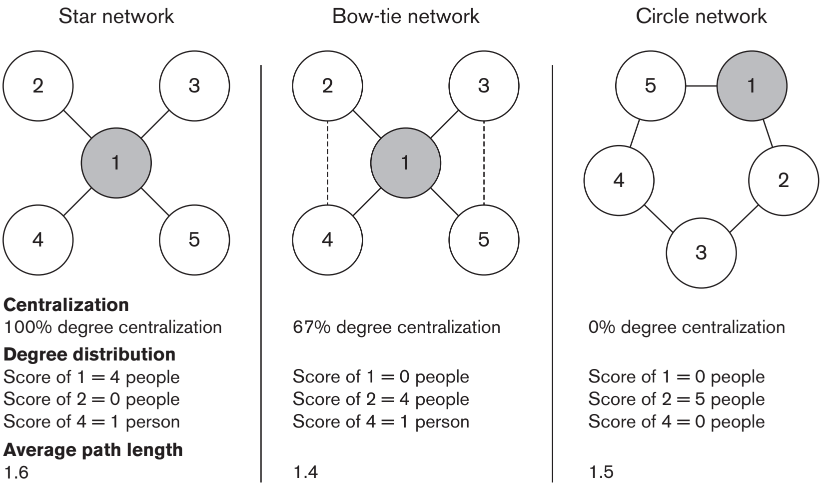


FIGURE 9.1. Hypothetical network shapes.

age, *more* direct connections (normed average degree equals 2.1) than individuals from the Al Qaeda communications network (normed average degree equals 1.6). Much like what happens when raw crime counts are compared to crime rates, taking into consideration the size of the group changes our understanding of the overall pattern of communications.

Degree centralization reveals the proportion of the network directly connected to the node with the highest individual degree centrality score. (I will spare you the calculation.) This information provides some indication of the starlike quality of the overall structure. To illustrate this concept, I created some small hypothetical networks, each containing five nodes but with varying ties. If we eyeball the visuals, it is easy to determine that the first panel of figure 9.1 exhibits a highly centralized network comprising five actors. Take a look at person number 1 depicted in gray. This highly central individual is directly linked to four people, forming the center of the star. I am sure it is not surprising to learn that the degree centralization in this network is 100 percent—the whole network is centered on one person. At the other extreme is the circle network, where everyone has the same number of direct contacts. Here, degree centralization is 0 percent, indicating that no one stands out on this characteristic. Notice the bow tie network situated between the two extreme examples. This network differs from the star formation owing to the two additional ties illustrated by the dashed lines. These relations have the effect of reducing the centrality of the person shown in gray as a result, the degree centralization of the network declines.

To exemplify this concept with a real network, let's consider the terror communications network (see box 9.4). By symmetrizing and dichotomizing the network, we make two assumptions: (1) if a communication occurred between two people, information could flow either way (hence the need to symmetrize the network); and (2) the volume of communication observed is an artifact of the amount of intelligence available (to avoid biasing the results, we dichotomize the network). The transformed Al Qaeda communications subnetwork exhibits a greater starlike structure in the sense that it is more likely that one person is quite central—one person is directly connected to 32.72 percent of the people in the network. Degree centralization is substantially lower in the ISI subnetwork. I find the whole network score to be interesting. Including ISI communication reduces centralization as expected, but given that eight key leaders are present in both networks, the most central Al Qaeda figure may be centrally positioned overall. From this finding, I could hypothesize that a person central to Al Qaeda communications is positioned to be central to the whole terrorist community that was observed with this data. I will return to the issue of calculating and interpreting individual actor scores shortly.

The final statistic to report when describing the connectivity of a whole network is the average path length. **Average path length** refers to average **geodesic** distance—average length of shortest paths linking any two nodes in the network. In essence, the geodesic is calculated for each pair of people in the network. Then all of the geodesics are averaged. Since all permutations are considered, this statistic is hard to calculate by hand for a large network! Two factors will significantly influence the average geodesic distance: (1) the presence of star formations relative to chainlike formations, and (2) cohesion. Star formations shorten the average distance between actors, whereas chains lengthen distances. For example, figure 9.1 reports the average path length for three hypothetical network shapes. Since the networks are small, we see minor differences; however, if the relation between nodes 1 and 5 were to dissolve in the circle network (on the right-hand side of the figure), the network would become a line graph and the average path length would increase to a value of 2. Cohesion will also influence average path length: high levels of cohesion will shorten geodesics. To visualize what I am talking about, compare the circle network of figure 9.1 to the complete network in figure 9.2. Paths are shorter in the complete network because of all of the direct ties (greater cohesion).

Recall that in a small world (described in chapter 3), we generally expect to find an average geodesic of about six steps, meaning that on average, all of the people in the network could send a message to anyone else using five

BOX 9.5. DESCRIPTION OF THE STRUCTURAL DISTANCES

Network	*Average path length*	*Maximum distance*
AL Qaeda network	3.59	8.00
ISI network	4.44	10.00
Combined (all data)	4.54	10.00

intermediaries. Short communication paths might indicate that, under optimal conditions, something could travel quickly from one node to another. Flow is efficient. Returning to the terror communication example, while both subnetworks fall within the threshold of small worlds, the average distance within the Al Qaeda community is shorter, with geodesics averaging 3.6 steps. For ISI affiliates, however, the distance is about 4.4 (see box 9.5). The maximum distance (or longest geodesic) is also larger within ISI, indicating the diameter of the network is wider, with at least one set of people positioned at quite a distance from each other. Considering the average and maximum (longest) geodesics, it is likely that the network exhibits some chainlike structures, particularly within the ISI community.

To recap, in this section of the chapter I talked about statistics that, when reported together, tell us about two aspects of connectivity—cohesiveness of the group and centralization. Using the terror-group communications data, we saw an example of how to interpret the scores. But remember that interpreting these basic descriptive statistics can be tricky. You must interpret values in light of the social behavior under investigation (what nodes and ties represent), the length of the observation period, the limitations of the data, and the qualities of ties (e.g., directed versus symmetric). As we learned previously, the cohesiveness of a network (and, alternatively, the sparseness of its structure) has important implications for network criminology. For example,

- Group cohesiveness might account for why some outreach programs are unable to change illicit drug use—cohesive groups of criminally minded people with similar characteristics will continually reinforce existing group norms, constraining individuals and nullifying the influence of outreach workers. (Think about Charlie's heroin addiction described in chapter 1.)

- Sparse networks or changing levels of cohesion may explain how exposure to crime opportunity varies throughout a person's life. (Several criminological theories reviewed in chapters 4 and 5 would argue that is the case.)
- Criminal networks, such as illicit drug distribution networks, ringing operations, and terror groups that have high centralization and low cohesion may be particularly vulnerable to targeted attacks. (Chapter 6 reviewed some research supporting this hypothesis.)
- Criminal networks form through routine activities. Members of a criminal group are likely to have greater cohesion when they have overlapping activity space, and this makes the group more resilient to targeted attack. (Recall the P.I.V.O.T. project described in chapter 1.)

Clustering

With the basic descriptive statistics about overall cohesion and centralization explained, I now turn to clustering. Networks are bound to exhibit different patterns of clumping, possibly indicating the existence of important subgroups. While I will get into subgroup analysis more formally in the next chapter, when describing a whole network, it is generally important to provide the reader with some sense of how sets of nodes cluster.

Clustering is a key feature of small worlds. Small worlds tend to exhibit clustering and short average path lengths. To test for small-world properties, Borgatti, Everett, and Johnson (2013) suggest that we compare the clustering coefficient from an observed network to what we find in a randomly generated graph of the same size. Then we should compare average path length. Networks with small-world properties will have more clustering but the same, or shorter, average path lengths, compared with random networks.

As with many of the metrics we will discuss in this chapter, multiple forms of the statistic exist. The global version, or overall clustering coefficient, reports on the overall clustering observed in the network, whereas local calculations indicate the embeddedness of specific actors within local neighborhoods. Local neighborhoods include all nodes that are adjacent to the focal actor (meaning all those immediately or directly connected to the person of interest). For example, when calculating the local clustering coefficient for node 1 in the kite network shown in figure 9.2, we would consider everyone except node 5, because this person is not directly adjacent to node 1.

Calculating the clustering coefficient raises the notion of transitivity. It is the first statistic discussed that reflects the connectivity observed among sets of three actors. The relations among a set of three people are said to be transi-

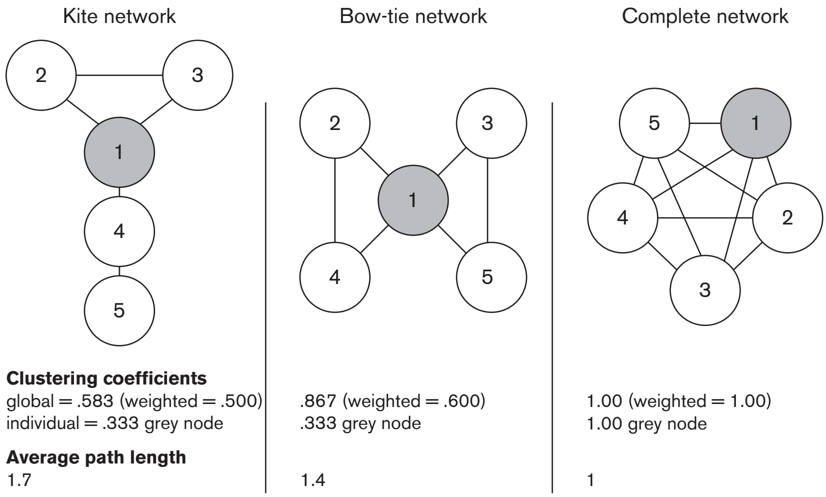


FIGURE 9.2. Global clustering coefficients for hypothetical network shapes.

tive if ties exist linking everyone. For instance, in the kite network you just examined, nodes 1, 2, and 3 are transitive, but nodes 1, 3, and 4 are not. Similarly, nodes 1, 2 and 5 are not transitive. Without getting into too much detail (we will discuss transitivity in greater depth in the next chapter), I will point out that the clustering coefficient is the proportion of triads in the network—sets of three people who are all connected, relative to the number of potential triads (what could exist). In essence, we use this statistic to understand how often threesomes appear in the network—for example, Alex is friends with Bruce, Bruce is friends with Charlie, and Charlie and Alex are friends too. Because we are investigating three-way relations instead of dyads, clustering metrics are generally considered to be higher-order statistics compared to their simpler cousins, which explore dyadic relations.

Formally, we calculate the global-clustering coefficient by averaging the density for each ego network. This calculation involves four steps. (1) Count just the number of ties that a focal individual has with others and then add the ties among people directly connected to the ego. (2) Weight this score. Weighting is important as egonets will vary in size. Weighting is accomplished by dividing the density by the number of pairs of people in the ego network (or, more simply put, just divide by the ego's degree centrality). (3) Repeat steps 1 and 2 for each person in the network. This will produce a weighted density for each person in the network. (4) Finally, average these weighted scores to come up with a weighted mean clustering coefficient. Interestingly, the mean weighted clustering coefficient turns out to be the same as the transitivity index (Borgatti Everett, and Johnson 2013). The transitivity index counts simply the number of observed triangles and divides by the number of times a set of three people has a two-star formation. Since this equation is simpler than the weighted mean clustering coefficient, it appears below.

$$\textit{Transitivity Index} = \frac{\sum i,j,k \quad x_{ij} x_{ik} x_{jk}}{\sum i,j,k \; x_{ij} x_{ik}}$$

The numerator of this equation simply tallies for each person the number of three-way relations they are part of, and the denominator tallies the number of two-way relations each person is part of. The resulting score is a proportion ranging from 0 to 1, with higher scores indicating more clustering and a value of 1 denoting a complete graph, in which everyone is connected to everyone else. This calculation generates a proportion because the numerator is the number of transitive triads and the denominator incorporates all two-star formations (structures where one individual is connected to two other people), including those that are in transitive triads. This means that the denominator contains the special subset of instances that are transitive: all of the transitive triads counted in the numerator are also in the denominator.

In the real world, there is a limit on the number of relations people can maintain. For large networks, therefore, we do not expect to observe scores of 1; instead, scores for clustering coefficients typically range from .3 to .6. Scores of 1 indicate that all the actors in the group are embedded in all possible triads. In other words, a score of 1 for a set of a hundred co-offenders would indicate that they are very active and change partners so frequently that, during the period observed, everyone in the group, ends up co-offending with everyone else. The astute observer will notice that a clustering coefficient of 1 means that the network is complete—no subnetwork variation exists. On the other extreme, a score of 0 also means that there is no subnetwork variation, but in this situation no one in the network is part of a transitive triad. Three-way relations do not exist.

Figure 9.2 illustrates clustering among three hypothetical networks—global clustering and local clustering scores for actor 1 are reported. Notably, star, line, and circle graphs are not shown, as they have clustering coefficients of 0. Depicting these structures would not help to explain clustering. The kite network has the lowest overall clustering of the three forms portrayed—among the five nodes, we observe only one triad. In the complete network located on the right side of the figure, everyone is connected to all others; therefore, by definition, a complete network has a score of 1, indicating that

BOX 9.6. LOCAL CLUSTERING COEFFICIENT OF ACTOR 1

A comparison of the kite and bow tie networks in figure 9.2 shows that actor one exhibits the same level of embeddedness in local clusters, despite the appreciable difference in observed network structure. Remember that a set of three nodes is classed as transitive if a relation exists among all actors—meaning node 1 is connected to 2, 2 is connected to 3, and 3 is connected to 2. This connectivity is observed in the kite network; however, it is the only transitive triad in the network. The other possible sets of three relations within actor 1's local neighborhood are not transitive. Below, the potential triads associated with the local neighborhoods of actor 1 are compared.

Kite network	Bow tie network
1, 2, and 3 are transitive	1, 2, and 4 are transitive
1, 2, and 4 are not transitive	1, 3, and 5 are transitive
1, 3, and 4 are not transitive	1, 2, and 3 are not transitive
	1, 2, and 5 are not transitive
	1, 3, and 4 are not transitive
	1, 5, and 4 are not transitive
One of three are transitive	**Two of six are transitive**

Adding the number of transitive triads and dividing by the sum of potential triads provides a score of local clustering for node 1. Notice how both equations result in a value of .333 (1/3 = 2/6); the level of local clustering observed for node 1 in each network is the same in these two examples.

all possible triads are observed. What is not apparent by examining these global statistics is how nodes clump together differently in subsections in each of the graphs. Finding distinct subgroups or clusters within a network is important, but we will not discuss this until chapter 10. As a preview of what is to come, examine node 1 in the kite and bow tie networks. You will notice that this person is observed to be embedded in about a third of the possible triads in those local neighborhoods, even though the kite and bowtie networks have different global structures. Box 9.6 explains why this is the case.

Ready for something a bit more complicated? Let's take a look at the terror-group communication networks. Examining the whole network, I found a clustering coefficient of .289. This value suggests that some clustering exists. While the value is a bit lower than what we usually observe (recall that clustering coefficients in the real world typically range from .3 to .6), it is not exceptionally different from what we might expect. Turning to the subsectors, the Al Qaeda network exhibits less clustering, as indicated by a value of .277, than the ISI network, which has a global clustering coefficient of .294. I will return to clustering in chapter 10. In the meantime, if you would like more information on this topic, there is an excellent video (Systems Academy 2015c).

By exploring descriptive statistics, we have already discovered several characteristics of the whole network: (1) The main component includes 90.7 percent of the 302 actors named in the intelligence sources, and the density is very low—there are only 541 unique ties, with 1 percent of all possible communication paths observed. (2) There is a moderately high level of degree centralization and short geodesics—about 21 percent of the actors are linked to one individual, and the average path length is 4.5. (3) The clustering coefficient nearly reaches .3, suggesting that some sectors of the network exhibit greater cohesion than other sectors. Taking these findings together we could argue that the network is sparse; there must be some centrally positioned actors (starlike formation) that shortened social distances between people linked in a chainlike fashion; and there must be at least one distinctive cluster with greater cohesion than the rest of the network. Had we compared this network configuration to a random graph, it is likely we would conclude that the observed network has small-world properties.

To recap, in the first half of this chapter, we reviewed a set of descriptive statistics, which when considered together present a general understanding of the global structure of the network. While these values are sufficient for most experienced network investigators, you may feel a bit cheated. People who are curious about networks can feel a little disappointed when a report does not include a visual representation of the whole structure. Before moving on to the next section, go ahead and satisfy your curiosity. Using the analysis we just completed, visualize the graph in your mind and then take a look at the mapped communications network found in Bichler and Bush (2016). I suspect that you will not be surprised! The descriptive statistics have given you a sense of what to expect.

INVESTIGATING ACTOR POSITIONING

Investigations of criminal networks often seek to identify key persons or relations that appear to be vital to the group's illicit behavior. Because it is often hypothesized that centrally positioned individuals are more likely to be influential or more important to the criminal success of the group, it follows that targeting key persons may be an efficient way to disrupt criminal operations. Efforts to target key persons requires using statistics that can provide some indication of the relative centrality of prominent individuals. The notion of centrality is not new to you. Recall that the mean and median are used to find the central value in a distribution of scores, and the mean can be used to understand the relative position of individual scores within the distribution (I am talking about z scores). This background should provide some intuitive understanding of the utility of exploring a distribution of values to discover what is typical and extreme.

In the case of criminal networks, there tends to be a greater interest in identifying the most extreme, rather than typical, values, in part because the distribution of scores rarely exhibits a normal curve. Most centrality metrics exhibit a curve better described by power laws, according to which a small number of people have exceptionally high scores and most people have relatively low scores. This suits networked criminology well. In chapter 8, you saw how derivation from this pattern can be used to discern the quality of data on hand: Benford's law of first significant digits can be used to investigate whether the data are artificially limited in some way (e.g., there are exceptional levels of missing or falsified data). Because there is a tendency to uncover highly skewed distributions, investigation of centrality metrics either reports the distribution of values or the highest values, typically the top five or ten.

Many different centrality metrics exist, each designed to answer a specific question about the relative position of nodes within the network. Matching the appropriate metric to the research question is fairly easy, as information about the purpose of each statistic is readily available. Readers should note that, while most software applications list about twenty metrics, many are calculated differently depending on the nature of the network examined (directed versus nondirected), focus (global versus individual scores), number of modes (one-mode versus two-mode networks), and targets (nodes or edges). Factoring for all of these variations, there are more than fifty centrality statistics to choose from. Decisions should be based on the research question, with consideration of the overall properties of the network under investigation.

Among criminal network studies, two centrality metrics dominate—degree centrality and betweenness centrality. I suspect that these metrics preoccupy network criminologists for four reasons: (1) Several of the more nuanced centrality measures are based on these foundational metrics, and thus they are a good starting point for any inquiry; (2) they are easier to understand and calculate than some of the more elaborate metrics; (3) most

SNA instruction on centrality begins with these metrics; and (4) of greatest importance, these metrics get at the heart of what many of us want to know about criminal networks—who has the most contacts and who functions as an important bridge between others?

Degree Centrality

Your introduction to degree centrality began at the start of this chapter when I reviewed average degree centrality. Recall that this metric counts the number of direct ties a node has, and it is calculated as

$$d(n_i) = \sum_{i=1}^{g} x_{ij}$$

where, g indicates the number of nodes in a network and x^{ij} are the nodes connected to i. Degree centrality is the simplest measure of centrality and is the basis for many other metrics. Individuals with high scores, relative to others in the network, are described as hubs. Hubs are likely to directly influence, and be influenced by, many others. High-scoring positions are also taken to reflect the extent to which a node is directly exposed to something flowing through the network (e.g., gossip). Depending on what the network ties represent, scores could also be indicative of power, leadership, knowledge, or involvement (e.g., high-scoring individuals are the most active). As we learned earlier, degree centrality is highly sensitive to the size of the network; thus, it is important to normalize values to compare different networks. Raw scores can be compared directly only if there are the same number of actors across all networks being examined. Additionally, network configuration must be considered when interpreting degree centrality—that is, directed versus nondirected networks. Following are three different scenarios that illustrate how scores can be interpreted in light of the network configuration.

Scenario 1. Interpreting Degree Centrality for a Dichotomous, Symmetric Network

Figure 9.3, graph A illustrates a hypothetical network with twenty-five actors and thirty-nine unique ties among them, which becomes seventy ties when directed links are symmetrized. Let's suppose that the group is involved in the illicit transnational trade of avocados. (This is not to suggest that the avocado industry is inherently shady!) In this hypothetical network, ties represent co-offending activity—observed events, in which people worked together in furtherance of illicit trade. We can see that this

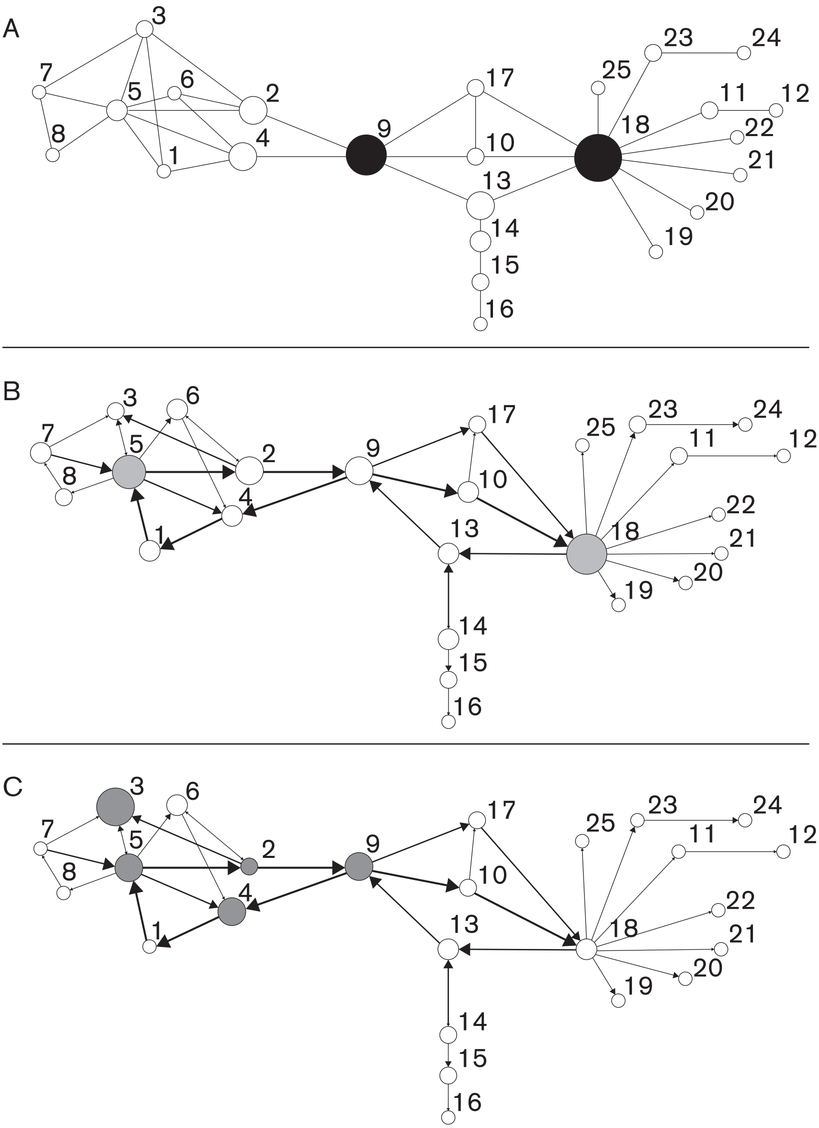


FIGURE 9.3. Degree centrality for a hypothetical group trafficking illicit produce. *A*, dichotomous network, where *g* equals 25 people, *l* equals 39 unique ties (70 when symmetrized), and symbol size varies by degree centrality; *B,* graph depicts out-degree centrality for a valued network; *C*, graph depicts in-degree centrality for a valued network, where *g* equals 25 people, *l* equals 136 total ties (39 unique ties), symbol size varies by degree centrality, line width varies by tie strength as measured by a count of coactivity in furtherance of illicit trade, and arrows originate with the person initiating activity.

group clearly exhibits some classic subgroup structures. A more cohesive group on the left side is connected through a narrow channel to an off-shooting chain and a starburst. The size of each circular symbol varies to indicate degree centrality scores, and the people with the highest scores are shaded gray. It is evident from looking at the visualization that person 18 has a lot of direct contacts—the starburst formation that has this person in the middle is a prominent feature of the graph. What is a bit less apparent at first sight is the influence person 5 potentially exerts. This is why it is important to use centrality measures to identify central actors in a network. Visualization techniques alone are untrustworthy, particularly for large, complex networks. Important features of the network will be hidden in a mosh pit of ties.

The patterns identified regarding degree centrality through this visual representation of the graph are confirmed by the statistics reported in the first column of table 9.1. The smallest nodal symbols represent a score of 1, indicating that the node has a single direct tie. The largest symbol, representing a score of 10, shows that the node is directly connected to ten other nodes. Notably, almost 33 percent of this network is centralized on actor 18. The two gray symbols identify the actors with the two highest scores—actors 18 and 5. As you can see from the values reported in the table, their degree centrality scores differ by three. Since I reported raw scores, and this is a dichotomous, undirected network, we know that actor 18 co-offended with three more people than actor 5 during the period under observation. I used raw scores instead of normalized values to aid understanding.

Scenario 2. Dichotomous, Directed Networks

As with most measures of centrality, degree centrality calculated on a dichotomous, directed network generates two different scores—*out-degree centrality* counts the number of ties extended toward others and *in-degree centrality* counts the number of ties received. Since this is a network of co-offending, you can consider the initiator of the activity to be the person who originates an outgoing tie, while the other party receives (incoming)—that is, one person sends a shipment and the other receives it. The directed version of the network produced the results appearing under the headings "out-degree centrality" and "in-degree centrality" in table 9.1. Take a moment to examine these values.

Table 9.1 reports that there is more variability among members of the group regarding their out-degree centrality—number of outgoing ties—compared to ties received. While the mean number of outgoing and incoming ties is the same (1.6, or less than two co-offending partnership each),

TABLE 9.1. Comparison of raw, undirected and directed centrality scores

	Symmetric, dichotomized network (25 nodes, 70 ties)	*Dichotomized, directed network* (25 nodes, 39 ties)		*Valued, directed network* (25 nodes, 136 ties)	
	Degree centrality	*Out-degree centrality*	*In-degree centrality*	*Out-degree centrality*	*In-degree centrality*
Mean (SD)	2.8 (2.1)	1.6 (1.8)	1.6 (.9)	5.4 (6.7)	5.4 (5.1)
Minimum—maximum	1–10	0–8	1–4	0–22	1–16
Degree centralization	32.6%	28.0%	10.6%	7.2%	4.6%
Highest-scoring nodes (raw score)	Actor 18 (10)	Actor 18 (8)	Actor 3 (4)	Actor 9 (22)	Actor 5 (16)
	Actor 5 (7)	Actor 5 (5)	Actor 2 (3)	Actor 5 (20)	Actor 9 (14)
	Actor 9 (5)	Actor 9 (4)	Actor 4 (3)	Actor 18 (17)	Actor 18 (14)
	Actor 2 (4)		Actor 5 (3)	Actor 2 (13)	Actor 4 (13)
	Actor 3 (4)			Actor 1 (12)	Actor 2 (11)
	Actor 4 (4)				
Interpretation	Number of co-offending partners	Number of people ego led in a co-offense (instigated)	Number of people ego joined to commit a co-offence	Number of co-offenses ego initiated	Number of co-offense activities ego joined

the standard deviation among out-degree scores is double that of the in-degree ones. This variation is also evident in the maximum score differential—the highest out-degree score is 8 (actor 18 initiated co-offending partnerships with eight people), and the highest in-degree score is 4 (actor 3 joined activities initiated by four other people). If we mapped communications instead of co-offending activities, actor 18 could send information directly to eight people, meaning they could influence the greatest number of people. Actor 3, a different kind of hub, receives information from the most people.

Considering degree centralization tells us a bit more about the respective roles of prominent people in the network. The level of degree centralization for outgoing ties is almost three times greater than incoming centralization. Since actor 18 has the highest out-degree score, we know that this person stands to have the greatest direct influence and has initiated the most co-offending activity. We also know that the network structure optimizes outgoing relations over incoming relations, because the inward relations are less concentrated. The greatest level of inward concentration, representing only 10.6 percent of situations in which someone joins in on a co-offending activity, involves actor 3. Keep in mind that these are raw values.

Scenario 3. Valued, Directed Networks

Interpreting degree centrality can be tricky when used with valued networks. Graphs B and C, of figure 9.3 exhibit a valued, directed, illicit avocado-trafficking network. Imaginary intelligence suggests that the cohesive cluster on the left side of the graph appears to originate the flow of goods, and thus might represent the start of the illicit avocado-distribution chain. Since this is a valued network, the line width varies to illustrate how much activity occurs among pairs of actors; arrowheads show the direction of flow. Symbol size also varies to depict degree centrality scores, and the highest-scoring individuals are represented with colored nodes. Graph B illustrates out-degree centrality, and graph C shows in-degree centrality.

Let's examine out-degree centrality first, looking at the final two columns of table 9.1. Notice how the average out-degree score is now 5.4. (Also notice that the variability of scores, as indicated by the standard deviation, also increased.) But if you look at the network size, there are still twenty-five people in the network. So why is the average out-degree centrality score so high? Notice we now have 136 ties. This is the sum of ties, and it includes situations wherein the same pair of people committed multiple co-offending activities. The value 5.4 is not, therefore, telling us the number of co-offending partners that people had on average, instead, this score tells us the average number of co-offenses. This means that, if two people had an out-degree centrality score of 5, one person could have had only one partner for five activities at the same time that another person engaged in three activities with one person and two activities with another. Now out-degree tells us the volume of activity but not the number of partners. Interpreting in-degree centrality also changes in this way. We know only how many times someone joined in an activity led by others, not how many co-offending partners the person had.

Another interesting change is that, by valuing this network, degree centralization dropped. When we factor for the volume of activity, the importance of the most central individual declines. This change makes sense. Look at graph B of figure 9.3. Considering the arrowheads and line thickness connecting colored nodes, we can see fairly prominent patterns of directed activity linking central nodes in a chainlike fashion. The existence of several high-volume links in this distribution chain are diminishing the centralization of the network.

Turning to the prominence of actors, we find that a number of people remain in the top five ranking, albeit they are ranked differently, and we find that a new person (number 1) has been added to the list. Generally, when comparing the prominence of actors across network configurations, I look first for stability—that is, people who appear prominent because they interact with a lot of people (dichotomized version of the network) and because they have a lot of interactions (valued networks). The second thing I look for is whether anyone stands out in only one network configuration. And finally, when I have a directed network, I like to investigate the structural roles of these prominent actors. This involves comparing in-degree and out-degree scores.

As a general rule of thumb, when applied to criminal activity, the differences between a node's in- and out-degree centrality can be thought of as follows (here, I am extrapolating criminological implications from Wasserman and Faust [1994] to explain empirical results of criminal network research):

- **Peripherally located.** When the in-degree centrality and out-degree centrality scores are both valued at 0 or are very low, the actor is somewhat isolated from criminal activity. Either their involvement is not known (missing data), or their involvement in the network has changed (e.g., has ended or just begun), or they have intentionally obscured their role by manipulating the structure around them. Notably, a relatively peripheral person can also mark the boundary of group membership or be positioned at a communications transition—that is, the data map email exchanges and the peripheral actor has a minimal number of email messages, but unbeknown to you, the person is a major transmitter of information through in-person meetings.

- **Transmitter (distributor or source/leader).** Transmitters have null or very low scores on in-degree centrality, but relatively high scores on out-degree centrality. Someone who buys illicit drugs in wholesale quantities, and then sells small quantities to many people at a retail level, would exhibit this pattern of scores—the person is a distributor. However, if the network is valued, a high out-degree score does not necessarily mean the individual has a lot of contacts; rather, the person might send a high volume of messages or materials to a few people, perhaps even one other person—the individual is potentially a source who can introduce materials or initiate actions (leader). Comparing the individual's out-degree score, as computed on a valued network, to their score as calculated on a dichotomous network will help to determine if the person is a distributor or a source.
- **Carrier (transshipment).** Major carriers have relatively high scores for both in-degree and out-degree activity—a large amount of materials, information, or what have you, moves through them. Because they receive and transmit a lot—either high volumes from a few others or small volumes from many others—their exact roles require additional exploration. These nodes are central to maintaining flow through the network, and they are often located toward the core of the network.
- **Receiver (consolidator).** Receivers have high in-degree centrality scores and exceptionally low out-degree scores. They may, in fact, have an out-degree value of 0. Receivers have the opposite function of transmitters—they funnel the network flow by consolidating from many contacts or large volumes from a small number of contacts. If their out-degree score is 0, and they are not likely the terminus of the network flow, the peripheral nature of their position could be an artifact of missing data, a product of intentional social engineering, or a boundary of the group based on the type of relation mapped.

With this information in mind, take another look at the valued graphs depicted in figure 9.3. What is your assessment of the role of person 18? This person has a large out-degree centrality score and a relatively low in-degree score. If you guessed transmitter, you are correct. But is person 18 likely to be a distributor or a leader of this avocado-trade network? I would argue distributor—making orders through actor 13 and receiving from two suppliers (actors 10 and 17); perhaps this is a fail safe in case customs interferes by seizing shipments. Being at the center of a starburst is also suggestive of taking a role in dispersing materials. Had this individual been positioned on the left side of the graph, I might be inclined to be more reserved in my conclusion, as this could indicate greater involvement in originating the distribution chain.

Actors 5 and 3 also appear to occupy influential positions within this network, relative to all others known to be involved in the illicit avocado-trafficking ring. Actor 5 also has a high out-degree score but is positioned differently. Embedded in a cohesive subsection, and exhibiting high out- and in-degree centrality scores, this individual is positioned to be a relatively important carrier. Actors 18 and 5 are both engaged in consolidating and transmitting, but their position is clearly different from actor 9, who sits in the middle of the illicit avocado-supply chain. Actor 9 is a different kind of carrier, which we will talk more about shortly. Actor 3 has an interesting position as well. This individual receives much more than he or she initiates and is situated on the production side of the supply chain, which likely means actor 3 plays an important role in the overall operation of the organization.

Before explaining the second centrality measure of interest to criminal network investigators, I want to share five thoughts on analyzing degree centrality.

1. Individuals with a lot of known contacts are highly visible in the network. While this may indicate that the person occupies an influential role, criminal networks operate within a hostile environment, suggesting caution should be exercised when identifying targets with degree centrality alone. The individuals controlling operations might be partially hidden, meaning they may be directly or indirectly tied to others with high-degree centrality. Use multiple metrics to find pivotal actors. For example, a statistic related to degree centrality that might identify partially hidden but highly central people is **eigenvector centrality** (see table 9.2 for notes on this metric).
2. Flow and architecture are two different facets of social networks. When possible, examine both a valued, directed network (flow) compared to a dichotomized, symmetric version of the group structure (potential architecture). Together, these analyses will provide greater insight into how the group might operate.
3. Networks are dynamic. Comparing the degree centrality scores for specific actors over time can tell us something about how the group

responds to evolving conditions—that is, reorganizing after the loss of a leader, launching a major operation, or responding to a conflict between factions.

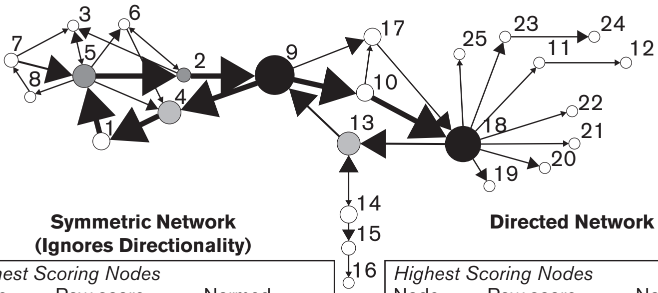
4. Raw metrics are easier to interpret, but when comparing different networks, such as how actor centrality changes over time, normalized scores should be used, as degree centrality is highly sensitive to network size.
5. Theoretically, actors with high centrally scores can be interpreted as hubs. Recall that owing to preferential attachment, the influence of hubs will grow, and they will connect to more people in time. Hubs also tend to link to other hubs. As a consequence, the average path length among all members of the collective will shrink the longer the network exists. Level of centralization and presence of hub-to-hub structures, therefore, might indicate the longevity of the network (a.k.a. length of its criminal career).

Betweenness Centrality

As previously noted, actor 9 is situated in a very unique position, as is actor 18. They sit between many other pairs of people in such a way that they may function as instrumental brokers. Finding people who are positioned between many others is a common exercise in the study of criminal networks. Many definitions and metrics are available to identify actors or ties that join parts of a network, referred to as bridges, brokers, and boundary spanners. A systematic review of twenty-four studies conducted by Long, Cunningham, and Braithwaite (2013) uncovered fifteen unique definitions, and approximately seventeen different analytics to identify key nodes or links. One of the most commonly used metrics was nodal betweenness centrality.

Freeman's (1977) nodal betweenness centrality is generally interpreted to be an indicator of control over the exchange of resources flowing through the network. Individuals positioned between sets of others indirectly link pairs of individuals or subgroups, who would otherwise not be able to interact. Before I go any further in explaining this metric, take a look at figure 9.4. Examine the structural positions occupied by actors 9 and 18, relative to the positions occupied by other people in the hypothetical avocado-trafficking network.

While betweenness centrality is only calculated on a dichotomous network (directed or undirected), I illustrated how scores vary among individuals on a valued graph so that you can better visualize how these actors might control the exchange of resources within the group. Variability in line and arrowhead size reflects the number of activities pairs of people



Symmetric Network (Ignores Directionality)

Highest Scoring Nodes

Node	Raw score	Normed
18	173	63
9	132	48
13	93	34
2	54	20
4	49	18

Whole Network Metrics

Mean (SD)	28.8 (43.6)	10.4 (15.8)
Range	0–173	0–62

Betweenness Centralization index = 54.4%

Size 25 nodes, 70 ties

Directed Network

Highest Scoring Nodes

Node	Raw score	Normed
9	197	36
18	180	33
4	142	26
5	107	19
13	87	15

Whole Network Metrics

Mean (SD)	41.3 (57.0)	7.5 (10.3)
Range	0–197	0–35

Betweenness Centralization index = 29.3%

Size 25 nodes, 39 ties

FIGURE 9.4. Betweenness centrality.

participate in, with arrows originating with the initiator and terminating with the "receiving" accomplice. Activities include a range of tasks in support of the illicit avocado trade, including communications, shipments, money exchange, and so on. While the tie values are depicted in the image, they are not included in the calculations. The colored nodes will be explained later. The insets report key statistics: highest-scoring nodes, network average and standardization, range, and the betweenness centralization index. The whole-network metrics are included to provide context for interpreting the scores of the most central actors. We know from an earlier discussion that the symmetric network has twenty-five people and seventy ties, and the directed network includes twenty-five people and thirty-nine arcs. Two sets of analyses are included, so that we can compare betweenness if the network were treated as symmetric (if a tie exists, either party could initiate activity) or directed (there is an inherent order to who initiates activity).

Actors 18 and 9 are black to highlight their importance as brokers. What I hope you notice immediately is that, even though actor 18 has one of the highest scores, because he or she is positioned in the middle of the largest starburst (I will get to how degree and betweenness centrality can be correlated at the end of this section), actor 9 might be more instrumental to the operation. Situated in the middle of the supply chain, actor 9 has a relatively high betweenness score and, more importantly, directly links others, who have a high brokerage potential but are also somewhat redundant with each other—losing one of them will not disrupt operations. Losing actor 9, however, will bring down the operation, at least for a while. Losing actor 18 after that person takes a delivery may cause a lot of avocados to ripen in a warehouse, but since each stage of the supply chain before that point was uninterrupted and money for product/services was likely to have been received, the bulk of the supply chain remains solvent.

Now that you have a basic theoretical overview of brokerage in this network, let's turn to the metrics. Freeman's (1977) nodal betweenness centrality calculates the potential brokerage function a node may exert by calculating the number of times it sits on the shortest geodesic distance between other pairs of nodes, such that

$$b(n_i) = \sum_{j<k} \frac{g_{jk}(Ni)}{g_{jk}}$$

Where g_{jk} refers to the number of geodesic paths between two nodes, represented as j and k, and $g_{jk}(Ni)$ is the number of geodesics between j and k that includes node i. The score will be 0 if node i does not fall on any geodesic paths (among all potential pairs of actors) and 1, if node i sits along all geodesics. To normalize (standardize), Wasserman and Faust (1994) suggest dividing individual scores by the maximum possible number of geodesics i could sit on.

$$C'_B(N_i) = \frac{C_B(N_i) \times 2}{(g-1)(g-2)}$$

Where g is the number of nodes, and $C_B(N_i)$ is the betweenness score for node i calculated with the prior equation. Of note, this calculation assumes that only one geodesic is possible between each set of nodes.

Akin to degree centrality, an index of group centralization can be calculated to determine the extent to which one node dominates. Centralization scores nearing 0 suggest that all actors have an equivalent potential to moderate relations among others, whereas, a value of 1 (or 100 percent, if multiplied by 100) suggests that one dominant node mediates between all others as it sits on all geodesics.

Betweenness, Nondirected (Symmetric) Relations

Returning to the first set of values reported in figure 9.4, we see that, on average, actors in this network sit along about twenty-nine geodesics (rounding up from 28.8), but the range is considerable, 0 to 173. If we ignore directionality, among the highest-scoring individuals, there are natural breaks. Actors 18 and 9 form one set, actor 13 is another, and a third set is made up of actors 2 and 4. Comparatively, the first set containing actors 18 and 9 is roughly three times as central as the latter set (actors 2 and 4). The underlying organizational structure is highly centralized around one central actor; actor 18 sits on 54 percent of all geodesics.

Betweenness, Directed Network

Taking the directionality of activity into consideration, we find shuffling among the most central actors. Actor 9 now stands out as the most central, albeit this individual's control potential is closer to the second-ranked broker, while four of the five previously named actors remain among the top five highest scores. Two individuals swapped positions (dark gray nodes). Actor 2 is an important broker if we consider simply the underlying structure of the network, but when we factor for directionality, that individual is replaced by actor 5. These five actors rise to the top because they sit between exchanges going in both directions (geodesics toward retail operations at the far right side of the graph and toward the supply side on the left side of the graph). Controlling flow in one direction would diminish each one's relative standing in the group (e.g., actor 2). Another notable difference is in the betweenness centralization index. Previously, we saw that the most central broker controlled exchanges along about 54 percent of the geodesics. When directionality is considered, centralization drops and the most critical broker (now actor 9) has a controlling position for about 29 percent of geodesics.

Again, readers should see that comparing analyses produced under different conditions tells us more about actor positioning than a single investigation. Most research finds that, when identifying targets on which to focus efforts to disrupt criminal and deviant activity, it is essential to compare the results of different analyses—in particular the results of different centrality metrics. Each metric defines centrality in a unique way, and each metric has its own limitations. Recent investigations go further, suggesting that target prioritization should integrate multiple centrality measures with other information, such as offender characteristics or role within the organization (e.g., Bright et al. 2014; Schwartz and Rousselle 2009). A particularly good example of a study using centrality scores in conjunction with offender attributes is Hashimi and Bouchard (2017). In this study, the authors introduce readers to a new metric—called network capital—which integrates positional and actor attributes to improve the rigor of target prioritization.

While it is not feasible to go through more centrality metrics in depth, table 9.2 summarizes considerations for using degree and betweenness

TABLE 9.2. Summary of centrality statistics

Concept	*Statistic*	*Finds*	*Interpretation*	*Issues*	*Other metrics to compare*
Hubness	Degree centrality Out-degree centrality	Number of direct connections Node influence	*Dichotomous, symmetric network*: HS = being connected to more people; *Valued network:* HS = more ties (possibly to the same person) *Dichotomous network:* HS = emits, extends, or sends more nodes *Valued network:* HS = emits, extends, or sends high volume of transmissions	Sensitive to network size	*Eigenvector centrality:* indicates association with popular people—counts adjacent nodes but weights them by their degree centrality; sensitive to actors with low degree scores that are connected to others with high degree; uses symmetric data only and degree scores must vary; valued or binary *Beta centrality:* identifies channels of indirect influence, where indirect channels are weighted (inversely) by their length; uses walks rather than geodesic paths; should select medium length for weighting lengths (weights that are too low produce equivalent to degree centrality, and weights that are too high generate eigenvectors)
	In-degree centrality	Node prominence or popularity	*Dichotomous network:* HS = receives from or is chosen by more nodes *Valued network:* HS = receives more or is more highly valued by others		

Brokering / bridges	Betweenness centrality Directed betweenness centrality	Nodes positioned to control indirect exchanges among others	*Symmetric network:* HS = node positioned to control indirect exchanges among others along the shortest geodesics *Asymmetric network:* HS = node positioned to control flow in either direction or is particularly dominant in one direction	Default calculations on dichotomous (binary) networks; can be calculated on nodes (*node betweenness*) or relations (*edge betweenness*)	*Closeness/farness centrality (efficiency):* high-scoring actors sit at the origin (closeness) or terminus (farness) of the shortest paths to all others; measure of overall social distance; all nodes must be reachable (unlikely in a directed network); must be connected network (isolates and small components ignored); large networks have low scores

NOTE: HS = high score.

centrality and offers some alternative metrics you can use to compare your results. It is a good idea to review the table before embarking on the final section of this chapter. An excellent reference book that offers more information about these metrics is Wasserman and Faust (1994). Alternatively, there are many YouTube videos. I particularly like the videos produced by Systems Innovation (for more information, see https://systemsinnovation.io/). While the focus of the videos pertain to all types of networks, it is relativity easy to see how the concepts can be applied to criminal networks. For useful videos on decentralized networks, network centrality, network degree distribution, and network theory topology, see the Systems Academy (2015a, 2015b, 2015d, 2015e).

EGOCENTRIC NETWORKS

An egocentric network can be thought of as a network extracted from a larger graph representing a group or community. **Egonet** extractions focus the inquiry on a specific individual. The purpose is to investigate a node's *local neighborhood*—that is, the set of other actors that is most likely to influence behavior or to be influenced by the ego. More specifically, an egonet includes the focal actor (the ego), the ego's direct contacts (alters), and the relations among the ego's contacts (connections among alters). An egonet extending two steps will also include actors that alters connect to but have no relations with the ego. (If testing contagion hypotheses, you might want to generate egonets extending four steps from the ego.) To explain some of the essential descriptive statistics and common measures of centrality, we are leaving the fictitious world of avocado trafficking for the real world and returning to communications involving Al Qaeda and ISI command staff.

Combining all communications, the whole terror communications network includes 302 people. But only a small portion of these individuals is made up of active operatives occupying positions within the executive command of the terror groups. Even fewer have critical influence over operations. Seeking to better understand the local structure surrounding key leaders and support staff positioned to exert a controlling influence on operations, we identified the most central individuals (Bichler and Bush 2016). Four measures of centrality were used—degree and betweenness centrality, which you know about, and two other metrics, Bonacich's power and eigenvector centrality. Actors who consistently scored among the highest on multiple measures of centrality were selected from each network (Al Qaeda and ISI were treated separately). Removing duplicates, resulted in 41 centrally positioned people—15 named in Al Qaeda intelligence, 18 named in ISI intelligence, and the 8 people named in both sets of source documents. An egonet for each person was extracted, and this sample of egonets is used in the analysis reported below. To provide a point of comparison, summary statistics for all egonets (302 people) are included in the table, so that readers can see how the local neighborhoods of the subset of influential actors differ from the community as a whole.

Descriptive Statistics

To understand how one person's egonet differs from all others, we need to know what the typical egonet structure looks like. If you run the basic descriptive statistics for egonets using UCINET software, the default settings generate sixteen different descriptive metrics. (I will talk more about this software in chapter 10.) Some of these metrics are calculated using the nodes directly connected to the ego (e.g., size, number of pairs), and other metrics include nodes at two steps from the ego (e.g., two-step reach, which is the number of nodes within two links of the ego). Since our time is limited, I will report on only a few.

The values reported in table 9.3 were calculated on a dichotomized network. (Egonets can be undirected or they can capture directionality—*in-neighborhoods* include actors the ego receives ties from, and *out-neighborhoods* include actors the ego extends ties to). When describing egonets, start with size, degree centrality, and number of ties. **Size** indicates the number of alters present in the egonet (excluding the ego). At this point, I will digress to explain a bit of network magic. When you are working with dichotomized, symmetric networks, the number of alters will equal the ego's degree centrality. (Go ahead, reread the prior sentence. It will click if you give it a minute.) This is so because each alter contributes one tie to the ego's degree centrality score. Since the values are the same, it is customary to report one of these statistics. In this case, I opted to drop degree centrality. Only size appears in the table. The number of **ties** reported indicates the number of relations observed among alters: ties to the focal ego are not included, which explains why the minimum number of ties in at least one egonet is zero.

Notice that, on average, the subset of influential actors has larger ego networks compared to all actors in the network. The command staff of terror groups was observed to have about nine alters, whereas people involved in this set of communications more generally had on average three alters. More dramatic is the difference in the number of ties—10 compared to 1.5. Alters of command staff exhibited more communications than were generally observed for the network. The next descriptive statistic to report is egonet density.

TABLE 9.3. Description of egonet structures

Measures	*Average*	*SD*	*Minimum*	*Maximum*
Subset of 41 most influential actors				
Descriptive statistics				
Size	9.3	10.2	2	66
Ties	10.0	14.1	0	76
Density (%)	20.6	25.6	0	100
Ego centrality				
Betweenness	80.5	317.1	0	2,041
Effect size (whole network)	8.35	0.2	0.3	1
All Egonets				
Descriptive statistics				
Size	3.0	5.0	1	66
Ties	1.5	4.4	0	44
Density (%)	21.4	29.7	0	100
Ego centrality				
Betweenness	14.0	143.4	0	2,455
Effect size (whole network)	2.7	4.5	1	61

Egonet density reports the cohesiveness of the ego's neighborhood by dividing the number of ties in the network by the pairs of alters (alters connected to each other). The values reported are multiplied by a hundred to convert the proportion to a percent. A high density, a value close to 1 (or 100 percent), indicates that all alters are connected to each other—ego is embedded in a complete subgraph. Lower values, such as the average density of 20.6 percent reported in the table, indicate that some connectedness is observed among alters. Most egos were not ensconced in highly cohesive neighborhoods. Moreover, the set of forty-one influential actors was not embedded in neighborhoods that were more cohesive than those of members of the observed terror community more generally. Now that we have a sense of how the focal set compares to the community, we turn our attention to exploring centrality.

Measures of Central Positioning

Common sense suggests that the ego would be the most central individual in the individual's own local neighborhood and should have maximum influence and control on surrounding persons. For once, common sense

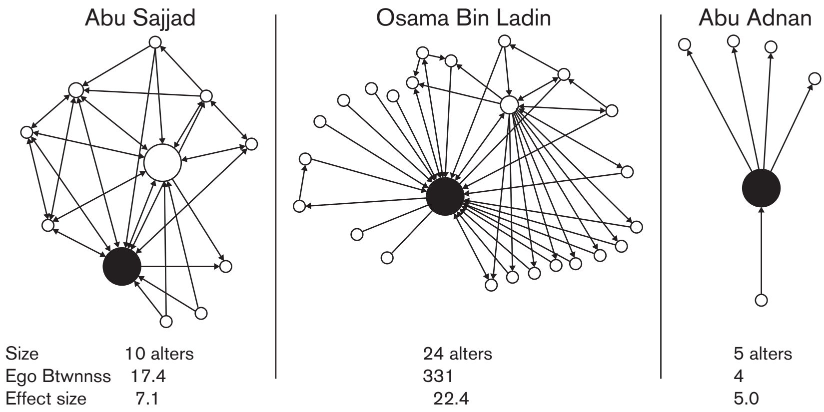


FIGURE 9.5. Comparison of three egonets.

should be ignored. Egos can be completely redundant and virtually powerless in their own social neighborhoods. To explain how this occurs, I will explore three metrics. But first, take a good look at Figure 9.5.

The ego of each network is shown as a black symbol, and alters are white. Arrowheads indicate the direction of communications, such that arrows originate with the initiator and terminate with the receiver. Double-headed arrows indicate reciprocal communication channels. Notice how different the egonets look. First, the networks exhibit fairly dramatic variation in size. Second, each ego seems to be entangled in webs of varying cohesion. So how do the egos' relative positions within their respective networks vary? To understand relative position, we turn to betweenness centrality and effect size.

Betweenness centrality is another metric that you are now familiar with. As used here, betweenness tells us the degree to which an ego brokers exchanges among the individual's alters. Someone with a low score is not in a position to control interactions within his or her own local world, and someone with a high score is positioned in the center of a starlike formation. Figure 9.5 illustrates what this might look like—symbol size varies by betweenness score. As you can see, Abu Sajjad has the same betweenness score as someone else in his own ego network (Shaykh Abu Hind for those of you who are curious). Both actors have a score of 17.4. If Abu Sajjad were removed from the network, his alters would be able to continue communicating—he is redundant in his own circle of contacts. Osama bin Laden is in a better position in the sense that he is more central; however, the node located above him with many of the same ties (Atiyya Abdul Rahman) has a score of 201, and the next highest score is 20. Structurally, this communication network suggests that Atiyya Abdul Rahman, who was acting as a secretary at the time, was in a very advantageous position within Osama bin Laden's local neighborhood—he was a key broker. Much of the network would continue to operate without Osama bin Laden. The third person profiled, Abu Adnan, exhibits the lowest raw betweenness score, since his egonet is small, but he appears to be positioned in such a way that he could exert the most control over exchanges between alters. But does this translate to greater social capital?

Burt (1992) argued that, in order to have a lot of social capital through which to broker exchanges between alters and seize opportunities discovered through contacts, egos must be positioned in such a way that their alters link them to unique sets of people. Alters who are connected to each other do not provide unique social advantages. (See chapter 3 for a richer explanation of social capital.) Not only did Burt provide us with an interesting theory, but he also suggested how we could operationalize social capital constructs, offering four different metrics. The metric we will review is effect size because it considers degree centrality—a statistic you already know about.

Effect size is used to assess the nonredundant nature of ego's contacts. The information obtained through a contact is likely to be redundant if the person you communicated with is also tied to others in your local social neighborhood—that is, the information has little unique value, as alters already know. Effect size is calculated by summing the number of alters, minus their average degree centrality scores: ties to the ego are not included. In a valued network, the calculation would be

$$Effect\ Size_i = \sum_j \left[1 - \sum_q p_{iq} m_{jq}\right], q \neq i,j$$

where, p is ego's contacts linking to q (q is every other actor in the egonet) divided by the sum of ego's contacts; m is the strength of j's connection with q divided by the strongest other association. Higher scores suggest that the direct contacts ego has are less redundant. As a result, the ego is likely to be able to exert more unique influence or control over the flow of resources in the individual's personal network. In other words, higher effect sizes translate to more structural holes. It is notable that if the egonet is small, the score can be depressed.

Returning to figure 9.5, and the values reported below each egonet, we find that factoring for the redundancy introduced in the first two networks by the presence of an alter with a lot of ties, Osama bin Laden had a higher effect size than Abu Sajjad. Visually, Abu Adnan appears to have a higher proportion of nonredundant ties, but if the alters are peripheral to the network and themselves have low degree centrality, Abu Adnan is unlikely to have access to much unique information, which causes his social capital to suffer, as we can see in this example.

Comparing these actor scores to the average effect size found among the group of forty-one influential operatives, as well as the communications network as a whole (see table 9.3), we discover that these three actors have more social capital than the average person in the whole network. Further, among the most influential operatives, Osama bin Laden has about 2.7 times more social capital than the others (I divided 22.4 by 8.35).

In sum, the scientific method stipulates that basic descriptive statistics are required when reporting results; SNA is not exempt from this foundational tenet. I had two aims in writing this particularly long chapter: (1) to introduce readers to some basic descriptive statistics for whole networks and egonets, and (2) to introduce the multifaceted concept of centrality. In the next chapter, we will swim a bit deeper into the pool of statistics by learning how to identify subgroups and test hypotheses. Be sure to enjoy life a bit before moving on to the next chapter. SNA is important, but life is short. Go spend some time with your loved ones. Chapter 10 will be there when you get back.

10. Advanced Analytic Options

> The law enforcement community, being largely unaware of the methods and concepts developed within the discipline of network analysis, has not yet had the opportunity to enunciate its needs for more sophisticated tools.
>
> SPARROW, "The Application of Network Analysis to Criminal Intelligence"

Describing whole networks and investigating the relative position of target actors, generates insight into how the organizational structure of social groups might influence behavior. While these first steps are critical, they are not enough to fully investigate a network's local and global properties. As we learned in the first couple of chapters, networks are not random constellations of relations among actors. Instead, networks exhibit social tendencies suggesting they evolve—even if they start with a random distribution of relations among nodes, over time clumpiness or local clustering will emerge. Looking closely at local properties within subsections of a network helps us to better understand global properties, as the network is a byproduct of substructures.

When studying networks, we often develop hypotheses about how network structures shape behavior, how networks evolve, and how nonrandom distributions of personal attributes and positioning, resources, and subgroup characteristics interact with social structure to shape behavior. To conduct these more advanced, and often multivariate, investigations, we need another set of analytics to tease out how each factor accounts for observed patterns of behavior.

The suggestion posed by Sparrow (1991) in the quotation that opens this chapter is that the true value of social network analysis (SNA) for criminological inquiry has yet to be fully realized. As we learned in chapter 6, network criminology is still an emerging field, despite the value of incorporating SNA theory into our efforts to understand crime (chapter 5) and criminality (chapter 4). This brings me to the three objectives driving chapter 10. First, I want to spend a bit more time reviewing the concept of transitivity to make certain you are comfortable with triads. In furtherance of this objective, the first section delves deeper into transitivity by explaining how a triad census is used to uncover local structures in gang violence. Second, building on this base, I aim to introduce you to some of the analytics used to explore local clustering within networks. After briefly reviewing top-down and bottom-up approaches, I showcase a couple of methods from each set of techniques with data from a study that mapped drug trafficking. Finally, the remaining section of Chapter 10 provides a short description of advanced analytic options for testing hypotheses about how network structure influences behavior. My aim in this section is modest because this is an introductory primer. I hope that by linking advanced analytic strategies with suggested software, you will have the resources necessary to guide your exploration of multivariate models.

TRANSITIVITY

In preparation for the final section of this chapter, let's start with a quick chat about triads and transitivity. Remember that a triad is a set of three nodes and all ties or arcs (if directed) among them. When examining triads, it is important to keep in mind that these configurations are embedded within a constellation of relations, so many combinations of three are possible. The number of permutations, or possible sets of three, is a function of two factors—the number of nodes in the graph being examined and whether the graph captures the directionality of relations. It is this latter factor that shapes the discussion that follows.

Nondirected Graphs

Returning to the story of Charlie and his quest to become drug free that was introduced in the first chapter, it is possible to examine his network for all potential sets of three people and the relations observed among them. Remember that Charlie, a Seattle resident, has a heroin addiction that he is trying to kick by enrolling in the LEAD program. The network of his relations is nondirected (if a relation exists, both people are party to it), and it is not valued (either a relationship exists or it does not). Figure 10.1, under "tracing triads sets," depicts Charlie (gray face) and his four friends—Smith, Jason, Robert, and Valentina. (You may recognize some people from the entity resolution example in chapter 7.) Recall that a smiling happy face represents those of Charlie's friends who support his efforts, unhappy faces represent individuals who are disappointed with Charlie's decision, and a face with no smile or frown depicts an associate who is ambivalent about LEAD.

Notice that the figure does not include the observed relations. Instead, it depicts possible sets of three people. The exercise of tracing triads requires

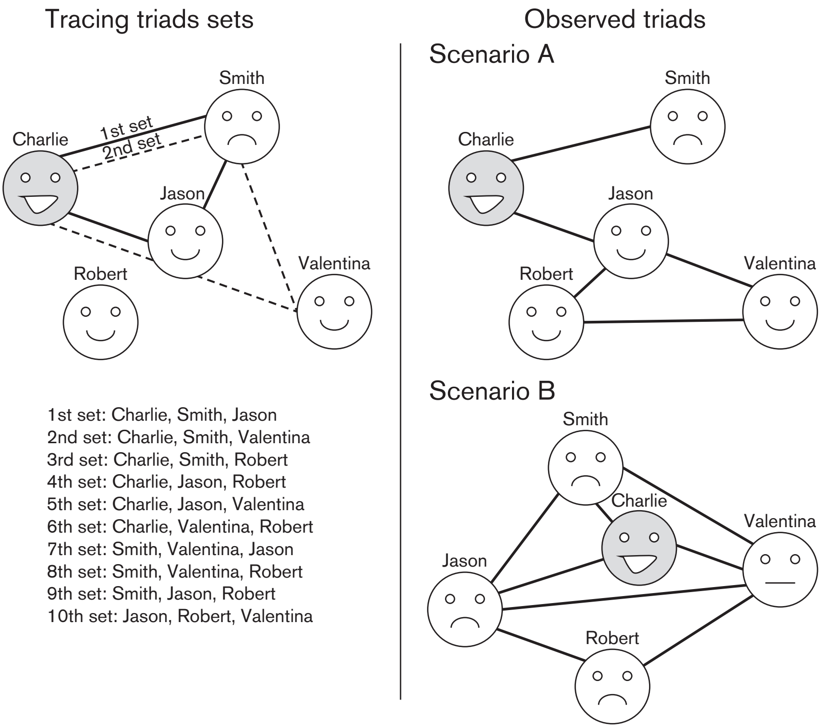


FIGURE 10.1. Transitivity.

that we temporarily ignore the observed ties. Map what is possible first. Then look to see which relations exist among members of each set. To begin, notice that the figure illustrates two possible triads—the first considers Charlie, Smith, and Jason, and the second looks at Charlie, Smith, and Valentina. All possible combinations of three are listed below the graphic. Since the number of permutations is ultimately a function of two factors—the number of actors in the graph and whether the network is directed or not—the maximum number of combinations can be staggering. In this example, a nondirected graph of five people, we can trace ten *possible* combinations of three actors. Note the emphasis on the word possible.

Now that we have traced all of the possible triad sets, we need to take a census of the patterns that are actually observed. Four types of triadic relations are possible among nodes of a nondirected graph—**empty set** (no one is connected), **one edge, two-star** (two edges), and **triangle** (all three are connected). These structures are arranged from least (empty set) to most cohesive (triangle). If graphs have a higher proportion of triangles, the structure of relations has greater connectivity, and actors are embedded within a more cohesive group that is likely to impose greater constraint on their behavior. Graphs with more empty sets are less dense and are likely to impose less constraint on group members. Recall that this section of the chapter will lead us to a conversation about identifying subgroups. Some of you may already be thinking that finding regions in a network that have greater levels of triangles might be useful. Regions with greater connectivity might indicate the existence of a subgroup of the larger network that has a different local structure. For now, just park this thought at the back of your mind.

To conduct a census, simply count the number of observed structures. On the right side of figure 10.1, you will see illustrations of the two possible social scenarios for Charlie that we discussed earlier. Recall from chapter 1 that in one scenario Charlie was connected to two friends, one who supported his decision to try to get clean and the other who did not. Scenario A depicts what appears to be a relatively sparse network, compared to the graph of social relations illustrated as scenario B. The structure of social ties observed in scenario B suggests that Charlie is ensnared within a local network that will make it hard to be successful in the LEAD program—Charlie's friends do not support his idea to stop using heroin. But how much difference is there? Is Charlie really at a disadvantage if his social world looks like scenario B? To answer these questions we can perform a triad census for each scenario and compare the results.

Counting the number of triadic structures appearing in each scenario shows a clear difference between the two situations. Let's start with scenario A. Examine the observed ties for the first set, which includes Charlie, Smith, and Jason. In scenario A, ties exist between Charlie and Smith and between Charlie and Jason. There is no link between Smith and Jason. This configuration is a two-star structure, as two edges are observed. Now examine the second set. Among Charlie, Smith, and Valentina, we find only one edge—a relation exists between Charlie and Smith, but neither Charlie nor Smith knows Valentina. This systematic examination continues for all possible combinations of three people. Box 10.1 reports the results of a census of observed structures. In total, scenario A includes zero empty sets, six one-edge relations, three two-stars, and only one triangle.

Now take a look at the observed structures reported for scenario B. What do you notice? Scenario B includes more triangle formations. In fact, box 10.1 reports that compared to scenario A there are 4 times as many triangles and

BOX 10.1. TRIAD CENSUS FOR TWO GRAPHS

Triad Sets	**Scenario A**	**Scenario B**
First set: Charlie, Smith, Jason	Two-star	Triangle
Second set: Charlie, Smith, Valentina	One edge	Triangle
Third set: Charlie, Smith, Robert	One edge	One edge
Fourth set: Charlie, Jason, Robert	Two-star	Two-star
Fifth set: Charlie, Jason, Valentina	Two-star	Two-star
Sixth set: Charlie, Valentina, Robert	One edge	Two-star
Seventh set: Smith, Valentina, Jason	One edge	Triangle
Eighth set: Smith, Valentina, Robert	One edge	Two-star
Ninth set: Smith, Jason, Robert	One edge	Two-star
Tenth set: Jason, Robert, Valentina	Triangle	Triangle

Census A Summary	**Census B Summary**
Zero empty sets	Zero empty sets
Six one-edge relations	One one-edge relation
Three two-stars	Five two-stars
One triangle	Four triangles

1.7 times as many two-stars in scenario B. These counts are useful for purposes other than simply comparing sums. For example, we could use them to calculate the **global transitivity index** (TI), using the equation from chapter 9. Doing so, we find that the TI for scenario A is 25 percent (TI = 1/[1+3]; then multiply by 100 to generate a percent), and the TI for scenario B is substantially larger at 55.6 percent (TI = 4/[4+5]; then multiply by 100 to generate a percent). Scenario B has just over 2 times as much transitivity. The transitivity index ranges from 0 to 1 (or 0 to 100 if it is converted to a percent), but it is normally somewhere between .3 and .6. Comparing our findings to this convention, we can say that scenario A has fairly low transitivity, and scenario B has relatively high transitivity. Consequently, scenario B is likely to constrain Charlie more. With the majority of this network disapproving of LEAD, Charlie would face greater difficulty in achieving his objective. He would be much better off were his friendship network to look like scenario A.

Directed Graphs

Conducting a triad census on a directed graph is slightly more complicated because there are many more possible configurations of actors. Here we take into consideration the direction of the tie, so that A → B, B → C, and C → A is not the same as A → B, B → C, and C ← A. Notice that the direction of the relation between C and A is different. In the second set, the arrowhead originates with actor A (ego) and terminates with actor C (alter). The directionality of this relation is opposite that of the first set (where actor C extends a relation to actor A, or C → A). This minor difference generates a different pattern of relations among the threesome. In this example, each pattern of ties generates a triangle, but when we consider the directionality of relations, the triangles are not equivalent.

In total, it is possible to observe sixteen different triadic structures in a directed graph. The sixteen structures are organized using a system developed by Holland and Leinhardt (1970). Referred to as the **MAN labeling scheme,** each configuration has a three-digit code that is based on the types of ties observed: the first digit counts the number of mutual ties (reciprocal), the second digit counts the asymmetric ties (one direction), and the third digit counts the number of missing ties. To see how this coding scheme works, compare the structure of 003 to 300, shown, respectively, at the far left and far right of figure 10.2. The simplest structure, 003, is an empty set with no ties (recall that in a directed network we often refer to ties as arcs); consequently, there are zero mutual arcs, zero asymmetric arcs, and three null or missing arcs. At the other extreme, 300 has three mutual arcs, as you can see from the three double-headed arrows, and no asymmetric or missing arcs.

Take a good look at the other structures in the figure before continuing. You will notice two things. At first glance, you may find that organizing structural patterns by arc count is a bit odd. For instance, the first pattern visible in the column titled "2 arcs" appears to have only one tie. Note that this one tie goes in both directions; hence there really is one arc aimed at the top node and another arc aimed at one of the bottom nodes. By separating this tie by its component parts that indicate directionality, we discover that there are really two arcs. Now that I have explained how arcs are counted, you will notice that all of the columns beyond the one labeled "3 arcs" contain structures that have at least one double-headed arrow.

The second thing you may have noticed is that some of the codes for structures illustrated in figure 10.2 contain a fourth digit. The fourth digit is present only when asymmetric ties occur. (Remember that asymmetric ties go in only one direction.) The coding system adds letters to describe

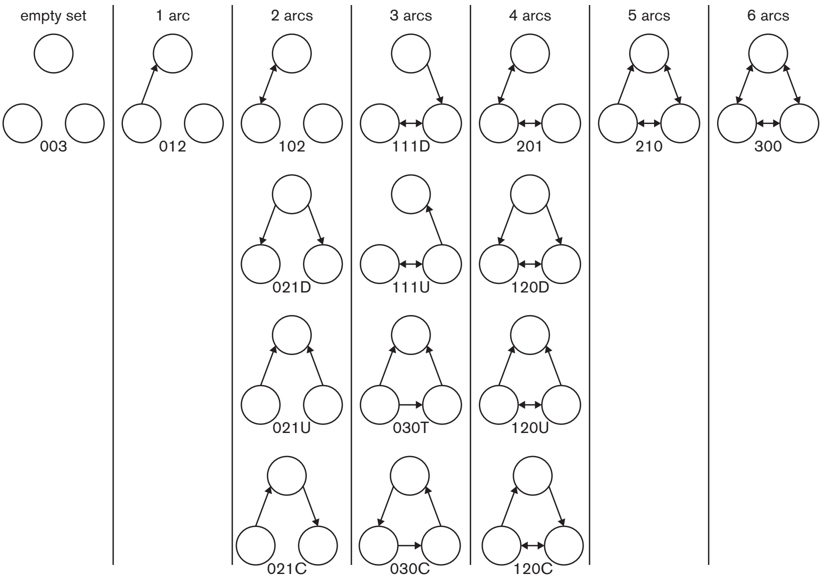


FIGURE 10.2. Triadic structures found in directed graphs.

asymmetry: "D" for down; "U" for up, "T" for transitive, and "C" for cyclic. Take a moment to compare the structure of 111D to 120D. At the top of the column of structures with three arcs, we find 111D. This triadic structure includes one mutual relation (reciprocal arcs), one asymmetric arc, and one null or missing arc. The notation "D" indicates that the asymmetric relation is headed down. In the column of structures with four arcs, you will see 120D in the second row. A visual inspection reveals only one difference—the presence of one additional arc. The coding reveals that, as before, there is only one mutual relation, but now we find two asymmetric arcs and no null or empty paths. Both asymmetric ties point downward. (You might be wondering if "up" and "down" are arbitrary designations within a network. The answer is no. If you were actually looking at the underlying matrix that contains all of the relational data, you would notice the order. Since we are looking at a visual depiction rather than the matrix, it just seems arbitrary. This is all you need to know for our purposes, but if you are interested in learning more, see Wasserman and Faust (1994, chapter 14).

The process of tabulating the number of observed triadic structures is called a **triad census.** While the census itself tells us a lot about the structure of a network, repeating a census and comparing the results are particularly informative, as the analysis might reveal the development of transitive and cyclic structures. Transitive and cyclic tie formations tend to interest criminologists, particularly when they are investigating how criminal networks evolve. For example, these structures can be used to identify how indirect relations form when a direct path between two actors is blocked. Let's briefly consider border controls and smuggling operations.

Improved border controls between two nations may cause smuggling activity to divert through a third party—perhaps another country or a free port (a network with a lot of 012 structures might change to one with many 111U structures). Looking at border control from another perspective, border controls that are weakened as a result of political instability may cause previously indirect smuggling operations to become more direct (a prevalence of 021C structures may shift to 120C structures). Below, I demonstrate how repeating a census can help to track the evolution of gang violence after the imposition of a civil gang injunction.

Gang Violence Example

Interested in how gang violence spreads through communities like a social contagion, Jasmine Randle and I studied the network of violence involving 158 Los Angeles street gangs over a nine-year period, from January 1, 2002, to December 31, 2010 (Randle and Bichler 2017). Counting only gang-on-gang attacks, we found 205 victimizations. Notably, this value does not include 421 victim-offender dyads, where the victims had no known gang affiliation.

Focusing on gangs affiliated with the Bloods and Crips consortiums, we sought to investigate whether serious violence (murder, attempted murder, and assault with a deadly weapon) among these historic rivals played out as interconsortium (group rivalries) or intragroup conflict (struggles for leadership or displays of local dominance). More to the point for our discussion, we were interested to see if there was evidence of Papachristos's (2009) hypothesis that gang-on-gang violence exhibits a pecking order. According to this, one group attacks another, who is unable to retaliate, and then the victimized group becomes the aggressor and attacks another, less dominant group. You will recognize this as structure 021C—a directed line or chain that suggests some sort of hierarchical structure. We were also interested in other hierarchical structures, such as situations involving two gangs attacking the same gang (structure 021U capturing multiple victimizations) or one gang attacking multiple gangs (structure 021D capturing outward attacks). These are two-star structures. You may be interested to know that we found that, out of 205 gang-on-gang victimizations, 42 percent involved a directed line (021C), 27 percent were outward attacks (021D), and 17 percent occurred in situations in which one gang was victimized by two others

(021U). In total, 86 percent of the triadic structures observed did not involve reciprocation.

Investigating intergang violence also revealed some other notable findings. About 38 percent of these attacks reflected the Bloods and Crips rivalry. Viewing this result another way, most gang-on-gang violence (62 percent) occurred within the consortium. Crips gangs tended to attack other Crips, and Blood gangs tended to attack other Bloods. Moreover, almost all gangs attacked nongang-affiliated members of the community.

Building on this study, we have joined with two other researchers to investigate whether patterns of violence changed with the imposition of a civil gang injunction (CGI). CGIs impose behavioral restrictions on gang-involved individuals in an effort to reduce the types of social interactions that often give rise to conflict—for example, hanging out in public areas, engaging in open-air drug sales, and carrying weapons in the neighborhood. We aim to investigate how gang violence, involving seventy-two gangs with CGIs, evolved over a twenty-year period. This research, which focuses on Los Angeles gangs, uses a two-step, snowball-sampling process that starts with the seventy-two enjoined gangs (seeds). This study may sound familiar, as I described it briefly in chapter 7. While as of this writing the research is still in progress, with funding from the Office of Juvenile Justice and Delinquency Prevention, we have some preliminary findings from a pilot study of twenty-three enjoined Bloods and Crips gangs that are pertinent to our discussion of the utility of a triad census (Bichler, Norris, Dmello, and Randle 2019).

Comparing the centrality of gangs and changing structure in attack behavior, the pilot study tested whether our research protocol would capture the effects of CGIs. We used four inclusion criteria to draw a sample of cases. To be included in the study, the case must have involved at least one conviction for a violent crime (robbery, assault with a deadly weapon, attempted murder, or murder); a defendant tried as an adult; at least one defendant or victim who was known to be a member of a seed gang based in the City of Los Angeles at the time of the incident; and a crime that occurred between January 1, 1997, and December 31, 2015. Appling these criteria to sort through cases reported on LexisNexis, we found 272 cases prosecuted in the City of Los Angeles (1997–2015). Linking each defendant to each victim, we mapped 1,002 victimization dyads. You might be curious to know that 45 percent of the cases named a lone offender, and the most prevalent co-offending behavior involved a pair of attackers. About 93 percent of cases involved gun violence, and most incidents occurred in public areas—10 percent in a park, 47 percent on a street or in a parking lot, 28 percent just outside of a business, and 15 percent just outside of a residence.

To investigate gang-on-gang violence we aggregated links to the gang level. Coding all nongang-involved people to either a law enforcement organization (Los Angeles Police Department or Los Angeles County Sheriff's Department) or a nongang community group, we were left with a network of 109 groups (106 gangs). The master network included three components, with 96 percent of groups in the main component; the master pre-CGI network included 68 groups, with 90 percent in the main component; and the master post-CGI network included 74 groups, with 92 percent in the main component. Density for the master and pre-CGI networks was 8 percent, and the post-CGI network density was about 11 percent. After reading chapter 9, you should recognize these descriptive statistics as indicative of sparse networks.

One of the analyses we did to explore the effects of CGIs was to generate pre- and post-CGI ego networks for the most aggressive and most victimized groups—ten groups in total. Of this set, two Blood and three Crips gangs appeared in both sets of analyses. Examining attacking behavior with a triad census of the egonets, we discovered the following:

1. Conflict networks exhibited dramatic change post-CGI, and most change involved new conflict.
2. Egonets tended to increase in size post-CGI, meaning that conflict expanded to include more gangs.
3. At least 60 percent of the gangs experienced an intensification of conflict involving in-star formations (i.e., they were victimized by two other gangs as captured by triadic structure 021U), whereas 40 percent of gangs exhibited an intensification of out-star aggression (as captured by triadic structure 021D).

These findings led us to conclude that the most active gangs tended to become more deeply enmeshed in violence post-CGI. And this small set of gangs was centrally positioned within a dynamic web of nonreciprocated conflict exhibiting complex hierarchical structures. What remains to be seen when the main study is complete is how prevalent this pattern is and whether we will discover a more nuanced evolutionary shift with a larger set of data. One of the structural characteristics we plan to explore is whether distinct subgroups form as more gangs are enjoined.

SUBGROUP IDENTIFICATION

Network analytics provide us with many different ways to search for sets of actors who exhibit local relational structures that differ from the larger network within which they are located. Sometimes the local patterns of connectivity differ so much from the larger community that the set of actors is classified as a subgroup or cluster. Within any network, we are bound to find groups of actors who interact with each other so much that they constitute an identifiable subset.

Since we think highly cohesive groups tend to share norms and have common ideals and goals, it is likely that these substructures have a lot of influence on members—and greater constraint leads to a high potential for peer pressure. If the group is really tight, it might behave as a single unit. In these situations, particularly if the network exhibits a number of tight clusters, we might want to aggregate each distinctive set of individuals to a **supernode,** to reduce network complexity and better understand the way the system operates. For instance, we can look at how unique cells fit within terror networks. Investigating supernodes requires first recoding each tight subgroup into a single supernode, after which the structure of connections linking supernodes can be reexamined.

Even though subgroups can be highly cohesive, it is likely that you will find some positional differences among members. Sometimes, therefore, you will want to focus on the topographic features of cores and spanners. Highly cohesive groups often have a small set of **core actors** who are positioned in such a way that they might control the group and dictate norms. Since core actors are positioned differently from others in the subgroup, divisions between core and peripheral members may develop. Alternatively, if the core splits, and group members are divided between an "in-group" and an "out-group," the singular unit might split into warring factions. Another set of actors that is important to identify is made up of spanners. **Spanners** have ties to other subgroups and may act as critical conduits of information.

Subgroup analytics are generally divided into two groups—top-down methods and bottom-up methods. **Top-down** methods preform functions on the whole network to see how it might split into subgraphs. **Bottom-up** strategies build groupings or clusters by gathering nodes, using inclusion rules. Six different analytic strategies are described below.

Top-Down Approaches

Component analysis looks at the composition of the graph with a detailed exploration into the structure of each component found. The process starts with identifying the number and size of components in the network. A network with one component is a **connected graph.** If the network has two or more components, it is classed as a **disconnected graph.**

Disconnected graphs must be described in greater detail, treating each component as a unique network. As we learned in chapter 9, this description might include information about the percentage of nodes appearing in each component. Additional analysis might investigate the patterns of relations in concert with nodal or tie attributes. For instance, subgroup analysis of components might explore density or examine the level of star and chain-like formations, along with the types of mechanisms linking members of each component (assuming multiplex relations are observed).

If the network is directed, then subgroup investigation will determine the strength of the component. **Weak components** are configured in such a way that not every member is reachable. This means the pattern of arcs—in terms of with whom the ties originate and terminate—are arranged in a way such that some people may not be able to receive or send information. Arc patterns may identify subsectors of unreachable people. On the other hand, components might be strongly configured—double-headed arrows join all members of the **strong component** directly (all links are reciprocal).

Investigations of criminal networks often generate networks that are characterized as disconnected graphs, hence, component analysis is likely to be familiar to you. Some of the published analysis described in chapter 6, involves careful study of the main component. What makes the work reviewed in the prior chapter different from a true component analysis is the lack of attention paid to investigating the characteristics of the other, smaller components. A true component analysis of subgroups would report on all components.

Before moving on to the next technique for top-down subgroup identification, I want to reiterate something important: criminal networks might appear to be disconnected as a result off missing data. We reviewed the problem of missing data extensively in chapter 8, but it is relevant to mention again because, by default, a component analysis provides some indication of where gaps in information might be. A component analysis, therefore, can also aid your efforts to assess the quality of network information.

Density varies throughout sections of graphs—this information can be used to partition the network into factions (subgroups), maximizing within-group density and minimizing between-group linkages. **Faction analysis** seeks to find an optimal solution (best fit). Essentially, investigators start with the number of subgroups they expect to find in the network. The starting number could be based on some information about the network, a research question, or a theory. If no target number is specified, most software uses a default setting—that is, two groups. This target number is a partition or split expected within the network. Once a target number of factions is identified, the algorithm sets about fitting groups. It starts by arbitrarily assigning nodes to groups, calculating fit, and making adjustments (i.e., reassigns some nodes). Then it recalculates fit and makes adjustments, trying nodes in different factions. The fitting process continues until the maximum number of iterations is reached (set by the researcher or a software), or there is no improvement in fit. If there is no improvement in fit, then the best configuration of the designated number of groups has been reached.

If engaged in an exploratory exercise, you would then try another target value (I usually increase the target number of expected factions to try to fit more subgroups, rather than starting high and decreasing expectations.) Then repeat this exercise with another target number of factions, continuing to try partitions until the fit value stabilizes. Stabilization is reached when you find little change after adjusting the number of target factions. In other words, the optimal solution has been found that splits the network into subgroups with maximum within-group density.

Since faction algorithms generally try to maximize density within the factions and minimize links between subgroups, it would not be a viable method of identifying subgroups if the overall network were completely connected or very dense. If the graph were really dense, any partition would be equivalent to any other. If it were to have no natural breaks, the algorithm would not be able to accurately assess the significance of any change in fit.

Generally running the algorithm a number of times from different starting configurations will help to establish the robustness of the solution. Remember, the algorithm begins each run with an arbitrary assignment of nodes to factions. So each run will start with a different configuration. If separate runs agree, meaning the resulting allocation of nodes to specific factions is consistent, it is likely that there are clear partitions in the network, suggesting the existence of distinct subgroups. Since some software offers a choice between different measures of fit, runs using different metrics to assess fitness could be compared.

Before running an analysis, consider the sharpness of the network boundary. Efforts to map criminal networks can include people who are not fully embedded in the network. Sometimes these peripheral actors have a single link to the group. Referred to as **pendants,** individuals with a single direct connection might represent a fuzzy boundary. Since faction-identifying algorithms use density to detect groups, pendants will never be included. Because they have only one connection, pendants cannot by definition be cohesively embedded in any subgroup. Further, if you are working with a really large network, these additional people will just slow down computing time. So it is best to remove pendants before launching the analysis.

Next in our tour of top-down subgroup identification techniques is block and cut point algorithms. **Block and cut point** algorithms find structurally important nodes, which are weak spots in the graph (a.k.a. split points). The nodes constitute weak points in the sense that if they were to be strategically removed, the network would divide into parts, called blocks or bicomponents. The structure falls apart without cut points because these nodes function as brokers among otherwise disconnected sections of the network. In other words, removing the person acting as a cut point would split a connected network into components. Since this person has ties to unique sets of actors, who would not be able to share information or resources if the person were removed from the network, removing the target cut point would require the termination of the ties to each set of actors. (So at least two relationships would need to be terminated in order to remove the cut point from the network.)

If you are interested in this technique, see McGloin 2005. For this study, McGloin examined the web of co-offending among individual gang members active in Newark, New Jersey. Mapping multiple layers of associations, she found that gangs were loosely organized with pockets of cohesion, and certain individuals acted as cut points linking to individuals and groups of gang members. These findings led McGloin to suggest that the cut points within gangs are notable for two reasons. First, cut points might be useful as communication agents for a deterrence message, and second, these individuals might be vulnerable to the pulling levers strategy.

The **Girvan-Newman algorithm** finds structurally important edges (edge-split points). In this instance, structural importance is such that removing an edge would fragment the network into separate components. The first step of the process is to set the maximum number of cohesive subgroups expected in the network. The threshold value is referred to as *k*—meaning k groups. Then the algorithm deploys an iterative process that starts with identifying edges with the highest betweenness centrality—this is edge betweenness centrality. Notice that the focus is not on a node; rather, the objective is to find a channel or tie between nodes that is central to the network. **Edge betweenness centrality** calculates the number of times each edge lies on a geodesic path between a pair of nodes. The tie with the highest score is deleted. The network is reassessed by counting the number of components that now exist. If the value exceeds k groups, then the algorithm stops. If it doesn't, edge betweenness is recalculated and the next highest score edge is removed, and so on. The iterations will continue so long as the number of components identified is less than a user-defined maximum. Overall, the objective is to fit a solution that generates cohesive subgroups representing distinct communities.

Modularity (a.k.a. the fit of the solution) can be assessed with a **Q score.** Q compares the number of ties observed in the "new components" to what might be expected if ties were distributed randomly among the nodes in each component. Positive values suggest the algorithm found significant groupings, whereas negative values suggest that the components are less cohesive than random. When the Q score stabilizes, meaning that removing more edges does not generate more cohesive clusters, the optimal number of subgroups in the network has been located.

A related analytic process identifies **lambda sets.** As above, the focus is to find connections that, if removed, would result in splitting one connected graph into unconnected components. While both lambda sets and the Girvan-Newman algorithm identify high betweenness edges, the lambda algorithm ranks each relationship in the network on the basis of how important it is in maintaining flow among all actors. The edges selected for removal are those that most disrupt the network as a whole. This means that it prioritizes which of the high betweenness ties to remove for maximum disruptive impact. I suspect that by now, criminal justice applications are already coming to mind, so rather than elaborate further, I will move on to bottom-up subgroup identification methods.

Bottom-Up Approaches

Cliques are subsets of a network, wherein every actor is adjacent to (directly connected) to every other actor. This means that the group is "complete" or completely connected. The default is three, which means that among a set of at least three actors, everyone is connected. (Remember our conversation about transitivity?) In essence, therefore, the algorithm starts by identifying all transitive triads, after which it adds those people to each subset who are also connected to everyone else in the subset. The clique has reached its "maximal" size when no other actors can be added because they lack ties to all members of the group. In essence, the function used to classify nodes into subgroups is an inclusion density criterion. Keep in mind that, since the process involves gathering sets of nodes with complete connectivity, it is feasible that some nodes will not be part of a clique.

Cliques can overlap—for example, when a person belongs to more than one clique. Clique identification can be done on directed networks, but the analysis will include only reciprocal ties, so fewer cliques will be found in the graph than if the network were undirected. If you are having difficulty

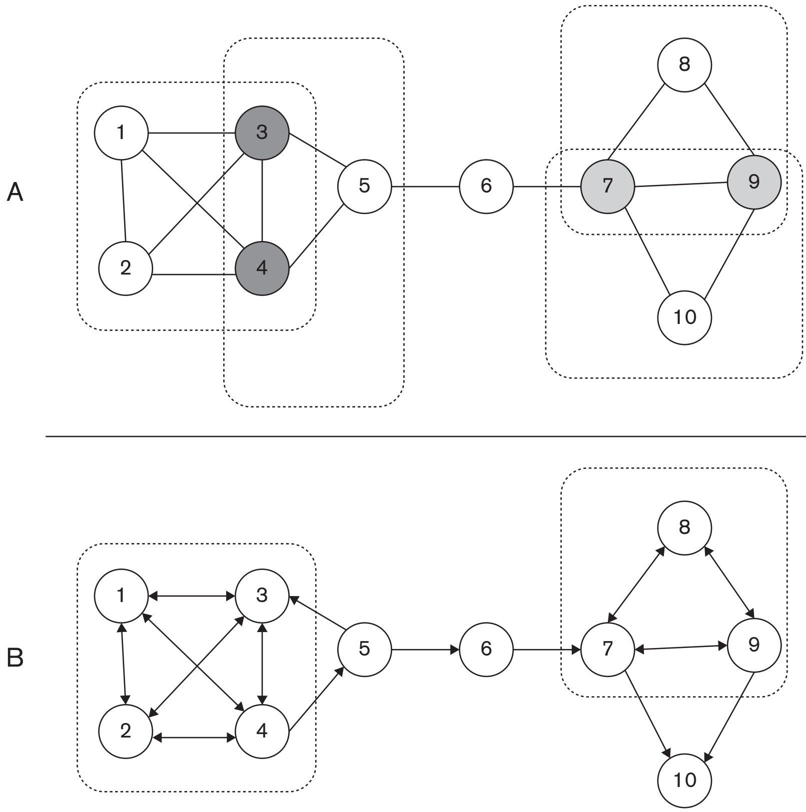


FIGURE 10.3. Cliques visualized. *A*, four cliques are found in this nondirected network, and four people have overlapping membership in two cliques; *B*, two cliques are found in this directed network, and no overlapping membership is observed.

visualizing this, take a good look at figure 10.3. The figure depicts a network with ten actors and fourteen ties. (I borrowed this configuration from Borgatti, Everett, and Johnson [2013, 183], because it always makes sense to my students.) Cliques are identified with dashed symbols. The first panel illustrates overlapping cliques in the nondirected network, and the second panel identifies cliques if the network were directed. Persons 3 and 4 (dark gray) are both members of the same two cliques, and persons 7 and 9 (light gray) share overlapping membership in a different set of two cliques. In total, panel A shows four cliques. Contrast this level of subgroup cohesion with panel B, which exhibits two cliques with no overlapping membership. Notice that in the directed network illustrated in panel B, clique membership requires that pairs be reachable—meaning relations must extend both ways.

N-cliques involve a similar group-identification protocol, the difference being that the researcher defines the minimum number of people who must be interconnected to constitute a subgroup. So if we conducted an n-clique analysis specifying four relations, only the subgroup including persons 1, 2, 3, and 4 would be counted as a clique. A clique comprises at least four completely connected actors. If we specified five relations, then, at a minimum, five people would have to be completely connected. In this instance, no cliques would be found within the network depicted in figure 10.3.

K-Core procedures use degree centrality to identify subgroups. Thus, for example, people are considered a k-core subgroup of four if all members have degree centrality scores of four (raw value) and all members are connected to each other. The same formula pertains to a k-core of five, a k-core of six, and so on. K-core algorithms can identify subgroups within the main component and all smaller components. In short, K-core functions gather sets of nodes that are connected to each other and exhibit criterion levels of degree centrality (hubness). There is an intuitive appeal to this method, in that having a sufficient number of connections within a group might make a person a more fundamental part of the group, even if that person is not connected to most other members. Core membership in a subgroup depends on connectivity instead of immersion. I find K-cores more useful visually when I am working with a dense network. This subgroup identification technique can highlight tight clusters of varying sizes, exposing what might constitute the structural cores of the network.

Example Application

To illustrate the benefits of subgroup identification when applied to criminal networks, I examined a drug-trafficking community mapped using information extracted from a threat assessment of criminal enterprise. Box 10.2 provides a short description of the data—a full report about the data source (and analysis) is in Malm, and Bichler (2011, 2013) and Malm, Bichler, and Van De Walle (2010). Please take a few minutes to review box 10.2 before proceeding.

While the source data described in box 10.2 map a multiplex graph of 2,198 people linked through 2,748 ties, only co-offending activity was used to illustrate subgroup identification. Co-offending activity was the most prevalent type of tie found among a subset of individuals named in the

BOX 10.2. MAPPING A DRUG-TRAFFICKING COMMUNITY

On an annual cycle, the Criminal Intelligence Service Canada and the Royal Canadian Mounted Police generate a National Threat Assessment Report consolidating recent intelligence about known criminal groups involved in organized crime. Criterion behavior is here defined as a *crime committed by any group of at least three people that is punishable by more than five years in prison and has a material benefit—meaning the primary motive is profit* (Canadian Criminal Code, section 467.1).

Data were extracted from the 2007 "E" Division Provincial Threat Assessment report, which compiled information about individuals and groups operating within British Columbia and the Yukon Territory from 2004 to 2006.

Sampling originated with a list of known crime groups and individuals from the preceding threat assessment; then about forty RCMP intelligence officers examined all information sources to update relational information. They began by querying national information systems (reviewing thousands of police intelligence reports and crime reports), after which they interviewed law enforcement personnel and prosecutors, examined wiretap transcripts, and conducted offender interviews. Intelligence was integrated into a narrative for each group. Narratives included details about all members of the crime group, as well as co-offending information and all known associates, including legitimate business partners, relatives, and friends (this was a multiplex network). About 79 percent of people had co-offending ties to others in the network.

Most of the criminal activity captured in this threat assessment involved some association with illicit drug production and trafficking. Of the 2,198 people identified in the threat assessment, 92 percent were themselves directly involved in the drug trade or were associated with someone who was directly involved. Readers should also note that, of the 186 criminal groups examined, 24 percent were outlaw motorcycle groups, predominantly Hells Angels.

threat assessment and associated with organized crime activity from 2004 to 2006. Of the 1,654 actors (1,808 ties), 62 percent were in the main component. Removing pendants (people with only one co-offending relationship) generated a subgraph of the main component, which included 316 people linked through 483 co-offending relations.

Active subgroups are likely to exist within this community of co-offenders. So to get an initial idea about the inherent clumpiness of the graph, I would start with a faction analysis. Running the procedure several times, I found fifteen unique factions. Recall that faction analysis identifies subgroups of within-group density. Faction analysis provides an interesting way of finding cohesive groups or communities operating within the larger set of co-offenders. This information aides investigations of highly active people (prolific offenders can be identified with a positional analysis described in chapter 9) because we now have information about the cohesive local network encapsulating the target individuals. While this intelligence is useful, it does not tell us how to disrupt the criminal enterprise—remember these individuals are involved in organized crime.

Moving on, two subgroup identification procedures highlighted in figure 10.4. Panel A visualizes a cut point analysis. The individuals identified in black, if removed from the network, split the graph into disconnected components. Cut points in a co-offending community are important because they may partner with others from different sets of people. Notice that node size varies by degree centrality calculated on the full graph. It is clear that while some cut points are hubs in the larger network, many are not once we remove pendants and focus on the main component. Targeting cut points may have strategic value, particularly if a number of such individuals can be pursued simultaneously. Being mindful that we are investigating a subgraph of co-offending, when a multiplex network exists, we can only tentatively suggest that these cut points are vulnerable points in the network. If kinship or legitimate partnerships exist with other structural patterns, it may be very difficult to target cut points for removal from the graph. Even so, this analysis provides some insight into vulnerabilities that might be targeted to disrupt criminal operations.

Figure 10.4, panel B visualizes the results of a K-core analysis of the main component subgraph with pendants removed. Cores of one are white; cores of two are represented with gray symbols; and cores of three are illustrated with black symbols. Several distinct subgroups with varying levels of cohesion appear in the graph. Notice that the dominant hubs (larger symbols) do not appear in the most cohesive cores. Without exploring the actor attributes, but knowing that most of the criminal enterprise activities listed in the threat assessment involve illicit drug trafficking, we can see that these cells represent different stages of drug production and distribution, regional divisions in drug trafficking, or distinct communities. Now that we have our sights on subgroups, the natural inclination is to start layering other bits of information. How old are group members? What is the nature of their co-offending activity? Do they have any other types of ties? Are they members of the same formal organized crime group? The desire to layer additional variables leads us to advanced analytics.

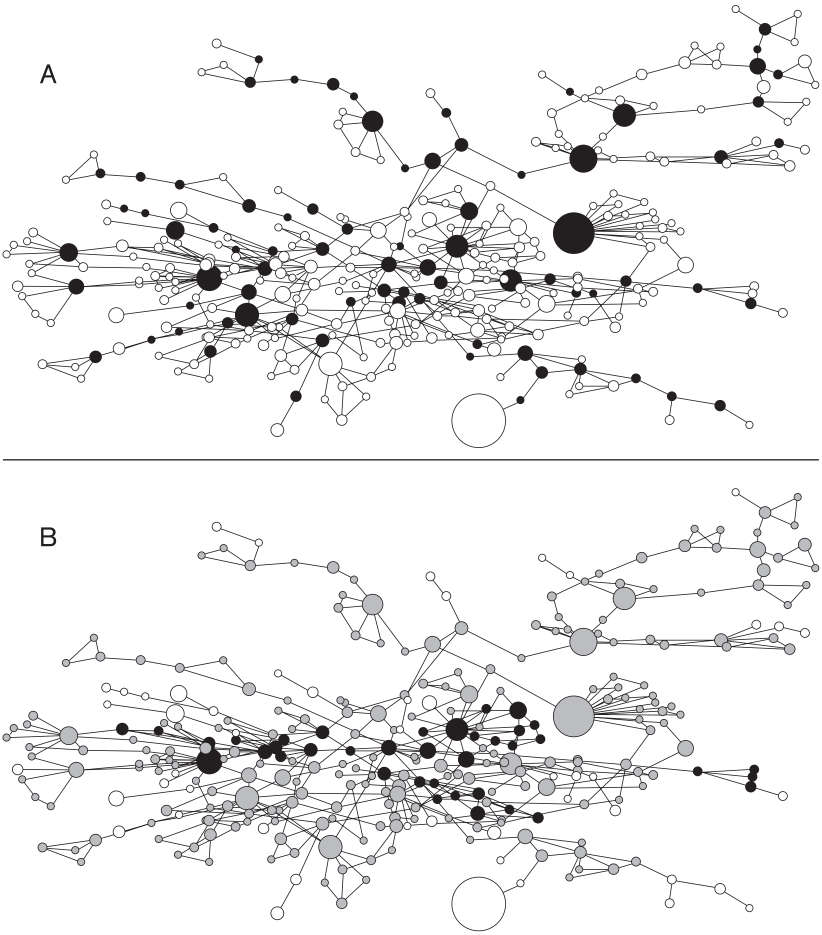


FIGURE 10.4. Subgroup identification of co-offenders with cut points and K-cores. *A*, cut points illustrated in black; *B*, K-cores illustrated in black. Node size varies by degree centrality, which was calculated on the full graph before pendants were removed.

ADVANCED ANALYTICS

Advancing one more step into the analytic realm, we find ourselves considering hypotheses about the distribution of relations among actors, such as whether prolific co-offenders have more nonredundant ties. It is also likely that you will be interested in looking into inferential explanations for network patterns or emerging structures. Inferential analysis gauges whether findings are simply a chance occurrence or are illustrative of patterns that exist in a larger "population." For instance, we might want to predict how other international drug-trafficking networks might respond to repeated targeting of smugglers with high betweenness centrality. This section of the chapter will be relatively short, for my aim is to simply introduce some techniques and point you toward resources where you can find more information. To begin, let us consider hypothesis testing.

When investigating networks, we might be interested in whether two groups have different densities—for example, we might want to know if marijuana smugglers are part of less dense networks than groups running growing operations (producers). If we discover that groups of smugglers have networks with an average density of .14, and growers are part of networks with an average density of .39, we might want to know if these observed values are significantly different, so that we can gauge how likely the differences are to characterize other operations that are not part of our original study.

Hypotheses tests in SNA typically use a bootstrapping process to generate a sampling distribution against which to test the observed estimates. In general terms, the software executes a protocol to compute an estimated sampling distribution of density measures by randomly extracting thousands of subsamples from the observed network. With each subsample, a density is calculated. Typically, you can set the number of subgraphs to draw, but five thousand or more is a good threshold. Calculating the standard deviation of densities indicates how much variation we might find in any sample just by chance (called the standard error). We can use this figure to estimate how likely it is that our observed difference in mean densities are the result of chance.

We can also look at correlations between two networks containing the same people. We could, for example, compare the illicit drug activity among people who are part of a trafficking network to the legitimate business relations among all actors, to see how much illicit and legitimate enterprises overlap. Measures of association can be calculated for nominal, ordinal, and interval associations between the relations in two graphs. As with conventional statistics, several measures of association are available. Which to choose will depend on the level of measurement characterizing the two sets of relations. For instance, the Jaccard coefficient is used for dichotomous networks (e.g., binary relations coded 0 for no relation and 1 for a relation). If both networks have valued relations (i.e., for each relation you have an estimate of the strength of the tie), then use Pearson correlations. In this example, of drug-trafficking and legitimate-enterprise networks, I would use the Jaccard coefficient, as the networks record simply the existence of relations. Measures of association generally range from 0 to 1, with higher scores indicating more overall structural similarity between the networks examined.

To test for the significance of the association, standard errors are generated using quadratic assignment procedures (QAP)—our example would be called a QAP correlation. QAP is a resampling method, much like bootstrapping, that is used to estimate standard errors to assess significant of findings. Essentially, the algorithm generates a set number of graphs by randomly shuffling the dependent variable (redistributes the dependent variable, or, in our example, one of the networks is resampled to generate a distribution). This set of permutations is then used to generate an expected distribution of patterns against which it is possible to compare the observed data (to determine how likely the observed pattern is). For this example, we would be determining whether the association found between illicit and legal relations (pairs of people engaged in illicit activity are also involved in legal enterprise) is significantly different from chance. In other words, if we found a Jaccard coefficient of .63, indicating that 63 percent of pairs of people engaged in illicit business also conduct legal business activity, would it be likely that this moderately strong association is real? If the pattern were deemed significant, then it would not likely be a chance finding.

Moving on to prediction. If you want to use other information to predict whether a relation exists—for example, the existence of another relation or an actor's characteristic—you can use a QAP regression procedure. If you think relations are nested within groups, try a multilevel model. If you are interested in predicting how networks evolve in time, use a stochastic actor-based model. Since these analytics are outside the scope of this introductory text, I will not elaborate further. Just know that there is a full range of statistical models available.

Boxes 10.3 and 10.4 report on nine of the most widely used software programs. Practitioners will quickly notice that professional software like Sentinel Visualizer®, Visallo, Case Closed Cloud™, Sintelix, Link Explorer, Wynyard Advanced Crime Analytics, among others, are not discussed here are three reasons. First, they cost too much and require expensive technical support. Most people starting out in SNA do not have access to these programs, so there is no point profiling them here. Second, professional practitioner-based software programs are designed to run off data management systems. Since network criminologists advocate integrating information from multiple sources, professional software tied to a single system is not necessarily the best place to start learning about criminal networks. And third, academic-based software provides a broader range of advanced analytic techniques with greater transparency of functions and capabilities. You will notice that the recommended programs are free, or nearly so, and offer full technical support through many platforms at no cost. More importantly, because they were designed for scientific use, programmers did not hide algorithmic details behind proprietary excuses—you can find out what each button does. This is not to say that professional software is simply a useless black box spitting out meaningless figures; rather, if you want to learn how to do SNA properly, you need to begin with scientific software. What I value most about the programs showcased is that the developers provided access to a range of training opportunities that marry software instruction with SNA theory, methods, and analytics. The software links provided are portals to software downloads and technical resources. Please explore these sites thoroughly, as they are information gold mines. Admittedly, some websites are easier to navigate compared to others.

I draw attention to general purpose software, as well as to advanced analytic programs designed to do multivariate hypothesis testing. Each of the general purpose software programs have data management tools, can import and export in different formats, and provide a range of analytic options. They vary with respect to the "friendliness" of the user interface, capacity to deal with large data, and quality of visualizations. All things considered, I recommend that you start with UCINET and its partner visualization program, NetDraw. This software set uses a windows interface that resembles SPSS and is therefore the easiest for most people just starting to examine networks. A free textbook accompanies the software, but I usually suggest that people consult Borgatti, Everett, and Johnson (2013). When you are ready to graduate to more advanced visualizations and exploration of large data, try GEPHI.

Once they are comfortable with basic SNA analytics, practitioners who are crossing over from academic SNA to intelligence applications may opt to migrate to ORA. ORA was developed by CASOS under the leadership of Kathleen Carley at Carnegie Mellon, with support from the Office of Naval Research, the Department of Defense, the Army Research Lab, NASA, the National Science Foundation, and the US government. Designed for the intelligence community, this software has a capacity and functionality that is well suited to criminal investigation. Everton's (2012) *Disrupting Dark Networks* is a very useful resource for those making the transition. Everton provides technical instruction on the software while interweaving discussions of dark network disruption theory with practical applications. Several chapters demonstrate how to use UCINET, NetDraw, Pajek, and ORA.

BOX 10.3. OVERVIEW OF GENERAL ANALYSIS SNA SOFTWARE

UCINET (www.analytictech.com/) is a general purpose software package for the generation, manipulation, and analysis of networks. It is used for exploration and hypothesis testing. UCINET runs with a Windows interface, and its functionality resembles that of SPSS. The program is inexpensive and comes with a free visualization tool (NetDraw). Free support includes a textbook with software examples, user group, tutorials and teaching materials, a listing of PhD courses, training opportunities, and sample data, including criminal networks. Resources are available in Spanish. The software was developed by Borgatti, Everett, and Freeman (2002).

GEPHI (gephi.org/) is useful for the exploratory analysis and visualization of big data and can work with up to a hundred thousand nodes and a million edges. The 3-D rendering is very effective. (Dramamine is optional!) The software is free for noncommercial use, but donations are accepted. While a limited set of metrics is available, it offers a full range of display and printing functionality, including timelines and dynamic filtering. There is extensive tutorial support with useful videos, supplemented by forums, blogs, papers, WIKI, and a bug tracker. Support materials are available in French, Spanish, and Chinese. The software was developed by Bastian, Heymann, and Jacomy (2009).

ORA (www.casos.cs.cmu.edu/projects/ora/software.php) is a general purpose program for big data (up to 10^6 nodes) management and analysis, with statistical and visualization capabilities that can handle time and space variables. Free (ORA-LITE) and professional versions are both available. The program permits analysis of one, two, and multimode networks. It also has interoperable capabilities—it can function in a standalone mode or a service within a web architecture. User support includes training and sampling data. The software was developed by Carley (2001–2018).

PAJEK (mrvar.fdv.uni-lj.si/pajek/) enables the exploratory analysis of large data and has the capacity to integrate spatial references. It is also useful for analyzing signed networks (edges have positive and negative values). The software is free for noncommercial use. User support is available in many languages, but web support sources are a bit harder to navigate compared to other software, though a complete range of materials is available. Be aware that mathematical language is heavily used. The software was developed by Batagelj and Mrvar. (1998).

Advanced analytic software is designed for specific types of modeling and hypothesis testing. You should not approach these programs until you are comfortable generating, manipulating, and exploring networks with at least one of the general purpose software programs. One of the best explanations for exponential random graph models (ERGMs) is in a textbook

BOX 10.4. OVERVIEW OF ADVANCED MODELING SNA SOFTWARE

Exponential Random Graph Models (ERGMs) (www.melnet.org.au/pnet/) is a free software for noncommercial use that supports hypothesis testing with a suite of programs.

- **PNET.** Simulation and estimation of ERGMs for one-mode networks
- **MPNET.** Multilevel networks of ERGMs for two-mode and two-level networks that run autologistic actor-attribute models
- **XPNET.** Simulation and estimation of ERGMs for two- and one-mode networks

Programs come with excellent user manuals (at no cost). The software was developed by Wang, Robins, and Pattison (2009).

RSiena (www.stats.ox.ac.uk/~snijders/siena/) runs stochastic actor-based models that enable multivariate, dynamic (longitudinal) analysis of complete networks (not egonets) and test hypotheses about purposeful actor behavior. The software is free for noncommercial use. It runs on the R statistical platform. There is a substantial amount of support materials, available at no cost, including a user manual, scripts, example data, example research, and applications. The software was developed by Ripley et al. (2018).

NetLogo (ccl.northwestern.edu/netlogo/index.shtml) is a multiagent programmable modeling software (simulation modeling) that can be used to study the behavior of heterogeneous, autonomous agents over time. The program is free for noncommercial use, but donations are accepted. Simulations generate results that can be imported and analyzed with other software to test hypotheses. Web support includes many materials at no cost, including a user manual, a library of sample models, a user community, many support groups, and a user blog. The software was developed by Wilensky (1999).

coauthored by one of the originators of the recommended software (Lusher, Koskinen, and Robins 2013). It is well worth the money.

Armed with the knowledge gained in the first ten chapters of this book, you are now ready to launch some of your own investigations into criminal network structure. But first I recommend that you read the next chapter carefully. I offer some advice for packaging your work so as to appease the critics and, more importantly, for making your analysis accessible to readers.

11. Producing Professional Products

> [S]ocial structure is not random but patterned. It is patterned behavior in which each social unit is seen as embedded in a network or web of other social units who respond to it and to whom it responds. At whatever level we choose to tap in to the study of social structure, our interest is always in patterning.
>
> FREEMAN, "Social Networks and the Structure Experiment."

Because social network analysis (SNA) is relatively new to crime science, most people do not understand what they are reading. Do not underestimate the wisdom of this statement. Unfortunately, I have plenty of examples to illustrate this point. To impress on you the magnitude of the problem, I thought that it might be useful to share snippets of comments I received from two anonymous reviewers. It may be surprising to learn that the comments refer to the same manuscript.

> This paper also falls flat on its face on a number of levels. Most troubling is the overall tone of the research, which appears to be conducted by scholar(s) with extensive expertise in and knowledge of scripting, network analysis, and hyperquantification of social phenomena . . . but zero intimate knowledge and/or expertise on the illicit art and antiquities trade. The authors, in other words, seem all too eager to try to quantify the hell out of cultural property issues with little sensitivity to the complexities of these issues. . . .
>
> Most importantly, however, since the operationalization of concepts and the description of how decisions were made regarding operationalization are both fundamentally flawed, this makes the methodology and subsequent findings and conclusions problematic. In short, I got bored and stopped reading, because if the method is problematic, so are the results.
>
> Based on the critique offered above, my recommendation is to reject this paper. It is fundamentally flawed in both its theoretical grounding and methodological execution, and it is not of the caliber expected for this journal.
>
> Anonymous Reviewer A

> This study employs a combination of script and network analysis to examine the illicit trade of fine art. Building on Morselli and Roy in

> particular, the study seeks to identify which actor groups are positioned to control the flow of art in the market and what tools foster market interlock between illicit and legal art markets. The study is quite interesting. It examines a form of crime which has never been a topic of interest in criminology but which shares, at the same time, many characteristics with other types of organized crime.
>
> The methodology is clear and well-described and the findings clearly presented. The method is obviously characterised by a number of limitations but the authors do well in presenting them at the end of the study. Another interesting point in this study lies in its attempt to design preventative strategies—the most important benefit of crime scripts often forgotten by researchers. The quality of this study is of a very good standard and should be published in [journal name omitted]. A few and very minor comments that might improve the study are outlined below.
>
> Anonymous Reviewer B

Fortunately for this manuscript, I am happy to report that the piece eventually found a home. The lesson learned from my experience is that professional products are not easy to develop because most audiences have no formal training or exposure to network theory, methods, or analytics. Networked criminologists, practitioners, and analysts operate within a somewhat hostile environment. Generating professional products that describe network research in the face of such challenges is difficult, which leads me to the objective of chapter 11.

Criminal network analysis is not part of the mainstream science of criminology. We are still trying to figure out how our understanding of crime and deviance will change with the integration with SNA, which methods and analytics are of greatest utility, and whether a networked understanding will offer crime control gains over the long run. Adding the criticisms of skeptics to the equation led me to decide that concluding the book with a traditional summarizing chapter seemed a bit premature, and much less useful, than offering some advice. So instead of a summary proclamation, this chapter provides tips on publishing in peer review outlets, generating graphics, and making presentations about networks. I also provide a short introduction to professional associations and organizations dedicated to SNA scholarship, so that you know where to find courses and other instruction, as well as opportunities to connect with SNA-oriented investigators. While some readers will lament the lack of quotable summary statements, I hope that the advice offered here will help placate the skeptics.

MAKING CRITICS HAPPY

Networked criminology is an emerging discipline, and we are only beginning to get around to developing a set of reporting standards. Assessing the current state of knowledge regarding the structure of criminal organizations involved in illicit drug trafficking, Bichler, Malm and Cooper (2017) argue that inconsistencies in reporting habits, such as omitting details about methodology and failure to include descriptions of networks, as well as the lack of reporting of standardized metrics, constrains the advancement of networked criminology. But more immediately, it gives critics things to complain about and excuses to reject manuscripts. To correct this problem, we outline seven suggestions (note that I modified the explanation accompanying each point to better explain the statement within the present context).

1. Researchers and analysts must clearly describe how they generated the networks. Network generation involves making decisions about what constitutes a link between actors and where they obtained information about these connections, as well as whether the relations have an inherent value or directionality. Each decision can significantly influence what the graph maps. This means that any report of findings must explain to readers what constitutes a tie, whether ties were valued (or binary), and whether the network was directed (or symmetrical). A related issue is the composition of subgraphs. After generating the initial network, researchers often extract a subsample for analysis (i.e., a principal component). A clear explanation of subsample extraction is necessary. It is important to provide these details, as decisions made here may radically influence the results. Even though editorial preferences will relegate some of these details to footnotes or a technical appendix, it is important not to cut this information in the final edits, as it reveals how methodological decision making influences the results and helps to promote replication.
2. Structural differences among groups are often associated with how individuals are connected; thus, investigations must be specific about what constitutes a connection between people (or groups). Criminal network analysis is still in its infancy, and we are still figuring out what the important binding mechanisms are and what advantages different types of connections have for criminal behavior. Consequently, it is advisable to provide technical details

about the link-coding process and to repeat analysis using different types of ties, so as to advance our understanding of relationships that sustain criminal behavior.

3. Sampling procedures and network boundaries should be described. SNA research applies many different sampling strategies, including hybridized techniques using multiple procedures. Even within research using a case study approach, focal individuals referred to as seeds are often selected as the starting point around which a network is generated. Using selection criteria, we add individuals to the network who have some type of association with the seed individuals. While most authors provide this information, they do not always explain where the network stops—the network boundary is not well defined. Conceivably, one could continue for several steps out from a focal individual. For example, should the friend-of-a-friend-of-a-friend still be part of the same group? Does someone belong to a crime group if they attended only one meeting or partook in only one crime incident? While it is good to explore different sampling and inclusion rules, these details must be provided alongside results. Without these details, replication and cross-network comparison are limited.
4. Irrespective of the stated research objectives, a set of basic descriptive statistics must be reported. As we learned in chapter 9, the scientific method stipulates that basic descriptive statistics are required when reporting results; SNA is not exempt from this foundational tenet. Generally, the set of basic descriptive statistics to report includes the number of nodes and ties (or arcs), density, number of components, average path length, average degree, and degree centralization for each network under examination. If a subset is drawn, two sets of values may be necessary—descriptive statistics for the full network and descriptive statistics for the subset.
5. Standardized values for all metrics used to test hypotheses or answer research questions need to be reported. Because network size influences many statistics, standardized versions of the key metrics, referred to as normalized values, are available. While raw values have an inherent interpretability, and are therefore widely preferred, results should also include normalized values where possible (i.e., normalized centrality measures). Reporting normalized values enable cross-network comparisons that will advance knowledge.

6. With such a rich body of existing research, replication should take precedence. In our excitement to explore all of the analytic possibilities that SNA offers to criminology, many of us are racing ahead and forgetting a fundamental tenet of the scientific method—the importance of replication. Repeating investigations using different samples or different inclusion criteria will meaningfully advance networked criminology. We need to see if different networks share similar properties when the same analytics are applied.
7. It is important to construct titles, select keywords, and write abstracts using standard terms and phrases to ensure that related research is identified, irrespective of the search engine used. Including standard terms and phrases improves the research process by making relevant material more visible, while building a more cohesive scientific foundation. For instance, if we all included "networked criminology" or "criminal network analysis" as a key term, it would be much easier to identify materials relevant to the field.

Following this advice will help to ensure that your reporting is complete, but unfortunately this is not enough. It is important to make your manuscripts, reports, and presentations accessible to consumers.

Pay attention to flow

Criminal networks are inherently interesting. Make sure that your narrative does not become too "square." Consider including some anecdotal examples to break up the heavy material and illustrate the findings.

Do not overly complicate the report

I regularly see two mistakes. First, people try to fit too many analyses into one report. Pick just one focal analysis and save your other great ideas for the next report or presentation. Second, the equations are important, but what is more at issue are the model parameters and network-generation protocol. Technical details that explain equations make good footnotes or appendices. Spend more of your limited time explaining the methods.

Cite the software and user manual

Software is continually evolving, and bugs are being exterminated regularly. Consequently, each program may use slightly different algorithms and default settings to calculate key metrics. So be specific about which tools you employ in your analysis, as readers need to know which version of the software you are using.

Do not skimp on the details about source materials

The caliber of criminal network inquiries rests on the completeness, reliability, and accuracy of the data. And most critics are eager to pounce on any real (or imagined) methodological shortcoming to discredit the work. Do not give them ammunition by failing to explain what you have done. Remember, most critics have little knowledge of SNA, but they will all intuitively pick up on data limitations.

Follow standard scientific conventions when reporting findings

Decades of scientific exploration have established standards regarding table construction, notations for significant findings, figure legends and keys, estimates of model fit, and the like. While some discipline-specific derivations exist, networked criminology, and SNA more generally, is inherently interdisciplinary and multidisciplinary. For this reason, you must keep with general conventions that are common to most disciplines. And all critics—from editors to bosses—have some expectations about how information will be presented. Do not get overly creative and reinvent the wheel; instead, find some good examples and follow their lead.

Include at least one network map or subgraph

SNA, like spatial analysis, is afflicted (and aided) by the audiences' presumption that they will get to "see" the findings. Given how important it is to provide a good illustration of the network (or a network concept you are trying to explain), the next section of this chapter provides some instruction on network cartography.

VISUALIZING CONNECTIONS

Slices and Layouts

It is difficult to produce a professional-grade visualization of an observed network. A difficult task is made even more daunting when dealing with a large network. Too often the dots and lines are so dense that the graph looks like a mosh pit. To improve visual permeability consider the two options—slices and layouts.

If you are working with a valued network, it is possible to illustrate the network with **n-slices**. This means that you set a threshold, usually something meaningful, and depict only ties of that level and above. For example,

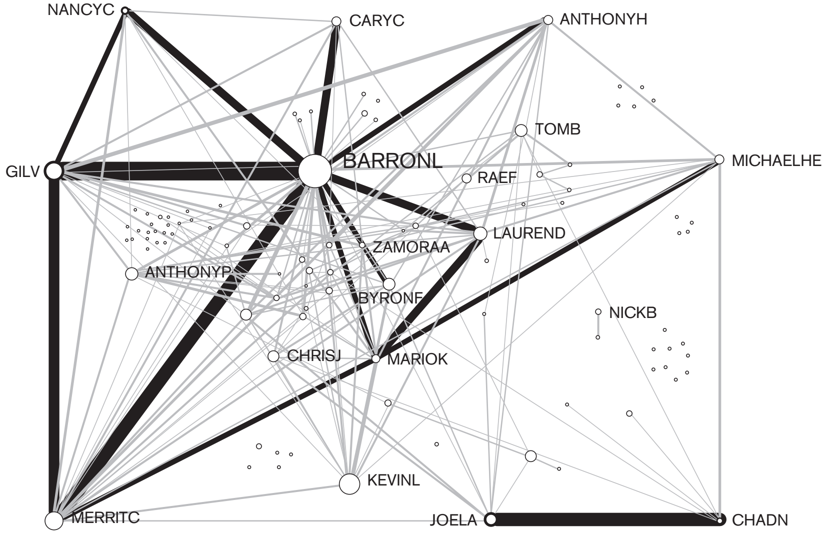


FIGURE 11.1. Officers attending five or more vagrancy calls, January 1, 2012—June 4, 2014 (Bichler, Bush, and Hagala 2014).

a five-slice graphic depicts ties valued at 5 and above. Figure 11.1 is part of an image generated for a workload analysis of the Palm Springs Police Department. The image shows a network of officers who attended vagrancy calls for service over a 2.5-year period. The analysis was done in preparation for a problem-oriented policing project. Line thickness varies with the number of calls jointly attended. The thinnest line represents five incidents, and the thickest lines indicates that two officers attended ninety-nine vagrancy calls together. While the full network includes 2,216 ties among 101 officers and fire department personnel, the five-slice subnetwork graphic illustrates 362 ties among 53 officers. Notice the free-floating isolates. They are isolates only because they responded to fewer than five calls for service with others. Normally, you would not illustrate the isolates. I did so to make the point that there are many other officers responding infrequently to vagrancy calls. Symbol size varies by degree centrality, with the most central hubs identified. Had I included all ties, the graph would have looked like a blob of ink. What becomes readily apparent are the established working relations among sets of officers (thick black lines) and which key personnel work the most with vagrants. If I were forming a problem-oriented policing team to address vagrancy-related issues, I would start by recruiting these centrally positioned officers with strong working relations.

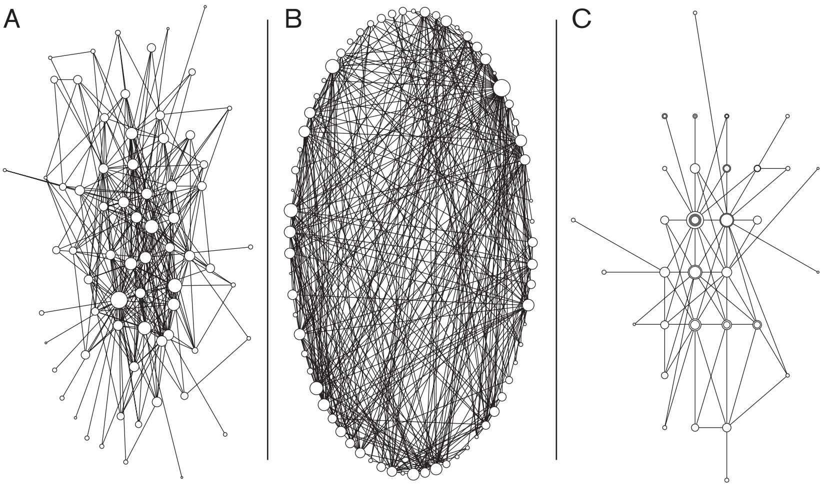


FIGURE 11.2. Comparison of layouts. *A*, layout arranged by geodesic distance; *B*, circle layout; *C*, nonmetric MDS layout.

Visualization software offers many different layout options, which can provide ways of exploring the network. Programs vary in their functionality, so start with something simple (e.g., Netdraw) before moving on to more sophisticated software programs that offer an extensive array of options. For ideas, check the web. A simple Internet search of social networks will reveal illustrations ingenious enough to impress Edward Tufte. I highly encourage you to explore what a web search reveals before creating an image. In the meantime, three layouts are illustrated in figure 11.2, drawing on the vagrancy calls for service data used in the prior example. This time it is a three-slice network (the full network is too much of a mosh pit to be useful for this illustration). This subgraph includes sixty-nine officers and 708 ties. Again, the symbol size varies by degree centrality. Tie values are not depicted (to save ink).

In this example, panel A illustrates relative positioning by **geodesic distance.** Actors directly tied to each other are positioned closer to each other. A central core of responding officers is visible. **Circle layouts** (panel B) evenly space nodes in a curved shape—usually a circle but sometimes an ellipse, as seen here. This type of graphic is useful for illustrating density when comparing different networks. Circle layouts can also reveal who is most highly connected (has the most ties). While **nonmetric multidimensional scaling** appears to reduce the number of nodes depicted, it actually does not. Instead, panel C shows a graphic where nodes are placed on top of each other if they are similar relative to others. In this example, people who are similarly positioned, meaning they have short paths to the same people, are located on top of each other. This "reduction" in visual complexity might be useful for showing how similarly positioned subgroups of actors are connected to other subsets. Now that you have a sense of the variability of layout options, we need to discuss image elements.

Image Elements

Images, whether appearing in a presentation or a report, must follow convention, so that audiences will more easily understand what they are looking at. Moreover, as with all scientific communication, graphics and charts should be sufficiently detailed, such that a person who picked them up off the street could understand them. General suggestions follow.

Symbols

- While graphics made for presentation may include color, most printed images must be in black and white, or gray scale. Color symbols used to indicate what is most "important" should be in a color that attracts attention. For instance, if my point were to illustrate noncriminal actors who were central to drug trafficking, I would color them black, and leave criminal actors to be represented with white symbols.
- Vary symbol size to illustrate centrality scores. (Most software has this capability.)
- Symbol shapes or colors can be modified (changed) to represent an actor attribute but only if there are a few categories of actors. If you use more than four or five symbols or colors, not even the most detailed key will make the graphic readable.
- If generating a set of images to depict changing relational structures, set up the first image with all nodes and relations. Then toggle relations on and off, saving images for each set of relations to be depicted. There will be a lot of symbols with no ties (isolates). Afterward, use image-editing software to erase the isolates. By reverse engineering, you will create a set of images that can be "flipped through." As if they are animated, nodes will appear stationary, but ties will appear and disappear.
- Lines depicting ties can be modified to illustrate different types of relations or values. To go beyond black and white, or gray scale, use

line thickness to indicate value and dashes and dotted lines to indicate less permanent or new ties.

- Use arrowheads for directed relations. Symbols can be resized to represent tie strength (values).

Labels

- Label symbols only when depicting small networks. Otherwise use color coding to highlight important nodes or subgroups. (See above for notes on symbol color.) Make labels two to three times larger than you think is necessary.
- Create the image with the SNA visualization software, adding labels and legend details afterward with an illustration software so the font and font size are appropriate.
- Use Arial or another simple font to enhance readability when using a small font size.
- Report the nature of relations (in the title or legend if different relations are depicted).
- Don't move symbols to make labels visible; move the labels.

Descriptive Statistics

- Always report the number of nodes and ties—in a footnote, in a title, or as text on the graphic.
- Where possible include density and other notable metrics, but do not go overboard.

Presentation Tips

One of the most sensible set of presentation tips for criminal justice audiences was put together by Professor Jerry Ratcliffe (2014), Temple University. Having spent considerable time as a law enforcement officer and decades speaking with criminal justice practitioners, he is an expert in crossing the academic-practitioner divide. Jerry always gives an exceptional presentation, using the right mix of humor and visuals to share complex ideas in an accessible manner. For the following list of his main tips, I have interjected some of my own comments. These ideas can be applied to areas other than social network research.

Tip 1: Strike a sensible contrast between text and background

Colors should have a maximum contrast, both to be ADA compliant and to avoid lighting problems (often you will not know ahead of time how the room will be set up). If you are presenting to a criminal justice audience, think about the colors you usually see. Use black, blue, shades of hunter green, grays, or white for backgrounds. Highlight and main font colors will vary, but try not to use more than two colors (one to highlight and one for the rest of the text). Never use fuchsia or pink as a text color, to highlight, or as a background color. Fire engine red can be used effectively to highlight. But please be careful. Red is an inflammatory color and it is not recommended by ADA guidelines. Use it with caution.

Tip 2: Use simple titles and points

Simpler words with fewer letters are better than polysyllabic terms. Written text should be short and in note form, not full sentences. It is okay to drop the technical language and use more accessible terms, even if they do not perfectly express the concept.

Tip 3: Get the font size and type right

Jerry suggests using a standard font (commonly used as a default setting in most other software) of size sixteen or more. The choice of font type will influence readability. Arial (regular not narrow) is a very good choice, as is any other font that is simple and has no flourishes. Calibri is another good choice.

I think this general rule of thumb regarding size is useful. But a better suggestion is to make the font size large enough to be easily readable by those at the back of the room. Seating configurations, lighting, and room shapes differ so much that a sixteen-point font might work in one setting but not in another. Moreover, people often provide handouts with three or more slides per page. All text on the handout should be readable. When you are not familiar with the room, use a font size that is larger than you think you need. As Jerry rightly states, if the text does not fit on the slide in a large font, do not reduce the font. Instead, cut the number of words.

Tip 4: Limit the number of bullet points

The general rule is six by six—no more than six bullet points with no more than six words per bullet (Jerry suggests four to six bullets is ideal). Please note that this rule can be violated without destroying the presentation. But violate it sparingly. No one wants to sit through a text-heavy presentation. Also, while we are on the topic of bullet points, no one will remain interested in a presentation that has slide after slide after slide of bulleted lists. Mix things up by including graphics, word art charts, or some text without bullets. Your audience will thank you.

Tip 5: Avoid the trap of fancy builds and dimming

Essentially, this tip suggests that animation and slide transitions should be used only when necessary. For instance, having bullets appear on the slide as you speak about the item is useful for keeping the audience engaged and for moving the presentation forward. If you use this strategy, however, animate the full bullet (not one word or line at a time) and use it only a couple of times during the presentation. Repeating this animation sequence for every bulleted slide gets old fast. Jerry also suggests refraining from dimming sections of graphics or adding sound—such as drum rolls—for emphasis.

Tip 6: Don't rely on the spel chequer

As I like to say in class, Bill Gates's team did not know all of the technical terms or proper nouns you will need in your presentation. It is all too easy to develop the bad habit of ignoring the red underline below text. Also, a correctly spelled word may not be the right one to use.

Jerry's point about spelling is an important one and cannot be overstated. Typos are problematic. I usually recommend that absent a professional editor, the next best proofreader is someone from another line of work, who is unfamiliar with the project. A person with no knowledge of the project is likely to question content—their skepticism and "freshness" to the topic will be an asset. Also, know that you should ask a person to review the material only once, as they might be less sharp in a subsequent review. If a colleague is not available, ask a relative or friend who cares about you and will thus be more vigilant in looking for mistakes.

Tip 7: How to bore—include technical detail

Technical details like equations have a place but should never be used when presenting to a nontechnical audience. And even audiences at technical forums will lose interest in a talk with overly complex equations, graphics, or results. Keep it simple and leave the details in the accompanying report or technical appendix. For example, rather than showing a network metric in equation format, show a simple graphic illustrating the idea with a small network—you will be surprised what you can explain with a simple diagram of three nodes.

Tip 8: Maps and graphs speak volumes

Audiences watching a network presentation expect to see a network map, even if it is just a subgraph. But keep the list of essential elements of network visualization in mind. If the image is of an observed network, elements must be included to aid readability—at a minimum, report the number of nodes and ties and explain any important symbolization with a key.

Tip 9: Continuity across slides is essential

Design format is important. Jerry suggests presenters decide on a color scheme and font upfront (one for titles and one for text, if you want to use two fonts) and use them on all slides. Varying design elements like font, color, and styles throughout the presentation will annoy the audience. Consistency is good and so is brand. If your organization has a corporate style, use it.

Tip 10: Finish on your title slide or a black screen

For example, you could duplicate the title slide, replacing the details about the presentation date and forum with your contact details. This will remind people who have drifted off what your talk was about and inform them how to get a copy if they missed something important. Repeating the title slide is also a good habit for professional marketing.

DRIVERS

Now that you know what to do and how best to communicate your findings, following are some influential supernodes. Several research centers, practitioner-oriented programs, and professional associations are significantly advancing the field. In this section, some of these important drivers are profiled. Note that many offer resources and training in network analytics.

Academic Research Centers

Centre International de Criminologie Comparée

Université de Montréal, Canada

Founded in 1969, the center advances a multidisciplinary, global approach to understanding the processes by which to intervene and regulate criminal behavior. Advancing five lines of inquiry, the center aims to support a broad research agenda involving a micromacro continuum of analysis—individual offender, social control agencies, community and network configurations, state issues, and the digital society. Research under the Criminology, Networks and Community division appears to generate most, but not all, of the networked criminology supported by the center. Among the topics investigated by researchers are co-offending, gangs and organized crime, and collusion and corruption. Notable network-oriented researchers are Carlo Morselli, David Décary-Hétu, and Remi Boivin, from the École du Criminologie, Université de Montréal, Canada.

Center for Computational Analysis of Social & Organizational Systems

Carnegie Mellon University

Merging computer and social science, the Center for Computational Analysis of Social & Organizational Systems aims to develop a new scientific field investigating sociocultural behavior by developing metrics, technologies, and algorithms for mining large data; using agent-based model and network science to anticipate and explain behavioral change; and evaluating social and organizational policies and procedures. Through summer institutes, conferences, projects, hosting visiting scholars, and developing analytic tools and software, the center promotes new network science and agent-based modeling tools, many of which were designed for aiding intelligence operations. (For examples of work produced by this center, see Carley 2011; and Carley, Ju-Sung, and Krackhardt 2002).

International CyberCrime Research Centre

Simon Fraser University, Canada

Founded in 2008, the International CyberCrime Research Centre promotes research and education in cybercrime detection, response, and e-crime prevention. Launched in 2008 with joint support from the School of Criminology, the Minister of Labour and Citizen's Services (British Columbia), and the Society for the Policing of Cyberspace, the center is housed at Simon Fraser University's Surrey campus. In collaboration with public and private sectors, the center investigates a range of crime, including but not limited to cyberterrorism and online extremism, child exploitation, hacker forums, financial crime, organized crime, and critical infrastructure protection. Many of the studies initiated at the center use a networked computer science perspective. Notable researchers are Martin Bouchard and Richard Frank, School of Criminology, Simon Fraser University, Canada.

Transcrime, Joint Research Centre on Transnational Crime

Università Cattolica del Sacro Cuore, Alma Mater Studiorum Università di Bologna, and Università degli Studi di Perugia, Italy

Transcrime, under the leadership of Ernesto Ugo Savona, uses a multidisciplinary approach to analyzing crime, evaluating crime prevention policy, developing risk assessment models, and testing legislation for criminogenic effects (crime-proofing analysis). Notably, staff at Transcrime are particularly focused on organized crime, economic crime, money laundering, illegal markets, and urban crime. Francesco Calderoni, one of the faculty researchers, publishes widely using SNA to investigate organized crime.

Yale Institute for Network Science

Yale University

Founded in 2013, under the leadership of codirectors Dan Spielman and Nicholas Christakis, the Yale Institute for Network Science facilitates interdisciplinary research, supports education in network science, and promotes the application of network science through conferences, a weekly seminar series, and working papers.

Practitioner-Oriented Programs

Common Operational Research Environment Lab

Post Naval Graduate School, Monterey, CA

Founded in 2007, the Common Operational Research Environment Lab supports US and international field operatives by providing advanced training in geospatial, temporal, and social network analyses. Filling the gap between theory and operational capacity, it engages in training and research. Recent projects include methodologies to exploit social media to map dark networks, developing web applications to mine and analyze twitter feeds and open sources in support of Military Information Support Operations analysts. Sean Everton, codirector of the Lab and faculty of the Department of Defense Analysis, published an instructional book on using SNA to support strategies to disrupt dark networks (Everton 2012).

DHS Center of Excellence for Criminal Investigations and Network Analysis

George Mason University

Over the next ten years, the Department of Homeland Security, Science and Technology Directorate, will invest forty million dollars on a consortium of programs led by a team from George Mason University. Integrating criminology, engineering, and mathematics, the center COE is mandated to use an interdisciplinary approach to engage in research to enhance investigation of homeland security–related crimes, including transnational criminal organizations' activities and cybercrime. The center develops knowledge products, analytic tools, and technology-based solutions for agents, police officers, and investigators, so they can better predict, prevent, and prosecute homeland security–related crimes.

National Network for Safe Communities

John Jay College of Criminal Justice

Founded in 2009, the National Network for Safe Communities works to develop and promote strategic interventions to reduce violence, strengthen communities, and reduce incarceration, while also improving relationships between law enforcement and communities. It promotes SNA by making the work of Andrew Papachristos accessible. As part of the Group Violence Initiative, an SNA software and sample data set are available to examine violence for the purposes of crime prevention. It is available at nnscommunities.org/our-work/innovation/social-network-analysis.

Professional Associations and Training Opportunities

Duke Network Analysis Center

Duke University

The center seeks to promote network science and analysis in the social, physical, and biological sciences. Among its stated purposes is to train new researchers and promote collaborations in network science. Toward this aim, the center hosts seminars and training workshops. Its web page also links to assorted resources, including data set repositories, tutorials, software, and other notable groups and associations dedicated to network science.

Illicit Networks Workshop

The Illicit Networks Workshop and its partner, the Policing Flows Workshop, annually brings network scholars together to discuss current research. Frequent sponsors of this workshop are (1) Andrew Goldsmith and Russell Brewer, through the Crime and Security Research Centre, housed in the Law and Criminology Program, Flinders University, Australia; and Carlo Morselli, through the Centre international de Criminologie Comparée, Université de Montréal, Canada. Hosts and meeting agendas evolve, but three constants remain—the meeting moves to different parts of the world each year to make the event accessible to practitioners, great effort is extended to encourage practitioners to participate or attend the workshop, and the materials produced through the workshop are made accessible to the public via journal special issues and edited books. Among the products springing from the workshop are the following

Journal Special Issues

- Martin Bouchard, ed., special issue, *Global Crime* 14, nos. 2–3 (2013).
- Aili Malm and Gisela Bichler, eds., special issue, *Journal of Contemporary Criminal Justice* 31, no. 3 (2015).
- Carlo Morselli and Remi Boivin, special issue, *Social Networks* 51 (2017).

Edited Books

- C. Morselli, *Crime and Networks* (New York: Routledge, 2014).
- M. Bouchard, *Advances in Research on Illicit Networks* (New York: Routledge, 2015).
- G. Bichler and A. Malm, *Disrupting Criminal Networks: Network Analysis in Crime Prevention*, Crime Prevention Studies 28 (Boulder, CO: FirstForumPress, 2015b).

International Network for Social Network Analysis

The International Network for Social Network Analysis is an interdisciplinary, multidisciplinary professional association made up of researchers interested in SNA. It advances social network research through its publications (*Connections* and the *Journal of Social Structure*), its support of a network journal (*Social Networks*), hosting of an annual international conference (Sunbelt conference), and its maintenance of electronic services (e.g., SOCNET, an electronic discussion forum, and REDES, a Spanish-language listserv). Before the annual meeting, the organization organizes a series of workshops, many of which offer training in specific software or analytic techniques.

LINKS Center

University of Kentucky housed in the Gatton College of Business and Economics, the LINKS Center for Social Network Analysis engages in research, runs summer training programs, offers consulting services, and runs conferences in order to promote the study and optimization of social networks in organizations. While the center does not focus on criminal matters, you may find the summer workshop series is worth the fee.

Admittedly, the preceding list of drivers is not exhaustive. Many more individuals and organizations are involved in developing networked criminology and advancing knowledge of criminal network structure and topography, testing the applicability of SNA techniques and pushing the discipline from descriptive to predictive modeling, and holding the field accountable for methodological limitations and gaps in the empirical foundation of this emerging discipline. For instance, while this book was in production Martin Bouchard launched CaINLab (twitter.com/thecainlab) and Andrew Papachristos set-up the Northwestern Network and Neighborhood Initiative (N^3) (twitter.com/n3initiative).

CONCLUDING REMARKS

In an effort to assist your dive into studying criminal networks, this concluding chapter provided the following: (1) a few publishing tips to supplement the reporting standards supplied in earlier chapters; (2) an examination of the importance of visualization in reporting on investigations of criminal networks, including various techniques for making images and graphics readable and accessible; and (3) a look toward the future, focusing on the crucial role played by centers of scientific enterprise and research collaborations in the diffusion of networked criminology. With respect to the latter, the chapter reviewed some of these critical drivers.

Our conversation about criminal networks is drawing to a close. If you are reading this, I suspect that our paths will cross at some future conference, workshop, or meeting. Perhaps you have become a network convert, and, if so, I predict that you may never look at data the same way again. I hope you found this book useful and that you now have the tools and confidence to embark on your own investigation of criminal networks. Thank you for reading what I have to say and best of luck with your next inquiry.

Before I sign off, I have one final caveat to share. Despite my best efforts, in a book of this scope, you will have undoubtedly come across an omission or something that is unclear. If you think of it, please drop me a line. Reference the issue and page number and I will do my best to fix it in the next edition (gbichler@csusb.edu: subject line “Understanding Criminal Networks Correction”).

Cheers.

References

Aral, S. 2011. "Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion." *Marketing Science* 30 (2): 217–23.

Baker, W.E., and R.R. Faulkner. 1993. "The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry." *American Sociological Review* 58:837–60.

Barabási, A.-L., and R. Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–12.

Bastian M., S. Heymann, and M. Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Weblogs and Social Media.

Batagelj, V., and A. Mrvar. 1998. "Pajek Program for Large Network Analysis." *Connections* 21 (2): 47–57.

Benford, F. 1938. "The Law of Anomalous Numbers." *Proceedings of the American Philosophical Society* 78:551–72.

Berlusconi, G. 2013. "Do All the Pieces Matter? Assessing the Reliability of Law Enforcement Data Sources for the Network Analysis of Wire Taps." *Global Crime* 14 (1): 61–81.

Bichler, G., and D.J. Birks. 2015. "Exposed: Agent-Based Simulation of Military Weapons Transfers." *ResearchGate* (May). https//doi.org/10.13140/RG.2.1.1630.9927.

Bichler, G., D.J. Birks, J. Van de Merghel., J. Hagala., and B. Rios. 2015. "The X Factor: Validating Agent-Based Models of Transnational Crime." Presented at the Seventh Annual Illicit Networks Workshop, Montreal, Canada, November.

Bichler, G., and S. Bush. 2016. "Staying Alive in the Business of Terror." In *Disaster Forensics: Understanding Root Cause and Complex Causality,* edited by Anthony J. Masys, 271–94. New York: Springer.

Bichler, G., S. Bush, and J. Hagala. 2014. *Palm Springs Police Department Resource Constraint Study, Final Report.* Technical Report. Prepared for

Chief Alberto Franz and Captain Dennis Graham, Palm Springs Police Department, City of Palm Springs, September.

Bichler, G., S. Bush, and A. Malm. 2013. "Bad Actors and Faulty Props: Unlocking Legal and Illicit Art Trade." *Global Crime* 14 (4): 359–85.

———. 2015. "Regulatory Foresight: Estimating Policy Effects on Transnational Illicit Markets." *Journal of Contemporary Criminal Justice* 31 (3): 297–318.

Bichler, G., and J. Franquez. 2014. "Conflict Cessation and the Emergence of Weapons Supermarkets." In *Networks and Network Analysis for Defence and Security,* edited by Anthony Masys, 189–215. New York: Springer.

Bichler, G., S. Lim, and E. Larin. 2017. "Tactical SNA: Using Affiliation Networks to Aid Serial Crime Investigation." *Homicide Studies* 21 (2): 133–58. First published September 27, 2016. https//doi.org/ 10.1177/1088767916671351.

Bichler, G., and A. Malm. 2015a. *Disrupting Criminal Networks: Network Analysis in Crime Prevention.* Crime Prevention Studies 28. Boulder, CO: FirstForumPress.

———. 2015b. "The Routine Nature of Transnational Crime." In *The Criminal Act: The Role and Influence of Routine Activity Theory,* edited by Martin A. Andresen and Graham Farrell, 33–58. Basingstoke, UK: Palgrave Macmillan.

Bichler, G., A. Malm, and J. Christie-Merrall. 2011. "Urban Backcloth and Regional Mobility Patterns as Indicators of Juvenile Crime." In *Patterns, Prevention, and Geometry of Crime,* edited by Martin A. Andresen and J. Bryan Kinney, 118–36. Routledge Crime Science Series. New York: Routledge.

Bichler, G., A. Malm, and T. Cooper. 2017. "Drug Supply Networks: A Systematic Review of the Organizational Structure of Illicit Drug Trade." *Crime Science* 6 (2): 1–23.

Bichler, G., A. Malm, and J. Enriquez. 2014. "Magnetic Facilities: Identifying Convergence Settings of Juvenile Delinquents." *Crime and Delinquency* 60 (7): 971–98. First published October 1, 2010. https//doi.org/10.1177/0011128710382349.

Bichler, G., A. Norris, J. R. Dmello, and J. Randle. 2019. "The Impact of Civil Gang Injunctions on Networked Violence between the Bloods and the Crips." *Crime and Delinquency* 65(7): 875-915. https//doi.org/10.1177/0011128717739607.

Borgatti, S. P., M. G. Everett, and L. C. Freeman. 2002. *Ucinet for Windows: Software for Social Network Analysis.* Harvard, MA: Analytic Technologies.

Borgatti, S. P., M. G. Everett, and C. J. Johnson. 2013. *Analyzing Social Networks.* Thousand Oaks, CA: Sage.

Borgatti, S. P., and V. Lopez-Kidwell. 2011. "Network Theory." In *The SAGE Handbook of Social Network Analysis,* edited by J. Scott and P. J. Carrington, 40–54. Thousand Oaks, CA: SAGE.

Boston Police Department. 1998. "Operation Ceasefire: Deterring Youth Firearm Violence." Submission to the Herman Goldstein Award for Excellence in Problem-Oriented Policing. www.popcenter.org/library/awards/goldstein/1998/98–08(W).pdf.

Bouchard, M. 2015. *Advances in Research on Illicit Networks.* New York: Routledge.

Braga, A.A. 2008. "Pulling Levers Focused Deterrence Strategies and the Prevention of Gun Homicide." *Journal of Criminal Justice* 36: 332–43.

Braga, A.A., R. Apel, and B.C. Welsh. 2013. "The Spillover Effects of Focused Deterrence on Gang Violence." *Evaluation Review* 37:314–42.

Braga, A.A., D.M. Hureau, and A.V. Papachristos. 2014. "Deterring Gang-Involved Gun Violence: Measuring the Impact of Boston's Operation Ceasefire on Street Gang Behavior." *Journal of Quantitative Criminology* 30 (1): 113–39.

Brantingham, P.J., and P.L. Brantingham. 1984. *Patterns in Crime.* New York: Macmillan.

———. 2008. "Crime Pattern Theory." In *Environmental Criminology and Crime Analysis*, edited by R. Wortley and L. Mazerolle, 78–93. Cullompton, UK: Willan.

Brantingham, P.J., P.L. Brantingham, and M.A. Andresen. 2017. "The Geometry of Crime and Crime Pattern Theory." In *Environmental Criminology and Crime Analysis*, edited by R. Wortley and M. Townsley, 98–115. 2nd ed. New York: Routledge.

Brantingham, P.L., and P.J. Brantingham. 2015. "Understanding Crime with Computational Criminology." In *The Criminal Act: The Role and Influence of Routine Activity Theory*, edited by Martin A. Andresen and Graham Farrell, 131–45. Basingstoke, UK: Palgrave Macmillan.

Bright, D.A., and J.J. Delaney. 2013. "Evolution of a Drug Trafficking Network: Mapping Changes in Network Structure and Function across Time. *Global Crime* 14:238–60.

Bright, D.A., C. Greenhill, M. Reynolds, A. Ritter, and C. Morselli. 2014. "The Use of Actor-Level Attributes and Centrality Measures to Identify Key Actors: A Case Study of an Australian Drug Trafficking Network." *Journal of Contemporary Criminal Justice* 31 (3): 262–78.

Bright, D.A., C.E. Hughes, and J. Chalmers. 2012. "Illuminating Dark Networks: A Social Network Analysis of an Australian Drug Trafficking Syndicate." *Crime, Law and Social Change* 57 (2): 151–76.

Burcher, M., and C. Whelan. 2018. "Social Network Analysis as a Tool for Criminal Intelligence: Understanding Its Potential from the Perspectives of Intelligence Analysts." *Trends in Organized Crime* 21 (3): 278–94.

Burt, R.S. 1992. *Structural Holes: The Social Structure of Competition.* Cambridge, MA: Harvard University Press.

Calderoni, F. 2015. "Network Analysis of Organized Crime Meetings as a Tool for the Identification of Leadership Positions." In *Disrupting Criminal Networks: Network Analysis in Crime Prevention*, edited by G. Bichler and A. Malm, 89–110. Crime Prevention Studies 28. Boulder, CO: First Forum Press.

Campana, P. 2011. "Eavesdropping on the Mob: The Functional Diversification of the Mafia Activities across Territories." *European Journal of Criminology* 8 (3): 1–16.

Campana, P., and F. Varese. 2012. "Listening to the Wire: Criteria and Techniques for the Quantitative Analysis of Phone Intercepts." *Trends in Organized Crime* 15:13–30.

Carley, K.M. 2001–2018. ORA. Center for Computational Analysis of Social and Organizational Systems (CASOS), Institute for Software Research International (ISRI), School of Computer Science, Carnegie Mellon University.

———. 2011. *Organizational Risk Analyzer (ORA).* Pittsburgh: Center for Computational Analysis of Social and Organizational Systems (CASOS), Carnegie Mellon University.

Carley, K.M., L. Ju-Sung, and D. Krackhardt. 2002. "Destabilizing Networks." *Connections* 24 (3): 79–92.

Carrington, P.J. 2011. "Crime and Social Network Analysis." In *The SAGE Handbook of Social Network Anal*ysis, edited by J. Scott and P.J. Carrington, 236–55. Thousand Oaks, CA: SAGE.

Center for Problem-Oriented Policing. n.d. Watts College of Public Service and Community Solutions, Arizona State University (website). Accessed 2/21/2019.

Centola, D. 2018. *How Behavior Spreads: The Science of Complex Contagions.* Vol 3. Princeton, NJ: Princeton University Press. https://popcenter.asu.edu/about/?p=triangle.

Christakis, N.A. 2010a. "The Hidden Influence of Social Networks." Filmed February 2010, in New York, NY. TED Video, 18:07. www.ted.com/talks/nicholas_christakis_the_hidden_influence_of_social_networks.

———. 2010b. "How Social Networks Predict Epidemics." Filmed June 2010, in Cannes, France. TED Video, 17:48. https://www.ted.com/talks/nicholas_christakis_how_social_networks_predict_epidemics.

Christakis, N.A., and J.H. Fowler. 2009. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives.* New York: Little, Brown.

Christen, P. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* New York: Springer.

Cincinnati Police Department. 2017. "P.I.V.O.T.: Place-Based Investigations of Violent Offender Territories." Submission to the Herman Goldstein Award for Excellence in Problem-Oriented Policing.

Clarke, R.V. 1999. "Hot Products: Understanding, Anticipating and Reducing Demand for Stolen Goods." Police Research Series Paper 112, British Home Office Research Publications, London.

———. 2008. "Situational Crime Prevention." In *Environmental Criminology and Crime Analysi*s, edited by R. Wortley and L. Mazerolle, 178–94. Cullompton, UK: Willan.

Clarke, R.V., and J. Eck. 2003. *Become a Problem-Solving Crime Analyst in 55 Small Steps.* London, UK: Ashford Colour Press.

Cohen, L.E., and M. Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 4:588–608.

Cohen-Cole, E., and J.M. Fletcher. 2008. “Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic.” *Journal of Health Economics* 27 (5): 1382–87.

Combatting Terrorism Center. 2012. Letter SOCOM-2012–0000003-HT. In *Letters from Abbottabad: Bin Laden Sidelined?* West Point, NY: Westpoint.

Cornish, D. 1994. “The Procedural Analysis of Offending and Its Relevance for Situational Prevention.” In *Crime Prevention Studies,* edited by R. Clarke, 151–96. Monsey, NY: Criminal Justice Press, 1994.

Cornish, D.B., and R.V. Clarke. 2016. “The Rational Choice Perspective.” In *Environmental Criminology and Crime Analysis,* edited by R. Wortley and M. Townsley, 29–61. London: Routledge.

Décary-Hétu, D., and B. Dupont. 2012. “The Social Network of Hackers.” *Global Crime* 13 (3): 160–75.

Dickinson, J.R. 2014. “A Mathematical Law for Assessing Outcome Values of Games.” *Simulation and Gaming* 45 (3): 318–31.

Dodds, P.S., R. Muhamad, and D.J. Watts. 2003. “An Experimental Study of Search in Global Social Networks.” *Science* 301 (8): 827–29.

Duijn, P.A.C., V. Kashirin, and P.M.A. Sloot. 2014. “The Relative Ineffectiveness of Criminal Network Disruption.” *Scientific Reports* 4 (4238): 1–15.

Eck, J.E. 1995. “A General Model of the Geography of Illicit Retail Market Places. In *Crime and Place,* edited by J.E. Eck and D. Weisburd, 67–94. Crime Prevention Studies. Monsey, NY: Criminal Justice Press.

———. 2002. “Preventing Crime at Places.” In *Evidence-Based Crime Prevention,* edited by L.W. Sherman, D.P. Farrington, B. Welsh, and D.L. MacKenzie, 241–94. New York: Routledge.

Eck, J.E., and E.B. Eck. 2012. “Crime Place and Pollution: Expanding Crime Reduction Options through a Regulatory Approach.” *Criminology and Public Policy* 11:281–317.

Eck, J.E., and W. Spelman. 1987. *Problem Solving: Problem-Oriented Policing in Newport News.* Washington, DC: Police Executive Research Forum.

Ennett, S.T., V.A. Foshee, K.E. Bauman, A. Hussong, R. Faris, J. Hipp, L. Cai, R. DeRant, and H.L. McNaughton Reyes. 2008. “The Social Ecology of Adolescent Alcohol Misuse.” *Child Development* 79 (6): 1771–91.

Everton, S.F. 2012. *Disrupting Dark Networks.* New York: Cambridge University Press.

Farrell, G., and J. Roman. 2006. “Crime as Pollution: Proposal for Market-Based Incentives to Reduce Crime Externalities.” In *Crime Reduction and the Law,* edited by K. Moss and M. Stephens, 135–55. Abingdon, UK: Routledge.

Felson, M. 1995. “Those Who Discourage Crime. In *Crime and Place,* edited by J.E. Eck and D. Weisburd, 53–66. Crime Prevention Studies 4. Monsey, NY: Criminal Justice Press.

———. 2003. “The Process of Co-offending.” In *Theory for Practice in Situational Crime Prevention,* edited by Martha J. Smith and Derek B. Cornish, 149–68. Crime Prevention Studies 16. Monsey, NY: Criminal Justice Press.

———. 2006. *Crime and Nature.* Thousand Oaks, CA: Sage.

Felson, M., and R.V. Clarke. 1998. "Opportunity Makes a Thief." Police Research Series Paper 98, British Home Office Research Publications, London.

Felson, M., and L.E. Cohen. 1980. "Human Ecology and Crime: A Routine Activity Approach." *Human Ecology* 8 (4): 389–406.

Fowler J.H., and N.A. Christakis. 2008. "Estimating Peer Effects on Health in Social Networks: A Response to Cohen-Cole and Fletcher; Trogdon, Nonnemaker, and Pais." *Journal of Health Economics* 27 (5): 1400–1405.

Freeman, L.C. 1977. "A Set of Measures of Centrality Based on Betweenness." *Sociometry* 4 (1): 35–41.

———. 1992. "Social Networks and the Structure Experiment." In *Research Methods in Social Network Analysis,* edited by Linton C. Freeman, Douglas R. White, and A. Kimball Romney, 11–40. New Brunswick, NJ: Transaction.

Gaviria, Marcella, dir. 2016. *Chasing Heroin.* Season 4, episode 6, of *Frontline*. Aired February 23, on PBS. https://www.pbs.org/wgbh/frontline/film/chasing-heroin/transcript/.

Goldsmith, A., and R. Brewer. 2015. "Digital Drift and the Criminal Interaction Order." *Theoretical Criminology* 19 (1): 112–30.

Goldstein, H. 1979. "Improving Policing: A Problem-Oriented Approach." *Crime and Delinquency* 25 (2): 236–58.

Granovetter, M. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–80.

———. 1983. "The Strength of Weak Ties: A Network Theory Revisited." *Sociological Theory* 1:201–33.

Green, B., T. Horel, and A.V. Papachristos. 2017. "Modeling Contagion through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014." *JAMA Internal Medicine* 177 (3): 326–33. https//doi.org/10.1001/jamainternmed.2016.8245.

Gunnel, S., and K. Todter. 2009. "Does Benford's Law Hold in Economic Research and Forecasting?" *Empirica: Journal of Applied Economics and Economic Policy* 36 (3): 273.

Hashimi, S., and M. Bouchard. 2017. "On to the Next One? Using Social Network Data to Inform Police Target Prioritization." *Policing: An International Journal of Police Strategies and Management* 40:768–82.

Haynie, D.L. 2001. "Delinquent Peers Revisited: Does Network Structure Matter?" *American Journal of Sociology* 106:1013–57.

Haynie, D.L., N.J. Doogan, and B. Soller. 2014. "Gender, Friendship Networks, and Delinquency: A Dynamic Network Approach." *Criminology* 52 (4): 688–722.

Heider, F. 1958. *The Psychology of Interpersonal Relations.* New York: Wiley.

Hipp, J.R., and A. Boessen. 2013. "Egohoods as Waves Washing across the City: A New Measure of 'Neighborhoods.'" *Criminology* 51 (2): 287–327.

Hipp, J.R., R.W. Faris, and A. Boessen. 2012. "Measuring 'Neighborhood': Constructing Network Neighborhoods." *Social Networks* 34 (1): 128–40.

Holland, P.W., and S. Leinhardt. 1970. "A Method for Detecting Structure in Sociometric Data." *American Journal of Sociology* 70:492–513.

Kadushin, C. 2012. *Understanding Social Networks: Theories, Concepts, and Findings*. New York: Oxford University Press.

Kennedy, D. 1997. "Pulling Levers: Chronic Offenders, High-Crime Settings, and a Theory of Prevention." *Valparaiso University Law Review* 31:449–84.

Knoke, D., and S. Yang. 2008. *Social Networks Analysis*. 2nd ed. Quantitative Applications in the Social Sciences Series 154. Thousand Oaks, CA: SAGE.

Kossinets, G. 2006. "Effects of Missing Data in Social Networks." *Social Networks* 28 (3):247–68.

Krebs, V.E. 2002. "Mapping Networks of Terrorist Cells." *Connections* 24 (3): 43–52.

Long, J.C., F.C. Cunningham, and J. Braithwaite. 2013. "Bridges, Brokers and Boundary Spanners in Collaborative Networks: A Systematic Review." *BMC Health Services Research* 13 (158): 1–13.

Lusher, D., J. Koskinen, and G. Robins. 2013. *Exponential Random Graph Models for Social Sciences: Theory, Methods, and Applications*. New York: Cambridge University Press.

Madensen, T. 2007. "Bar Management and Crime: Toward a Dynamic Theory of Place Management and Crime Hotspots." PhD diss., University of Cincinnati.

Madensen, T., and J.E. Eck. 2008. "Violence in Bars: Exploring the Impact of Place Manager Decision-Making." *Crime Prevention and Community Safety* 10:111–25.

Madensen, T., M. Herold, M.G. Hammer, and B.R. Christenson. 2017. "Place-Based Investigations to Disrupt Crime Place Networks." *Police Chief*, April 14–15.

Madensen, T., and W. Sousa. 2008. "Integrating Place Managers and Broken Windows Theory." Presented at the Seventeenth International Symposium on Environmental Criminology and Crime Analysis, Anchorage, Alaska, July.

Malm, A., and G. Bichler. 2011. "Networks of Collaborating Criminals: Assessing the Structural Vulnerability of Drug Markets." *Journal of Research in Crime and Delinquency* 48 (2): 271–97.

———. 2013. "Using Friends for Money: The Positional Importance of Money-Launderers in Organized Crime." *Trends in Organized Crime* 16 (4): 365–81.

Malm, A., G. Bichler, and S. Van De Walle. 2010. "Comparing Ties That Bind Criminal Networks: Is Blood Thicker Than Water?" *Security Journal* 23: 52–74.

McGLoin, J.M. 2005. "Policy and Intervention Considerations of a Network Analysis of Street Gangs." *Criminology and Public Policy* 4 (3): 607—35.

McGloin, J.M., and A.R. Piquero. 2010. "On the Relationship between Co-offending Network Redundancy and Offending Versatility." *Journal of Research in Crime and Delinquency* 47 (1): 63–90.

Medina, R., and G. Hepner. 2008. "Geospatial Analysis of Dynamic Terrorist Networks." In *Values and Violence: Intangible Aspects of Terrorism*, edited by I. Karawan, W. McCormack, and S. E. Reynolds, 151–67. Berlin: Springer.

Milgram, S. 1967. "The Small-World Problem." *Psychology Today* 1 (1): 61–67.

Morselli, C. 2009. "Hells Angels in Springtime." *Trends in Organized Crime* 12 145–58.

———. 2014. *Crime and Networks*. New York: Routledge.

Morselli, C., and J. Roy. 2008. "Brokerage Qualifications in Ringing Operations." *Criminology* 46:71–98.

Morselli, C., P. Tremblay, and B. McCarthy. 2006. "Mentors and Criminal Achievement." *Criminology* 44 (1): 17—43.

Natarajan, M. 2000. "Understanding the Structure of a Drug Trafficking Organization: A Conversational Analysis." In *Illegal Drug Markets: From Research to Policy*, edited by M. Natarajan and M. Hough, 273–98. Crime Prevention Studies 11. Monsey, NY: Criminal Justice Press.

———. 2006. "Understanding the Structure of a Large Heroin Distribution Network: A Quantitative Analysis of Qualitative Data." *Journal of Quantitative Criminology* 22:171–92.

Ouellet, M., M. Bouchard, and M. Hart. 2017. "Criminal Collaboration and Risk: The Drivers of Al Qaeda's Network Structure before and after 9/11." *Social Networks* 51:171–77.

Papachristos, A.V. 2009. "Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide." *American Journal of Sociology* 115 (1): 74—128.

———. 2011. "The Coming of a Networked Criminology?" In *Measuring Crime and Criminality: Advances in Criminological Theory*, edited by J. MacDonald, 101–40. Vol. 17. New Brunswick, NJ: Transaction.

———. 2013. "The Importance of Cohesion for Gang Research, Policy, and Practice." *Criminology and Public Policy* 12 (1): 49–58.

Papachristos, A. V., A. A. Braga, and D. Hureau. 2012. "Social Networks and the Risk of Gunshot Injury." *Journal of Urban Health* 89:922–1003. https//doi .org/10.1007/s11524–012–9703–9.

Papachristos, A. V., A. A. Braga, E. Piza, and L. S. Grossman. 2015. "The Company You Keep? The Spillover Effects of Gang Membership on Individual Gunshot Victimization in a Co-offending Network." *Criminology* 53 (4): 624–49.

Papachristos, A. V., D. M. Hureau, and A. A. Braga. 2013. "The Corner and the Crew: The Influence of Geography and Social Networks on Gang Violence." *American Sociological Review* 78 (3): 1–31.

Papachristos, A. V., and C. M. Smith. 2014. "The Embedded and Multiplex Nature of Al Capone." In *Crime and Networks*, edited by C. Morselli, 97–115. New York: Routledge.

Radil, S. M., C. Flint, and G. E. Tita. 2010. "Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles." *Annals of the Association of American Geographers* 100 (2): 307–26.

Randle, J., and G. Bichler. 2017. "Uncovering the Social Pecking Order in Gang Violence." In *Crime Prevention in the 21st Century,* edited by Benoit Leclerc and Ernesto Savona, 165–86. New York: Springer.

Ratcliff, J. 2014. "Jerry's Top Ten PowerPoint Tips." *Evidence-Based Policing, Criminal Intelligence, and Crime Science* (blog). October 20, 2014. www.jratcliffe.net/blog/powerpoint-tips/.

Reynald, D.M. 2010. "Guardians on Guardianship: Factors Affecting the Willingness to Supervise, the Ability to Detect Potential Offenders, and the Willingness to Intervene." *Journal of Research in Crime and Delinquency* 47 (3): 358–90.

Ripley, R., T. Snijders, Z. Boda, A. Vörös, and P. Preciado. 2018. *Manual for SIENA Fersion 4.0* (May 7). Oxford: University of Oxford, Nuffield College, Department of Statistics.

Robins, G., P. Pattison, and J. Woolcock. 2005. "Small and Other Worlds: Global Network Structures from Local Processes. *Journal of Sociology* 110 (4): 894–36.

Rossmo, D.K. 2000. *Geographic Profiling.* New York: CRC Press.

Sageman, M. 2004. *Understanding Terror Networks.* Philadelphia: University of Pennsylvania Press.

Sampson, R., J.E. Eck, and J. Dunham. 2010. "Super Controllers and Crime Prevention: A Routine Activity Explanation of Crime Prevention Success and Failure." *Security Journal* 23 (1): 37–51.

Sarnecki, J. 2001. *Delinquent Networks: Youth Co-offending in Stockholm.* Cambridge: Cambridge University Press.

Schaefer, D.R. 2012. "Youth Co-offending Networks: An Investigation of Social and Spatial Effects." *Social Networks* 34:141–49.

Schwartz, D.M., and T. Rousselle. 2009. "Using Social Network Analysis to Target Criminal Networks." *Trends in Organized Crime* 12 (2): 188–207.

Shukla, R. 2016. *Methamphetamine: A Love Story.* Oakland, CA: University of California Press.

Sparrow, M.K. 1991. "The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects." *Social Networks* 13:251–74.

Sutherland, E.H. 1947. *Principles of Criminology.* 4th ed. Chicago: J.B. Lippincott.

Systems Academy. 2015a. "Decentralized and Small World Networks." April 27. Video, 5:51. www.youtube.com/watch?v=QUWds9gt6aE.

———. 2015b. "Network Centrality." April 20. Video, 5:29. www.youtube.com/watch?v=NgUj8DEH5Tc.

———. 2015c. "Network Clustering and Effectiveness." April 24. Video, 4:02. www.youtube.com/watch?v=2Oa7mef77nM.

———. 2015d. "Network Degree Distribution. April 25. Video, 4:19. www.youtube.com/watch?v=0CHTuTq_EOI.

———. 2015e. "Network Theory Topology. April 21. Video, 5:38. www.youtube.com/watch?v=bWaCB8IEgEY.

Tenti, V., and C. Morselli. 2014. "Group Co-offending Networks in Italy's Illegal Drug Trade." *Crime, Law and Social Change* 6 (1): 21–44.

Tita, G., and S. Radil. 2011. "Spatializing the Social Networks of Gangs to Explore Patterns of Violence." *Journal of Quantitative Criminology* 27 (4): 521–45.

Travers, J., and S. Milgram. 1969. "An Experimental Study of the Small World Problem." *Sociometry* 32 (4): 425–43.

Tremblay, P. 1993. "Searching for Suitable Co-offenders." In *Routine Activity and Rational Choice,* edited by R.V. Clarke and M. Felson, 17–36. Vol. 5. New Brunswick, NJ: Transaction.

Uzzi, B. 1996. "The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect." *American Sociological Review* 61:674–98.

Uzzi, B., and J. Spiro. 2005. "Collaboration and Creativity: The Small World Problem." *American Journal of Sociology* 111:447–504.

Van Mastrigt, S.B., and P.J. Carrington. 2014. "Sex and Age Homophily in Co-offending Networks: Opportunity or Preferences?" In *Crime and Networks,* edited by C. Morselli, 28–51. New York: Routledge.

Varese, F. 2011. *Mafias on the Move.* Princeton, NJ: Princeton University Press.

Wang, P., G. Robins, and P. Pattison. 2009. *PNet: Program for the Simulation and Estimation of Exponential Random Graph Models.* Melbourne School of Psychological Sciences, University of Melbourne.

Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

Watts, D.J. 2003. *Six Degrees: The Science of a Connected Age.* New York: W.W. Norton.

Watts, D., and S. Strogatz. 1998. "Collective Dynamics of Small-World Networks." *Nature* 393:440–42.

Wellman, B. 1988. "Structural Analysis: From Method and Metaphor to Theory and Substance." In *Social Structures: A Network Approach,* edited by B. Wellman and S.D. Berkowitz, 15–61. New York: University of Cambridge.

Westlake, B., M. Bouchard, and R. Frank. 2011. "Finding the Key Players in Online Child Exploitation Networks." *Policy and Internet* (May): 1–32.

Wilensky, U. 1999. NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wortley, R. 2008. "Situational Precipitators of Crime." In *Environmental Criminology and Crime Analysis,* edited by R. Wortley and L. Mazerolle, 48–69. Cullompton, UK: Willan.

———. 2016. "Situational Precipitators of Crime." In *Environmental Criminology and Crime Analysis,* edited by R. Wortley and M. Townsley, 62–86. 2nd ed. London: Routledge.

Xu, J., and H. Chen. 2008. "The Topology of Dark Networks." *Communications of the ACM* 51:58–65.

Young, J.T.N. 2011. "How Do They 'End Up Together'? A Social Network Analysis of Self-Control, Homophily, and Adolescent Relationships." *Journal of Quantitative Criminology* 27 (3): 251–73.

Index

Founded in 1893,

UNIVERSITY OF CALIFORNIA PRESS
publishes bold, progressive books and journals on topics in the arts, humanities, social sciences, and natural sciences—with a focus on social justice issues—that inspire thought and action among readers worldwide.

The UC PRESS FOUNDATION
raises funds to uphold the press's vital role as an independent, nonprofit publisher, and receives philanthropic support from a wide range of individuals and institutions—and from committed readers like you. To learn more, visit ucpress.edu/supportus.