1 Introduction to Social Networks Research

Imagine someone tells you that the government has become dangerous, and you need to flee. If that someone were a total stranger, would you believe the news? Would you seek a second opinion? Would you leave immediately? If that someone were your friend, would your response change? What about someone who wasn't a friend per se, but a coworker, or attends your place of worship? Would your reaction be identical, or would you be more inclined to believe this surprising news or act on it if it came from certain sources? Now imagine the source is not your friend, but is a friend of a friend, or a friend of a friend of a friend. Would it make a difference if this person was a friend not only of yours but also of five of your other friends? Now, what if the source was the person who was most popular in your town – would that status bear on whether you believe that the government was truly dangerous or found this information worthy of action?

Now picture a completely different scenario: someone just won ten million dollars. How likely do you think it is that you would receive any of that money if the person who came into the windfall were a stranger? If instead they were your friend, or a friend of a friend of a friend, or the most popular person in your town, would your likelihood of receiving any of the winnings change?

In both scenarios, the possible sources of information and money are different kinds of personal connections. The sources vary in the way they are connected to you and the way they are connected to others. The idea that connections can matter in scenarios like these, and that exactly how they matter can depend on both the type and the arrangement of connections, is the principal motivation of social networks research.

Studying Social Networks

Social networks research is premised on the idea that connections between actors can matter for things we care about. An empirical social networks research project aims to document this idea and to better understand it in the real world. The key unit of analysis in these studies is a social network, which is a set of relationships among a group of actors.

1

The foundational step in any networks project is to precisely specify the two constituent parts of a network, the group of actors (called "nodes" or "vertices") and the relationships (called "links" or "edges"). Both require careful thought. Typically, the actors of interest that will serve as a network's nodes are individuals, for instance citizens, students, politicians, or voters. Although they are often unitary actors, they need not be; nodes could instead be aggregates such as committees, organizations, households, or countries.

Even once a set of nodes is selected, the researcher faces a large menu of options for the relationships that will serve as a network's links. After all, the opening scenarios make clear that there is a wide variety of ways in which people can be connected. Friends, acquaintances, coworkers, teammates, relatives, and fans of the same band are all interpersonal relationships of some form.

The choice of which links to focus on can be daunting, especially considering how consequential the choice will be for a study's conclusions. Moreover, links need to be defined with sufficient precision in order to properly guide data collection. This book helps researchers arrive at a useful definition of relevant links by using a framework that separately identifies a link's type and its function. Table 1.1 previews this framework by showing some questions that can steer a researcher toward the appropriate nodes and links for a study.

Table 1.1 Framework for defining a network with sufficient precision and relevance. Topic of Chapter 4.

Attribute	Question to Ask	Examples
Nodes	What is the set of things that are connected?	Villagers; adult females; members of an organization; voters; committees; rebel groups; countries
Link Type	What is the relevant connection between two nodes?	Friendship; trust; shared geographic space; experience as colleagues; kinship; overlapping membership; past interactions
Link Function	What does a link between two nodes do? Why is the presence of such a link important?	Transmits information; spreads disease; exposes one to others' opinions; conveys peer pressure; aligns interests; shares resources

The idea that networks may matter hinges on the notion that social connections – links – may *do* something. If I am connected to the person sharing news of government danger by friendship, then our link type is friendship. But whether this link means I am likely to pick up my belongings and flee in response depends in part on what a friendship link might *do*. A useful way to think about this is to see a link as a potential channel through which something can spread. Exactly *what* might spread will vary by context and the relationship type.

Consider a few examples of functions that links could serve in a social network. A link may serve as a channel of information. Perhaps that is what friendship does in a particular setting; it opens a line of information sharing. A link may serve as a channel of peer pressure, establishing whose esteem I value and whose approbation I prefer to avoid. A link may serve as a channel of goods exchange, allowing money, technology, or other things to flow from person to person. Or, if it provides sufficient contact, a link may serve as a channel of disease spread.

A key theme of this book is that a link is not completely defined unless both its type and its function have been specified. Knowing that a link indicates friendship only identifies its type. Why friendship might matter for any study hinges on what it is that friendship may do in that context.

An empirical social networks study is concerned with what a well-defined social network is like, and whether and how it matters in some context of interest. Designing a successful one requires serious thinking on the front end about what the network is and what it does in theory. This book aims to help researchers do just that.

The Vast Scope of Social Networks Research

Exactly what context is of interest in a networks study is up to the researcher. This can be daunting, as social networks have been shown to be relevant to a staggering variety of outcomes. Classic network studies span nearly all topics and timelines, stretching from hunter-gatherer groups and their patterns of social organization (Apicella *et al.*, 2012), to medieval Florentine elite and their political maneuvering (Padgett & Ansell, 1993), and modern-day adolescents and their romantic entanglements (Bearman *et al.*, 2004).

It turns out that social networks are relevant to lots of different outcomes that social scientists study. The presence, quality, and pattern of relationships have been found to affect outcomes such as voting behavior (Sinclair, 2012; Sokhey & McClurg, 2012), political attitudes (Lazer *et al.*, 2010), public opinion (McClurg, 2006), policymaking (Scholz *et al.*, 2008), legislative

politics (Ringe *et al.*, 2013), social capital (Jackson *et al.*, 2012), protests (González-Bailón & Wang, 2016), and political violence at a multitude of levels (Parkinson, 2013; Larson & Lewis, 2018; Dorff *et al.*, 2023), as well as public health decisions (Rao *et al.*, 2007) and technology adoption (Conley & Udry, 2010), to name just a few.

Although the options for the types of networks to focus on and the contexts in which to study them are numerous, a common footing will help any network study. This book aims to help organize thinking early in the planning stages of any empirical study of networks, whatever the topic area may be. The more carefully a theory of which network matters and why is laid out at the start, the stronger the study will be.

Having some examples of theory-driven research in mind can be helpful. Consider a few findings from such research in political science.

When a rebel group is just starting to form, its few members plan and train in hiding from the government, but often in view of citizens of remote towns or villages. Whether these citizens keep the nascent rebel group's presence a secret from the government is the key determinant of whether the rebel group eventually grows to become a viable threat (Lewis, 2020). Larson and Lewis (2018) show that the citizens' social networks are a key part of this process. When the nearby villagers' networks allow everyone to hear information from everyone quickly, people coordinate easily on secret-keeping and the rebels are likely to succeed. When these social networks are more fragmented into separate pockets, different groups of civilians have different views of the rebels – and someone is likely to spill information to the government, leading to the rebel group's demise. Whether trusted social networks are fragmented or unified makes all the difference.[1]

Switching gears, Cruz *et al.* (2017) studied elections in the Philippines. The authors show that political exchange is an important motivation for vote choice in this context. Voters prefer candidates who are likely to provide them with perks. It turns out that a voter's access to the perks that a politician may dole out is determined by their placement in the family marriage network. Someone whose family is married to the politician's family has better access to perks than someone whose family is more distant, say married to someone who is married to someone in the politician's family. Moreover, candidates who are close in this sense to the most voters in an area are more likely to win elections. In summary, the location of candidates in marriage networks determines their likelihood of electoral success.

---

[1] Cruz *et al.* (2020) show that this same network feature – the extent of fragmentation into separate social groups – also matters in a different network, intermarriage among families, for a different outcome, public goods provision. Fragmentation can thwart elite capture, which has the consequence of improving public goods provision in certain areas.

In yet a different context, Eubank and Kronick (2021) show that social networks can facilitate peer pressure that encourages political behavior. The context of this study is political activity in Venezuela. It turns out that individual actions, such as signing a petition demanding a referendum on recalling President Nicolas Maduro and attending a protest demanding the petition be honored, also have social motivation. The authors show that although any peer can apply peer pressure, some people occupy positions in a social network that afford especially ample opportunity for many of their peers to do so. Venezuelans with lots of friends who could pressure them to sign and sanction them if they refrained were especially likely to sign, and the same goes for joining the protest.[2]

The above three examples all feature nodes that are individuals. Networks can also matter when the nodes are instead aggregates. In a study of fighting tactics in civil conflict, Dorff *et al.* (2023) consider networks in which the nodes are groups of armed actors – "warring groups." Two warring groups are considered linked if they have battled one another. Warring groups have a choice of battle tactics and can choose to selectively target civilians when doing so helps to secure territory without losing too many supporters in the process. It turns out that the benefits of this tactic outweigh the costs when the network features high "network competition" – lots of pairs of armed actors that battle one another. Warring group networks that have more links are those that generate more violence against civilians in civil war.

In all of these contexts, a well-defined social network matters, but it functions somewhat differently. In the rebel context, networks matter because they allow villagers to share information with one another and coordinate on a plan.[3] A feature of the network's shape – whether it is fragmented or not – affects how well the networks can perform this role. In the Philippine candidate context, a different network is relevant. Here, intermarriage is what ties families together, and these connections provide access to the spoils of politics. Position within this network, being closer to or more distant from a candidate, determines a person's likelihood of receiving some of the spoils, which shapes their preferences over candidates. In the civil war setting, networks of battles among warring groups determine the costs and benefits of using certain tactics.

In all of these empirical network studies, there is a set of links among a set of nodes that is hypothesized to matter for some outcome. The theory for why

---

[2] Larson (2017c) finds that this same use of networks may also explain why behavior was so civil in a very different context: boomtowns in the so-called "wild west." People could watch what others were doing, spread the word through their networks, and the threat of the resulting punishment kept most people behaving quite well.

[3] Arias *et al.* (2019) also consider the coordinating role of networks but in the context of Mexican voters evaluating new information about candidates.

the links matter appeals to what exactly the particular links in the network can do. Each study establishes not only that links matter, but that the particular arrangement of the links is especially important. If they are arranged into separate fragments, we expect a different outcome from a more unified arrangement. Links arranged to bring a voter closer to a politician change that voter's preference and, in the aggregate, influence the election outcome more than long chains of links separating voters from candidates. Arrangements that let many others monitor and peer-pressure someone result in different behavior compared to arrangements that let someone go barely noticed. Arrangements that contain more battle links result in more civilian violence compared to those that contain fewer.

This book sets out to help researchers design an empirical social network study. It will help them identify the nodes and links that are most relevant, develop a theory for why the network matters, make a plan for data collection, and describe the measured network's key features. Naturally, there is no one-size-fits-all design; the optimal design will vary study by study, and hinges on the theory underlying the study. Consequently, this book pays special attention to carefully specifying a network-related theory that will guide complex design decisions.

Types of Empirical Social Networks Research

As the above examples show, social networks can vary in a number of ways. Of course, the type of link can vary – a network that includes friendship links is quite different from one that includes battle links. But once we start thinking about the relationships of the actors involved in a relationship, that is, the *other* links of the two people who are linked, it becomes apparent that the arrangement of links within a network can vary too.

Think about it from the perspective of one person within a network. That person may have one relationship or many. Even if she has only one relationship, just the one link, she may have many indirect ones if the person she is linked to has many. Moreover, exactly how she is situated in the network may be very different from how someone else is situated in the same network. She may have few links while someone else has many. Or she and someone else may have the same number of links, but she is connected in two steps (linked to someone who is linked to someone) to many people while someone else is connected in two steps to few people. Or it may be that the people she is linked to are all linked to one another while the people someone else is linked to are not. These are just a few examples of possible variation in individual positions within networks.

When we zoom out further, networks can also vary as a whole, so that one network may be quite different from another network. One may have more links than the other, or even if the links are present in equal number, one may have the links arranged so that people are to a greater extent arranged into separate social pockets. Or one may have links arranged so that in general people are indirectly linked to everyone else in few steps while another may have people quite far apart in the network.

Putting these pieces together, there are many different sets of nodes that we could study in any context. Then, given a set of nodes, there are many different links that we could focus on. Given a choice of nodes and links to study, those links could be arranged in many different ways, creating variance across nodes and across whole networks. Consequently, there are lots of ways for an empirical study to advance our understanding of social networks.

It can be helpful to think in terms of categories of options for theory-driven empirical social networks research, organized by a project's goal. Some research takes the network as the outcome of interest and aims to describe it, determine why it looks the way it does, or explain how it would change in response to a given stimulus. Such projects usually have one of the following types of goals:

1. Describe the network.
2. Establish how the network changes in response to something.

A network can also serve as an independent variable theorized to explain some outcome of interest. We can break up projects with these goals into three more categories:

3. Establish that the network is relevant to some outcome.
4. Establish that particular node-level features within networks matter for some outcome.
5. Establish that particular network-level features matter for some outcome.

Chapters 2 and 3 introduce a wealth of node-level and network-level features that could be of interest. It is also possible to conduct an empirical networks study with the point of generating more careful theory. The above categories assume that the researcher can stipulate the network that is likely to matter and a way in which it could. Of course, empirical investigation can help generate a knowledge base from which such a theory can be constructed in the first place. A final category of empirical networks research is then:

6. Learn about features of networks and their relationship to outcomes in pursuit of theory-building.

This book will help researchers design projects that fall into any of these six broad categories – all of which are strengthened by a theory-forward approach.

Why Networks Warrant Special Consideration

Once you are sold on studying a network as part of a broader study, why read a whole book about it? As it happens, a social networks study has some features that set it apart from more traditional research designs.

First, a social network is its own kind of object, with properties that are unique to its kind. Whole fields have been devoted to describing and categorizing a network.[4] This means that networks research requires learning a new vocabulary for describing the positions of nodes within networks and attributes of networks overall. The concept of "fragmentation" in the rebels example, and "closeness" in the voters example, are just two of many network features that can be described and studied. This book introduces this vocabulary. Because networks are collections of nodes and links, some of the concepts need a fair bit of notation to express them precisely – but the book leans heavily on examples and intuition to make this as efficient as possible for a new user of networks. By the end of this book, the reader will have a broad sense of the ways that networks can vary and how these differences can be related to political outcomes.

Second, a social network study aims to empirically study something – a social relationship – that can be conceptually nebulous. The researcher must decide what counts as a link in theory, and also how to operationalize that concept in order to collect the appropriate data. It is tempting to say the social relationship of interest is "friendship" and move on. However, whether the real nodes that are studied understand friendship to be the same thing the researcher understands it to be, or whether the proxies for friendship that generate data pick up a sufficiently rich notion of friendship, are issues that can make or break a study. Researchers need to take care to thoroughly specify which relationships might matter, and why, in order to properly design an empirical study of networks. Doing so requires having a clear network theory in mind. This book walks the researcher through crafting and operationalizing their theory.

Third, social network studies depart from traditional research designs that rely on sampling of independent observations. By definition, a researcher

---

[4] A branch of mathematics called graph theory does so in great depth. See Trudeau (2013) and Murty and Bondy (2008).

including a measure of a network is interested in *dependent* observations, since it is the interconnections among nodes that are the object of study. One implication is that drawing inferences from a small, randomly selected sample is often not a viable option. A broader implication is that special care must be taken to design the study well ex ante since the standard battery of fixes that are available for studies using randomly drawn samples of independent observations are unavailable to the networks researcher (for instance, simple multiple imputation of missing data).

This book helps the researcher make careful design choices in advance. It also offers practical guidance for designing original network data collection, by focusing heavily on designs that entail fieldwork in which original data about networks are elicited using surveys. Of course, this is not the only way to collect new network data, but it is perhaps the most complicated. The principles of design that guide original data collection in field settings are also those that guide good practice in other data assembly techniques (scraping links from social media, for instance). Even if a researcher never plans to collect original data in field settings, thinking through the design that *would be best in such a case* is a constructive way to establish the gold standard to which other collection efforts can be compared. Researchers using a dataset where links are likes on Facebook or trade flows between countries can think through the hypothetical, theory-driven survey that would collect the ideal data to better assess their actual data plan.

Collecting new network data via surveys in the field is an increasingly popular option, especially for studying rural or developing country settings. Eliciting networks via surveys has been central to advances in knowledge about the consequences of network position for individual behavior (Barr *et al.*, 2009; Baldassarri & Grossman, 2013; Baldassarri, 2015; Breza *et al.*, 2015; Atwell & Nathan, 2022; Larson *et al.*, 2022); the spread of information through a group (Banerjee *et al.*, 2014; Alatas *et al.*, 2015; Larson & Lewis, 2017); the reach of shared ideas among elites (Ingram, 2016); the presence and magnitude of peer influence (Conley & Udry, 2010; Banerjee *et al.*, 2013; Paluck *et al.*, 2016); and the way social networks change in response to shocks (Fafchamps & Gubert, 2007; Jackson *et al.*, 2012). The contexts for these studies vary widely, for example, from villages in the Philippines and India to high schools in the US.

In any context, however, measuring networks in the field is not a straightforward task. Since respondents in a network are interrelated – and thus cannot be taken as independent and identically distributed observations – problems that would be minor under the usual independent sampling procedures can compound into large problems that generate misleading results in a network context. To measure any one network, the way nodes and links are sampled and the way survey questions are worded and ordered have large consequences

for the resulting data. Moreover, for a given group of people, numerous different networks could be measured. Because resources, time in the field, and the attention span of respondents are all finite, researchers measuring networks face a large set of high-stakes tradeoffs that can be challenging to evaluate prior to collecting data.

This book helps alert researchers to pitfalls that could arise at all stages of an empirical network study to help guide the most promising study right out of the gate.

Should I Undertake Networks Research?

When a study is designed to have a network as the dependent variable, the answer is straightforward: The network should be included in the study. For studies in which a network *could* be an independent variable, considering a couple of questions can help inform the final choice of whether to include it. The primary guiding question is:

1. When I think about how the outcome relates to (or depends on) any one actor, could something about other actors matter too?

If the answer to this question is no even in principle, then adding networks to the research design is unlikely to be fruitful. We might expect actions to be determined exclusively by an actor alone when the decision is especially personal, when strict rules constrain the ability for others to weigh in, or when there is no time for social ties to operate. A person's immediate reaction to some stimulus, say a survey prime, may be strictly individual if she has no time to interact with and be influenced by any other actors before her reaction is recorded.

Another way to frame this consideration is to ask: Is there a network that my actor is a part of such that if that network were different in some way, I would expect the outcome to be different? Suppose we are interested in studying international recruitment to the Islamic State of Iraq and Syria. The actors of interest are citizens of European countries, and we want to understand what drives some to relocate to join ISIS. If we believe all that matters are individual incentives such as ideology, life experiences, expected compensation, and outside economic options, then including a network may be unnecessary. However, if we could imagine a situation in which trusted friends and family members personally recruit people to ISIS, so that people with more connections to ISIS supporters are more likely to be recruited, then including this network in a study would be valuable.

Of course, we should be open to the possibility that our reasoning about the irrelevance of networks is wrong. We might have thought that whether a person turns out to vote, decides to get a flu shot, or contributes to a voluntary retirement plan are decisions an individual makes for herself, but it appears that the behavior of social ties also matters for all three (Bond *et al.*, 2012; Rao *et al.*, 2007; Duflo & Saez, 2003). Therefore, another guiding question is:

2. If an actor can be affected by other actors, are the relevant other actors a specific subset or all of them?

The simple fact that other actors may affect an actor is not a sufficient reason to include a network in a study. A network can be thought of as a record of which actors are relevant to which other actors. If *all* other actors are relevant to all actors, then measuring networks will add little value. Returning to our ISIS example, a person's choice to leave Europe and join ISIS could be affected by other people in the sense that a person is more likely to do so if that seems like a popular choice for others in her country. The more other people join, the more likely she is to join.

While this does mean that her behavior depends on other people, it does not depend on specific other people. If instead she is more likely to join if her trusted friends and family join, and her trusted friends and family are possibly a different set of people from someone else's trusted friends and family, then a network will add value. The network, by recording who are the trusted friends and family of whom, would keep track of who could be affected by whom. The former scenario is one in which other people matter "globally" while the latter is one in which others in a network matter.[5]

3. Would a record of which actors are connected to which other actors explain variation in my outcome?

Believing that links in a network are relevant to an outcome makes that network a good candidate to consider including in a study of that outcome. Relevance is a higher standard than mere existence. Lots of networks exist among actors. After all, networks are everywhere. As Chapter 4 describes, we could define many different networks among any group of actors. The people sitting in the same room right now are interconnected in a co-room-sitting network. People born in the same city are linked in a shared-birth-city net-work. Just because we *can* define a network among a group of actors does not mean that studying that network will be productive for learning about them.

---

[5] We could split hairs and note that the global view is technically also a network in which shared country of residence is the link. But everyone being linked to everyone else is not a very interesting network, and would not need to be measured carefully using the tools of this book. Chapter 5 discusses this further.

The explanatory power of any independent variable is based on how variation in it corresponds to variation in the dependent variable. We should evaluate independent variables involving networks the same way. The more variation in a network explains variation in the dependent variable, the more explanatory power the network has. This book covers lots of different ways that networks can vary. A network is valuable to include in a study when changes in its structural features (which nodes and links are present and how they are arranged, see Chapter 2), node and link attributes (what the nodes and links are like, see Chapter 3), or link usage (who uses links and why, see Chapter 4) are theoretically connected to changes in some outcome. The stronger the connection to changes in some outcome, the more valuable its inclusion is likely to be.

When a study is more exploratory, more focused on theory-building, then collecting evidence to help evaluate the above three questions can be a useful way to settle on whether networks should be included moving forward. Exploration can probe for the existence of a network for which question 3 is answered in the affirmative.

How to Use This Book

This book is designed to take a reader who is curious about networks through the steps of defining a network, theorizing about why the network might matter, creating a data collection plan, preparing the data for analysis, visualizing it, and describing it. The more technical chapters conclude with exercises that offer additional practice along the way, with selected answers at the end of the book.

Chapter 2 begins by introducing the technical details of how to describe networks overall and the positions of nodes within them. This chapter covers the key terms that appear in many network studies – degree, centrality, network distance – and offers many opportunities to further develop intuition about their interpretation. It might be tempting to skip what feels like a light math lesson, but it is useful to develop the mental encyclopedia of network shapes that comes from working through this chapter. Practice connecting these to outcomes paves the way to creating and testing theory. Readers already comfortable with the definitions of network features might skip this chapter, or return to it for practice with intuition.

Chapter 3 introduces substantive features of networks. Knowing which nodes are included in the network and which pairs of nodes are linked is sufficient to know everything about the structure (shape) of the network. However, nodes and links may both have substantive attributes that could enrich

our view of a network. This chapter introduces node and link attributes and presents ways to incorporate these features to do things like establish that some links are more important than others or to make assessments about whether nodes are disproportionately connected to other, similar nodes.

Chapter 4 walks the reader through the process of precisely defining a network of interest, creating hypotheses for why that network might matter, and offering theory that explains the hypotheses. These concepts are picked up in Chapter 5, which guides choices about operationalizing the network. Determining what counts as the network in theory and then what will indicate a link to include in the data in practice are surprisingly subtle issues that warrant careful thought.

The nuts and bolts of data collection are covered in Chapter 6. Some data are already assembled and ready to use. In the event that the researcher needs to assemble a new dataset of nodes and links, this chapter helps with the key decisions, including how to evaluate different tradeoffs when time or budget is constrained. It focuses most heavily on collecting data in field settings, and then extrapolates principles that pertain to all original network data collection.

Chapter 7 covers the easily overlooked step of preparing data for analysis. Network data create special issues with storage and cleaning. This chapter is a guide to handling these issues, but should not be saved for the end of a study. Reading this chapter before collecting data helps design the collection in a way that can avoid difficulties later on.

Once you have network data to work with, visualizing it and describing it are key to assessing data quality and catching data entry problems. Chapter 8 is a tutorial for exploring network data in R using the package igraph. It covers importing the data, creating a network object, visualizing the network, calculating the structural features covered in Chapter 2, and incorporating the substantive features covered in Chapter 3.

At that point in the book, the next step is to read widely to collect inspiration for network studies and to learn more about methods of data analysis or other design tools. The Conclusion offers suggested readings to springboard the reader into the next phase of the study of networks.