Social Network Analysis

Katherine Ognyanova (Rutgers University)

Ognyanova, K. (2022). Social Network Analysis. In A. Ceron (Ed.), *Encyclopedia of Technology and Politics*. Edward Elgar Publishing.

Abstract

Social network analysis (SNA) gives us the tools to examine how social relationships form, how they are organized, and how their structure affects our choices and actions. This interdisciplinary line of research is informed by both the social and the natural sciences.

Early works in this area focused on descriptive characteristics, developing metrics to identify key individuals, capture important network attributes, or discover closely knit subgroups and communities. Today, network analysis often aims to explain how social ties influence behavior and how our actions change the structure of interpersonal networks.

Advances in information technology have also played an important role in the development of the field. Digital platforms have changed the nature of human social structures, as well as enabling the development of new analytical tools and data collection strategies.

Keywords: network, methodology, social structure, tie strength, centrality, contagion

Author Bio

Katherine Ognyanova is an associate professor at the School of Communication & Information, Rutgers University. Her research examines the effects of social influence on civic and political behavior, confidence in institutions, information exposure/evaluation, and public opinion formation. Ognyanova's methodological expertise is in computational social science, network science, and survey research. Her recent work examines the links between misinformation exposure and political trust.

Introduction

Social network analysis is an interdisciplinary research approach using mathematical, statistical, and computational methods to examine the structure of social relationships. The study of interconnected systems has a long tradition across academic fields including sociology, social psychology, anthropology, and political science (Freeman, 2004). More recently, important network analytical tools have also come from natural science disciplines such as physics and biology. In the social sciences, the use of network analysis marks a gradual shift from individualistic to relational scientific explanations.

In the twenty-first century, advances in computing and technology in general have contributed to the growth of network research. Networked forms of organization supported by new information and communication technologies have come to the forefront of social scientific explorations. At the same time, technological innovations also made possible the development of advanced and resource-intensive methods for data collection and analysis. Network science, previously perceived as being fairly descriptive and atheoretical, has also developed nuanced explanatory frameworks that support theory building in a variety of disciplines.

One of the benefits of social network research is that it captures both individual behavior and global structure, allowing scholars to make connections between micro- and macro-levels of analysis. In political science, researchers have explored networks with diverse composition, from social networks among voters, to congressional or lobbying networks, interorganizational networks, and even networks of nation states.

Network definition and structure

A network notation can describe a variety of systems composed of interconnected entities. The two types of elements that constitute a network are *nodes* (the connected actors, also known as *vertices*) and *ties* (the connections, also known as *edges*). A wide variety of entities can serve as nodes in a network – people, machines, groups, organizations, nations, documents, and locations, among others.

Network ties represent a relationship or exchange among actors. Links of friendship, advice-seeking, collaboration, or political discussion can form a network among people. Research focusing on digital connections has also explored ties based on phone calls, online message exchange, or connections between user accounts on online platforms. Depending on the data and research question, network ties can be binary (two actors are either connected or not), signed (actors can have positive, negative, or no connection), and valued (each tie is assigned a number based on its strength). For instance, a binary tie may tell us whether two people are friends or not, while a signed tie can show us whether they like or dislike each other. A valued tie can denote the strength of a friendship, or the number of weekly phone calls people exchange.

As evident from these examples, network ties can represent durable relationships (friend, colleague), behaviors (communication), transactions (money transfer, information exchange), or sentiments (liking, trust). Network relationships can also be classified as *directed* (liking someone, asking them for advice, following them on Twitter) or *undirected* (being married, being Facebook friends). Undirected networks are ones where connections must always be mutual, while directed networks have ties that may not always be reciprocated.

One key characteristic of social connections in the network literature is *tie strength.* In an influential study on this subject, Granovetter (1973) suggested that we rely on stronger, closer relationships for material and emotional support, while weaker ties are a better source of novel information. Our strong ties are expected to share our social circles and rely on the same information sources we use. In contrast, weak ties are often seen as more likely to connect us to new people and ideas. More recent research has suggested that strong ties may be useful sources of new information in a rapidly changing information environment, while weak ties are more helpful in cases when information is scarce (Aral, 2016). How we measure tie strength also varies from study to study, with common metrics focusing on emotional closeness, social support, reciprocity, communication frequency or volume, shared connections and social contexts (Brashears & Quintane, 2018).

Network data collection

The two most-often used approaches to representing and analyzing network data focus on *whole networks* and *ego networks* (also called egocentric networks). Whole networks typically require that the researchers select a well-defined group of interconnected actors and capture the relationships among them (see Figure 1). The nodes in such a network may, for instance, be students in a classroom, employees in an organization, participants in an online community, or members of Congress. The collected data describing their relationships may be self-reported, with each person naming from memory or selecting from a roster their social ties in the group. Alternatively, the data may come from offline archives (e.g., organizational charts), communication logs (e.g., phone call metadata), or digital behavior records (e.g., emails, social media data).

In many cases, however, we may not have a well-defined interconnected group. Additionally, we may want to do research representative of multiple different groups operating in different contexts. In such cases, we can select a set of actors who may not belong to the same group or have any ties to each other. For instance, we can use a random sample from the population of a country, or from the users of an online service. For each person in the sample (*ego*), the research captures key relationships to others (*alters*) who are often ***not*** part of the sample.

For example, a survey may ask a nationally representative sample of voters to each name 5 people with whom they discuss politics. The survey may further ask each respondent to describe the relationships among the five people they named, as well as list the party affiliation of each named person. This information would allow researchers to examine the local political conversation network of each respondent.

The networks described above focus on relationships between actors of the same type. Those could, for instance, be ties of friendship among people, or partnership links among organizations. Occasionally, we may be interested in the connections among two types of actors: for instance, examining how people are connected to organizations. Networks that describe relationships between nodes of two different types are called *two-mode networks* (also known as *affiliation networks* or *bipartite graphs*). We can use them to represent employees connecting to companies, voters connecting to information sources, politicians belonging to coalitions, or online users connecting to social media platforms.

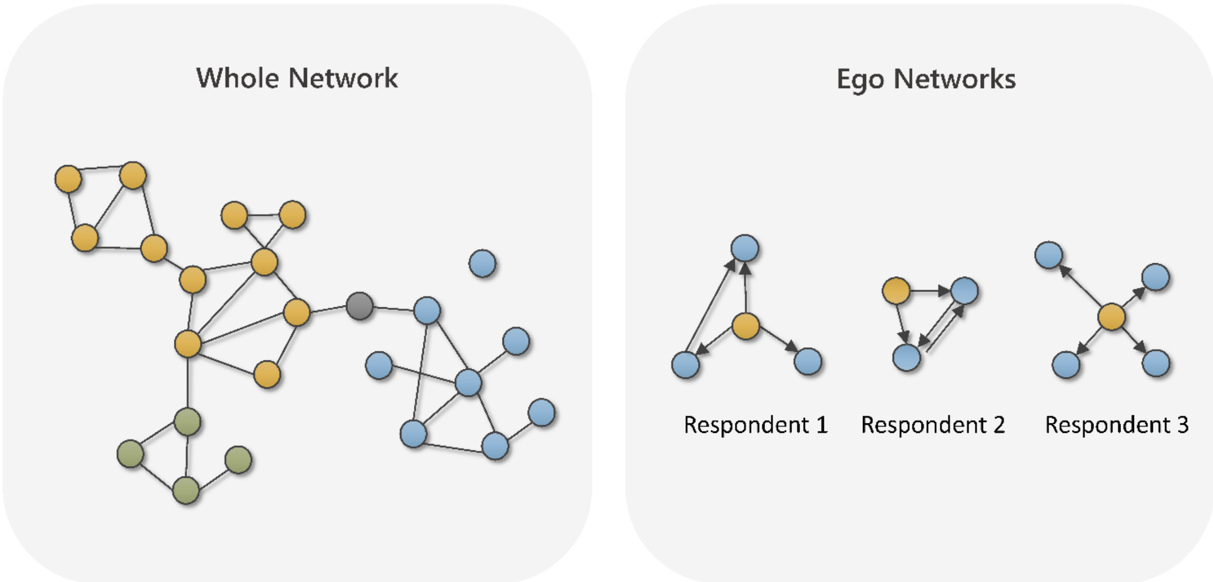

Figure 1. Whole and egocentric networks.

Network metrics

In the early days of network analysis, research often focused on describing the structural characteristics of actors and the overall properties of networks. Some of the key questions asked in such studies revolved around identifying important or influential individuals. A number of different *node centrality* metrics were developed to answer those questions (Wasserman & Faust, 1994).

The simplest way to identify important nodes in a network is to examine the number of connections they have. Well-connected actors with many social ties may be more visible and have better access to information, support, and other resources. Reflecting that idea, the popular *degree centrality* measure is based on the number of ties for each actor. In networks where ties have direction (e.g. Emma advises Alex, or Emma follows Alex on social media), we can calculate *in-degree centrality* as the number of incoming connections and *out-degree centrality* as the number of outgoing connections. People with many incoming connections may be popular or influential (Emma has many followers or colleagues asking her for advice). People with many outgoing connections may be especially active (Alex is following many people online or asking them for advice).

A related measure of importance, *eigenvector centrality*, was developed based on the idea that it matters not only how many people you know, but also who they are. Instead of assuming all ties are created equal, this metric gives more weight to ties with people who are also well-connected themselves.

A different way of conceptualizing importance is to look for brokers: people who connect others or bring groups together. Those are the individuals who create links between parts of the network that might otherwise be disconnected. One measure reflecting this brokering position is called *betweenness centrality*. It is based on the number of times people need to go through a node in order to reach someone else in the network. Actors who serve as bridges between different communities are likely to have high betweenness scores. A bridging position may offer various advantages, such as early access to new information or the ability to control the flow of resources (Burt, 2005).

A wide variety of additional metrics of centrality are available for specific purposes, such as identifying people who can easily reach others in the network, or actors who have key positions with regard to information flows (Borgatti et al., 2013). Many of the centrality measures tend to be correlated, though often the most important people based on one metric may not score highest using a different metric. For example, within the blue group on Figure 2, node B has the most connections, while node E is the only one with a bridging tie to a different group.

Descriptive metrics are also useful when we want to examine the properties of a network as a whole. *Density*, for example, is the proportion of ties that are present in a network (out of all theoretically possible ties). In directed networks, *reciprocity* is based on the proportion of ties that are reciprocated (e.g. Emma follows Alex and Alex follows Emma back). Network *centralization* metrics examine the extent to which node importance is evenly distributed among actors (*decentralized networks*) vs. cases where there is a small number of important and well-connected actors (*centralized networks*).

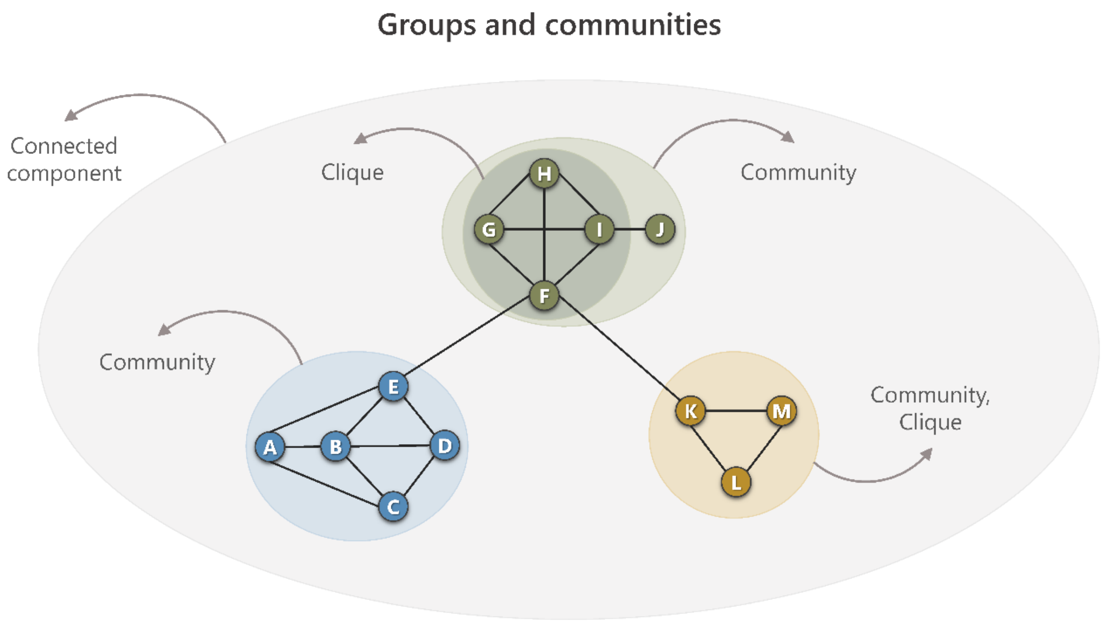

Figure 2. Network subgroups

Another key task in network research is identifying specific subgroups and communities. Those clusters of nodes may be based on physical proximity, shared social context, shared interests, or other factors. Very often, we can identify them without knowing why they formed by simply observing the network structure. For example, a clique in an undirected network represents a very tightly knit group where each node is connected to all others in the group (see Figure 2). A connected component is a group of actors that may not be directly linked but are reachable from each other. For example, perhaps Emma doesn't know Anna, but Emma knows Alex who knows Anna. Anna is thus reachable from Emma and belongs to the same connected component. We can also examine how the network can be split into communities – groups of actors that have many connections with others in the same community, but very few that go across to other communities. A number of algorithms can perform community detection in different ways, depending on the network type and research goals (Dao et al., 2020).

Advances and future directions

Since its early days, network analysis has moved beyond simple descriptives. Studies today can investigate the complex mechanisms of network evolution and the drivers behind tie formation and dissolution. Through empirical and simulated data, research has examined how information, behaviors, or diseases spread through social structures (Lehmann & Ahn, 2018).

Social science often uses network analysis to answer one of two questions: (1) how do networks influence our actions and attributes, and (2) how do our actions and attributes affect network structure. A key analytical challenge here is disentangling the processes of *social influence* and *social selection*. Influence causes people to adopt actions and opinions endorsed by their social ties. Social selection (also known as *homophily*) reflects the tendency of actors to connect to others who are similar to them (McPherson, Smith-Lovin, & Cook, 2001). Thus, we may select friends who agree with us politically – or we may influence our friends to change their opinions to match ours. Both of those options will result in networks that are homogenous with regard to political preferences.

A number of methods have been used to explore social influence and social selection. Among them are natural or lab *experiments* (Centola, 2010), which randomize the structure of networks or the treatment of actors. *Stochastic actor-oriented models* provide a tool for analyzing longitudinal network data to examine the co-evolution of social ties and individual behavior (Snijders, Van de Bunt, & Steglich, 2010). *Exponential random graph models* offer another way of exploring the social mechanisms and attributes that could generate specific network structures (Lusher et al., 2012) .

In addition to advances in methodology, the past few decades have brought about new forms of data collection. Large-scale, high-velocity digital data streams have enabled us to capture the minute-by-minute evolution of massive human networks. Researchers can now obtain data from mobile phones, email systems, online platforms, smart devices, as well as from digitized archives of previously inaccessible offline data. The abundance of networked information also creates new challenges. Network scholars need to interpret the digital signals, extract meaning from data that was not designed for research, and investigate whether traditional theoretical frameworks apply in a new setting (Lazer et al., 2021). Future work in network analysis will thus not only have to design tools that can handle these data, but also develop new frameworks explaining the role of social structure in a changing information environment.

References

Aral, S. (2016). The Future of Weak Ties. *American Journal of Sociology*, 9.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks* (1 edition). SAGE Publications Ltd.

Brashears, M. E., & Quintane, E. (2018). The weakness of tie strength. *Social Networks*, *55*, 104–115. https://doi.org/10.1016/j.socnet.2018.05.010

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. Oxford University Press.

Dao, V. L., Bothorel, C., & Lenca, P. (2020). Community structure: A comparative evaluation of community detection methods. *Network Science*, *8*(1), 1–41. https://doi.org/10.1017/nws.2019.59

Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press ; BookSurge.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K., & Radford, J. (2021). Meaningful measures of human society in the twenty-first century. *Nature*, *595*(7866), 189–196. https://doi.org/10.1038/s41586-021-03660-7

Lehmann, S., & Ahn, Y.-Y. (Eds.). (2018). *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks* (1st ed. 2018 edition). Springer.

Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications* (Illustrated edition). Cambridge University Press.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.