Analyzing online social networks

Tod Van Gunten
tvangun@ed.ac.uk

May 16, 2022

Abstract

Social network analysis (SNA) is a broad family of methods and theories for studying relations between social entities, such as individuals. The rapid growth of online communication, new forms of social interaction such as social media, and information dissemination on the internet makes SNA methods and theories a natural approach to online research. This 'how to' guide provides a brief overview of the fundamentals of social networks, common research questions in the field, and a discussion of data sources and the challenges of data collection. SNA methods involve a variety of tools, from relatively simple data visualization to sophisticated statistical models. Research questions often focus on 1) the positions occupied by particular nodes; 2) the relations between all pairs of nodes; 3) groups or 'communities' within the network and 4) the structure of networks as a whole. While the internet makes many new data sources available, collecting these data in a manner accessible for SNA methods remains challenging, particularly at large scales. Fortunately, many tools exist to address this challenge for online platforms such as Twitter, Youtube and Reddit, while other approaches are available in different settings. Twitter is probably the most popular site for current research, owing to the relatively easy access to data.

Forthcoming in *SAGE Research Methods: Doing Research Online*

Social network analysis (SNA) is a broad family of methods and theories for studying relationships among social entities. These entities may be individuals, but also organizations, digital communities, non-human users (e.g. bots), places, and even abstract objects (words or concepts), among other possibilities. Social relations are similarly broad: they include those based on friendship or acquaintance — in the colloquial sense of having "friends" or "followers" on social media platforms — but also a wide variety of other forms of social interaction and similarity. At the broadest level, SNA seeks to identify patterns in these relations, as well as correlations and causal influences between these patterns and other variables of interest.

The increasing availability of digital data online has opened new horizons for social network data collection, as well as creating new forms of social interaction amenable to SNA analysis. The internet itself, as its name suggests, is a large network (and a popular search engine, Google, uses an algorithm based on network theories). Social media sites such as Facebook, Twitter and Instagram are of course "social networks" in the colloquial sense, and much research on these platforms has used SNA methods. Many other online sources provide data that can fruitfully be analyzed using SNA theories and methods. Indeed, since many data sources are disseminated online, the boundary of what constitutes an "online network" is ambiguous. To limit the scope of this entry, I focus mainly on networks representing activity occurring through electronic media, though I will discuss other research where relevant for theoretical or methodological reasons.

This entry consists of three main sections. First, I provide some necessary foundations and terminology. Second, I consider representative research questions at different levels of analysis. Third, I review the challenges of collecting network data online and indicate various ways that previous research has addressed these challenges.

1 Foundations: what is a network?

At the simplest level, a network is a set of points (often called nodes or vertices) connected by lines (called ties or arcs). As noted, nodes represent some set of social entities, such as individuals or organizations, and ties represent relations between them. Ties may be *directed* or *undirected*, reflecting whether the relationship in question is inherently symmetric or not. For example, a

Twitter user may follow another without being 'followed back;' in this case the relationship has directional meaning and may point in one direction, the other, or both simultaneously. In contrast, on Facebook 'friendship' requires both users to agree to the relationship, and all friendships are reciprocated by definition (though this may not be true of the offline relationship); such ties are undirected Ties may also have a value, for example, representing the strength of the relationship. For example, the number of times one user retweets another may contain information about the strength of the relationship.

Networks represented in this way can be described mathematically in many different ways. A simple property of nodes is their *degree*: the number of ties that connect a particular node. For example, the count of a user's friends or followers on a social media platform is a measure of degree. In directed networks, degree may be disaggregated into *in degree* (the number of directed relations "pointing at" a focal node) and *out degree* (the number of relations pointing "away"). The count of a Twitter account's followers is a case of in degree (and can be interpreted as influence or prominence), while the number of accounts followed is a case of out degree (interpretable as susceptibility to external information). Degree distributions are often skewed, with very long tails: most nodes have a small number of connections, while a small number of nodes have very many connections. This "scale free" character is a revealing property of networks because of the underlying processes that often generate such skewed distributions. Interaction settings in which nodes tend to form relationships with other nodes who already have large numbers of relations — as in high school popularity contests, a process known as 'preferential attachment' (Barabási and Albert, 1999) — tend to generate these skewed degree distributions. This observation is relevant to the dynamics of online networks and the emergence of 'viral' phenomena, insofar as the ability to observe the behavior of other users (e.g. a high number of 'likes' on a user's content) can further increase the attention paid to a small subset of users.

At a higher levels of aggregation, important properties of networks include density, the degree of clustering, and the length of paths through the network, among others. In the simplest case, network density is simply the number of observed ties divided by the number of possible ties, or (stated differently) the proportion of nodes in the network that are connected. Density is a simple measure of cohesion, or how well a network "holds together." Other important properties of networks include the length of shortest paths between all pairs of nodes (dyads) in the network. Intuitively, these path lengths capture the possibile speed of transmission of information through a network. If we imagine information passed by word of mouth through the ties making up the network (as in the game of telephone), the shortest path lengths between every dyad indicates how quickly information can spread through the network. This concept of network "flow" is central to models of diffusion and contagion and the "small world" model, discussed below.

There are many other properties of networks at the micro (node), meso (cluster or group) and macro (the network as a whole) level. A classic review of network concepts is Wasserman and Faust (1994); more recent reviews include Yang et al. (2017) and Knoke and Yang (2020).

2 Research questions

Researchers should carefully consider the goals of analysis before starting data collection. The fact that some set of social entities are connected in a network is seldom interesting in its own right; rather, SNA theories generate hypotheses about the effects and determinants of networks at different levels of analysis. While the array of research questions is endless, the following provides some orienting ideas.

2.1 Positions

Often network analysts are interested in some aspect of the *positions* of particular nodes within a network. For example, we might be interested in knowing who the most popular users on a social media platform are. A simple approach is to count the number of friends, mentions, or likes that a user or account has. Such a measure could be interpreted as an indicator of influence or popularity, depending on the domain. This is a particular application of centrality analysis, and more particularly the concept of "degree centrality" as discussed above. Centrality is a major area of study in network theory and there are more sophisticated measures than this simple count. For example, some approaches take into account the number of ties of each node's network neighbors: for example, a node who is "liked" by other nodes who also receive many "likes" would be considered more popular. The Google PageRank algorithm is one application of this idea (Brin and Page, 1998). Another important measure is betweenness centrality, which captures the extent to which nodes are located at the interstices of a network, often seen as a strategically valuable position.

Centrality measures can be used descriptively to identify relevant (e.g. influential) nodes in a network (e.g. Tremayne, 2014; Burris et al., 2000). Other researchers wish to go beyond such descriptive approaches in order to analyze the effects or causes of centrality. In general, research designs based on network positions often explore the hypothesis that position in a network is associated with some other outcome, such as rewards or prestige. Rossman et al. (2010) use the Internet Movie Database (IMDB) to create a network-based measure of status (based on a logic similar to the PageRank algorithm) among film actors, which is associated with a higher probability of academy award nomination. Or, as in the case of Lewis et al. (2008), researchers explore the determinants of occupancy of these network positions, in other words, why some individuals are more central than others in a network.

2.2 Relations

Another set of research questions concern the relational level of analysis: that is, social processes observed among pairs of nodes (called dyads). A researcher might be interested, for example, in the factors that lead some users to select other users (as friends, or as objects of attention) on social media platforms. In this case the main unit of analysis is not the node (as with centrality analysis, as just discussed), but rather the relation or dyad. For example, an important research agenda seeks to understand the effects of social homophily, often expressed as the tendency of "birds of a feather to flock together" (McPherson et al., 2001). In other words, individuals and other units form social networks in part on the basis of pre-existing attributes, such as demographic characteristics or cultural tastes. Research on homophily seeks to disentangle the impact of this preference for similar individuals on the composition and structure of networks from other processes that shape network formation. For example, Wimmer and Lewis (2010) use data from Facebook to study racial homophily (the preference for same-race or ethnic friendships) on a university campus.

A related ongoing strand of research that can be fruitfully addressed at the relational level focuses on the role of networks in the diffusion of technologies, culture, collective action and other behaviors and, relatedly, various forms of political and cultural similarity. This research is highly relevant to the spread of culture, political information and disinformation, and 'viral' phenomena online. Theories of social contagion suggest that new behaviors from the adoption of new technologies to choice of cultural goods to participation in social movements spread through social networks (for a discussion, see Centola, 2018). Social actors are more likely to adopt a particular technology, product or political behavior if they are surrounded by contacts who have already done so. Studies of collective action online (González-Bailón et al., 2011; Larson et al., 2019) often formulate hypotheses at the relational level of analysis. As just noted, homophily implies that actors tend to select into relations with those with whom they share similar underlying characteristics — such as cultural tastes and political inclinations. This entanglement of homophily and social influence is a major challenge in network studies seeking to study either, and methodological innovations have focused on identification strategies to isolate these distinct social processes (e.g. Aral et al., 2009).

Analytical methods at the relational level have developed substantially over the past several decades; research in this area often involves substantial statistical sophistication. Relational data analysis must address the non-independence of observations: when analysing all pairs of dyads, the existence or value of a network tie between Janet and Satoshi is not independent of that between Janet and Maria (for example, both may be influenced by Janet's intrinsically outgoing character). This dependence violates the assumptions of standard methods of statistical inference. Because of this, network analysis requires specialized statistical tools. The state of the art approach is the exponential random graph models (ERGM) which can be used to study to occurrence, or formation of relations in a network at the dyadic level. The work of Wimmer and Lewis (2010) already mentioned is a good example of the utility of these models.

2.3 Groups

Visual inspection of networks often gives the impression of clusters or groups: we have the intuitive sense that some set of nodes "hang together" but apart from other sets of nodes. This tendency of nodes to cluster into discrete subgroups is the basis of the large set of methods of community detection. These methods have the common aim of partitioning a network into a set of subgroups, where each node is assigned to one (non-overlapping) subgroup. An important family of approaches is based on the concept of modularity, which is (stated informally) a measure of the extent to which the density of ties *within* groups is greater than *between* groups. Intuitively, a network with discrete subgroups is one in which most network ties are with members of the same group, rather than between groups. The analysis and visualization of the political blogosphere in the 2004 US elections (Adamic and Glance, 2005) is a classic illustration of this concept: right-wing blogs link to other right-wing blogs, and left-wing blogs to other left-wing blogs, but links across partisanship are rare. The existence of a strong group structure on social media platforms is one manifestation of the "echo chamber" phenomenon.

Researchers may be interested in identifying communities or subgroups within an online network in order to facilitate navigation or qualitative interpretation of the network structure. For example, community detection is often helpful in 'mapping' online communities. Olson and Neal (2015) use community detection alongside other methods to produce a map of reddit, identifying communities (and relations between communities) of users discussing such topics as electronic music, programming, and LGBT issues. Similarly, Darius and Urquhart (2021) studied twitter users communicating about lockdown policies, conspiracy narrative and other topics related to the Covid–19 pandemic and use community detection to identify areas of thematic interest. These kinds of approaches use community detection in a primarily descriptive and qualitative way, as a means to uncovering the content of social interaction. Other approaches use community approaches to study diffusion approaches as discussed above. For example, Weng et al. (2013) add a community layer to the diffusion of memes (some of which 'go viral') on twitter, showing that some memes are 'trapped' within particular communities of interest, while others have cross-community appeal, enabling them to reach broader audiences.

2.4 Whole networks

Finally, some studies seek to examine the global properties of networks as a whole. For example, a classic question in the networks scholarship concerns the cohesion of social networks. Intuitively, cohesive networks are those that are tightly bound together and difficult to break apart (Moody, 2001). Scholar might be interested, for example, in whether an online movement or community constitutes a single, well-connected unit, or multiple competing or hostile factions. Simple measures of cohesion include graph density (discussed above) and the average length between all pairs of nodes in the network (though see Moody and White 2001 for limitations of these approaches). Other questions at the whole-network level include the degree of centralization, hierarchy, or clustering, among other measures. Modularity-based community detection methods discussed above also provide a useful network-level metric in the modularity score, which indicates the extent to which the network is composed of discrete groups, rather than a single group.

Because global network measures often boil down to a single number (e.g. the overall density of a graph), whole network approaches are usually most useful where either longitudinal (a network observed over time) or comparative (equivalent networks observed in different populations) data are available. Social media data lend themselves naturally to longitudinal analysis given the constant flow of information. For example, researchers might wish to quantify the degree of clustering or modularity in an online community over time to evaluate whether polarization or fragmentation are increasing, decreasing or constant. Among longitudinal analyses, Óskarsdóttir and Mallett (2021) track density, diameter (the maximum path length across the network), clustering and other measures in the blockchain transaction network underlying bitcoin. The work of Shwed and Bearman (2010) is a good illustration of the utility of over-time modularity measures. Wüest et al. (2021) take a comparative approach, measuring the size, density and centralization of Twitter networks associated with Swiss political parties.

The influential small world model of network structure (Watts, 1999) suggests that many empirical networks have a high degree of "local" clustering (for example, a strong tendency for friends of friends to also be friends) combined with a small proportion of paths that connect these otherwise distinct clusters. Some studies seek to evaluate whether such as "small world" model fits the global network structure by calculating average path lengths and clustering coefficients (Watts, 1999; Moody, 2004). Researchers also evaluate this model in relation to other models of network structure, such as the "preferential attachment" model discussed above.

3 Data collection and analysis

As already noted, the growth of online communication has made new forms of network data available, but collecting these data remain challenging. These challenges include access, scale, ethics, and technical difficulty. The sheer size of many online platforms is a challenge for data collection and storage: datasets may rapidly reach multiple gigabytes in size, posing a challenge for both storage and manipulation. Many web platforms limit the data that researchers can collect by disallowing some data collection approaches by both legal and technical means. Thus, the ubiquity of online information does not equal ease of access, and for some projects custom coding is essential. Moreover, analyzing data generated or collected for reasons other than research purposes poses an additional set of challenges (Robins et al., 2021).

At time of writing, among major social media platforms Twitter is probably the most accessible to researchers. Twitter operates an Application Programming Interface (API) available to academic researchers, which can be used to collect data at large scales. Many scholars have used networks of followers, retweets and other online interactions on Twitter to study social movements (Barberá et al., 2015; González-Bailón et al., 2011; González-Bailón, 2016; Larson et al., 2019) and other political communication online (Tremayne, 2014; Darius and Urquhart, 2021), among other topics. In contrast, in the wake of controversy over online privacy, researcher access to Facebook data is now extremely limited and possibilities for network analysis are constrained (though see Lewis et al., 2008). Some studies (Brooks et al., 2014; Hofstra et al., 2017) have used survey approaches supplemented with data collected from Facebook itself.

Other researchers have used network data from interaction on massively multiplayer game environments (Burt, 2012), links between web pages of political movements (Doreian and Mrvar, 2021; Park and Thelwall, 2008; Burris et al., 2000), software development collaboration in Github (Gote et al., 2021), polarization on political blogs (Adamic and Glance, 2005), editorial conflict on wikipedia (Yasseri et al., 2012), racial prejudice in online dating sites (Lewis, 2013), diffusion of technologies on an instant messenging platform (Aral et al., 2009), and cryptocurrency transactions (Fischer et al., 2021), among many others. Combining data across multiple platforms is an emerging area: one recent study combined multiple digital communication flows including email, instant messages, and video calls among Microsoft employees to study the impact of home working during the Covid–19 pandemic (Yang et al., 2021), though obtaining such vast datasets is likely to be a challenge for academic researchers.

Beyond specific platforms and technologies, researchers can repurpose a wide range of online data for social network analysis. Many data sources can be usefully analyzed as social networks. A classic approach is to treat rosters of event participation as a network, with co-participation in events treated as a network tie. For example, Davis et al. (1941) collected data on the social events attended by a group of women; this can be analyzed as a network among the women, with ties representing the number of events co-attended by the women. (Alternatively, this can be treated as a 'bipartite' network, that is, one with two types of nodes — people and events — and connection only between opposing node types). This basic paradigm can be applied to many data sources. For example, the network of joint film appearances represented in the Internet Movie Database (e.g. Rossman et al., 2010) has a similar structure: actors participate in films, and joint participation is a network tie.

The mechanics of collecting network data online is an key challenge. Data collection on small networks — up to a few hundred ties, for example — can be done "by hand" (e.g. by typing or pasting information into a spreadsheet or text file). However, the amount of time required increases rapidly as the size of the network increases. Many applications require some automation of data collection, either through custom coding or "off the shelf" tools. Among the latter, the software package NodeXL (Smith et al., 2010) allows users to collect data from several social media platforms and perform some basic network analysis; YouTube Data Tools (Reider, 2015) facilitates searching and extraction of several network types.

Researchers gain more flexibility by coding in a flexible general purpose software environment such as R or Python, particularly for large-scale data collection. In R, access to the Twitter API for Academic Research is available with the AcademicTwitteR (Barrie and Ho, 2021) package, while the vosonSML package (Graham et al., 2020) facilitates data collection of hyperlink networks across the internet, as well as data collection and network extraction from Twitter, Youtube and Reddit, as well as collection of hyperlink networks on the broader internet. RedditExtractoR (Rivera, 2021) also provides access to the Reddit API and function for extracting networks. Another resource relevant for research on Reddit is the pushshift database (Baumgartner et al., 2020), which permits faster searches of historical data. Researchers can also write customized web scrapers to collect data from particular online resources; scholars pursuing this route need to review ethical and legal rules regarding bulk data collection online.

An important but overlooked point to bear in mind in devising data collection plans is that often knowing reliably which units are *not* connected is as important and difficult as knowing which units are connected. Scholars sometimes invest substantial time and resources into collecting network data, only to discover that the network is so densely connected that there seems to be no meaningful or interesting structure. Another issue is that information found online may be biased towards certain kinds of visible relations while omitting hard to observe relations, making information about the absence of ties unreliable. Research design and data collection should thus anticipate how non-relations can be reliably known.

3.1 Software for analysing SNA data

Once data are collected and structured in network form, scholars have additional options for analyzing network data. As with data collection, researchers who write code in (or collaborate with coders adept in) a general purpose statistical computing language such as R or Python have more flexibility in choosing analytical approaches. SNA is a relatively quantitative field of research, but the level of quantitative sophistication required varies widely, with some research questions requiring no knowledge of advanced mathematics or statistics. Network analysis is also a highly visual field in which drawing a picture (usually with aid of a computer) is often the first step of the research process. Network visualization most often uses the simple device of representing nodes as points and relations between them as lines. A wide variety of network layout algorithms are available to select positions for the points; often the choice of layout is relatively arbitrary and reflects aesthetic judgment, though in some settings there are scientific grounds for choosing a layout.

Researchers who do not wish to code can find many tools for network visualization and analysis in programs such as Ucinet (Borgatti et al., 2002), Gephi (Bastian et al., 2009) and Pajek (**?**).

These programs are very flexible, but often require data input in network form (though simple data transformations are possible). Thus, one of the main difficulties of conducting network analysis without coding is the generation of relational data in the first place; for some research questions, analysis is tractable without coding once data are available in network form.

4 Conclusion

Researchers considering undertaking analysis of online networks face a number of challenges, from formulating research questions to collecting and analyzing potentially large datasets. As this guide has indicated, studies can range from analysis of small datasets of a few dozen entities, up to very large networks observed repeatedly over time. A key decision researchers face is thus what skills to invest in, depending on their research questions and other motivations.

Notes

References

Adamic, L. and Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the the 3rd International Workshop on Link Discovery*, pages 36–43.

Aral, S., Muchnik, L., and Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5):21544–21549.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10):1531–1542.

Barrie, C. and Ho, J. C.-t. (2021). academictwitteR: Access the Twitter Academic Research Product Track V2 API Endpoint.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media.*

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., and Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020)*, pages 830–839.

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002). Ucinet for Windows: Software for Social Network Analysis. Analytic Technologies.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117.

Brooks, B., Hogan, B., Ellison, N., Lampe, C., and Vitak, J. (2014). Assessing structural correlates to social capital in Facebook ego networks. *Social Networks*, 38:1–15.

Burris, V., Smith, E., and Strahm, A. (2000). White Supremacist Networks on the Internet. *Sociological Focus*, 33(2):215–235.

Burt, R. S. (2012). Network-Related Personality and the Agency Question: Multirole Evidence from a Virtual World. *American Journal of Sociology*, 118(3):543–591.

Centola, D. (2018). *How Behavior Spreads: The Science of Complex Contagions*. Princeton University Press, Princeton.

Darius, P. and Urquhart, M. (2021). Disinformed social movements: A large-scale mapping of conspiracy narratives as online harms during the COVID-19 pandemic. *Online Social Networks and Media*, 26:100174.

Davis, A., Gardner, B. B., and Gardner, M. R. (1941). *Deep South Chicago*. University of Chicago Press, Chicago.

Doreian, P. and Mrvar, A. (2021). Hubs and Authorities in the Koch Brothers Network. *Social Networks*, 64:148–157.

Fischer, J. A., Palechor, A., Dell'Aglio, D., Bernstein, A., and Tessone, C. J. (2021). The Complex Community Structure of the Bitcoin Address Correspondence Network. *Frontiers in Physics*, 9:1–16.

González-Bailón, S. (2016). Networked discontent: The anatomy of protest campaigns in social media. *Social Networks*, 44:95–104.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, 1(1):197.

Gote, C., Scholtes, I., and Schweitzer, F. (2021). Analysing Time-Stamped Co-Editing Networks in Software Development Teams using git2net. *Empirical Software Engineering*, 26(4):75.

Graham, T., Ackland, R., Chan, C.-h., and Gertzel, B. (2020). vosonSML: Collecting Social Media Data and Generating Networks for Analysis.

Hofstra, B., Corten, R., van Tubergen, F., and Ellison, N. B. (2017). Sources of Segregation in Social Networks: A Novel Approach Using Facebook. *American Sociological Review*, 82(3):625–656.

Knoke, D. and Yang, S. (2020). *Social Network Analysis*. SAGE Publications, Inc., Thousand Oaks.

Larson, J. M., Nagler, J., Ronen, J., and Tucker, J. A. (2019). Social Networks and Protest Participation: Evidence from 130 Million Twitter Users. *American Journal of Political Science*, 63(3):690–705.

Lewis, K. (2013). The limits of racial prejudice. *Proceedings of the National Academy of Sciences*, 110(47):18814–18819.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4):330–342.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 2:415–444.

Moody, J. (2001). Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*, 107(3):679–716.

Moody, J. (2004). The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review*, 69(2):213–238.

Olson, R. S. and Neal, Z. P. (2015). Navigating the massive world of reddit: Using backbone networks to map user interests in social media. *PeerJ Computer Science*, 1:e4.

Óskarsdóttir, M. and Mallett, J. (2021). Strangely mined bitcoins: Empirical analysis of anomalies in the bitcoin blockchain transaction network. *PLOS ONE*, 16(9):e0258001.

Park, H. W. and Thelwall, M. (2008). Link analysis: Hyperlink patterns and social structure on politicians' Web sites in South Korea. *Quality & Quantity*, 42(5):687–697.

Reider, B. (2015). YouTube Data Tools (Version 1.22).

Rivera, I. (2021). RedditExtractoR: Reddit Data Extraction Toolkit.

Robins, G., Bright, D., Weissinger, L., and Stys, P. (2021). Data collection for social network research. *Social Networks*, page In Press.

Rossman, G., Esparza, N., and Bonacich, P. (2010). I'd Like to Thank the Academy, Team Spillovers, and Network Centrality. *American Sociological Review*, 75(1):31–51.

Shwed, U. and Bearman, P. S. (2010). The Temporal Structure of Scientific Consensus Formation. *American Sociological Review*, 75(6):817–840.

Smith, M., Ceni, A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., and Dunne, C. (2010). NodeXL: A free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016. Social Media Research Foundation.

Tremayne, M. (2014). Anatomy of Protest in the Digital Era: A Network Analysis of 'Twitter' and Occupy Wall Street. *Social Movement Studies*, 13(1):110–126.

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

Watts, D. J. (1999). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, 105(2):493–527.

Weng, L., Menczer, F., and Ahn, Y.-Y. (2013). Virality Prediction and Community Structure in Social Networks. *Scientific Reports*, 3(1):2522.

Wimmer, A. and Lewis, K. (2010). Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*, 116(2):583–642.

Wüest, B., Mueller, C., and Willi, T. (2021). Controlled networking: Organizational cohesion and programmatic coherence of Swiss parties on Twitter. *Party Politics*, 27(3):581–596.

Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., Joyce, C., Shah, N., Sherman, K., Hecht, B., and Teevan, J. (2021). The effects of remote work on collaboration among information workers. *Nature Human Behaviour*.

Yang, S., Keller, F. B., and Zheng, L. (2017). *Social Network Analysis: Methods and Examples.* SAGE Publications, Inc, Thousand Oaks.

Yasseri, T., Sumi, R., Rung, A., Kornai, A., and Kertész, J. (2012). Dynamics of Conflicts in Wikipedia. *PLoS ONE*, 7(6):e38869.