Social Network Analysis (Social Media)

ITAI HIMELBOIM
University of Georgia, USA

The study of social media networks has evolved in the last two decades into an interdisciplinary scientific area of research. Network analysis of social media data emerged toward the end of the 20th century when, for the first time in history, immense social interactions were recorded and became available for researchers. At that point in history, decades of social science literature, theoretical and empirical, about network analysis of small and medium-sized social (i.e., symbolic) networks intersected with a growing body of literature in hard sciences, from biology to computer science and physics, that examined much larger physical networks. Social scientists contributed the theoretical and conceptual foundations for understanding social network structures and the role of key players and communities in these communication networks, within the broader sociological, political, economic, and psychological levels, to name a few. Natural scientists introduced an understanding of the structures and dynamics of large networks (neurological, genetic, statistics, and computers, for instance) as well as the methods and algorithms required for analyzing social networks in a magnitude that had never been seen before. Network analysis of social media data is blooming within this unique intersection of social and natural sciences.

Systems in different areas of life and research take the form of networks. Examples include social networks, the Internet, the World Wide Web, neural networks, metabolic networks, food webs, distribution networks such as blood vessels or postal delivery routes, and citations networks. A necessary key starting point is a definition of a social network. A social network is a constellation of nodes and links. A *node* (also known as an actor or a vertex) is the fundamental unit of any network, social or otherwise. Nodes are found across social media spaces. Examples of nodes include a user on Twitter, Facebook, Wikipedia, a discussion forum, Instagram, or a blog, or a photo on Flickr. A *link* (also known as an edge or a tie) on social media is a symbolic connection between two nodes. A few examples include a "friendship" between two Facebook users, a follow relationship between two Twitter users, a hyperlink sent from one blog to another, a reply of one user in a discussion forum to a message posted by another, and a common tag that two photos on Flickr share. A network (also known as a graph) captures interconnected groups of nodes. When the domain of interaction is a social media space, these social connections create social media networks.

A *domain* defines the boundaries of the network. A domain can be a specific social media platform, such as Twitter or the blogosphere. A domain of an interaction can be determined by a section of a single social media platform. A fan page on Facebook captures social interactions among users who post, like, and share content within that

The International Encyclopedia of Communication Research Methods. Jörg Matthes (General Editor),
Christine S. Davis and Robert F. Potter (Associate Editors).

DOI: 10.1002/9781118901731.iecrm0236

page only. A single discussion forum focused on a single topic captures the interactions among its participants. A *topic-network* also defines a domain of interaction. A topic-network captures the social network created by a group of users on social media who posted content about a given issue. For instance, a topic-network on Twitter can be the President's State of the Union address. The domain of this network includes users who posted content containing the string "State of the Union" or its related hashtag (#sotu), and this domain also includes the Twitter links among them (mentions and/or follow, for instance). Within the blogosphere, a domain may include only political blogs or blog titles corresponding with a given set of keywords. Consequently, a single social media space can include an almost endless number of topic networks.

A social network approach to research shifts the focus from the characteristics and traits of a social actor (individual, organization, a country, etc.) to the type and pattern of relationships that are formed between actors. Network analysis characterizes nodes based on their location in the networks, rather than by their inherent traits. For instance, in a Twitter network, a node can be characterized based on the number of followers it has within the network, or the extent to which it bridges otherwise much less connected users. A link on Twitter can be formed by different types of social ties, such as follow, retweets, or replies. A single group of users can therefore form many networks, depending on the type of links the researcher is interested in examining. For instance, two very different breast cancer topic-networks on Twitter can be formed, depending on how links are defined (e.g., follow, mentions, replies, retweets).

Sociometric and egocentric networks

Social network literature identifies two distinct approaches to social network analysis. The *sociocentric network* approach focuses on the quantification of ties between people within a defined group or domain. Historically, this approach was influenced by the German sociologist Georg in the mid-20th century, and this approached was adopted by other areas such as subsequent work on networks in classrooms, residents in a village, or trade among nations. In social media, sociocentric networks include nodes and links formed in topic-networks on Twitter, a Facebook fan page, a newsgroup about a particular issue, or blogs that share a common topic, and so on. Network analysis measures patterns of those interactions, the overall structures of the network, and their implications. Much of the research about social networks analysis in social media takes this approach, and therefore hereafter when a network is discussed the reader should assume it is a sociometric network unless indicated otherwise.

The *egocentric* (i.e., personal) *network* approach focuses on a node (here: a seed node) and the relationships surrounding this node, typically an individual. The origins of this approach are often traced to English anthropologist Alfred Radcliffe-Brown. An egocentric network (or ego-network) is formed by nodes that connect to a seed node (i.e., its alters) and the links or relationships among them. For instance, an individual can be a seed node while a partner, children, co-workers, and members at their social clubs are the alters. The network is composed by all the alters and the ties among them. In social media, an egocentric network can be formed by a user, its alters, and the relationships among them. For instance, a Facebook ego-network will include a Facebook user, those he or she has Facebook ties with, and the ties among them. On Wikipedia, an egocentric network may include a seed article, all the articles that an article (seed) has hyperlinks to (alters), and all hyperlinks collecting these nodes.

Directed and undirected networks

A network is a constellation of nodes connected by links. In many social networks, these links are naturally mutual. If Jon and Jane are friends, for example, this relationship, or link, has no direction. Jon is Jane's friend and vice versa. Such a network of friendships will be considered an undirected network. Other relationships have a direction. Jon may ask for Jane's advice, but Jane may not ask for Jon's. Such a link has a direction, from Jon to Jane. A network of advice seeking, therefore, is composed of a directed network. Communication via social media creates both directed and undirected networks. In social media, threaded discussion forums are directed, as Bob may reply to a message Ron posted, but Ron does not have to reciprocate. On Twitter, Danielle may follow Michelle, creating a directed link from Danielle to Michelle. A link may not necessarily be mutual, as Michelle is not required to follow Danielle's Twitter account for a tie between them to form. When forming following relationships on Twitter, then, users form directed networks. A network of Facebook friends, in contrast, has no direction. One must approve a friendship request for a tie to form. Bob and Ron are friends on Facebook, and this link has no direction. The network of Facebook friends, therefore, is an undirected one.

The directionality of a network is therefore determined by the definition of links, not nodes. Users on a social media space may form several networks, some directed and some not. For instance, Facebook users form an undirected network by friending one another. The same users also form directed networks when they "Like," comment, or share a friend's post, as these activities are forms of directed ties. On the blogosphere, hyperlinks sent from one blog to another constitute directed links, and therefore the network of blogs connected via hyperlinks is a directed network. Blogs can also be connected by a mutual tag (e.g., SNA), creating an undirected network of blogs.

The significance of the directionality of links and networks is both methodological and conceptual. Conceptually, a directed link determines the direction of flow (e.g., information, influence, social capital, web traffic). Methodologically, as will be discussed at length later, some social network analysis metrics are better suited to capture key users and patterns of connections in directed networks while others are better suited for undirected networks.

Key social network analysis measurements in social media

On social media, large amounts of information are shared and exchanged. Attention to information and sources of information, however, has become scarce and therefore key to evaluate social media activity. Within this context, influential users became one of the most sought after, but also most criticized, concepts in social media analytics. The almost mythical idea that a minority of influential individuals on social media, and social networks in general, can have massive influence over the information flow and attitudes, relied primarily on anecdotes and examples of content spread widely on social media ("going viral"). The notion of "influentials" gained much criticism. Scholars pointed to the lack of conceptual or operational definitions for these key individuals, the problematic underlying assumption that key users will be influential across social media and across issues, and the overall lack of empirical support (see, for example, Watts & Dodds, 2007).

Social networks metrics provide unique understanding of patterns of information flow, and attention giving and receiving. It provides context for identifying key users, and their potential influence, at different levels of analysis. When applying social media network analysis, units of analyses vary and can include a node, a link, a cluster, and the whole network.

Node-level measurements and influential users on social media

Users on social media ("nodes"), can take a variety of structural positions in a network, holding different types and levels of influence in terms of information flow. These structural positions define the roles that users play in the network. Users in structural positions are identified by measurements of connectivity. One of the most common sets of measurements are centrality measurements. Centrality refers to how prominently connected an actor is in a network. As the name implies, actors who are more important in a social network are "central." Centrality helps to explain the extent to which an individual or organization is connected to others in their environment. There are several types of key structural roles that a user can take in a network.

Degree centrality and hubs. A common position that a social actor may take in social media networks is a hub. A hub is an actor with many relationships with other actors (i.e., alters) in the network. On many social media networks, many connections ("links") mean many channels of information flow or indicators of attention given to social media content. In social networks terminology, this hub is central as it has a high degree of centrality. On Facebook, where networks are formed by users and Facebook friendships, a user's degree of centrality measures the number of friends the user has on Facebook. In directed networks, where each link has a direction, two types of degree centrality emerge: *in-degree centrality* and *out-degree centrality*. In-degree centrality is based on ties or relationships that others have initiated with a user, and out-degree centrality is based on the relationships one initiated with others. Many social media networks are directed. On Twitter, relationships have a direction, as one user may follow, mention, or reply to another. An in-degree would be the number of mentions, replies, or followers one has in the network. Tweets posted by a high in-degree user enjoy a larger audience of followers or widespread via retweeting for example. One's out-degree is the number of other users it follows, mentioned, or replied to. In the blogosphere, a blog's in-degree would be the number of hyperlinks other blogs directed toward that blog, and the out-degree is the number of blogs it posted hyperlinks to. Similarly, the network of users and replies in discussion forums, email networks, and many other social media networks forms directed social networks, and the prominence of nodes in the network can be measured using their in- and out-degree centrality values. A node with high degree is therefore considered a hub in that network.

Betweenness centrality and bridges. Popular issues are discussed by large numbers of individuals on social media. Many of those are not directly connected to one another, relying on other users for information flow. Users on social media networks, then, can vary in terms of the extent to which they play the important role of network bridges. Such users hold power over the extent and type of information they share. Organizations, as will be discussed later, learn to rely on such key users to reach out to new publics.

Burt's theory of structural holes examines social actors (e.g., individuals and organizations) in unique positions in a social network and where they connect other actors that otherwise would be much less connected, if at all. In Burt's (2005, p. 24) words, "A bridge is a (strong or weak) relationship for which there is no effective indirect connection through third parties. In other words, a bridge is a relationship that spans a structural hole." A lack of relationships among social actors, or groups of actors, in a network gives those positioned in previously structural holes strategic benefits, such as control, access to novel information, and resource brokerage. Actors that fill structural holes are viewed as attractive relationship partners precisely because of their structural position and related advantages. These actors are called brokers (as they fill a brokerage positions) or bridges. Granovetter (1973) showed that these actors form nonredundant, often weak, ties among otherwise less connected actors.

A contemporary example is that Jon has a group of Facebook college friends with whom he has strong relationships. He also has a relationship with Jamie, with whom he briefly met during an internship in a large PR firm abroad. For Jon's friends, Jon's weak relationship with Jamie may be an important network path to a group of PR professionals abroad. It is also a nonredundant relationship, different from other relationships in his strong and immediate social networks, which are highly interconnected with one another, but are not connected to Jamie's group of PR professionals. Jon, then, is located in a powerful structural position in his networks as a broker or a bridge.

Betweenness centrality, therefore, measures the extent that an actor falls on the shortest path between other pairs of actors in the network. The more people that depend on an actor to make connections with other people, the higher its betweenness centrality value. This value is therefore associated with bridging actors in a network.

It should be noted that in directed networks, one should consider a node's in- and out-degree centrality measures when evaluating the direction in which information flows through a bridge within the network. Taking the Twitter following network as an example, Jon's high betweenness centrality alone does not indicate the direction in which information flows. If Jon has low out-degree centrality but high in-degree centrality, it indicates that Jon is bridging users in the network by providing them with common information as he posts or retweets. In contrast, if Noa has high out-degree centrality but low in-degree centrality, then she connects the networks by her exposure to information coming from loosely connected users or groups. Both Jon and Noa hold unique but different bridging positions in their networks.

Closeness centrality. A different approach to measuring centrality is *closeness centrality*, which measures the average distance between a node and every other node in the network. Unlike degree and betweenness centrality measures, the lower the values of closeness centrality, the more central a node is in the network. A user with low closeness centrality can reach information from users across the network directly or indirectly via a small number of users. A user with a high value of closeness centrality is more peripheral in the networks as it is distant, in terms of connectivity, from most users to reach and to be reached.

Eigenvector centrality measures one's importance in a network, based on one's proximity to other important nodes. Following an algorithm, a score is assigned to each node based on its relative connectivity in the network, which is then used to calculate the eigenvector centrality of each node. As will be discussed later, large self-organized networks, including many social media networks, often form a power-law distribution where only a few nodes are highly and disproportionally connected in the network while most are much less connected, if at all. In these networks, most users are unlikely to become one of the very few high-degree users (or in directed networks, users with high in-degree). In this ecology, becoming close to important users is particularly important. Taking Twitter, for example, being followed by President Obama, the New York Times, or Justin Bieber, for instance, makes one important in the network, as their tweets can reach those who follow these highly followed users.

Reciprocity. A relationship between two actors is reciprocal, or mutual, if each actor has initiated a tie with the other actor. Social media users seek to gain attention to their posted content. In an attempt to do so, they give attention to others, by sharing, tagging, retweeting, or linking, for instance. The success of this strategy can be measured by the level of mutuality of connections between a user and its alters. From a different angle, organizations, including news organizations, joining social media often attract large audiences (fans, followers) for their existing reputation. Reciprocity is a key measurement to evaluate the extent to which news organizations adapt to social media. Recent studies, it is worth noting, suggest that news organizations often apply old practices to new spaces, as illustrated by their low reciprocity values. Broadly speaking, reciprocal relationships between individuals may indicate a wide range of social attributes, such as cooperation, trust, exchange of opinions, and power balance.

At the user level, reciprocity is measured as the number of users one is connected with (alters) that are reciprocal over the total number of alters. On Twitter, for example, a reciprocal or mutual relationship between two users can be established if they follow one another. If Dan is connected with 10 other users on Twitter (whether following or being followed), and five of these users relationships are mutual (i.e., Dan follows five users who also follow him), Dan's reciprocity value will be .5. Reciprocity can be used to evaluate users' relationship building. When establishing a social media presence, users often aim to attract the attention of influential users by giving them attention (retweeting, posting hyperlinks, tagging, mentioning, etc.). Reciprocity is used to evaluate the success of such a strategy.

Link-level measurements

A dyad is constituted of a pair of nodes connected via a link. Two Facebook users connected via a friendship relationship, one user following another, two YouTube videos that share a tag, a blog posting a hyperlink to another blog, and two individuals who bought the same book on Amazon.com are just a few examples of dyads on social media. Putting together a group of dyads creates networks.

Links are often seen as binary: exist or absent. Link-level measurements, however, are key for understanding any network. Earlier, the *directionality* of links was discussed. Two other aspects of links hold key insights for the network: link type and link weight.

Link type. Any given group of nodes can create multiple social networks based on the researcher's definition or selection of links. For instance, nodes in Wikipedia networks are entry-pages. These pages can be connected by hyperlinks, as Wikipedia entries frequently link to one another. A very different network is formed when considering Wikipedia editors as the connecting elements. A user who edits two pages also serves as a link, creating a network of pages and editors. These two distinct networks consist of the same nodes—Wikipedia pages—and share the same domain—Wikipedia. The networks, however, are very different and result in different key users. Pages will show different values of centrality measurements, for example, across the two types of networks.

Many other social media domains hold the opportunity to be studied across types of links. On Twitter, a follow, a mention, a retweet, a reply, a favorite, lists, and a shared hashtag, are just a few examples of links; each link creating a different network. On Facebook, "friendship," liking a post, commenting on a post, posting on one's wall, and tagging a user in a photo all constitute types of Facebook links. On Instagram, following a photo stream is one type of link while liking a photo is another.

A helpful distinction between types of links on social media is dynamic versus static. A static link is usually created once, and typically does not change over time. Facebook friendships or following relationships on Twitter, Instagram, or Flickr are all static relationships, in that while they can be changed, they rarely are. Liking posts, photos, or videos, posting textual and video comments, and retweeting or sharing messages are all dynamic relationships. A link may exist today, but may not be repeated tomorrow. Static links, and consequently static networks, capture the *potential* of information flow, influence, or attention. Following someone on Twitter does not guarantee that one would actually read any tweet that person posted. Dynamic links and networks capture the actual attention giving or information flow, as they require users to continuously take actions indicating their involvement in the network.

*Link weight***.** Relationships on social networks vary in terms of their strength. The weight of a link can be classified by the researcher. For instance, kinship relationships are often considered stronger than work collegiality. The availability of large social media network data allows researchers to measure weight by measuring the frequency of links. In static networks, links cannot occur more than once (e.g., Jon can friend Joe only once, and Jane cannot follow the *NYTimes* on Twitter more than one time), whereas in dynamic networks, a link between two users can form multiple times, allowing scholars to quantify the weight of each link. For instance, Rachel liking ten photos posted by Michal on Instagram constitutes a link weight of 10, connecting Rachel to Michal. George liking only one photo posted by Julie would form a link weight of 1. The Instagram relationship from Rachel to Michal could therefore be considered stronger than the link from George to Julie. Many social media activities can be weighted, such as frequency of liking, mentioning, tagging, linking, and sharing.

Link reciprocity. A node's reciprocity, as mentioned earlier, captures the extent to which a node's ties with its alters are mutual. At the link level, a tie may or may not be reciprocated. Jane follows Dana, but only if Dana follows Jane back; in this instance, the Jane–Dana link would be considered mutual. If a link is mutual, its reciprocity value is 1; otherwise its value is 0. More advanced measurements of link reciprocity take into consideration the extent to which a dyad is reciprocal. For instance, Kate replying 10 times to a message posted by Kay on a discussion forum, and Kay replying only twice to messages posted by Kate results in a reciprocity of 0.2 of the tie between the two (2 divided by 10). In this case, values can range between 0 and 1. Other measurements of reciprocity exist.

Network-level measurements

In many areas of the social sciences, group-level measurements are often an aggregation of individual-level traits or characteristics. For instance, a classroom of students is often described in terms of the average of students' scores or attitudes (e.g., height, grades, political opinions) or by categorical measures (e.g., percentage of students by gender). One of the unique characteristics of social network analysis is treating a group as a unit of analysis, characterized by network structures; that is, patterns of social interactions. For instance, like a network, a classroom can be characterized by a level of interconnectedness of students or overall mutuality of ties among all classmates. Network measurements define the network as a whole, rather than of an aggregate of its members. The whole is, indeed, more than the sum of its parts.

Density. Networks vary in terms of their interconnectedness. Some are more tightly interconnected, while in others nodes are only sparsely connected. Network density is measured as the number of possible or potential connections (i.e., links), over the number of actual connections. Any two nodes in a network can be connected, regardless of whether or not they actually share a link. Density values range between zero and one. It should be noted that the calculation is slightly different for directed and undirected networks, as directed networks have twice as many possible links (i.e., from node A to node B, and from node B to node A).

The extent to which a network is densely interconnected impacts the rate of information flow within it. Carley (1991) shows that interaction between individuals leads to shared knowledge, and shared knowledge leads to even more interaction. This finding has important implications for the stability of a group and its interactions with individuals outside its boundaries. Granovetter (1973) states that tightly interconnected individuals are typically connected by strong and redundant ties. Burt (2005) notes that networks in which people are very highly interconnected are better at transmitting information. Similarly, Coleman (1990) demonstrates that an important outcome of strongly embedded close relationships is an increase in trust between individuals, which can lead to increased information transfer. Lerman and Ghosh (2010), in a study of the role of social network structures in the spread of information on Digg and Twitter, found that the rate at which information is spread through a network depended on its density.

Reciprocity. Reciprocity measurements at the node and link levels were discussed earlier. At the network level, reciprocity measures the extent to which ties among a group of nodes is mutual. Reciprocity is measured as a proportion of mutual links to the overall number of links in a network. Values range between 0 (i.e., no mutual ties in the network) to 1 (i.e., all links are mutual in the network). Other approaches to reciprocity measurements exist.

Centralization. The degree in which connections are aggregated around just a few actors in the network is measured by network centralization. Nodes in a network in which one or a few actors attract a large and disproportionate number of connections depend primarily on these members for information flow and sharing. These few people are hubs as well as gatekeepers in their networks. A less hierarchical distribution of connections is an indication of more egalitarian patterns of information sharing and flow. On social media, the U.S. news blogosphere was found to exhibit a highly centralized structure, where information from the most central bloggers in the network dominated. In contrast, a different study found the structure of the online community created by the hyperlinks between web sites of Iranians in the Netherlands to be low in centralization, with many focal points existing in that network (van den Bos, 2006).

High levels of network centralization often form a star-shaped (or spoke-and-hub) network in which a single charismatic or popular individual is linked to a varying number of other users who have very few connections among themselves. Himelboim and Han (2013) found that over time, star-shaped clusters disappeared as soon as their core actors—such as celebrities—stopped tweeting about the topic. Wang, Jiang, & Ma (2010) examined a network of Chinese scientist bloggers, revealing several loosely connected star-like clusters in the network that indicated the existence of topical subcommunities each centered around different bloggers who were "experts" in particular domains.

Groups and clusters

Social media hosts conversations about a wide and diverse range of topics. While many conversations are formed by large groups of people and posts, participants are rarely connected to or aware of the conversation as a whole. In large social media conversations, users define the boundaries of information flow and influence through their selection of connections. Given the opportunity to interact freely, individuals create tightly connected subnetworks. Within these clusters, information flows more freely, while outside of these clusters, information flows more loosely.

In social networks, smaller subgroups of densely interconnected users—clusters—often arise. Clusters, also called "communities," refer to subgroups in a network in which nodes are substantially more connected to one another than to nodes outside that subgroup. In social media networks, links often take the form of information channels or giving of attention.

In Twitter political networks, users' exposure to political tweets derives from the users they follow. Twitter users, then, are more likely to read content posted by their cluster-mates than by users in other clusters. Likewise, users in a given cluster choose to expose themselves to the same set of hubs, which serve as popular information sources among these users.

An understanding of social media activity, then, relies to a large extent on understanding the naturally occurring communities within a given topic. A cluster is an important unit of analysis. Himelboim, Smith, and Shneiderman (2013) developed a Selective Exposure Cluster method to study these connected networks and their discussion patterns on Twitter. In each cluster, they identified key distributors of information, the hub users, and the most frequently used hashtags and hyperlinks, as indicators of the political orientation of each subconversation. The findings indicate the existence of patterns of selective exposure, in which participation in fragmented interactions and forms divides users into groups of individuals who are tuned in to a narrow segment of a wider range of politically oriented information sources.

Homophily is a theory that asserts that "a contact between similar people occurs at a higher rate than among dissimilar people" (McPherson, Smith-Lovin, & Cook, 2001, p. 416). It captures a key characteristic of naturally occurring social networks and depicts a mechanism through which "distance in terms of social characteristics translates into network distance" (p. 416). Put simply, homophily suggests that similar individuals will be socially closer to one another than dissimilar people. Homophily has been documented in literature for almost a century, and is based on demographic characteristics such as age, sex, race, and education as well as psychological characteristics like intelligence, political attitudes and aspirations, and the selection of peers across a range of relationships such as marriage, friendship, and mere contact. In social media networks, evidences of homophily are found based on political attitudes on blogs, Twitter, Facebook, and book buying habits. In the context of homophily, network clusters define the boundaries of information flow across un-likeminded users on social media and elsewhere.

A cluster, then, is a subgroup of nodes. As such, network-level measurements are applicable. Density, centralization, and reciprocity matrices are often used to describe not only whole networks but also subsections within them.

Modularity. Modularity examines the extent to which clusters are distinct from one another. While modularity is a network-level measurement, it is presented here within the discussion about clusters. Clusters may share many connections among themselves, just a few, or none at all. The extent to which information flows freely, for instance, across the network, depends not only on the overall density of the user connectivity but also on the ties across clusters. Smith, Rainie, Shneiderman, and Himelboim (2014) used NodeXL to examined patterns of cluster connectivity across a wide range of topic-networks on Twitter. They found that political conversations about controversial issues tend to form two distinct clusters, each leaning to an opposite side of the political spectrum. The overall density of the network is high, as each cluster is heavily interconnected; however, information flow is very limited across the network as the two clusters share very few links among them. Such networks would have a high modularity value. In contrast, professional conferences often create networks that are highly interconnected, both among users and across clusters. While some subgroups of users are more interconnected than others, these users frequently establish links across clusters lines. Such networks will have high network density, as users are interconnected, but low modularity, as clusters are interconnected rather than distinct.

Like clusters, a wide range of algorithms are proposed to measure modularity. A commonly used algorithm is proposed by Newman (2006). It measures the quality of the divisions imposed on the network. Modularity measures the extent to which the divisions among clusters are good, in the sense that there are many links within clusters and only a few between. The value of modularity can range between 0 and 1. The higher the modularity value, the more distinct or separate the clusters are; that is, the clusters are less connected to one another. However, beyond a modularity value of 0.6, networks show little or no increase in the division of clusters. Newman therefore suggests 0.6 is a threshold for detecting meaningfully distinct clusters.

Social mediators, direct and indirect publics, and siloed mediated content

The different levels of network measurements provide researchers and practitioners with conceptual and methodological frameworks for understanding and evaluating social media activity. Taking a cluster approach to examining social media conversations, Himelboim, Golan, Suto, and Moon (2014) defined a public as a social media network cluster. *Social mediators* were used to describe key users who act as a bridge between an organization and publics the organization cannot reach directly. While traditionally bridges have been measured by their betweenness values alone, this approach does not take into consideration the directionality of links, and, therefore, does not consider information flow and influence. The scholars paired high in-degree at the cluster level and high betweenness at the network level, to define and identify social mediators.

Building on this work, Himelboim, Jin, Reber, and Grant (2015) defined Twitter network clusters that surround an organization as *direct publics* and defined publics that organizations must rely on social mediators to reach as *indirect publics*. The scholars found that direct publics form a star-shaped cluster (i.e., high centralization), in which the organization is the primary point of connection between users who rarely interact with one another. In contrast, indirect publics where found to be more interconnected (i.e., higher density), suggesting that indirect publics share and exchange more content among users.

A key interest of many organizations is to reach out to publics that the organization cannot reach directly. Social mediators are key users that can serve this purpose. The researchers also examined the unique characteristics of the content and found that *siloed content* describe messages exchanged within a cluster and *mediated content* capture posts that are shared between users across social media clusters or publics. The findings demonstrated that, in the case of Airline-related Twitter conversations, tweets that include a call to action, negative content, or are associated with an organization's social sponsored activities were associated more with mediated than with siloed content.

Self-organized networks and common network structures

Much of the traditional offline social network analysis relies on relatively small networks, which kept many characteristics of large networks hidden. The study of very large nonsocial networks, such as computer networks, networks in biology, neurology, physics, and geography, reveal that regardless of the immense differences between networks, large self-organizing networks (i.e., scale-free networks or "real life" networks) tend to show very similar structures and patterns of interaction. The Internet, and in particular social media, provides researchers an understanding of large scale-free social networks.

Power-law and the preferential attachment. As discussed earlier, a degree refers to the number of links a node has in a network. Degree distribution describes and measures the distribution of links or ties among nodes, such as social actors in a network. Because highly connected nodes have a greater influence within their respective networks, degree distribution is an important structural characteristic. It can be seen as a barometer for the egalitarian structure of networks. A hypothetical normal degree distribution would show that a few social actors are connected to a few other actors, a few are connected to many actors, and most are connected to about the average number of actors in the network. In contrast, a strongly positively skewed degree distribution of ties among actors would suggest that a few actors have a large and disproportionate number of ties in a network, whereas most actors enjoy very few ties, if any. The skewed distribution indicates a much more hierarchical structure than a normal distribution would. In a skewed distribution a few social actors benefit from the networks, access information, and potentially control the network flow more than most other participants.

Although in theory networks can differ in terms of their degree distributions, networks across all spectra were found to follow a specific form of positively skewed degree distribution—power-law—that can be represented by a regression line on a logarithmic graph (Simon, 1955). Power-law is considered to be an inherent attribute of many large and self-organized networks. In social media networks, Power-law is an inherent attribute for the distribution of hyperlinks in blogs, followers on Twitter, replies in online forums, and friends on Facebook. Power-law also accurately describes the distribution of links in neutral science networks, such as biology, chemistry, and astronomy (for examples, see Faloutsos, Faloutsos, & Faloutsos, 1999; Tang & Wu, 2009). This phenomenon sometimes carries a variety of labels. "In economics, [power-law] goes by the name 'fat tails,' in physics it is referred to as 'critical fluctuations,' in computer science and biology it is 'the edge of chaos,' and in demographics and linguistics it is called 'Zipf's law'" (Newman, 2000, p. 412). Power-law, however, is not found when random graphs were simulated (Jeong, Tombor, Albert, Oltvai, & Barabasi, 2000).

When attempting to describe the mechanism that explains the formation of the power-law, scientists often refer to preferential attachment: in large networks, new links are attached preferably to nodes that are already well-connected (Newman, 2001). Consequently, already highly connected nodes increase their connectivity faster than their less-connected peers. In other words, "the rich get richer." Preferential attachment is identified in many social networks, such as in sexual relationships, emails and hyperlinks on the web, and citations in academic work. Preferential attachment is not limited to social networks and is also found in biology, chemistry, and linguistics.

*The small world and the six (or so) degrees of separation***.** The small world study encompassed several experiments conducted by Stanley Milgram (1967), examining the average path length within social networks of individuals in the United States. Counterintuitively to many, he found that regardless of the size of networks (the entire U.S. population, in this case), society forms a small world-type network characterized by rather small average path lengths between any given two people. In other words, the number of steps one needs to take in order to reach anyone else in the network is rather small. In Milgram's study, the average path length was six people, hence "six degrees of separation."

Watts and Strogatz (1998) examined the small world phenomenon with greater methodological and mathematical scrutiny. They concluded that the reason for the short path length—short global separation—in many networks is related to high local clustering, or subgroups, within a network. For example, many people are associated with local organizations, such as houses of worship, clubs, or neighborhood watch groups. In these groups, the network is quite dense, as most people know one another. A few members of these groups, such as clergy members or community organizers, are also connected to other groups across a neighborhood, state, or even around the world. Through these people, group members can reach people in other groups in large, sparse, and decentralized networks.

Many networks that differ drastically from one another are shown to be small world networks. A few examples are the power grid of the western United States, the collaboration of film actors, the spread of epidemics and/or HIV infections, metabolism, sexual contacts, acquaintances "offline," and connections on.

SEE ALSO: Big Data, Analysis of; Big Data, Collection of (Social Media, Harvesting); NodeXL; Online Research Methods, Quantitative; Social Network Analysis

References

van den Bos, M. (2006). Hyperlinked Dutch-Iranian cyberspace. *International Sociology*, *21*(1), 83–99. doi:10.1177/0268580906059292

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

Carley, K. (1991). A theory of group stability. *American Sociological Review*, *56*, 331–354. doi:10.2307/2096108

Coleman, J. S. (1990). *Foundations of social theory*. New York: Belknap Press.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. In *SIGCOMM '99: Proceedings of the conference on applications, technologies, architectures, and protocols for computer communication* (pp. 251–262). New York: ACM. doi:10.1145/316188.316229

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

Himelboim, I., Golan, G. J., Suto, R. J., & Moon, B. B. (2014). A social networks approach to public relations on Twitter: Social mediators and mediated public relations. *Journal of Public Relations Research*, *26*(4), 359–379. doi:10.1080/1062726X.2014.908724

Himelboim, I., & Han, J. Y. (2013). Cancer talk on Twitter: Patterns of information seeking in breast and prostate cancer networks. *Journal of Health Communication*, *19*, 210–225. doi:10.1080/10810730.2013.811321

Himelboim, I., Jin, Y., Reber, B., & Grant, P. (2015). *Informing crisis communication preparation and response through network analysis: An elaboration of the social-mediated crisis communication model*. Paper presented at the annual Conference of the Association for Education in Journalism and Mass Communication, August 6–9, 2015, San Francisco, CA.

Himelboim, I., Smith, M., & Shneiderman, B. (2013). Tweeting apart: Applying networks analysis to explore selective exposure on Twitter. *Communication Methods and Measures*, *7*(3), 169–197. doi:10.1080/19312458.2013.813922

Jeong, H. Tombor, B. Albert, R., Oltvai, Z. N., & Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*, 65. doi:10.1038/35036627

Lerman, K., & Ghosh, R. (2010). Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In *Proceedings of 4th International AAAI Conference on Weblogs and Social Media (ICWSM-10)* (pp. 90–97). Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1509/1839/ (accessed November 4, 2016).

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444. doi:10.1146/annurev.soc.27.1.415

Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60–67. doi:10.1037/e400002009-005

Newman, M. (2000). The power of design. *Nature*, *405*, 412. doi:10.1038/35013189

Newman, M. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*, 025102. doi:10.1103/physreve.64.025102

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582. doi:10.1073/pnas.0601602103

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425–440.

Smith, M. Rainie, L., Shneiderman, B., & Himelboim, I. (2014). *Mapping Twitter topic networks: From polarized crowds to community clusters*. Retrieved from Pew Research Center website: http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/ (accessed November 4, 2016).

Tang, J. F., & Wu, D. J. (2009). Electron-cyclotron maser emission by power-law electrons in coronal loops. *Astronomy and Astrophysics*, *493*, 623–628. doi:10.1051/0004-6361:200810792

Wang, X., Jiang, T., & Ma, F. (2010). Blog-supported scientific communication: An exploratory analysis based on social hyperlinks in a Chinese blog community. *Journal of Information Science*, *36*(6), 690–704. doi:10.1177/0165551510383189

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, *34*(4), 441–458. doi:10.1086/518527

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*, 440–442. doi:10.1038/30918

Further reading

Himelboim, I., Smith, M., & Shneiderman, B. (2013). Tweeting apart: Applying networks analysis to explore selective exposure on Twitter. *Communication Methods and Measures*, *7*(3), 169–197. doi:10.1080/19312458.2013.813922

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. Oxford: Oxford University Press.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167–256. doi:10.1137/s003614450342480

Wasserman, S., & Faust, K. (1999). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wilson, F. (February 23, 2008). Slide, Zynga, and Rock You: Facebook apps follow power law. *Business Insider*. Retrieved from http://www.businessinsider.com/2008/2/slide--zynga--and-rock-you--facebook-apps-follow-power-law?IR=T (accessed February 7, 2017).

Itai Himelboim is an associate professor at the Department of Advertising and Public Relations and the Director of the SEE Suite: Social Media Engagement & Evaluation, at the Grady College of Journalism and Mass Communication, University of Georgia. His research involves examining social media networks, identifying pattern of information flow, key communities and users in influential positions, and drawing implications for political communication, international communication, advertising, and public relations.