# Social Network Analysis

THOMAS N. FRIEMEL
*University of Zurich, Switzerland*

## Subject and origins

Social network analysis (SNA) encompasses a large set of methodological, statistical, and theoretical approaches that are developed for the analysis of relational data. Relational data are given when the entities of interest (e.g., media content, people, organizations) are not assumed to be independent from each other and systematic information is collected about their relations. Hence, *a social network is defined as a set of nodes and their relations (ties)*. Including the relational information between the entities in the research design has a strong impact on every aspect of empirical research. This ranges from data collection to data management and the analytic approaches. How fundamentally different relational data are can be derived from the fact that most standardized methods put a lot of effort into eliminating and controlling relational effects. This starts with the widespread use of random sampling to ensure that all measures are independent from each other. Barton compares this to a "sociological meatgrinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it" (1968, p. 1). Whether such a meatgrinder makes sense has to be decided on a case by case level. However, since communication is relational by definition, SNA is the analytic approach of choice to study communication infrastructure, content, actors, and processes.

Social network analysis as a research field includes (i) structural intuition, (ii) relational data, (iii) visual representation, and (iv) mathematical and computational models (Freeman, 2004). Each of these aspects has their origin in the 19th century or earlier. The interest in kinship relations (as described in the Book of Genesis) and the visual representation of genealogical trees can be regarded as early examples of relational data and their visual representation. As a source of structural intuition it is often referred to early sociologists such as Comte, Maine, Simmel, and Tönnies. In the 1930s Jacob Moreno combined the four aspects and coined the term of sociometry that can be regarded as the precursor of today's SNA. In the past decades SNA has become increasingly relevant in social sciences while in natural sciences a similar tradition developed under the term *network science*. In the past decades the increasing computational power has stimulated the development of methodological tools. This was necessary since many statistical analysis of networks have to take a large number of possible network configurations into account that increases exponentially based on the number of nodes included.

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

DOI: 10.1002/9781118901731.iecrm0235

## Basic terms

As mentioned earlier, *a social network is defined by a set of nodes and ties*. The *nodes* are the instances that are analyzed in most other empirical approaches. In communication research the nodes are typically individuals using media, organizations that are producing, distributing, or regulating media, infrastructure through which communication is transmitted, or media content such as articles, statements, arguments, or websites. All information that is related to such a node (e.g., age and gender of media users or capitalization of media organizations) can be ascribed to these as *node attributes*. In a *sociogram*, which is the graphical representation of a network, these attributes can be visualized by node size, color, shape, or position in space. Any relation between the nodes can be defined as a *tie* which also can have attributes. Referring to the examples of nodes listed earlier, a tie can be a communicative act between two persons, a contract or equity participation between organizations, a syntactical relation between arguments, citations, or a link between websites. In the simplest case one distinguishes between existing and absent ties. However, one may also consider additional *tie attributes* such as information about the *direction* of a tie (e.g., information flow, influence, nomination for advice seeking), *strength* of a tie (e.g., weak vs. strong), and the *sign* of a relation (i.e., positive or negative).

The smallest network and basic element for any larger network is a *dyad* that is defined as two nodes and the tie(s) between them. In many instances it is sufficient to consider the direction of a tie disregarding its strength. Hence, the so called *MAN-typology* has become a standard to describe dyads. *M* stands for *mutual* and describes a dyad in which a directed tie is sent from both nodes to the respective other (Figure 1). *A* stands for an *asymmetric* relationship in which only one node sends an outgoing tie and the other receives an incoming tie. *N* as the third possibility is defined as a *null* relation in which no tie is present between the two nodes.

In many instances one is not only interested in dyads but also in higher order structures. Adding an additional node one speaks of a *triad* (three nodes and their ties). Disregarding the identity of the three nodes (i.e., isomorph nodes) and binary values of directed ties there are 16 different triads possible. Again the MAN-typology helps to label and distinguish these types. Hereby it is indicated how many ties of a specific kind are found in a triad. The first triad in Figure 2 is a 003-triad since it includes zero mutual, zero asymmetric and three null ties. For some types an additional letter is needed to distinguish similar triads. For 021-triads in which both ties are emerging in the same node a *D* indicates that the two asymmetric ties point *down* (type 4 – 021D), a *U* stands for *up* (type 5 – 021U), and *C* for a cyclic structure (type 6 – 021C). Finally, the 030T stands for a *transitive* structure (type 9) in which an indirect relation from $i \rightarrow j$ and $j \rightarrow k$ is closed by a direct tie $i \rightarrow k$.

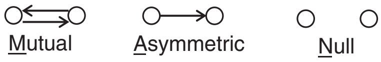


**Figure 1** MAN-typology of dyads.

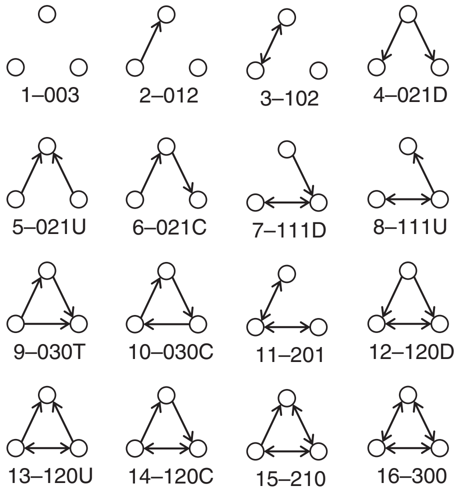


**Figure 2** MAN-typology of isomorph triads (Davis & Leinhardt, 1972).

## Data collection

In many instances data collection of social networks is more complicated than collecting unrelated data. Two main reasons for this are the exponential increase of complexity for increasing network size and the delineation of a network. First, the exponential increase of complexity is given by the fact that the number of possible ties between all actors in a network of size g is g*(g−1) in a directed network and g*(g−1)/2 in an undirected network. For example, increasing the network size from 10 to 100 nodes increases the number of all possible undirected ties from 45 to 4950. Second, in many instances the boundaries of a network are not given, but have to be defined by the researcher. If one is interested in interpersonal influence processes one would need to include all relevant nodes and ties. However, what is relevant is in many instances the subject of interest and cannot be defined beforehand. If sufficient resources would be available to researchers to collect data from a generously large number of nodes the problem described earlier of increasing complexity starts to mount. Ego-networks, snowball procedures, and whole networks are three approaches that handle these challenges in different ways.

The principle of *ego-networks* is to include additional information about the immediate network neighborhood for a node, which is referred to as ego (Crossley et al., 2015). Applied in surveys this method includes so-called name-generator and name-interpreter questions (Marsden, 2011). *Name-generator* questions are used to identify the alteri (Latin for "others") that are of relevance for ego. Methodological experiments have found that naming an upper limit of alteri has an effect on the number of nominated alteri. Hence, if the sheer number of nominations are of interest it is advisable not to give hints about an expected or maximum number. To avoid biases by a single name-generator question it is possible to generate names by multiple questions. If this results in too large a number of alteri, respondents sometimes are asked to select a specific number of most relevant alteri before proceeding with the questionnaire. However, the selection of the most relevant alteri can become rather complicated, which is why it is often necessary to conduct the survey in the form of a personal interview instead of a written questionnaire. In most instances a range of *name-interpreter* questions are subsequently asked regarding the attributes of the alteri, the ties between ego and the alteri, and the ties between the alteri. One of the major advantages of collecting ego-network data is the possibility to combine it with random sampling which makes it compatible with other research designs such as representative surveys. Nevertheless, there are two major limitations to ego-networks. First, in most instances it is not feasible or possible to collect data about absent ties (e.g., alteri whose names are unknown cannot be nominated). However, absent ties can be of major relevance as the literature on structural holes demonstrates. Second, it is found that not only the immediate neighborhood is of relevance but also indirect relations and the overall structure of a network.

One possibility to overcome these limitations is to extend the network by means of a *snowball procedure*. Hereby, all nominated alteri are treated in a second step as egos as well in order to collect information about their ego-network, and so on. This approach requires many resources as the number of egos increases by the power of average nominated alteri. Furthermore, it can become an almost endless endeavor since the *small world structure* often found in social networks would make it necessary to include the entire population unless the overall network is divided in disconnected components.

*Whole networks* solve the problem of the two previous approaches by including all nodes based on a given criteria. Ideally this criterion would be "the relevance" of the nodes for the respective research question. However, this may be too large a number or the subject of investigation itself. Therefore, pragmatic decisions have to be made based on known information and predefined criteria for belonging to a specific network. If the entire population is too large a kind of cluster sampling can be applied. The only prerequisite is that the identity of all nodes must be known beforehand. This enables the use of so-called roster questionnaires in which a list of all nodes is presented. In this instance questionnaires typically focus on the ties to the alteri and less on attributes of the alteri, because data on node attributes are collected by asking the respective nodes directly. An exception to this is when the researcher is interested in the influence of perceived node attributes rather than effective attributes. As is known from questionnaire design in general, long lists are fatiguing for participants. Therefore, lists of 50 or more alteri can cause increasing nonresponse and refusal to participate. If the researcher is able to overcome these challenges by defining meaningful networks of reasonable size this approach certainly provides the richest data.

The relevance of secondary data as a source for social network analysis has increased in the past decades due to the digitalization of interpersonal communication and mass media. This includes digital traces of communicative behavior, the analysis of link structures on the Internet (e.g., webometrics), but also semantic analysis of written language. In addition to these digital data, official registers also may be sources of relational information. In most instances these networks are treated as whole networks since the information encompasses a defined set of nodes and their relations. Finally, social network data may also be obtained from observations. Depending on the design, observations can be conceptualized as ego-networks, snowball procedures, or whole networks.

## Data management

In SNA two general types of networks can be distinguished based on the number of node sets which are included. The most intuitive way is to think of a network with a single set of nodes. In these *one-mode networks* (often called *unimodal networks*) every node may be connected to any other node. An example would be a group of persons who are talking to each other about media content (Figure 3a). In two-mode networks (also called *bipartite* or *bimodal networks*) two distinct node sets are included whereby ties are only permissible between nodes of the other set but not among nodes of the same set (Figure 3b: circles and squares as different node sets). An example is a network of people who are using different media content (e.g., TV programs). Hereby, people are only related to TV programs but not directly to each other nor are the TV programs. Of course, the information about media preferences could also be conceptualized as node attributes. However, most analytic approaches only allow inclusion of very few node attributes and also for the purpose of visualization it is difficult to include more than three variables or variables with many values.

If multiple networks of a specific type are analyzed at a time one speaks of a *multiplex network*. In a multiplex one-mode network multiple relations are considered simultaneously between the nodes (e.g., use of different communication channels among a set of persons or organizations (Figure 3c: full and dashed lines as separate types of ties). If two or more networks of different type are combined one speaks of a *multilevel network* (Lazega & Snijders, 2016). An example of a multilevel network would be the combination of a one-mode network of conversation ties among friends and a two-mode network of these persons and their preferred media content

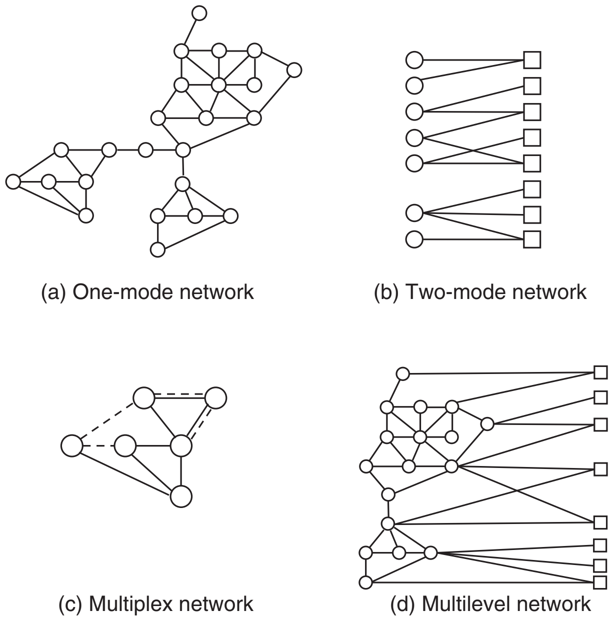


**Figure 3** General types of networks.

(a) Matrix

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | – | 2 | 0 | 0 | 1 |
| B | 2 | – | 0 | 1 | 0 |
| C | 0 | 0 | – | 1 | 1 |
| D | 0 | 0 | 0 | – | 0 |
| E | 0 | 0 | 0 | 1 | – |

(b) Edgelist

A B 2
A E 1
B A 2
B D 1
C D 1
C E 1
E D 1

(c) Sociogram

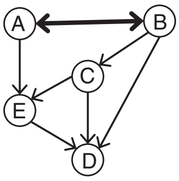


**Figure 4** Storage and visual representation of network data.

(Figure 3d). It is important to note that the term multilevel does not mean that the networks have a nested structure.

Relational data can be stored in two different ways: as a matrix or as edge lists. In a *matrix* all ties of a node are represented by a row and a column. A one-mode network necessarily forms a square matrix while a two-mode network may also result in a rectangular matrix. The convention is to read the matrices from rows to columns. Hence, in a row all outgoing ties of a node are listed while columns report all incoming ties. In most instances reflexive ties (i.e., a tie from a node to itself) are meaningless and therefore the diagonal of the matrix remains empty. In an undirected network the matrix is symmetrical (i.e., the transposed matrix $A'$ is identical to $A$). Figure 4(a) shows an asymmetric valued network with five nodes and tie values of 0, 1, and 2 indicating absent, weak, and strong directed ties. Storing network data in a matrix allows an easy and intuitive inspection of the data. However, in cases of large networks the necessary storage space increases exponentially (cf., the earlier illustrated calculation of the number of ties given a number of nodes). Since most large networks are rather sparse, a matrix is a very inefficient method of data storage. Alternatively, network data can be stored as *edge lists* in which only the non-null relations are included (Figure 4b). This list of all M- and A-dyads only includes the node-ID (in the order of source and target node). In a weighted network the list also includes an information about the tie strength. For example “A B 2” tells us that there is a tie from A to B with tie strength 2.

## Visualization

Visualization of network data as a sociogram is often the first and the last step in SNA. As a first step into analysis it provides an overview of the data and its structure, and may guide the statistical analysis. Additionally, sociograms may also help to report the complex nature of the findings after analysis have been finalized. A large number of algorithms are available to optimize sociograms. One should be aware of the fact that many algorithms include random seeds and are stochastic so that slightly divergent visualizations may result in each run. As a rule of thumb a good visualization places nodes in meaningful special relation to each other (e.g., dependent on shortest path) and minimizes tie crossing to enhance readability. Figure 4(c) shows a sociogram that additionally puts emphasis on the hierarchical structure of the network by placing mutual ties on the same y-coordinate and arranges the other nodes in a way that all arrows point in one direction.

Due to their intuitive appealing nature, sociograms are also suitable to be discussed with lay people. Especially in *qualitative approaches* it is also common to let people draw a kind of sociogram in an interview or to let people elaborate and interpret sociograms (for an overview of qualitative and mixed method applications, see Domínguez & Hollstein, 2014).

## Descriptive measures

Descriptive measures for social networks can be calculated on the level of the entire network as well as subsets or single nodes. The most general network descriptive refers to the number of nodes and ties in a network. The number of nodes is equal to the network *size* (*g*). Of course the number of ties between the nodes is limited by the network size and is not meaningful to be compared across networks of different size. Therefore, the number of ties is often put in relation to the maximum number of possible ties which gives the *density* of a network (*d*). But also density should not be compared if network size varies too greatly. Resources to create and maintain network ties are often limited and the number of possible ties increases exponentially. Hence, smaller networks are often found to be denser than larger networks, which inhibits a meaningful comparison just by means of density. If a network contains only non-null ties it is called a *complete network*.

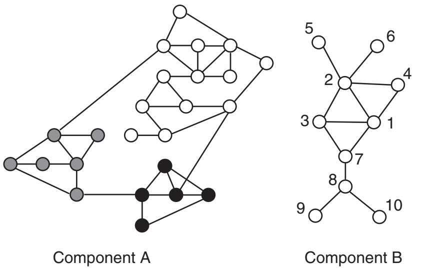
One of the most fundamental insights of SNA is that ties in a network are not randomly distributed. Typically there are subsets of nodes that are more densely connected than other subsets. If a subset of nodes is not connected to the rest of the network one speaks of a *component*. Figure 5 shows a network with two components. Since the detection of cohesive subgroups is of vital interest for many researchers, a large number of algorithms are available for their detection. The basis for subgroup detection may be the maximum diameter *n* within a subgroup (e.g., *n-cliques*) or the number *k* of adjacent alteri (e.g., *k-plexes* or *k-cores*).

Another measure to describe the overall structure of a network is its centralization. *Network centralization* can be defined as the ratio between the node with the highest centrality and the theoretical maximum given a specific network size. Alternative procedures are based on the variance of observed centrality measures of all nodes. Both definitions refer to node centralities. In fact, node centralities are much more used and reported. The most often applied centrality measures are degree centrality, closeness centrality, and betweenness centrality. This is not because they cover all possible applications but rather because they are intuitive to understand. With respect to communication networks one has to be very cautious to use the right measure and should restrain from simply calculating all three measures. As Borgatti has shown (2005) for flow processes in general, there are various types of flow (e.g., gossiping) for which no suitable centrality measure exists. Nevertheless the three measures will be briefly described to provide a general idea of their functionalities.

*Degree centrality* ($C_D$) is a very simple measure which counts the number of direct relationships of a tie. In an undirected network this is the number of ties to other nodes which are non-null. In a directed network indegree ($C_{Di}$) for incoming ties and outdegree ($C_{Do}$) for outgoing ties are distinguished. If the network data is represented as a matrix, $C_{Di}$ can simply be calculated as the sum of the column and $C_{Do}$ as the row sum. Depending on the kind of ties (e.g., advice seeking) indegree centrality can be interpreted as a measure of popularity or prestige, and outdegree centrality as a measure of activity. In component B of Figure 5 node 2 has the highest degree centrality ($\mathrm{C_D} = 5$).

In many instances the researcher is not only interested in the number of direct relations a node has but also in its structural embeddedness in the network. For example, one might be interested to learn how many steps a node has to take to reach all other nodes in the network. The idea is that the fewer steps that have to be made, the more central a node is. This so-called *closeness centrality* ($C_C$) is calculated by the sum of the inverses of the distances from a node to all other nodes. The usual interpretation is that people with a low closeness centrality in a network are very efficient at reaching all other nodes and are good starting points for a fast-spreading diffusion. If only the nodes of component B are regarded, node number 1 has the highest closeness centrality. Here reported as a standardized value $\mathrm{C_{C'}} = 0.56$.

The third centrality measure, *betweenness centrality*, provides an insight into how much control a node has by being part of the relation between others. From a structural point of view this is the case if two nodes are only indirectly related to one another.



Node centralities for component B

| Node | $C_D$ | $C_{C'}$ | $C_{B'}$ |
|---|---|---|---|
| 1 | 4 | 0.56 | 0.29 |
| 2 | 5 | 0.50 | 0.43 |
| 3 | 3 | 0.53 | 0.17 |
| 4 | 2 | 0.43 | 0.00 |
| 5 | 1 | 0.35 | 0.00 |
| 6 | 1 | 0.35 | 0.00 |
| 7 | 3 | 0.53 | 0.50 |
| 8 | 3 | 0.43 | 0.42 |
| 9 | 1 | 0.31 | 0.00 |
| 10 | 1 | 0.31 | 0.00 |

**Figure 5** Subsets and node centralities in a network.

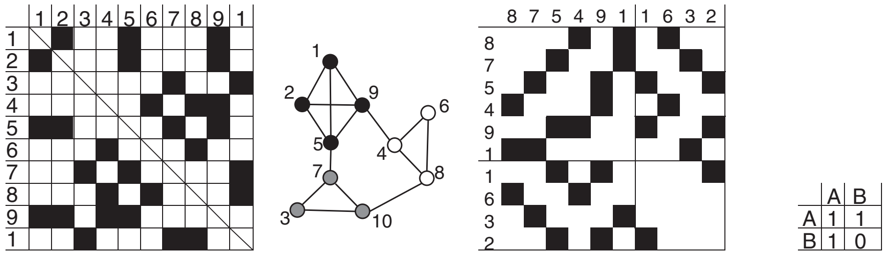
In such an instance the node in between (linking the two) has full control. Hence, the betweenness centrality of a node ($C_B$) is calculated as the sum of shortest paths (geodesics) between all possible pairs of nodes in a network that lead through that node. In the simplest case with unweighted ties it is assumed that geodesics of the same length are used with the same probability. Hence, if two alteri are connected by multiple geodesics ($g_{ij}$) of the same length ego will only be assigned the respective proportion ($1/g_{ik}$). The node with the highest standardized betweenness centrality in component B of Figure 5 is node 7 ($\mathrm{C}_{\mathrm{B'}} = 0.50$).

As can be seen from this example each measure guides our attention to another node in the network. Therefore, it is crucial to choose the right measure to avoid incorrect analysis and interpretation. Both the closeness and the betweenness centralities are good examples of why we should be careful when applying these measures to communication processes. Unless there is a good argument that the flow always follows the geodesics (i.e., the shortest path) these measures can be misleading. *Information centrality* ($C_I$) is a derivation of betweenness centrality and is less restrictive on the nature of the flow processes. It takes all possible paths between two nodes into account (also non-geodesics) and weights them by their length. Therefore, short paths still contribute more to a high centrality value but all other options are not neglected. Intriguing by its idea but a bit more complicated to calculate is the *eigenvector centrality ($C_E$)*. The principle idea is that a node is central if it is connected to central nodes.

To make the centrality measures comparable across different network sizes all of them can be *standardized* by taking the number of nodes into account. Furthermore, most of the centrality measure may also be calculated for ties, which then is called *tie centrality*.

## Analytic approaches

In most instances the information provided in a network is too complex to be interpreted at once. Beside the descriptives explained earlier there are various techniques to reduce the complexity by identifying typical or generalized positions in a network. If two nodes have exactly the same set of ties to other nodes they are *structurally equivalent* and can be thought of as having the same position in a network. However, this requirement is rather restrictive and is not often met in empirical research. A more relaxed definition is *regular equivalence* where it is not necessary to be tied to the same alteri but to a similar pattern of alteri. Simply speaking, two journalists within the same organization can be structurally equivalent since they work for the same editor. Two journalists in two distinct organization are a regular equivalent since they both have an editor they work for (but not the same). If the network data are stored as a matrix (Figure 6a) structural equivalence can be found by identifying rows and columns with the same entries. If one sorts the matrix by placing identical/similar columns and rows next to each other (permutation) a *blockmodel* results (Figure 6c). This example illustrates that the quest for regular equivalent positions is not guided by clique membership. The black, gray, and white cliques in Figure 6(b) are not forming the blocks. The blocks are built by separating the nodes with clique internal relations only (1, 6, 3, and 2) and



(a) Original matrix (b) Sociogram (c) Permuted matrix (d) Image

**Figure 6** Blockmodeling.

those building bridges between cliques. In a prototypical blockmodel the matrix is separated like a chessboard into square areas (i.e., blocks) with high similarities within the block and low similarity across the blocks. In further steps one can calculate the density within each block and by applying a threshold value reducing the matrix to a so-called binary *image matrix* (Figure 6d). In such an image matrix the original nodes are condensed to one node per block and one can see more abstract patterns of the network (i.e., on the level of positions instead of individuals).

Identifying more general positions in networks by blockmodeling (Doreian, Batagelj, & Ferligoj, 2005) can also help to compare different networks. However, if one is interested in a single value for how similar two networks are the *quadratic assignment procedures* (QAP) is the method of choice. The idea is to have an equivalent measure to a correlation coefficient but since network data are not independent from each other and sparse networks are causing zero-inflation problems, a special procedure has been developed. QAP makes use of permutation to calculate a null distribution and significance values. This means that the node labels of one network are randomly permuted before the correlation of the two networks is calculated again. Of course no substantial relation exists between the two networks after permutation. Repeating this procedure provides a null-distribution which can be used for significance tests. Similar to regression analysis there is also a multivariate extension of this procedure (*MR-QAP*) which allows to test the relation of several networks simultaneously.

For several decades the *triad census* was used to describe the composition of networks. Hereby, the idea is that the frequency of different triadic configurations (see the MAN-typology earlier) provides insights into the building blocks which make up the overall structure. For example, the up and down oriented triads (type 4, 5, 8, 12, and 13) as well as transitive triads (type 9) can be used as indicators for hierarchical structures whereas cyclical triads (type 6, 10, and 14) and mutual relations are indicators for non-hierarchical networks. Today the triad census is not used anymore since it does not take dependencies (e.g., nesting) between the various triads and higher order configurations sufficiently into account. However, the idea that the analysis of small configurations helps to understand complex network structures is the basis of the currently most elaborated approaches (ERGM and SAOM, which will be introduced in the next paragraphs).

*Exponential random graph models* (ERGM) are a class of statistical models to analyze and describe social network structures (Lusher, Koskinen, & Robins, 2013).

Based on theoretical arguments, a model with numerous parameters (e.g., the triadic configurations mentioned earlier) is fitted to the observed network. For each model parameter an estimate and standard error is calculated and goodness-of-fit parameters for the model can be used to decide on the overall model fit. Hence, this approach allows to test hypotheses about the underlying principles of network formation. Besides selected triadic configurations these models typically include the reciprocity parameter. Since reciprocity is very fundamental for most social networks this parameter is rather used as a control variable and not as a hypotheses to test against null. If node attributes are included in the analysis it is also possible to test the tendencies towards homophily. *Homophily* (or *network autocorrelation*) is the tendency that nodes with similar attributes are more likely to be tied to each other. For example, fans of a specific musician might be more likely to become friends. If attributes are visualized by node colors this results in sociograms such as Figures 5 or 6(b). ERGMs are available for all types of social networks (i.e., directed, bipartite, multiplex, and multilevel networks). Also longitudinal data can be analyzed which is referred to as *temporal exponential random graph models* (TERGM). While the temporal extension of ERGMs is rather new there is already a longer tradition to analyze dynamic processes with actor-oriented modeling.

Also *stochastic actor-oriented modeling* (SAOM) builds on the idea of small configurations which are used to model the emergence of larger network structures (Snijders, 2011). In contrast to ERGMs this approach is not tie based but assumes the *nodes* to make decisions about the formation of ties and their attributes. This possibility makes SAOM the method of choice for the analysis of co-evolution processes of network structure and node attributes over time. This is of major interest, if one is not only interested in the fact that there is network autocorrelation in a network but also in its underlying dynamic. Hereby, two very distinct processes have to be distinguished, both leading to network autocorrelation: social influence and social selection. *Social influence* is the process that is assumed to account for diffusion processes. That is, influence is exerted along existing ties and triggers a person to adapt to the behavior suggested by a related alter (e.g., opinion leader). However, assuming the network structure to be stable may be shortsighted. In many instances it is more plausible that network structures are changing over time as well and that the node attributes may be relevant for tie formation and dissolution. Hence, node attributes should not only be thought of as dependent variables but also as independent variables exerting an influence on the structure. This reversed perspective addresses the issue of *social selection*. For example, the use of specific media technologies or media content might be a prerequisite to get in contact and become acquainted. Hence, similar media preferences among befriended people are not necessarily the outcome of social influence but may also emerge due to social selection. Surprisingly this analytic approach has only recently been applied to communication research even though the respective research tradition of opinion leadership and diffusion research can be regarded as one of the antecedes of the discipline. Nevertheless, this conceptual and methodological extension is of major importance, since both selection and influence are likely to be overestimated if they are not controlled for each other (Friemel, 2015). Similar to ERGM it is possible to analyze bipartite, multiplex, and multilevel networks. For example, the co-evolution of media use and media-related conversation can either be analyzed as a one-mode network of conversational ties and including media use as node attribute (Friemel, 2012), or as a multilevel network in which media use is included as a bipartite network (Friemel, 2015).

The currently available software for TERGM and SAOM are designed to analyze network data that represent observations at discrete time points. Hence, the typical research design includes several snapshots of networks and the models try to estimate parameters that are likely to be relevant for the network change between two time points. For empirical research it is crucial to choose the right time interval between the observations because both too little and too much change between two observations contain insufficient information to fit a suitable model. The *jaccard index* can be used as a measure for network change and the respective software manuals guide the user to meaningful thresholds. Both TERGM and SAOM assume that network change is in fact a continuous process and the use of discrete observations is rather due to practical reasons of data collection. However, since more data become available that document communicative behavior on the level of single actions (i.e., digital traces) new analytic approaches will become of increasing relevance in the future to handle these event networks. The networks are called *event networks* because every event of network change is observed. If one wants to see the bigger picture of the whole network it is necessary to sum all events in a specific time frame. In fact this is what typically has been done before analyzing event network data with the aforementioned approaches. However, it is kind of pointless to sum the network changes and then try to model exactly these changes in a computational intensive process.

A further application of the analysis of small configurations is to use these insights to impute missing values. Under the assumption that missing values are random and not systematic, the missing information for nodes and ties (e.g., of people not participating in a survey) can be imputed based on the structural insights of the observed network.

## Ethical issues

Given the analytic power of social network analysis researchers have to be especially aware of ethical issues. Besides the established good practices in social science, an ambivalent discussion exists about the use of relational data that are linked to nonparticipants in a whole network design. In this case it can be argued that persons (or any other node type) who do not agree to participate in a study should not be included in the node set. From an opposite standpoint it can be argued that information about a tie to other nodes is not the property of the alteri but of ego. In fact one has to be aware that this information may not even be accurate but solely a subjective perception of ego. Given the assumption that the alteri cannot be identified at a later stage and do not encounter any negative effect of the research, it is therefore widely accepted to include information about nonparticipants.

Collecting data of whole networks the nodes have to be identified at a specific point. This often causes privacy concerns by ethical review boards, data protection agencies, and the participants. If it is not possible to provide researchers with a list of the members of a whole network (e.g., pupils in a school, employees in a company, or citizens in a community) it is advisable to compile the list of participants in a separate step prior to the collection of network data to grant data protection and informed consent. However, also under this procedure the collection of relational data can never be made anonymously. The only possibility is to guarantee *confidentiality* and to *anonymize data right after collection*. However, research designs such as panel surveys require storing key tables with the real identity and a de-identified ID to link answers of different panel waves. In this instance it is advisable to store the key tables and the data separately. For example, the key table can be retained by a third party, independent from the researcher and the participants.

SEE ALSO: Big Data, Analysis of; NodeXL; Online Ethnography; Panel Research Methods; Social Network Analysis (Social Media)

## References

Barton, A. H. (1968). Bringing society back in: Survey research and macro-methodology. *American Behavioral Scientist*, *12*(2), 1–9. doi:10.1177/000276426801200201

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. doi:10.1016/j.socnet.2004.11.008

Crossley, N., Bellotti, E., Edwards, G., Everett, M. G., Koskinen, J., & Tranmer, M. (2015). *Social network analysis for ego-nets*. Los Angeles: SAGE.

Davis, J. A., & Leinhardt, S. (1972). The structure of positive interpersonal relations in small groups. In J. Berger, M. Zelditch, Jr., & B. Anderson (Eds.), *Sociological theories in progress*, Vol. 2 (pp. 218–251). Boston: Houghton Mifflin.

Domínguez, S., & Hollstein, B. (Eds.). (2014). *Mixed methods social networks research: Design and applications*. Cambridge, UK: Cambridge University Press.

Doreian, P., Batagelj, V., & Ferligoj, A. (2005). *Generalized blockmodeling*. Cambridge, UK: Cambridge University Press.

Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science*. Vancouver: Empirical Press.

Friemel, T. N. (2012). Network dynamics of television use in school classes. *Social Networks*, *34*(3), 346–358. doi:10.1016/j.socnet.2011.08.002

Friemel, T. N. (2015). Influence versus selection: A network perspective on opinion leadership. *International Journal of Communication*, *9*, 1002–1022.

Lazega, E., & Snijders, T. A. B. (Eds.). (2016). *Multilevel network analysis for the social sciences: Theory, methods and applications*. Cham, Switzerland: Springer.

Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge, UK: Cambridge University Press.

Marsden, P. V. (2011). Survey methods for network data. In J. Scott & P. J. Carrington (Eds.), *The SAGE handbook of social network analysis* (pp. 370–388). London: SAGE.

Snijders, T. A. B. (2011). Network dynamics. In J. Scott & P. J. Carrington (Eds.), *The SAGE handbook of social network analysis* (pp. 501–513). London: SAGE.

## Further reading

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

Robins, G. (2015). *Doing social network research: Network-based research design for social scientists*. London: SAGE.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

**Thomas N. Friemel** (PhD 2008, University of Zurich), is professor of communication and media research at the University of Zurich, Switzerland. His research interests are media use and media effects with a special emphasis on the social context, audience research, social network analysis, health communication, and campaign evaluation. He has organized several conferences and workshops on social network analysis and has edited books and journal special issues on that topic.