A Critical Review of Centrality Measures in Social Networks

Social networks are currently gaining increasing impact especially in the light of the ongoing growth of web-based services like facebook.com. A central challenge for the social network analysis is the identification of key persons within a social network. In this context, the article aims at presenting the current state of research on centrality measures for social networks. In view of highly variable findings about the quality of various centrality measures, we also illustrate the tremendous importance of a reflected utilization of existing centrality measures. For this purpose, the paper analyzes five common centrality measures on the basis of three simple requirements for the behavior of centrality measures.

The Authors

Dipl.-Math. oec., M.Sc.
Andrea Landherr (✉)
Dipl.-Math. Bettina Friedl
Dr. Julia Heidemann
FIM Research Center
Finance & Information Management
University of Augsburg
Universitätsstraße 12
86159 Augsburg
Germany
andrea.landherr@wiwi.
uni-augsburg.de
bettina.friedl@wiwi.uni-augsburg.de
julia.heidemann@wiwi.
uni-augsburg.de

1 Introduction

Fundamental developments in information technology (IT) and especially the enormous growth of the Internet are essential drivers for the increasing global interconnectedness of companies and individuals. The targeted use of powerful IT in this process significantly facilitates the interaction of actors at different locations and information exchange in real time. In this context, services subsumed under the term Web 2.0, such as wikis, blogs, or online social networks in which individuals are connected to each other and share news, experiences, and knowledge, increasingly gain importance. The U.S. market researcher Hitwise, for instance, reported in March 2010 that – as measured by the number of visits – the online social network facebook.com has replaced the search engine giant google.com as the most visited U.S. website (Hitwise 2010). Moreover, according to a recent study by the Nielsen Company, about 66% of global Internet users actively use these new social communities each month (The Nielsen Company 2009, p. 2). Given this development, it is not surprising that web-based social networks have attracted the interest of many companies since a majority of their customers now regularly use these services and, in this way, exchange views on products and services (De Valck et al. 2009, p. 185).

The constitutive feature of social networks are the relationships between network members and hence the network structure induced by the mutual connections (Zinoviev and Duong 2009). This interconnectedness of actors – i.e., their structural integration into the network – significantly influences their communication and interaction, and therefore holds valuable information for companies with regard to various corporate issues. Concerning viral marketing, for instance, the integration of well-connected actors is of considerable importance in order to attract the attention of the largest possible audience to a brand, a product, or a campaign (Kiss and Bichler 2008, p. 233; De Valck et al. 2009, p. 187). In product development and in particular in the identification of trends, the integration of members who take a central position within their network is also of great advantage since these actors have access to information about a multitude of other actors (De Valck et al. 2009, p. 185).

The successful implementation of this exemplary list of business related issues and similar ones requires the identification of those members (*key persons*) who are structurally very well integrated into a social network. This identification is not only necessary for the success of business decisions, but is particularly important in view of time and budget constraints. In this context, it appears appropriate to take recourse to the social network analysis (SNA), which has already developed and discussed a variety of centrality measures (CM) for the quantification of the interconnectedness of actors in social networks. Therefore, the aim of this paper is (1) to show the current state of research with regard to CM in social networks and (2) to illustrate the enormous importance of a reflected utilization of existing CM in view of highly variable findings on the quality of various CM in SNA.

The paper is organized as follows: In Sect. 2, we first present the state of research on CM in social networks. On this basis, we exemplarily formulate three simple general requirements for CM in

Sect. 3, which are used in Sect. 4 to analyze five commonly applied CM from the literature on SNA. The paper concludes with a summary of results and an outlook in Sect. 5.

2 Social Networks

2.1 Structure and Characteristics of Social Networks

Based on Valente (1996), the term social network is understood in this paper as a "pattern of friendship, advice, communication or support" (Valente 1996) between individual members or groups of members within a social system (cf. also Burt and Minor 1983; Knoke and Kulinsik 1982; Scott 1991; Wellmann 1988). Usually, a common goal, interest, or need of the various persons involved constitutes the unifying element of such a network. Web-based social networks use the infrastructure of the Internet to provide basic functionality for identity management (i.e., the presentation of oneself), relationship management (i.e., managing one's own contacts or maintaining the network), and visualization of profiles and networks (Koch et al. 2007). In this way, the community feeling of the actors, which is a central characteristic of such networks, can be achieved also without their direct physical presence (Heidemann 2010). The features for relationship management and in particular the management of contacts via contacts lists in web-based social networks especially enable maintaining casual acquaintances which are often not kept alive in real life.

Taking a structural point of view, we can model the relationships within a social network as a graph G with a set V_G of nodes and a set E_G of edges between these nodes. The set V_G represents the members of the social network, while the set E_G refers to the relationships between them and thus describes social ties and interaction potentials between the actors (Sabidussi 1966; Wassermann and Faust 1994). The resulting network structure of a social network can also be represented by a matrix $A = (a_{ij}) \in \{0; 1\}^{nxn}$. The entry a_{xy} of this so-called adjacency matrix is 1, if $(x, y) \in E_G$ holds. Otherwise, the entry a_{xy} is 0. **Figure 1** illustrates an example of the representation of a social network as a graph.

Regarding the characteristics of social networks, which highly differ from biological or technical networks
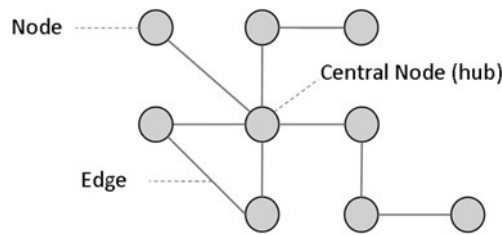

Fig. 1 Example of a social network

(Newman and Park 2003), we can draw on a variety of existing knowledge from SNA (for an overview cf. Wassermann and Faust 1994). Social networks can, e.g., be classified as to whether there are one-sided relationships ((un-)directed network) or different relationship intensities ((un-)weighted network) (Wassermann and Faust 1994, p. 44). Furthermore, the American psychologist Stanley Milgram realized already in the 1960s that every person is connected to everyone else in the world via a surprisingly short chain of on average six contacts (Milgram 1967). This so-called "small world phenomenon", which is also known under the heading "six degrees of separation", can be observed both in the offline and in the online world (e.g., Dodds et al. 2003; Leskovec and Horvitz 2008; Travers and Milgram 1969). Given these findings we can assume that the majority of actors in a social network form a single connected graph. In addition, there may also be other smaller groups of members which can be analyzed separately in terms of interconnectedness, as well as isolated members without any relationship to other actors (Kumar et al. 2006; Mislove et al. 2007). The following comments, however, focus on social networks or those subgraphs in which each person has a relation to every other person in a direct or indirect way. Furthermore, numerous studies show that social networks are mostly scale-free networks in which the number of contacts is not distributed homogeneously across all members (e.g., Barabási and Bonabeau 2003; Ebel et al. 2002; Kumar et al. 2006; Mislove et al. 2007). Instead, such networks are made up of many scarcely interconnected and only some highly integrated members – so-called hubs (see **Fig. 1**). These hubs act as a link between individual groups of strongly interconnected members. Overall, the interconnectedness of different individual members of a social network generally differs significantly. In order to identify those actors that play a central

role in a social network, it appears appropriate to make use of CM that have been developed within SNA. In the following section we therefore present the state of research as regards CM in social networks.

2.2 Interconnectedness and Centrality Measures in Social Networks

Since many years, the interconnectedness of actors in social networks has been a central issue of SNA. The discussion is often limited to undirected, unweighted social networks in a simplifying way. However, even for these relatively simple graphs there is no uniform understanding of an actor's centrality in a social network (Borgatti and Everett 2006, p. 467). Instead, some very different concepts and context-specific interpretations of the centrality of a node exist (Borgatti and Everett 2006, p. 467) that may result from different objectives for the use of CM. In the following, we therefore first present four basic concepts of centrality. In the simplest case, the number of a network member's direct contacts is a useful indicator of centrality. The advantage of this interpretation of an actor's centrality, with *degree centrality* (DC) as its standard representative (Nieminen 1974; Shaw 1954), is the fact that the results are relatively easy to interpret and communicate. A second approach is based on the idea that nodes that have a short distance to other nodes and consequently are able to disseminate information on the network very effectively, take a central position in the network (Beauchamp 1965; Sabidussi 1966). A representative of this approach is *closeness centrality* (CC), where a person is seen as centrally involved in the network if he requires only few intermediaries for contacting others and thus is structurally relatively independent. Accordingly, the calculation of this CM includes the length of the shortest paths to all other actors in the network. Further developments of CC even use the length of all paths between

the actors for the calculation (e.g., Newman 2005). A third approach, however, equates centrality with the control of the information flow which a member of the network may exert, based on his position in the network. Here, it is assumed implicitly that the communication and interaction between two not directly related actors depends on the intervening actors. The most prominent representative of this concept is *betweenness centrality* (BC), where the determination of an actor's centrality is based on the quotient of the number of all shortest paths between actors in the network that include the regarded actor and the number of all shortest paths in the network (Bavelas 1948; Freeman 1977; Shaw 1954). The common characteristic of all networking concepts presented so far is that only little or no attention is paid to indirect contacts, meaning they are not or only indirectly included in the quantification of an actor's centrality. This is where the so-called influence measures come into play. These CM consider actors to be centrally involved in the network if their directly connected network members are in relationship with many other well-connected actors. Some of the best known of these recursively defined CM are the eigenvector centrality (EC) (Bonacich 1972), the CM by Bonacich (1972), and the CM by Katz (1953). Besides these representatives of the four basic concepts of centrality, a plethora of other CM has been defined over the years (see, e.g., Bonacich and Lloyd 2001; Freeman et al. 1991; Lee et al. 2009; Rousseau and Zhang 2008) which, e.g., enable the integration of edge weights or of directional connections or are suitable for specific applications and network types. Usually, these CM represent modifications or enhancements of the already discussed CM and thus are not elaborated in more detail in this article. For the mathematical calculation of each CM, different algorithms have been developed which may vary significantly in terms of complexity. While the DC only requires to count the direct contacts of the n nodes in the network (complexity of $O(n)$), the complexity of BC in unweighted graphs amounts to $O(n \cdot m)$ (Brandes 2001), where m is the number of edges in the network.[1] At the same time, this algorithm allows the calculation of other distance-based CM, such as CC, for which Okamoto et al. (2008) discuss other algorithms and

heuristics. According to Kiss and Bichler (2008), the complexity of calculating the EC is $O(n^2)$, whereas in case of Katz's CM the inverting of the adjacency matrix initially induces a complexity of $O(n^3)$. However, this complexity can be reduced by applying the algorithm of Coppersmith and Winograd (1990) to $O(n^{2.376})$.

Starting from the definition of different CM, a lively discussion of the characteristics and the robustness (e.g., in case of incorrect or incomplete data on the network structure) of different CM has arisen. Accordingly, on the one hand numerous empirical studies exist that discuss the application of CM using different real or simulated networks. On the other hand, we can also find a great deal of research which – starting from the concept of different CM – derives conclusions about their properties or suitability for different applications. **Table 1** provides an overview of relevant contributions which are classified according to the dimensions of focus (empirical vs. conceptual), approach, and analyzed CM.

In the field of applying CM to real or simulated networks, e.g., Bolland (1988) discusses the robustness of DC, CC, BC, and the CM by Bonacich in random and systematic variation of the underlying network structure. This analysis shows that BC is generally very unstable with regard to the variation of the network structure. In contrast, for DC and CC the centrality score usually varies only slightly in case of a random or systematic change of the underlying network structure. However, according to the studies of Bolland (1988) the CM by Bonacich is the least sensitive one in terms of a random or systematic variation of the network structure. A further contribution to the discussion on the robustness of different CM is provided by Borgatti et al. (2006) who first define four different types of error (adding or deleting an edge or a node) and then compare the CM DC, CC, BC, and EC with regard to these different types of errors. The main finding of the study is that the four CM react very similarly to manipulations of the network structure, with BC performing slightly worse than the other three. Frantz et al. (2009) extend these investigations by differentiating five network topologies. They conclude that the robustness of the four CM also depends on the particular topology of the

network. Furthermore, Costenbader and Valente (2003) also analyze the stability of different CM in presence of incorrect or incomplete information on the structure of a network (e.g., for the analysis of a sample of the network). In addition to the classic CM DC, CC, BC, EC, and the CM by Bonacich, their investigation includes two more CM and they also extend their analysis to directed graphs. For undirected, unweighted social networks they come to the conclusion that the centrality scores of individual actors, which have been determined based on a sample of the overall network, have the highest average correlations with the centrality scores of the individual actors in the overall network in the case of EC (before DC, CC, and BC). Here, BC, however, comes off less successfully than the other three CM, indicating a fundamentally distinct conception of centrality for this CM (Bolland 1988). The investigations on the robustness of CM concerning the variation of the network structure discussed here are very important since the connections between actors that are considered for an analysis of social networks usually only present a distorted picture of the real social network for both the offline and the online context. Therefore, a user of CM is often confronted with the problem of incomplete information on the structure of the network or does not have the resources essential for measuring the structure of large, complex networks in total. Therefore, the information on the robustness of the used CM is highly relevant.

Besides considerations on the robustness of different CM, further research exists which identifies and analyzes the differences in results when applying different CM. Mutschke (2008), for instance, describes six anomalies (i.e., high centrality score of an actor using one CM but low centrality score when using other CM at the same time) when applying the CM DC, BC, and CC and gives a possible justification for each of these differences in the centrality of an actor. Further contributions focus on the partly significant differences in the rankings of the different actors in a social network when using different CM (e.g., Freeman 1979; Freeman et al. 1980; Kiss and Bichler 2008). Here, the ranking of the actors is defined by the descending order of the centrality score of the respective CM. In this context, when comparing

---

[1]If an approximation rather than a precise calculation of the values of BC is sufficient, the faster algorithm of Bader et al. (2007) can be used.

Table 1 Approaches to the analysis of centrality measures

Authors	Focus	Approach	Analyzed centrality measures
Bolland (1988)	Conceptional & empirical	Analysis of the robustness and sensitivity of different centrality models under conditions of random and systematic variation introduced into a network	DC, CC, BC, CM by Bonacich
Borgatti (2005)	Conceptional	Discussion of various CM regarding their matching for different types of network flow	DC, CC, BC, EC
Borgatti et al. (2006)	Empirical	Analysis of the robustness of CM under conditions of imperfect data	DC, CC, BC, EC
Costenbader and Valente (2003)	Empirical	Analysis of the stability of centrality measures when networks are sampled	DC, CC, BC, EC, CM by Bonacich, Integration, Radiality
Freeman (1979)	Conceptional & empirical	Discussion of different concepts of centrality and application of related CM to different exemplary networks	DC, CC, BC
Freeman et al. (1980)	Conceptional & empirical	Identification of an appropriate CM for the case of problem solution in groups	DC, CC, BC
Gneiser et al. (2010)	Conceptional	Development of requirements for a CM for online social networks	DC, CC, BC, PageRank-based CM
Kiss and Bichler (2008)	Empirical	Comparison of the performance of different CM regarding the identification of influential actors in a network of calls from a telecom company	DC, CC, BC, EC, PageRank-based CM, Edge-weighted DC, HITS-based CM, SenderRank CM
Mutschke (2008)	Conceptional	Discussion of various anomalies of CM	DC, CC, BC
Nieminen (1974)	Conceptional	Development of axioms for CM	DC
Sabidussi (1966)	Conceptional	Development of axioms for CM	Various indices

the CM DC, CC, and BC for all possible graphs with five actors, Freeman (1979), e.g., concludes that the order of the different actors varies enormously with the use of different CM. This observation is also confirmed by the work of Freeman et al. (1980), in which the CM DC, CC, and BC are applied to other sample networks. In addition, this article evaluates the suitability of the three CM to identify key persons in the context of "problem solving in groups". More recent contributions deal with the capability of different CM for other applications (e.g., Borgatti 2006; Hossain et al. 2007; Kiss and Bichler 2008; Lee et al. 2010; Gloor et al. 2009). For example, Kiss and Bichler (2008) investigate the quality of different CM in terms of news dissemination in a telecommunications network. Their analysis is based on a defined diffusion model. In addition to the classic CM DC, CC, BC, and EC, the authors also apply newer concepts (such as PageRank-based CM, the edge-weighted DC, a HITS-based CM and a SenderRank CM) (Kiss and Bichler 2008, pp. 236).

The main result of this investigation is that the centrality of individual actors significantly differs when using various CM, with the SenderRank CM and the relatively simple CM out-degree (a directed version of the DC) being suited best for the identification of key persons in this application case. Hossain et al. (2007) consider a similar issue by evaluating real-world data from the mobile sector as regards the four CM DC, CC, BC, and EC in order to assess the relationship between the centrality of an actor and his possibilities for disseminating information. It turns out that only by combining different CM the most important actors for the dissemination of information can be identified. Lee et al. (2010) deal with a related problem and analyze the suitability of the CM DC and BC as an indicator for the influence of individual customers on the behavior of the entire customer base. For this purpose, the authors conduct various field studies and evaluate the involved actors' self-assessment and the assessment by others in terms of their influence on other

clients. The analysis shows that BC is positively related to opinion leadership in both cases, whereas out-degree centrality is only a good indicator for the self-assessment of the surveyed actors. Moreover, Borgatti (2006) examines the quality of CM for the identification of key individuals for the purpose of optimally diffusing something through the network on the one hand and for the purpose of disrupting or fragmenting the network by removing nodes on the other hand. The author concludes that the traditional CM CC is suited best for the first case, while in the second case BC is preferable. Since these CM do not exhaustively solve the particular problems, Borgatti (2006) additionally developed new CM that are better suited for the studied issues. Comparing the results of current research which analyzes the centrality of individual actors in the application of various CM as discussed above, it remains to be noted that different CM in some cases lead to considerably different results in terms of the centrality of individual actors.

In addition to the previously discussed empirical work, SNA also offers some conceptual studies on the characteristics and underlying assumptions of CM. Bolland (1988) explains for each of the CM DC, CC, BC, and the CM by Bonacich two assumptions on the nature of network flow, one concerning the decay of resources (such as information) over distance and time and the other concerning the paths through which resources are able to flow. He comes to the conclusion that different CM are implicitly based on different assumptions on the losses that incur in transferring a resource from one actor to another. While DC assumes immediate deterioration of the transferred resource after a transfer starts, BC and CM by Bonacich assume no deterioration of the resource. In case of CC, however, a gradual loss of the resource with increasing number of transfers is assumed. Also Borgatti (2005) discusses different possibilities of network flow using some example cases for the application of CM and assigns appropriate CM to them. However, this assignment in Borgatti (2005) is merely argumentative, i.e. he does not provide quantitative criteria for intersubjective verification of the suitability of individual CM for certain applications. Other authors (e.g., Nieminen 1974; Sabidussi 1966) approach the question of the quality of a CM by formulating axiomatic requirements for the characteristics and the behavior of CM. Also for the special case of online social networks first contributions exist, aiming at a stronger focus on the characteristics (e.g., high relevance of indirect contacts of an actor) of these web-based social networks when deriving requirements for a CM in order to quantify the centrality of individual actors (see, e.g., Gneiser et al. 2010). However, this research largely lacks the motivation or justification for why and in which cases the CM should meet the requirements. Moreover, these requirements are partly of qualitative nature so that an intersubjective assessment of their validity for different CM proves difficult.

In summary, we can note that recent research on interconnectedness and CM in social networks has defined different concepts of centrality and, based on these concepts, has developed different CM. Furthermore, there are a number of both empirical and conceptual papers which compare different CM and discuss their suitability for various applications, network types, and network flows. In this context, the respective authors aim at the presentation and discussion of anomalies of different CM on the one hand or at the identification of the CM that is most appropriate for the particular application case or network flow on the other hand. In addition, the analysis of current research shows that different CM in some cases provide considerably differing results in terms of the centrality of individual actors. Therefore, the selection of a CM requires the consideration of both the specifics of different CM and the widely varying requirements of different application cases. Given the highly variable findings in view of quality of different CM, this article focuses on the illustration of the enormous importance of a reflective use of CM. Based on the findings of the SNA literature, in the following section we motivate and formulate three simple general, quantitative, and thus intersubjectively verifiable properties of CM in social networks, partly drawing on the work of Nieminen (1974) and Sabidussi (1966). The three properties are then used for the analysis of some of the most widely discussed and used CM of SNA.

3 Properties of Centrality Measures in Social Networks

Formally, a measure to quantify interconnectedness of a node x in a graph G is a mapping $\sigma^G : V_G \to IR_0^+$ which assigns a non-negative real number to each $x \in V_G$, where a higher value of σ^G indicates a better interconnectedness. In case of an identical network structure of two nodes x and y in the network the application of the CM should have the same value $\sigma^G(x) = \sigma^G(y)$ for both nodes (Nieminen 1974, p. 333; Sabidussi 1966, p. 592). Two nodes x and y are thereby considered as being identically structurally integrated into the network if a renaming of all nodes of the network is possible in such a way that all existing edges remain and x is mapped to y, i.e. if an automorphism[2] $\eta : V_G \to V_G$ with $y = \eta(x)$ exists. In **Fig. 2**, e.g., the nodes 1 and 5 as well as the nodes 2 and 4 are identically integrated into the network in terms of structure since these nodes can be mapped to each other through $2 \to 4$,

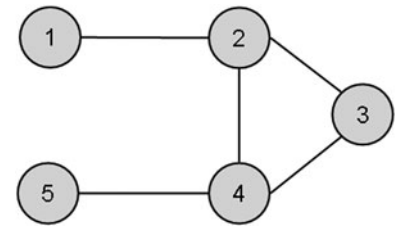**Fig. 2** Example of a network illustrating structural equivalence

$4 \to 2$, $3 \to 3$, $1 \to 5$, $5 \to 1$ and the edges (1, 2) (4, 5) (2, 3) (3, 4) and (2, 4) remain.

In the subsequent motivation of three simple general properties of CM in undirected, unweighted social networks we always assume a connected graph G. Furthermore, statements about the desired behavior of a CM when adding a new edge are made. Thus, the network is transformed from a state 1 (with associated graph G) into a state 2 (with associated graph G'). The removal of an edge corresponds exactly to the opposite operation and is associated with the reversal of the statement. For this reason, we only consider the case of adding an edge in the following.

With an additional relationship between a member x and another member y in the network, the opportunities for communication and interaction particularly increase if x gains a more direct connection (i.e. of lesser distance $d_G(x, y)$) to y. The distance $d_G(x, y)$ between the actors x and y is thereby defined as the minimum length of all paths in G that lead from x to y. In this context, Davis (1969, p. 549) assumes that the flow of information between two actors decreases in proportion to their connection length. Thus, both the extent and the quality of the transmitted information between two actors are usually higher the smaller their distance is. In addition, with a relatively low number of contacts between two actors, the contact is normally made faster and the individual actors tend to have a higher willingness to disclose relevant information (Algesheimer and von Wangenheim 2006). Moreover, the message passed through the network is usually trusted more when the actors are closer to each other. Overall, an actor x thus holds a higher potential in terms of information exchange when he is more directly connected to an actor y than without the additional connection. This should be

---

[2]An automorphism is an isomorphism of a graph to itself, with two graphs $G = (V_G, E_G)$ and $G' = (V_{G'}, E_{G'})$ being referred to as isomorphic if a bijection $\eta : V_G \to V_{G'}$ exists with $(a, b) \in E_G$ if and only if $(\eta(a), \eta(b)) \in E_{G'}$ for all $a, b \in V_G$.

Property 1 Monotonicity with respect to the distance of the actors

If the distance of the actor x to at least one other actor y is reduced through an additional relationship in the network, the centrality score of x increases.

Formally, this means:

If $V_G = V_{G'}$, $v, w, x, y \in V_G$, $v \neq w$, $x \neq y$, $(v, w) \notin E_G$, $E_{G'} = E_G \cup (v, w)$ and if for the distance between x and y due to the additional relationship between v and w $d_{G'}(x, y) < d_G(x, y)$ holds, it follows that $\sigma^{G'}(x) > \sigma^G(x)$.

Due to the symmetry of relations in this case it also follows that $\sigma^{G'}(y) > \sigma^G(y)$.

Property 2 Monotonicity with respect to the number of shortest paths

If the number of paths with shortest length from an actor x to at least one other actor y increases through an additional relationship in the network, the centrality score of x increases.

Formally, this means:

If $V_G = V_{G'}$, $v, w, x, y \in V_G$, $v \neq w$, $x \neq y$, $(v, w) \notin E_G$, $E_{G'} = E_G \cup (v, w)$ and if for the distance between x and y due to the additional relationship between v and w $d_{G'}^{(v,w)}(x, y) = d_G(x, y)$ (Here, $d_{G'}^{(v,w)}(x, y)$ refers to the length of the shortest path between x and y which contains the edge (v, w). Since the path with the length $d_{G'}^{(v,w)}(x, y)$ only exists after adding the edge (v, w), the number of shortest paths really increases in the case $d_{G'}^{(v,w)}(x, y) = d_G(x, y)$.) holds, it follows that $\sigma^{G'}(x) > \sigma^G(x)$.

Due to the symmetry of relations in this case it follows that $\sigma^{G'}(y) > \sigma^G(y)$.

Property 3 Receipt of the actors' ranking

With an additional relationship between two actors x and y, the ranking of the two members does not change with respect to the CM.

Formally, this means:

If $V_G = V_{G'}$, $x, y \in V_G$, $x \neq y$, $\sigma^G(x) > \sigma^G(y)$, $(x, y) \notin E_G$, $E_{G'} = E_G \cup (x, y)$ holds, it follows that $\sigma^{G'}(x) \geq \sigma^{G'}(y)$.

If even $\sigma^G(x) = \sigma^G(y)$ holds, it follows that $\sigma^{G'}(x) = \sigma^{G'}(y)$.

positively reflected in the value of the CM of x, as expressed in Property 1.

Furthermore, it is advantageous for the interaction in a social network if an actor can contact another member on various paths (Davis 1969, p. 549). In this way, on the one hand disruptions of the information flowing along a single path can be compensated. On the other hand, the actor usually receives more information on different paths from and about a larger number of indirect contacts. In addition, several paths to another member generally contribute to trust. This is due to the fact that in this case several, more direct contacts of an actor have a relationship to this member and thus independently indicate his trustworthiness. Due to the benefits of a smaller distance between two actors, as already described above, a path is more valuable the shorter it is. If there are multiple paths of shortest length from one network member to another, this actor also becomes more independent from the influence of individual actors in between (Freeman 1979, p. 221). Hence, an increase in the number of paths with shortest length should positively affect the centrality score of x. This is stated in Property 2.

Based on the assumption of symmetrical relations, an additional relationship

between the actors x and y is always of advantage for both parties involved as they may gain a better access to the network of each other due to the new relationship. If actor x was previously better connected than actor y, it is expected that this ranking of the actors in terms of their centrality score remains the same after adding the new contact. This results from the fact that actor y cannot benefit more from the network of actor x than x does, since x still has a more direct access to his (better evaluated) network than y and vice versa. This is expressed in Property 3.

The Properties 1 to 3 represent three simple general requirements for the behavior of CM in social networks which may be desirable in various applications. In Sect. 4 we now analyze some representatives of the CM already presented in section two in more detail.

4 Analysis of Centrality Measures

In the following we first formally define five CM one at a time and illustrate them by means of an example network. Subsequently we analyze them with respect to the previously formulated properties. The selection is limited to DC, CC, and

BC as well as the two influence measures EC and the CM by Katz as some of the most commonly used CM in SNA literature. Thus it provides a cross section of the different basic concepts of centrality as presented in Sect. 2.

4.1 Degree Centrality

DC σ_D represents the simplest CM and determines the number of direct contacts as an indicator of the quality of a network member's interconnectedness (Nieminen 1974, p. 333). Using the adjacency matrix $A = (a_{ij})$ it can be formalized as follows:

$$\sigma_D(x) = \sum_{i=1}^{n} a_{ix}. \tag{1}$$

As a consequence, the centrality score $\sigma_D(x)$ for a node x is higher, the more contacts a node x has. In the network of **Fig. 3**, e.g., it follows that $\sigma_D(1) = 1$ since actor 1 has only one direct relationship with actor 2. In contrast, actor 4 has the centrality score $\sigma_D(4) = 3$.

Table 2 shows the values of DC for all members of the example network. In addition, the actors' ranking (in short
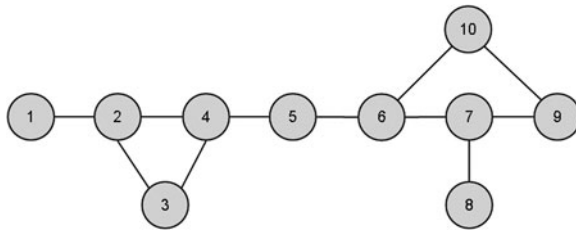
Fig. 3 Example of a network for the illustration of centrality measures

Table 2 Results for degree centrality

Degree Centrality										
Actor x	1	2	3	4	5	6	7	8	9	10
$\sigma_D(x)$	1	3	2	3	2	3	3	1	2	2
Rank	9	1	5	1	5	1	1	9	5	5

Table 3 Results for closeness centrality

Closeness Centrality										
Actor x	1	2	3	4	5	6	7	8	9	10
$\sigma_C(x)$	1/34	1/26	1/27	1/21	1/19	1/19	1/23	1/31	1/29	1/25
Rank	10	6	7	3	1	1	4	9	8	5

"rank") is stated, i.e. their order in descending value of DC. The actors 2, 4, 6, and 7 take rank 1 and thus are the best networked members when applying this CM.

With respect to Properties 1 to 3, the major disadvantage of DC is that indirect contacts are not considered at all. Therefore, a reduction of the distance from one actor x to another actor y resulting from an additional relationship in most cases does not increase the value of the CM.[3] The intensification of a connection of shortest length between x and y does also not increase the value of this CM, since DC only considers direct contacts. However, in an undirected, unweighted graph a direct connection between the actors x and y can exist only once. Overall, Properties 1 and 2 are therefore generally not met. In contrast, DC satisfies Property 3. Through a new relationship both actors involved win one additional direct contact. So the DC of both members equally increases by 1 and the ranking of the actors thus always remains the same.

4.2 Closeness Centrality

CC σ_C is based on the idea that nodes with a short distance to other nodes can spread information very productively through the network (Beauchamp 1965). In order to calculate the CC $\sigma_C(x)$ of a node x, the distances between the node x and all other nodes of the network are summed up (Sabidussi 1966, p. 583). By using the reciprocal value we achieve that the CC value increases when the distance to another node is reduced, i.e. when the integration into the network is improved. Formally, this means (e.g., Freeman 1979, p. 225)

$$\sigma_C(x) = \frac{1}{\sum_{i=1}^{n} d_G(x, i)} \quad (2)$$

For actor 4 in the network of **Fig. 3** results $\sigma_C(4) = 1/21$. This is due to the fact that for the actors $x = 2, 3, 5$ $d_G(4, x) = 1$, for the actors $x = 1, 6$ $d_G(4, x) = 2$, for the actors $x = 7, 10$ $d_G(4, x) = 3$ and for the actors $x = 8, 9$ $d_G(4, x) = 4$ holds. **Table 3** includes the centrality scores of all members in the network in **Fig. 3** and their ranking when applying CC.

For CC the reduction of the distance to at least one other actor when adding another relationship leads to a smaller value of the denominator in formula (3). Consequently, in this case the CM value of the considered actor increases and Property 1 is satisfied. However, in formula (3) only the distances between the

different actors are taken into account. Therefore, a larger number of paths with shortest length between two actors does not positively affect the value of this CM as illustrated by the network 4a of **Fig. 4**, where both before and after adding the additional connection (3, 4) $\sigma_C^G(1) = \sigma_C^{G'}(1) = 1/4$ holds, although in G' there are two paths of length 2 from actor 1 to actor 3. In the network 4b the ranking of the actors 1 and 2 also changes. While initially $\sigma_C^G(1) = 1/13 = \sigma_C^G(2)$ holds, $\sigma_C^{G'}(1) = 1/10 < \sigma_C^{G'}(2) = 1/9$ results after adding the connection (1, 2). Consequently, Property 3 is also not fulfilled.

4.3 Betweenness Centrality

In case of BC σ_B a network member is considered to be well connected if he is located on as many of the shortest paths as possible between pairs of other nodes. The underlying assumption of this CM is that the interaction between two non-directly connected nodes x and y depends on the nodes between x and y. According to Freeman (1979, p. 223) the BC $\sigma_B(x)$ for a node x is therefore calculated as

$$\sigma_B(x) = \sum_{i=1, i \neq x}^{n} \sum_{j=1, j < i, j \neq x}^{n} \frac{g_{ij}(x)}{g_{ij}} \quad (3)$$

with g_{ij} representing the number of shortest paths from node i to node j, and $g_{ij}(x)$ denoting the number of these paths which pass through the node x.

For actor 9 in the network of **Fig. 3**, e.g., $\sigma_B(9) = 1/2 + 1/2 = 1$ results since he is located on one of the two shortest paths from the actors 7 and 8 to actor 10. The values of the BC for the other actors and their ranking are listed in **Table 4**.

BC does not meet any of the required properties as is demonstrated by the networks of **Fig. 5**. In the network 5a actor 1 has a centrality score of $\sigma_B^G(1) = 3$ before adding the connection (4, 5) and a centrality score of $\sigma_B^{G'}(1) = 1$ afterwards, although the distance to actor 4 is reduced through the new relationship. Consequently, Property 1 is violated. Network 5b shows that Property 2 is also not satisfied for BC. For actor 1 holds first $\sigma_B^G(1) = 2$ and after adding the connection (3, 4) $\sigma_B^{G'}(1) = 0,5$ although there are two paths of length 2 from actor 1 to actor 3 due to the additional relationship. Moreover, the ranking of the actors 1 and 2 changes in net-

---

[3]An exception is the case that the new edge (x, y) is added and $d_{G'}(x, y) = 1$ results.
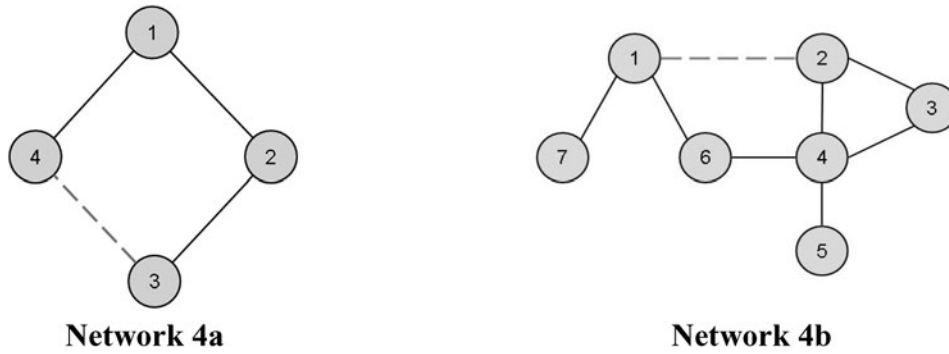
Fig. 4 Closeness centrality – Counterexamples regarding Properties 2 and 3
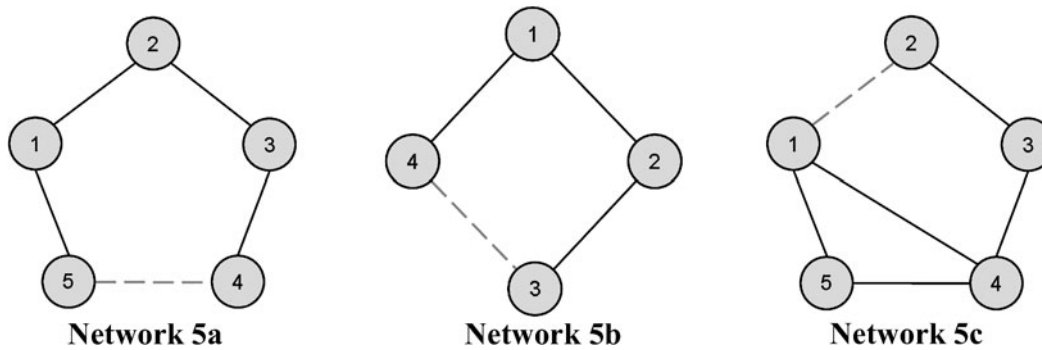

Fig. 5 Betweenness centrality – Counterexamples regarding Properties 1 to 3

work 5c. While both have the same centrality score ($\sigma_B^G(1) = 0 = \sigma_B^G(2)$) before adding the connection $(1, 2)$, afterwards $\sigma_B^{G'}(1) = 1,5 > \sigma_B^{G'}(2) = 0,5$ holds. Consequently, Property 3 is also not fulfilled for BC.

4.4 Eigenvector Centrality

EC σ_E is based on the idea that a relationship to a more interconnected node contributes to the own centrality to a greater extent than a relationship to a less well interconnected node. For a node x, the EC is therefore defined as (Bonacich and Lloyd 2001)

$$\sigma_E(x) = v_x = \frac{1}{\lambda_{\max}(A)} \cdot \sum_{j=1}^{n} a_{jx} \cdot v_j \quad (4)$$

with $v = (v_1, \ldots, v_n)^T$ referring to an eigenvector for the maximum eigenvalue[4] $\lambda_{\max}(A)$ of the adjacency matrix A.

In **Table 5** the values of the EC for the actors 1 to 10 in the network of **Fig. 3** and the resulting ranking of the actors are listed.

Table 4 Results for betweenness centrality

Betweenness Centrality

Actor x	1	2	3	4	5	6	7	8	9	10
$\sigma_B(x)$	0	8	0	18	20	21	11	0	1	6
Rank	8	5	8	3	2	1	4	8	7	6

Table 5 Results for eigenvector centrality (with $\lambda_{\max}(A) = 2,41$)

Eigenvector Centrality

Actor x	1	2	3	4	5	6	7	8	9	10
$\sigma_E(x)$	0,171	0,413	0,363	0,463	0,342	0,363	0,292	0,121	0,221	0,242
Rank	9	2	3	1	5	3	6	10	8	7

Just like BC, EC does not meet any of the required properties. This can be illustrated by means of the networks from **Fig. 6**.[5] In network 6a, actor 1 first has a centrality score of $\sigma_E^G(1) = 0,602$ and after adding the connection $(4, 6)$ of $\sigma_E^{G'}(1) = 0,417$ although the distance of actor 1 to actor 6 has been reduced. This contradicts Property 1. In network 6b the value of the CM decreases for actor 4

when the new connection $(1, 2)$ is added ($\sigma_E^G(4) = 0,604 > \sigma_E^{G'}(4) = 0,530$) although the relationship between actor 4 and actor 2 has been intensified. Therefore Property 2 is also not fulfilled. Regarding Property 3, network 6c can serve as a counterexample. Whereas before adding the connection $(4, 6)$ actor 4 has a lower centrality score than actor 6 ($\sigma_E^G(4) = 0,271 < \sigma_E^G(6) = 0,311$), the ranking of both actors changes due to

---

[4]Being a non-negative, irreducible matrix, A always has a positive eigenvalue, which is equal to the spectral radius, and an associated eigenvector with only positive entries (Graham 1987, p. 131).
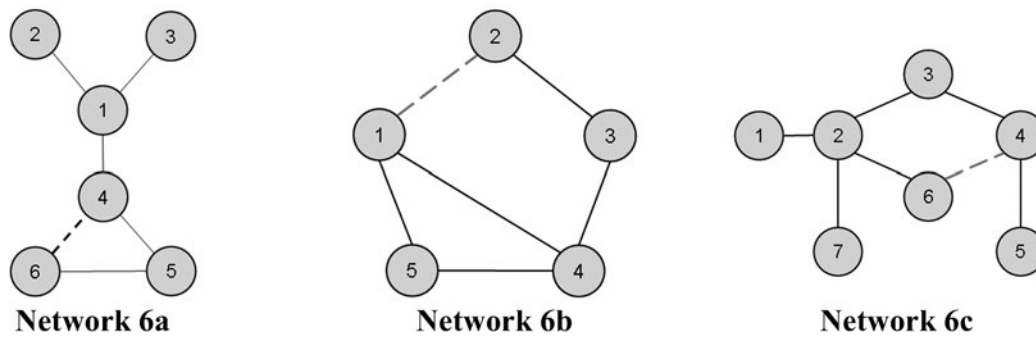[5]For detailed calculations and further details, see Appendix A.

Fig. 6 Eigenvector centrality – Counterexamples regarding Property 1 to 3

the new relationship ($\sigma_E^{G'}(4) = 0{,}435 > \sigma_E^{G'}(6) = 0{,}421$). Even these simple example networks demonstrate the additional problem that the results of EC are harder to interpret and less comprehensible than those of the previously described CM.

4.5 Katz's Centrality Measure

According to Katz not only the number of direct connections but also the further interconnectedness of actors plays an important role for the overall interconnectedness in a social network (Katz 1953). Therefore, Katz includes all paths of arbitrary length from the considered node x to the other nodes of the network in the calculation of his CM σ_K. The CM by Katz for the node x is thus defined as

$$\sigma_K(x) = 1^T \left(\sum_{i=1}^{\infty} k^i A^i \right) e_x \quad (5)$$

with $1 = (1, 1, \ldots, 1, 1)^T$ representing the $n \times 1$ vector consisting of ones only and $e_x = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ the unit vector as well as k an arbitrary (usually positive) weighting factor.[6] Since the corresponding adjacency matrix $A = (a_{ij})$ only contains the values 0 and 1, the entry $\tilde{a}_{xy}$ of the matrix $\tilde{A} = A^i$ represents the number of paths of length i from x to y (Katz 1953, p. 40). For the convergence of the series, k must be smaller than the reciprocal value of the maximum eigenvalue $\lambda_{max}(A)$ of the adjacency matrix A (Katz 1953, p. 42). This simplifies σ_K to

$$\sigma_K(x) = 1^T \left((I_n - kA)^{-1} - I_n \right) e_x \quad (5')$$

Table 6 Results for the centrality measure by Katz (with $k = 1/3$, $\lambda_{max}(A) = 2{,}41$)

Centrality measure by Katz									
Actor x	1	2	3	4	5	6	7	8	9	10
$\sigma_K(x)$	1,91	4,72	3,99	5,25	4,06	4,91	4,30	1,77	3,23	3,38
Rank	9	3	6	1	5	2	4	10	8	7

with I_n referring to the identity matrix of the dimension $n = |V_G|$. The weighting factor k can then sometimes be interpreted as the probability that a single relationship is useful for node x. This results (assuming independence of probabilities) in a probability of k^2 for a relation of second degree, and so forth (Katz 1953, p. 41). In **Table 6**, the values of the CM by Katz and the resulting rankings for the actors 1 to 10 of the example network in **Fig. 3** are listed.

In each case, the CM by Katz meets the Properties 1 and 2 since adding any new relationship in a connected graph always leads to an increase in the interconnectedness of all actors in the network. This is due to the fact that in a connected graph, a new relationship for any actor opens up additional paths to all other actors in the network. The validity of the third property for the CM by Katz can be proven formally only under certain conditions. However, extensive simulation studies show that the ranking of two actors in the CM σ_K does not change by adding an additional relationship.[7]

4.6 Summary and Comparison of Analysis Results

Table 7 summarizes the resulting rankings of the actors when applying the five considered CM to the example network in **Fig. 3**. It becomes obvious that the actors 1 and 8 have the worst centrality scores for all CM investigated. Apart from this, however, there are significant differences regarding the actors' rankings when using different CM. Actor 3, for instance, is seen as poorly interconnected when applying BC or CC, while he ranks in the midfield using DC or the CM by Katz and reaches a top position when applying EC. In addition, it is striking that DC and BC generally do not sufficiently differentiate the interconnectedness of individual members. Thus, e.g., DC provides the same value for the actors 2, 4, 6, and 7 although with all other CM the actors 4 and 6 are seen as (partly significantly) better interconnected than actor 7. This is due to the fact that DC only considers the number of direct contacts and not their further interconnectedness (i.e., their indirect contacts). In addition, BC does not distinguish between actors who have only one contact and actors whose contacts are completely interconnected. In both cases, such actors have a centrality score

---

[6]In contrast to the work of Katz we abstain from normalizing the column sum of the adjacency matrix by multiplication with $1/(n-1)$. As a consequence, the result differs by a multiplicative constant from the outcome in the original work by Katz. In addition, under the assumptions described here, the CM by Katz differs only by a constant of the alpha centrality. Further details are outlined in Appendix B.
[7]For detailed descriptions, see Appendix C.

Table 7 Ranking applying different centrality measures

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	Centrality measure by Katz
1	2, 4, 6, 7	6	5, 6	4	4
2		5		2	6
3		4	4	3, 6	2
4		7	7		7
5	3, 5, 9, 10	2	10	5	5
6		10	2	7	3
7		9	3	10	10
8		1, 3, 8	9	9	9
9	1, 8		8	1	1
10			1	8	8

Table 8 Analysis of centrality measures – Summary

Centrality Measure	Property 1	Property 2	Property 3
Degree Centrality	×	×	✓
Betweenness Centrality	×	×	×
Closeness Centrality	✓	×	×
Eigenvector Centrality	×	×	×
Centrality measure by Katz	✓	✓	✓

of 0 and thus the last rank (see, e.g., actors 1, 3, and 8). This analysis shows that an actor's centrality score can vary considerably depending on the CM.

Table 8 summarizes the results regarding the validity of the three properties for all CM presented in this paper. It shows that the majority of the frequently used CM in SNA literature does not or not fully meet the properties discussed in this article. Both BC and EC, for instance, satisfy none of the desired properties, while DC and CC only fulfill one of the three properties. According to the findings of the authors so far, among the five CM considered in this paper, only the CM by Katz meets all three properties. However, the validity of the third property could only be validated by a simulation study (see Appendix C). This result is all the more astonishing as the properties presented here constitute relatively generic, intuitively plausible requirements for the behavior of CM when adding a supplementary relationship between two actors. It makes clear that the unreflective use of existing CM may often lead to not considered and possibly undesirable side effects. Against this background, responsible decision makers should always be aware of the information a CM may provide and its limitations. A CM thus should not be selected carelessly or arbitrarily. Instead, a careful analysis regarding the requirements resulting from the particular application case is necessary.

5 Summary and Outlook

The comprehensive IT penetration of all areas of life and the enormous growth of the Internet are major causes of fundamental changes in the communicative behavior of individuals. Hence, a significant proportion of the world population is actively using web-based social networks today (just facebook.com had about 400 million members at the beginning of 2010 (Facebook 2010)). This induces an increase in the number of companies that are interested in the use of such networks for selected activities in areas such as marketing or product development. In this context, the identification of actors who are structurally well integrated into the network is of major importance. For this purpose, many CM have been developed and discussed in the SNA in recent decades. On account of the currently high importance of web-based social networks, this paper aimed at (1) presenting the current state of research regarding CM in social networks and (2) illustrating the enormous relevance of a reflected use of the existing CM in social networks given the widely varying findings on the quality of different CM.

The paper shows that in the SNA literature a variety of CM exist to quantify the interconnectedness of individual actors in social networks. Here, four basic concepts can be distinguished according to the underlying concept of centrality. In addition, numerous empirical and conceptual contributions analyze and compare the properties and robustness of different CM. These contributions from SNA literature allow the overall conclusion that different CM often lead to significantly different results for the centrality of individual actors. This is also illustrated in this paper by means of an example network. Furthermore, depending on the application case, different requirements for CM may exist. In this context, this paper exemplarily presents three simple general properties which may be desirable in various applications. On the basis of these properties we analyzed five different CM and showed that surprisingly only one of the studied CM meets all three properties, while the other four CM satisfy them only partially. This result is all the more astonishing as the properties can be considered as relatively generic, intuitively plausible requirements for the behavior of CM when adding a supplementary relationship. Consequently, decision makers should not uncritically rely on intuitively obvious statements from the application of CM. Instead, the widely varying results provided by different CM require an accurate analysis in relation to the relevant application.

The properties used to compare the CM, however, are derived based on some restrictive assumptions. First, an undirected, unweighted network is assumed. In doing so, both the existence of one-sided relations and relationships with different intensity or emotional ties between the actors are neglected. Secondly, we also do not consider the interaction frequency of each member separately in this paper. This, however, is an indicator of the actor's actual contact intensity. Due to the fact that such phenomena are difficult to observe in practice, it is often only possible to include such issues under extremely high cost. Third, the paper does not take cannibalization and saturation effects into account, which sometimes arise when an actor can devote less time to maintaining existing relationships as

a result of adding new contacts. However, since new contacts may also lead to an increase of activity in the network for many members, possible cannibalization effects are partially compensated and are therefore generally difficult to consider. Overall, the described limitations result in a variety of possible starting points for future research examining the properties and behavior of different CM in a broader context. In addition, further research is needed with regard to different concrete application scenarios of CM, the resulting requirements for the CM, and the concrete integration of the results into each application scenario. Although the results from the application of a CM – as presented in detail in this paper – thus should be considered in a differentiated manner, CM still succeed in providing an idea of the involvement of different actors in a social network and can provide valuable information for various application scenarios when used in a reflected way.

Appendix A: Counterexamples to the Properties 1 to 3 for Eigenvector Centrality

In the following we present the calculations regarding the counterexamples to the Properties 1 to 3 when applying EC in more detail. The calculations were carried out using the software Octave.

A.1 Counterexample to Property 1

The following calculations refer to the network 6a. First, we determine the adjacency matrix A of the initial network (before adding the connection $(4, 6)$). For this matrix, we calculate the maximum eigenvalue and a corresponding eigenvector.

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_{\max}(A) = 1{,}902$$

$$v(A) = \begin{pmatrix} 0{,}602 \\ 0{,}316 \\ 0{,}316 \\ 0{,}512 \\ 0{,}372 \\ 0{,}195 \end{pmatrix}$$

In the next step, we modify the network to the extent that a new relationship between the actors 4 and 6 is added. This

results in the following modified adjacency matrix A' for which the maximum eigenvalue and a corresponding eigenvector are calculated:

$$A' = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\lambda_{\max}(A') = 2{,}278$$

$$v(A') = \begin{pmatrix} 0{,}417 \\ 0{,}183 \\ 0{,}183 \\ 0{,}584 \\ 0{,}457 \\ 0{,}457 \end{pmatrix}$$

For actor 1 its centrality score first is $\sigma_E^G(1) = 0{,}602$ and after adding the connection $(4, 6)$ it is $\sigma_E^{G'}(1) = 0{,}417$ although the distance of actor 1 to actor 6 has been reduced. This violates Property 1.

A.2 Counterexample to Property 2

For network 6b we also initially determine the corresponding adjacency matrix A of the initial network (before adding the connection $(1, 2)$) and calculate the maximum eigenvalue and a corresponding eigenvector.

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_{\max}(A) = 2{,}214$$

$$v = \begin{pmatrix} 0{,}497 \\ 0{,}155 \\ 0{,}342 \\ 0{,}604 \\ 0{,}497 \end{pmatrix}$$

After adding the connection $(1, 2)$ the following modified adjacency matrix A' results, for which the maximum eigenvalue and a corresponding eigenvector are calculated:

$$A' = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_{\max}(A') = 2{,}481$$

$$v = \begin{pmatrix} 0{,}530 \\ 0{,}358 \\ 0{,}358 \\ 0{,}530 \\ 0{,}427 \end{pmatrix}$$

For actor 4 its centrality score first is $\sigma_E^G(4) = 0{,}604$ and after adding the connection $(1, 2)$ it is $\sigma_E^{G'}(4) = 0{,}530$ although the contact of actor 4 to actor 2 is intensified by the new connection $(1, 2)$. This violates Property 2.

A.3 Counterexample to Property 3

For network 6c we first also determine the associated adjacency matrix A of the initial network (before adding the connection $(4, 6)$) and calculate the maximum eigenvalue and a corresponding eigenvector.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_{\max}(A) = 2{,}101$$

$$v = \begin{pmatrix} 0{,}311 \\ 0{,}653 \\ 0{,}440 \\ 0{,}271 \\ 0{,}129 \\ 0{,}311 \\ 0{,}311 \end{pmatrix}$$

After adding the connection $(4, 6)$ the following modified adjacency matrix A' results, for which the maximum eigenvalue and a corresponding eigenvector are calculated:

$$A' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_{\max}(A') = 2{,}359$$

$$v = \begin{pmatrix} 0{,}236 \\ 0{,}557 \\ 0{,}421 \\ 0{,}435 \\ 0{,}185 \\ 0{,}421 \\ 0{,}236 \end{pmatrix}$$

Whereas before adding the connection $(4, 6)$ actor 4 has a lower centrality score than actor 6 ($\sigma_E^G(4) = 0{,}271 < \sigma_E^G(6) = 0{,}311$), the ranking of the two actors changes as a result of the additional relationship ($\sigma_E^{G'}(4) = 0{,}435 > \sigma_E^{G'}(6) = 0{,}421$), which is a contradiction to Property 3.

The calculations also show that the Properties 1 to 3 are often even not met if you multiply each entry of the eigenvector with the maximum eigenvalue $\lambda_{max}(A)$ resp. $\lambda_{max}(A')$ of the respective adjacency matrix, which may increase by adding a new relationship, but will never decrease (Bermann and Plemmons 1994).

Appendix B: Alpha Centrality

In the context of the outlined assumptions the CM by Katz differs only by a constant from the alpha centrality (AC) σ_α, which is defined as

$$\sigma_\alpha(x) = e_x^T (I_n - \alpha A)^{-1} c$$
$$= c^T (I_n - \alpha A)^{-1} e_x \qquad (6)$$

Here, I_n represents the unity matrix of dimension $n = |V_G|$ and A is the adjacency matrix of the network of relationships between the actors of the considered network. In addition, the vector c allows the consideration of the structural influences on the interconnectedness that are independent of the structure of relationships. The parameter α specifies the relative weighting of those influences that are induced by the structure of relationships and the exogenous influences (Bonacich and Lloyd 2001). If the exogenous influences do not differ, one can choose $c = 1$. In this case, for $\alpha = k$ the values of AC and the CM by Katz consequently vary only by the constant 1. The statements about the CM by Katz in Sect. 4.5 therefore also apply to AC. In order to avoid redundancy we refrain from a separate representation of AC.

Appendix C: Comments on the Validity of Property 3 for the CM by Katz

C.1 Description of the Simulation

The verification of whether the CM by Katz satisfies Property 3 was analytically not possible in the general case. Therefore, we simulated contiguous networks with 5 to 1000 nodes and examined the effect of adding an edge (x, y) on the ranking of two previously not directly connected nodes x and y. As the ranking of the newly connected nodes did not change after having carried out approximately 1 million tests, we assumed that Property 3 is met by the CM by Katz in general.

The networks and their modifications have been generated according to the following procedure: First, we created a random binary matrix B. Since the adjacency matrix of the social network must be symmetrical due to the symmetry of the relations, we mirrored the upper triangular part of the matrix B downwards

and obtained a symmetric matrix A. In a next step, all matrices were rejected which contained isolated nodes (at least one row or column sum <1) and thus apparently do not represent connected graphs. For the remaining matrices we proceeded as follows: First the adjacency matrix A' for a modified network was calculated by identifying an entry with $a_{xy} = 0$ and set $a'_{xy} = 1$ as well as $a'_{yx} = 1$, i.e., the graph received a new edge (x, y). Subsequently, we calculated the CM by Katz for both the matrix A and matrix A' for the nodes x and y and determined the ranking of the nodes before and after modification of the network. Based on the described simulation study we assume that Property 3 is satisfied when using CM by Katz.

C.2 Partially Analytical Evidence

For some special cases an analytical proof that Property 3 is met by the CM by Katz is possible. Therefore, we further present a formal proof of the validity of Property 3 in more detail for these cases, i.e. under certain restrictive assumptions, in the following. This consideration does not consider the constant -1, which would be added as a result of the subtraction of the unit matrix in formula $(5')$, because it is eliminated when calculating the difference in terms of the comparison of the ranking.

For the further deliberations the following denotations are used:

$$a = (I_n - kA) = (a_{ij})$$
$$b = (I_n - kA)^{-1} = (b_{ij})$$
$$\tilde{A} = (I_n - kA - kE^{xy}) = (\tilde{A}_{ij})$$

and $\tilde{B} = (I_n - kA - kE^{xy})^{-1} = (\tilde{B}_{ij})$, with $E^{xy} = (E_{ij}^{xy})$ describing the matrix whose entries – except of the two entries E_{xy}^{xy} and E_{yx}^{xy} – are null. The entries E_{xy}^{xy} and E_{yx}^{xy} have the value 1 and thus represent the added edge (x, y).

According to Sherman and Morrison (1950), by using the above denotations and applying the formula twice to calculate the inverse of a matrix when changing one matrix entry in each case (by adding the edge (x, y) the entries change at positions a_{xy} and a_{yx}), we obtain the following expression:

$$\tilde{B}_{ij} = \frac{(1 - kb_{xy})(b_{ij} - kb_{ij}b_{xy} + kb_{ix}b_{yj}) - (b_{ij} - kb_{ij}b_{xy} + kb_{ix}b_{yj})(kb_{xy} - k^2 b_{xy}^2 + k^2 b_{xx}b_{yy})}{(1 - kb_{xy})[(1 - kb_{xy})^2 - k^2 b_{xx}b_{yy}]}$$
$$+ \frac{(b_{iy} - kb_{iy}b_{xy} + kb_{ix}b_{yy})(kb_{xj} - k^2 b_{xj}b_{xy} + k^2 b_{yj}b_{xx})}{(1 - kb_{xy})[(1 - kb_{xy})^2 - k^2 b_{xx}b_{yy}]}$$

Hence, for $\sigma_K^{G'}(x) = \sum_{i=1}^n \tilde{B}_{ix}$ and $\sigma_K^{G'}(y) = \sum_{i=1}^n \tilde{B}_{iy}$ it results that

$$\sigma_K^{G'}(x) = \sum_{i=1}^n \tilde{B}_{ix}$$
$$= \sum_{i=1}^n \frac{[(1 - kb_{xy})b_{ix} + kb_{iy}b_{xx}]}{[(1 - kb_{xy})^2 - k^2 b_{xx}b_{yy}]}$$

and

$$\sigma_K^{G'}(y) = \sum_{i=1}^n \tilde{B}_{iy}$$
$$= \sum_{i=1}^n \frac{[(1 - kb_{xy})b_{iy} + kb_{ix}b_{yy}]}{[(1 - kb_{xy})^2 - k^2 b_{xx}b_{yy}]}$$

The difference between the centrality score of x and y in graph G', which results from adding the edge (x, y), therefore is:

$$\Delta^{G'} = \frac{1}{(1 - kb_{xy})^2 - k^2 b_{xx}b_{yy}}$$
$$\times \left[(1 - kb_{xy} - kb_{yy}) \sum_{i=1}^n b_{ix} \right.$$
$$\left. - (1 - kb_{xy} - kb_{xx}) \sum_{i=1}^n b_{iy} \right]$$

The denominator of this expression is abbreviated by $N := (1 - kb_{xy})^2 - k^2 b_{xx}b_{yy}$ in the following.

Here, $b_{xx} - 1 \geq 0$ holds. This can be justified by the fact that $b_{xx} - 1$ represents the number of paths in G that are weighted according to their length that begin and end in x. Since this value is always positive in a connected graph, $b_{xx} - 1 \geq 0$ and thus also $b_{xx} > 0$ holds. Analogously, it follows that $b_{yy} > 0$.

Property 3 comprises two statements. Sub-statement 1 refers to the ranking of two nodes that initially have a different centrality score. The second part of Property 3 makes a statement for nodes to which the CM in the original graph assigns the same centrality score. Below, the two sub-statements are proven separately from each other for some special cases.

Sub-statement 1 To be shown:

$$\sigma_K^G(x) = \sum_{i=1}^{n} b_{ix} > \sigma_K^G(y) = \sum_{i=1}^{n} b_{iy}$$

$$\Rightarrow \quad \sigma_K^{G'}(x) \geq \sigma_K^{G'}(y),$$

$$\text{i.e. } \Delta^{G'} = \sigma_K^{G'}(x) - \sigma_K^{G'}(y) \geq 0$$

The four cases in which the first sub-statement of Property 3 is satisfied in general for the CM by Katz are outlined below:

(a) *Assumptions:* $b_{xx} > b_{yy}$ and $1 - kb_{xy} - kb_{xx} > 0$

It follows from both assumptions that $1 - kb_{xy} - kb_{yy} > 0$.

$$\Rightarrow \quad N = (1 - kb_{xy})^2 - k^2 b_{xx} b_{yy}$$
$$> (1 - kb_{xy})^2 - k^2 b_{xx}^2 > 0$$

$$\Rightarrow \quad \Delta^{G'} = \frac{1}{N}\left[(1 - kb_{xy} - kb_{yy}) \sum_{i=1}^{n} b_{ix} \right.$$
$$\left. - (1 - kb_{xy} - kb_{xx}) \sum_{i=1}^{n} b_{iy} \right]$$

$$\underset{\sum_{i=1}^{N} b_{iy} < \sum_{i=1}^{N} b_{ix}}{\geq} \frac{1}{N}[1 - kb_{xy} - kb_{yy}$$

$$- 1 + kb_{xy} + kb_{xx}] \sum_{i=1}^{n} b_{iy}$$

$$= \underbrace{\frac{1}{N}}_{>0} \underbrace{[kb_{xx} - kb_{yy}]}_{>0 wg.Ann} \underbrace{\sum_{i=1}^{n} b_{iy}}_{>0} > 0$$

(b) *Assumptions:* $b_{xx} > b_{yy}$ and $1 - kb_{xy} - kb_{xx} = 0$

It follows from both assumptions that $1 - kb_{xy} = kb_{xx} > 0$ (as $k, b_{xx} > 0$)

$$\Rightarrow \quad N = (1 - kb_{xy})^2 - k^2 b_{xx} b_{yy}$$
$$> (1 - kb_{xy})^2 - k^2 b_{xx}^2 = 0$$

$$\Rightarrow \quad \Delta^{G'} = \frac{1}{N}\left[(1 - kb_{xy} - kb_{yy}) \sum_{i=1}^{n} b_{ix} \right.$$
$$\left. - (1 - kb_{xy} - kb_{xx}) \sum_{i=1}^{n} b_{iy} \right]$$

$$= \underbrace{\frac{1}{N}}_{>0} \underbrace{(1 - kb_{xy} - kb_{yy})}_{>(1-kb_{xy}-kb_{xx})=0}$$
$$\times \underbrace{\sum_{i=1}^{n} b_{ix}}_{>0} > 0$$

(c) *Assumptions:* $b_{xx} = b_{yy}$ and $1 - kb_{xy} > kb_{xx}$

$$\Rightarrow \quad N = (1 - kb_{xy})^2 - k^2 b_{xx} b_{yy}$$
$$= (1 - kb_{xy})^2 - k^2 b_{xx}^2 > 0$$

$$\Rightarrow \quad \Delta^{G'} = \frac{1}{N}\left[(1 - kb_{xy} - kb_{yy}) \sum_{i=1}^{n} b_{ix} \right.$$
$$\left. - (1 - kb_{xy} - kb_{xx}) \sum_{i=1}^{n} b_{iy} \right]$$

$$= \underbrace{\frac{1}{N}}_{>0} \underbrace{(1 - kb_{xy} - kb_{xx})}_{>0}$$

$$\times \underbrace{\left(\sum_{i=1}^{n} b_{ix} - \sum_{i=1}^{n} b_{iy} \right)}_{>0} > 0$$

(d) *Assumptions:* $b_{xx} = b_{yy}$, $1 - kb_{xy} < kb_{xx}$ and $|1 - kb_{xy}| < kb_{xx}$

$$\Rightarrow \quad N = (1 - kb_{xy})^2 - k^2 b_{xx} b_{yy}$$
$$= (1 - kb_{xy})^2 - k^2 b_{xx}^2 < 0$$

$$\Rightarrow \quad \Delta^{G'} = \frac{1}{N}\left[(1 - kb_{xy} - kb_{yy}) \sum_{i=1}^{n} b_{ix} \right.$$
$$\left. - (1 - kb_{xy} - kb_{xx}) \sum_{i=1}^{n} b_{iy} \right]$$

$$= \frac{1}{N}(1 - kb_{xy} - kb_{xx})$$

$$\times \left(\sum_{i=1}^{n} b_{ix} - \sum_{i=1}^{n} b_{iy} \right)$$

$$= \underbrace{\frac{1}{N}}_{<0} \underbrace{(1 - kb_{xy} - kb_{xx})}_{<0}$$

$$\times \underbrace{\left(\sum_{i=1}^{n} b_{ix} - \sum_{i=1}^{n} b_{iy} \right)}_{>0} > 0$$

In summary, the validity of the first sub-statement of Property 3 can be proven for the CM by Katz, if either (1) $b_{xx} > b_{yy}$ and $1 - kb_{xy} - kb_{xx} \geq 0$ or (2) $b_{xx} = b_{yy}$ and $1 - kb_{xy} > kb_{xx}$ or alternatively $1 - kb_{xy} < kb_{xx}$ and $|1 - kb_{xy}| < kb_{xx}$ holds.

Sub-statement 2 To be shown:

$$\sigma_K^G(x) = \sum_{i=1}^{n} b_{ix} = \sigma_K^G(y) = \sum_{i=1}^{n} b_{iy}$$

$$\Rightarrow \quad \sigma_K^{G'}(x) = \sigma_K^{G'}(y), \quad \text{i.e.}$$

$$\Delta^{G'} = \sigma_K^{G'}(x) - \sigma_K^{G'}(y) = 0$$

Abstract

Andrea Landherr, Bettina Friedl, Julia Heidemann

A Critical Review of Centrality Measures in Social Networks

Social networks are currently gaining increasing impact in the light of the ongoing growth of web-based services like facebook.com. One major challenge for the economically successful implementation of selected management activities such as viral marketing is the identification of key persons with an outstanding structural position within the network. For this purpose, social network analysis provides a lot of measures for quantifying a member's interconnectedness within social networks. In this context, our paper shows the state of the art with regard to centrality measures for social networks. Due to strongly differing results with respect to the quality of different centrality measures, this paper also aims at illustrating the tremendous importance of a reflected utilization of existing centrality measures. For this purpose, the paper analyzes five centrality measures commonly discussed in literature on the basis of three simple requirements for the behavior of centrality measures.

Keywords: Social network, Interconnectedness, Centrality measures, Social network analysis

Assumptions: $b_{xx} = b_{yy}$ and
$|1 - kb_{xy}| \neq kb_{xx}$

$$\Rightarrow \quad N = (1 - kb_{xy})^2 - k^2 b_{xx} b_{yy}$$
$$= (1 - kb_{xy})^2 - k^2 b_{xx}^2 \neq 0$$

$$\Rightarrow \quad \Delta^{G'} = \frac{1}{N} \left[(1 - kb_{xy} - kb_{yy}) \sum_{i=1}^{n} b_{ix} \right.$$
$$\left. - (1 - kb_{xy} - kb_{xx}) \sum_{i=1}^{n} b_{iy} \right]$$

$$= \frac{1}{N} (1 - kb_{xy} - kb_{yy} - 1$$
$$+ kb_{xy} + kb_{xx}) \sum_{i=1}^{n} b_{ix}$$

$$= \frac{1}{N} \underbrace{(kb_{xx} - kb_{yy})}_{=0} \sum_{i=1}^{n} b_{ix} = 0$$

Under the conditions $b_{xx} = b_{yy}$ and $|1 - kb_{xy}| \neq kb_{xx}$ the second substatement of Property 3 is satisfied when using the CM by Katz.

References

Algesheimer R, von Wangenheim F (2006) A network based approach to customer equity management. JRM 5(1):39–57

Bader DA, Kintali S, Madduri K, Mihail M (2007) Approximating betweenness centrality. Algorithms and models for the web-graph. In: 5th international workshop, WAW 2007, San Diego

Barabási A, Bonabeau E (2003) Scale-free networks. Sci Am 288(5):50–59

Bavelas A (1948) A mathematical model for group structures. Human Organization 7:16–30

Beauchamp MA (1965) An improved index of centrality. Behav Sci 10(2):161–163

Bolland JM (1988) Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. Social Networks 10(3):233–253

Bonacich P (1972) Factoring and weighing approaches to status scores and clique identification. J Math Sociol 2(1):113–120

Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Social Networks 23(3):191–201

Borgatti SP (2005) Centrality and network flow. Social Networks 27(1):55–71

Borgatti SP (2006) Identifying sets of key players in a social network. Computational and Mathematical Organization Theory 12(1):21–34

Borgatti SP, Everett MG (2006) A graph-theoretic perspective on centrality. Social Networks 28(4):466–484

Borgatti SP, Carley KM, Krackhardt D (2006) On the robustness of centrality measures under conditions of imperfect data. Social Networks 28(2):124–136

Brandes U (2001) A faster algorithm for betweenness centrality. J Math Sociol 25(2):163–177

Burt RS, Minor MJ (1983) Applied network analysis. Sage, Newbury Park

Coppersmith D, Winograd S (1990) Matrix multiplication via arithmetic progressions. JSC 9(3):251–280

Costenbader E, Valente TW (2003) The stability of centrality measures when networks are sampled. Social Networks 25(4):283–307

Davis JA (1969) Social structures and cognitive structures. In: Abelson et al (eds) Theories of cognitive consistency. Rand McNally, Chicago

De Valck K, van Bruggen GH, Wierenga B (2009) Virtual communities: a marketing perspective. Decis Support Syst 47(3):185–203

Dodds PS, Muhamad R, Watts DJ (2003) An experimental study of search in global social networks. Science 301(5634):827–829

Ebel H, Mielsch LI, Bornholdt S (2002) Scale-free topology of e-mail networks. Phys Rev 66(3):035103-1-035103-4

Facebook (2010) Statistiken. http://www.facebook.com/press/info.php?statistics. Accessed 2010-03-16

Frantz TL, Cataldo M, Carley KM (2009) Robustness of centrality measures under uncertainty: examining the role of network topology. Comp Math Organ Theory 15(4):303–328

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41

Freeman LC (1979) Centrality in social networks: conceptual clarification. Social Networks 1(3):215–239

Freeman LC, Roeder D, Mulholland RR (1980) Centrality in social networks. II: Experimental results. Social Networks 2(2):119–141

Freeman LC, Borgatti SP, Douglas RW (1991) Centrality in valued graphs: a measure of betweenness based on network flow. Social Networks 13(2):141–154

Gloor PA, Krauss J, Nann S, Fischbach K, Schoder D (2009) Web science 2.0: identifying trends through semantic social network analysis. In: IEEE international conference on computational science and engineering, Vancouver

Gneiser M, Heidemann J, Landherr A, Klier M, Probst F (2010) Valuation of online social networks taking into account users' interconnectedness. Appears in: ISeBM Special Issue

Graham A (1987) Nonnegative matrices and applicable topics in linear algebra, 1st edn. Ellis Horwood, Chichister

Heidemann J (2010) Online social networks – Ein sozialer und technischer Überblick. Informatikspektrum 33(3):262–271

Hitwise O (2010) Facebook reaches top ranking in US. http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html. Accessed 2010-07-02

Hossain L, Chung KSK, Murshed STH (2007) Exploring temporal communication through social networks. In: Baranauskas et al (eds) INTERACT 4662(I):19–30

Katz L (1953) A new status index derived from sociometric analysis. Psychometrika 18(1):39–43

Kiss C, Bichler M (2008) Identification of influencers – measuring influence in customer networks. Decis Support Syst 46(1):233–253

Knoke D, Kulinsik J (1982) Network analysis. Sage, Newbury Park

Koch M, Richter A, Schlosser A (2007) Produkte zum IT-gestützten Social Networking in Unternehmen. WIRTSCHAFTSINFORMATIK 49(6):448–455

Kumar R, Novak J, Tomkins A (2006) Structure and evolution of online social networks. In: Proc of the 12th ACM SIGKDD internat conf on knowledge discovery and data mining, pp 611–617

Lee S, Yook SH, Kim Y (2009) Centrality measure of complex networks using biased random walks. Eur Phys J B 68(2):277–281

Lee SHM, Cotte J, Noseworthy TJ (2010) The role of network centrality in the flow of consumer influence. Journal of Consumer Psychology 20(1):66–77

Leskovec J, Horvitz E (2008) Worldwide buzz: planetary-scale views on a large instant-messaging network. In: Proc of the 17th international world wide web conference, Beijing

Milgram S (1967) The small world problem. Psychology Today 2(1):60–67

Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proc of the 7th ACM SIGCOMM conf on internet measurement, San Diego

Mutschke P (2008) Zentralitätsanomalien und Netzwerkstruktur. Ein Plädoyer für einen "engeren" Netzwerkbegriff und ein community-orientiertes Zentralitätsmodell. In: Stegbauer C (ed) Netzwerkanalyse und Netzwerktheorie. Ein neues Paradigma in den Sozialwissenschaften. VS Verlag für Sozialwissenschaften, Wiesbaden

Newman MEJ (2005) A measure of betweenness centrality based on random walks. Social Networks 27(1):39–54

Newman MEJ, Park J (2003) Why social networks are different from other types of networks. Physical Review E68(3):36122

Nieminen J (1974) On the centrality in a graph. Scand J Psychol 15(1):332–336

Okamoto K, Chen W, Li XY (2008) Ranking of closeness centrality for large-scale social networks. In: Preparata FP, Wu X, Yin J (eds) Frontiers in algorithmics. Springer, Berlin, pp 186–195

Rousseau R, Zhang L (2008) Betweenness centrality and Q-measures in directed valued networks. Scientometrics 75(3):575–590

Sabidussi G (1966) The centrality index of a graph. Psychometrika 31(4):581–603

Scott J (1991) Network analysis: a handbook. Sage, Newbury Park

Shaw ME (1954) Group structure and the behavior of individuals in small groups. Journal of Psychology 38:139–149

The Nielsen Company (2009) Global places and Networked Places – A Nielsen report on social networking's new global footprint. http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen_globalfaces_mar09.pdf. Accessed 2009-06-29

Travers J, Milgram S (1969) An experimental study of the small world problem. Sociometry 32(4):425–443

Valente T (1996) Social network thresholds in the diffusion of innovations. Social Networks 18(1):69–89

Wassermann S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

Wellmann B (1988) Structural analysis: from method and metaphor to theory and substance. In: Wellmann B, Berkowitz SD (eds) Social structures: a network approach. Cambridge University Press, Cambridge

Zinoviev D, Duong V (2009) Toward understanding friendship in online social networks. International Journal of Technology, Knowledge and Society 5(2):1–8

to: Appendix A

Bermann A, Plemmons RJ (1994) Nonnegative matrices in the mathematical sciences, 1st edn. SIAM, Philadelphia

to: Appendix B

Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Social Networks 23(3):191–201

to: Appendix C

Sherman J, Morrison WJ (1950) Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. Ann Math Statist 21(1):124–127