Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths

Tore Opsahl[*,a], Filip Agneessens[b], John Skvoretz[c]

[a] *Imperial College Business School, Imperial College London, London SW7 2AZ, UK*
[b] *Department of Organization Sciences, VU University Amsterdam, 1081 HV Amsterdam, The Netherlands*
[c] *College of Arts & Sciences, University of South Florida, Tampa, FL 33620, USA*

Abstract

Ties often have a strength naturally associated with them that differentiate them from each other. Tie strength has been operationalized as weights. A few network measures have been proposed for weighted networks, including three common measures of node centrality: degree, closeness, and betweenness. However, these generalizations have solely focused on tie weights, and not on the number of ties, which was the central component of the original measures. This paper proposes generalizations that combine both these aspects. We illustrate the benefits of this approach by applying one of them to Freeman's EIES dataset.

Key words: degree, closeness, betweenness, weighted networks

[*] Corresponding author. Tel: +44 20 7594 3035.
Email addresses: t.opsahl@imperial.ac.uk (Tore Opsahl), f.agneessens@fsw.vu.nl (Filip Agneessens), skvoretz@cas.usf.edu (John Skvoretz)

1. Introduction

Social network scholars are increasingly interested in trying to capture more complex relational states between nodes. One of these avenues of research has focused on the issue of tie strength, and a number of studies from a wide range of fields have begun to explore this issue (Barrat et al., 2004; Brandes, 2001; Doreian et al., 2005; Freeman et al., 1991; Granovetter, 1973; Newman, 2001; Opsahl and Panzarasa, 2009; Yang and Knoke, 2001). Whether the nodes represent individuals, organizations, or even countries, and the ties refer to communication, cooperation, friendship, or trade, ties can be differentiated in most settings. These differences can be analyzed by defining a weighted network, in which ties are not just either present or absent, but have some form of weight attached to them. In a social network, the weight of a tie is generally a function of duration, emotional intensity, intimacy, and exchange of services (Granovetter, 1973). For non-social networks, the weight often quantifies the capacity or capability of the tie (e.g., the number of seats among airports; Colizza et al., 2007; Opsahl et al., 2008) or the number of synapses and gap junctions in a neural network (Watts and Strogatz, 1998). Nevertheless, most social network measures are solely defined for binary situations and, thus, unable to deal with weighted networks directly (Freeman, 2004; Wasserman and Faust, 1994). By dichotomizing the network, much of the information contained in a weighted network datasets is lost, and consequently, the complexity of the network topology cannot be described to the same extent or as richly. As a result, there has been a growing need for network measures that directly account for tie weights.

The centrality of nodes, or the identification of which nodes are more "central" than others, has been a key issue in network analysis (Freeman, 1978; Bonacich, 1987; Borgatti, 2005; Borgatti et al., 2006). Freeman (1978) argued that central nodes were those "in the thick of things" or focal points. To exemplify his idea, he used a network consisting of 5 nodes (see Figure 1). The middle node has three advantages over the other nodes: it has more ties, it can reach all the others more quickly, and it controls the flow between the others. Based on these three features, Freeman (1978) formalized three different measures of node centrality: degree, closeness, and betweenness. Degree is the number of nodes that a focal node is connected to, and measures the involvement of the node in the network. Its simplicity is an advantage: only the local structure around a node must be known for it to be calculated (e.g., when using data from the General Social Survey; McPherson et al., 2001). However, there are limitations: the measure does not take into consideration the global structure of the network. For example, although a node might be connected to many others, it might not be in a position to reach others quickly to access resources, such as information or knowledge (Borgatti, 2005; Brass, 1984). To capture this feature, closeness centrality was defined as the inverse sum of shortest distances to all other nodes from a focal node. A main limitation of closeness is the lack of applicability to networks with disconnected components: two nodes that belong to different components do not have a finite distance between them. Thus, closeness is generally restricted to nodes within the largest component of a network[1]. The last of the three measures, betweenness, assess the degree to which a

[1] A possible method for overcoming this limitation is to sum the inversed distances instead of the inverse sum of distances as the limit of 1 over infinity is 0.

node lies on the shortest path between two other nodes, and are able to funnel the flow in the network. In so doing, a node can assert control over the flow. Although this measure takes the global network structure into consideration and can be applied to networks with disconnected components, it is not without limitations. For example, a great proportion of nodes in a network generally does not lie on a shortest path between any two other nodes, and therefore receives the same score of 0.

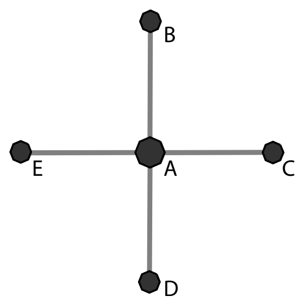

Figure 1: A star network with 5 nodes and 4 edges. The size of the nodes correspond to the nodes' degree. Adapted from Freeman (1978).

Freeman's (1978) measures are only designed for binary networks. There have been a number of attempts to generalize Freeman's (1978) three node centrality measures to weighted networks (Barrat et al., 2004; Brandes, 2001; Newman, 2001). However, all these attempts have solely focused on tie weights, and not on the number of ties, which formed the basis of the original measures. First, degree was extended to weighted networks by Barrat et al. (2004) and defined as the sum of the weights attached to the ties connected to a node. An outcome of 10 could either be a result of 10 ties with a weight of 1, 1 tie with a weight of 10, or a combination between those two extremes.

Second, the extensions of the closeness and betweenness centrality measures by Newman (2001) and Brandes (2001), respectively, rely on Dijkstra's (1959) shortest path algorithm. This algorithm defines the shortest path between two nodes as the least costly path. Brandes' (2001) and Newman's (2001) implementations suggest costs are only based on tie weights. In so doing, these three generalizations do not take into account a key feature, which the original measures were defined around, the number of ties (Freeman, 1978).

This raises a crucial question about the relative importance of tie weights to the number of ties in weighted networks. One can view the number of ties as more important than the weights, so that the presence of many ties with any weight might be considered more important than the total sum of tie weights. However, ties with large weights might be considered to have a much greater impact than ties with only small weights. This trade-off is the main motivation for this paper and drives the need for defining novel measures that enable researchers to set the relative importance between the number of ties and tie weights.

The rest of the paper is organized as follows. We start by proposing a generalization of degree centrality for weighted networks where the outcome is a combination of the number of ties and the tie weights. Then, in order to extend the closeness and betweenness centrality measures, we propose a generalization of shortest distances for weighted network that takes into account both the number of intermediary nodes and the tie weights. Subsequently, we suggest how the closeness and betweenness measures can take advantage of this generalized shortest distance algorithm. In Section 4, we evaluate the benefits of the proposed measures and explore the trade-off further by applying the degree measure to the well-known EIES dataset (Freeman and Freeman, 1979). In particular, we conduct a sensitivity analysis of the relative importance between the number of ties and the tie weights. Finally, we conclude with a discussion on the measures and various levels of the tuning parameter.

2. Degree

Freeman (1978) asserted that the degree of a focal node is the number of adjacencies in a network, i.e. the number of nodes that the focal node is connected to. Degree is a basic indicator and often used as a first step when studying networks (Freeman, 2004; McPherson et al., 2001; Wasserman and Faust, 1994). To formally describe this measure and ease the comparison among the different measures introduced in this paper, this measure can be formalized as:

$$k_i = C_{\mathrm{D}}(i) = \sum_j^N x_{ij} \tag{1}$$

where i is the focal node, j represents all other nodes, N is the total number of nodes, and x is the adjacency matrix, in which the cell x_{ij} is defined as 1 if node i is connected to node j, and 0 otherwise.

Degree has generally been extended to the sum of weights when analyzing weighted networks (Barrat et al., 2004; Newman, 2004; Opsahl et al., 2008), and labeled *node* strength. This measure has been formalized as follows:

$$s_i = C_{\mathrm{D}}^w(i) = \sum_j^N w_{ij} \tag{2}$$

where w is the weighted adjacency matrix, in which w_{ij} is greater than 0 if the node i is connected to node j, and the value represents the weight of the tie. This is equal to the definition of degree if the network is binary, i.e. each tie has a weight of 1. Conversely, in weighted networks, the outcomes of these two measures are different. Since node strength takes into consideration the weights of ties, this has been the preferred measure for analyzing weighted networks (e.g., Barrat et al., 2004; Opsahl et al., 2008). However, node strength is a blunt measure as it only takes into consideration a node's total level of involvement in the network, and not the number of other nodes to which it connected. To exemplify this, node A and node B have the same strength in Figure 2, but node B is connected to twice as many nodes as node A, and is therefore, involved in more parts of the network. Since degree and strength can be both indicators of the level of involvement of a node in the surrounding network, it is important to incorporate both these measures when studying the centrality of a node.

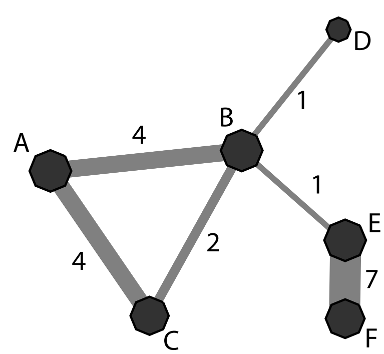

Figure 2: A network with 6 nodes and 6 weighted edges. The size of the nodes correspond to the nodes' strength.

In an attempt to combine both degree and strength, we use a tuning parameter, α, which determines the relative importance of the number of ties compared to tie weights. More specifically, we propose a degree centrality measure, which is the product of the number of nodes that a focal node is connected to, and the average weight to these nodes adjusted by the tuning parameter. We formally propose the following measure:

$$C_{\mathrm{D}}^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i}\right)^{\alpha} = k_i^{(1-\alpha)} \times s_i^{\alpha} \tag{3}$$

where α is a positive tuning parameter that can set according to the research setting and data. If this parameter is between 0 and 1, then having a high degree is taken as favorable, whereas if it is set above 1, a low degree is favorable. In Sections 4 and 5, we elaborate on the different levels of α.

Table 1 illustrates the effect of the α on the value of this measure for the nodes of the network in Figure 2. As shown by this table, when $\alpha = 1$ the measure's value equal the node's strength (eq. 2). When $\alpha < 1$ and the total node strength is fixed, the number of contacts over which the strength is distributed increases the value of the measure. For example, when $\alpha = 0.5$, node B attains a higher score than node A, despite having the same node strength. Conversely, when $\alpha > 1$ and the total node strength is fixed, the number of contacts of which the strength is distributed decrease the value of the measure in favor of a greater concentration of node strength on only a few nodes. Hence, node A attains a higher value of the measure than node B. Moreover, with an $\alpha = 1.5$, node F attains a higher value than node A and node B, even though it has a lower node strength.

Directed networks add complexity to degree as two additional aspects of a node's involvement are possible to identify. The activity of a node, or its gregariousness, can be quantified by the number of ties that originate from a node, k^{out}. While the number of ties that are directed towards a node, k^{in},

			$C^{w\alpha}_{\text{D}}$ when α=			
Node	C_{D}	C^{w}_{D}	0	0.5	1	1.5
A	2	8	2	4	8	16
B	4	8	4	5.7	8	11.3
C	2	6	2	3.5	6	10.4
D	1	1	1	1	1	1
E	2	8	2	4	8	16
F	1	7	1	2.6	7	18.5

Table 1: Degree centrality scores when different values of α are used.

is a proxy of its popularity. Moreover, since not all ties are not necessarily reciprocated, k^{out} is not always equal to k^{in}. For a weighted network, s^{out} and s^{in} can be defined as the total weight attached to the outgoing and incoming ties, respectively. However, these two measures have the same limitation as s in that they do not take into account the number of ties. In a similar spirit as $C^{w\alpha}_{\text{D}}$, we propose the following two measures to assess a node's activity and popularity, respectively:

$$C^{w\alpha}_{\text{D-out}}(i) = k^{out}_i \times \left(\frac{s^{out}_i}{k^{out}_i}\right)^{\alpha} \tag{4a}$$

$$C^{w\alpha}_{\text{D-in}}(i) = k^{in}_i \times \left(\frac{s^{in}_i}{k^{in}_i}\right)^{\alpha} \tag{4b}$$

The value of α in these equations is similar to the one in Equation 3. If two nodes have the same s^{out} and different k^{out}, the measure would assign a higher score to the node with the highest k^{out} if α is below 1, whereas the node with the lowest k^{out} would get the highest score if α is greater than 1.

3. Closeness and Betweenness

The closeness and betweenness centrality measures rely on the identification and length of the shortest paths among nodes in the network. Therefore, in an effort to generalize these measures for weighted networks, a first step is to generalize how shortest distances are identified and their length defined. There has been great interest in the shortest distances among nodes in networks (Katz, 1953; Newman, 2001; Peay, 1980; Wasserman and Faust, 1994; Yang and Knoke, 2001). In a binary network, the shortest path is found by minimizing the number of intermediary nodes, and its length is defined as the minimum number of ties linking the two nodes, either directly or indirectly. We define it here as the *binary* shortest distance to add clarity to our argument:

$$d(i,j) = \min\left(x_{ih} + ... + x_{hj}\right) \tag{5}$$

where h are intermediary nodes on paths between node i and j. For instance, if the two nodes are not connected, but are connected to the same other node, the shortest distance between them would be 2.

An important assumption implied when analyzing the shortest distances is that the intermediary nodes increases the cost of the interaction. First, a higher number of intermediary nodes, increases the time taken for the interaction between the two nodes. Second, the intermediary nodes are in a position of tertius gaudens or powerful third-party, and can distort information or delay interaction between the nodes (Simmel, 1950; Burt, 1992). Since all ties have the same weight in binary networks, the shortest path for interaction between two nodes is through the smallest number of intermediary nodes.

Different aspects of the shortest distances among nodes in a network are used in the closeness and betweenness measures. Closeness centrality relies on the length of the paths from a node to all other nodes in the network, and is defined as the inverse total length. Betweenness relies on the identification of the shortest paths, and measures the number of them that passes through a node. Freeman (1978) asserted that closeness and betweenness were, respectively:

$$C_{\mathrm{C}}(i) = \left[\sum_{j=1}^{N} d(i,j)\right]^{-1} \tag{6a}$$

$$C_{\mathrm{B}}(i) = \frac{g_{jk}(i)}{g_{jk}} \tag{6b}$$

where g_{jk} is the number of binary shortest paths between two nodes, and $g_{jk}(i)$ is the number of those paths that go through node i.

A complication arises when the ties in a network are differentiated (i.e., have a weight attached to them). For example, diseases are more likely to be transferred from one person to another if they have frequent interaction (Valente, 1995). This complication has implications for diffusion in networks, especially if a backbone of strong ties exists. In fact, it has been shown that the nodes with the highest node strength are likely to be strongly connected in networks from a range of different domains (Opsahl et al., 2008).

The network in Figure 3 illustrates three paths between two nodes, node A and B, which are composed of different number of intermediary nodes and ties with different weights. The binary shortest path would be the direct connection ($\{A, B\}$). However, in a weighted network, one could wonder whether this is the quickest path for flow. Although the path through node D and node E contains two intermediary nodes ($\{A, D, E, B\}$), it could be quicker or more likely since it is composed of stronger ties. For example, information could be transmitted through a longer chain of strong ties more quickly, and diseases might have higher probability of being transmitted through a chain containing more individuals connected through more frequent ties than through a weak direct connection.

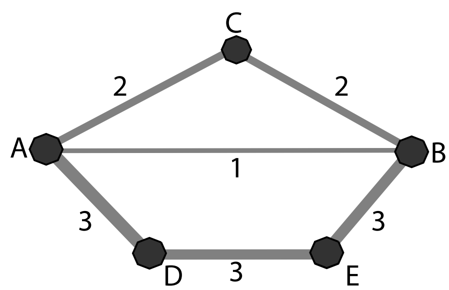

Figure 3: A network with three paths between two nodes (node A and node B): directly, $\{A, B\}$; through one intermediary node, $\{A, C, B\}$; or through two intermediary nodes, $\{A, D, E, B\}$.

There have been several attempts to identify shortest paths in weighted networks (Dijkstra, 1959; Katz, 1953; Peay, 1980; Yang and Knoke, 2001). Dijkstra (1959) proposed an algorithm that finds the path of least resistance, and was defined for networks where the weights represented costs of transmitting (e.g., distance in GPS devices or time to route Internet traffic). Since weights in most weighted networks are operationalizations of tie strength and not the cost of them, the tie weights need to be reversed before directly applying Dijkstra's algorithm to identify the shortest paths in these networks. Both Newman (2001) and Brandes (2001) separately proposed to invert the tie weights while extending closeness and betweenness centrality, respectively.[2] In so doing, the weights can be considered as costs since high values represented weak or costly ties, whereas low values represented strong or cheap ties. For example, if the tie between two nodes has a weight that is twice as large as the tie between another pair of nodes, the distance between the former pair is half of the distance between the latter pair. Moreover, absent ties (weight of 0) would be assigned an infinite large distance with this method. Newman's (2001) and Brandes' (2001) implementation of Dijkstra's algorithm can be formally defined as:

$$d^w(i,j) = \min\left(\frac{1}{w_{ih}} + ... + \frac{1}{w_{hj}}\right) \tag{7}$$

To illustrate the effect of taking tie weights into account when calculating distance, Table 2 shows the distance calculated by this algorithm for the three distinct paths in Figure 3. As can be seen from the table, the distance between node A and node B is not affected by the number of nodes on the shortest path between two nodes. In fact, Dijkstra's algorithm implicitly assumes that the number of intermediary nodes only represent a negligible cost.

Following Dijkstra (1959), Brandes (2001), and Newman (2001), we extend the shortest path algorithm by taking into consideration the number of intermediary nodes. We transform the inverted weights by a similar tuning parameter used in the proposed degree measure, Eq. 3, before using Dijkstra's algorithm to find the least costly path. This ensures that both the tie

[2]Brandes (2008) suggested alternative methods for transforming positive tie weights into costs by subtracting the true tie weight from an upper bound like the maximum plus one, or using a negative exponent of the true tie weight.

			$d^{w\alpha}(A,B)$ when α=			
path	$d(A,B)$	$d^{w}(A,B)$	0	0.5	1	1.5
{A,B}	1	1	1	1	1	1
{A,C,B}	2	1	2	1.4	1	0.7
{A,D,E,B}	3	1	3	1.8	1	0.5

Table 2: Lengths of the paths in Figure 3 when defined by the binary distance and Dijkstra's distance as well as when different values of α are used.

weights and the number of intermediary nodes affects the identification and length of paths. Formally, we define the length of the shortest path between two nodes, which incorporates the method for identifying it, as:

$$d^{w\alpha}(i,j) = \min\left(\frac{1}{(w_{ih})^{\alpha}} + ... + \frac{1}{(w_{hj})^{\alpha}}\right) \tag{8}$$

where α is a positive tuning parameter.

To illustrate the effect of various tuning parameters, the last four columns of Table 2 show the length of the three paths between node A and node B in Figure 3 when different values of α are used. When $\alpha = 0$, the proposed measure produces the same outcome as the binary distance measure, whereas when $\alpha = 1$, the outcome is the same as the one obtained with Dijkstra's algorithm. When Dijkstra's algorithm produces the same distance score for paths with different number of intermediary nodes (as it does for the three paths in Figure 3), a value for $\alpha < 1$ assigns the path with the greatest number of intermediary nodes the longest distance. Hence, for $\alpha < 1$, a shorter path composed of weak ties (e.g., {A,B}) is favored over a longer path with strong ties (e.g., {A,D,E,B}). Conversely, when $\alpha > 1$, the impact of additional intermediary nodes is relatively unimportant compared to the strength of the ties and paths with more intermediaries are favored.

The proposed shortest path algorithm can be used to allow the closeness centrality measure to take into account both the number of intermediary nodes and the tie weights. By combining Equations 6a and 8, we get the following measure:

$$C_{\mathrm{C}}^{w\alpha}(i) = \left[\sum_{j=1}^{N} d^{w\alpha}(i,j)\right]^{-1} \tag{9}$$

Moreover, we can also take advantage of the proposed shortest path algorithm to extend betweenness centrality. In so doing, betweenness will be based on a combination of the number of intermediary nodes and tie weights. Formally, we propose the following measure for betweenness centrality:

$$C_{\mathrm{B}}^{w\alpha}(i) = \frac{g_{jk}^{w\alpha}(i)}{g_{jk}^{w\alpha}} \tag{10}$$

The proposed generalizations can also be applied to directed networks. The identification of the shortest paths, and their length, in directed networks is similar to the process in undirected networks with a constraint. A path from one node to another can only follow the direction of present ties. For example, information cannot be passed from a node to another if the first node is not connected to the second node, irrespective of whether the second node is connected to the first node. This implies that the distance from a node i to another node j is not necessarily equal to the distance from node j to node i. This constraint can also easily be applied to our proposed measure.

4. The generalized degree centrality measure applied to Freeman's EIES network

To illustrate the effect of various levels of α, we apply our measures to a commonly used network datasets, the Freeman's EIES dataset (Freeman and Freeman, 1979; Opsahl and Panzarasa, 2009; Wasserman and Faust, 1994). This dataset was collected in 1978 and contains three different network relations among researchers working on social network analysis. While the first two networks are the inter-personal relationships among the researchers at the beginning and at the end of the study, the ties in the third network are defined as the number of messages sent among 32 of the researchers on an electronic communication tool. We focus on the third network as the tie weights in this network is based on a ratio scale (i.e., 0 implies the absent of a tie and a weight of 10 is twice a weight of 5).

Table 3 ranks the 32 researchers according to the degree centrality score for different values of α. Looking at the most central individuals, we see that Lin Freeman is the most central researchers, irrespective of α, as he is connected to the most people and has sent the highest number of messages. Sue Freeman and Nick Mullins are connected to the same number of other researchers; however, when we increase the α from 0 to 0.5, Barry Wellman and Russ Bernard replace them in the top three as they sent considerably more messages to other researchers, albeit to a smaller number of them.

In addition, most researchers maintain a relatively stable ranking across the diverse α. Nonetheless, a number of individuals provide exemplary results. In particular, the ranks of Phipps Arabie and Maureen Hallinan change considerably when using different α's. On the one hand, Phipps Arabie ranks

rank	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$
1	Lin Freeman (31)	Lin Freeman (314)	Lin Freeman (3171)	Lin Freeman (32071)
2	Sue Freeman (31)	Barry Wellman (249)	Barry Wellman (2208)	Barry Wellman (19607)
3	Nick Mullins (31)	Russ Bernard (200)	Russ Bernard (1596)	Russ Bernard (12752)
4	Phipps Arabie (28)	Sue Freeman (180)	Doug White (1121)	Lee Sailer (8382)
5	Barry Wellman (28)	Doug White (177)	Lee Sailer (1061)	Doug White (7093)
6	Doug White (28)	Nick Mullins (142)	Sue Freeman (1044)	Sue Freeman (6059)
7	Russ Bernard (25)	Lee Sailer (134)	Pat Doreian (838)	Pat Doreian (5424)
8	Ron Burt (20)	Pat Doreian (129)	Nick Mullins (652)	Nick Mullins (2990)
9	Pat Doreian (20)	Ron Burt (85)	Ron Burt (360)	Ron Burt (1527)
10	Richard Alba (18)	Richard Alba (77)	Richard Alba (331)	Richard Alba (1419)
11	Jack Hunter (18)	Steve Seidman (70)	Steve Seidman (305)	Maureen Hallinan (1396)
12	Lee Sailer (17)	Phipps Arabie (66)	Al Wolfe (283)	Steve Seidman (1332)
13	Steve Seidman (16)	Jack Hunter (65)	Carol Barner-Barry (239)	Al Wolfe (1272)
14	Carol Barner-Barry (15)	Al Wolfe (63)	Jack Hunter (233)	Carol Barner-Barry (954)
15	Al Wolfe (14)	Carol Barner-Barry (60)	Maureen Hallinan (227)	Jack Hunter (838)
16	Paul Holland (12)	Paul Holland (49)	Paul Holland (198)	Paul Holland (804)
17	John Boyd (11)	John Boyd (45)	John Boyd (188)	John Boyd (777)
18	Davor Jedlicka (11)	Davor Jedlicka (45)	Davor Jedlicka (188)	Davor Jedlicka (777)
19	Charles Kadushin (7)	Maureen Hallinan (37)	Phipps Arabie (155)	Don Ploch (430)
20	Nan Lin (7)	Don Ploch (28)	Don Ploch (109)	Phipps Arabie (365)
21	Don Ploch (7)	Claude Fischer (22)	Claude Fischer (82)	Claude Fischer (303)
22	Claude Fischer (6)	Mark Granovetter (22)	Mark Granovetter (81)	Mark Granovetter (298)
23	Mark Granovetter (6)	Charles Kadushin (21)	Joel Levine (70)	Joel Levine (293)
24	Maureen Hallinan (6)	Nan Lin (20)	Nick Poushinsky (68)	Nick Poushinsky (251)
25	Nick Poushinsky (5)	Nick Poushinsky (18)	Charles Kadushin (65)	Charles Kadushin (198)
26	Sam Leinhardt (4)	Joel Levine (17)	Nan Lin (58)	Nan Lin (167)
27	Joel Levine (4)	John Sonquist (10)	John Sonquist (24)	Gary Coombs (63)
28	John Sonquist (4)	Sam Leinhardt (9)	Sam Leinhardt (22)	John Sonquist (59)
29	Brian Foster (3)	Brian Foster (7)	Gary Coombs (20)	Sam Leinhardt (52)
30	Ev Rogers (3)	Gary Coombs (6)	Brian Foster (15)	Brian Foster (34)
31	Gary Coombs (2)	Ev Rogers (6)	Ev Rogers (14)	Ev Rogers (30)
32	Ed Laumann (2)	Ed Laumann (4)	Ed Laumann (8)	Ed Laumann (16)

Table 3: Ranking of different scientists according to their degree centrality scores (in parenthesis) when different values of α are used.

fourth when α is set to 0 as he sent messages to all, but three, in the network. However, as illustrated in the left panel (A) of Figure 4, the mean number of messages he sent to others is relatively low as compared to individuals who sent roughly the same total amount of messages, such as John Boyd (middle panel, B). As a result, Phipps Arabie's ranking drops considerably when α increases (to 12^{th} when $\alpha = 0.5$, 19^{th} when $\alpha = 1$, and 20^{th} when $\alpha = 1.5$), and he is, in fact, becoming less central than John Boyd. On the other hand, Maureen Hallinan had a strikingly different communication pattern (right panel, C). While she sent approximately the same total number of messages as Phipps Arabie and John Boyd, she sent messages to only six people. As the number of contacts is relatively low, she is only ranked 24^{th} when an α of 0 is used. Since the mean number of messages to her contacts is relatively high, she becomes the eleventh most central person when α is set to 1.5. This illustrates that the measure considers both the number of ties and tie strength as well as being sensitive to the average tie weight of a node.

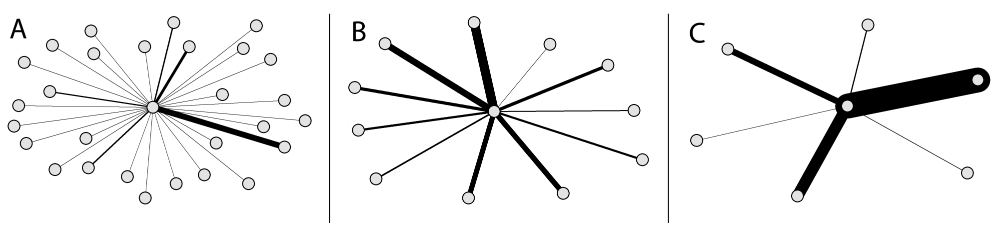

Figure 4: Ego networks of Phipps Arabie (A), John Boyd (B), and Maureen Hallinan (C) from Freeman's third EIES network. The width of a tie corresponds to the number of messages sent from the focal node to their contacts.

5. Discussion and Conclusion

This paper was motivated by the need for centrality measures to incorporate both the number of ties and their tie weights when applied to weighted networks, and to allow researchers to define the relative importance they want to give to each of these two aspects. The original measures proposed by Freeman (1978) solely consider the number of ties and disregard tie weights. Conversely, the existing generalizations of Freeman's (1978) node centrality measures exclusively focus on tie weights. In particular, Barrat et al.'s (2004) generalization equates two ties with a weight of 1 and a single tie with a weight of 2, and the implementation of Dijkstra's (1959) shortest path algorithm by Newman (2001) and Brandes (2001) implicitly assumes that the number of intermediary nodes only represent a negligent cost. This might be a valid assumption for servers routing Internet traffic as information is transferred without alteration or delay. However, in a social network, this assumption is likely to be invalid. In fact, the quality of the resources flowing through paths with more intermediary nodes is likely to be lower than for paths with fewer intermediary nodes, even if both paths have the same distance according to Dijkstra's algorithm.

To take both the number of ties and tie weights into consideration, all the proposed measures included a tuning parameter, α. This parameter controls for the relative importance of these two aspects. More specifically, there are two benchmark values (0 and 1), and if the parameter is set to either of these values, the existing measures are reproduced. If the parameter is set to the benchmark value of 0, the outcomes of the measures are solely based on the number of ties, and are equal to the ones found when applying

Freeman's (1978) measures to a binary version of a network where all the ties with a weight greater than 0 are set to present. In so doing, the tie weights are completely ignored. Conversely, if the value of the parameter is 1, the outcomes of the measures are based on tie weights only, and are identical to the already proposed generalizations of degree (Barrat et al., 2004), closeness (Newman, 2001), and betweenness (Brandes, 2001). This implies that the number of ties is disregarded. For example, for degree, the outcome is equal to the sum of weights attached to all the ties of a node, irrespectively of whether the node is involved with many or few nodes in the network. Similarly, for closeness and betweenness, the identification and length of the shortest paths is based on the sum of the inverted tie weights. Thus, the number of intermediary nodes is ignored.

For other values of α, alternative outcomes are attained, which are based on both the number of ties and tie weights. In particular, two ranges of values can be distinguished. First, a parameter set between 0 and 1 would positively value both the number of ties and tie weights. This implies that, for the degree centrality measure, both increments in node degree and node strength will then increase the outcome. While for closeness and betweenness centrality, paths with a lower number of intermediary nodes will be considered to be shorter. Second, if the value of the parameter is above 1, the measures would positively value tie strength and negatively value the number of ties. More specifically, for degree, nodes with on average stronger ties will get a higher score, and the shortest paths will be composed of stronger ties than weaker ones, even though they might have a higher cost accordingly to Dijkstra's algorithm.

Our measures have direct applicability to knowledge networks, such as information and advice networks. A number of researchers have argued that the transfer and sharing of tacit knowledge requires strong ties (Hansen, 1999). Therefore, when focusing on the effects of tacit knowledge, an α greater than 1 might be more appropriate than an α lower than 1. In this case, fewer strong ties would increase the degree centrality as compared to more weak ties, and for closeness and betweenness centrality, increase the importance of longer paths composed of stronger ties over shorter and weaker paths.

On the contrary, if the focus is on explicit or easily codified knowledge, where weak ties are important (Granovetter, 1973), an α lower than 1 might be more suitable. For degree, such an α will increase the importance of the number of contacts. In so doing, the measure favors having many weak ties over having a few strong ones. In a similar spirit, when focusing on the fact that intermediary nodes on a path between two nodes can be in a position of control over the interaction, the number of them might be more important for calculating the distance than tie weights. Therefore, when strong ties are not a requirement for transfer of knowledge, closeness and betweenness centrality measures should mainly take into account the number of intermediary nodes.

A main limitation of the proposed generalizations in this paper, as with other measures for weighted networks, is that they assume that the tie weights are based on a ratio scale (Opsahl and Panzarasa, 2009). If this is not the case, the mean tie weight has no real meaning, and therefore, the proposed centrality measures can in principle not be used. Moreover, although certain features have been associated with specific ranges of α, it is difficult to determine the exact value of α to use. This leads to another area of potential research, which involves identifying the optimal α for various outcome variables, such as intra- and inter-organizational performance, using a regression framework. For example, we could ask the question whether it is better to have many weak ties ($\alpha \in [0, 1]$) or few strong ties ($\alpha > 1$). Such studies would allow for a better understanding of the appropriate α to use in certain settings.

Note

The proposed measures are implemented in the *R*-package *tnet*.[3] This package is available through the Comprehensive R Archive Network (CRAN).

References

Barrat, A., Barthélémy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences 101 (11), 3747–3752.

Bonacich, P., 1987. Power and centrality: A family of measures. American Journal of Sociology 92, 1170–1182.

Borgatti, S. P., 2005. Centrality and network flow. Social Networks 27 (1), 55–71.

Borgatti, S. P., Carley, K., Krackhardt, D., 2006. Robustness of centrality

[3] `http://opsahl.co.uk/tnet/`

measures under conditions of imperfect data. Social Networks 28 (2), 124–136.

Brandes, U., 2001. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25, 163–177.

Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. Social Networks 30, 136–145.

Brass, D. J., 1984. Being in the right place: A structural analysis of individual influence in an organization. Administrative Science Quarterly 29 (4), 518–539.

Burt, R. S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Colizza, V., Pastor-Satorras, R., Vespignani, A., 2007. Reaction-diffusion processes and metapopulation models in heterogeneous networks. Nature Physics 3, 276–282.

Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271.

Doreian, P., Batagelj, V., Ferligoj, A., 2005. Generalized Blockmodeling. Cambridge University Press, Cambridge, UK.

Freeman, L. C., 1978. Centrality in social networks: Conceptual clarification. Social Networks 1, 215–239.

Freeman, L. C., 2004. The Development of Social Network Analysis: A Study in the Sociology of Science. BookSurge, North Charleston, SC.

Freeman, L. C., Borgatti, S. P., White, D. R., 1991. Centrality in valued graphs: A measure of betweenness based on network flow. Social Networks 13 (2), 141–154.

Freeman, S. C., Freeman, L. C., 1979. The networkers network: A study of the impact of a new communications medium on sociometric structure. Social Science Research Reports S 46. University of California, Irvine, CA.

Granovetter, M., 1973. The strength of weak ties. American Journal of Sociology 78, 1360–1380.

Hansen, M. T., 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly 44, 232–248.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18, 39–43.

McPherson, J. M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. Annual Review of Sociology 27, 415–444.

Newman, M. E. J., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E 64, 016132.

Newman, M. E. J., 2004. Analysis of weighted networks. Physical Review E 70, 056131.

Opsahl, T., Colizza, V., Panzarasa, P., Ramasco, J. J., 2008. Prominence and control: The weighted rich-club effect. Physical Review Letters 101 (168702).

Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. Social Networks 31 (2), 155–163.

Peay, E. R., 1980. Connectedness in a general model for valued networks. Social Networks 2, 385–410.

Simmel, G., 1950. The Sociology of Georg Simmel (KH Wolff, trans.). Free Press, New York, NY.

Valente, T., 1995. Network Models of the Diffusion of Innovations. Hampton Press, Cresskill, NJ.

Wasserman, S., Faust, K., 1994. Social Network Analysis. Cambridge University Press, Cambridge, MA.

Watts, D. J., Strogatz, S. H., 1998. Collective dynamics of “small-world” networks. Nature 393, 440–442.

Yang, S., Knoke, D., 2001. Optimal connections: strength and distance in valued graphs. Social Networks 23, 285–295.