Network Science **2** (2): 277–297, 2014. © Cambridge University Press 2014
doi:10.1017/nws.2014.15

Subgroup centrality measures

JOCELYN R. BELL
Department of Mathematical Sciences, United State Military Academy, West Point, NY, 10996, USA
(e-mail: jocelyn.bell@usma.edu*)*

Abstract

In this paper we examine natural generalizations of four widely used centrality measures to subgroups of nodes in a network. This allows for a division into local and global influence. As an example, we analyze a classic network and discuss previously hidden features made visible by these new techniques. Network-wide measures and centralization formulae are derived.

Keywords: *centrality*, *attribute subgroups*, *centralization*

1 Introduction

Centrality measures are traditionally used to detect important nodes in a network. In many real world networks, nodes are divided by their attributes into subgroups. This paper details new methods for locating significant subgroup nodes in both the local and global sense. In a social network of both sexes, for instance, which women are most influential over other women, and which women are most influential over the men?

One potential solution is to examine the subgraph generated by a subgroup. However, calculating centrality measures on a subgraph ignores relationships outside the subgroup. When studying the spread of a cold virus through children at a school, it may be unreasonable to assume children are only infected by other children and not by adults as well. We explore a framework in which nodal centralities relative to a subgroup are calculated with no changes to the network. Because the network is not modified in any way, there is no loss of information.

We define both local measures (within the subgroup) and global measures (outside the subgroup). In contrast to the combined internal and external score of Everatt & Borgatti (2012), our approach considers local (internal) and global (external) centrality as separate pieces, allowing for rank order comparisons as well as comparisons to measures calculated on subgraphs. Furthermore we explicitly introduce normalized versions which take the size of the subgroup under investigation into account. As an illustrative example, in section 7 the classic dolphin data set (Lusseau et al., 2003) is analyzed using these new techniques. We discuss information which may have been missed by using only existing network science tools. Finally, in section 8 network-wide measures are discussed and the necessary theoretical maximums for centralization are computed.

2 Centrality

The networks considered in this paper are assumed to be connected, unweighted, and undirected. Let (V, E) be a network, where V is the set of nodes and E is the set of edges. We say a network is of size n if $|V| = n$. If node a is connected to node b, we write $(a, b) \in E$.

2.1 Centrality measures

A centrality measure assigns numerical values to nodes in a network based on structural properties. We are interested in the idea of centrality relative to subgroups of nodes. This is not a measure of the centrality of the subgroup as a whole; for group centrality see Everett & Borgatti (1999).

Many centrality measures can be described as follows. Suppose a particular relationship between nodes in a network can be quantified (for example, distance). This can be represented as a function of two variables

$$f : V \times V \to \mathbb{R}$$

where the value of the relationship between node a and node b is $f(a, b)$. For each node a, we may sum the values of its relationships to other network nodes

$$c(a) = \sum_{x \in V} f(a, x)$$

This function c is a type of centrality measure. Different relationships yield different centrality measures, however for this definition to make sense the relationship must be "node to node". There are centrality measures which are not of this variety. For example, betweenness is based on relationships between nodes and *pairs* of nodes. This kind of measure is *medial* (Everett & Borgatti, 2006) while a node to node type measure is *radial*. We postpone a discussion of medial centrality measures to section 6.

2.2 Subgroup centrality measures

Now suppose some subgroup S of nodes in a network (V, E) is of interest. To calculate measures on the nodes in S, we may form the subgraph generated by S. However relationships between nodes are redefined: the relationship between nodes a and b in the whole network may be different than their relationship in the subgraph. For example, in the subgraph the distance between two nodes may no longer be defined.

Instead of working with the subgraph, we will define a *subgroup measure* according to S as follows: for $a \in V$

$$c_S(a) = \sum_{x \in S} f(a, x)$$

where here the sum is restricted to only those nodes in S. The distinction between a subgroup measure and a subgraph measure is important. We say a subset $S \subseteq V$ of nodes is a *subgroup* of the network; note this is not the subgraph induced by S as it contains no edge information.

There are two special cases when $a \in S$. The subgroup measure of a according to S is a *local measure*, which measures how central node a is inside the subgroup. The subgroup measure of a according to S^c (everything except the nodes in S) is a *global measure*, which measures the centrality of a with respect to nodes outside the subgroup. It is evident from the definitions that the sum of a node's local measure and its global measure is the original measure.

Centrality measures are often normalized to produce values between 0 and 1. In the case of subgroup measures, normalization may be especially important as a small subgroup of a large network will likely have members with tiny subgroup centrality measures. To remedy this dependence on the size of the subgroup, normalized subgroup measures are introduced. Additionally, this allows for comparisons: does a node's subgroup centrality increase or decrease when the subgroup is changed?

3 Degree

Perhaps the most intuitive of the centrality measures is that of degree centrality. This measure has mathematical roots, as it is equivalent to (or proportional to, if normalized) graph-theoretic degree. The concepts of what we call local and global degree have appeared previously: see Wasserman & Faust (1994) for local degree and Arney & Peterson (2010) for global degree. A combination of the two, known as the E-I index, may be found in Krackhardt (1988).

The function which describes the degree relationship between node a and node b has value 1 if a and b are connected in the network, otherwise the value is 0. Symbolically, suppose (V, E) is a network and define $h : V \times V \to \mathbb{R}$ as

$$h(x, y) = \begin{cases} 1 & \text{if } (x, y) \in E \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

since then

$$deg(a) = \sum_{x \in V} h(a, x)$$

More precisely, this is out degree; however we have assumed our networks are undirected, thus out degree is identical to in degree. Subgroup degree centrality is as follows.

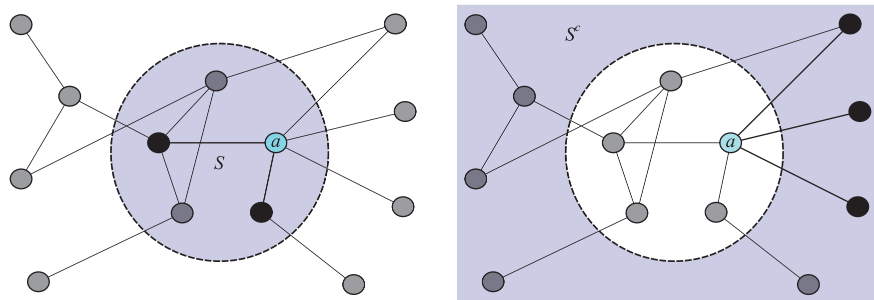
Definition 1

In a network (V, E) with subgroup S, for $a \in V$ the *subgroup degree* of a according to S is

$$deg_S(a) = \sum_{x \in S} h(a, x)$$

This is the subgroup measure defined by h (as in Equation (1)) and S, and is denoted by $deg_S(a)$. To normalize, divide by $(|S| - 1)$ if $a \in S$; otherwise divide by $|S|$.

This definition states that the subgroup degree (according to a subgroup S) of a node is the number of nodes, only from inside S, which are adjacent to it; see Figure 1. Subgroup degree of node a is equivalent to its standard degree in the subgraph generated by $S \cup \{a\}$.



(a) Local degree = 2. Normalized: $\frac{2}{4}$ (b) Global degree = 3. Normalized: $\frac{3}{8}$

Fig. 1. Local and global degree. Node *a*'s degree is split into two pieces. (color online)

4 Closeness

A node may be considered central if information can travel quickly from it to all others (Wasserman & Faust, 1994). This is the motivation for defining closeness centrality, which, as its name suggests, is based on the *distance* between nodes in a network. The smaller the total distance from a node to the rest of the network, the more central it is with respect to this measure.

Unlike degree, subgroup closeness may differ from subgraph closeness. Edge information is deleted in a subgraph, and so two nodes once reachable from each other may no longer be so. Let $d(x, y)$ denote the distance between node x and node y. Any distance function may be used; here we will only consider geodesic distance. "Farness" is perhaps a better description, since a smaller value means "closer". This is addressed by inverting when normalizing.

Definition 2
In a network with subgroup S, *subgroup closeness centrality* according to S of node a is

$$cl_S(a) = \sum_{y \in S} d(a, y)$$

To normalize, divide into $|S| - 1$ if $a \in S$; otherwise divide into $|S|$.

That is, subgroup closeness adds the distances from a only to nodes in S. It does not matter if a geodesic uses nodes from outside S. This is one possible advantage versus the subgraph approach; subgroup closeness is always defined (for a connected network) regardless of whether or not the subgraph is connected.

How does subgroup closeness differ from closeness in the subgraph? Subgraph paths between nodes can only be longer (if they exist at all) than subgroup paths. However, rankings may be different. The closest subgraph node need not be the closest subgroup node; see Figure 2.

5 Eigenvector centrality

In a social network, an individual may be described as influential if she has powerful friends. Here, the status of one's connections rather than just the number of ties is what matters; a node's rank is proportional to the sum of the ranks of its

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

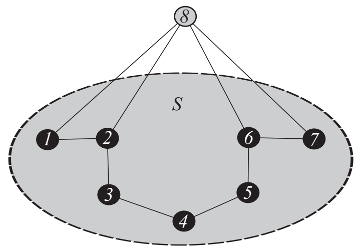


Fig. 2. Subgroup and Subgraph Closeness Centrality. Nodes 2 and 6 are top ranked by the subgroup closeness measure, whereas node 4 is ranked highest in the subgraph.

neighbors. This circular definition leads to a system of equations and unknowns, the solution to which is commonly termed eigenvector centrality. This measure is originally attributed to Seeley (1949), and subsequent generalizations to Katz (1953) and Bonacich (1987) among others.

Suppose a network (V, E) has adjacency matrix A with leading eigenvalue λ and corresponding unit eigenvector $\overrightarrow{v}$, and node a's standard eigenvector centrality is denoted by $e(a)$. The function required is

$$g(x, y) = \begin{cases} \frac{1}{\lambda} \cdot e(y) & \text{if } (x, y) \in E \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

since then

$$e(a) = \sum_{y \in V} g(a, y)$$

More precisely, this is "left" eigenvector centrality, however we have assumed our networks are undirected so this is not a concern.

Definition 3
For a subgroup of nodes S in a network, *subgroup eigenvector centrality* according to S is

$$e_S(a) = \sum_{y \in S} g(a, y)$$

where g is as in Equation (2).

It may be desirable to rescale the values to take into account the size of the subgroup. Since eigenvector centrality is traditionally defined by a unit vector, similarly we ensure the squares of the subgroup measures sum to 1. In this case a distinction must be made between the local and global cases.

Definition 4
Suppose S is a subset of nodes in a network. For $a \in S$, *normalized local eigenvector centrality* according to S is

$$e'_S(a) = \frac{e_S(a)}{\sqrt{\sum_{a_i \in S} e_S(a_i)^2}}$$

The *normalized global eigenvector centrality* according to S is

$$Ge'_S(a) = \frac{e_{S^c}(a)}{\sqrt{\sum_{a_i \in S} e_{S^c}(a_i)^2}}$$

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Does subgroup eigenvector differ from subgraph eigenvector centrality? The answer is it may, and strikingly so; we will see such an example in section 7 (Figure 6).

6 Betweenness

Betweenness centrality is fundamentally different in its calculation than degree, closeness or eigenvector. It measures to what extent a particular node impacts connections between other pairs of nodes. This concept is generally attributed to L.C. Freeman, whose work on this type of measure may be found in Freeman (1977).

Because betweenness is a medial centrality measure (Everett & Borgatti, 2006), it cannot be described using a node to node relationship, but rather requires a node to *pair of nodes* relationship. How is a node a in a network (V, E) related to a pair of nodes $\{b, c\}$? Let $P(b, c)$ be the list of all shortest paths between nodes b and c. For a path $p \in P(b, c)$ and node $a \in V$ let

$$Q(a, p) = \begin{cases} 1 & \text{if } a \text{ is on } p \text{ and } a \neq b \text{ and } a \neq c \\ 0 & \text{otherwise} \end{cases}$$

and

$$B(a, \{b, c\}) = \sum_{p \in P(b,c)} \frac{Q(a, p)}{|P(b, c)|}$$

Then betweenness centrality of node a is

$$b(a) = \frac{1}{2} \sum_{(b,c) \in V \times V} B(a, \{b, c\})$$

where the factor of $\frac{1}{2}$ appears due to counting each path twice.

6.1 Subgroup betweenness

Subgroup betweenness is defined by only summing over pairs of nodes which are both members of the subgroup.

Definition 5

In a network with a subgroup S, the *subgroup betweenness* of a according to S is given by

$$b_S(a) = \frac{1}{2} \sum_{(b,c) \in S \times S} B(a, \{b, c\})$$

To normalize, divide into the total number of such possible pairs, which is $\frac{1}{2}(|S| - 1)(|S| - 2)$ if $a \in S$ and $\frac{1}{2}(|S|)(|S| - 1)$ if $a \notin S$.

If $a \in S$, the *local betweenness* of a is subgroup betweenness according to S, while *global betweenness* is subgroup betweenness according to S^c. However, the local and global versions of betweenness do not tell the whole story, as they do not account for pairs of nodes for which one is within the subgroup and one is outside. To this end, we introduce a third subgroup betweenness measure we call *boundary betweenness*, as the paths must cross the "boundary" of S. See Figure 3.

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

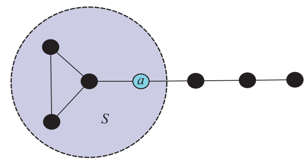


Fig. 3. Boundary Betweenness. Node a has zero local and global betweenness, but every path starting inside S and ending outside must pass through a, making a's normalized boundary betweenness equal to 1. (color online)

Definition 6
In a network (V, E) with subgroup S, the *boundary betweenness* of a according to S is

$$Bb_S(a) = \sum_{(b,c)\in S\times S^c} B(a, \{b, c\})$$

To normalize, divide by the total number of such pairs, which is $(|S| - 1)(|V| - |S|)$ if $a \in S$ and $(|S|)(|V| - |S| - 1)$ if $a \notin S$.

If S is the entire network or a single point, boundary betweenness is zero. Notice for any node a and any subgroup S,

$$b_S(a) + b_{S^c}(a) + Bb_S(a) = b(a)$$

This equation may fail to hold if normalized versions are used instead.

6.2 Local clustering coefficient

In classic social network analysis, the local clustering coefficient of a node is a measurement of the likelihood that any two of its neighbors are connected (Newman, 2010). It can be used to locate structural holes. For a node a in a network, the *local clustering coefficient* C_a of a is the ratio

$$C_a = \frac{\#\text{ of pairs of connected neighbors of } a}{\#\text{ of pairs of neighbors of } a}$$

There is a relationship to local betweenness. For each node a in a network, let S_a be the subgroup containing the node a and all of its neighbors. Then

$$b'_{S_a}(a) = 1 - C_a$$

where the prime indicates normalization. Thus, local betweenness is a generalization of local clustering.

7 Example analysis: Dolphin network

To illustrate the ideas put forth in this paper, we will analyze the classic dolphin network in Lusseau et al. (2003). This network consists of what we will call "friendship" ties between 62 bottle-nose dolphins in Doubtful Sound. This network was chosen due to its small size and low density and our analysis is not intended to draw conclusions about dolphin relationships.

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

7.1 Attribute subgroups

In this section, our subgroup of interest will be the group of male dolphins; see Figure 4. This represents a division according to node attribute data, in contrast to divisions based on structural communities. Tables including rankings by centrality measure are located in Appendix B while the algorithms used may be found in Appendix A.

A visual inspection of Figure 4 reveals a few immediate concerns. There are male dolphins who have become isolates in the subgraph (Cross, Fork, and Zig). Furthermore, it would appear that SN96 should be quite important, as his removal would split the subgraph network into three disjoint pieces: the Topless group, the Beescratch group and the Bumper group. However, looking at the network as a whole, female dolphins hold these three groups of males together; SN96's removal no longer causes the network to splinter.

The top four dolphins by subgroup and subgraph betweenness are represented in Figure 5. The subgraph overestimates the importance of SN96, PL, Beak and Haecksel due to deleting relationships to female dolphins. What might this mean? For studying rumor spread among male dolphins, the subgroup rather than subgraph approach to betweenness would be appropriate, assuming communications may spread through female dolphins as well. All gossip need not go through SN96; in fact Beescratch is more important in this regard. For a communicable disease affecting and carried solely by male dolphins the subgraph approach is preferable; vaccinating SN96 is indeed a good idea. However, if females are also carriers then Beescratch is a better choice.

There are less obvious differences between subgraph and subgroup measures. The most striking of these is eigenvector centrality (Figure 6). In the subgraph of male dolphins (Figure 6(a)), members of the Beescratch group have higher eigenvector centrality than the members of the Topless group. However the subgroup measures display the opposite trend: the Topless group members are ranked higher than the Beescratch group (Figure 6(b)). The power structure is vastly different for these two measures. How can this be interpreted? As eigenvector centrality measures a node's value based on the value of its neighbors, in the subgraph case the values of a node's neighbors are due exclusively to relationships with other males; in the subgroup case, the values of neighbors include relationships with females as well. For example, Topless has six male and five female friends. His high importance in turn impacts each of his male neighbors, increasing their eigenvector centrality in the subgroup sense. However in the subgraph sense, only his six male friends are considered, so there his influence is less. In Jet we see the complementary situation: all of Jet's nine friends are male, so his initial value in the subgraph is high and does not increase by examining the total network.

Some general information may be found by computing correlation coefficients. For subgroup versus subgraph eigenvector this coefficient is −.36457, indicating (not surprisingly, in light of Figure 6) the lack of a clear linear relationship between the two. Subgroup and subgraph betweenness are also poorly correlated (.321381), consistent with Figure 5. Normalized closeness is better, with a subgroup to subgraph coefficient of .803556. Interestingly, boundary betweenness very closely

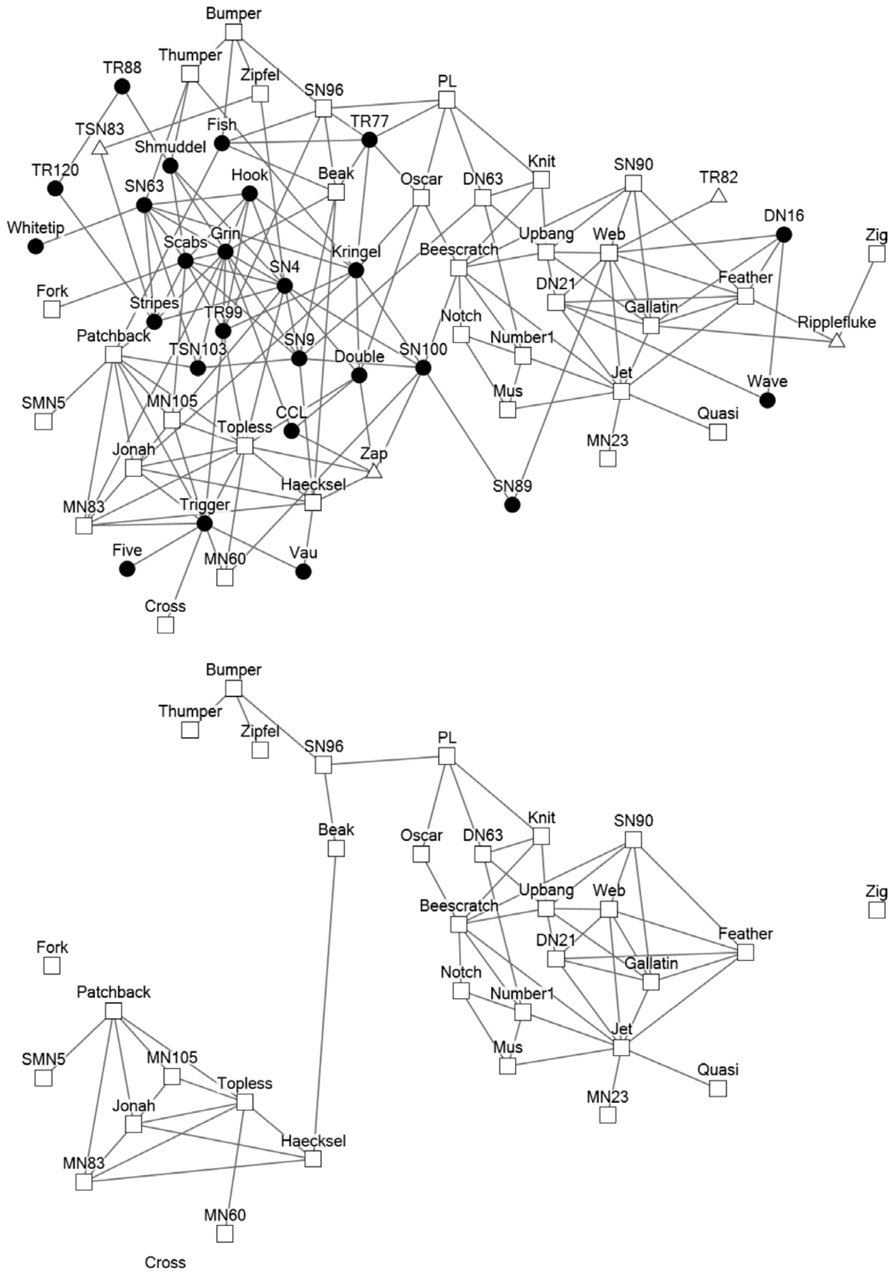


Fig. 4. Dolphin Network. The top graph is the entire dolphin network. Males are represented by squares, females by solid circles, and those of unknown gender by triangles. The bottom graph is the subgraph generated by only male dolphins. In the subgraph, Fork, Cross and Zig have become isolates, while SN96 appears to hold the Topless (lower left), Beescratch (upper right) and Bumper (middle) groups together.

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

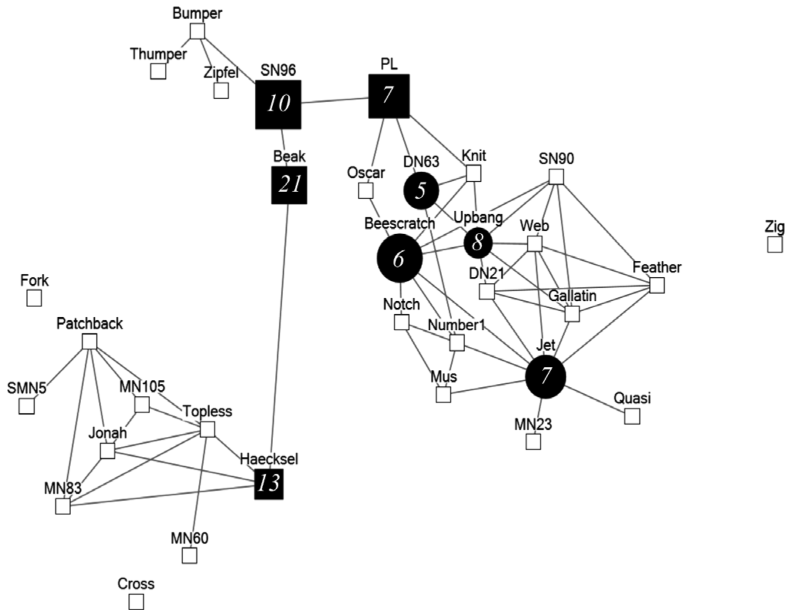


Fig. 5. Subgraph and Subgroup Betweenness, Top Four. Subgraph betweenness is represented by solid squares and subgroup by solid disks, sized according to rank. The values indicate the rank in the other measure.

follows overall betweenness with a correlation coefficient of .995942. It is unclear whether or not this is to be expected or an artifact of this network.

How do local and global measures compare? A dramatic difference exists for closeness. Figure 7 is a visual representation of top ranked dolphins by local and global closeness. The dolphins in the Topless group are generally closer in a global sense, while those in the Beescratch group are closer in a local sense. As a possible interpretation, a communication from Topless will spread through the female dolphin community quickly, while one from Beescratch will spread rapidly through the male community. If a particular message is highly important to female dolphins but not to males, Topless is an ideal messenger.

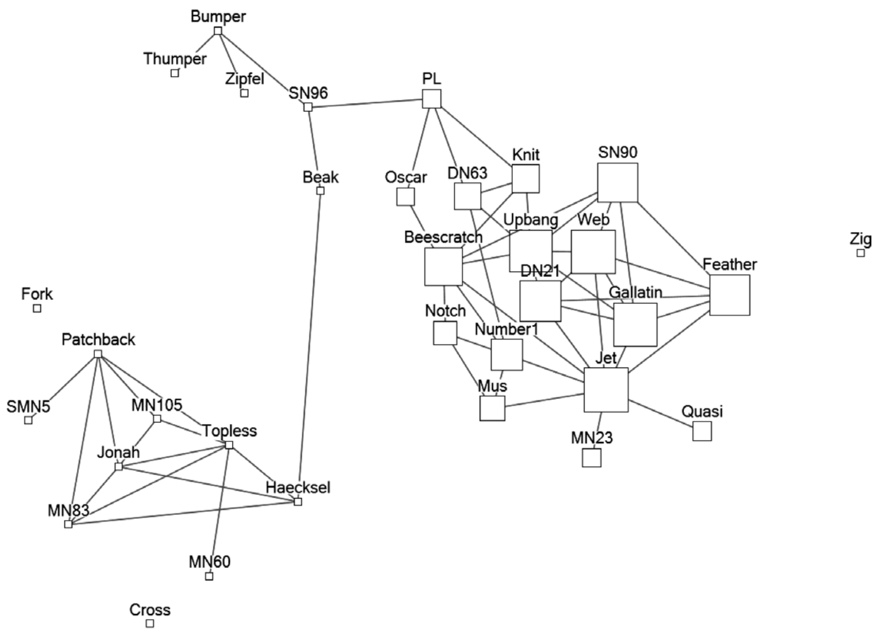
The top five ranked males in each type of closeness measure is given in Table 1. There are interesting individuals listed in this table. Patchback is only close in the global sense (to female dolphins- a sort of "ladies' man"); neither subgroup nor subgraph nor overall closeness ranks him among the top ten. Number 1 is only close to other males; he is ranked 20^{th} globally and 14^{th} overall. Without this subgroup analysis, the importance of these two may have been overlooked.

7.2 Structural subgroups

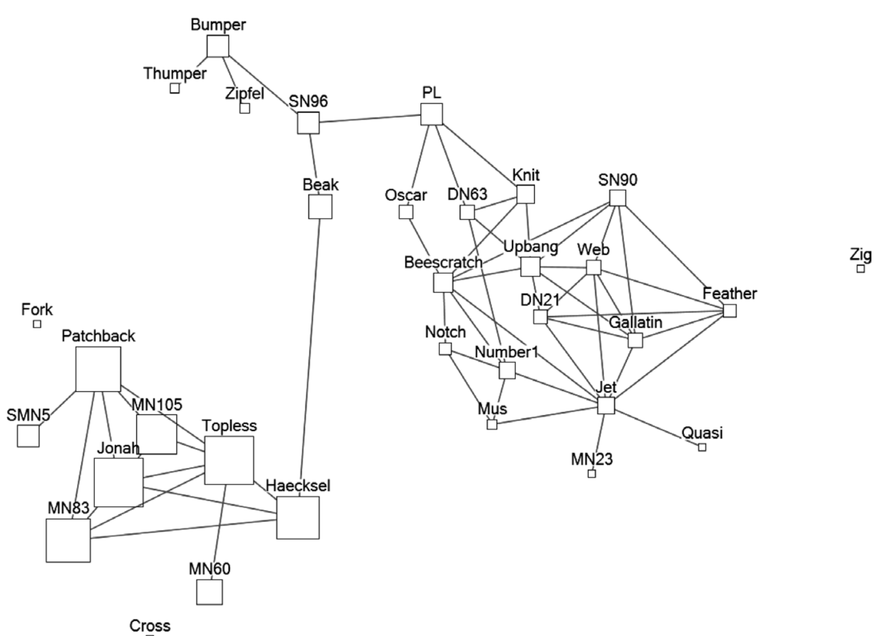
In the first example, the dolphin network was divided by nodal attributes into subgroups. What happens if divisions are made based on network structure? In Lusseau & Newman (2004), two distinct communities are identified using a betweenness-based algorithm. These communities are depicted in Figure 8.

As might be expected, there is much less difference between the subgroup and subgraph measures when divisions are made based on structure. In fact, for

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press



(a) Subgraph Eigenvector



(b) Subgroup Eigenvector

Fig. 6. Eigenvector Measures. Nodes are sized according to their eigenvector centrality. In figure (a), the members of the Beescratch group are ranked higher than the Topless group, whereas in figure (b) the Topless group is higher ranked.

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Table 1. *Closeness. Top five ranked, listed in decreasing order. An asterisk or double asterisk indicates a tie.*

Local	Subgraph	Global	Overall
Beescratch	PL	Beak*	Beescratch
DN63	DN63	Topless*	DN63*
Oscar	Knit	MN105	Oscar*
Number 1*	SN96	Jonah*	Beak**
Upbang*	Oscar	**Patchback***	Topless**

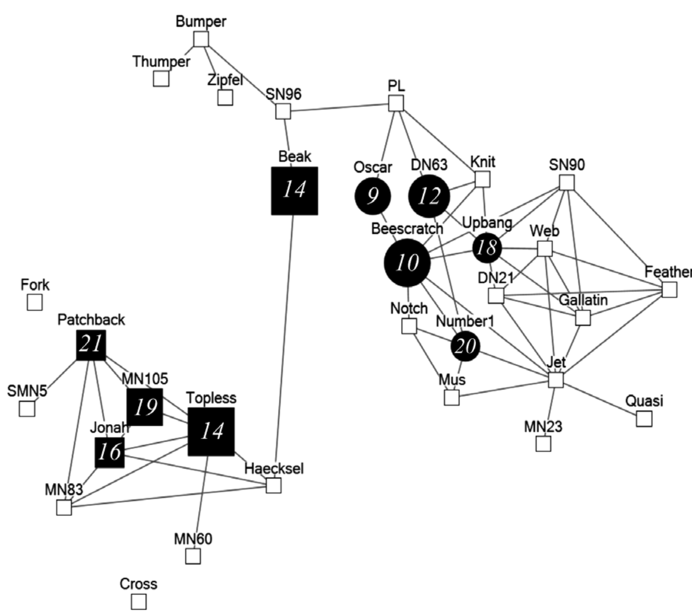


Fig. 7. Local and Global Closeness: Top Five. Local closeness is represented by solid disks and global by solid squares, sized according to rank. The values indicate the node's rank in the other measure.

Community 1 in Figure 8, there is a perfect correlation between both subgroup versus subgraph closeness and subgroup versus subgraph betweenness. This is because there are *no* shortest paths between Community 1 members which pass through Community 2 members. For Community 2, the respective correlations are .9998 and .9993.

Interestingly, however, there is still a difference for Community 1 with respect to eigenvector centrality, with a correlation coefficient of .7342. Such a distinction is not present in Community 2, where this coefficient is .9999. In Community 1, Gallatin is ranked 1^{st} in subgraph eigenvector and 6^{th} in subgroup, while Upbang is ranked 6^{th} in subgraph eigenvector and 1^{st} in subgroup. This is likely due to the fact that Upbang is connected to Knit, Beescratch and DN64, who in turn are connected to Community 2. Gallatin has no neighbors with connections outside Community 1.

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

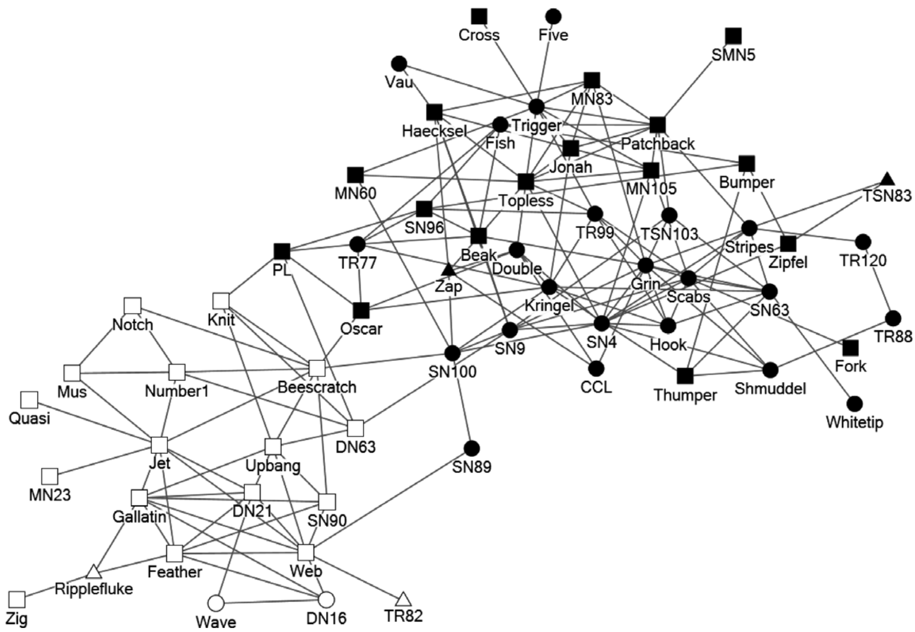


Fig. 8. Structural Dolphin Communities. Shape indicates gender: Males are square, females circles, and unknowns are triangles. Community 1 appears in white, Community 2 in black.

8 Network-wide measures

We define average subgroup degree and other standard network-wide measures in the obvious way. Because some information may be gleaned from examining correlation coefficients between subgroup and subgraph or global to overall measures, these could be considered network-wide measures as well. For instance, a perfect correlation between subgroup and subgraph closeness means not only is the subgraph connected, but also for every pair of nodes in the subgroup there is a shortest path between them which passes through only subgroup nodes.

The remainder of this section deals with centralization. Network centralization is a way to measure how much variation in centrality is present in the network. We use the classic Freeman definition (see Freeman, 1979). In order to calculate it, the maximum possible variation must be known. This maximum is calculated over all possible networks of the same size and is the denominator in the centralization formula. We calculate this quantity for local as well as global measures.

Proposition 1
In a network with subgroup S, maximum local degree centralization occurs when the subgraph induced by S is a star graph, and this maximum is $(|S|-1)(|S|-2)$.

Proof
If every node in S has non-zero local degree, the result follows. Suppose S contains m many nodes with zero local degree, and let b be a node with maximum local degree. Then $deg_S(b) \leqslant |S|-m-1$ and the remaining $|S|-m-1$ many nodes have local degree at least 1. So $\sum_{i=1}^{|S|} deg_S(b) \leqslant |S|\cdot(|S|-m-1)$ and $\sum_{v\in S} deg_S(v) \geqslant 2(|S|-m-1)$ and therefore the maximum difference is at most $(|S|-m-1)(|S|-2)$. Clearly this is less than $(|S|-1)(|S|-2)$ if $m > 0$. □

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

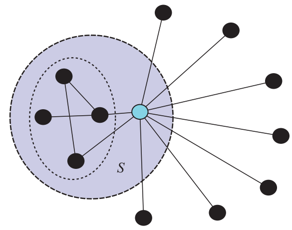
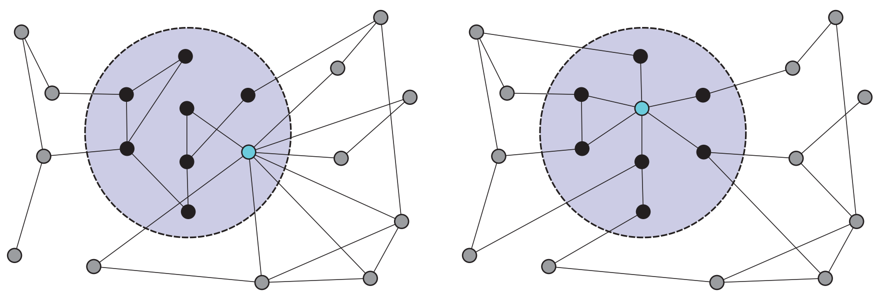


Fig. 9. Maximum Global Degree Centralization. It does not matter how the nodes in S are connected to each other, so long as none of them except the highlighted node are connected to any nodes outside of S. (color online)



(a) Local: .1905, Global: .5476 (b) Local: .7619, Global: .0952

Fig. 10. Local and Global Degree Centralization. Figure (a) has a clear global leader with no local leaders, while Figure (b) depicts the opposite situation. (color online)

If using normalized local degree centrality, this value becomes $|S|-2$. The global argument is obvious; see Figure 9.

Proposition 2
In a network (V, E) with subgroup S, the theoretical maximum for global degree centralization is $(|V|-|S|)(|S|-1)$. Normalized, this becomes $|S|-1$.

If S is the entire network, then local degree centralization is the same as degree centralization for the whole network in the usual sense. It is possible to have high local degree centralization and low global degree centralization or vice versa. For example, high global and low local degree centralization could mean the network has a small number of powerful liaisons between the group and the rest of the network, yet within the group there are no local leaders. See Figure 10.

It is easy to see that maximum subgroup betweenness centralization occurs when the subgroup betweenness of one subgroup node is maximal and all other subgroup nodes have zero subgroup betweenness. This happens, for example, in a network which is a star graph and the subgroup contains the center of the star.

Proposition 3
In a network (V, E) with subgroup S, the local betweenness centralization maximum is $\frac{1}{2}(|S|-1)^2(|S|-2)$, the global maximum is $\frac{1}{2}(|S|-1)(|V|-|S|)(|V|-|S|-1)$ and the boundary maximum is $(|S|-1)^2(|V|-|S|)$. Normalized, all three become $(|S|-1)$.

Global betweenness centralization is zero if there is only one connection between the subgroup S and the rest of the network (a bridge).

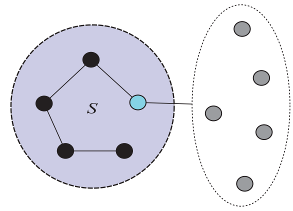


Fig. 11. Maximum Global Closeness Centralization. It does not matter how the nodes outside S are connected, as long as the only element of S they are connected to is the highlighted node. (color online)

For local closeness centralization, the argument in Everett et. al. (2004) goes through using normalized local closeness, and hence this maximum is

$$\frac{\left(|S|-1\right)\left(|S|-2\right)}{2\,|S|-3}$$

Unfortunately, this trick does not work for normalized global closeness centralization. Instead, we work with non-normalized global closeness, in which case a smaller value means "closer". Thus instead of summing the difference from the maximum, we sum the difference to the minimum. Also we add an additional assumption.

A subgroup is *locally connected* provided the subgraph it induces is connected.

Proposition 4

The maximum global closeness centralization for a locally connected subgroup S in a network (V, E) is $\frac{1}{2}(|V|-|S|)(|S|\cdot(|S|-1))$.

Proof

Denote global closeness by Gcl_S. Suppose $a \in S$ is such that $\sum_{y\notin S} d(a,y)$ is as small as possible and $b \in S$ and $y \notin S$. Then $d(b,y)-d(a,y) \leqslant d(a,b)$ and there are $(|V|-|S|)$ many such y's, hence

$$\sum_{y\notin S}\left(d\left(b,y\right)-d\left(a,y\right)\right) \leqslant \left(|V|-|S|\right)\cdot d\left(a,b\right)$$

from which we see $Gcl_S(a)-Gcl_S(b) \leqslant (|V|-|S|)\cdot d(a,b)$. Now sum over all $b \in S$ to get

$$\sum_{b\in S}\left(Gcl_S\left(b\right)-Gcl_S\left(a\right)\right) \leqslant \left(|V|-|S|\right)\sum_{b\in S} d\left(a,b\right) = \left(|V|-|S|\right)\cdot cl_S\left(a\right)$$

The maximum for the right hand side is attained when $cl_S(a)$ is as large as possible; in a locally connected subgroup this is $\sum_{i=1}^{|S|-1} i = \frac{1}{2}\cdot(|S|\cdot(|S|-1))$, when the subgraph is a line graph. Equality is attained by a network as in Figure 11. □

A schematic of maximum global closeness centralization is given in Figure 11. The assumption of local connectedness is necessary; see Figure 12.

To complete the analysis of the male subgroup of the dolphin network, centralization values appear in Table 2. Notably, local and global closeness centralization are quite low, indicating little variation in these measures, while the local and boundary betweenness measures display the most variation.

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Table 2. *Centralization: Male Dolphin Network. Global closeness centralization is calculated using the denominator for locally connected subgroups.*

	Degree	Closeness	Betweenness
Local	.1825	.02540	.2519
Global	.1228	.05414	.1283
Boundary	N/A	N/A	.2096

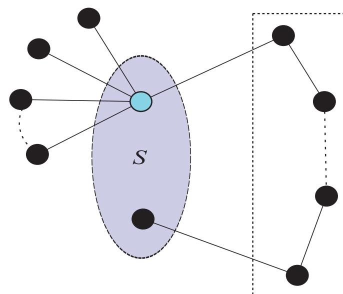


Fig. 12. Locally Disconnected: Global Closeness Centralization. Suppose the box on the right contains l many nodes and the total number of nodes is n. The global closeness centralization is $(n - l - 2)(l + 1)$, which is larger than the theoretical maximum for locally connected subgroups, in this case $(n - 2)$, provided $l > 0$ and $n > l + 3$. (color online)

9 Conclusion

We have proposed straightforward methods for calculating the local and global influence of a node relative to a subgroup of a network. The subgroup measures defined are natural generalizations of degree, closeness, betweenness, and eigenvector centrality. These new measures have the potential to uncover important information regarding network subgroup properties not visible in the total graph nor in the subgraph. Our analysis of male dolphins in the classic dolphin network has revealed such information.

We have focused exclusively on unweighted and undirected networks however this restriction is not necessary. Our flexible definitions also create subgroup versions of many other commonly used centrality measures. Alternatives in place of geodesic distance for closeness and only counting certain paths for betweenness will yield such measures. For example, subgroup decay and subgroup k-path centrality may be defined this way.

We have assumed all edges carry the same weight, however simple modifications may be made to, for example, weight edges so that communications “prefer” to travel within the subgroup if possible. This may be accomplished by using the subgroup to establish weights on the entire network and then computing subgroup measures. A natural weighting system would be to give highest weight to edges within the group, less weight to those between the group and its complement, and least weight to edges entirely outside the group.

References

Arney, D. C., & Peterson, E. (2010). Cooperation in a network: Agent metrics and cooperation space. In *Proceedings of the Army Science Conference.*

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Bonacich, P. F. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, **92**, 1170–1182.
Everett, M. G., & Borgatti, S. P. (1999). The centrality of groups and classes. *Journal of Mathematical Sociology*, **23**, 181–201.
Everett, M. G., & Borgatti, S. P. (2012). Categorical attribute based centrality: E-I and G-F centrality. *Social Networks* **34**(4), 562–569.
Everett, M. G., & Borgatti, S. P. (2006). A Graph-theoretic perspective on centrality. *Social Networks*, **28**(4), 466–484.
Everett, M. G., Sinclair, P., & Dankelmann, P. (2004). Some centrality results new and old. *Journal of Mathematical Sociology*, **28**, 215–227.
Freeman, L. C. (1977). A set of measures based on betweenness. *Sociometry*, **40**, 35–41.
Freeman, L. C. (1979). Centrality in social networks: I. Conceptual clarification. *Social Networks*, **1**, 215–239.
Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, **18**(1), 39–43.
Krackhardt, D., & Stern, R. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, **51**(2), 123–140.
Lusseau, D., & Newman, M. E. J. (2004). Identifying the role that animals play in their social networks. *Proceedings of the Royal Society of London*, Series B: Biological Sciences, **271**(Suppl 6), S477–S481.
Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long lasting associations. *Behavioral Ecology and Sociobiology*, **54**, 396–405.
Newman, M. E. J. (2010). *Networks, an introduction*. Oxford: Oxford University Press.
Seeley, J. R. (1949). The net of reciprocal influence; a problem in treating sociometric data. *Canadian Journal of Psychology/Revue canadienne de psychologie*, **3**(4), 234–240.
Wasserman, S., & K. Faust (1994). *Social Network Analysis*. Cambridge: Cambridge University Press.

Appendix A: Algorithms

Since subgroup degree is simply subgraph degree we do not include an algorithm. All algorithms are written in python and require networkx. G is a connected graph and S is a subgroup of network nodes. The measures are normalized by default but this is optional.

Algorithm 1 Closeness

```
import networkx as nx

def subgroup_closeness(G, S, normalized=True):
    subgroup_closeness = {}
    for b in nx.nodes(G):
        sum = 0
        for a in S:
            sum += nx.shortest_path_length(G, a, b)
        subgroup_closeness[b] = sum
        if normalized:
            if sum > 0.0 and b in S:
                subgroup_closeness[b] = (len(S)-1.0)/sum
            if sum > 0.0 and b not in S:
                subgroup_closeness[b] = (len(S))/sum
    return subgroup_closeness
```

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Algorithm 2 Betweenness

```
import networkx as nx

def subgroup_betweenness(G, S, normalized=True):
    subgroup_betweenness = {}
    npairs = (len(S)-1)*(len(S)-2)
    npairs2 = (len(S))*(len(S)-1)
    for c in nx.nodes(G):
        sum = 0
        for a in S:
            for b in S:
                Pab = list(nx.all_shortest_paths(G, a, b))
                Pabc = [p for p in Pab if c in p and a!=c and b!=c]
                sum += len(Pabc) / len(Pab)
        subgroup_betweenness[c] = sum / 2
        if normalized and c in S and npairs > 0:
            subgroup_betweenness[c] = sum / npairs
        if normalized and c not in S and npairs2 > 0:
            subgroup_betweenness[c] = sum / npairs2
    return subgroup_betweenness
```

Algorithm 3 Boundary Betweenness

```
import networkx as nx

def boundary_betweenness(G, S, normalized=True):
    boundary_betweenness = {}
    B = [i for i in nx.nodes(G) if i not in S]
    denom = (len(S)-1)*(len(B))
    denom2 = (len(S))*(len(B)-1)
    for c in nx.nodes(G):
        sum = 0
        for a in S:
            for b in B:
                Pab = list(nx.all_shortest_paths(G, a, b))
                Pabc = [p for p in Pab if c in p and a!=c and b!=c]
                sum += len(Pabc)/len(Pab)
        boundary_betweenness[c] = sum
        if normalized and c in S:
            boundary_betweenness[c] = sum / denom
        if normalized and c not in S:
            boundary_betweenness[c] = sum / denom2
    return boundary_betweenness
```

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Algorithm 4 Eigenvector

```
import networkx as nx
from math import sqrt
import numpy as np

def subgroup_eigenvector(G, S):
    subgroup_eigenvector = {}
    A = nx.adj_matrix(G,nodelist=G.nodes())
    eigenvalues = np.linalg.eigvals(A)
    lead = max(eigenvalues).real
    eig = nx.eigenvector_centrality_numpy(G)
    for a in nx.nodes(G):
        sum = 0
        for b in nx.neighbors(G, a):
            if b in S:
                sum += eig[b]
        subgroup_eigenvector[a] = sum / lead
    return subgroup_eigenvector

def local_eigenvector(G, S, normalized=True):
    local_eigenvector = {}
    C = subgroup_eigenvector(G, S)
    V = {k: C[k]*C[k] for k in S}
    denom = sqrt(sum(V.values()))
    for a in S:
        local_eigenvector[a] = C[a]
        if normalized:
            local_eigenvector[a] = C[a] / denom
    return local_eigenvector

def global_eigenvector(G, S, normalized=True):
    global_eigenvector = {}
    B = [i for i in nx.nodes(G) if i not in S]
    C = subgroup_eigenvector(G, B)
    V = {k: C[k]*C[k] for k in S}
    denom = sqrt(sum(V.values()))
    for a in S:
        global_eigenvector[a] = C[a]
        if normalized:
            global_eigenvector[a] = C[a] / denom
    return global_eigenvector
```

https://doi.org/10.1017/nws.2014.15 Published online by Cambridge University Press

Appendix B: Data tables

		DEGREE			CLOSENESS				BETWEENNESS					EIGENVECTOR			
ID	NAME	L	G	O	L	S	G	O	L	S	G	B	O	L	S	G	O
0	**Beak**	23	2	11	14	9	1	4	21	3	9	19	18	8	20	3	7
1	**Beescratch**	2	16	5	1	6	10	1	1	6	4	1	1	13	8	16	14
2	**Bumper**	14	16	21	27	14	17	22	19	11	23	25	24	11	19	18	16
4	**Cross**	31	16	28	31	31	25	29	28	23	25	28	28	31	31	14	19
6	**DN21**	7	16	11	18	13	28	26	25	20	5	12	13	22	5	23	28
7	**DN63**	13	16	16	2	2	12	2	3	5	6	2	2	19	11	15	13
9	**Feather**	7	10	7	25	19	29	28	18	23	14	17	17	24	6	20	27
12	**Fork**	31	16	28	30	31	15	25	28	23	25	28	28	31	31	13	17
13	**Gallatin**	4	10	5	17	11	27	24	9	18	7	8	8	20	2	20	26
15	**Haecksel**	13	4	7	11	16	8	6	13	4	8	11	11	5	21	9	6
17	**Jet**	1	24	2	7	10	20	18	2	7	13	3	3	16	1	24	22
18	**Jonah**	7	10	7	16	21	4	7	17	15	22	21	19	2	25	8	4
19	**Knit**	13	24	21	6	3	19	14	15	9	24	23	21	15	10	24	21
21	**MN105**	14	4	11	19	28	3	8	20	23	17	22	22	6	28	2	3
22	**MN23**	25	24	28	28	21	31	31	28	23	25	28	28	29	14	24	31
23	**MN60**	25	10	24	11	29	10	8	11	23	11	9	9	7	29	11	8
24	**MN83**	13	10	11	23	24	6	16	27	16	20	24	25	4	26	5	5
25	**Mus**	14	24	24	21	17	30	27	26	21	25	27	27	26	12	24	30
26	**Notch**	14	24	24	19	18	26	23	23	22	25	26	26	25	13	24	29
27	**Number1**	7	24	16	4	7	20	14	8	12	25	15	14	17	9	24	24
28	**Oscar**	23	4	16	3	5	9	2	6	13	12	6	6	23	17	7	11
29	**Patchback**	7	2	2	21	27	4	12	5	14	3	7	7	3	27	4	2
30	**PL**	13	16	16	8	1	15	12	7	2	21	13	12	9	16	17	15
31	**Quasi**	25	24	28	28	21	31	31	28	23	25	28	28	29	14	24	31
35	**SMN5**	25	24	28	32	30	24	30	28	23	25	28	28	10	30	24	18
41	**SN90**	7	24	16	13	14	23	21	14	17	15	16	16	18	7	24	25
42	**SN96**	14	4	11	9	4	13	10	10	1	19	14	15	12	18	9	9
44	**Thumper**	25	4	21	25	25	6	17	22	23	18	20	23	27	22	6	10
45	**Topless**	4	1	1	14	20	1	4	12	10	10	10	10	1	24	1	1
54	**Upbang**	2	24	7	5	8	18	11	4	8	2	4	4	14	4	24	20
57	**Web**	4	4	2	9	11	20	19	16	19	1	5	5	21	2	19	23
60	**Zig**	31	16	28	33	31	33	33	28	23	25	28	28	31	31	22	33
61	**Zipfel**	25	10	24	23	25	13	19	24	23	16	18	20	27	22	12	12

Subgroup measures: by ranking. L stands for local, G for global, S for subgraph, O for overall and B for boundary. In the event of a tie, identical ranks are awarded and the next rank is skipped.

		DEGREE			CLOSENESS				BETWEENNESS					EIGENVECTOR			
ID	NAME	L	G	O	L	S	G	O	L	S	G	B	O	L	S	G	O
0	**Beak**	2	4	6	101	105	75	176	3.702	168.000	15.003	16.216	34.921	0.034	0.002	0.094	0.129
1	**Beescratch**	7	1	8	80	98	84	164	142.494	76.353	24.188	223.702	390.384	0.024	0.274	0.018	0.042
2	**Bumper**	3	1	4	120	117	96	216	6.767	55.000	2.300	7.537	16.603	0.029	0.002	0.010	0.040
4	**Cross**	0	1	1	137	0	108	245	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.029
6	**DN21**	5	1	6	106	116	122	228	2.264	3.663	22.776	28.711	53.752	0.012	0.318	0.000	0.012
7	**DN63**	4	1	5	81	91	86	167	78.645	84.944	22.628	115.104	216.377	0.014	0.133	0.029	0.043
9	**Feather**	5	2	7	116	132	126	242	9.171	1.167	9.088	19.978	38.237	0.011	0.311	0.001	0.012
12	**Fork**	0	1	1	132	0	95	227	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.039
13	**Gallatin**	6	2	8	104	115	121	225	25.427	4.830	21.772	49.510	96.709	0.014	0.364	0.001	0.015
15	**Haecksel**	4	3	7	99	119	81	180	14.097	154.000	16.370	30.458	60.925	0.112	0.001	0.052	0.164
17	**Jet**	9	0	9	93	110	104	197	93.971	74.786	10.410	104.788	209.169	0.018	0.374	0.000	0.018
18	**Jonah**	5	2	7	103	138	78	181	10.955	27.167	3.044	13.185	27.184	0.148	0.000	0.055	0.202
19	**Knit**	4	0	4	92	92	101	193	13.742	67.462	1.762	8.862	24.365	0.021	0.147	0.000	0.021
21	**MN105**	3	3	6	107	162	76	183	6.082	0.000	5.545	11.616	23.242	0.097	0.000	0.110	0.207
22	**MN23**	1	0	1	124	138	133	257	0.000	0.000	0.000	0.000	0.000	0.002	0.068	0.000	0.002
23	**MN60**	1	2	3	99	165	84	183	22.168	0.000	13.283	41.743	77.194	0.040	0.000	0.048	0.087
24	**MN83**	4	2	6	115	139	80	195	1.321	15.333	3.884	8.306	13.511	0.120	0.000	0.073	0.193
25	**Mus**	3	0	3	111	120	128	239	2.122	1.900	0.000	0.887	3.009	0.006	0.119	0.000	0.006

		DEGREE			CLOSENESS				BETWEENNESS					EIGENVECTOR			
ID	NAME	L	G	O	L	S	G	O	L	S	G	B	O	L	S	G	O
26	**Notch**	3	0	3	107	122	113	220	3.248	1.226	0.000	4.736	7.983	0.009	0.104	0.000	0.009
27	**Number1**	5	0	5	89	100	104	193	29.918	42.796	0.000	23.585	53.503	0.016	0.182	0.000	0.016
28	**Oscar**	2	3	5	85	97	82	167	33.456	31.321	12.051	76.659	122.165	0.012	0.061	0.057	0.068
29	**Patchback**	5	4	9	111	160	78	189	35.622	28.333	25.643	58.654	119.919	0.128	0.000	0.084	0.212
30	**PL**	4	1	5	94	89	95	189	32.294	209.500	3.523	24.665	60.482	0.030	0.064	0.011	0.041
31	**Quasi**	1	0	1	124	138	133	257	0.000	0.000	0.000	0.000	0.000	0.002	0.068	0.000	0.002
35	**SMN5**	1	0	1	142	188	107	249	0.000	0.000	0.000	0.000	0.000	0.029	0.000	0.000	0.029
41	**SN90**	5	0	5	100	117	105	205	13.743	8.074	6.085	22.723	42.550	0.015	0.300	0.000	0.015
42	**SN96**	3	3	6	98	93	87	185	25.072	231.000	4.111	24.176	53.359	0.029	0.012	0.052	0.081
44	**Thumper**	1	3	4	116	145	80	196	3.351	0.000	5.400	13.278	22.029	0.006	0.000	0.072	0.078
45	**Topless**	6	5	11	101	137	75	176	21.719	55.167	14.309	38.399	74.427	0.148	0.000	0.137	0.285
54	**Upbang**	7	0	7	89	101	99	188	58.600	70.149	32.083	90.710	181.393	0.023	0.344	0.000	0.023
57	**Web**	6	3	9	98	115	104	202	11.261	4.830	60.897	81.937	154.095	0.013	0.364	0.004	0.017
60	**Zig**	0	1	1	165	0	177	342	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001
61	**Zipfel**	1	2	3	115	145	87	202	2.493	0.000	6.003	17.481	25.977	0.006	0.000	0.046	0.052

Subgroup measures: Non-normalized values. L stands for local, G for global, S for subgraph, O for overall, and B for boundary.