Relative Centrality Measures

Jocelyn R. Bell
United States Military Academy, West Point NY 10996
(e-mail: jocelyn.bell@usma.edu)

Abstract

In this paper we present natural and compelling generalizations of four widely used centrality measures to local and global measures relative to subsets of nodes in a network. These new relative centrality measures can provide insight into network properties with respect to subsets. Definitions of these measures are given and various properties are investigated.

Keywords: Centrality, Closeness, Betweenness, Eigenvector, Centralization

1 Introduction

Centrality measures are traditionally used to determine which nodes in a network are most important. In some cases, it may be necessary to determine which nodes in some smaller neighborhood of a network are most influential. However, it is no longer clear what “most influential” should mean. Is it most influential with respect to the nodes within the neighborhood, most influential over the nodes outside the neighborhood, or some combination of the two? This is our motivation for defining relative centrality: a node’s centrality depends on which nodes are “measuring” it.

One solution, which has been used previously in network analysis, is to modify the network in some way, and then calculate centralities on the resulting new network. In particular, one may generate a new graph from the subset of nodes of interest, or collapse this subset to a single vertex. However, there are potentially significant drawbacks to each of these methods. By changing the network, information is lost. By calculating centralities on a graph generated by a subset of nodes, the relationships to nodes outside the subset are ignored, and thus global information is lost. By calculating centralities on a graph generated by collapsing a subset to a single node, the ties between nodes within the subset are removed, and therefore local information is lost. We present a new framework in which nodal centralities relative to a subset of interest can be calculated without changing the underlying network. Because the network is not modified in any way, there is no loss of information.

For example, consider a social network of students at a university. We may wish to determine which members of the freshman class are most influential over their fellow freshman. If we calculate centrality measures over the entire network, we are only able to find the most important freshmen with respect to all students. If we generate the subgraph consisting only of freshman and calculate centrality measures there, then we are ignoring influence that spreads through the freshman class via upperclassmen. A similar
problem arises if we want to determine the most influential freshman with respect to only upperclassmen.

This paper is based on the observation that for any given subset of nodes in a network, the centrality of a node is due to both local influence (within the subset) and global influence (outside the subset). Each new measure we define in this paper will be comprised of two pieces: a local component and a global component. We will define local and global versions of four classic centrality measures: degree, closeness, betweenness and eigenvector. These local and global centrality measures assign ranks to individual nodes, rather than one value to the entire subset. They are constructed to be consistent with standard centrality measures, in the sense that if the subset is the entire network, then the local measure agrees with the standard measure. At the other extreme, if the subset consists of a single node, then the global measure reduces to the standard measure. Normalized versions are also introduced in the same consistent manner; these normalized measures take the size of the subset under investigation into account, thereby allowing meaningful comparisons across differing subsets of various sizes and among different networks.

2 Preliminaries

The networks considered in this paper are assumed to be connected, unweighted and undirected. Let $V$ represent the set of vertices or nodes of a network. We say a network is of size $n$ if $|V| = n$. The distance between node $x$ and node $y$ is the length of a geodesic between the two, and is denoted by $d(x, y)$.

2.1 Subsets of Nodes

We say a subset $S \subseteq V$ of nodes is a subset of the network; note this is not the same as the subgraph generated by $S$. A subset $S$ is locally connected if there is a path between every pair of nodes in $S$ which remains inside $S$. More formally:

Definition 2.1
Suppose $S$ is a subset of nodes in a network. $S$ is locally connected provided for every pair of nodes $a \in S$ and $b \in S$, there exist nodes $c_1, \ldots, c_n$ all from $S$ such that $d(a, c_1) = 1$, $d(c_i, c_{i+1}) = 1$ for all $i < n$, and $d(c_n, b) = 1$.

A subset $S$ is locally disconnected if it is not locally connected.

From this definition, it is immediate that certain kinds of collections of nodes cannot be locally connected. For example:

Proposition 2.1
Suppose $S$ is a subset of nodes in a network and there exists a node $a \in S$ such that $d(a, b) > 1$ for every other $b \in S$, so $S$ contains no neighbors of $a$. Then $S$ is locally disconnected.

Certain network properties of subsets may fail to hold if the subsets are locally disconnected. However, as such subsets do arise naturally, we will not assume our subsets are locally connected unless otherwise specified.
2.2 Local and Global Centrality

A centrality measure is a real valued function defined on $V$. We are interested in extending the idea of centrality relative to subsets of nodes. Strictly speaking, this is not a centrality measure since we are dealing with subsets rather than subgraphs. It is also not a measure of the centrality of a subset as a whole; for group centrality measures see (Everett & Borgatti, 1999).

**Definition 2.2**
In a network with vertices $V$, a *local centrality measure* is a family of real valued functions $\{L_S : S \subseteq V\}$, one for each subset $S \subseteq V$, such that the domain of $L_S$ is $S$. If $a \in S$, the *local $S$ centrality* of $a$ is $L_S(a)$.

Notice that if node $a$ is not in the subset $S$, then $L_S(a)$ is not defined. For our purposes, it does not make sense to consider it “local centrality” if the node is not in $S$.

**Definition 2.3**
In a network with vertices $V$ with a local centrality measure $\{L_S : S \subseteq V\}$, if $a \in S$, the *global $S$ centrality* of $a$ is $L_{S \cup \{a\}}(a)$.

In words, the global centrality of a node is the local centrality according to $S \cup \{a\}$, where the node $a$ is added to the complement of $S$ so that local centrality is defined. Since the notation for global $S$ centrality of a node $a$ is cumbersome, we will denote it by $G_S(a)$. Throughout this paper, we will provide separate definitions of local and global centrality, even though global centrality is always defined from local centrality so technically one definition would suffice. Separate definitions are included for readability and notational convenience.

If we already have a centrality measure defined on a network, we would like to define a local centrality measure so that if the subset $S$ is the entire set of vertices, then for each node the local $S$ centrality is the same as the given centrality measure. This is formalized in the next definition.

**Definition 2.4**
In a network with vertices $V$ and given centrality measure $c$, a local centrality measure $\{L_S : S \subseteq V\}$ is *compatible* with $c$ provided $c = L_V$.

As an example, for a network and centrality measure $c$, if for each subset $S$, $L_S$ is defined to be the same centrality measure $c$ calculated on the subgraph generated by $S$, then $\{L_S : S \subseteq V\}$ is a local centrality measure compatible with $c$. Note, however, that $c$ must be able to be calculated on every such subgraph. This may be a concern if $S$ is not locally connected (hence the subgraph generated by it is disconnected) and the centrality measure $c$ requires a connected network (e.g., closeness). Thus, compatible local centrality is a generalization of this approach.

It follows from the definition of compatibility that for any node $a$, if the subset $S$ is just the singleton set $\{a\}$, then the global $S$ centrality of $a$ agrees with the given centrality measure. This is precisely what the following proposition states.

**Proposition 2.2**
If $\{L_S : S \subseteq V\}$ is a local centrality measure which is compatible with a given centrality measure $c$, then for every node $a \in V$, $G_{\{a\}}(a) = c(a)$. 
Proof

\[ G_{\{a\}}(a) = L_{V \setminus \{a\}}(a) = L_V(a) = c(a). \]

Notice there may be many possible ways to define a compatible local centrality measure for a given centrality measure; we present natural definitions of some of the most widely used measures in the next four sections.

3 Degree

Various definitions of local and global degree have appeared elsewhere: see (Wasserman & Faust, 1994) for local degree and (Arney & Peterson, 2010) for a version of global degree. We introduce new notation here to be consistent with the following sections. Throughout this paper, the prefix \( L \) indicates a local measure is being defined, the prefix \( G \) indicates global, and the prime denotes a normalized measure.

3.1 Local and Global Degree

We define local and global degree centrality as follows.

**Definition 3.1**

In a network of size \( n \) with a subset \( S \) of size \( k \), for each \( a \in S \) the *local S degree* of \( a \) is

\[ Ldeg_s(a) = |\{y \in S : d(a,y) = 1\}| \]

To normalize, divide by \( k - 1 \):

\[ Ldeg'_s(a) = \frac{Ldeg_s(a)}{k - 1} \]

In words, this definition states that the local degree (according to the subset \( S \)) of a node is the number of nodes, only from inside \( S \), which are adjacent to it. The normalized local degree is just the proportion of nodes inside \( S \) which are connected to \( a \). Both are valid local centrality measures, compatible with degree and normalized degree respectively. Local degree is equivalent to standard degree in the subgraph generated by \( S \).

**Definition 3.2**

In a network of size \( n \) with a subset \( S \) of size \( k \), for each \( a \in S \) the *global S degree* of \( a \) is

\[ Gdeg_s(a) = |\{y \notin S : d(a,y) = 1\}| \]

To normalize, divide by \( n - k \)

\[ Gdeg'_s(a) = \frac{Gdeg_s(a)}{n - k} \]

That is, the global degree (according to the subset \( S \)) of a node is the number of nodes from outside of \( S \) which are adjacent to it. It is equivalent to standard degree in the subgraph generated by \( S \cup \{a\} \). A comparison of local and global degree is pictured in Figure 1.

Alternatively, we could have defined the local \( S \) degree of a node \( a \) to be the degree of \( a \) restricted to the subgraph \( S \), and the global degree as the degree of \( a \) in the entire network less the local \( S \) degree; this is precisely the definition found in (Arney & Peterson,
Relative Centrality Measures

(a) $L_{deg_S}(a) = 2$. Normalized: $\frac{2}{4}$

(b) $G_{deg_S}(a) = 3$. Normalized: $\frac{3}{8}$

Figure 1: Examples of Local and Global $S$ Degree

2010). However, this is not equivalent when using normalized degree measures. In Figure 1, for example, the normalized local $S$ degree of node $a$ is larger than its normalized degree. If normalized global $S$ degree were defined by subtraction in this way, the resulting value would be negative. Perhaps there are applications where negative degree values are appropriate; however for the purposes of this paper all of the subset centrality measures defined will be non-negative.

3.2 Properties

For every subset $S$, the sum of a node’s local $S$ and global $S$ degree is its standard degree. The next proposition is a consequence of this fact, and provides the circumstances under which local or global versions of degree reduce to standard degree centrality.

**Proposition 3.1**

Suppose $S$ is a subset of a network and $a \in S$. Then $L_{deg_S}(a) = deg(a)$ and $G_{deg_S}(a) = 0$ if and only if $S$ contains all neighbors of $a$. On the other hand, $G_{deg_S}(a) = deg(a)$ and $L_{deg_S}(a) = 0$ if and only if $S$ contains no neighbors of $a$.

How does a fixed node’s local and global degree change as the subset $S$ varies? If $S \subseteq S'$ and $a \in S$, we say node $a$’s local (global) degree increases if the local (global) $S$ degree of $a$ is less than or equal to the local (global) $S'$ degree of $a$. We say node $a$’s local (global) degree decreases if the local (global) $S$ degree of $a$ is at least as large as the local (global) $S'$ degree. Without normalizing, increasing the size of the subset increases the local degree and decreases the global. This is precisely what the next proposition states.

**Proposition 3.2**

Suppose $S$ and $S'$ are subsets of a network such that $S \subseteq S'$. Then for each $a \in S$

$$L_{deg_S}(a) \leq L_{deg_{S'}}(a)$$

$$G_{deg_S}(a) \geq G_{deg_{S'}}(a)$$

Strict inequality holds if and only if there exists $b \in S' \setminus S$ such that $d(a,b) = 1$.

**Proof**
Suppose \( S \subseteq S' \), and let \( B = S' \setminus S \). Then \( L_{deg}'(S \cup B) \geq L_{deg}'(S) \) (that is, the normalized local degree of \( a \) increases) if and only if
\[
L_{deg}'_{B \cup \{a\}}(a) \geq L_{deg}'_{S}(a)
\]
and \( G_{deg}'(S \cup B) \geq G_{deg}'(S) \) (that is, the normalized global degree of \( a \) decreases) if and only if
\[
L_{deg}'_{B \cup \{a\}}(a) \leq G_{deg}'_{S}(a)
\]
Proof
Straightforward manipulation of inequalities. Note that \( L_{deg}(a) + L_{deg}_{B \cup \{a\}}(a) = L_{deg}'(a) \). Suppose \( |B| = q \), and \( p = |z \in B : d(a,z) = 1| \). Then \( L_{deg}'_{B \cup \{a\}}(a) = \frac{p}{q} \). If we write \( w = L_{deg}(a) \), then \( L_{deg}'(a) = w + p \). \( S' = B \cup S \) and so \( S' \) contains \( k + q \) many nodes. Hence \( \frac{w}{k} \leq \frac{w + p}{k + q} \) if and only if \( \frac{w}{k} \leq \frac{p}{q} \).

For global degree, we have \( G_{deg}'(a) = G_{deg}(a) - L_{deg}_{B \cup \{a\}}(a) \) and so writing \( x = G_{deg}(a) \) gives \( G_{deg}'(a) = x - p \). The complement of \( S' \) contains \( n - k - q \) many nodes, and thus \( \frac{x}{n-k} \leq \frac{x - p}{n-k-q} \) if and only if \( \frac{x}{n-k} \leq \frac{p}{n-k-q} \).

In particular, if we consider only adding one additional point to our subset \( S \), the above proposition becomes

**Corollary 3.1**
Suppose \( S \) is a subset of a network with \( a \in S \) and \( b \notin S \). Then if \( a \) and \( b \) are adjacent (i.e. \( d(a,b) = 1 \)) then the normalized local degree of \( a \) according to \( S' \cup \{b\} \) will increase and the global degree will decrease. If \( d(a,b) \neq 1 \), then the normalized local degree of \( a \) will decrease and the global degree will increase.

**Proof**
Here, \( B = \{b\} \) and so if \( d(a,b) = 1 \), \( L_{deg}'_{\{b\}}(a) = 1 \) and so by the preceding, \( L_{deg}'_{\{b\}}(a) \leq L_{deg}'(a) \) and \( G_{deg}'_{\{b\}} \leq L_{deg}'(a) \). If \( d(a,b) \neq 1 \), then \( L_{deg}'_{\{b\}}(a) = 0 \).

This corollary provides an easy way to generate examples in which a node’s local and global normalized degree shift in opposite directions. It is possible, however, for a node’s normalized local and global degree to either both increase or both decrease. Examples of these phenomena are pictured in Figure 2.

### 3.3 Centralization

Network centralization is a way to measure how much variation in centrality is present in the network. In this paper, we will be using the classic Freeman definition; see (Freeman, 1979). In order to calculate it, the maximum possible variation must be known.
Relative Centrality Measures

Figure 2: Example of Local and Global Degree Increasing or Decreasing

**Proposition 3.4**
In a network of size $n$ with a subset $S$ of size $k$, the theoretical maximum for local degree centralization occurs when the subset $S$ (as a network) is the star graph, and hence this maximum is $(k - 1)(k - 2)$.

**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 $Ldeg_S(b) \leq k - m - 1$ and the remaining $k - m - 1$ many nodes have local degree at least 1. So $\sum_{i=1}^{k} Ldeg_S(b) \leq k \cdot (k - m - 1)$ and $\sum_{v \in S} Ldeg(v) \geq 2(k - m - 1)$ and therefore the maximum difference is at most $(k - m - 1)(k - 2)$. Clearly this is less than $(k - 1)(k - 2)$ if $m > 0$. \[\square\]

**Proposition 3.5**
In a network of size $n$ with a subset $S$ of size $k$, the theoretical maximum for global degree centralization is $(n - k)(k - 1)$. This occurs when one node inside $S$ is connected to every node outside $S$, and every other node inside $S$ has no connection to nodes outside $S$.

**Proof**
If $b$ is a node with maximum global degree, then $Gdeg_S(b) \leq n - k$. The minimum global degree is 0 for the $k - 1$ remaining nodes, so the maximum difference is $(n - k)(k - 1)$. This maximum is attained by a network which is as described. \[\square\]

Figure 3 is a schematic of this maximum situation for global degree. It does not matter how the interior nodes are connected to each other, so long as none of them except the highlighted node are connected to any nodes outside of $S$.

With the theoretical maximums, we can now define local and global degree centralization.

**Definition 3.3**
In network of size $n$ with subset $S = \{n_1, \ldots, n_k\}$ of size $k$, ordered so that $n_1$ is a node with highest local degree, the local degree centralization of $S$ is

$$\frac{\sum_{i=1}^{k} (Ldeg_S(n_1) - Ldeg_S(n_i))}{(k - 1)(k - 2)}$$
Figure 3: Example Schematic of Maximum Global Degree Centralization

Normalized, this denominator becomes \((k - 1)\).

If \(S\) is the entire network, then local degree centralization is the same as degree centralization for the whole network in the usual sense. In general terms, high local degree centralization means there are only a few individuals with many local connections. Low local degree centralization indicates members of the group have roughly the same number of ties to one another.

**Definition 3.4**

In network of size \(n\) with subset \(S = \{n_1, \ldots, n_k\}\) of size \(k\), ordered so that \(n_1\) is a node with highest global degree, the *global degree centralization* of \(S\) is

\[
\frac{\sum_{i=1}^{k} (Gdeg_S(n_1) - Gdeg_S(n_i))}{(k - 1)(n - k)}
\]

Normalized, this denominator becomes \((k - 1)\).

It should be noted that it is possible to have high local degree centralization and low global degree centralization or vice-versa. For example, high global and low local centralization implies the existence of a small number of liaisons between the group and the rest of the network, however within the group there are no local leaders. Such an example given in Figure 4(a). Figure 4(b) depicts the opposite situation.

### 3.4 Combining Local and Global Degree

Suppose we are interested in combining local and global degree into one score. This can be accomplished choosing some parameter \(0 \leq t \leq 1\) and defining for each node \(a\) in a subset \(S\)

\[
degs(a) = t \cdot Ldeg_S(a) + (1 - t) \cdot Gdeg_S(a)
\]

The parameter \(t\) determines how much weight is given to the local component. It is easy to see that \(0 \leq degs(a) \leq 1\). In a network of size \(n\) with a subset \(S\) of size \(k\), if the parameter \(t\) is chosen to be

\[
t = \frac{k - 1}{n - 1}
\]

then the resulting total \(degs(a)\) is standard normalized degree.
Relative Centrality Measures

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

Figure 4: Differences in Local and Global Degree Centralization

4 Closeness

Closeness centrality measures the distance from a particular node to all other nodes in the network. Here we define local and global versions. Recall that $d(x,y)$ denotes the distance between node $x$ and node $y$.

4.1 Local and Global Closeness

Definition 4.1
In a network of size $n$ with a subset $S$ of size $k$, where $k > 1$, for $a \in S$ the local $S$ closeness of node $a$ is

$$LCl_S(a) = \frac{1}{\sum_{y \in S} d(a,y)}$$

To normalize, multiply by $k - 1$:

$$LCl'_S(a) = \frac{k - 1}{\sum_{y \in S} d(a,y)}$$

That is, local closeness adds the distances from $a$ only to nodes in $S$ and then inverts this sum. It does not matter if a geodesic uses nodes from outside $S$. Local closeness is always defined (for a connected network) regardless of whether or not $S$ is locally connected. Even if $S$ is locally connected, the shortest path between two points of $S$ may in fact go outside of $S$, and is not equivalent to closeness in the subgraph generated by $S$.

Definition 4.2
In a network of size $n$ with a subset $S$ of size $k$, where $k < n$, for $a \in S$ the global $S$ closeness of $a$ is

$$GCl_S(a) = \frac{1}{\sum_{y \in S} d(a,y)}$$

To normalize, multiply by $n - k$:

$$GCl'_S(a) = \frac{n - k}{\sum_{y \in S} d(a,y)}$$
Jocelyn R. Bell

Global closeness adds the distances from \( a \) to each node outside \( S \) and inverts this quantity. A representation is depicted in Figure 5.

\[
\begin{align*}
\text{(a) } LCl_S(a) &= \frac{1}{6} \text{. Normalized: } \frac{1}{5} \\
\text{(b) } GCl_S(a) &= \frac{1}{116} \text{. Normalized: } \frac{8}{116}
\end{align*}
\]

Figure 5: Examples of Local and Global \( S \) Closeness

### 4.2 Properties

Unlike degree, it is not true that the sum of the local and global \( S \) closeness is standard closeness for every subset \( S \).

**Proposition 4.1**

Suppose \( S \) and \( S' \) are subsets of a network such that \( S \subseteq S' \). Then for each \( a \in S \)

\[
LCl_S(a) \geq LCl_{S'}(a)
\]

\[
GCl_S(a) \leq GCl_{S'}(a)
\]

Strict inequality holds if and only if \( S \not\subseteq S' \).

**Proof**

Suppose \( b \in S' \setminus S \). Then \( \sum_{y \in S} d(a,y) + d(a,b) = \sum_{y \in S \setminus \{b\}} d(a,y) \leq \sum_{y \in S'} d(a,y) \). Also, \( \sum_{y \not\in S \cup \{b\}} d(a,y) = \sum_{y \in S} d(a,y) - d(a,b) \). Strict inequality since \( d(a,b) > 0 \)

As before, the preceding does not hold if normalized versions of the measures are used. Normalized local closeness of a node \( a \) will increase provided the new nodes to be added are “close enough” to \( a \), and the normalized global closeness will increase if the new nodes are “far enough” away from \( a \).

**Proposition 4.2**

Suppose \( a \in S \subseteq S' \), and let \( B = S' \setminus S \). Then the normalized local closeness of \( a \) will increase (i.e. \( LCl'_{B \cup \{a\}}(a) \geq LCl'_S(a) \)) if and only if

\[
LCl'_{B \cup \{a\}}(a) \geq LCl'_S(a)
\]

The normalized global closeness of \( a \) will increase (i.e \( GCl'_{S'}(a) \geq GCl'_S(a) \)) if and only if

\[
LCl'_{B \cup \{a\}}(a) \leq GCl'_S(a)
\]
Relative Centrality Measures

Proof
Straightforward manipulation of inequalities. Suppose we are adding \( q \) many new nodes, and let \( b = \sum_{y \in B} d(a, y) \). Then \( LCl'_{B \cup \{a\}}(a) = \frac{b}{q} \) since \( B \cup \{a\} \) has \( q+1 \) many nodes. If we write \( w = \sum_{z \in S} d(a, z) \), we have \( \sum_{z \in S'} = w + b \). Then if \( \frac{k-1}{w} \leq \frac{k+q-1}{w+b} \) if and only if \( \frac{k-1}{w} \leq \frac{q}{b} \).

The complement of \( S' \) contains \( n - k - q \) many nodes, and writing \( x = \sum_{z \in S} d(a, z) \), we see \( \sum_{z \notin S'} = x - b \). Then \( \frac{n-k}{x} \leq \frac{n-k-2}{x-b} \) if and only if \( \frac{q}{b} \leq \frac{n-k}{x} \).

However, unlike the case for normalized degree, adding a single point to a subset does not simplify the problem much. It is true that:

**Corollary 4.1**
For a node \( a \) in a subset \( S \) of a network, if \( y \notin S \) and \( d(a, y) = 1 \), then for \( S' = S \cup \{y\} \), the following inequalities are true:

\[
LCl'_{S}(a) \leq LCl'_{S'}(a) \quad \text{and} \quad GCl'_{S}(a) \geq GCl'_{S'}(a)
\]

Proof

\( LCl'_{S \cup \{y\}}(a) = \frac{1}{\pi_{S \cup \{y\}}} = 1 \) and so \( LCl'_{S}(a) \leq LCl'_{S \cup \{y\}}(a) \) by above proposition. Similarly, \( GCl'_{S}(a) \leq LCl'_{S \cup \{y\}}(a) \).

This corollary says that adding a node adjacent to \( a \) to the subset \( S \) causes \( a \)'s local closeness to increase and its global closeness to decrease. An example of all four possibilities when adding a single new node to a subset is depicted in Figure 6. The dashed circle represents \( S \). According to \( S \cup \{y\} \), node \( a \) sees both normalized local and global increase, while node \( b \) sees both decrease. According to \( S \cup \{x\} \), node \( a \)'s normalized local closeness decreases while its global increases, and node \( b \)'s normalized local closeness increases while its global decreases.

Figure 6: Examples of Differing Closeness Measures
4.3 Centralization

Closeness centralization is traditionally defined using inverse sums of distances. For a rigorous proof of the standard closeness centralization formula, see (Everett et. al., 2004). We depart from this convention and use a slightly modified version we call inverse centralization, which allows for simpler calculations.

Definition 4.3

In a network of size \( n \) with nodes \( \{a_1, \ldots, a_n\} \), if \( c \) is any positive centrality measure, and \( a^* \) is any node such that \( c(a^*) \) is maximum, define the inverse centralization index as

\[
\sum_{i \in n} \left( \frac{1}{c(a_i)} - \frac{1}{c(a^*)} \right)
\]

Proposition 4.3

The maximum local closeness inverse centralization for a locally connected subset occurs when the subset \( S \) is the star graph. In this case the maximum is \( (k - 1)(k - 2) \).

Proof

Suppose \( a \in S \) is fixed such that \( LCl_S(a) \) is maximum, (so \( \sum_{y \in S} d(a,y) \) is minimum), and \( b \in S \) and \( y \in S \) where \( a, b \) and \( y \) are all distinct. Then \( d(b,y) - d(a,y) \leq d(a,b) \), where equality holds if and only if \( a \) lies on a geodesic from \( b \) to \( y \). There are \( k - 2 \) many such \( y \)'s, hence

\[
\sum_{y \in S \setminus \{a,b\}} (d(b,y) - d(a,y)) \leq (k - 2) \cdot d(a,b)
\]

where equality holds if and only if \( a \) lies on a geodesics from \( b \) to \( y \) for each \( y \in S \setminus \{a,b\} \).

Since \( \sum_{y \in S \setminus \{a,b\}} (d(b,y) - d(a,y)) = \frac{1}{LCl_S(b)} - \frac{1}{LCl_S(a)} \), summing over all \( b \in S \) gives

\[
\sum_{b \in S} \left( \frac{1}{LCl_S(b)} - \frac{1}{LCl_S(a)} \right) \leq (k - 2) \sum_{b \in S} d(a,b) = (k - 2) \cdot \frac{1}{LCl_S(a)}
\]

Equality holds if and only if \( d(x,a) + d(a,y) = d(x,y) \) for each \( x \in S \setminus \{a\} \) and \( y \in S \setminus \{a,x\} \). If there exists \( z \in S \) such that \( d(a,z) > 1 \), then since \( S \) is locally connected there exists \( w \in S \), \( w \neq z \), such that \( d(z,w) = 1 \). Therefore \( a \) does not lie on a geodesic from \( z \) to \( w \), and hence equality does not hold. Therefore, equality holds if and only if \( d(a,z) = 1 \) for all \( z \in S \setminus \{a\} \) and \( a \) lies on every geodesic between pairs of points in \( S \), that is, \( S \) is the star graph. In this case, \( \frac{1}{LCl_S(a)} = k - 1 \).

The assumption of local connectedness is necessary; see the “weighted star” example schematic in Figure 7. The dotted links indicate more nodes connected in the same manner can be added. Here, if the subset has \( k \) many nodes, and the network has a total of \( n \), then the local closeness inverse centralization is \( (k - 2)(n - 1) \). The highlighted node has local closeness of \( \frac{1}{n - 1} \), and so this is an example where equality is attained, and yet the local closeness inverse centralization is greater than \( (k - 2)(k - 1) \).

Proposition 4.4

The maximum global closeness inverse centralization for a locally connected subset occurs when one node in \( S \) has maximum global closeness \( \frac{1}{n - 1} \), and all other nodes of \( S \) are as far away from the nodes outside \( S \) as possible. In this case, the maximum is

\[
\frac{1}{4} (n - k)(k \cdot (k - 1))
\]
Relative Centrality Measures

Figure 7: Locally Disconnected: Local Closeness Inverse Centralization

Proof
Suppose \( a \in S \) is such that \( \sum_{y \in S} d(a, y) \) is as small as possible, and \( b \in S \) and \( y \not\in S \). Then \( d(b, y) - d(a, y) \leq d(a, b) \) and there are \( n - k \) many such \( y \)'s, hence

\[
\sum_{y \neq S} (d(b, y) - d(a, y)) \leq (n - k) \cdot d(a, b)
\]

from which we see

\[
\frac{1}{GCl_S(b)} - \frac{1}{GCl_S(a)} \leq (n - k) \cdot d(a, b)
\]

Now sum over all \( b \in S \) to get

\[
\sum_{b \in S} \left( \frac{1}{GCl_S(b)} - \frac{1}{GCl_S(a)} \right) \leq (n - k) \sum_{b \in S} d(a, b) = (n - k) \cdot \frac{1}{LCl_S(a)}
\]

The maximum for the right hand side is attained when \( \frac{1}{LCl_S(a)} \) is as large as possible; in a locally connected subset this is \( \sum_{i=1}^{k-1} i = \frac{k(k+1)}{2} \). Equality is attained by a network as described in the proposition.

A schematic of maximum global closeness inverse centralization is given in Figure 8. It does not matter how the nodes outside the subset are connected, as long as the only element of the subset they are connected to is the highlighted node.

Again, the assumption of local connectedness is necessary; see the schematic in Figure 9. Here, the subset \( S \) contains just two elements, and suppose the box on the right contains \( l \) many nodes and the total number of nodes is \( n \). Then the global closeness inverse centralization is \( (n - l - 2) (l + 1) \), which is larger than the theoretical maximum for connected subsets, in this case \( n - 2 \), provided \( l > 0 \) and \( n > l + 3 \). The local closeness of the highlighted node is \( l + 1 \), and so we do have that the inverse centralization is less than \( (n - 2) (l + 1) \), as shown in the proof of the proposition.

The inverse centralization formulas given below use the locally connected subset maximums. These formulas can be applied even if local connectedness is not assumed; in that case, on rare occasion, the inverse centralization can be greater than 1.

Definition 4.4
In network of size \( n \) with subset \( \mathcal{S} = \{n_1, \ldots, n_k\} \) of size \( k \), ordered so that \( n_1 \) is a node with highest local closeness, the local closeness inverse centralization of \( \mathcal{S} \) is

\[
\sum_{i=1}^{k-1} \left( \frac{1}{\bar{LCL}(n_i)} - \frac{1}{\bar{LCL}(n_1)} \right) \frac{1}{(k-1)(k-2)}
\]

In general terms, high local closeness centralization means only a few nodes are very close to the rest of the group. Low local closeness centralization indicates members of the group are roughly equidistant.

Definition 4.5
Relative Centrality Measures

In network of size $n$ with subset $S = \{n_1, \ldots, n_k\}$ of size $k$, ordered so that $n_1$ is a node with highest global closeness, the global closeness inverse centralization of $S$ is

$$2 \sum_{i=1}^{k} \left( \frac{1}{GC_S(n_i)} - \frac{1}{GC_S(n_1)} \right) \frac{k \cdot (k-1)(n-k)}{k}$$

4.4 Combining Local and Global Closeness

The two measures can be combined parametrically as was presented for degree centrality; that is choose $0 \leq t \leq 1$ and for each node $a$ in a subset $S$ define

$$CL_S(a) = t \cdot LCL'_S(a) + (1-t) \cdot GCL'_S(a)$$

However in this case, there is no way to choose the parameter $t$ for every subset $S$ so that the resulting $CL_S(a)$ is always equal to the standard normalized closeness centrality of node $a$.

5 Betweenness

Betweenness centrality is fundamentally different in its calculation than degree or closeness. It measures to what extent a particular node impacts connections between other pairs of nodes. Early work on this type of measure was done by Freeman in (Freeman, 1977). The local and global behavior of the betweenness centrality measure is also different from the two previously mentioned centrality measures. The challenge is categorizing pairs of nodes for which one is inside and one is outside of a given subset. To this end, we introduce a third subset centrality measure, called boundary centrality.

5.1 Local, Global and Boundary Betweenness

The main idea is to separate pairs of nodes in the network into three non-overlapping groups.

Notation 5.1

In a network of size $n$ with a subset $S$ of size $k$, denote

$$L_S = \{\{x, y\} : x \in S \text{ and } y \in S \text{ and } x \neq y\}$$

$$G_S = \{\{x, y\} : x \notin S \text{ and } y \notin S \text{ and } x \neq y\} = L_S^c$$

$$B_S = \{\{x, y\} : x \in S \text{ and } y \notin S\}$$

Definition 5.1

In a network of size $n$ with a subset $S$ of size $k$, the local $S$ betweenness of $a \in S$ is given by

$$Lb_S(a) = \sum_{\{x, y\} \in L_S(a)} \frac{\text{# of geodesics including } a}{\text{total # of geodesics between } x \text{ and } y}$$
Since there are $\frac{1}{2} (k - 1) (k - 2)$ many pairs in $L_S\{a\}$, to normalize divide $Lb_S(a)$ by this quantity:

$$Lb'_S(a) = \frac{2 \cdot Lb_S(a)}{(k - 1)(k - 2)}$$

The difference between local betweenness and standard betweenness is that the sum is restricted to only those pairs of nodes both of whose elements are members of the subset. Notice we use $S \setminus \{a\}$ instead of $S$, to prevent the node $a$ from appearing in the possible pairs.

**Definition 5.2**

In a network of size $n$ with a subset $S$ of size $k$, the **global $S$ betweenness** of $a \in S$ is given by

$$Gb_S(a) = \sum_{\{x,y\} \in G_S} \frac{\# \text{ of geodesics including } a}{\text{total } \# \text{ of geodesics between } x \text{ and } y}$$

Since there are $\frac{1}{2} (n - k) (n - k - 1)$ many pairs in $G_S$, to normalize divide $Gb_S(a)$ by this quantity:

$$Gb'_S(a) = \frac{2 \cdot Gb_S(a)}{(n - k)(n - k - 1)}$$

Again, the difference is in the restriction to pairs of nodes both outside the subset. Similarly, we define the following:

**Definition 5.3**

In a network of size $n$ with a subset $S$ of size $k$, the **boundary $S$ betweenness** of $a \in S$ is given by

$$Bb_S(a) = \sum_{\{x,y\} \in B_S\{a\}} \frac{\# \text{ of geodesics including } a}{\text{total } \# \text{ of geodesics between } x \text{ and } y}$$

Since there are $(k - 1)(n - k)$ many pairs in $B_S\{a\}$, to normalize divide $Bb_S(a)$ by this quantity:

$$Bb'_S(a) = \frac{Bb_S(a)}{(k - 1)(n - k)}$$

In the extreme cases, where $S$ is the entire network or a single point, the boundary betweenness is zero. It is possible to have zero local and global betweenness; see Figure 10. The highlighted node $a$ has only boundary betweenness.

### 5.2 Properties

For every subset $S$, the sum of local, global, and boundary $S$ betweenness values is standard betweenness centrality. If we have nested subsets $S \subseteq S'$, then local increases while global decreases. However, the boundary case is different. The number of boundary pairs over which the sum is taken can increase or decrease, depending on the size of the subset $S$ relative to the size of the network.

**Proposition 5.1**

In a network of size $n$ with a subset $S$ of size $k$, if $S \subseteq S'$ then $S'$ defines more boundary pairs, that is $|B_S| \leq |B_{S'}|$, whenever $|S'| \leq n - 2k + 1$. 

*Jocelyn R. Bell*
Relative Centrality Measures

Figure 10: Example of non-zero Boundary Betweenness with zero Local and Global Betweenness

In particular, we see that if the size of the subset $S$ is more than half the size of the network (so the subset is large), then the number of pairs in the boundary calculation for any other subset containing $S$ will never increase. However, even if the sum is taken over fewer pairs of nodes, this does not necessarily mean the boundary betweenness must decrease; see Figure 11. If node $b$ is added to the subset in Figure 11(a), node $a$'s boundary $S$ betweenness decreases even as more boundary pairs are added. In Figure 11(b), node $a$'s boundary $S$ betweenness increases according to $S \cup \{b\}$, even though the sum is taken over fewer pairs.

(a) Node $a$: Boundary Betweenness Decreases  (b) Node $a$: Boundary Betweenness Increases

Figure 11: Examples of Boundary Betweenness Increasing or Decreasing

5.3 Centralization

The theoretical maximum centralization calculations are simple to prove. In fact, all three are maximized when $S$ is a star graph; see Figure 12.

Proposition 5.2

In a network of size $n$ with a subset of size $k$, maximum local $S$ betweenness centralization occurs when the subset is a star graph. This maximum is $(k - 1)^2 (k - 2)$.

Proof
Jocelyn R. Bell

The maximum for a node \( a \in S \) occurs when \( a \) sits on all paths between pairs of the remaining nodes in \( S \). In a star graph, this maximum is attained and all other nodes have a local betweenness score of zero. □

The argument for global and boundary maximums is similar.

**Proposition 5.3**

In a network of size \( n \) with a subset of size \( k \), maximum global betweenness centralization is \((k-1)(n-k)(n-k-1)\) and maximum boundary betweenness centralization is \((k-1)^2(n-k)\).

![Figure 12: Maximum Betweenness Centralization](image)

It may be simpler to work with normalized centralization formulas, since in that case all three maximums are \((k-1)\). This fact allows us to define all three centralization formulas at once.

**Definition 5.4**

In network of size \( n \) with subset \( S = \{n_1, \ldots, n_k\} \) of size \( k \), let \( b'_S \) denote normalized local, global or boundary betweenness. If \( n_1 \) is a node with highest \( b'_S \) betweenness, the \( b_S \) betweenness centralization of \( S \) is

\[
\sum_{i=1}^{k} \frac{(b'_S(n_1) - b'_S(n_i))}{(k-1)}
\]

If \( S \) is the entire network, then local betweenness centralization is the same as standard betweenness centralization. The following is an interesting characterization of zero global betweenness centralization.

**Proposition 5.4**

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

**Proof**

Since there is only one connection, no node inside the subset \( S \) lies on any geodesic between nodes outside, and therefore the global \( S \) betweenness is zero for each node in \( S \). □
Relative Centrality Measures

5.4 Combining Local, Global and Boundary Betweenness

To combine the three quantities into one, choose non-negative parameters \( t_1 \) and \( t_2 \) satisfying \( 0 \leq t_1 + t_2 \leq 1 \) and define for each node \( a \) in a subset \( S \)

\[
B_S(a) = t_1 \cdot Lb^S_S(a) + t_2 \cdot Gb^S_S(a) + (1 - (t_1 + t_2)) \cdot Bb^S_S(a)
\]

In a network of size \( n \) with subset \( S \) of size \( k \), if the parameters are chosen as

\[
t_1 = \frac{(k - 1)(k - 2)}{(n - 1)(n - 2)} \quad \text{and} \quad t_2 = \frac{(n - k)(n - k - 1)}{(n - 1)(n - 2)}
\]

the resulting \( B_S(a) \) is standard normalized betweenness centrality.

5.5 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; see (Newman, 2010). It can be used to locate structural holes. It is defined as:

**Definition 5.5**

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 between the local clustering coefficient and local betweenness of a node. For each node \( a \) in a network, let \( S_a = \{ b \in S : d(a, b) \leq 1 \} \). In words, \( S_a \) is the network subset containing the node \( a \) and all of its neighbors. Then

\[
Lb^S_S(a) = 1 - C_a
\]

and thus local betweenness is a generalization of local clustering.

6 Eigenvector Centrality

Eigenvector centrality measures a node’s importance by making its rank proportional to the sum of the ranks its neighbors. This centrality measure was originated by Bonacich and later generalized in (Bonacich, 1987). A thorough explanation of this centrality measure can be found in (Newman, 2010).

6.1 Local and Global Eigenvector Centrality

In this section, suppose our network has adjacency matrix \( A \) with leading eigenvalue \( \lambda \) and corresponding unit eigenvector \( \vec{v} \), and denote node \( a \)’s standard eigenvector centrality by \( e(a) \).

**Definition 6.1**

For a node \( a \) in a subset \( S \) of a network, local \( S \) eigenvector centrality is

\[
L_e^S(a) = \frac{1}{\lambda} \sum_{b \in S, d(a, b) = 1} e(b)
\]
Global $S$ eigenvector centrality is

$$\text{Ge}_S(a) = \frac{1}{\lambda} \sum_{b \in S, d(a,b)=1} e(b)$$

Notice the sum of local and global returns the standard centrality:

$$\text{Le}_S(a) + \text{Ge}_S(a) = e(a)$$

### 6.2 Properties

In addition to the summation property mentioned previously, we have the following:

**Proposition 6.1**

If $S$ is a subset in a network and $a \in S$, then $\text{Le}_S(a) = e(a)$ and $\text{Ge}_S(a) = 0$ if and only if $S$ contains all neighbors of $a$. Furthermore, $\text{Ge}_S(a) = e(a)$ and $\text{Le}_S(a) = 0$ if and only if $S$ contains no neighbors of $a$.

If the subset is the whole network or a singleton, the local and global measures are as required.

**Proposition 6.2**

Suppose $S$ and $S'$ are subsets of a network such that $S \subseteq S'$. Then for each $a \in S$

$$\text{Le}_S(a) \leq \text{Le}_{S'}(a)$$

$$\text{Ge}_S(a) \geq \text{Ge}_{S'}(a)$$

Strict inequality holds if and only if there exists $b \in S' \setminus S$ with $d(a, b) = 1$.

### 6.3 An Example Network

In this section we provide example calculations. For the network in Figure 13, the standard eigenvector centralities have been calculated and are indicated next to the nodes. Additionally, one can verify that the leading eigenvalue for this network is $\lambda = 2.8737$. The local $S$ eigenvector centrality for node $a$ is

$$\text{Le}_S(a) = \frac{1}{\lambda} (e(b) + e(c)) = \frac{1}{2.8737} (.3799 + .2588) = .2223$$

and the global $S$ eigenvector centrality for node $a$ is

$$\text{Ge}_S(a) = \frac{1}{\lambda} e(e) = \frac{.4070}{2.8737} = .1416$$

To verify, $\text{Le}_S(a) + \text{Ge}_S(a) = .2223 + .1416 = .3639$ which rounds to $e(a)$.

Nodes $b$ and $c$ have no connections outside of $S$, and therefore their local $S$ eigenvector centralities must match their standard eigenvector centralities:

$$\text{Le}_S(b) = \frac{1}{2.8737} (.2588 + .3639 + .4690) = .3799 = e(b)$$

$$\text{Le}_S(c) = \frac{1}{2.8737} (.3799 + .3639) = .2588 = e(c)$$
6.4 Normalization

The local and global versions of eigenvector centrality are already between 0 and 1 so normalization may not be necessary. It may be desirable, however, to rescale the values to take into account the size of the subset.

**Definition 6.2**
Suppose \( S = \{a_1, \ldots, a_k\} \) is a subset of nodes of size \( k \) in a network \( n \). For \( a \in S \), define normalized local \( S \) eigenvector centrality

\[
Le'_S(a) = \frac{Le_S(a)}{\sqrt{e(a_1)^2 + \ldots + e(a_k)^2}}
\]

Note that the denominator is never zero if \( S \) is non-empty, since our networks are assumed connected. This denominator is always less than or equal to 1, so the normalized local eigenvector centrality is always greater than or equal to local eigenvector centrality. Equality holds only if \( S \) is the entire network.

**Definition 6.3**
Suppose \( S = \{a_1, \ldots, a_k\} \) is a subset of nodes of size \( k \) in a network \( n \), and suppose \( S' = \{a_{k+1}, \ldots, a_n\} \). For \( a \in S \), define normalized global \( S \) eigenvector centrality

\[
Ge'_S(a) = \frac{Ge_S(a)}{\sqrt{e(a)^2 + e(a_{k+1})^2 + \ldots + e(a_n)^2}}
\]

Notice if \( S = \{a\} \), the denominator is equal to 1.
6.5 Combining Local and Global Eigenvector Centrality

The two measures can be combined parametrically as before. Choose a parameter $0 \leq t \leq 1$ and for each node $a$ in a subset $S$ define

$$e_S(a) = t \cdot L e_S'(a) + (1-t) \cdot G e_S'(a)$$

However, there is no way to choose the parameter $t$ for every subset $S$ so that $e_S(a)$ is always equal to the standard normalized eigenvector centrality.

7 Conclusion

We have proposed methods for easily calculating the local and global influence of a node relative to some subset of nodes. The local and global measures defined are powerful generalizations of degree, closeness, betweeness and eigenvector centrality. These new centrality measures have the potential to uncover important information regarding network subset properties. For example, in a social network consisting of both men and women, we now have several methods for calculating an individual’s centrality with respect to each sex.

We have focused exclusively on unweighted and undirected networks, however many of the definitions provided can be manipulated to handle directed and/or weighted networks.

References


