Maintaining the Duality of Closeness and Betweenness Centrality[☆]

Ulrik Brandes

Department of Computer & Information Science, University of Konstanz

Stephen P. Borgatti

Gatton College of Business and Economics, University of Kentucky

Linton C. Freeman

Institute of Mathematical Behavioral Sciences, University of California at Irvine

Abstract

Betweenness centrality is generally regarded as a measure of others' dependence on a given node, and therefore as a measure of potential control. Closeness centrality is usually interpreted either as a measure of access efficiency or of independence from potential control by intermediaries. Betweenness and closeness are commonly assumed to be related for two reasons: First, because of their conceptual duality with respect to dependency, and second, because both are defined in terms of shortest paths.

We show that the first of these ideas – the duality – is not only true in a general conceptual sense but also in precise mathematical terms. This becomes apparent when the two indices are expressed in terms of a shared dyadic dependency relation. We also show that the second idea – the shortest paths – is false because it is not preserved when the indices are generalized using the standard definition of shortest paths in valued graphs. This suggests that closeness-as-independence is different from closeness-as-efficiency, and we propose an adjustment that maintains the duality with betweenness also on valued relations.

Keywords: Betweenness centrality, closeness centrality, duality, dependency, derived relations

[☆]Research partially supported by DFG under grant Br 2158/6-1.

Email addresses: `Ulrik.Brandes@uni-konstanz.de` (Ulrik Brandes), `sborgatti@uky.edu` (Stephen P. Borgatti), `lin@aris.ss.uci.edu` (Linton C. Freeman)

1. Introduction

A number of attempts have been made to bring order to the universe of centrality measures, including Sabidussi (1966), Koschützki, Lehmann, Peeters, Richter, Tenfelde-Podehl, and Zlotowski (2005), and Borgatti and Everett (2006). By far the most influential of these has been Freeman (1979). Since the publication of that paper, degree, closeness and betweenness centrality have been regarded as prototypical measures that capture most important aspects of centrality. The only other measure as well-known as these is eigenvector centrality (Bonacich, 1972), along with its variants (Bonacich, 1987).

In this paper, we focus on closeness and betweenness, which are based on an underlying concept of something flowing through a network along optimal paths. Consistent with the imagery used in Freeman's seminal paper, we assume the ties in our networks can be viewed as communication channels, although it should be clear that our results apply to any kind of network for which flows, geodesics, closeness and betweenness have meaningful interpretations.

Betweenness is generally employed with the understanding that it captures the potential for control of communication between actors. For closeness, Freeman (1979) actually outlines two different possible interpretations: either as independence from such control by others (*closeness as independence*) or as a measure of access or efficiency (*closeness as efficiency*). Here we focus on the interpretation of independence as it is referred to in many empirical studies such as Brass (1984), Rowley (1997), and Powell, Koput, and Smith-Doerr (1996).

Freeman (1980) shows that the interpretive duality of closeness and betweenness as measures of independence and control is quantitatively justified. It has been widely overlooked, though, that this justification is established via a shared underlying dependency relation. Instead, it is often stated that the measures are related because both are defined in terms of geodesics. We will argue that this view is rather misleading, and that closeness-as-independence and closeness-as-efficiency are actually two different concepts that happen to agree on non-valued networks. The common generalization of closeness to valued networks is in line with the efficiency interpretation only. We therefore propose new generalizations of closeness to directed, disconnected, and valued networks that maintain the independence interpretation and thus the duality with (common generalizations of) betweenness.

We start by defining necessary terminology and introducing the basic concept of a dependency cube in Section 2. The relations between dependencies and the dual indices of closeness and betweenness are derived in Section 3, leading to our re-definition of closeness-as-independence in Section 4. In Sections 5 and 6, we show how this generalizes to directed and valued networks while maintaining the duality with betweenness. We conclude in Section 7.

2. Preliminaries

We assume that networks are represented as graphs and use standard terminology such as found in Bollobás (1998) or Diestel (2010).

An (undirected) *graph* $G = (V, E)$ consists of a set V of *vertices* (also called nodes) representing actors and a set $E \subseteq \binom{V}{2}$ of (undirected) *edges* (also called links) representing ties between actors. An edge is thus an unordered pair of vertices representing a symmetric relationship. If there exists an edge $e = \{u, v\} \in E$, we say that u and v are *adjacent* and that u and v are *incident* to e. We will use $n = |V|$ for the number of vertices and $m = |E|$ for the number of edges of a graph.

A *path* from a *sender* $s \in V$ to a *receiver* $r \in V$, or (s, r)-path for short, is an alternating sequence of vertices and edges that starts with s, ends with r, and in which every vertex is incident to both the edges that come before and after it in the sequence. A graph is *connected*, if every pair of vertices is linked by a path.

In this and the following section, all graphs are assumed to be undirected and connected. The definitions will be extended to directed and valued graphs in Sections 5 and 6, where we also consider disconnected graphs.

2.1. Distance and Closeness Centrality

Closeness centrality, as the name suggests, is an index defined in terms of a distance. Let the *length* of an (s, r)-path be the number of edges contained in it. We define the *(shortest-path) distance*, $dist(s, r)$, of $s, r \in V$ as the minimum length of any (s, r)-path. Recall that we consider only connected graphs for now and observe that $dist(s, s) = 0$ for all $s \in V$.

The *distance matrix* $D = (dist(s, r))_{s,r \in V}$ of an undirected graph is symmetric, so that the total distance, $dist(v)$, of a vertex $v \in V$ is obtained as either the row and column sums

$$dist(v) = \sum_{r \in V} dist(v, r) = \sum_{s \in V} dist(s, v) \ .$$

The larger the associated distance sum, the farther a vertex is from the others, which is why a vertex is considered more central, in terms of closeness, if its associated value is smaller (Sabidussi, 1966).

Because of this reversal in ranking, *closeness centrality* of a vertex $s \in V$ is usually defined as the inverse of the total (or, equivalently, average) distance (Bavelas, 1950; Beauchamp, 1965),

$$c_C(s) = \left[\sum_{r \in V} dist(s, r) \right]^{-1} = dist(s)^{-1} \ ,$$

but sometimes also by subtraction from an upper bound on the maximum distance (Valente and Foreman, 1998).

2.2. Dependency and Betweenness Centrality

Betweenness centrality is based on the idea that brokering positions between others provide the opportunity to intercept or influence their communication. Again, the assumption is that communication is happening along shortest paths.

Denote by $\sigma(s, r)$ the number of shortest (s, r)-paths, and let $\sigma(s, r|b)$ be the number of shortest (s, r)-paths passing through some brokering vertex $b \in V \setminus \{s, r\}$. For consistency, let $\sigma(s, s) = 1$, and $\sigma(s, r|b) = 0$ if $b \in \{s, r\}$. If all shortest paths are equally likely to be chosen, the ratio $\delta(s, b, r) = \frac{\sigma(s,r|b)}{\sigma(s,r)}$ gives the probability that b is involved in the indirect communication of s with r. The term $\delta(s, b, r)$ is well-defined because $\sigma(s, r) > 0$ (for now, we assume connected graphs) and referred to as the *dependency* of a sender s and a receiver r on a broker b. From the broker's perspective it represents the degree of control that b has over the communication from s to r.

Betweenness centrality is defined as the total dependency of communicating pairs on a broker $b \in V$,

$$c_B(b) = \sum_{s,r \in V} \delta(s, b, r) \ ,$$

and thus corresponds to b's overall potential for control.

In the next section we recall and extend a largely unknown result of Freeman (1980) showing that the dependencies give rise to a dyadic relation that relates closeness and betweenness quantitatively.

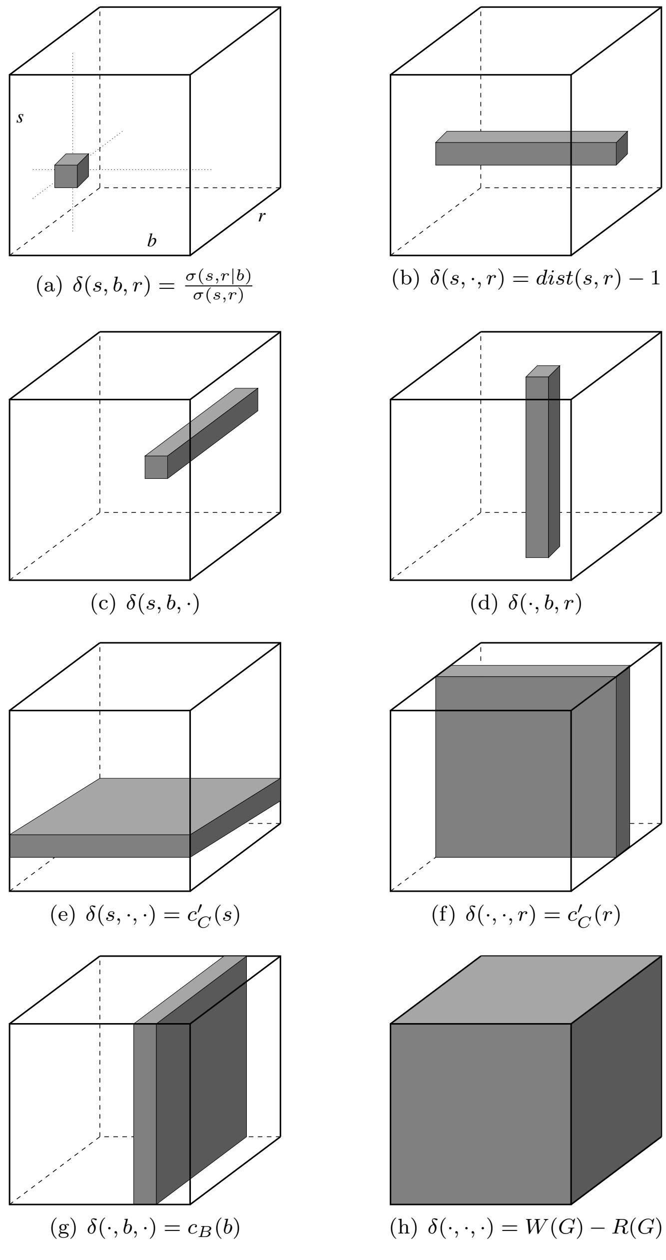
3. Dyadic Dependencies and Duality

The dependencies defined above form a three-way tensor, i.e., a generalized matrix $\Delta = (\delta(s, b, r))_{s,b,r \in V}$, the *dependency cube*. It has first been considered explicitly by Borgatti and Bonacich (1989), who referred to it as the geodesic cube. The cube assumes the role of a repository of elementary information about all communication triples consisting of a sender, a receiver, and a potential broker in between. If all n^3 entries are required, a straightforward algorithm of Batagelj (1994) can used to determine them in time $\mathcal{O}(n^3)$.

The above definition of betweenness corresponds to a summation over the (s, r)-plane in the dependency cube, and a number of other interesting quantities and insights can be obtained by summing over other subsets of elements of Δ. These are detailed next and summarized in Figure 1.

First observe that any summation of dependencies $\delta(s, b, r)$ over either the senders, brokers, or receivers yields a valued, asymmetric and dyadic relation. It relates either brokers and receivers, or senders and receivers, or senders and brokers in a square matrix and thus defines a valued network.

Consider, for example, the dependencies $\delta(s, b, \cdot)$ of senders s on brokers b obtained from summation over all receivers. These can be interpreted as quantifying how likely it is that b is involved in a communication originating at s and directed at any r, i.e., to which extent s depends on b in sending to the rest of



(a) $\delta(s,b,r) = \frac{\sigma(s,r|b)}{\sigma(s,r)}$
(b) $\delta(s,\cdot,r) = dist(s,r) - 1$
(c) $\delta(s,b,\cdot)$
(d) $\delta(\cdot,b,r)$
(e) $\delta(s,\cdot,\cdot) = c'_C(s)$
(f) $\delta(\cdot,\cdot,r) = c'_C(r)$
(g) $\delta(\cdot,b,\cdot) = c_B(b)$
(h) $\delta(\cdot,\cdot,\cdot) = W(G) - R(G)$

Figure 1: Marginals of the dependency cube

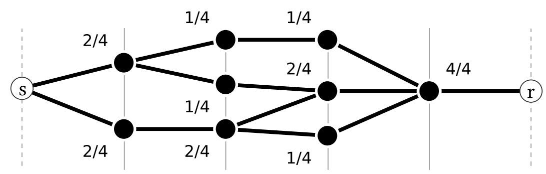


Figure 2: The fractions $\frac{\sigma(s,r|b)}{\sigma(s,r)}$ of shortest (s,r)-paths passing through those vertices b that have the same distance from s add up to 1, so that $\delta(s,\cdot,r) = \sum_{b\in V} \delta(s,b,r) = dist(s,r) - 1$.

the network by the efficient paths. These one-sided dependencies[1] thus form a new asymmetric and valued relation between senders and brokers derived from the original adjacency relation. Since

$$c_B(b) = \sum_{s,r\in V} \delta(s,b,r) = \sum_{s\in V} \delta(s,b,\cdot) \ ,$$

betweenness centrality can also be interpreted as indegree in the derived network. It thus quantifies the extent to which senders depend on b. It is interesting to note that, for a given sender s, one-sided dependencies $\delta(s,b,\cdot)$ can be computed by accumulating dependencies on brokers farther away from s, so that it is computationally more efficient to determine them directly rather than by explicitly determining all entries of Δ and subsequent summation (Brandes, 2001).

Similarly, marginals $\delta(\cdot,b,r)$ can be interpreted as the dependencies of receivers r on gatekeepers b to let incoming information through. By symmetry, betweenness in the original graph corresponds to outdegree in the graph defined by the valued relation $\delta(\cdot,b,r)$.

A key observation is that the third matrix of marginals $\delta(s,\cdot,r)$, the dependency of each pair s,r on the rest of the network, is almost identical to the matrix of shortest-path distances (recall that we are considering non-valued networks for now). This was already observed in Freeman (1980), Borgatti and Bonacich (1989), and independently in Büchel (2009, Lemma 4.5.1). We include a straightforward proof illustrated in Figure 2.

Lemma 1. *In a connected graph,* $\delta(s,\cdot,r) = dist(s,r) - 1$ *for all* $s \neq r \in V$.

Proof. At every distance $i = 1,\ldots,dist(s,r)-1$ from s, each shortest path from s to r passes through exactly one broker b, so that for $s \neq r$ we have

$$\delta(s,\cdot,r) = \sum_{b\in V} \delta(s,b,r) = \sum_{i=1}^{dist(s,r)-1} \underbrace{\sum_{b\in V\,:\,dist(s,b)=i} \frac{\sigma(s,r|b)}{\sigma(s,r)}}_{=1} = dist(s,r) - 1 \ .$$

[1] Freeman (1980) uses the term *pair-dependencies* which we avoid as it is prone to misinterpretation in our more general context.

This observation leads to two interesting insights: the quantitative duality of closeness and betweenness, and the reason for the alternative interpretations of closeness-as-efficiency and closeness-as-independence.

Closeness and betweenness centrality are dual to each other conceptually. While one quantifies the independence from control of others, the other quantifies the potential control over (communication between) others. The following lemma shows that the relation is not only conceptual but holds quantitatively.

Corollary 2. *In a connected graph,* $\sum_{b \in V} \delta(s,b,\cdot) = c_C(s)^{-1} - (n-1)$.

PROOF. Using Lemma 1, we get $\sum_{b \in V} \delta(s,b,\cdot) = \delta(s,\cdot,\cdot) = \sum_{r \in V} \delta(s,\cdot,r) = \sum_{r \in V} [dist(s,r) - 1] = c_C(s)^{-1} - (n-1)$.

Closeness and betweenness are thus dual in the sense that they are obtained as (the inverse of) row and column sums (i.e., outdegree and indegree) of the dependency relation $\delta(s,b,\cdot)$,

$$c_B(b) = \sum_{s \in V} \delta(s,b,\cdot) \quad \text{and} \quad c_C(s)^{-1} = (n-1) + \sum_{b \in V} \delta(s,b,\cdot) \; .$$

Backed by formal arguments we can therefore state that betweenness is in exactly the same sense a measure of control, or the dependency *of* others on an actor, as closeness is a measure of independence, or the lack of dependency *on* others. As exemplified in Figure 3, both are directly related via the dependency relation $\delta(s,b,\cdot) = \sum_{r \in V} \frac{\sigma(s,r|b)}{\sigma(s,r)}$, for which inverse closeness corresponds to weighted outdegree (up to a constant) and betweenness to weighted indegree.

Moreover, as demonstrated by the examples in Figure 4, the rankings obtained from these dual notions may coincide but may also be quite different from each other.

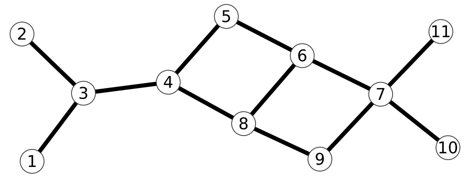
While no deep mathematics are involved, and despite an explicit derivation in Freeman (1980), this relationship has been largely overlooked. We deem it important, however, because it adds strong support for the interpretation duality that empirical researchers have been relying on, and even more so because it has important consequences for the generalization of closeness to unconnected, directed, and valued networks as discussed in the subsequent sections.

4. Interpretation and Adjustment of Closeness Centrality

For the interpretation of closeness centrality we focus on total distances, $dist(s) = c_C(s)^{-1}$, because the sole purpose of taking inverses is to reverse the order of values and this could be achieved in any of a number of ways.

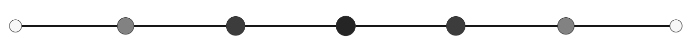
An interpretation obtained directly from its definition is one of efficiently reaching others:

> "...a point is central to the degree that the distances associated with all its geodesics are minimum. Short distances mean fewer message transmissions, shorter times and lower costs." (Freeman, 1979, p. 225)

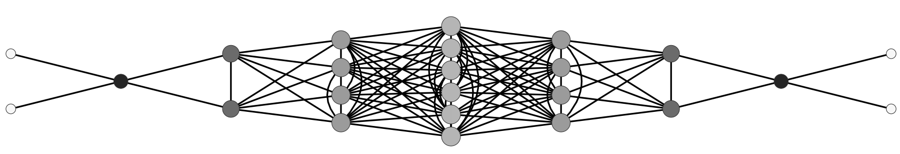


$\delta(s,b,\cdot)$	1	2	3	4	5	6	7	8	9	10	11	$c_C(s)^{-1} - 10$
1	0	0	9	7	$1\frac{1}{2}$	2	2	$3\frac{1}{2}$	1	0	0	26
2	0	0	9	7	$1\frac{1}{2}$	2	2	$3\frac{1}{2}$	1	0	0	26
3	0	0	0	7	$1\frac{1}{2}$	2	2	$3\frac{1}{2}$	1	0	0	17
4	0	0	2	0	$1\frac{1}{2}$	2	2	$3\frac{1}{2}$	1	0	0	12
5	0	0	2	$3\frac{5}{6}$	0	$4\frac{1}{6}$	$2\frac{1}{3}$	$\frac{2}{3}$	0	0	0	13
6	0	0	2	3	2	0	$2\frac{1}{2}$	$2\frac{1}{2}$	0	0	0	12
7	0	0	2	3	$1\frac{1}{3}$	$4\frac{1}{6}$	0	$2\frac{1}{3}$	$1\frac{5}{6}$	0	0	15
8	0	0	2	$3\frac{1}{2}$	0	2	2	0	$1\frac{1}{2}$	0	0	11
9	0	0	2	$3\frac{1}{3}$	0	$\frac{2}{3}$	$2\frac{5}{6}$	$5\frac{1}{6}$	0	0	0	14
10	0	0	2	3	$1\frac{1}{3}$	$4\frac{1}{6}$	9	$2\frac{2}{3}$	$1\frac{5}{6}$	0	0	24
11	0	0	2	3	$1\frac{1}{3}$	$4\frac{1}{6}$	9	$2\frac{2}{3}$	$1\frac{5}{6}$	0	0	24
$c_B(b)$	0	0	34	$43\frac{2}{3}$	12	$27\frac{1}{3}$	$35\frac{2}{3}$	$30\frac{1}{3}$	11	0	0	194 $= W(G) - R(G)$

Figure 3: In connected undirected graphs, inverse closeness and betweenness are the row and column sums of the asymmetric dependency relation $\delta(s,b,\cdot) = \sum_r \delta(s,b,r)$



(a) path



(b) path with vertices expanded to cliques of different size, plus pendants

Figure 4: Actors that others depend on are not necessarily independent: darkness indicates betweenness whereas size indicates independence. In (b), the independent actors in the middle are easier to avoid than the outer ones with the exception of the pendants

As discussed in more detail in Section 6, this is the interpretation on which common generalizations of closeness to valued networks are based.

It appears, however, that the interpretation of closeness as being dual to the potential for control associated with betweenness,

> "...a point is viewed as central to the extent that it can avoid the control potential of others." (Freeman, 1979, p. 224)

is the more prominent (see, e.g., Brass, 1984; Rowley, 1997; Powell et al., 1996). Since dyadic dependencies add up to shortest path distance minus one, they actually correspond to the number of intermediaries on a shortest path. The higher this number, the more dependent an actor is on others. From the point of view of the independence interpretation, the row sums in the matrix of $\delta(s, b, \cdot)$ thus reflect the intended meaning even better than the distances.

It will prove useful beyond the elimination of constants to define the following variant of closeness centrality.

Definition 3. *For a graph $G = (V, E)$,* closeness-as-independence *is defined via*

$$c'_C(s)^{-1} = \delta(s, \cdot, \cdot)$$

for all $s \in V$.

Clearly, when compared to $c_C(s)$, this does not affect the ranking of vertices in a connected undirected graph because all values are shifted equally by $(n-1)$,

$$c'_C(s)^{-1} = \delta(s, \cdot, \cdot) = \sum_{b \in V} \delta(s, b, \cdot) = dist(s) - (n-1) = c_C(s)^{-1} - (n-1) \ .$$

The modification does have, however, important consequences for networks of valued relations, and also for directed and disconnected networks as shown in the next section.

5. Reachability and Disconnected Networks

Closeness centrality is ill-defined on disconnected graphs because some distances are undefined and it is not clear how to compare partial sums having different numbers of defined distances. We will address this problem together with the generalization of the above results to networks of asymmetric relations.

A graph $G = (V, E)$ is called *directed*, or *digraph* for short, if its edges are defined as ordered rather than unordered pairs of vertices. We write $(v, w) \in E \subseteq V \times V$ or $v \to w$ to distinguish directed edges from undirected edges $\{v, w\} \in \binom{V}{2}$.

A path is called *directed*, if each of its edges is directed from its preceding to its succeeding vertex. If there exists a directed (s, r)-path, r is said to be *reachable* from s. Reachability is the reflexive and transitive closure of adjacency and thus a reflexive and transitive, but not necessarily symmetric, relation $s \to^*$

r. We say that a digraph is *strongly connected*, if every vertex is reachable from every other vertex.

The definition of closeness centrality generalizes to strongly connected digraphs, although asymmetry now forces us to decide whether distances should be measured from or to the focal vertex. For convenience, we will only consider closeness centrality in terms of distances *from* a vertex, the other case is symmetric.

For digraphs that are not strongly connected, the number of intermediaries an actor depends on is meaningful only in relation to the number of possible receivers that can actually be reached. For a digraph $G = (V, E)$ let $R(G) = |\{(s, r) \in V \times V \,:\, s \neq r,\, s \rightarrow^* r\}|$ be the number of ordered non-loop pairs in the reachability relation. For a vertex $v \in V$, we define $R^+(v) = |\{r \in V \,:\, v \rightarrow^* r\}|$ to be the number of reachable vertices and $R^-(v) = |\{s \in V \,:\, s \rightarrow^* v\}|$ to be the number of reaching vertices, so that $R(G) = \sum_{s \in V} R^+(s) = \sum_{r \in V} R^-(r)$. Furthermore, we let

$$W(G) = \sum_{s,r \in V:\, s \rightarrow^* r} dist(s, r)$$

be the sum of all defined distances. This sum is known as the *Wiener index* (Wiener, 1947) and often used as a network-level characteristic. Note that *characteristic path length* (Watts and Strogatz, 1998), one of the dimensions to assess whether a network is considered a small world, is defined as the average distance of any pair of vertices, $\frac{W(G)}{R(G)}$, and therefore yields simply a normalized version of the Wiener index.

The following result shows that the total (and thus also average) closeness-as-independence and betweenness are equal, and that they correspond to the Wiener index corrected for reachability.

Theorem 4. *For a directed graph $G = (V, E)$,*

$$\sum_{s \in V} c'_C(s)^{-1} \;=\; W(G) - R(G) \;=\; \sum_{b \in V} c_B(b) \;.$$

Proof. Again, we express all quantities in terms of three-way dependencies, and thus obtain $\sum_{s \in V} c'_C(s)^{-1} = \sum_{s \in V} \sum_{b \in V} \delta(s, b, \cdot) = \sum_{b \in V} \sum_{s \in V} \delta(s, b, \cdot) = \sum_{b \in V} c_B(b)$, i.e., both totals, for closeness and betweenness, are equal to the total of all dependencies $\sum_{s,b,r \in V} \delta(s, b, r) = \sum_{s \neq r \in V:\, s \rightarrow^* r} dist(s, r) - 1 = W(G) - R(G)$.

Closeness-as-independence and betweenness centrality are thus partitions of the same total volume, which is smaller for more compact graphs. However, they divide this volume up in different ways as described in the previous section. This common scale may also be of interest in a comparative analysis of closeness and betweenness in terms of endogenous and exogenous centrality (Everett and Borgatti, 2010).

For disconnected undirected graphs or non-strongly connected digraphs, in which the reachability relation is not complete, betweenness retains its interpretation as the total potential for control of shortest-path connections. From the

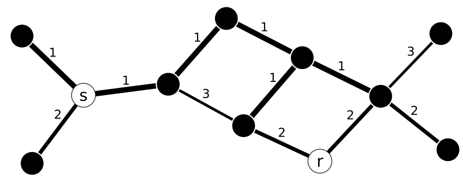


Figure 5: In this valued graph, $dist(s,r) = 6$, and there are three shortest (s,r)-path with 2, 4, and 4 inner vertices for an average of $3\frac{1}{3} = \delta(s,\cdot,r)$ intermediaries that s and r depend on.

would-be broker's point of view, it may not make make a difference whether a given pair can connect via better paths that do not involve the broker, or can't connect at all: either way the broker will not be brokering between them.

From the point of view of an actor avoiding dependence on brokers, however, it may in fact make a difference whether the same number of intermediaries control the connections to many or few reachable receivers. In the absence of a substantive justification for combining dependency with the number of reachable receivers, closeness centrality should therefore be treated as a bi-criterial index, i.e., the two values for the total number of intermediaries and the number of reachable receivers should not be combined into a single quantity.

6. Generalized Distance and Valued Networks

During generalization to valued graphs the difference between closeness and our variant closeness-as-independence becomes most apparent. In fact, the duality with betweenness is maintained only by closeness-as-independence. This difference highlights the fact that the interpretations of closeness as either efficiency or independence are actually distinct indices that happen to coincide on non-valued networks.

Let a *valued* graph be defined as a graph $G = (V, E; \lambda)$ with edge values $\lambda : E \to \mathbb{R}$. Such values typically have a positive sign and represent a distance or lag in the connection between adjacent vertices, so that the length of a path in a valued graph is generally defined as the sum of the values of its edges (Flament, 1963). Note that the definition of path length as the number of edges given in Section 4 is the special case in which all edges have a length of 1. Distances $dist(s,r)$ are then defined as before, i.e., as the minimum length of any (s,r)-path. While other values are possible and other generalizations of shortest or best paths exist (e.g., Yang and Knoke, 2001; Opsahl et al., 2010), this appears to be the most frequently employed.

From any generalization of path lengths to valued graphs we obtain straightforward generalizations of closeness and betweenness. Since the dependencies $\delta(s,b,r)$ are defined as the fraction of optimal (s,r)-paths passing through b, independent of the value associated with such paths, the interpretation of betweenness is preserved in any such generalization.

The interpretation of closeness as indicating access efficiency is also preserved as long as distances still represent an effort necessary for the sender to reach the receiver. This is not necessarily true, however, for the interpretation of closeness as indicating the independence of an actor from others, because Lemma 1 no longer holds. This is illustrated in Figure 5.

The rationale behind the latter interpretation of closeness was the independence from intermediaries. Intermediaries, however, feature in the standard definition of closeness only because their number corresponds to the number of edges in a path minus one. It is thus rather by coincidence than by design that distance and dependence almost agree in non-valued graphs.

Our variant closeness-as-independence index, on the other hand, preserves the duality with betweenness established above for the case of non-valued networks. Recall that $c_B(b) = \sum_{b \in V} \delta(s, b, \cdot)$ and $c'_C(s)^{-1} = \sum_{s \in V} \delta(s, b, \cdot)$ in non-valued networks. Since the interpretations of $\delta(s, b, r)$ and thus $\delta(s, b, \cdot)$ are not affected by generalization to valued networks, both betweenness and closeness-as-independence generalize straightforwardly with the other generalized quantities, and their duality as indegree and outdegree is maintained.

With a suitably defined distance, closeness-as-independence can still be regarded as the inverse of a total distance as shown in the following Lemma 1.

Lemma 5. *In a valued graph $G = (V, E; \lambda)$, $\delta(s, \cdot, r)$ equals the average number of inner vertices in shortest (s, r)-paths for any $s, r \in V$.*

Proof. For any pair of vertices $s, r \in V$, $\delta(s, \cdot, r) = \sum_{b \in V} \frac{\sigma(s,r|b)}{\sigma(s,r)}$ by definition. Since every inner vertex b contributes 1 to each shortest path it is contained in, $\sum_{b \in V} \sigma(s, r|b)$ is the total number of occurrences of inner vertices in any shortest path, and division by the number $\sigma(s, r)$ of shortest (s, r)-path yields the average.

Therefore, since $c'_C(s) = \delta(s, \cdot, \cdot) = \sum_{r \in V} \delta(s, \cdot, r)$ by Definition 3, our adjustment of closeness to maintain the independence interpretation can also be seen as replacing shortest-path distances $dist(s, r)$ with the average numbers of intermediaries $\delta(s, \cdot, r)$ on shortest paths. Note that this also holds in non-valued networks because all shortest (s, r)-paths have the same number $dist(s, r) - 1$ of inner vertices (Lemma 1).

Like betweenness and closeness-as-efficiency, our variant closeness-as-independence index can be determined in $\mathcal{O}(nm + n^2 \log n)$ time because a single-source shortest-paths computation from a sender s yields the dependencies $\delta(s, b, \cdot)$ for all $b \in V$ (Brandes, 2001).

7. Discussion

We have pointed out a formal duality between closeness and betweenness centrality that, although known, has long been used only in conceptual terms. The duality is expressed in terms of a derived relation, the dyadic dependency of senders on brokers. Betweenness and closeness are in fact the weighted indegrees and outdegrees in the network of this derived relation. Since total betweenness and closeness in a graph thus equal the total of the dyadic dependencies, they also equal the sum of distances in a graph minus the number of reachable pairs.

Closeness and betweenness yield the same ranking on paths, star graphs, cliques, and a number of other graphs. It will be interesting to investigate by how much they can actually differ. We gave an example (a path of cliques of varying size) in which the rankings are almost the reverse of each other.

The observed duality generalizes to directed and non-connected networks, no matter whether closeness is generalized by introducing a finite distance for unreachable pairs or by considering total distance and number of reachable vertices as a two-dimensional index. By reversing edge directions, it is easily confirmed that the corresponding dependency of receivers on brokers corresponds to closeness defined by distances to, rather than from, an actor.

In valued networks it becomes apparent that the two interpretations of closeness centrality as efficiency and as independence actually refer to two different concepts that happen to coincide (up to an additive constant) in non-valued networks. Duality is maintained in valued networks only if the definition of closeness is adapted. By replacing the sum of distances with a sum of dependencies, we effectively replace shortest-path distance with the expected number of intermediaries on a shortest path. Since this is in line with the original motivation for closeness centrality as an indicator of independence (Freeman, 1979), we consider it a strong argument for our new variant in cases where closeness is interpreted as independence.

Finally, we note that the concept of dual centrality indices applies more generally to all indices that are co-determined by an asymmetric relation derived from the original network. Closeness and betweenness are row and column sums of dyadic dependencies, so indegree and outdegree on other relations are obvious extensions. For certain derived relations, however, we also expect meaningful dualities to arise from left and right singular vectors.

Bibliography

Batagelj, V., 1994. Semirings for social network analysis. Journal of Mathematical Sociology 19 (1), 53–68.

Bavelas, A., 1950. Communication patterns in task-oriented groups. Journal of the Acoustical Society of America 22 (6), 725–730.

Beauchamp, M. A., 1965. An improved index of centrality. Behavioral Science 10, 161–163.

Bollobás, B., 1998. Modern Graph Theory. Vol. 184 of Graduate Texts in Mathematics. Springer-Verlag.

Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology 2, 113–120.

Bonacich, P., 1987. Power and centrality: A family of measures. American Journal of Sociology 92 (5), 1170–1182.

Borgatti, S. P., Bonacich, P., 1989. The geodesic cube: A 3-dimensional object underlying standard measures of centrality. Presented at the Sunbelt IX Social Networks Conference.

Borgatti, S. P., Everett, M. G., 2006. A graph-theoretic framework for classifying centrality measures. Social Networks 28 (4), 466–484.

Brandes, U., 2001. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25 (2), 163–177.

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.

Büchel, B., 2009. Advances in strategic network formation: Preferences, centrality, and externalities. Ph.D. thesis, University of Bielefeld, Department of Economics.

Diestel, R., 2010. Graph Theory, 4th Edition. Vol. 173 of Graduate Texts in Mathematics. Springer-Verlag.

Everett, M. G., Borgatti, S. P., 2010. Induced, endogenous and exogenous centrality. Social Networks 32 (4), 339–344.

Flament, C., 1963. Applications of Graph Theory to Group Structure. Prentice Hall.

Freeman, L. C., 1979. Centrality in social networks: Conceptual clarification. Social Networks 1 (3), 215–239.

Freeman, L. C., 1980. The gatekeeper, pair-dependency and structural centrality. Quality and Quantity 14 (4), 585–592.

Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D., Zlotowski, O., 2005. Centrality indices. In: Brandes, U., Erlebach, T. (Eds.), Network Analysis: Methodological Foundations. Vol. 3418 of Lecture Notes in Computer Science. Springer-Verlag, pp. 16–61.

Opsahl, T., Agneessens, F., Skvoretz, J., 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 32 (3), 245–251.

Powell, W. W., Koput, K. W., Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. Administrative Science Quarterly 41 (1), 116–145.

Rowley, T. J., 1997. Moving beyond dyadic ties: A network theory of stakeholder influences. Academy of Management Review 22 (4), 887–910.

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

Valente, T. W., Foreman, R. K., 1998. Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. Social Networks 20 (1), 89–105.

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

Wiener, H., 1947. Structural determination of paraffin boiling points. Journal of the American Chemical Society 69 (1), 17–20.

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