Social Networks 44 (2015) 153–159

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Maintaining the duality of closeness and betweenness centrality☆

Ulrik Brandes[a,*], Stephen P. Borgatti[b], Linton C. Freeman[c]

[a] Department of Computer & Information Science, University of Konstanz, Germany
[b] Gatton College of Business and Economics, University of Kentucky, United States
[c] Institute of Mathematical Behavioral Sciences, University of California at Irvine, United States

ARTICLE INFO

Keywords:
Betweenness centrality
Closeness centrality
Duality
Dependency
Derived relations

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 unveils that closeness-as-independence is in fact different from closeness-as-efficiency, and we propose a variant notion of distance that maintains the duality of closeness-as-independence with betweenness also on valued relations.

1. Introduction

A number of attempts have been made to bring order to the universe of centrality measures, including Sabidussi (1966), Koschützki et al. (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; Brin and Page, 1998).

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 et al. (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 2.1. In Sections 5 and 6, we

☆ Research partially supported by DFG under grant Br 2158/6-1.
* Corresponding author. Tel.: +49 7531 88 4433; fax: +49 7531 88 3577.
E-mail addresses: Ulrik.Brandes@uni-konstanz.de (U. Brandes), sborgatti@uky.edu (S.P. Borgatti), lin@aris.ss.uci.edu (L.C. Freeman).

http://dx.doi.org/10.1016/j.socnet.2015.08.003
0378-8733/

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 Fig. 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 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

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