UC Irvine
UC Irvine Previously Published Works

Title

Betweenness centrality measures for directed graphs

Permalink

Journal

Social Networks, 16(4)

ISSN

Authors

White, Douglas R
Borgatti, Stephen P

Publication Date

1994-10-01

DOI

Copyright Information

Peer reviewed

Betweenness centrality measures for directed graphs

Douglas R. White [*,a], Stephen P. Borgatti [b]

[a] *University of California, Irvine, CA 92717, USA*
[b] *University of South Carolina, Columbia, SC 29208, USA*

This paper generalizes Freeman's geodesic centrality measures for betweenness on undirected graphs to the more general directed case. Four steps are taken. The point centrality measure is first generalized for directed graphs. Second, a unique maximally centralized graph is defined for directed graphs, holding constant the numbers of points with reciprocatable (incoming and outgoing) versus only unreciprocatable (outgoing only or incoming only) arcs, and focusing the measure on the maximally central arrangement of arcs within these constraints. Alternatively, one may simply normalize on the number of arcs. This enables the third step of defining the relative betweenness centralities of a point, independent of the number of points. This normalization step for directed centrality measures removes Gould's objection that centrality measures for directed graphs are not interpretable because they lack a standard for maximality. The relative directed centrality converges with Freeman's betweenness measure in the case of undirected graphs with no isolates. The fourth step is to define the measures of this concept of graph centralization in terms of the dominance of the most central point.

1. Introduction

Betweenness centrality (Freeman 1977, 1979, 1980) is a fundamental measurement concept for the analysis of social networks. The recent book by Hage and Harary (1991) demonstrates some of its many descriptive and predictive uses. It was originally defined, however, only for undirected graphs. This constitutes a rather severe limit on its potential utility for directed (nonsymmetric) graphs and social networks. Gould (1987) argues that measures of betweenness centrality are possible for the directed case, but that owing to lack of a unit of measurement and of a unique definition of the maximally centralized graph for the directed case, the measure remains uninterpretable.

* Corresponding author.

The present study identifies a unit of measurement and a uniquely maximal centralized graph for the directed case, and so provides for a proper generalization of betweenness centrality to directed graphs.

2. Generalization of betweenness centrality to directed graphs

Freeman (1980) showed how betweenness centrality for undirected graphs is derived from the column totals of a single matrix of the numbers of pairwise dependencies of each point on every other point in terms of mediating access in reaching third points.

The following paragraphs parallel Freeman's (1980: 587–588) derivation of the measure of pair-dependency, generalized here (with appropriate modifications) to directed graphs.

"Consider a directed graph representing the nonsymmetric relation 'communicates to ' for a set of people. When point i communicates to point j, j is said to be adjacent from i. A set of arcs linking i to j to k constitute a path from p_i to p_k. The shortest path linking one point to another is called a geodesic. There can, of course, be more than one geodesic for any ordered pair of points."

"Now let g_{ik} = the number of geodesics from p_i to p_k, and $g_{ik}(p_j)$ = the number of geodesics that contain point p_j as an intermediary in the geodesics from p_i to p_k, then:

$$b_{ik}(p_j) = \frac{g_{ik}(p_j)}{g_{ik}}$$

Thus, $b_{ik}(p_j)$ is the proportion of geodesics linking p_i to p_k that contain p_j: it is an index of the degree to which p_i and p_k need p_j in order to communicate along the shortest path linking them together. Since it is a proportion, $0 \le b_{ik}(p_j) \le 1$. Moreover, when $b_{ik}(p_j) = 1, p_j$ is strictly between p_i and p_k; they cannot communicate along the geodesic(s) linking them without its support in relating messages. In such a situation the communication from p_i to p_k is completely at the whim of p_j: he can distort or falsify any information passing through him."

"Now we can define pair-dependency as the degree to which a point, p_i, must depend upon another, p_j, to relay its messages along

geodesics to other reachable points in the network. Thus, for a network containing n points,

$$d_{ij}^* = \sum_{k=1}^{n} b_{ik}(p_j)(i \neq j \neq k)$$

is the pair-dependency of p_i on p_j."

"We can calculate the pair-dependencies of each point on every other point in the network and arrange the results in a matrix,

$$D = (d_{ij}^*)$$

Each entry in D is an index of the degree to which the point designated by the row of the matrix must depend on the point designated by the column to relate messages to others. Thus D captures the importance of each point as a gatekeeper with respect to each other point—facilitating or perhaps inhibiting its communication."

"Whenever any person is in a position to be a gatekeeper for communications, others must depend on that person. A gatekeeper position, however, may be either rather wide or quite narrow in its impact..."

"Obviously, such local pair effects are of great potential importance to the points affected. In large networks, where individuals may be at considerable distance from one another, global patterns may be submerged and pair effects may be the main factors determining information flows."

Freeman (1980) shows for undirected graphs that the sums down the columns of the pair-dependency matrix D is a measure of betweenness centrality of the points:

$$\sum_{i=1}^{n} d_{ij}^* = 2C_B(p_j)$$

$C_B(p_j)$, the betweenness centrality of point j, is twice the value of its pair dependency column sum: twice because the upper diagonal and lower diagonals are equal in the pair-dependency matrix of an undirected graph.

To generalize Freeman's measure of betweenness point centrality, $C_B(p_j)$, for directed graphs, we use the equality:

$$C_B(p_j) = \sum_{i=1}^{n} d_{ij}^*$$

We divide further by two for comparability with Freeman's measure for undirected graphs.

3. The maximally centralized graph

It is desirable for some purposes to have a measure of betweenness centralities that is not affected by the number of points in the graph (Leavitt 1951). Freeman (1977: 37–40) derived such a measure by dividing $C_B(p_j)$ by the maximal value it can take relative to points in the graph. In showing the existence of the maximum, he also showed that the unit of measurement for increases in a point's centrality in a graph was the addition of an arc to the graph that increased the number of paths through the point. The measure of centrality monotonically increases as such arcs are added to a graph, and reaches its maximum when no more such arcs can be added. He also proved by this procedure that an undirected graph with maximal betweenness centrality for one of its points is always a star, where the central point is adjacent to all of the other points and none of the other points are adjacent to one another.

For meaningful interpretation of a measure of centrality, one needs to be able to derive its *minimum* and *maximum*, and a *unit of measurement* with respect to which it is monotone. Gould (1987) examined this question and was unable to define either a maximum value or a unit of measurement for directed graphs. Thus, he argued that while it was possible to *define* a betweenness centrality measure for directed graphs, one could not *interpret* the measure. Borgatti and Bonacich (1989) also arrived at the measure of $C_B(p_j)$ for directed graphs, without the normalization and unit of measurement.

In a directed graph, if n_O is the number of points with outgoing arcs and n_I the number with incoming arcs, then in a star graph the product $(n_I - 1)(n_O - 1)$ is the number of paths through the center. The directed graph with maximal point betweenness, *given n_I and n_O*,

is a star in which points with incoming and outgoing arcs are maximally disjoint. In such a star, each arc contributes to the betweenness centrality of the center point. Let n_S = the number of points with reciprocated arcs. The betweenness centrality of the most centralized star for a directed graph is:

$$C_B(p_j) = (n_I - 1)(n_O - 1) - (n_S - 1)$$

Again we divide by 2 for consistency with Freeman's measure. In the case of an undirected graph, $n_I = n_O = n_S = n$, and the formula for betweenness centrality of a star simplifies to:

$$C_B(p_j) = (n - 1)(n - 1) - (n - 1) = (n^2 - 3n + 2)$$

Divided by 2, this converges with Freeman's result for the betweenness centrality of the most central point in the most centralized (undirected) graph.

Now let us duplicate for the directed case Freeman's procedure for adding arcs to a graph, beginning with isolated points, so as to increase the betweenness centrality of a single most central point until its maximum is reached. To do so, we must add arcs that connect peripheral points in a star to the central point. In the directed case, we must add the arcs so as to form the maximum number of (directed) paths that run through the center point. The simplest graph with a directed star pattern, as shown in Fig. 1, begins with two arcs, from i to j and j to k, j being the central point that sits on the path from i to k. To obtain the maximal value of $C_B(p_j)$ with each new arc with additional points, one must alternate between addition of incoming and outgoing arcs with respect to the central point j. Figure 2 shows two ways of adding an arc to the graph in Fig. 1 to span another point and retain maximal betweenness centrality. For the first $n - 1$ arcs added to a graph with n points, reciprocated connections between
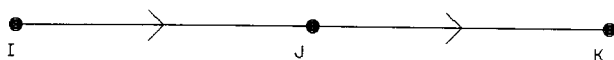
$C_B(p_j) = 1$ is the maximal betweenness

Fig. 1. Simplest directed star.
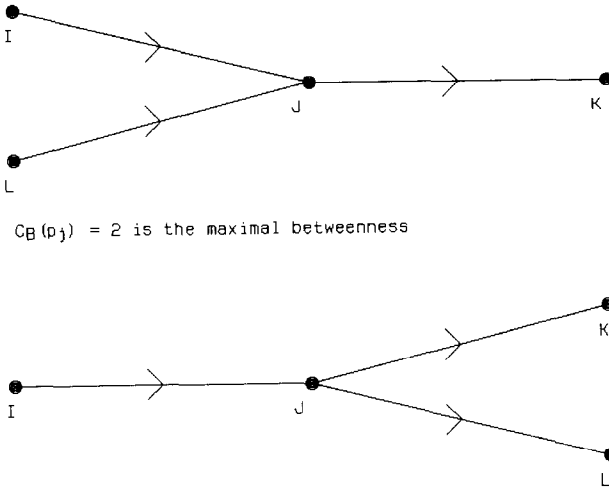
$C_B(p_j) = 2$ is the maximal betweenness

Fig. 2. Next larger directed stars.

center and periphery are avoided because they produce less of an increase in betweenness centrality than adding only arcs that are unreciprocated. Figure 3 shows a graph with maximal betweenness centrality of point j with 5 points and 4 arcs. It has 4 (directed) paths mediated by point j and gives $C_B(p_j) = 4$. Figure 4 shows two graphs where instead of the last arc $m \rightarrow j$ in Fig. 3 we add arcs that reciprocate the relation either between i and j or between j and k. Here $C_B(p_j) = 2$ in one case and $C_B(p_j) = 3$ in the other. These are not graphs of maximal centralization.

Once the first $n - 1$ directed arcs are added to maximize the betweenness centrality of the center point with the addition of each arc, the star pattern so formed connects each of the graph's $n - 1$
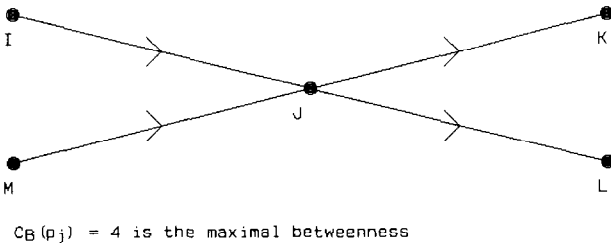

$C_B(p_j) = 4$ is the maximal betweenness
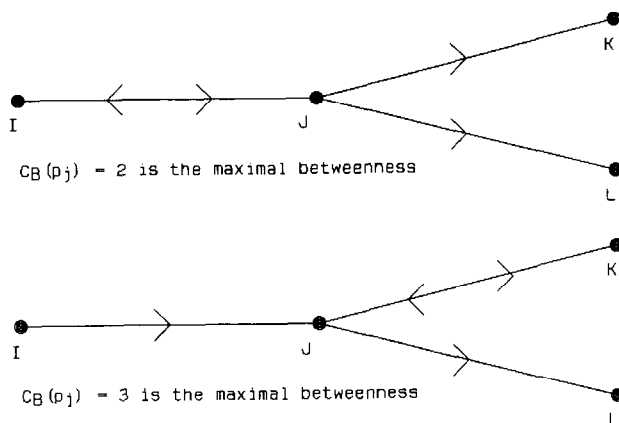
Fig. 3. Successively larger directed star.

Fig. 4. Alternate directed stars.

peripheral points to the central point j. This is the *maximal directed star* pattern of an asymmetric graph. If the number of points is odd, one has $n_I = n_O = (n+1)$ and $n_S = 0$, and

$$\max C_B(p_j) = ((n+1)/2 - 1)((n+1)/2 - 1) = (n^2 - 2n + 1)/4$$

In Fig. 3, for example, $\max C_B(p_j) = (25 - 10 + 1)/4 = 4$. If the number of points is even, the maximal centrality is

$$\max C_B(p_j) = (n/2 - 1)(n/2 + 1 - 1) = (n^2 - 2n)/4$$

In Fig. 2, $\max C_B(p_j) = (16 - 8)/4 = 2$. In general,

$$\max C_B(p_j) = (n/2 - 1)(n/2 + 1 - 1) = (n^2 - 2n + \text{modulo}(n, 2))/4$$

This is the maximum of star-pattern centralities for asymmetric graphs, but it is not the maximum for a directed (nonsymmetric) graph, since one can continue to increase the centrality of the central point by adding reciprocated arcs (again, in an alternating fashion), in which case one arrives at Freeman's undirected star as the limiting case.

Since we have established a procedure for adding arcs so as to monotonically and maximally increase the measure of centrality, the question of the meaningfulness of the measurement for betweenness centrality in directed graphs is not a matter of a lack of a unit of

measurement. It is that while a maximum is well defined for the asymmetric and undirected (symmetric) graphs, the maximum is not yet uniquely defined for the general case of the directed (nonsymmetric) graph.

Defining betweenness centrality for directed graphs must depend on a specification of the parameters n_I, n_O and n_S. At one extreme we have asymmetric graphs where $n_S = 0$, and at the other we have undirected graphs where $n_S = n$. Maximal betweenness is in each case a specialization of the general formula:

$$\max\ C_B(p_j) = (n_I - 1)(n_O - 1) - (n_S - 1)$$

What is the maximal betweenness centrality for intermediate case where $0 < n_S < n$? We need to define a maximal value that takes into account a limit on the number of points with reciprocatable arcs having both incoming and outgoing arcs (only for such points can arcs be reciprocated). This implies a limit on the remaining points with unreciprocatable arcs (with only incoming or only outgoing arcs). The simplest way to do this is to hold constant, as observed in the actual graph, the number n_S of points with reciprocatable arcs and those with only unreciprocatable arcs ($n_I - n_S$ and $n_O - n_S$) [1]. Then maximum point centrality is obtained by our procedure for adding arcs up to the limit on the number of points with only in- or outwardly oriented arcs, and then up to the limit on the separate set of points with reciprocatable arcs (the unit of measurement procedure for constructing the maximal graph requires we add arcs at first as unreciprocated arcs between the center and each peripheral point and then proceed to make them reciprocal, following the rule of alternating orientation). Then our general formula will apply where the parameters n_I, n_O, and n_S are fixed from the empirical graph under study. [2]

---

[1] Alternately, we may simply normalize on the number of arcs, as follows. Take the k arcs, allocate $k + 1 - n$ of them to n_S reciprocatable points, and divide the remainder among the remaining points so that $n_I + n_O - n_S = n$ and n_I and n_O differ at most by one. This defines the most centralized graph holding constant only the number of arcs, but does not control for the degree of symmetry in the graph.

[2] The $C_B(p_j)$ measure for the maximum graph, in the case of undirected graphs with no isolates, will converge with Freeman's measure (multiplied by 2).

4. Relative betweenness centrality for directed graphs: interpretability

The relative betweenness centrality of a point j is its betweenness centrality $C_B(p_j)$ divided by the maximum betweenness centrality, max $C_B(p_j)$, holding constant the number of nodes in the graph with reciprocatable arcs n_S or only unreciprocatable arcs, $n_I - n_S$, $n_O - n_S$:

$$C_B'(p_j) = C_B(p_j)/\text{max } C_B(p_j)$$

$$= C_B(p_j)/(n_I - 1)(n_O - 1) - (n_S - 1)$$

The measure $C_B'(p_j)$ converges with Freeman's relative betweenness centrality measure in the case of undirected graphs with no isolates. Normalizing the point centrality measure establishes comparability between graphs not only of different sizes but of differing degrees of symmetry, and prevents the measure of relative centrality in graphs with very few reciprocated relations from being deflated because of their lack of symmetry. This step (or the alternate one in note 1) in normalizing the directed centrality measure removes Gould's (1987) objection that centrality measures for directed graphs are uninterpretable because they lack a standard for maximality.

What is the interpretation for the betweenness centrality of a point relative to one with maximal betweenness in a graph of the same number of points and of points with reciprocatable or only unreciprocatable arcs of each type? By holding constant the number of points in terms of their types of arcs, we are measuring the extent to which the existing arcs between these types of points could be rearranged—adding arcs to the center and deleting other arcs to peripheral points—to increase the centrality of the dominant point.

5. Measures of graph centralization

Following Freeman (1977: 39), the overall centralization of a graph is defined as the average difference in centrality between the most central point and all others. We define $C_B(p_k^*)$ and as the largest
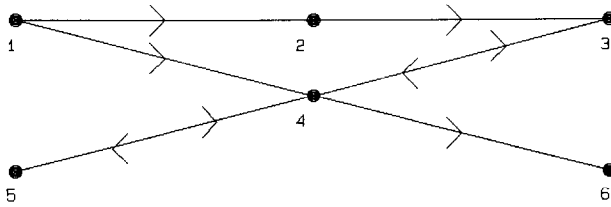
CB (p) 4= 6 is the largest betweenness

Parameters: NO = 5, NI = 5, NS = 4, n = 6

Fig. 5. Example of empirical graph.

centrality value associated with any point in the graph. Then we have as the measures of dominance of the most central point:

$$C'_B = \frac{\sum_{i=1}^{n} [C_B(p_k^*) - C_B(p_i)]}{n-1}$$

This measure of $C'B$ converges with Freeman's measure of betweenness graph centralization in the case of undirected graphs with no isolates.

6. Example

Consider the graph in Fig. 5. Here $C_B(p_2) = 1$ and $C_B(p_4) = 8.5$. This graph has the parameters $n_1 = 5$ (5 points with incoming arcs), $n_O = 5$ (five with outgoing arcs), and $n_S = 4$ (4 with both incoming and outgoing arcs). A maximally centralized graph with these parameters is shown in Fig. 6, where $C_B(p_k^*) = (n_1 - 1)(n_O - 1) - (n_S - 1) = (5 - 1)(5 - 1) - (4 - 1) = (16 - 3) = 13$. The relative centrality for the most central point is

$$C'_B(p_4) = C_B(p_4)/\max C_B(p_j) = 8.5/13 = 0.65$$

```
max CB (*) = 13 is the maximal betweenness
```
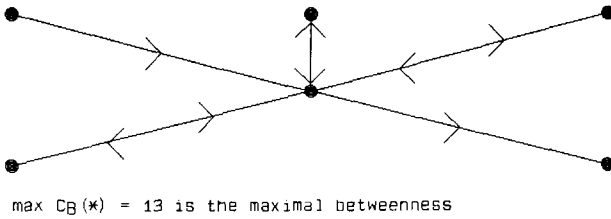
```
Parameters: NO = 5. NI = 5. NS = 4. n = 6
```

Fig. 6. Maximally central directed star corresponding to empirical graph.

The relative centralities for the other two points with betweenness centrality are

$$C_B'(p_3) = C_B(p_3)/\max\, C_B(p_j) = 3/13 = 0.23$$

$$C_B'(p_2) = C_B(p_2)/\max\, C_B(p_j) = 0.5/13 = 0.04$$

The graph centralization is

$$C_B' = \frac{\sum\limits_{j=1}^{n} [C_B(p_k^*) - C_B(p_i)]}{n-1}$$

$$= \frac{3(0.65 - 0) + (0.65 - 0.23) + (0.65 - 0.04)}{5} = 0.50$$

7. Conclusion

Development of a standardized measure of betweenness for directed graphs is an important step in the development of scientific propositions concerning social networks. Centrality is a fundamental property of actors and has numerous predictive consequences. Freeman's measures of betweenness centrality are conceptually well defined and known to provide correspondence between theoretical predictions and substantive findings about network centralities where the

social relations are symmetric, as with the Bavelas small—group experiments (Bavelas 1950; see Freeman et al. 1980). Extending them to directed graphs increases the interpretability of similar findings and the testability of theoretical propositions in the larger arenas of directed graphs and social networks with directed relations.

Acknowledgments

References

Bavelas, Alex.
  1950 "Communication patterns in task oriented groups." *Journal of the Acoustical Society of America 22*: 271–282.
Borgatti, Stephen P. and Phil Bonacich.
  1989 "The geodesic cube: A 3-dimensional object underlying standard measures of centrality." Paper presented at the International Sunbelt Social Networks Conference, Tampa, Fl.
Freeman, Linton C.
  1977 "A set of measures of centrality based on betweenness." *Sociometry 40*: 35–41.
  1979 "Centrality in social networks. 1. Conceptual clarification." *Social Networks 1*: 215–239.
  1980 "The gatekeeper, pair-dependency and structural centrality." *Quantity and Quality 14*: 585–592.
Freeman, Linton C., D. Roeder and Robert R. Mulholland.
  1980 "Centrality in social networks. II. Experimental results. *Social Networks 2*: 119–141.
Gould, Roger V.
  1987 "Measures of betweenness in nonsymmetric networks." *Social Networks 9*: 277–282.
Hage, Per and Frank Harary.
  1991 *Exchange in Oceanea: A Graph Theoretic Analysis*. Oxford: Clarendon Press.
Leavitt, Harold J.
  1951 "Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology 46*: 38–50.