This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial
Entropy as a measure of centrality in networks characterized by path-transfer flow

Frank Tutzauer *

Department of Communication, 359 Baldy Hall, University at Buffalo, Buffalo, NY 14260, USA

Abstract

Recently, Borgatti [Borgatti, S.P., 2005. Centrality and network flow. Social Networks 27, 55–71] proposed a taxonomy of centrality measures based on the way that traffic flows through the network—whether over path, geodesic, trail, or walk, and whether by means of transfer, serial duplication, or parallel duplication. Most of the extant centrality measures assume that traffic propagates via parallel duplication or, alternatively, that it travels over geodesics. Few of the other flow possibilities have centrality measures associated with them. This article proposes an entropy-based measure of centrality appropriate for traffic that propagates by transfer and flows along paths. The proposed measure can be applied to most network types, whether binary or weighted, directed or undirected, connected or disconnected. The measure is illustrated on the gang alliance network of Kennedy et al. [Kennedy, D.M., Braga, A. A., Piehl, A.M., 1998. The (un)known universe: mapping gangs and gang violence in Boston. Crime Prevention Studies 8, 219–262].

© 2006 Elsevier B.V. All rights reserved.

Keywords: Centrality; Entropy; Flow; Methods; Paths; Traffic; Transfer

1. Introduction

It is patently obvious that different measures of centrality will capture different aspects of what it means for a node to be “central” to the network. As an example, Freeman argued in his seminal paper (1978/1979) that degree centrality indexes a node’s activity, whereas betweenness centrality measures a node’s control, and closeness centrality measures its communication efficiency. Over the years, researchers have proposed a number of different measures, such as eigenvector centrality (Bonacich, 1972) and information centrality (Stephenson and Zelen, 1989), each focusing on slightly different features of central nodes.

* Tel.: +1 716 645 2141x1178; fax: +1 716 645 2086.
E-mail address: comfrank@buffalo.edu.
Until now, the various measures have been difficult to compare conceptually. In a recent Social Networks article, Borgatti (2005) created a typology of centrality measures based on the ways that traffic flows through the network. He considered not only the routes that traffic may follow, but also the method by which the traffic spreads. More specifically, he argued that traffic may flow along paths (sequences of linked nodes in which neither nodes nor links are repeated), geodesics (shortest paths), trails (which allow repetition of nodes), or walks (which allow repetition of both nodes and links). Moreover, how traffic propagates along the routes may also vary. Traffic may spread by parallel duplication, in which traffic flows from a node to its neighbors simultaneously, as in an email virus alert. Traffic may also spread by serial duplication, as in person-to-person gossip, where a node passes the gossip to other nodes one at a time. Finally, some traffic flows can be characterized as transfer processes, such as a package delivery system, where an item can be only one place at one time, and moves from node to node in the network. These propagation methods are similar to the various exchange relations in the study of exchange networks. Transfers are analogous to a special case of exclusion connections, namely negative connections, in which a node benefits from exchanging with a single network partner, but not with more than one. Parallel and serial duplication, on the other hand, are analogous to various types of inclusion and exclusion-inclusion connections in which nodes benefit from exchanging with all or many partners. (For a good introduction to the various types of exchange connections, see Willer, 1992.)

These two characteristics – the route the traffic follows (geodesics, paths, trails, or walks) and the method of propagation (parallel duplication, serial duplication, or transfer) – define a two-dimensional typology. Each measure of centrality makes assumptions about the importance of the various types of traffic flow, and thus each measure of centrality can be assessed by where it falls in the typology. For example, betweenness centrality, which indexes a node’s control by counting the number of shortest paths that pass through it, is perfect for flows involving geodesics. A node with high betweenness centrality is essentially a traffic checkpoint, and can shut down the flow. On the other hand, betweenness is inappropriate if the traffic is not constrained to follow geodesics. Non-geodesic paths avoid the checkpoints altogether, making an alternative measure essential.

Quite surprisingly, Borgatti found that almost all of the extant centrality measures possess one of two properties: either they involve parallel duplication or they emphasize the importance of geodesics. Most of the cells of the typology are empty. Borgatti found no examples of centrality measures for serial duplication via paths, trails, or walks, and none for transfers over paths or trails, meaning that “most of the sociologically interesting processes are not covered by the major measures” of centrality (p. 63).

The purpose of this paper is to begin filling the empty cells of the typology. I propose a measure of centrality for networks characterized by path-based transfer flows. Path-transfers are an important class of flow processes, and a corresponding measure of centrality is not just a logical possibility, but an important development, particularly at a time when many flows – of nuclear material, of arms, of explosive devices – probably follow path-transfers. The measure proposed in this article is based on information theory, Claude Shannon’s mathematical theory of communication (Shannon, 1948; see also Shannon and Weaver, 1964). For a flow beginning at a specific node, the centrality of that node is related to the distribution of the probabilities that the flow stops at each of the nodes in the network. It is equivalent to the information content of a communication system in which symbols are transmitted with the indicated probabilities.

The structure of the paper is as follows: after introducing mathematical fundamentals and specifying the path-transfer flow process, an entropy-based measure of centrality is introduced and
illustrated. The measure is then generalized to a variety of network types and the generalizations
are illustrated on a directed, weighted, and disconnected network. Then computational efficiency
is considered, and the measure is distinguished from Stephenson and Zelen’s (1989) information
centrality. Finally, some conclusions are offered.

2. Mathematical preliminaries

The mathematical representation of a network is a (looped, undirected) graph $G = (V,E)$, where
$V = \{1, 2, \ldots, N\}$ is a finite, nonempty set of vertices or nodes, and $E$ is a symmetric, reflexive
relation on $V$. The elements of $E$ are called edges. The edge $(i, j) \in E$ is incident with the vertices
$i$ and $j$, and $i$ and $j$ are incident with the edge $(i, j) \in E$. Moreover $(i, j) \in E$ is a link if $i \neq j$
and a loop if $i = j$. Requiring $E$ to be reflexive means that each vertex has a loop, and the symmetry
assumption means that directed graphs are excluded. Both of these assumptions will be relaxed
later.

If two vertices are incident with the same edge, then they are adjacent. Adjacent vertices are
called neighbors. Define the $N \times N$ adjacency matrix $A = [a_{ij}]$ by setting $a_{ij}$ equal to 1 if $(i, j) \in E$
and 0 if not. The row sums of the adjacency matrix give the degrees of the vertices.

A path (of length $n$) from $i$ to $j$ is an ordered sequence of distinct vertices $P = \{v_0, v_1, \ldots, v_n\}$
with $v_0 = i$, $v_n = j$, and $(v_t, v_{t+1}) \in E$ for $t = 0, 1, \ldots, n-1$. If there is no path from $i$
to $j$ of length less than $n$, then the path $P$ is a geodesic. If the requirement of distinctness
of the vertices is removed, then $P$ is a walk. A trail allows non-distinct vertices, but
requires that every ordered pair of consecutive vertices $(v_t, v_{t+1})$ appears only once in the
sequence.

A graph is connected if there is a path from each vertex to every other (distinct) vertex; otherwise
it is disconnected. Note that in the path $P$, whereas $n$ is the length of the path, $t$ is the serial position
(in the path) of vertex $v_t$, assuming the count starts at 0. To prevent cumbersome notation in the
sequel, I make the identity $v(t) = v_t$, and freely switch between the symbols depending on whether
I want to avoid multiple parentheses on the one hand or subscripts of subscripts on the other. Note
that the path $\{1,2,3\}$, say, is distinct from the path $\{3,2,1\}$. Maintaining this distinction will ensure
easy generalization to directed graphs.

3. Path-transfer flows

3.1. Description

There are many networks in which traffic flows by transference of items over paths. Borgatti
nicknamed such flows “mooch” processes. He did not give an example, but it is easy to imagine
one. Consider a group of people linked by friendship ties, and suppose one of them has a magazine.
When done reading the magazine, one of this person’s friends borrows – mooches – the magazine.
Of course, to continue the flow, one of the friend’s friends mooches the magazine, and then one
of that friend’s friends, and so forth. As various friends mooch the magazine, it follows paths
(because once one has read the magazine, one does not reread it). It is also physically transferred
from one person to another. The flow stops when someone elects to keep the magazine or has no
friend wanting to read it.

Another example of a path-transfer process is the sending of a chain letter. At each step of
the flow, the current holder of the letter has a choice of either breaking the chain or continuing
it by putting his or her name at the bottom of the letter and sending it to a friend. Note
that the flow process of the classic chain letter differs from an email chain letter, which propagates via parallel duplication not transfer (the email chain letter gets sent at once to everyone in an address book). In the classic case, the letter is physically transferred from one node to another, and it follows paths. The letter is not “going” anywhere – it has no specific destination – so it does not necessarily follow geodesics. Moreover, it does not follow trails or walks because each prior recipient’s name is at the bottom of the letter. The flow stops only when a node decides to break the chain or when the node has no friends to whom the letter can be forwarded.

There are many other examples of path-transfer flows. Trading and smuggling networks (Mackenzie, 2002) often involve path-transfers, particularly if the traded or smuggled commodity is discrete. Possibilities include exotic animals (Warchol, 2004), nuclear weapons material and parts (Broad et al., 2004), fossils (Yimin and Stokstad, 2002), artworks and antiquities (Steiner, 1994), and even trafficking in humans (Surtees, 2005).

To model a path-transfer process, it is helpful to think of a specific object being passed from one node to another. The flow originates at a particular vertex (the one whose centrality is under consideration), and one of the node’s neighbors (including the node itself) is randomly chosen. If the starting node is chosen, the flow is over before it begins; otherwise, the object passes to the chosen node. The next node then randomly chooses from among its neighbors (including itself, but not including the starting node), and again the flow either stops (if the node itself is chosen) or continues (if a different node is chosen). At each step, a choice is made from among eligible neighbors—that is, including the node itself, but excluding neighbors that have been chosen earlier. The object thus traverses a path in the network, traveling along links, stopping when a loop is chosen. At this point in the analysis, each of the eligible neighbors is assumed to be chosen with equal likelihood; preferential choices, which require the use of a weighted graph, are considered in a later section of the article.

3.2. Centrality

What properties should a measure of centrality possess if it is to describe nodes in a path-transfer network? Thinking of the network of magazine-sharing friends, it is difficult to predict where the magazine’s journey will end if it starts at a highly central node. Because the node is central, the magazine might ultimately end up virtually anywhere in the network. In contrast, with a less central node it is easier to predict the ultimate destination of the magazine—there are a few nodes that, with high likelihood, will ultimately receive the magazine, whereas for many of the other nodes in the network the probability of ending up with the magazine is close to zero. In fact if the node is so peripheral as to have no friends whatsoever, the situation is perfectly predictable—the magazine travels nowhere.

The key feature here is that a flow beginning at a highly central node will stop with nearly equal probability at all other nodes, and a flow beginning at a less central node will have a much more uneven distribution of probabilities. This idea can be more easily understood if one considers an isolate, at one extreme, and the center node of a star, at the other extreme. Recall that an isolate is a vertex with no incident links, and a star is a graph in which one vertex – the hub – is adjacent to every vertex in the network, and the other vertices are linked to the hub but not to each other. Because the hub of a star is adjacent to each of the other vertices as well as itself, it thus has degree \( N \), and, following the flow process outlined above, it is clear that for a flow starting at the hub, the probability that it ends at any given vertex is \( 1/N \). Thus, a flow beginning at the hub is equally likely to stop at any vertex in the network. At the other extreme, a flow beginning at an isolate
has nowhere to go, and thus the probability that the flow ends at the isolate is unity, whereas the probability that the flow ends elsewhere is zero. In this context, equal probabilities translate to high centrality. If the probabilities are nearly equal, the centrality will be high, becoming maximum when they are exactly equal. As the probabilities become more unequal, centrality drops, reaching its minimum when one probability is unity and the rest are zero. Thus, the first property the proposed measure of centrality should have is that of equiprobability: the more nearly equal the likelihood that the flow stops at the various vertices in the network, the higher the centrality.

Moreover, if all the probabilities are the same then centrality should grow with network size. For example, the hub of a star with 20 vertices should receive a higher centrality score than the hub of a star with five vertices. This property amounts to assuming monotonicity: for situations in which the probabilities are all equal to $1/N$, the centrality measure should be a monotonic increasing function of $N$.

Those familiar with communication theory will recognize the existence of a measure in a different context that possesses analogs of these properties, namely Shannon’s (1948) measure of information (see also Shannon and Weaver, 1964). Given a communication system with $N$ symbols, in which the $i$th symbol is transmitted with probability $p_i$, the information of the system is defined to be

$$H = -\sum_{i=1}^{N} p_i \log p_i.$$ 

If the base of the logarithm is chosen to be 2, then the unit of information is the bit. Informally, a bit is the amount of information in one yes–no question. The rationale for this rule of thumb is that if we have two symbols (A and B), each transmitted with equal probability, it takes only one yes-no question to determine the transmitted symbol. (Ask, “Is it A?” Then “Yes” implies that A was transmitted, and “No” implies that B was transmitted.) In such a situation, recognizing that $\log 1/2 = -1$, the above becomes $H = -(1/2 \log 1/2 + 1/2 \log 1/2) = 1/2 + 1/2 = 1$. Similarly, if there are four, eight, or sixteen equally likely symbols to transmit, then it would take, respectively, two, three, or four questions (each question eliminating half the alternatives) to determine which symbol was sent.

Information is also a measure of the uncertainty, or entropy, in a system—roughly, the amount of randomness or freedom of choice. When choices are highly constrained, as when it is very likely that certain symbols will be sent (or, in the present application, when a flow is highly likely to end at a just a few nodes), then information (and centrality) is low. When there is much freedom of choice, as when all symbols are equally likely to be sent (or a flow is equally likely to stop anywhere), then information (and centrality) is high.

Regardless of logarithmic base, the information, $H$, increases as the probabilities become equal, becoming maximum when all the $p_i$ are exactly equal. Moreover, when all the $p_i$ are equal, $H$ grows with increasing $N$. As mentioned, information is a measure of entropy, and I will use this name for the present measure of centrality so as not to confuse it with Stephenson and Zelen’s (1989) information centrality (which is not based on Shannon’s information-theoretic measure of uncertainty).

To summarize, we assume that the flow begins at a node $i$ and transfers from node to node along paths. Using the traffic-flow process described above, one computes the probabilities that the flow stops at each node in the network. The entropy of these probabilities is then taken as node $i$’s centrality.
4. An entropy-based measure of centrality

Fig. 1 shows the main component of Boston’s gang alliance network (Kennedy et al., 1998). These data were collected as part of the Boston Gun Project (National Institute of Justice, 2001), and the alliances are based on the perceptions of Boston police officers, probation officers, and streetworkers. These practitioners knew a great deal about the Boston gang problem, and there was “strong agreement” (Kennedy et al., 1998, p. 234) as to the nature of the alliances between the gangs.

It is reasonable to think that the flow of small arms, such as handguns, is a path-transfer process following gang alliance ties. At least a portion of the gun trafficking in Boston involved “the illegal sale and bartering of guns among youths themselves” (Kennedy et al., 2001, p. 7). The gang alliance network has the additional advantage of being very sparse, making it ideal for illustrating the key ideas involved in an entropy-based centrality measure.

For the sake of this section, I assume a connected, looped, undirected network represented by a graph $G = (V,E)$, with $V$ being a set of $N$ vertices numbered 1 through $N$, and with $E$ being a symmetric, reflexive relation on $V$. Consistent with these assumptions, Fig. 1 shows only the main component of the gang alliance network and I will treat it as though it were the entire network (i.e., ignoring the isolates, isolated dyads, and a lone isolated triad). Moreover, to avoid cluttering the figure, the loops, though assumed to be present, are not depicted, and the distinct edges $(i, j)$ and $(j, i)$ are depicted as a single line between $i$ and $j$.

Because the flow process is such that a transfer is made to nodes not previously chosen, it will be useful to develop some terminology. For a given path $P = \{v_0, v_1, \ldots, v_n\}$, a downstream edge of a vertex $v_t \in P$ is any edge $(v_t, w) \in E$ with $w \neq v_s$ for $s < t$. A downstream vertex of $v_t$ is any vertex $w$ such that $(v_t, w)$ is a downstream edge. The downstream degree, $D(v_t)$, of a vertex is the number of downstream edges it has and can be calculated from the adjacency matrix as:

$$D(v_t) = \sum_{j=1}^{N} a_{v(t),j} - \sum_{s<t} a_{v(t),v(s)}.$$ (1)
The first term on the right-hand side of Eq. (1) counts the total number of edges incident with \( v_t \) (i.e., its degree), and the second term subtracts from that total the number of edges incident with vertices appearing earlier in the path, hence giving the number of downstream edges incident with vertex \( v_t \).

At any point in the flow, a node receiving the transfer can pass the object to any node to which it is adjacent that does not appear on the path up to that point, that is to any downstream vertex. The downstream degree thus gives the number of eligible vertices to which the transfer can be next made. For example, in Fig. 1, consider a flow beginning at vertex 6, traversing the path \{6,7,10\}, and thus ending at vertex 10. At the beginning of the flow, vertex 6 has three choices. The object can be passed to vertices 5 or 7, or vertex 6 can stop the flow. There are thus three downstream edges – (6,5), (6,7), and (6,6) – and vertex 6’s downstream degree is equal to its ordinary degree of 3 (remember, loops are not shown in the figure). Suppose transfer is passed to vertex 7. Its downstream degree is also 3 (because the flow can continue to vertices 8 or 10, or stop at 7), one less than its ordinary degree because for \( v_t = 7 \) there is just one adjacent vertex, \( v_s = 6 \), with \( s < t \). Assume that the flow then passes to vertex 10, and note finally that its downstream degree is 4, because transfer can pass to vertices 8, 9, 11, or can stop at 10.

Observe that the downstream degree depends on the path followed. For example, the downstream degree of vertex 10 in the path \{6,7,10\} is equal to 4. But in the path \{6,7,8,9,10\} vertex 10’s downstream degree drops from 4 to 2 (because 8 and 9 are eliminated as possibilities—they are “upstream,” i.e., they have already been chosen). However, the particular path is usually apparent, so to avoid cluttering the notation the dependence will not be made explicit in the symbols.

The downstream degrees allow calculation of the probability that a flow originating at one vertex stops at any particular vertex of the network. For example, in Fig. 1, what is the probability that a flow beginning at vertex 5 ends at vertex 2? At this stage of the analysis a node’s choices are made with equal likelihood. Consequently there is a 1/4 probability that vertex 5 passes control to vertex 3 (because 5 chooses with equal likelihood from among nodes 3, 4, 5, and 6). Once node 3 receives control, the flow will not pass back to node 5, so there is a 1/2 probability that node 3 stops the flow, and a 1/2 probability that control passes to node 2. Likewise, vertex 2 chooses between stopping the flow or continuing it to vertex 1. As a result, the probability that a flow beginning at vertex 5 ends at vertex 2 is \((1/4)(1/2)(1/2) = 1/16\). The denominators are easily seen to be the downstream degrees. Stated differently, at any step in the process, the probability that transfer is passed to a downstream node is equal to 1 divided by the downstream degree. The probability that a node stops the flow is likewise equal to 1 divided by the downstream degree.

In general, there are \( K(i, j) \) paths from \( i \) to \( j \). Let \( P_k \) be such a path, and let it be of length \( n_k = n(k) \). The transfer probability of a vertex \( v_t \in P_k \) is given by:

\[
\tau_k(v_t) = \frac{1}{D(v_t)},
\]

and the stopping probability is

\[
\sigma_k(v_t) = \frac{1}{D(v_t)}.
\]

At this stage of the analysis the stopping probability and the transfer probability are defined by the same formula, but I give them different symbols in anticipation of generalization.
To obtain the single path probability – i.e., the likelihood that a flow beginning at \( i = v_0 \) ends at \( j = v(n_k) \) by traveling along the path \( P_k = \{v_0, \ldots, v(n_k)\} \) – simply multiply by the transfer probability of each of the first 0, \ldots, \( n_k-1 \) vertices and the stopping probability of the last vertex in the path. Then the overall probability that a flow starting at \( i \) ends at \( j \) is given by the combined path probability, which is simply the single path probabilities summed across the \( K(i, j) \) paths from \( i \) to \( j \):

\[
p_{ij} = \sum_{k=1}^{K(i,j)} \sigma_k(j) \prod_{t=0}^{n(k)-1} \tau_k(v_t).
\]

(4)

For looped, unweighted graphs, the stopping and transfer probabilities of vertex \( j \) are the same, so the above can be simplified to

\[
p_{ij} = \sum_{k=1}^{K(i,j)n(k)} \prod_{t=0}^{n(k)-1} \tau_k(v_t).
\]

(5)

The path-transfer centrality of vertex \( i \) is then given by the entropy

\[
C_H(i) = -\sum_{j=1}^{N} p_{ij} \log p_{ij}.
\]

(6)

where the symbol \( C_H(i) \) is chosen to be consistent with Freeman’s (1978/1979) notation. Eq. (6) is analogous to the Shannon information equation except that the probabilities, rather than being probabilities of transmission, are instead the probabilities that a path-transfer flow originating at node \( i \) stops at each of the nodes \( j = 1, 2, \ldots, N \). For a given \( N \), the function reaches its maximum when all the \( p_{ij} \) are equal, and for equal \( p_{ij} \) the function is an increasing function of \( N \). To put this centrality score on a zero-one scale, simply divide by the maximum entropy, which is well known to be \( \log(N) \) (Shannon, 1948). Hence the relative centrality of vertex \( i \) is given by

\[
C'_H(i) = \frac{C_H(i)}{\log N}.
\]

(7)

Table 1 shows the entropy centrality of each of the nodes in the Boston gang alliance network along with Freeman’s (1978/1979) measures for comparison,\(^1\) and Table 2 show the correlations among the measures. Differences between centrality measures are often discussed in terms of the way that the nodes are ranked by the measures, so Spearman’s rank-order correlation, \( \rho \), is shown in the table. But, if one is using centrality as a variable in another analysis, the actual centrality score, rather than simply the ranking, may be of interest, so Pearson’s \( r \) is shown in the table as well. (For computation and limitations of \( r \) and \( \rho \), see, e.g., Thompson, 2006.)

As would be expected, the measures are highly, but not perfectly correlated. Degree centrality, of course, provides the least discriminative ability with many ties in the ranking of the nodes. Its ranking is also least like the other measures. Consequently, its correlations with other measures are the lowest in Table 2. Betweenness centrality does not distinguish among the nodes that are

\(^1\) Software for calculating entropy in moderately sized networks is available from the author on request. All other centralities were computed using UCINET 6 (Borgatti et al., 2002).
Table 1
Centrality statistics for the gang alliance network

<table>
<thead>
<tr>
<th>Node</th>
<th>Entropy</th>
<th>Degree</th>
<th>Betweenness</th>
<th>Closeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3.028</td>
<td>2 (5–8)</td>
<td>25 (2)</td>
<td>23 (1)</td>
</tr>
<tr>
<td>7</td>
<td>3.006</td>
<td>3 (2–4)</td>
<td>24 (3)</td>
<td>24 (2–3)</td>
</tr>
<tr>
<td>5</td>
<td>2.906</td>
<td>3 (2–4)</td>
<td>27 (1)</td>
<td>24 (2–3)</td>
</tr>
<tr>
<td>10</td>
<td>2.809</td>
<td>4 (1)</td>
<td>12.5 (5)</td>
<td>28 (4)</td>
</tr>
<tr>
<td>8</td>
<td>2.806</td>
<td>3 (2–4)</td>
<td>3.5 (7)</td>
<td>29 (5–6)</td>
</tr>
<tr>
<td>3</td>
<td>2.683</td>
<td>2 (5–8)</td>
<td>16.0 (4)</td>
<td>29 (5–6)</td>
</tr>
<tr>
<td>9</td>
<td>2.658</td>
<td>2 (5–8)</td>
<td>0.0 (8–11)</td>
<td>36 (8–9)</td>
</tr>
<tr>
<td>4</td>
<td>2.396</td>
<td>1 (9–11)</td>
<td>0.0 (8–11)</td>
<td>33 (7)</td>
</tr>
<tr>
<td>11</td>
<td>2.304</td>
<td>1 (9–11)</td>
<td>0.0 (8–11)</td>
<td>37 (10)</td>
</tr>
<tr>
<td>2</td>
<td>2.300</td>
<td>2 (5–8)</td>
<td>9.0 (6)</td>
<td>36 (8–9)</td>
</tr>
<tr>
<td>1</td>
<td>2.037</td>
<td>1 (9–11)</td>
<td>0.0 (8–11)</td>
<td>45 (11)</td>
</tr>
</tbody>
</table>

Note. Numbers in parentheses are ranks. Betweenness centrality counts the number of geodesics passing through each vertex (appropriately weighted if there are multiple geodesics). Closeness is the sum of the distances from each node to every other node. Smaller numbers indicate more centrality. Entropy is computed on the basis of the looped network. Loops are irrelevant to the betweenness and closeness measures. Degree centrality is computed on the unlooped network. If loops had been counted, each degree would have increased by one, but the rank ordering would have been unchanged.

on no geodesics (such as nodes 9 and 11), whereas entropy and closeness centrality do. Entropy, betweenness, and closeness all agree on the top three nodes, though they rank them differently. Perhaps the biggest discrepancy among the measures occurs with node 2 which, being an interior node on nine geodesics, is ranked sixth on the betweenness measure but only tenth on the entropy measure and is placed in a tie for eighth/ninth by the closeness measure. Closeness produces the rankings most similar to the rankings of the entropy centrality. Such a result is to be expected because in the network of Fig. 1 probability is inversely correlated with distance. If a node has a high closeness score (i.e., low centrality), it will be near some nodes and far away from others, giving an uneven distribution of probabilities and thus a small entropy, hence the very large correlations between entropy and closeness for this network.

Table 2
Correlation matrices of centrality measures for the Boston gang network

<table>
<thead>
<tr>
<th></th>
<th>Entropy</th>
<th>Degree</th>
<th>Between</th>
<th>Close</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman’s ρ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>–</td>
<td>.702</td>
<td>.803</td>
<td>.966</td>
</tr>
<tr>
<td>Degree</td>
<td>–</td>
<td>–</td>
<td>.648</td>
<td>.690</td>
</tr>
<tr>
<td>Betweenness</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>.883</td>
</tr>
<tr>
<td>Closeness</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Pearson’s r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>–</td>
<td>.730</td>
<td>.762</td>
<td>.944</td>
</tr>
<tr>
<td>Degree</td>
<td>–</td>
<td>–</td>
<td>.544</td>
<td>.666</td>
</tr>
<tr>
<td>Betweenness</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>.834</td>
</tr>
<tr>
<td>Closeness</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note. Correlations with closeness are actually negative because closeness centrality is indexed with a sum of distances, hence the higher the score the lower the centrality. Negative signs have been omitted in the table. Moreover, as the observations making up the correlations are not independent, significance tests are not reported.
5. Generalizations

5.1.Disconnected networks

In a disconnected network, one can find two vertices $i$ and $j$ with no path from $i$ to $j$. In this case, the probability of the flow beginning at $i$ and ending at $j$ is zero, which is problematic because the log of zero is undefined. In information theory, symbols with a transmission probability of zero contribute nothing to the information sum. For example, if a communication system contains exactly two symbols, each transmitted with equal likelihood, then there is one bit of information. If there are three symbols, the first two of which are transmitted with probability $1/2$ and the third of which is transmitted with a probability of zero, then the information is likewise one bit. The zero-probability symbol added nothing to the information. Similarly, for disconnected graphs, in Eq. (6) the sum is to be taken across all nonzero combined path probabilities. This idea can be more precisely expressed as

$$ C_H(i) = - \sum_{j \in V} p_{ij} \log p_{ij}. $$

(8)

5.2. Nonsymmetric relations

If the edge relation is nonsymmetric, then the mathematical representation is a directed graph. Here, the row sums of the adjacency matrix are generally called the out-degrees; column sums are the in-degrees. Links are typically referred to as arcs. Paths, downstream edges and vertices, and the downstream degree were all defined keeping the direction of the edges into account, so all of the formulae for undirected graphs carry over for directed graphs.

5.3. Nonreflexive relations

A nonreflexive edge relation admits loopless vertices. Usually, reflexivity is assumed even if the underlying network is not looped. For example, one is not usually considered one’s own friend, so a friendship network would be loopless. But if a chain letter is transferred via friendship ties, people receiving the letter nonetheless have the option of breaking the chain. Each person’s ability to stop the flow is modeled by putting loops in the graph representing the network. The flow stops only when a loop is followed.

Occasionally, however, one wants to examine a path-transfer process in which nodes cannot elect to stop the flow. Consider, for example, the transfer of leadership in an organization. If the current leader appoints the successor, and if leaders are restricted to non-repeating terms, then the appropriate mathematical representation is a loopless graph. In a looped graph, the flow stops when a loop is followed, but if some of the vertices are loopless, then the flow stops either when a loop is followed or when a loopless vertex has no downstream edges. Alternatively, if the flow comes to a loopless vertex with downstream edges, the stopping probability is zero. The appropriate transfer and stopping probabilities can be written:

$$ \tau_k(v_t) = \begin{cases} 
0, & D(v_t) = 0 \\
\frac{a_{x(t),y(t+1)}}{D(v_t)}, & D(v_t) \neq 0
\end{cases} $$

(9)
and

\[ \sigma_k(v_t) = \begin{cases} 
1, & D(v_t) = 0 \\
\frac{a_{v(t),v(t)}}{D(v_t)}, & D(v_t) \neq 0 
\end{cases} \]  

(10)

for the transfer and stopping probabilities respectively. Observe that the numerators are zero or one according to the appropriate entry in the adjacency matrix. For looped graphs, the above equations reduce to Eqs. (2) and (3). Whether looped or not, the above quantities can be used in Eq. (4) to obtain the combined path probabilities, from which entropy (Eq. (8)) and relative entropy (Eq. (7)) are calculated.

5.4. Preferential choices

Up to this point, each vertex has chosen from among its downstream neighbors with equal probability. Perhaps, however, a close friend is chosen with a higher probability than an acquaintance. Such preferential choosing can be represented by a weighted graph in which each edge has associated with it a positive, real number, or weight. The weights represent the strengths of the ties, and choices are made in proportion to the weight placed on the tie. For example, if a vertex has two downstream neighbors, the first of whom is linked to the vertex with a weight of 6 and the second of whom has a weight of 2, then the probability of the flow transferring to the first neighbor is $6/(6+2) = 3/4$ and the probability of transferring to the second is $2/(6+2) = 1/4$. Using the convention that the adjacency matrix $A$ contains not zeroes or ones (according to whether or not the vertices are adjacent), but instead contains the weights of the edges from each $i$ to each $j$, then Eqs. (1), (9) and (10) carry through directly. In this case the downstream “degree,” $D(v_t)$, is not the number of downstream neighbors but instead is the sum of the weights of the downstream edges.

For convenience, the relevant formulae (Eqs. (1) through (5) and Eqs. (7) through (10)) are summarized in Table 3.

6. A more general example

To show the flexibility of entropy as a measure of centrality, I modified the gang alliance network of Fig. 1. First, I added a second component so that the network is disconnected (see Fig. 2). Second, I turned many of the undirected ties into directed ties. Finally, I assigned weights to the arcs. The weights can be thought of as the strength of the alliance on a scale of 1 to 10, with 10 being the strongest. I arbitrarily assigned the loops a weight of 4, which is just over the mean tie strength of 3.56.

Other than entropy, none of the measures of centrality discussed in this article are completely appropriate for this new network. The disconnected nature of the network creates infinite distances, which makes both closeness and information centrality undefined. Betweenness centrality ignores the weights. Degree centrality can be computed, but one must distinguish between in- and out-degrees. Table 4 shows the in- and out-degree centrality scores as well as the entropy scores for the nodes in Fig. 2.

As is to be expected, the centrality rankings for in- and out-degree bear no resemblance to the entropy rankings. The degree measures look only at what happens one step away, whereas the entropy measure looks at the pattern of the probabilities across all vertices. It is also clear that the network modifications significantly changed the entropies from the original network.
Table 3
Formule for calculating entropy-based centrality

<table>
<thead>
<tr>
<th>General formulae</th>
<th>Simplified formulae for looped, unweighted graphs</th>
<th>Centrality and relative centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D(v_t) = \sum_{j=1}^{N} a_{v(t),j} - \sum_{s&lt;t} a_{v(t),v(s)}$</td>
<td>$\sigma_k(v_t) = \frac{1}{D(v_t)}$, $D(v_t) = 0$</td>
<td>$C_H(i) = \sum_{j \in V} p_{ij} \log p_{ij}$</td>
</tr>
<tr>
<td>$\tau_k(v_t) = \begin{cases} 0, &amp; D(v_t) = 0 \ \frac{a_{v(t),v(t+1)}}{D(v_t)}, &amp; D(v_t) \neq 0 \end{cases}$</td>
<td>$p_{ij} = \sum_{k=1}^{K(i,j)} n(k-1)$</td>
<td></td>
</tr>
<tr>
<td>$p_{ij} = \prod_{t=0}^{n-1} \tau_k(v_t)$</td>
<td>$\tau_k(v_t) = \frac{K(i,j)}{n(k-1)}$</td>
<td>$C_H'(i) = \frac{C_H(i)}{\log N}$</td>
</tr>
</tbody>
</table>

**Note.** $V$ is the vertex set, and $N$ is the number of vertices. $A = [a_{ij}]$ is the adjacency matrix for unweighted graphs and the matrix of weights for weighted graphs. $D$, $\tau$, and $\sigma$ are, respectively, the (weighted) downstream degree, and the transfer and stopping probabilities, each of which depends on the length-$n(k)$ path $P_k$. The quantity $p_{ij}$ is the combined path probability formed by summing across the $K(i,j)$ paths from $i$ to $j$. $C_H$ and $C_H'$ stand for centrality and relative centrality, respectively.

change is that vertex 6, which was the most central in the original network, drops into a tie for last. The reason, of course, is that it has no outgoing arcs, so any flow beginning at vertex 6 stays there, and the entropy is zero. Two other nodes – 4 and 11 – similarly have no outgoing arcs, so their entropies are zero as well.

---

Fig. 2. Modified gang alliance network. For entropy calculations, loops are assumed but not shown. Loop weight equals 4 for all loops. This network is hypothetical.
Table 4
Centrality statistics for modified (hypothetical) gang alliance network

<table>
<thead>
<tr>
<th>Node</th>
<th>Entropy</th>
<th>In-degree</th>
<th>Out-degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.172 (1)</td>
<td>2(10–13)</td>
<td>5(4)</td>
</tr>
<tr>
<td>7</td>
<td>2.154 (2)</td>
<td>5(4–6)</td>
<td>8(3)</td>
</tr>
<tr>
<td>8</td>
<td>2.082 (3)</td>
<td>4(7–8)</td>
<td>4(5–7)</td>
</tr>
<tr>
<td>10</td>
<td>2.059 (4)</td>
<td>5(4–6)</td>
<td>13(2)</td>
</tr>
<tr>
<td>2</td>
<td>1.855 (5)</td>
<td>4(7–8)</td>
<td>4(5–7)</td>
</tr>
<tr>
<td>15</td>
<td>1.623 (6)</td>
<td>0(15)</td>
<td>4(5–7)</td>
</tr>
<tr>
<td>9</td>
<td>1.581 (7)</td>
<td>1(14)</td>
<td>2(9–10)</td>
</tr>
<tr>
<td>5</td>
<td>1.510 (8)</td>
<td>2(10–13)</td>
<td>17(1)</td>
</tr>
<tr>
<td>12</td>
<td>1.295 (9)</td>
<td>2(10–13)</td>
<td>3(8)</td>
</tr>
<tr>
<td>14</td>
<td>1.247 (10)</td>
<td>5(4–6)</td>
<td>2(9–10)</td>
</tr>
<tr>
<td>1</td>
<td>1.000 (11)</td>
<td>2(10–13)</td>
<td>1(11–12)</td>
</tr>
<tr>
<td>13</td>
<td>0.906 (12)</td>
<td>3(9)</td>
<td>1(11–12)</td>
</tr>
<tr>
<td>4</td>
<td>0.000 (13–15)</td>
<td>8(2–3)</td>
<td>0(13–15)</td>
</tr>
<tr>
<td>6</td>
<td>0.000 (13–15)</td>
<td>13(1)</td>
<td>0(13–15)</td>
</tr>
<tr>
<td>11</td>
<td>0.000 (13–15)</td>
<td>8(2–3)</td>
<td>0(13–15)</td>
</tr>
</tbody>
</table>

Note. Numbers in parentheses are ranks. Entropy is computed on the basis of the looped network. Degree centralities are computed on the unlooped network. If loops had been counted, each in- and out-degree would have increased by 4, but the rank ordering would have been unchanged.

The directionality of the arcs coupled with the weights means that some of the nodes remain about as central as before—for example nodes 7 and 10 keep their second and fourth place spots—whereas others are different—for example node 3 moves from sixth place to first. Six nodes are reachable from node 3, which is tied for most in the network, and the weights are such that the probabilities are more evenly distributed than, say, node 7.

7. Computational issues

7.1. Computational efficiency

Although entropy-based centrality has a fairly straightforward formula, one practical difficulty is that computing centrality on the basis of the formula requires finding all paths emanating from the node in question. And if one wants to compute entropy for every node in the network, every path in the underlying graph must be found. For large and dense networks, the total number of paths can be exceedingly large. Fortunately, there are a number of methods available for finding all paths in a graph, including algorithms based on subgraphs (Misra and Misra, 1980), spanning trees (Misra, 1979), and Petri nets (Hura, 1983). Fratta and Montanari (1975) have a very interesting approach. They developed a path algebra that allows one to reduce the problem of finding all paths in a graph into the problem of solving a system of linear equations. If matrix techniques are used to solve the equations, the computational complexity is \(O(n^4)\), and it is \(O(n^3)\) if Gaussian elimination is used (Fratta and Montanari, 1975).

Perhaps the best method—for entropy calculations at least—is employing a simple, depth-first search tree (Migliore et al., 1990). Strategic pruning of the tree can considerably shorten the search. If the network is small or sparse, there are few enough paths that an exhaustive search can be done, but in a large, dense network an exhaustive search may be computationally prohibitive.
Fortunately, an exhaustive search is not needed. Consider that in a large, dense network, there will be very many paths that pass through vertices of high degree. Also, many of the paths will be very long. Both of these factors drive down the path probability. If a path passes through a vertex of high degree, the transition probability will be small (because one is dividing by a large number). Similarly, if the path is very long, then the path probability will be small because many multiplications are performed. And although such paths will not have associated probabilities that are precisely zero, they nonetheless might be zero to many digits past the decimal. The irony of the situation is that in an exhaustive search one spends an inordinate amount of computing time finding paths that contribute virtually nothing to the final entropy values.

The solution is to compute the path probability as the path is being found in the depth-first search. If at some point, the simple path probability falls below a researcher-defined threshold, then the search tree is pruned at this point. Extending the path by adding vertices would only lower the already close-to-zero path probability, so paths deeper in the search tree need not be examined. Using such a pruning strategy, it is possible to compute entropy to five significant digits, processing on the order of 9200 paths per second, and turning a problem that might normally take hours to complete into one that takes minutes.²

7.2. Entropy or information?

Based as it is on Shannon’s (1948) information theory, entropy centrality might be confused with Stephenson and Zelen’s (1989) information centrality, which is likewise said to be based on information theory. The confusion stems from two different uses of the term information theory. The first, of course, is Shannon’s mathematical theory of communication; the second use is in statistical estimation and the design of experiments. Entropy centrality is based on the former, Stephenson and Zelen’s measure on the latter.

When one examines the computational details, the measures are clearly seen to be different. For Stephenson and Zelen, “information” is inversely proportional to the distance traveled, which is equivalent to the variance of an observation from a normal distribution. More specifically, for a pair of vertices, Stephenson and Zelen’s information is calculated to be the appropriately summed reciprocals of the lengths of the paths between the vertices. The information centrality is then taken to be the reciprocal of the mean information. The key quantities are the path lengths. Entropy, in contrast, takes not only the path lengths into account but also the probabilities with which the individual edges are traversed, and these probabilities in turn are based on the downstream degrees of the vertices in the path. Consider, for example, paths \{6,5,3\} and \{3,2,1\} in Fig. 1.

As both of these paths are of length 2, Stephenson and Zelen’s measure treats them as having identical information content, and thus they contribute equally to the centralities of vertices 6 and 3. But the entropy measure treats these two paths differently. In the path \{6,5,3\}, vertex 6 has three downstream edges, as does vertex 5, and vertex 3 has only two, so the path probability is \((1/3)(1/3)(1/2) = 1/18\). In contrast, for the path \{3,2,1\}, vertex 3 has three downstream edges, vertex 2 has two, and vertex 1 has one, so the path probability is \((1/3)(1/2)(1) = 1/6\), and thus these paths contribute differentially to the centralities of vertices 6 and 3. In practice, very sparse networks such as in Fig. 1 give rise to probabilities that are highly negatively correlated with path length, so information centrality and entropy centrality produce similar rankings. In more complex networks, the two measures will differ on which vertices are more central.

² I am indebted to Ben Elbirt for providing these estimates. Tests were done on a 3.4 GHz AMD Athlon 64 Processor 3400+ with 3 gigabytes of RAM.
The mathematical differences between the two measures derive from the different assumptions made about traffic flow. Whereas entropy assumes a path-transfer flow, Borgatti (2005) puts the Stephenson and Zelen measure in the “walk-parallel duplication” cell of the typology. Stephenson and Zelen explicitly assume that the “signal at \( i \) is transmitted to all of the” (p. 8) neighbors of \( i \) and thus the flow process is best characterized by parallel duplication. In contrast, entropy assumes that only one neighbor of \( i \) continues the flow and thus transference is more appropriate.

Finally, there are differences in the types of networks the two measures can handle. Like entropy, Stephenson and Zelen’s information centrality can cope with weighting. Unlike entropy, information centrality is “applied to nondirected networks” (pp. 26–27). Moreover, Stephenson and Zelen explicitly assume that “all points are reachable” (p. 30), and thus information centrality requires connected networks. In contrast, entropy can be computed for both directed and disconnected networks.

8. Conclusion

Quite often, a researcher will choose a centrality measure without regard for the assumptions made about the underlying flow process. All measures of centrality are correlated with each other to some degree, so one might ask why it matters. The answer of course, is that the closer our measures match the underlying network processes, the better our data and theories. This point is made forcefully by Borgatti (2005, p. 56):

What happens when we apply a [centrality] measure that assumes a given set of flow characteristics to a flow with different characteristics? One of two things must happen: either we lose the ability to fully interpret the measure (as when we compute the mean of a nominal-scaled variable) or we get poor answers (as when we use linear regression to predict values of a dependent variable when the relationship is actually non-linear).

There may also be practical implications, as when the wrong node in a criminal network is targeted or overlooked. In the gang alliance network, for example, node 9 has a betweenness centrality of zero, as it is not an interior node on any geodesics whatsoever. If the flow process indeed follows geodesics, then one might reasonably ignore this node as being unimportant. But for a transfer process process that follows paths, the node should definitely not be overlooked. It has an entropy of 2.658, which places the node seventh, and is in fact not much different from other, similarly ranked nodes.

The contribution of Borgatti’s (2005) article is that he gives us a language to discuss such matters. As researchers create centrality measures to populate the empty cells of his typology, we approach the day when one can match the centrality measure to the underlying network process by simply choosing the appropriate measure from the typology. Towards this end, the present article has presented a measure of centrality based on Shannon’s (1948) information-theoretic measure of uncertainty. The measure is specifically designed for networks in which traffic moves along paths and does so by a process of transference, rather than duplication. For a given node, the probabilities are computed that the flow stops at the various nodes in the network, and the entropy of these probabilities is taken as the measure of centrality. Entropy has the additional advantage that it can be computed on virtually any type of network, whether connected or disconnected, weighted or unweighted. About the only types of networks for which entropy cannot be computed are networks with negative weights and networks in which the arcs are labeled with nonnumeric information, such as edge-colored networks or signed networks.
Entropy, in fact, can provide a unifying centrality framework for many, if not all, flow processes. Assessing centrality by measuring the distribution of the probabilities is an extremely general idea: highly central nodes will have a very even distribution of probabilities whereas nodes low in centrality will have a very uneven distribution of probabilities. This conceptualization does not depend on the underlying flow process. What does depend on the flow, however, is how the probabilities are computed. Whereas a path-transfer flow stops at the nodes with the probabilities indicated in Section 4, if traffic instead were transferred over walks, the flow would be a Markov process. Transfers that take place over geodesics or trails would have different probabilities as well. Once the underlying probabilities are determined, it is a simple matter to compute the entropy of the probabilities, giving a centrality measure for the desired flow process.

Acknowledgements

An earlier version of this paper was presented at Sunbelt XXVI, the annual meeting of the International Network for Social Network Analysis, Vancouver, BC, Canada, April 2006. I would like to thank Ben Elbirt, Tom Feeley, and Carol Tutzauer for their insightful comments on earlier drafts.

References