



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Social Networks xxx (2005) xxx–xxx

SOCIAL
NETWORKSwww.elsevier.com/locate/socnet

A Graph-theoretic perspective on centrality[☆]

Stephen P. Borgatti^{a,*}, Martin G. Everett^{b,1}

^a *Department of Organization Studies, Boston College, Chestnut Hill, MA 01451, USA*

^b *University of Westminster, 35 Marylebone Road London NW1 5LS, UK*

Abstract

The concept of centrality is often invoked in social network analysis, and diverse indices have been proposed to measure it. This paper develops a unified framework for the measurement of centrality. All measures of centrality assess a node's involvement in the walk structure of a network. Measures vary along four key dimensions: type of nodal involvement assessed, type of walk considered, property of walk assessed, and choice of summary measure. If we cross-classify measures by type of nodal involvement (radial versus medial) and property of walk assessed (volume versus length), we obtain a four-fold polychotomization with one cell empty which mirrors Freeman's 1979 categorization. At a more substantive level, measures of centrality summarize a node's involvement in or contribution to the cohesiveness of the network. Radial measures in particular are reductions of pair-wise proximities/cohesion to attributes of nodes or actors. The usefulness and interpretability of radial measures depend on the fit of the cohesion matrix to the one-dimensional model. In network terms, a network that is fit by a one-dimensional model has a core-periphery structure in which all nodes revolve more or less closely around a single core. This in turn implies that the network does not contain distinct cohesive subgroups. Thus, centrality is shown to be intimately connected with the cohesive subgroup structure of a network.

© 2005 Published by Elsevier B.V.

1. Introduction

Centrality is a fundamental concept in network analysis. Bavelas (1948, 1950) and Leavitt (1951) used centrality to explain differential performance of communication networks and network members on a host of variables including time to problem solution, number of errors, perception of leadership, efficiency, and job satisfaction. Their work led to a great deal of

[☆] Earlier versions of this paper were presented at the 1991 *NSF Conference on Measurement Theory and Networks* (Irvine, CA), 1992 annual meeting of the *American Anthropological Association* (San Francisco), and the 1993 *Sunbelt XII International Social Network Conference* (Tampa, FL).

* Corresponding author. Tel.: +1 617 552 0452; fax: +1 617 552 4230.

E-mail addresses: borgatts@bc.edu (S.P. Borgatti), M.Everett@wmin.ac.uk (M.G. Everett).

¹ Tel.: +44 2083 318 716.

29 experimental, empirical, and theoretical research on the implications of network structure for
30 substantive outcomes, particularly in the context of organizations. Centrality has been used to
31 investigate influence in interorganizational networks (Laumann and Pappi, 1973; Marsden and
32 Laumann, 1977; Galaskiewicz, 1979), power (Burt, 1982; Knoke and Burt, 1983), advantage in
33 exchange networks (Cook et al., 1983; Marsden, 1982), competence in formal organizations (Blau,
34 1963), employment opportunities (Granovetter, 1974), adoption of innovation (Coleman et al.,
35 1966), corporate interlocks (Mariolis, 1975; Mintz and Schwartz, 1985; Mizruchi, 1982), status in
36 monkey grooming networks (Sade, 1972, 1989), power in organizations (Brass, 1984 and differ-
37 ential growth rates among medieval cities (Pitts, 1979). In addition, many other studies use well-
38 known measures of centrality but do not identify them as such. For example, researchers working
39 with ego-networks use the term “network size” (Campbell et al., 1986; Deng and Bonacich, 1991)
40 to refer to a variable that in another context we would recognize as degree centrality.

41 While many measures of centrality have been proposed, the category itself is not well defined
42 beyond general descriptors such as node prominence or structural importance. In addition, people
43 propose all kinds of interpretations of centrality measures, such as (potential for) autonomy,
44 control, risk, exposure, influence, belongingness, brokerage, independence, power and so on. The
45 one thing that all agree on is that centrality is a node-level construct. But what specifically defines
46 the category? What do all centrality measures have in common? Are there any structural properties
47 of nodes that are not measures of centrality?

48 Sabidussi (1966) tried to provide a mathematical answer to these questions. He suggested a set
49 of criteria that measures must meet in order to qualify as centrality measures. For example, he felt
50 that adding a tie to a node should always increase the centrality of the node, and that adding a tie
51 anywhere in the network should never decrease the centrality of any node. These requirements are
52 attractive: it is easy to see the value of separating measures that are “well-behaved” from measures
53 that behave less intuitively. However, there are problems with Sabidussi’s approach. For one thing,
54 it turns out that his criteria eliminate most known measures of centrality, including betweenness
55 centrality. This is clearly unsatisfactory. Furthermore, while his criteria provide some desirable,
56 prescriptive, characteristics for a centrality measure, they do not actually attempt to explain what
57 centrality is.

58 Freeman (1979) provided another approach to answering the ‘what is centrality’ question. He
59 reviewed a number of published measures and reduced them to three basic concepts for which he
60 provided canonical formulations. These were degree, closeness and betweenness. He noted that
61 all three attain their maximum values for the center of a star-shaped network, such as shown in
62 Fig. 1. It can be argued that this property serves as a defining characteristic of proper centrality
63 measures.

64 Borgatti (2005) has recently proposed a dynamic model-based view of centrality that focuses
65 on the outcomes for nodes in a network where something is flowing from node to node across the
66 edges. He argues that the fundamental questions one wants to ask about individual nodes in the
67 dynamic flow context are (a) how often does traffic flow through a node and (b) how long do things
68 take to get to a node. Once these questions are set, it becomes easier to construct graph-theoretic
69 measures based on the structure of the network that predict the answers to these questions. Hence,
70 in this approach, measures of centrality are cast as predictive models of specific properties of
71 network flows.

72 In this paper, we present an alternative perspective that eschews the dynamic element and is
73 fundamentally structural in character. It is a graph-theoretic review of centrality measures that
74 classifies measures according to the features of their calculation. Whereas the model-based view is
75 centered on the outcomes of centrality, the graph-theoretic view is centered on the way centrality

76 measures are calculated. In short, the present perspective is a means-based classification rather
 77 than the ends-based classification presented by Borgatti (2005).

78 2. Terminology

79 For simplicity (and in accordance with centrality convention), we will assume that all networks
 80 on which we might compute centrality measures consist of undirected graphs $G(V, E)$, in which
 81 V is a set of nodes (also called vertices, points or actors) and E is a set of edges (also called ties or
 82 lines) that connect them. Many centrality measures can be discussed in terms of directed graphs
 83 as well, but this topic is not treated here. It will be helpful to represent a graph in terms of its
 84 adjacency matrix A , in which $a_{ij} = 1$ if (i, j) is in E .

85 Nodes that are not adjacent may nevertheless be reachable from one to the other. A walk from
 86 node u to node v is a sequence of adjacent nodes that begins with u and ends with v . A trail is a
 87 walk in which no edge (i.e., pair of adjacent nodes) is repeated. A path is a trail in which no node
 88 is visited more than once.

89 The length of a walk is defined as the number of edges it contains, and the shortest path between
 90 two nodes is known as a geodesic. The length of a geodesic path between two nodes is known
 91 as the geodesic or graph-theoretic distance between them. We can represent the graph theoretic
 92 distances between all pairs of nodes as a matrix D in which d_{ij} gives the length of the shortest
 93 path from node i to node j .

94 3. Comparison of methods

95 To explain the graph-theoretic perspective, we begin by considering a sample of centrality
 96 measures and examining how they are computed. In a process similar to the anthropological
 97 technique of componential analysis, we extract dimensions along which measures vary. These are
 98 then used to develop a three-way typology of measures. We organize the discussion around the
 99 three best-known measures of centrality: degree, closeness and betweenness (Freeman, 1979).

100 3.1. Degree-like measures

101 We begin by considering the simplest and best-known measures of centrality: degree centrality.
 102 As defined by Freeman (1979), degree centrality is a count of the number of edges incident upon
 103 a given node. As shown in Eq. (1), it can be computed as the marginals of the adjacency matrix
 104 A :

$$105 \quad c_i^{\text{DEG}} = \sum_j a_{ij} \quad (1)$$

106 We can express this in matrix notation as $C^{\text{DEG}} = A\mathbf{1}$, where $\mathbf{1}$ is a column vector of ones.

107 It is useful to recognize that every edge is a walk of length 1. Consequently, we can think
 108 of degree centrality as counting the number of paths of length 1 that emanate from a node.²
 109 Degree centrality is therefore a special case of the measure proposed by Sade (1989) called k -

² As noted earlier, we assume an undirected graph. Hence, we can equally well describe degree in terms of the number of paths of length 1 that terminate at a node.

110 *path centrality*³ which counts all paths of length k or less that emanate from a node. When $k = 1$
 111 (its minimum value), the measure is identical to degree centrality. When $k = n - 1$ (its maximum
 112 value), the measure counts the total number of paths of any length that originate at a given
 113 node.

114 In order to establish the commonality of structure across measures of centrality, it is useful to
 115 note that k -path centrality may be computed as the marginals of a matrix W in which w_{ij} is the
 116 number of paths of length k or less from node i to node j . That is, $C^{K-path} = W\mathbf{1}$.

117 Other variations on this theme may be obtained by choosing different restrictions on the kinds
 118 of paths counted. For example, if we are only interested in shortest paths, we can define *geodesic*
 119 *k-path centrality* as the number of *geodesic* paths up to length k emanating from a given node. We
 120 can think of this as measuring the amount of direct involvement that a node has in the geodesic
 121 structure of the network.

122 Another variation is to count only edge-disjoint paths. Edge-disjoint paths are paths, which
 123 share no edges. Counting the number of edge-disjoint paths up to length k that originate or
 124 terminate at a given node yields a centrality measure we shall call *edge-disjoint k-path centrality*.
 125 Disjoint k -path centrality measures can be thought of as inverse measures of vulnerability. This
 126 interpretation is based on a theorem by Ford and Fulkerson (1956) which states that the number
 127 of edge-disjoint paths linking two nodes is equal to the minimum number of edges that must be
 128 deleted in order to disconnect the two nodes.⁴ In a network in which ties are subject to destruction
 129 (as in roads in a war zone), a disjoint k -path centrality measure assesses how difficult it would be
 130 to isolate a given node.

131 A variant of disjoint k -path centrality counts the number of vertex-disjoint paths up to length
 132 k rather than edge-disjoint paths. Vertex-disjoint paths are those which share no vertices (except
 133 the two end nodes). The set of such paths in a graph is a subset of the set of edge-disjoint paths.
 134 Menger (1927) showed that the number of vertex-disjoint paths linking two nodes is equal to the
 135 number of nodes that must be removed from a graph in order to isolate the two nodes from each
 136 other. The measure of social proximity developed by Alba and Kadushin (1976), and used as a
 137 basis for detecting *social circles*, is a (normalized) count of all vertex-disjoint paths of length 2
 138 or less connecting any two nodes. We call a measure counting the number of vertex-independent
 139 paths that originate or terminate at a given node a *vertex-disjoint k-path centrality* measure. The
 140 GPI power measure of Markovsky et al. (1988) is a vertex-disjoint k -path centrality measure,
 141 which subtracts the number of even-length vertex-disjoint paths emanating from a node from
 142 the number of odd-length vertex-disjoint paths emanating from the same node. All *reachability*
 143 measures (Higley et al., 1991), which count the number of nodes a given node can reach in a given
 144 number of links, are vertex-disjoint k -path centrality measures.

145 Thus far, we have only considered variations of degree centrality, which count true graph-
 146 theoretic paths. However, a number of measures count all walks, including those that visit the
 147 same nodes repeatedly. Katz's (1953) measure of centrality is a weighted count of the number
 148 of walks originating (or terminating) at a given node. The walks are weighted inversely by their
 149 length so that long, highly indirect walks count for little while short, direct walks count for a great
 150 deal. The extent to which the weights attenuate with length is controlled by an arbitrary parameter

³ Sade actually used the term *n-path centrality*, but since n is usually reserved for the number of nodes in a network, we have used k instead.

⁴ The number of edge-disjoint paths between two nodes is also equal to the maximum flow between them (Ford and Fulkerson, 1962).

151 supplied by the researcher. Katz's measure is defined as follows:

$$152 \quad w_{ij} = ba_{ij} + b^2(a^2)_{ij} + \dots + b^k(a^k)_{ij} + \dots = \sum_{k=1}^{\text{infinity}} b^k(a^k)_{ij} \quad (2)$$

$$153 \quad c_i = \sum_j w_{ij}$$

153 In matrix notation, $C^{\text{KATZ}} = W1$. Eq. (2) is based on the fact that the number of walks of length
 154 k between all pairs of nodes is given by the k th power of the adjacency matrix. The series is
 155 guaranteed to converge only if b is chosen to be smaller than the reciprocal of the largest eigenvalue
 156 of A . Hubbell (1965) proposed a measure very similar to Katz's, but which allows for the possibility
 157 of taking a weighted row sum of W . The weights are potentially but not necessarily derived from
 158 the network itself. If the weighting vector e is chosen to be all ones, Hubbell's measure equals
 159 Katz's minus 1. If e is chosen to be the degree of each node, as Hoede (1978) suggested, the result
 160 is that Katz's and Hubbell's measures are identical. Friedkin (1991) has developed a measure
 161 called *total effects centrality* which is equal to Katz's divided by the constant $1 - b$.

162 Bonacich (1987) writes a variant of Katz's measure in slightly more general terms as follows:

$$163 \quad w_{ij} = \delta(a_{ij} + b(a^2)_{ij} + b^2(a^3)_{ij} + \dots) = \delta \sum_{k=1}^{\text{infinity}} b^k(a^{k+1})_{ij} \quad (3)$$

$$164 \quad c_i = \sum_j w_{ij}$$

$$165 \quad C = W1$$

164 When b is positive, Bonacich's and Katz's measure are perfectly correlated. When b is negative,
 165 the two measures are perfectly negatively correlated. A key contribution of Bonacich's was to
 166 realize that b could be negative and that this would have a substantive interpretation in exchange
 167 networks (Cook et al., 1983; Markovsky et al., 1988). Indeed, Bonacich's measure predicts power
 168 use in experimental exchange networks very nicely. This is interesting because a negative value
 169 for b means that Eq. (3) effectively subtracts the number of even-length walks from the number
 170 of odd-length walks. This is exactly the same as the other well-known measure of power that
 171 emerges from the experimental exchange network literature, the GPI index of Markovsky et al.
 172 As Markovsky et al., point out, having many alters one link away from a node enhances that
 173 node's bargaining power, but having many alters two links away enhances the power of the node's
 174 first-order alters, and so on. More generally, a basic principle in exchange networks is that a node
 175 is powerful to the extent that it is connected to weak alters. In turn, a node is weak if it is connected
 176 to powerful alters. Interestingly, these descriptions resemble the hub and bridge distinctions of
 177 Mintz and Schwartz (1981a,b) and Mizruchi et al. (1986). The principal difference between GPI
 178 and the Bonacich power measures is that the former counts only vertex-disjoint paths while the
 179 latter counts all walks (weighted inversely by length).

180 Another way of interpreting the walk-based measures is in terms of an intuitive notion⁵ that
 181 a person's centrality should be a function of the centrality of the people he or she is associated
 182 with. In other words, rather than measure the extent to which a given actor "knows everybody",

⁵ Probably originating with Alexander (1963), but clearly evident in Bonacich (1972) as well.

183 we should measure the extent to which the actor “knows everybody who is anybody”. Hubbell’s
 184 measure can be written as follows:

$$185 \quad c^{\text{HUB}} = Xc^{\text{HUB}} + e \quad (4)$$

186 where X is a matrix derived from A , and e is the weighting vector of possibly exogenous con-
 187 tributions to status. Katz’s and Hoede’s measures are a special case of Hubbell’s in which e is
 188 equal to the row sums of X , and $X = bA$. Thus, both Katz and Hubbell can be seen as “implicit”
 189 centrality measures in which the centrality of a node is given by the weighted row sums of an
 190 adjusted adjacency matrix, where the weights are the centralities of the columns.

191 Bonacich (1972) noted the similarity of Eq. (4) to the definition of an eigenvector (Eq. (5)) and
 192 recommended that the principal eigenvector (associated with the largest eigenvalue) be used as a
 193 centrality measure. He has shown (Bonacich, 1991) that the eigenvector of A is the limit of Katz’s
 194 measure as b approaches $1/\lambda$ from below. Thus, the eigenvector can be regarded as an elegant
 195 summary of Katz, Hoede’s and Hubbell’s measures.

$$196 \quad v = \lambda^{-1}Av \quad (5)$$

197 Having defined the eigenvector v of adjacency matrix A , we can calculate the W matrix in Eq. (3)
 198 more simply, as follows:

$$199 \quad w_{ij} = a_{ij}v_j \quad (6)$$

200 Coleman’s (1973) Power and Burt’s (1982) prestige are application of the eigenvector measure
 201 to specific types of data. In Coleman’s case, the matrix A' is restricted to the relation “depends
 202 on”, while in Burt’s case A' is a non-symmetric relation such as “likes”.⁶

203 It is apparent that the variations among the degree-based measures are due entirely to the kinds
 204 of restrictions placed on the kinds of walks counted. This defines one typological dimension that
 205 we can use to classify measures. We refer to this dimension as *Walk Type*.

206 3.2. Closeness-like measures

207 It is also apparent that all of the measures considered so far count the number or volume of
 208 walks (of some kind) joining each node to all others. We shall refer to these as *volume measures*.
 209 Another set of centrality measures assesses the *lengths* of the walks that a node is involved in.
 210 We call these *length measures*. The distinction between volume measures and length measures
 211 forms another classificatory dimension, which we call *Walk Property*. It refers to what property
 212 of paths (their number or their length) is being measured.

213 The best known distance measure is Freeman’s (1979) *closeness centrality*, which is defined
 214 as the total geodesic distance from a given node to all other nodes. As shown in Eq. (5), it is
 215 computed as the marginals of a geodesic distance matrix D :

$$216 \quad c_i^{\text{CLO}} = \sum_j d_{ij} \quad (7)$$

217 In matrix notation $C^{\text{CLO}} = D\mathbf{1}$. This is clearly parallel to the degree-based measures discussed in
 218 the previous section, with D playing the role of W . Since the number of nodes is fixed in a network,

⁶ The descriptions are written in terms of A' instead of A because Coleman and Burt take column sums rather than row sums as we do here.

219 the measure is equivalent to the mean distance of a node to other nodes. Closeness centrality is an
 220 inverse measure of centrality since larger values indicate less centrality. In this sense, it technically
 221 measures farness rather than closeness⁷.

222 Direct measures of closeness (rather than farness) can be obtained by transforming the distance
 223 matrix into a “nearness” matrix prior to computing row marginals. For example, Høivik and
 224 Gleditsch (1975) recommend a linear transformation, as do Valente and Foreman (1998). The
 225 latter approach is to take either the row sums (which they call radiality) or the column sums
 226 (integration) of the geodesic distance matrix subtracted from a constant.⁸ In contrast, Burt (1991)
 227 recommends⁹ the following exponential transformation:

$$228 \quad s_{ij} = \alpha^{d_{ij}} \quad (8)$$

229 Other variants of closeness can be obtained by varying the way the initial distance matrix is
 230 defined. In Freeman’s measure, closeness is based on geodesic distances. Each d_{ij} entry in the
 231 geodesic distance matrix can be viewed as the minimum of the vector of lengths of all paths from
 232 i to j . However, if we do not believe that a given substantive phenomenon, such as diffusion of
 233 information, always makes use of the shortest paths, it makes sense to take into account all paths
 234 from i to j , perhaps by taking the median or mean length of all paths. The latter option is in fact
 235 the approach taken by Friedkin (1991) in developing his *immediate effects centrality*, which is
 236 defined as the reciprocal of the average distance from a given node to all others, where the distance
 237 between two nodes is defined (apart from scaling constants) as the average length of all paths
 238 between them. The problem with this, as we have discussed for other measures, is that many of
 239 the paths we shall be averaging together will not be totally distinct from each other. The question
 240 is, should we give full weight to all paths, or should we try to take into account the fact that some
 241 paths are largely redundant?

242 If we think of each path from i to j as a vector, we can see that what we are looking for is the
 243 length of a linear combination of the vectors in which some vectors are weighted more heavily
 244 than others according to their distinctiveness. Thus, we seek a set of weights or coefficients that
 245 are optimal with respect to some well-specified criterion. The question is, what criterion? In a way,
 246 the linear combination we want is the opposite of a factor or principal component. As Nunnally
 247 (1967) observes, the variance of a linear combination is high if the vectors are highly (positively)
 248 correlated, and low if they are not. Hence, we are looking for a linear combination that has as little
 249 variance as possible, given some constraint on the weights. If we denote the k th path between
 250 two nodes as p_k , the variance of the linear combination $w_1 p_1 + w_2 p_2 + \dots + w_k p_k = \sum_k w_k p_k$ is
 251 given by

$$252 \quad \text{Var} \left(\sum_k w_k p_k \right) = \sum_k \sum_l w_k w_l \sigma_{kl} \quad (9)$$

253 where σ_{kl} refers to the covariance between the k th and l th paths. Thus, we seek a set of weights w
 254 that minimize Eq. (9), subject to the constraint that $\sum w_k = 1$. After differentiating and rearranging

⁷ A normalized version of closeness, in which the reciprocals of c^{CLO} are multiplied by the number of nodes minus 1, solves this terminological problem.

⁸ It is sometimes claimed that this linear transformation enables us to measure closeness in disconnected graphs (i.e., those containing undefined distances) but this is not the case. In fact, if the constant is taken to be n , this approach gets the same results (linearly rescaled) as simply replacing undefined distances with n , which is clearly unsatisfactory.

⁹ In somewhat different context.

terms, we find that the optimal weights are the row marginals of the inverse of the covariance matrix, divided by the sum of all entries. Using these weights, we can theoretically construct a combined path from i to j whose length can be used as the distance between i and j . Computing this distance for all pairs of nodes, a new measure of closeness centrality can be constructed by computing the row marginals of the distance matrix, or, as above, from the reciprocals of the distance matrix.

The difficulty in all this, of course, is that we have not said how paths are to be represented as vectors. One possibility is a 0/1 indicator matrix X in which paths are columns, rows are edges, and values x_{ij} of the matrix indicate whether or not the i th edge occurs in the j th path. Note that the length of a path is given by the column sums minus 1. We can then compute covariance in the usual way, solve for the optimal weights, and construct a minimum variance linear combination. The length of the combined path is then given by the sum of its values.

A different approach is taken by Stephenson and Zelen (1989), who propose that we simply declare the covariance of two paths to be the number of edges they have in common¹⁰. If we pretend that Eq. (8) still holds, we can then solve for the linear combination of paths with minimum “variance”. Further, the variance of the best linear combination is then interpreted as its length, and therefore gives us the distance between i and j . The distances are converted to “nearnesses” by taking reciprocals, and a closeness measure is constructed by taking the harmonic mean of each row of the nearness matrix. Stephenson and Zelen invoke information theory to interpret the nearness matrix as “information”, and so they name their measure “information centrality”.

A variation on closeness is what we shall call *centroid centrality*. The idea is that one first identifies one or more nodes as the network *centroid*. Then, to calculate the centrality of any node, one measures the distance from that node to the centroid. One obvious choice for the centroid is the graph-theoretic center (Harary, 1969), which is the node (or pair of nodes if not unique) that has the least eccentricity. A node’s eccentricity is the length of its longest geodesic path to another node. We can then measure closeness as a node’s geodesic distance from the center. Of course, any criteria could be used to identify the central node(s), including other centrality measures. Another approach is to embed the graph into a multidimensional metric space (Freeman, 1983), find the location least distant from all nodes, and use the distance from that point to all others as centrality. This approach was used by Laumann and Pappi (1973). Here, the term “distance” refers to Euclidean distance, or any other distance metric used to define the vector space.

3.3. Betweenness-like Measures

All of the measures considered so far—including both the volume and the length measures—assess walks that emanate from or terminate with a given node. We shall refer to these as *radial* measures. Another class of centrality measures exists which are based on the number of walks

¹⁰ Interestingly, when all paths are the same length, the covariance between paths (represented by the edge-path incidence matrix described previously), really is related to the number of edges they have in common. Recall that covariance between paths k and l is defined as

$$\frac{1}{n} \sum_i x_{ik} x_{il} - \left(\frac{1}{n} \sum_i x_{ik} \right) \left(\frac{1}{n} \sum_i x_{il} \right)$$

When all paths have the same length, the term to the right of the subtraction becomes a constant, so the covariance is a linear transformation of $\sum_i x_{ik} x_{il}$, which is the number of edges shared by paths k and l .

that pass through a given node. We call these *medial* measures. The distinction between radial and medial measures forms the third classificatory dimension, which we call Walk Position.

The best known medial measure is Freeman's *betweenness centrality*. Loosely described, the betweenness centrality of a node is the number of times that any actor needs a given actor to reach any other actor. A more precise definition is as follows: Let g_{ikj} denote the number of geodesic paths from node i to node j , and let g_{ikj} denote the number of geodesic paths from i to j that pass through intermediary k . Then the betweenness centrality is defined as follows:

$$C_k^{\text{BET}} = \sum_i \sum_j \frac{g_{ikj}}{g_{ij}} \quad (10)$$

The measure is, in effect, k 's share of all shortest-path traffic from i to j , summed across all choices of i and j . If there is only one shortest path from any point to any other, the measure is equal to the number of geodesic paths that pass through a given node k . One can easily imagine that when the network being studied consists of ties that are very costly to build, betweenness will indeed index an ability to extort benefits from flows through the network. For example, if the network represents trade routes between medieval cities, the cities with high betweenness centrality have opportunities for amassing wealth and exerting control that other cities would not have (Pitts, 1979).

Several variations on betweenness centrality are possible. First of all, the reliance on geodesic paths alone may be undesirable. While inter-city trade might well take only shortest paths (in order to minimize costs), information might flow equally well across all possible paths. Hence we would modify g_{ikj} to record the number of paths of any kind that link i and j via k . Borgatti (2002, 2005) considers betweenness for all possible paths, as well as all possible trails, as well as walks (weighted inversely by length). However, he uses simulation to estimate the betweenness values rather than formulas. Newman (2004) provides closed-form equations for the case of random traversal via walks.

Another set of variants is obtained by limiting the length of paths, on the idea that very long paths are seldom used and should not contribute to a node's betweenness. Such measures might be called *k-betweenness centrality*, where k gives the maximum length of paths counted. Friedkin (1991) proposes a measure that is essentially k -betweenness measures with $k = 2$. Similarly, Gould and Fernandez (1989) develop brokerage measures that are specific variants of 2-betweenness measures. A sophisticated variant would be a class of betweenness measures that count paths of all lengths, but weight them inversely in proportion to their length.

Of course, if we count all paths or walks we are in danger of double-counting since many paths can share the same subset of edges. If this is a concern, we would want to count only edge-disjoint paths. This is exactly what Freeman et al. (1991) have done. Their measure is called *flow betweenness*¹¹. It is called flow betweenness because of the well-known relationship between the number of edge-independent paths between a pair of nodes, and the amount of material that can flow from one node to another via all possible edges (Ford and Fulkerson, 1956). Since flow betweenness assesses the proportion of edge-independent paths that involve a given node, it is in effect measuring the amount of flow in the network that would not occur if the node were not present (or were choosing not to transmit). This is really the essence of any betweenness measure: the potential for withholding flow, otherwise known as gatekeeping.

¹¹ Flow betweenness bears the same relationship to geodesic betweenness that information centrality bears to Freeman's closeness measure. Both information centrality and flow betweenness take into account all edge-disjoint paths, whereas Freeman's closeness and geodesic betweenness consider only geodesic paths.

330 An interesting aspect of flow betweenness is how it is computed. Because the sets of edge-
 331 independent paths between any two nodes are not unique, flow betweenness cannot be calculated
 332 directly by counting paths. Instead, programs like UCINET (Borgatti et al., 2002) essentially
 333 simulate the gatekeeping process by calculating flows between all pairs of nodes in the net-
 334 work, then removing the node whose centrality is being measured, and then recalculating the
 335 flows.

336 More specifically, let us denote by W the matrix of maximum flows between nodes (i.e., the
 337 number of edge independent paths between them) and denote by ${}^k W$ the matrix derived from W by
 338 deleting row and column k . In addition, let us denote by ${}^k W^*$ the matrix obtained by deleting node
 339 k from the original network, and recalculating the flow matrix. We can then define flow centrality
 340 as follows:

$$341 \quad c_k = \sum_{i,j} \frac{{}^k w_{i,j} - {}^k w_{i,j}^*}{{}^k w_{i,j}} \quad (11)$$

342 We can reformulate all betweenness measures in accordance with Eq. (11), simply by changing
 343 the W matrix. For example, to calculate Freeman's betweenness, we take W to be the geodesic
 344 count matrix in which w_{ij} gives the number of geodesic paths from i to j . Applying Eq. (11) gives
 345 scores exactly equal to twice Freeman's values.

346 Thus, betweenness-type measures might be thought of as “proportion reduction in cohesion”
 347 (PRC) measures, analogous to proportion reduction in error (PRE) measures in statistics. Where
 348 something flows over the links of a network, PRC measures quantify the potential of a node to
 349 disrupt flows throughout the network by ceasing its own transmissions.

350 If we define cohesion in terms of reachability, a PRC measure essentially indexes the network
 351 fragmentation that results from removing a node. Such a measure, called fragmentation (F),
 352 was introduced by Borgatti (2003, forthcoming). The F measure is simply the proportion of
 353 disconnected pairs of nodes that results when a given node is removed from a network. The bigger
 354 the value, the more important the node in maintaining cohesion. As implemented in UCINET
 355 (Borgatti et al., 2002), the measure is accompanied by a normalization in which the level of
 356 fragmentation of the original network is subtracted from the fragmentation after removing a
 357 given node, divided by the fragmentation of the original network. Large positive values indicate
 358 nodes that contribute to the cohesion of the network, while large negative values indicate nodes
 359 that reduce the cohesiveness of the network.

360 Defining cohesion in terms of distance yields a different set of measures. As a specific exam-
 361 ple, let us define a cohesion matrix W as the reciprocals of the geodesic distance between all
 362 pairs of nodes in a network (with the convention that the reciprocal of an undefined distance
 363 is 0). Now, we remove a given node (whose centrality we are measuring) from the W matrix,
 364 yielding ${}^k W$, and also remove the node from the original network and re-compute the recip-
 365 rocals of geodesic distances among all pairs of remaining nodes, yielding ${}^k W^*$. Then we apply Eq.
 366 (11), subtracting the values of ${}^k W^*$ from the values of ${}^k W$ and dividing by ${}^k W$. Summing all of
 367 these adjusted values, we obtain the relative decrease in cohesion obtained by removing a given
 368 node.

369 A similar measure was proposed by Borgatti (2003, forthcoming). Called “distance-weighted
 370 fragmentation” (DF), this measure is defined as the average reciprocal distance among nodes after
 371 removal of a given node. The measure is 1 when all nodes are distance 1 from each other (i.e., a
 372 complete graph), and 0 when all nodes are isolates. Intermediate values index the extent to which
 373 the presence of a node tends to reduce distances in the network.

Table 1
Cross-classification of centrality measures

	Radial	Medial
Volume	Freeman degree, Sade k-path, Bonacich eigenvector, Katz status, Hubbell status, Hoede status, Doreian iterated Hubbell, Markovsky et al. GPI, Friedkin TEC, Coleman power, Bonacich power, Burt prestige	Anthonisse rush Freeman betweenness, Freeman et al. flow between, Friedkin MEC, Newman RWB
Length	Freeman closeness, Stephenson-Zelen information Friedkin IEC	Borgatti DF

374 4. A Typology of measures

375 It is apparent in this review of measures that all of the measures evaluate a node's involvement in
 376 the walk structure of a network. That is, they evaluate the volume or length of walks of some kind
 377 that originate, terminate, or pass through a node. Furthermore, all are based on the marginals of an
 378 appropriately constructed node-by-node matrix, although the method of calculating marginals can
 379 vary from simple sums to averages and weighted averages to harmonic means, and so on. Thus
 380 four basic dimensions distinguish between centrality measures: the types of walks considered
 381 (called Walk Type, such as geodesic or edge-disjoint), the properties of walks measured (called
 382 Walk Property, namely volume or length), the type of nodal involvement (called Walk Position,
 383 namely radial or medial), and type of summarization (called Summary Type, such as sum or
 384 average).¹²

385 The Walk Type dimension concerns the restrictions that some measures impose on the kind of
 386 walks considered, such only geodesics, only true paths, limited length walks, and so on. The Walk
 387 Property dimension distinguishes between measures that evaluate the number of walks a node is
 388 involved in from measures that evaluate the length of those walks. The Walk Position dimension
 389 distinguishes between measures that evaluate walks emanating from a node from measures that
 390 evaluate walks passing through a node. The Summary Type dimension distinguishes measures
 391 using different ways to summarize rows of the walk matrix¹³. For simplicity, Table 1 gives a list
 392 of centrality measures cross-classified by just two of the dimensions: Walk Property (volume vs.
 393 length), and Walk Position (radial vs. medial).

394 While each centrality measure uses a different W matrix, in all cases the W matrix is an indicator
 395 of social proximity/cohesion among nodes.¹⁴ Almost all of the W matrices we have seen have
 396 been identified specifically in the network literature as measures of cohesion. The i, j th cell of

¹² A fourth attribute, less important, is the choice of summary statistic. All centrality measures can be computed by summarizing the rows of an actor-by-actor matrix W (or A). Typically, this statistic is a simple sum or average. However, other summary statistics, such as weighted means, medians, modes, minimums and maximums, are used as well. This decision point is well known in cluster analysis, where the difference between some clustering methods is whether the minimum, maximum or median is used to compute the distance from a point to a set of points (Johnson, 1967; D'Andrade, 1978). An important example of a weighted mean is the eigenvector, which is used by Bonacich (1972), Burt (1982), and Doreian (1986).

¹³ This decision point is well known in cluster analysis, where the difference between some clustering methods is whether the minimum, maximum or median is used to compute the distance from a point to a set of points (Johnson, 1967; D'Andrade, 1978). An important example of a weighted mean is the eigenvector, which is used by Bonacich (1972), Burt (1982), and Doreian (1986).

¹⁴ We use the terms proximity and cohesion interchangeably to refer to social closeness or strength of connection between pairs of nodes.

any W matrix indicates the “connectedness” or “relatedness” of i and j , which is often assumed to correspond to a potential for transmission of attitudes, diseases, resources, etc. Of course, the values of W can be scaled such that large numbers indicated greater cohesion (as in a matrix of valued strengths of ties) or lesser cohesion (as in a geodesic distance matrix).

If the W matrix reflects dyadic cohesion, it is no surprise that summarizing the values of the W matrix gives the well-known measures of network cohesion such as density and characteristic path length. Density is simply the average of the adjacency matrix across all cells, while characteristic path length is the average (or other summary measure) of the geodesic distance matrix.

The cohesion matrix is also used as the basis for all methods of detecting cliques and other cohesive subgroups (e.g. Alba and Kadushin, 1976; Burt, 1991). The myriad definitions and techniques of cohesive subgroups can be viewed as identifying clusters within the cohesion matrix. Interestingly, the Walk Property dimension that we have identified for centrality measures is also a key dimension in the classification of cohesive subgroups (Borgatti et al., 1990). As is well-known, the cohesive subgroup known as a clique is characterized by having maximum density (number of ties) and minimum distance (length of paths). Since its formalization by Luce and Perry (1949), subsequent researchers have sought to relax the definition of clique by relaxing either the distances among members of the subgroup, or the number of ties (i.e., walks of length 1). Concepts such as n -cliques, n -clans, and n -clubs relax the minimal distance property of the clique, while concepts such as k -plexes, k -cores, l s-sets, and λ -sets relax the maximum density property of cliques. In short: length versus volume.

Similarly, what centrality measures do is summarize the amount (and type) of dyadic cohesion that each node is involved in. In effect, centrality measures are indices of the share of dyadic cohesion attributable to each node. For comparison, whole network measures of cohesion, such as density or characteristic path length, are exactly like centrality measures except that instead of breaking out the summary by node, the entire cohesion matrix is summarized, much like the grand marginal in a contingency table.

Thus, a basic claim of the graph-theoretic typology presented here is that dyadic cohesion provides a common basis for not only centrality measures but subgroups and network cohesion. In this way, our analysis provides a sense of continuity with the other major areas of network analysis.

Another benefit that the typology provides is a partial answer to the commonly asked question of how to choose among centrality measures. The typology essentially divides measures into groups that, to put it in marketing terms, are more competitive with each other than with other measures. Our claim is that measures within the same box in Table 1 are similar enough on key attributes that they can be thought of as competitive, i.e., as potentially substitutable alternatives for each other. Among measures within each box, we can reasonably ask which is better. In contrast, measures in different boxes differ in fundamental ways, and are perhaps best viewed as complementary.

5. Radial measures and the core periphery assumption

It is apparent that all radial measures are constructed the same way. First one defines an actor-by-actor matrix W that records the number or length of walks of some kind linking every pair of actors. Then one summarizes each row of W by taking some kind of mean or total. Thus, centrality provides an overall summary of a node’s participation in the walk structure of the network. It is a measure of how much of the walk structure is due to a given node. It is quite literally the node’s share of the total volume or length of walks in the network.

442 Thus, the essence of a radial centrality measure is this: radial centrality summarizes a node's
 443 connectedness with the rest of the network. To the extent that dyadic cohesion is seen as index-
 444 ing influence (e.g., Katz, 1953; Hubbell, 1965; Friedkin, 1991), the centrality measure judges
 445 the overall influence of an actor. If the cohesion matrix is adjacency (as in degree centrality),
 446 and adjacency matrix represents the “friend of” relation, then centrality summarizes an actor's
 447 “friendliness”.

448 This raises a question. Under what conditions does it make sense to summarize, with a single
 449 value, a node's cohesion with all others? Consider the mean of any list of numbers. It can always
 450 be computed, but only serves as a summary when the distribution of the numbers is unimodal.
 451 Indeed, if the list is known to be normally distributed, the mean and standard deviation alone
 452 can generate the entire distribution. But if the shape of the distribution is bimodal, the mean is
 453 a very poor summary. If the ideal serving temperature of tea is in the range of 35–50° for much
 454 of the population (because they like iced tea) and the ideal serving temperature is in the range
 455 of 130–160° for the other half of the population (because they like hot tea), is the average of the
 456 ideal temperatures provide a good assessment of the population's tastes? That is, does a luke-warm
 457 temperature of 97.5° provide a good picture of what the people's tastes are? Probably not.

458 A radial centrality measure is clearly interpretable in a network in which dyadic cohesion is
 459 unimodal, but not in one which is multimodal. That is to say, radial centrality makes sense in
 460 networks which have, at most, one center. This means that a cohesive subgroup analysis would
 461 find only one subgroup (a core) to which all nodes belong to a greater or lesser extent. The network
 462 would not be divided in two or more subgroups. In that case, radial centrality would effectively
 463 serve as a measure of “coreness” (Borgatti and Everett, 1999; Everett and Borgatti, 2004), which
 464 is to say, strength of membership in the one and only group. Bonacich (1972) has noted that
 465 if a network contains more than one component (i.e., a maximal set of nodes that are mutually
 466 reachable), eigenvector centrality will assign zeros to all nodes not in the largest component,
 467 even if they are highly central in their own component. Rather, those nodes load highly on the
 468 remaining eigenvectors. In other words, the eigenvectors of a cohesion matrix measure strength
 469 of involvement of each node to each major subgroup (component).

470 Before we can interpret a radial measure of centrality, we must determine whether the network
 471 satisfies the one-group requirement. A network will exhibit a core-periphery pattern whenever its
 472 cohesion matrix can be modeled as a non-negative function of its marginals (Borgatti and Everett,
 473 1999). For example, if the cohesion matrix W is a valued adjacency matrix in which w_{ij} gives the
 474 number of interactions observed between i and j , and the model of independence holds (i.e., no
 475 interaction), then the network has a core-periphery structure. This is because the independence
 476 model specifies that the extent of dyadic cohesion between nodes i and j is proportional to the
 477 product of their general proximity to anyone (i.e., their centrality). Hence, the only region of the
 478 network with high densities of proximity will be populated by high centrality nodes, and there will
 479 only be one such region. This pattern of distribution of proximities is precisely a core-periphery
 480 structure.

481 Interestingly, one measure of centrality “comes with” such a test built-in: Bonacich's eigen-
 482 vector centrality. The eigenvectors and eigenvalues of any symmetric matrix can be multiplied to
 483 recreate the matrix, as shown in Eq. (12):

$$A = V'DV$$

$$a_{ij} = \sum_k v_{ik} e_k v_{jk} \quad (12)$$

485 In the equation, V is an $n \times n$ matrix whose columns are the eigenvectors of A , the matrix D is
 486 a diagonal matrix of eigenvalues, and e rewrites those eigenvalues as a simple vector such that
 487 $e_k = d_{kk}$. If the summation is performed for all k ranging from 1 to n , the approximation is exact.
 488 If the summation is performed using fewer than n eigenvectors, the approximation is as close as
 489 possible under a least squares criterion. Since eigenvector centrality is defined as the eigenvector
 490 of A with the largest eigenvalue, it is the single best vector for estimating the values of A under
 491 the following simple model:

$$492 \quad a_{ij} = \lambda_i^{\text{EIG}} c_j^{\text{EIG}} \quad (13)$$

493 According to the equation, the existence (or strength) of the tie between nodes i and j is approxi-
 494 mated by the product of their centralities (adjusted by λ , which serves as a scaling constant). The
 495 accuracy of the approximation is roughly indexed by the size of the eigenvalue, relative to the
 496 others. If all other eigenvalues are near zero, the approximation will be nearly perfect. Thus, the
 497 model fits when the eigenvector centrality alone is sufficient to reproduce the observed pattern of
 498 ties. When this occurs, the network necessarily exhibits a core-periphery pattern¹⁵. The relative
 499 size of the largest eigenvalue can therefore be interpreted as indicating the extent to which the
 500 network has a core-periphery structure.

501 The fit of a core-periphery model to an observed network may be seen as a generalized measure
 502 of network centralization. Freeman (1979) defines centralization as the sum of differences between
 503 the centrality of the most central node and all other nodes, divided by the same sum calculated
 504 on a star graph with the same number of nodes. The idea is that the star epitomizes the ideal of
 505 a centralized network, and Freeman's statistic gives the extent to which the observed network
 506 conforms to the ideal type. It is, therefore, a measure of fit between a network and an ideal model.
 507 The difference is in the choice of ideals. The star is only one example of a network structured
 508 as a core and periphery. What about networks with more than one node in the core? What about
 509 networks with nodes that are neither central nor peripheral, but somewhere in between? The
 510 core-periphery model can include networks with all of these characteristics.

511 Since the generic formula for a radial centrality measure is

$$512 \quad c_i = \sum_j w_{ij}, \quad (14)$$

513 radial centrality can be seen as a partitioning of total network cohesion ($\sum \sum w_{ij}$) by actor. The
 514 centrality of a node is its share of, or contribution to, total cohesion. If the number of nodes in
 515 the network is fixed, the total cohesion is a measure of overall cohesive density, and centrality is
 516 a node's contribution to that density.

517 However, in general we cannot view centrality as *generating* cohesion. The core-periphery
 518 model does not necessarily carry with it a *process* model. Consider, for example, fitting the
 519 independence model to a W matrix whose cells record the number of paths from each node to
 520 every other. If the model fits well, we are safe in concluding that the data form a core-periphery
 521 structure, and therefore a k -path centrality measure has a reasonable interpretation. But we have not
 522 specified a theory that explains how an underlying attribute of actors causes the observed pattern
 523 of paths. Such a theory is difficult to construct. If the cohesion matrix were simple adjacency for

¹⁵ Of course, all nodes can have equal centrality and satisfy the model, in which case we might either choose to regard all nodes as core or all nodes as periphery. The big point is that such a model cannot generate a structure with two or more cores (i.e. subgroups).

524 the “friend of” relation, it is plausible that a property of actors (e.g. friendliness) might determine
 525 the probability of forming a tie with another actor (cf., Holland and Leinhardt, 1981). But it
 526 is difficult to understand how friendliness works to create a specific number of *paths* or walks,
 527 since walks are as much functions of all the other nodes as they are of the two endpoints. Thus
 528 most centrality measures should be thought of as summaries of a node’s position in a one-group
 529 core-periphery structure, but not as parameters that generate that structure.

530 It should be noted that medial measures of centrality do not make the same one-group assump-
 531 tion. These measures correctly assign particularly high centrality scores to nodes serving as bridges
 532 between subgroups. However, it is still the case that it is difficult to interpret a given value of
 533 medial centrality without knowing the group’s cohesive structure. For example, the center of a
 534 sociometric star (a core-periphery structure) is not only highly medial, it is also central in more
 535 conventional, radial, ways as well (i.e., it is “in the thick of things”). In contrast, the liaison
 536 between several different subgroups can be very high on a medial centrality measure, and yet be
 537 only peripheral to each subgroup.

538 Some empirical evidence on the fundamental difference between medial and radial measures
 539 may be found in a study by Nakao (1990). She computed Freeman’s graph centralization mea-
 540 sures on all possible graphs of 4, 5, 6, 7, and 8 nodes. After computing correlations among the
 541 measures for each size of network, she concludes: “These correlations show that the betweenness-
 542 based centrality measure behaves somewhat differently from the other two measures, indicated
 543 by lower correlations. The degree-based measure and the closeness-based measure are related
 544 very closely in a linear manner.” She also notes that “The graphs whose order of centrality values
 545 is $C^{\text{BET}} > C^{\text{CLO}} > C^{\text{DEG}}$ tend to be divided into sub-clusters that are connected to each other by
 546 a line or via a focal point. This type of graph may be characterized as a decentralized network,
 547 in the manner defined in organizational research. On the other hand, the graphs which produce
 548 $C^{\text{CLO}} > C^{\text{DEG}} > C^{\text{BET}}$ would be described as one-cluster networks which contain a circle involving
 549 a large proportion of points in the network.”

550 6. Discussion

551 The differences between radial and medial measures discussed in the last section suggest that
 552 this distinction more important than the volume versus length distinction. In choosing between
 553 volume and length measures, one is choosing between different conceptions of cohesion. It seems
 554 plausible to suggest that, for a given theoretical application, it is possible to say that one is better
 555 than the other. For example, if one is studying risk of receiving in a timely manner something
 556 flowing through the network, it would seem that length measures makes the most sense since
 557 they map directly to the expected arrival times (Borgatti, 1995). When the concern is with the
 558 certainty of arrival of something flowing through the network, volume measures would seem like
 559 an obvious choice.

560 In contrast, the choice between radial and medial measures can be seen in terms of the distinct
 561 roles played by nodes in the network. For example, consider a cohesion matrix W defined as the
 562 number of the paths between all pairs of nodes. Thus, there exists an inventory of every path in
 563 the network. A radial measure of volume counts the number of these paths in which a given node
 564 serves as an endpoint. A medial measure counts the number of these paths in which the node
 565 serves as an interior point. Together, the radial and the medial add up to the total number of paths
 566 that a node is involved with in any role. In this sense, we can speak of decomposing a node’s
 567 total involvement in the paths of a network into radial and medial portions, as shown in Eq. (15).
 568 If so, radial and medial measures are complementary and both are needed to deliver a complete

569 picture of a node's contribution to the network (cf., [Friedkin, 1991](#)). Whereas radial measures
 570 assess group membership, medial measures assess bridging, reminiscent of the distinction in the
 571 social capital literature of bonding social capital and bridging social capital, or closed versus open
 572 ego networks.

$$573 \quad \text{Total Involvement} = \text{Radiality} + \text{Mediality} \quad (15)$$

574 It should be noted that, for most of the field of centrality, this decomposition is metaphorical
 575 rather than literal. It is literally accurate for the specific example given, but not elsewhere. For
 576 example, while nodes can only be in one position on a path, they can occur in multiple positions
 577 in a trail or walk, so that radial and medial counts no longer add to the total number of sequences.
 578 In addition, measures like Freeman's betweenness do not just count the number of times a node
 579 occupies an interior position of a geodesic, but weight those times according to exclusivity.

580 7. Conclusion

581 Following [Sabidussi \(1966\)](#), we have described the notion of centrality in purely graph-theoretic
 582 terms: what all measures of centrality do is assess a node's involvement in the walk structure
 583 of a network. This is the graph-theoretic answer to the question 'What do centrality measures
 584 measure?' We have suggested that centrality measures differ along four key dimensions: choice of
 585 summary measure, type of walk considered, property of walk assessed, and type of involvement.
 586 The choice of summary dimension has the least variance, consisting mostly of simple sums and
 587 averages, along with a few exemplars of weighted sums (e.g., eigenvectors) and centroids. The
 588 type of walk dimension distinguishes measures based on edges, geodesics, paths, trails and walks.
 589 The property of walk dimension distinguishes between volume and length measures. The type of
 590 involvement dimension distinguishes between radial and medial measures.

591 It can be seen that the single distinction made by [Borgatti \(2004\)](#) between frequency (measuring
 592 how often something flows across a node) and time (how soon something flows to a node) can
 593 be derived as a collapsing of the property of walk and type of involvement dimensions. That
 594 is, the frequency-based measures in [Borgatti \(2004\)](#) are medial-volume measures in the present
 595 terminology while the time-based measures correspond to radial-length measures. In addition, the
 596 cross-classification of measures by type of involvement and property of walk results in a four-fold
 597 classification that is not inconsistent with [Freeman's \(1979\)](#) three-fold categorization.

598 Radial measures in particular are reductions via aggregation operators of pairwise proximities
 599 to attributes of nodes or actors. These aggregations range from simple marginals (degree)
 600 to weighted marginals (eigenvectors) to distances along euclidean axes (centroid). The types of
 601 reductions correspond to standard statistical scaling and modeling techniques: simple marginals
 602 correspond to fitting the log-linear model of quasi-independence to a square, symmetric table
 603 with missing diagonals; eigenvectors correspond to factoring a correlation or covariance matrix;
 604 and euclidean axes correspond to a one-dimensional MDS scaling of a cohesion matrix. Not sur-
 605 prisingly, the usefulness and interpretability of radial measures depends on the fit of the cohesion
 606 matrix to the one-dimensional model, just as in univariate statistics the mean is most interpretable
 607 when applied to a unimodal distribution.

608 As discussed by [Borgatti and Everett \(1999\)](#), a network that is fit by a one-dimensional model
 609 has a core-periphery structure in which all nodes revolve more or less closely around a single
 610 core. Thus, radial centrality measures are most interpretable when the cohesion matrix passes a
 611 test of core-peripheriness, in which case the measures can be viewed as measures of "coreness".

612 Just as radial measures were shown to largely reduce to marginals of a cohesion matrix W ,
 613 medial measures were also reduced to a common formulaic structure that we referred to as
 614 proportion reduction in cohesion. As such, medial measures essentially measure the impact of the
 615 presence of a node on the dyadic cohesion among all pairs of nodes. In other words, they measure
 616 the change in cohesion that would result from removing a given node. As such, medial measures
 617 do not depend on core/periphery structures for interpretability, and in fact are particularly useful
 618 when networks have “clumpy” structures characterized by wide variation in local density.

619 At a general level, we note the relationship of centrality concepts with the concepts of graph
 620 cohesion and cohesive subgroups. The key underlying concept is that of dyadic cohesion—the
 621 social proximity of pairs of actors in a network. Dyadic cohesion is what is measured by the W
 622 matrix that undergirds all measures of centrality. There are two fundamental ways of analyzing
 623 cohesion. One is to seek regions of the network that are more cohesive than others—a focus on
 624 the pattern of cohesion. This constitutes the field of cohesive subgroups. The other is to attribute
 625 to individual nodes their share of responsibility for the cohesion of the network—a focus on the
 626 amount of cohesion. This constitutes the field of centrality measures. Within that, two fundamental
 627 approaches are discernable—the radial approach that directly partitions total cohesion by node,
 628 and the medial approach that assesses a node’s contribution to cohesion by removing it. The other
 629 fundamental distinction—between volume and length measures—is essentially an argument about
 630 the meaning of cohesion.

631 **Uncited references**

632 Erickson (1988), Everett and Borgatti (1999), Freeman (1980), Laumann and Pappi (1976),
 633 Newman (2005), Tam (1989) and Taylor (1969).

634 **Acknowledgements**

635 We thank Phil Bonacich, Lin Freeman, Noah Friedkin, and Bill Stevenson for extensive com-
 636 ments on this and earlier versions.

637 **References**

- 638 Alba, R., Kadushin, C., 1976. The intersection of social circles. *Sociological Methods and Research* 5, 77–102.
 639 Alexander, C.N., 1963. A method for processing sociometric data. *Sociometry* 26, 268–269.
 640 Bavelas, A., 1948. A mathematical model for group structure. *Human Organization* 7, 16–30.
 641 Bavelas, A., 1950. Communication patterns in task orientated groups. *Journal of the Acoustical Society of America* 22,
 642 271–288.
 643 Blau, P.M., 1963. *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies*. Univ.
 644 of Chicago Press, Chicago.
 645 Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical*
 646 *Sociology* 2, 113–120.
 647 Bonacich, P., 1987. Power and centrality: a family of measures. *American Journal of Sociology* 92, 1170–1182.
 648 Bonacich, P., 1991. Simultaneous group and individual centralities. *Social Networks* 13, 155–168.
 649 Borgatti, S.P., 1995. Centrality and AIDS. *Connections* 18 (1), 112–115.
 650 Borgatti, S.P., 2002. Types of network flows and how to destabilize terrorist networks. In: *Sunbelt International Social*
 651 *Networks Conference*, New Orleans.
 652 Borgatti, S.P., 2003. The Key Player Problem. 241–252 in *Dynamic Social Network Modeling and Analysis: Workshop*
 653 *Summary and Papers*, R. Breiger, K. Carley, P. Pattison (Eds.), National Academy of Sciences Press.
 654 Borgatti, S.P. (forthcoming) Identifying sets of key players in a network. *Computational, Mathematical and Organizational*
 655 *Theory*.

- 656 Borgatti, S.P., 2005. Centrality and network flow. *Social Networks* 27 (1), 55–71.
- 657 Borgatti, S.P., Everett, M.G., Shirey, P., 1990. LS sets, lambda sets and other cohesive subsets. *Social Networks* 12,
658 337–357.
- 659 Borgatti, S.P., Everett, M.G., 1999. Models of core/periphery structures. *Social Networks* 21, 375–395.
- 660 Brass, Daniel J., 1984. Being in the right place: a structural analysis of individual influence in an organization. *Administrative Science Quarterly* 29, 518–539.
- 661 Burt, R.S., 1991. *Structure Version 4.2 Reference Manual*. Center for the Social Sciences, Columbia University.
- 662 Burt, B.S., 1982. *Toward a Structural Theory of Action*. New York, Academic Press.
- 663 Campbell, K.E., Marsden, P.V., Hurlbert, J., 1986. Social resources and socioeconomic status. *Social Networks* 8, 97–117.
- 664 Coleman, J.S., 1973. Loss of power. *American Sociological Review* 38, 1–17.
- 665 Coleman, J.S., Katz, E., Menzel, H., 1966. *Medical Innovation: A Diffusion Study*. Bobbs-Merrill, Indianapolis.
- 666 Cook, K.S., Emerson, R.M., Gillmore, M.R., Yamagishi, T., 1983. The distribution of power in exchange networks: theory
668 and experimental results. *American Journal of Sociology* 89, 275–305.
- 669 D'Andrade, R., 1978. U-statistic hierarchical clustering. *Psychometrika* 43, 59–67.
- 670 Deng, Z., Bonacich, P., 1991. Some effects of urbanism on black social networks. *Social Networks* 13, 35–50.
- 671 Doreian, P., 1986. Measuring relative standing in small groups and bounded social networks. *Social Psychological Quarterly* 49, 247–259.
- 672 Erickson, B.H., 1988. The relational basis of attitudes. In: Wellman, B., Berkowitz, S.D. (Eds.), *Social Structures*. Cambridge University Press, New York, pp. 99–122.
- 673 Everett, M.G., Borgatti, S.P., 1999. Peripheries of Cohesive Subsets. *Social Networks* 21: 397–407.
- 674 Everett, M.G., Borgatti, S.P., 2004. Extending centrality. In Carrington, P., Scott, J., Wasserman, S. (Eds.) *Models and Methods in Social Network Analysis* (forthcoming).
- 675 Ford, L.R., Fulkerson, D.R., 1956. Maximal flow through a network. *Canadian Journal of Mathematics* 8, 399–404.
- 676 Ford, L.R., Fulkerson, D.R., 1962. *Flows in Networks*. Princeton University Press, Princeton.
- 677 Freeman, L.C., 1979. Centrality in networks: I. conceptual clarification. *Social Networks* 1, 215–239.
- 678 Freeman, L.C., 1980. The gatekeeper, pair-dependency, and structural centrality. *Quality and Quantity* 14, 585–592.
- 679 Freeman, L.C., 1983. Spheres, cubes and boxes: graph dimensionality and network structure. *Social Networks* 5,
682 139–156.
- 683 Freeman, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: a measure of betweenness based on network
684 flow. *Social Networks* 13, 141–154.
- 685 Friedkin, N.E., 1991. Theoretical foundations for centrality measures. *American Journal of Sociology* 96, 1478–1504.
- 686 Galaskiewicz, J., 1979. *Exchange Networks and Community Politics*. Sage Publications, Beverly Hills.
- 687 Gould, R.V., Fernandez, R.M., 1989. Structures of mediation: a formal approach to brokerage in transaction networks.
688 In: Clogg, C.C., Ann, Arbor (Eds.), *Sociological Methodology*. Blackwell, MI, pp. 89–126.
- 689 Granovetter, M.S., 1974. *Getting a Job: A Study in Contacts and Careers*. Harvard University Press, Cambridge, MA.
- 690 Harary, F. 1969. *Graph Theory*. Reading, Mass: Addison-Wesley.
- 691 Higley, J., Hoffman-Lange, U., Kadushin, C., Moore, G., 1991. Elite integration in stable democracies: a reconsideration.
692 *European Sociological Review* 7, 35–53.
- 693 Hoede, C. 1978. *A New Status Score for Actors in a Social Network*. unpublished manuscript, dept. of mathematics,
694 Twente University.
- 695 Høivik, T., Gleditsch, N.P., 1975. Structural parameters of graphs: a theoretical investigation. In: Blalock, H.M., et al.
696 (Eds.), *Quantitative Sociology*. Academic Press, New York, pp. 203–223.
- 697 Holland, P.W., Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs. *Journal of the*
698 *American Statistical Association* 76, 33–65.
- 699 Hubbell, C.H., 1965. An input–output approach to clique identification. *Sociometry* 28, 377–399.
- 700 Johnson, S.C., 1967. Hierarchical clustering schemes. *Psychometrika* 32, 241–254.
- 701 Katz, L., 1953. A new index derived from sociometric data analysis. *Psychometrika* 18, 39–43.
- 702 Knoke, D., Burt, R.S., 1983. Prominence. In: Burt, R.S., Minor, M. (Eds.), *Applied Network Analysis: A Methodological*
703 *Introduction*. Sage Publications, Beverly Hills, pp. 195–222.
- 704 Laumann, E.O., Pappi, F.U., 1973. New directions in the study of community elites. *American Sociological Review* 38,
705 212–230.
- 706 Laumann, E.O., Pappi, F.U., 1976. *Networks of Collective Action*. Academic Press, New York.
- 707 Leavitt, H.J., 1951. Some effects of certain communication patterns on group performance. *Journal of Abnormal and*
708 *Social Psychology* 46, 38–50.
- 709 Luce, R.D., Perry, A.D., 1949. A method of matrix analysis of group structure. *Psychometrika*. 20, 319–327.
- 710 Mariolis, P., 1975. Interlocking directorates and control of corporations. *Social Science Quarterly* 56, 425–439.
- 711

- 712 Markovsky, B., Willer, D., Patton, T., 1988. Power relations in exchange networks. *American Sociological Review* 53,
713 220–236.
- 714 Marsden, P.V., 1982. Brokerage behavior in restricted exchange networks. In: Marsden P.V., Lin, N. (Eds.), *Social*
715 *Structure and Network Analysis*. Sage Publications, Beverly Hills, pp. 201–218.
- 716 Marsden, P.V., Laumann, E.O., 1977. Collective action in a community elite: exchange, influence resources, and issue
717 resolution. power. In: Liebert, R.J., Imershein, A.W. (Eds.), *Paradigms and Community Research*. Sage Publications,
718 Beverly Hills, pp. 199–250.
- 719 Menger, K., 1927. Zur Allgemeinen Kurventheorie. *Fund. Math.* 10, 96–115.
- 720 Mintz, B., Schwartz, M., 1981a. Interlocking directorates and interest group formation. *American Sociological Review*
721 46, 851–868.
- 722 Mintz, B., Schwartz, M., 1981b. The structure of intercorporate unity in american business. *Social Problems* 28, 87–103.
- 723 Mintz, B., Schwartz, M., 1985. *The Power Structure of American Business*. University of Chicago Press, Chicago.
- 724 Mizruchi, M. 1982. *The Structure of the American Corporate Network: 1904–1974*. Beverly Hills: Sage.
- 725 Mizruchi, M.S., Mariolis, P., Schwartz, M. Mintz, B., 1986. Techniques for Disaggregating Centrality Scores in Social
726 Networks. In: 26–48 in *Sociological Methodology 1986*, Tuma, N.B. (Eds.). San Francisco, Jossey-Bass.
- 727 Nakao, K., 1990. Distribution of measures of centrality: enumerated distributions of freeman's graph centrality measures.
728 *Connections* 13, 10–22.
- 729 Newman, M.E.J., 2005. A measure of betweenness centrality based on random walks. *Social Networks* 27 (1), 39–54.
- 730 Nunnally, J.C., 1967. *Psychometric Theory*. McGraw Hill, New York.
- 731 Pitts, F.R., 1979. The medieval river trade network of Russia revisited. *Social Networks* 1, 285–292.
- 732 Sabidussi, G., 1966. The centrality index of a graph. *Psychometrika* 31, 581–603.
- 733 Sade, D.S., 1972. Sociometrics of macaca mulatta I: linkages and cliques in grooming networks. *Folia Primatologica* 18,
734 196–223.
- 735 Sade, D.S., 1989. Sociometrics of macaca mulatta III: N-path centrality in grooming networks. *Social Networks* 11,
736 273–292.
- 737 Stephenson, K., Zelen, M., 1989. Rethinking centrality: methods and examples. *Social Networks* 11, 1–37.
- 738 Tam, T. 1989. Demarcating the Boundaries Between Self and the Social: The Anatomy of Centrality in Social Networks.
739 11:387–401.
- 740 Taylor, M., 1969. Influence structures. *Sociometry* 32, 490–502.
- 741 Valente, T.W., Foreman, R.K., 1998. Integration and radiality: measuring the extent of an individual's connectedness and
742 reachability in a network. *Social Networks* 20, 89–105.