Re-conceptualizing centrality in social networks†

D. SCHOCH[1,2] and U. BRANDES[1,2]

[1]*Departement of Computer & Information Science, University of Konstanz, Konstanz, Germany*
emails: david.schoch@uni-konstanz.de, ulrik.brandes@uni-konstanz.de
[2]*Graduate School of Decision Sciences, University of Konstanz, Konstanz, Germany*

In the social sciences, networks are used to represent relationships between social actors, be they individuals or aggregates. The structural importance of these actors is assessed in terms of centrality indices which are commonly defined as graph invariants. Many such indices have been proposed, but there is no unifying theory of centrality. Previous attempts at axiomatic characterization have been focused on particular indices, and the conceptual frameworks that have been proposed alternatively do not lend themselves to mathematical treatment.

We show that standard centrality indices, although seemingly distinct, can in fact be expressed in a common framework based on path algebras. Since, as a consequence, all of these indices preserve the neighbourhood-inclusion pre-order, the latter provides a conceptually clear criterion for the definition of centrality indices.

Key words: network science, social networks, centrality, ranking, positional dominance

1 Introduction

Social network analysis [13, 27, 29, 48] is an area of applied network science [17] with a long tradition [23]. Starting with the concept of status in studies of sociometric choice [37] and popularized by small-group communication experiments [2, 3, 35], indices evaluating the position of nodes in a network have become a signature form of network analysis [16].

Depending on context and terminology, such indices operationalise various substantive concepts referred to as, e.g., centrality, status, prestige, importance, or power, by means of graph invariants. Applications in other areas add to the list of interpretations. We here use the term centrality as an umbrella concept to subsume the variety of indices that capture instantiations of a broadly construed notion of structural importance.

While research on particular centrality indices – including their characterization and computation – abounds, we are not aware of any substantial overarching results other than empirical and experimental comparison [11, 18, 25, 28]. In the absence of agreement even on the minimum requirements for centrality indices, this is not surprising. Attempts

at organizing the space of centrality indices are therefore mostly conceptual in nature [10, 12, 22].

We propose a strikingly simple formal characterization of centrality: If an actor has the same (and possibly more) ties, it can never be less central. The criterion was derived not only to capture a shared intuition underlying common definitions of centrality, but also to discriminate against other indices that do not. We motivate the proposal by means of a unifying framework in which existing indices can be expressed. The framework not only aids the proof that the indices actually satisfy the criterion, but it also suggest the definition of further indices based on the same principles.

We start by defining some of the more commonly used centrality indices in Section 2. Our framework is derived in Section 3 and based on path algebras with a special property that captures the effect of indirect relations among vertices. We prove in Section 4 that any index based on this framework favours vertices which dominate the neighbourhood of others, and discuss some implications in Section 5.

2 Preliminaries

For ease of exposition, we consider only simple undirected and unweighted graphs $G = (V, E)$ on a finite set of vertices V and edges $E \subseteq \binom{V}{2}$ without loops or multiple edges. We do note, however, that our results can be generalized to other classes including weighted, directed, and multi-graphs. We use $n = |V|$ and $m = |E|$ to denote the number of vertices and edges.

Adjacent vertices $\{u, v\} \in E$ are called neighbours, and we denote the neighbourhood of a vertex $v \in V$ by $N(v) = \{u \in V : \{u, v\} \in E\}$. The closed neighbourhood is defined as $N[v] = N(v) \cup \{v\}$.

An (s, t)-path (or walk or trail) is an alternating sequence of vertices and edges starting with $s \in V$ and ending with $t \in V$ such that each edge consists of the two vertices next to it. If $s = t$, the path is closed, and if no vertex other than possibly $s = t$ appears twice it is called simple. A graph is connected, if every pair $s, t \in V$ is connected by a path.

Again, without limitation, we consider only connected graphs so that complications with the definition of some indices are avoided. The two graphs on nine vertices in Figure 1 serve as running examples.

2.1 Centrality indices

A centrality index assigns non-negative real numbers to the vertices of a graph. Before arguing which such assignments are admissible, we review a number of the more common examples. The simplest such index is *degree centrality*,

$$c_D(v) = \deg(v),$$

where $\deg(v) = |N(v)|$ is the degree of $v \in V$.

Let $A(G)$ be the adjacency matrix of graph G and observe that the kth power of $A(G)$ gives, for every pair of vertices, the number of paths of length k between them. *Subgraph*

	degree	closeness	betweenness	eigenvector	subgraph
A	2	0.063	1.83	0.580	2.452
B	3	0.071	4.83	0.857	3.404
C	3	0.067	2.83	0.843	3.756
D	3	0.067	1.83	1.000	3.956
E	3	0.071	3.67	0.997	3.872
F	3	0.071	2.83	0.937	3.843
G	3	0.071	3.83	0.949	3.844
H	3	0.077	4.50	0.955	3.509
I	3	0.067	2.83	0.975	3.936

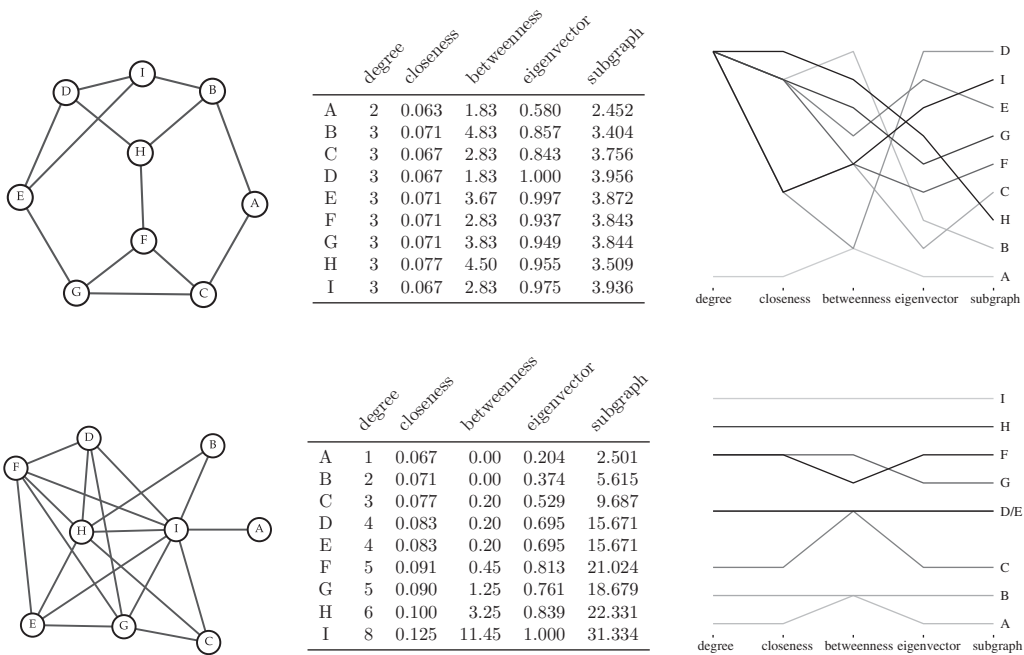|	degree	closeness	betweenness	eigenvector	subgraph
A	1	0.067	0.00	0.204	2.501
B	2	0.071	0.00	0.374	5.615
C	3	0.077	0.20	0.529	9.687
D	4	0.083	0.20	0.695	15.671
E	4	0.083	0.20	0.695	15.671
F	5	0.091	0.45	0.813	21.024
G	5	0.090	1.25	0.761	18.679
H	6	0.100	3.25	0.839	22.331
I	8	0.125	11.45	1.000	31.334

FIGURE 1. Centrality indices compared on two example graphs. The tables summarise the centrality scores of all vertices according to standard indices, and the parallel coordinate plots in the right compare the resulting rankings. While the indices disagree widely on the top graph, the bottom graph yields fairly consistent rankings.

centrality [19] is defined as

$$c_S(v) = \sum_{k=0}^{\infty} \frac{A(G)_{vv}^k}{k!},$$

and thus the weighted sum of all closed paths containing v. The scaling reduces the contribution of long paths but also guarantees convergence. Note that $A(G)_{vv}^2 = \deg(v)$ since there is a one-to-one correspondence between edges and closed paths of length 2.

Other generalisations of degree centrality also take indirect relations into account. The length of a shortest (s,t)-path defines the distance, dist(s,t), between s and t. *Closeness centrality* [4],

$$c_C(v) = \left(\sum_{t \in V} \text{dist}(v,t) \right)^{-1},$$

is then defined as the inverse of the total distance between a vertex $v \in V$ and all other vertices. The inverse is taken to maintain the interpretation that higher scores indicate greater centrality. Alternative order-reversing transformations such as subtraction from an upper bound have been used [14, 46] but the functional form of the transformation will not matter here as long as it is monotonic.

A different generalisation is based on shortest paths passing through, rather than emanating from, a vertex. Let $\sigma(s,t)$ be the number of shortest (s,t)-paths, and $\sigma(s,t|v)$ the number of shortest (s,t)-paths that contain $v \in V \setminus \{s,t\}$ as an inner vertex. The fraction

$\delta(s,t|v) = \frac{\sigma(s,t|v)}{\sigma(s,t)}$ is called the dependency of s and t on v, and *betweenness centrality* [21] is defined as

$$c_B(v) = \sum_{s,t \in V} \delta(s,t|v),$$

where we set $\delta(s,t|v) = 0$ if $v \in \{s,t\}$ for convenience.

Feedback centralities (seemingly) are not defined in terms of indirect relations but values of neighbours. Since the adjacency matrix of a connected undirected graph is real and symmetric, the Perron–Frobenius Lemma guarantees that the eigenvector associated with the largest eigenvalue λ is unique up to scaling and all entries have the same sign. We can thus define *eigenvector centrality* [8] as

$$c_E(v) = \frac{1}{\lambda} \cdot \sum_{w \in N(v)} c_E(w),$$

and assume that c_E positive and normalized such that, say, all entries sum to 1.

These are just a few of the more common examples, and many others have been proposed [32, 44].

2.2 Characterization and classification

Centrality indices such as those listed in the previous section are typically defined *ad hoc*, either with a particular application scenario in mind or as a variation of previously proposed indices to eliminate some perceived deficiency.

As Freeman notes, there is "no unanimity on what a centrality is, its conceptual foundations, and proper procedure of measurement". [22] Irritated by the growing number of minor variations on centrality indices already in the 1970s, he organized many of them around three key concepts. As a consequence, degree, closeness, and betweenness centrality have since been considered prototypical. Freeman does not, however, provide a criterion that delineates the scope of centrality, and the class of feedback centralities such as eigenvector centrality is left out completely.

Borgatti and Everett [12] and Borgatti [10] provide classifications for centralities organised around the mathematical ingredients in their definition rather than resulting properties. The classifications are therefore conceptual as well, and not intended to be comprehensive.

Formal attempts at delineation and classification are generally based on axiomatization [7,31,34,39–41,47]. Sabidussi [42] is the first to propose an axiomatic characterization of centrality. Its key elements are invariance under graph isomorphisms and a form of monotonicity under graph modification: A vertex receiving a new incident edge (via operations called *edge addition* and *edge switching*) can only become more central.

Sabidussi's axiom system is designed with the intuition of closeness in mind and thus ends up ruling out other indices (most of which had not been proposed at the time) [33]. Similarly, subsequent axiom systems have been designed largely to characterize particular indices in terms of their properties, with invariance under isomorphisms and monotonicity under graph modification as recurring concepts [7].

Neither the conceptual classifications nor the axiomatic characterizations proposed to date establish a scope on which a general theory of centrality could be built. The only uncontested assertion appears to be Freeman's *star property:* "A person located in the centre of a star is universally assumed to be structurally more central than any other person in any other position in any other network of similar size". [22]

3 Unification via path algebras

We now show that, despite their differences in rationale and definition, common centrality indices can be cast in a unifying framework that will enable us to substantially strengthen the requirements of the star property in the next section.

In one way or another, each centrality index evaluates vertices by aggregating their relationships with others. These relationships can be direct (such as adjacency in degree centrality) or indirect (such as distance in closeness centrality). We therefore break down the definition of centrality indices into three generic steps from which any particular index is obtained via suitable instantiations. These steps are as follows:

 (1) Definition of an indirect relation via some path algebra.
 (2) Definition of vertex positions via coordinates that evaluate indirect relations.
 (3) Definition of centrality scores as aggregate values from positions.

We briefly discuss the underlying formalisms and then re-formulate the centrality indices from the previous section in these terms. The unified formulation will prove useful for general statements about centrality indices in the next section.

3.1 Indirect relations

We argue that the indirect relations on top of which centrality indices are typically built can be obtained from certain path algebras. A comprehensive treatment of the concepts used in the following is given, for instance, by Gondran and Minoux [24].

A semiring $(S, \oplus, \odot, \bar{0}, \bar{1})$ is a set of values S including $\bar{0}, \bar{1} \in S$ together with two closed, associative binary operations $\oplus, \odot : S \times S \to S$ with neutral elements $\bar{0}$ (called zero) and $\bar{1}$ (called unity), respectively. In addition, $\oplus$ is commutative, $\bar{0}$ is an absorbing element for $\odot$, and $\odot$ distributes over $\oplus$.

A path algebra characterizes indirect relationships between vertices $s, t \in V$ of a graph $G = (V, E)$ by associating a value from a semiring with every (s, t)-path, and then aggregating them. Given a graph $G = (V, E)$ and a semiring $(S, \oplus, \odot, \bar{0}, \bar{1})$, we obtain a matrix $A \in S^{V \times V}$ by choosing an element $\bar{e} \in S$ (called edge value) and setting

$$a_{st} = \begin{cases} \bar{e} & \text{if } \{s, t\} \in E \\ \bar{0} & \text{if } \{s, t\} \notin E \text{ (including } s = t) \end{cases}$$

for all $s, t \in V$.

An (s, t)-path P along vertices $s = v_0, v_1, \ldots, v_{k-1}, v_k = t$ evaluates to the product $a(P) = \bigodot_{i=1}^{k} a_{v_{i-1} v_i}$, where $a(P) = \bar{1}$ if $k = 0$. The relationship between s and t is obtained from $a_{st}^* = \bigoplus_P a(P)$, where the summation extends over all (s, t)-paths P, and $a_{st}^* = \bar{0}$

if there is no such path. Multiplication and summation of the semiring thus capture the result of concatenation and aggregation of paths.

The following is a convenient joint formulation in terms of matrices. Let $A \in S^{V \times V}$ be the matrix of direct relations a_{st} as defined above and let $\mathbf{1} \in S^{V \times V}$ be the matrix with $\bar{1}$'s on the diagonal and $\bar{0}$'s elsewhere. Replacing the usual addition and multiplication with the corresponding semiring operations, we obtain a new semiring on the matrices $S^{V \times V}$ with zero $\mathbf{0}$ and unity $\mathbf{1}$ from the generalised matrix operations. Letting $A^0 = \mathbf{1}$ and $A^k = A \odot A^{k-1}$ for $k \geqslant 1$, we obtain the closure $A^* = \bigoplus_{k=0}^{\infty} A^k$ which contains the elements a_{st}^* defined above. We assume that A^* exists and is unique, i.e., our semirings are assumed to be closed.

3.1.1 *Reachability and distance*

An example is the path algebra giving rise to the reachability relation in a graph $G = (V, E)$. Consider the semiring $(S = \{0, 1\}, \oplus = \max, \odot = \min, \bar{0} = 0, \bar{1} = 1)$. We define $a_{st} \in S$, $s, t \in V$, as above with edge value $\bar{e} = 1$, so A is in fact the adjacency matrix of G. Since two vertices in a simple undirected graph are connected by a path if and only if they are connected by a simple path of length at most $n - 1$, it follows that $A^* = A^{n-1}$ and $a_{st}^* = 1$ if and only if there is at least one (s, t)-path.

For shortest-path distances, we modify the reachability semiring by extending $S = \mathbb{N}_0 \cup \{\infty\}$ and substituting $\oplus = \min$, $\odot = +$ (the usual addition), and $\bar{0} = \infty$. Then, concatenation yields the number of edges in a combined path whereas aggregation gives the minimum number of edges in either path, so that $a_{vw} = 1 = \bar{e}$ for $\{v, w\} \in E$ gives rise to $A^* = A^{n-1}$, again, with $a_{st}^* = \text{dist}(s, t)$.

3.1.2 *Shortest-path counts*

The semiring that will be used for the derivation of betweenness centrality, the geodetic semiring, is due to Batagelj [1]. Let $S = (\mathbb{N}_0 \cup \{\infty\}) \times \mathbb{N}_0$, $\bar{0} = (\infty, 0)$ and $\bar{1} = (0, 1)$. For $(a, b), (c, d) \in S$, let

$$(a, b) \odot (c, d) = (a + c, b \cdot d)$$

$$(a, b) \oplus (c, d) = \left(\min\{a, c\}, \begin{cases} b & \text{if } a < c \\ b + d & \text{if } a = c \\ d & \text{if } a > c \end{cases} \right).$$

Setting $a_{vw} = (1, 1) = \bar{e}$ for $\{v, w\} \in E$, we obtain the closure $A^* = A^{n-1}$ with $a_{st}^* = (\text{dist}(s, t), \sigma(s, t))$, where $\sigma(s, t)$ is the number of shortest (s, t)-paths. Note that $n - 1$ is an upper bound on the length of a shortest path, and that $\bar{0}$ is indeed absorbing.

3.1.3 *Walk counts*

The final group of path algebras we are considering is designed for walk-based centrality indices such as subgraph centrality, Katz status [30], and, as it turns out, eigenvector centrality. For $s, t \in V$ and $k \in \mathbb{N}_0$, let $\omega_{st,k}$ denote the number of (s, t)-walks with

exactly k edges, and $\omega_{st}^{(k)} = \sum_{i=0}^{k} \omega_{st,k}$ the number of (s,t)-walks with at most k edges. We represent sequences $(\omega_k)_{k \in \mathbb{N}_0}$ by their generating functions $\Omega(\beta) = \sum_{k=0}^{\infty} \omega_k \cdot \beta^k$. A generating function [43] is a formal power series that converges to a limit function for sufficiently small $\beta \in (-r, r)$ and is divergent for $|\beta| > r$. The actual radius of convergence $0 \leqslant r \leqslant 1$ depends on the growth of the elements in the sequence ω.

Let $\bar{0}$ and $\bar{1}$ be the generating functions associated with the infinite sequences $(0, 0, \dots)$ and $(1, 0, 0, \dots)$. Then, the usual addition and multiplication of functions yield a semiring on the set of generating functions restricted to non-negative arguments $\beta \geqslant 0$. Since these operations correspond to element-wise addition and convolution of the sequences, the path algebra we obtain by setting $\bar{e}$ to the generating function of $(0, 1, 0, 0, \dots)$ has the walk-generating functions as its closure $A^* = (\Omega_{st}(\beta))_{s,t \in V}$. The joint convergence radius is $\frac{1}{\lambda}$, where λ is the largest eigenvalue of the adjacency matrix.

For reasons discussed below, we will rather be interested in the sequences $(\omega_{st}^{(k)})_{k \in \mathbb{N}_0}$, $s, t \in V$, of the number of (s, t)-walks up to length k. The generating function of these prefix sums of walk counts is obtained simply from scaling $\frac{1}{1-\beta} \cdot \Omega(\beta)$. We define the corresponding path algebra by substituting $\bar{1} = (1, 1, \dots)$ and $\bar{e} = (0, 1, 1, \dots)$ and adjusting multiplication to $(1 - \beta)$ times the product of the two generating functions (otherwise, the product would represent the prefix sums of prefix sums).

A variant for fast-growing sequences and with a larger convergence radius are exponential generating functions $\omega_{st}(\beta) = \sum_{k=0}^{\infty} \frac{\omega_{st,k}}{k!} \beta^k$. Semiring and path algebra for exponential walk-generating functions are constructed as before and with the same sequences defining zero, unity, and the edge value. The joint convergence radius for the exponential generating functions of the closure, though, is infinite. This is the semiring underlying subgraph centrality and total communicability [5].

3.2 Positions

To obtain a centrality index from a path algebra, we next transform the elements of the semiring (which characterize indirect relationships) into non-negative real numbers $x_{st} \in \mathbb{R}_{\geqslant 0}$ for all $s, t \in V$. These will serve to define the relational position of a vertex $v \in V$ as a vector

$$\text{pos}(v) = (x_{vt})_{t \in V} .$$

Since we here restrict our attention to unweighted undirected graphs, this is but a special case of a recently introduced notion of position that applies to multiplex relations and any number of attributes on the vertices and edges [15]. We will come back to this in the final section but would like to point out already that the restricted type of position considered here constitutes what has been referred to as nodal statistic elsewhere [6]. From position vectors, centrality scores are obtained by an index-specific summarization defined in the next section.

The closure obtained from the path algebra on the shortest-path semiring ($\mathbb{N}_0 \cup \{\infty\}, \min, +, \infty, 0$) already contains the shortest-path distances $\text{dist}(s, t) = a_{st}^*$. Since these are the quantities used in the definition of closeness centrality, the identity transform $x_{st} = \text{dist}(s, t)$, $s, t \in V$, suffices.

For betweenness centrality, however, we transform the closure (dist, σ) of the path algebra obtained from the geodetic semiring into dependencies

$$\delta(s,v) = \sum_{t \in V} \begin{cases} \frac{\sigma(s,v) \cdot \sigma(v,t)}{\sigma(s,t)} & \text{if } \mathrm{dist}(s,t) = \mathrm{dist}(s,v) + \mathrm{dist}(v,t) \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

for all $s, v \in V$. We let $x_{vs} = \delta(s,v)$ and thus transpose the matrix of dependencies to obtain positions from its columns. In other words, the degree to which s depends on v defines how much v is in the role of a broker for s, and betweenness centrality is the degree to which v is a broker for everyone else.

Values from walk-generating functions are obtained by providing an argument $0 \leqslant \beta < r$ within the radius of convergence. Katz' status is based directly on prefix sums $x_{vt} = \frac{1}{1-\beta} \Omega_{vt}(\beta)$, where β is known as the attenuation parameter, and subgraph centrality and communicability [5] are based on the exponential walk-generating functions $x_{vt} = \omega_{vt}(\beta)$ which converge for any choice of $\beta \geqslant 0$.

Eigenvector centrality is generally defined as a feedback measure in which the centrality of a vertex depends on the centrality of its neighbours. An equivalent formulation, however, is the limit of the share of all walks of length k that start at the vertex. We can obtain values proportional to the desired shares $\lim_{k \to \infty} \frac{\omega_{vt,k}}{\lambda^k}$ from the walk-generating functions as $\lim_{k \to \infty} \frac{\Omega_{vt}(\frac{1}{\lambda})}{k}$ and therefore also from prefix sums $\lim_{k \to \infty} \frac{\lambda}{(\lambda-1)k^2} \Omega_{vt}(\frac{1}{\lambda})$.

In all these cases, we obtain for every vertex $v \in V$ a vector $\mathrm{pos}(v) \in \mathbb{R}^V_{\geqslant 0}$ of non-negative real numbers describing the relationships of that vertex with every vertex in the graph. The final step is to find a single number describing the structural importance manifest in a position.

3.3 Centrality scores

For a specific type of relation, the position of a vertex quantifies all its relationships. Depending on interpretation, however, these relationships may combine in different ways.

Closeness centrality, for instance, was defined as the inverse of the sum over all position entries, $\sum_{t \in V} \mathrm{dist}(v,t)$. Eccentricity centrality [26], on the other hand, is a centrality defined as the inverse of the maximum entry, $\max_{t \in V} \mathrm{dist}(v,t)$, rather than the sum.

Betweenness centrality is again defined as a sum, although of dependencies, $\sum_{s \in V} \delta(s,v)$.

Since walk-generating functions in the path algebras defined above sum over all walks of any length, Katz' status [30] is obtained as $\sum_{t \in V} \frac{1}{1-\beta} \cdot \Omega_{vt}(\beta)$. While the sum over exponential walk-generating functions, $\sum_{t \in V} \omega_{vt}(\beta)$, is known as total communicability, subgraph centrality is obtained via projection to position entry $\omega_{vv}(\beta)$ rather than summation.

Any combination of a closed semiring, transformation into coordinates, and summarisation might define a centrality index; we thus do not only unify existing indices but provide a cornucopia for new ones. While other summarisations are conceivable, the most commonly used are sum, extremum, and projection, possibly followed by an order-reversing transformation such as taking the inverse or subtracting from an upper bound in cases where the interpretation of more or less central is the reverse of the ordering of the aggregate quantities.
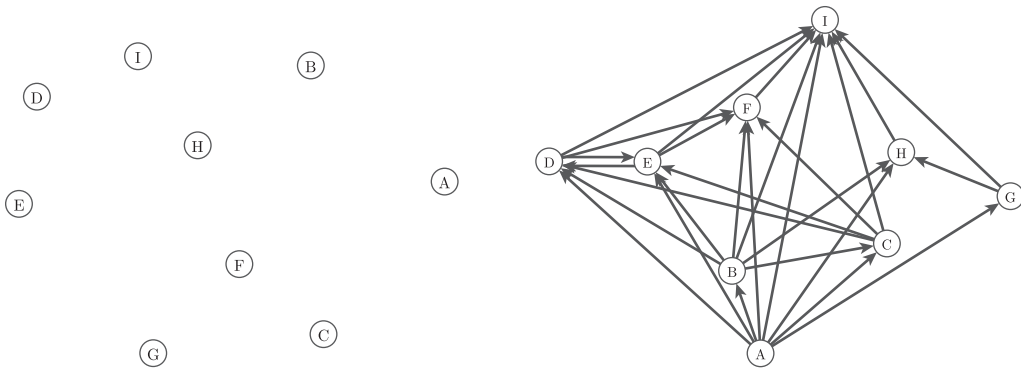
FIGURE 2. Dominance relation among vertices in example graphs from Figure 1. A directed edge indicates that the target dominates the source vertex. While no two vertices are comparable in the almost regular graph, the neighbourhood-inclusion pre-order is almost complete in the other.

The rationale for breaking down the definition of centrality indices into exactly these steps will become more apparent in the next section, where we derive a general statement about all centrality indices defined in this way.

4 Preservation of neighbourhood inclusion

Our goal is to show that the following criterion expresses the essence of centrality. Note that it effectively introduces requirements that tighten the star property.

Definition 1 (Neighbourhood inclusion) *Let $G = (V, E)$ be a simple undirected graph and $u, v \in V$. The relation*

$$u \leqslant v \quad if \quad N(u) \subseteq N[v]$$

indicates that the neighbourhood of v includes that of u. We say that u is dominated by v.

Neighbourhood inclusion defines a pre-order, i.e., a reflexive and transitive binary relation, on the vertices of a graph. It is sometimes referred to as the vicinal pre-order [20]. The closed neighbourhood $N[v]$ is used to ensure that the relation covers the case $\{u, v\} \in E$. Figure 2 depicts the neighbourhood-inclusion pre-orders for the example graphs from Figure 1.

In the previous section, we argued that the definition of centrality indices can be decomposed into three steps. A matrix representing the adjacencies (direct relationships) of a graph is first transformed into a closure matrix representing indirect relationships, then into position vectors quantifying the indirect relationships for each vertex, and finally into centrality scores summarising these position vectors in a single value.

We want to show that the neighbourhood-inclusion pre-order is preserved by centrality indices, and do so by showing that it is preserved in each of the three defining steps. Therefore, we need pre-orders also on semirings and positions. The canonical pre-order

associated with a semiring $(S, \oplus, \odot, \overline{0}, \overline{1})$ is given by

$$a \leqslant b \quad \text{if} \quad a \oplus c = b \quad \text{for some } c \in S.$$

For positions $x, y \in \mathbb{R}^n_{\geqslant 0}$, we define

$$x \leqslant y \quad \text{if} \quad x_i \leqslant y_i \quad \text{for all} \quad i = 1, \ldots, n$$

as a special case of positional dominance [15]. Depending on the indirect relation from which positions are derived, the relation may actually be modified as follows: When comparing the positions of two vertices $u, v \in V$, the comparison of reflexive entries $\text{pos}(u)_u$ and $\text{pos}(v)_v$ with $\text{pos}(u)_v$ and $\text{pos}(v)_u$ may not be meaningful, so that, for instance, the latter are pitted against each other or all four entries are ignored.

A simple property of semirings captures the central intuition behind centrality indices, namely that moving an actor away from another by an additional indirection can only reduce the value of their relationship. We say that an edge value

$$\overline{e} \text{ is decreasing,} \quad \text{if} \quad \overline{e} \odot a \leqslant a \quad \text{for all} \quad a \in S.$$

A path algebra that is constructed from a semiring using a decreasing edge value is called decreasing as well.

All centrality indices discussed above are based on decreasing path algebras. Note that the usual order of the integers is reversed in the shortest-path semiring $(\mathbb{N}_0 \cup \{\infty\}, \min, +, \infty, 0)$: If $a \geqslant b$ as integers, then $a \leqslant b$ in the semiring of shortest-path distances because $a \oplus c = \min\{a, c\} = b$ for the choice $c = b$.

While the semiring of walk-generating functions does not give rise to a decreasing path algebra, the semiring based on prefix sums does for $\beta \geqslant 0$. This observation captures the intuition shared by all walk-based centrality indices that having fewer long walks can be compensated for by additional short ones. The alternating contributions of walks for $\beta < 0$ defy this idea.

Our technical result can now be stated as follows.

Theorem 2 *Let $G = (V, E)$ be a simple undirected graph, and $c : V \to \mathbb{R}_{\geqslant 0}$ a centrality index that is obtained from a decreasing path algebra via monotone quantification and summarization of positions. Then, for all $u, v \in V$,*

$$u \leqslant v \implies c(u) \leqslant c(v) .$$

In other words, all standard centrality indices share the property that if a vertex v dominates a vertex u, then v is at least as central as u. Since this matches the intuition that it does not hurt to have more direct relationships it suggests to consider neighbourhood inclusion as the defining property of centrality.

The proof of the theorem rests on the following lemma which states that neighbourhood inclusion is preserved in the semiring pre-order.

Lemma 3 *Let $G = (V, E)$ be a simple undirected graph and A^* the closure of a decreasing path algebra on G. For every pair of vertices $u, v \in V$,*

$$N(u) \subseteq N[v] \implies \begin{cases} a_{ut}^* \leqslant a_{vt}^* & \forall t \in V \setminus \{u\} \\ \wedge & a_{uu}^* \leqslant a_{vu}^* \oplus \overline{1} \\ \wedge & a_{uu}^* \leqslant a_{vv}^*. \end{cases}$$

Proof Let $u \leqslant v \in V$, i.e., $N(u) \subseteq N[v]$. Consider first any $t \in V \setminus \{u\}$. By definition, $a_{ut}^* = \bigoplus_P a(P)$ where the sum extends over all (u, t)-paths and $a(P) = \bigodot_{e \in P} a(e)$ where the product is over all edges in P. We define an injective mapping of (u, t)-paths P to (v, t)-paths Q with $a(P) \leqslant a(Q)$. If a (u, t)-path P starts with an edge $\{u, w\} \in E$ such that $w \neq v$, neighbourhood inclusion guarantees that there is a unique path Q starting with $\{v, w\} \in E$ and continuing like P. Since $a_{uw} = \overline{e} = a_{vw}$ by definition, $a(P) = a(Q)$. If, however, $w = v$, then the continuation Q of P is a (v, t)-path itself, and because the edge value $\overline{e}$ is decreasing, $a(P) = a_{uv} \odot a(Q) = \overline{e} \odot a(Q) \leqslant a(Q)$, so the value of the (v, t)-path Q is at least as large as that of P. We have thus replaced every (u, t)-path in the definition of a_{ut}^* with a corresponding (v, t)-path of at least the same value, and there may be further (v, t)-paths starting with edges to some $w \in N(v) \setminus N(u)$. It follows that $a_{vt}^* = a_{ut}^* \oplus c$ for some $c \in S$ and hence $a_{ut}^* \leqslant a_{vt}^*$.

Now consider the case of (u, u)-paths. The trivial (u, u)-path without edges has value $\overline{1}$ by definition. All other (u, u)-paths start with an edge $\{u, w\}$ and can be matched by (v, u)-paths as above by replacing or omitting this edge, which implies $a_{uu}^* \leqslant a_{vu}^* \oplus \overline{1}$.

Since the trivial (u, u)-path can be matched with the corresponding (v, v)-path, and all other (u, u)-paths can be matched with (v, v)-paths by substituting the first and last edge, we also have $a_{uu}^* \leqslant a_{vv}^*$. $\square$

The theorem now follows from the assumed monotonicity of quantification and summarisation. Recall that monotonicity of positional dominance is generally established via special treatment of the four entries involving the two reflexive relationships. In some cases, monotonicity of positional dominance is then a consequence of symmetry $a_{uv}^* = a_{vu}^*$.

The longest-path algebra based on semiring $(\mathbb{N}_0, \max, +, 0, 0)$ and edge value $\overline{e} = 1$ is an example of a path algebra that is not decreasing and does not preserve neighbourhood inclusion. This is, however, a desired outcome as it is consistent with the idea that a vertex should not be more central, if it is farther away from others. Note that an index based on longest paths also violates the star property.

Similarly, prefix sums of walk-generating functions are not decreasing for $\beta < 0$. This provides a formal argument for the intuition that Bonacich's power index [9] captures properties that are indeed different from those built into centrality indices.

To assert that the theorem applies to common centrality indices, we only have to verify that both quantification and summarisation are monotone as well, i.e., the canonical pre-order of the semiring is preserved in positional dominance which in turn is preserved in the centrality ranking.

The identity and values of generating functions for non-negative parameters, clearly, are monotone transforms.

Of the common indices, betweenness centrality might be the one for which preservation of the neighbourhood-inclusion pre-order is least expected. For the dyadic dependencies derived from the geodetic semiring first observe that $N(u) \subseteq N[v]$ implies $(\text{dist}(u,t), \sigma(u,t)) \leqslant (\text{dist}(v,t), \sigma(v,t))$ for all $t \in V$ by Lemma 3. Since distances and paths are symmetric in undirected graphs, $(\text{dist}(s,u), \sigma(s,u)) \leqslant (\text{dist}(s,v), \sigma(s,v))$ for all $s \in V$. For any given $s \in V$, consider now the sums in equation (3.1) defining $\delta(s,u)$ and $\delta(s,v)$, and fix any $t \in V$. If $\text{dist}(s,t) = \text{dist}(s,u) + \text{dist}(u,t)$, then $\text{dist}(s,t) = \text{dist}(s,v) + \text{dist}(v,t)$ as well because the semiring order implies that both distances involving v are at most as large as those involving u. The semiring order also implies $\sigma(s,u) \leqslant \sigma(s,v)$ and $\sigma(u,t) \leqslant \sigma(v,t)$, so that each $t \in V$ contributes at least as much to $\delta(s,v)$ as it does to $\delta(s,u)$.

To summarise positions in a single value, standard centrality indices use summation, selection of an extremum, or projection to a component. If no single component of a position is less than another's, then so is the summary.

An instructive boundary case is alter-based centrality [38], which is defined as $c_A(v) = \sum_{w \in N(v)} \deg(w)$ but can be re-written in terms of walks of length two, $c_A(v) = \sum_{t \in V} \omega_{vt,2}$. Quantification of the walk-generating function ($\beta = 1$, projection to $k = 2$) and summarisation (sum over all $t \in V$) are so restrictive that they compensate for the fact that the path algebra is not decreasing. However, the index does not even distinguish the centre of a star from the peripheral vertices.

A degenerate case of monotonicity is the entropy of a position vector as used in an index called path-transfer centrality [45]. Since the position vector is normalised to sum to 1, the transformation is not monotone and no position dominates another unless they are equal.

5 Discussion

We argued that common centrality indices share one intuition: A vertex that is more connected to more others in more direct ways may not wind up being considered less central.

As a formalisation of this intuition, we proposed decreasing path algebras and showed that indices monotone in the indirect relations obtained from such algebras preserve the neighbourhood-inclusion pre-order.

In recent independent work [6], monotonicity and additivity are identified as properties shared by many centrality indices. These are properties of the transformations that turn positions (nodal statistics in their terminology) into numbers. We argued, however, that the essence of centrality is in the appropriate definition of positions while taking monotonicity for granted and dismissing additivity as a requirement.

Figure 2 seems to indicate that neighbourhood inclusion is a weak requirement. In the almost regular example graph in the top row of Figure 1, no two vertices are comparable. On the other hand, this is precisely the reason why we see subtle differences in structural position detected by the various centrality indices.

Moreover, from the other example graph we may get the impression that the criterion is rather strong as there are many pairs of vertices comparable by neighbourhood inclusion,

so that the ranking obtained from any centrality index is largely pre-determined. Even more extreme are star graphs where every pair of vertices is comparable by neighbourhood inclusion with the centre dominating all others, which are equivalent. Consistent with Freeman's star property, no centrality index preserving the neighbourhood-inclusion pre-order may rank a peripheral vertex above the centre.

The class of graph for which the neighbourhood-inclusion pre-order is complete is much larger than the class of star graphs, though, and it is known under various names including threshold graphs [36]. They can be seen as prototypical core-periphery graphs in which neighbourhood-inclusion ranks vertices from the core down to the periphery. By definition, no two centrality indices that respect neighbourhood inclusion contradict each other on a threshold graph. This leads us to propose the following.

Proposition 4 *A vertex index is a centrality if and only if it preserves the neighbourhood-inclusion pre-order.*

Our criterion thus generalizes the star property, and incompleteness of the neighbourhood-inclusion pre-order, or distance from a threshold graph, becomes an indicator of the degree to which differences in centrality indices can be attributed to their particular definition.

We hinted at two indices for which neighbourhood inclusion is not preserved, one based on longest paths and the other Bonacich's β-centrality [9] with negative β. This actually is a desired outcome because there is an apparent mismatch with the above centrality intuition.

With regard to other axiomatic approaches to the characterization of centrality, we would like to point out that an inherent criterion such as neighbourhood inclusion is non-quantitative, relatively easy to test, and eliminates a number of the technical difficulties incurred previously. Axioms requiring an index to be monotone under edge addition or edge switching lead to complicated proofs and necessitate that a class of graphs on which an index is defined be closed under these operations. The comparison of vertices by neighbourhood inclusion subsumes these graph modifications in the same graph.

The unification achieved via decreasing path algebras also leads to generalisation. We can systematically construct new indices by specifying semiring elements, semiring operations for concatenation and aggregation of paths, and an edge value that can be shown to be decreasing.

While we restricted our attention to connected simple undirected graphs, the ideas generalise rather straightforwardly to other classes of graphs. Moreover, via generalisation of neighbourhood inclusion to positional dominance of indirect relations [15], our approach extends to multi-layer networks of any kind. In fact, we did not include the common requirement of automorphism invariance in the centrality proposition to allow for more general notions of homogeneity based, for instance, on attribute data.

References

[1] BATAGELJ, V. (1994) Semirings for social network analysis. *J. Math. Sociol.* **19**(1), 53–68.

[2] Bavelas, A. (1948) A mathematical model for group structures. *Human Organizations* **7**(3), 16–30.

[3] Bavelas, A. (1950) Communication patterns in task-oriented groups. *J. Acoust. Soc. Am.* **22**(6), 725–730.

[4] Beauchamp, M. A. (1965) An improved index of centrality. *Behav. Sci.* **10**, 161–163.

[5] Benzi, M. & Klymko, C. (2013) Total communicability as a centrality measure. *J. Complex Netw.* **1**(2), 124–149.

[6] Bloch, F., Jackson, M. O. & Tebaldi, P. (2016) Centrality measures in networks. *SSRN Electron. J.*. Available at: http://dx.doi.org/10.2139/ssrn.2749124.

[7] Boldi, P. & Vigna, S. (2014) Axioms for centrality. *Internet Math.* **10**(3–4), 222–262.

[8] Bonacich, P. (1972) Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**, 113–120.

[9] Bonacich, P. (1987) Power and centrality: A family of measures. *Am. J. Sociol.* **92**(5), 1170–1182.

[10] Borgatti, S. P. (2005) Centrality and network flow. *Soc. Netw.* **27**(1), 55–71.

[11] Borgatti, S. P., Carley, K. M. & Krackhardt, D. (2006) On the robustness of centrality measures under conditions of imperfect data. *Soc. Netw.* **28**(2), 124–136.

[12] Borgatti, S. P. & Everett, M. G. (2006) A graph-theoretic perspective on centrality. *Soc. Netw.* **28**(4), 466–484.

[13] Borgatti, S. P., Everett, M. G. & Johnson, J. C. (2013) *Analyzing Social Networks*. Sage, London.

[14] Botafogo, R. A., Rivlin, E. & Shneiderman, B. (1992) Structural analysis of hypertexts: Identifying hierarchies and useful metrics. *ACM Trans. Inform. Syst.* **10**(2), 142–180.

[15] Brandes, U. (2016) Network positions. *Methodological Innov.* **9**, 2059799116630650.

[16] Brandes, U. & Erlebach, T. (editors) (2005) *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science, Vol. 3418, Springer-Verlag, Berlin.

[17] Brandes, U., Robins, G., McCranie, A. & Wasserman, S. (2013) What is network science? *Netw. Sci.* **1**(1), 1–15.

[18] Costenbader, E. & Valente, T. W. (2003) The stability of centrality measures when networks are sampled. *Soc. Netw.* **25**(4), 283–307.

[19] Estrada, E. & Rodríguez-Velázquez, J. A. (2005) Subgraph centrality in complex networks. *Phys. Rev. E* **71**(5), 056103.

[20] Foldes, S. & Hammer, P. L. (1978) The Dilworth number of a graph. *Ann. Discrete Math.* **2**, 211–219.

[21] Freeman, L. C. (1977) A set of measures of centrality based on betweenness. *Sociometry* **40**(1), 35–41.

[22] Freeman, L. C. (1979) Centrality in social networks: Conceptual clarification. *Soc. Netw.* **1**(3), 215–239.

[23] Freeman, L. C. (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science*, Empirical Press, Vancouver, BC.

[24] Gondran, M. & Minoux, M. (2008) *Graphs, Diods and Semirings*. Springer-Verlag, Berlin.

[25] Guimera, R., Mossa, S., Turtschi, A. & Nunes Amaral, L. A. (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Nat. Acad. Sci. USA* **102**(22), 7794–7799.

[26] Hage, P. & Harary, F. (1995) Eccentricity and centrality in networks. *Soc. Netw.* **17**, 57–63.

[27] Hennig, M., Brandes, U., Pfeffer, J. & Mergel, I. (2012) *Studying Social Networks – A Guide to Empirical Research*. Campus, Frankfurt/New York.

[28] Junker, B. H., Koschützki, D. & Schreiber, F. (2006) Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics* **7**(219).

[29] Kadushin, C. (2011) *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford University Press, New York, NY.

[30] Katz, L. (1953) A new status index derived from sociometric analysis. *Psychometrika* **18**(1), 39–43.

[31] KITTI, M. (2016) Axioms for centrality scoring with principal eigenvectors. *Soc. Choice Welf.* **46**(3), 639–653.

[32] KOSCHÜTZKI, D., LEHMANN, K. A., PEETERS, L., RICHTER, S., TENFELDE-PODEHL, D. & ZLOTOWSKI, O. (2005) Centrality indices. In: U. Brandes & T. Erlebach (editors), *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science, Vol. 3418, Springer-Verlag, Berlin, pp. 16–61.

[33] KOSCHÜTZKI, D., LEHMANN, K. A., TENFELDE-PODEHL, D. & ZLOTOWSKI, O. (2005) Advanced centrality concepts. In: U. Brandes & T. Erlebach (editors), *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science, Vol. 3418, Springer-Verlag, Berlin, pp. 83–111.

[34] LANDHERR, A., FRIEDL, B. & HEIDEMANN, J. (2010) A critical review of centrality measures in social networks. *Bus. Inform. Syst. Eng.* **2**(6), 371–385.

[35] LEAVITT, H. J. (1951) Some effects of certain communication patterns on group performance. *J. Abnormal Soc. Psychol.* **46**(1), 38.

[36] MAHADEV, N. V. R. & PELED, U. N. (1995) *Threshold Graphs and Related Topics*, Annals of Discrete Mathematics, Vol. 56, North Holland, Amsterdam.

[37] MORENO, J. L. (1953) *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Beacon House, New York, NY. First published in 1934.

[38] NEAL, Z. (2013) A computationally efficient approximation of beta centrality. *Connections* **33**(1), 11–17.

[39] NIEMINEN, J. (1973) On the centrality in a directed graph. *Soc. Sci. Res.* **2**(4), 371–378.

[40] NIEMINEN, J. (1974) On the centrality in a graph. *Scand. J. Psychol.* **15**, 332–336.

[41] RUHNAU, B. (2000) Eigenvector-centrality—a node-centrality? *Soc. Netw.* **22**(4), 357–365.

[42] SABIDUSSI, G. (1966) The centrality index of a graph. *Psychometrika* **31**(4), 581–603.

[43] SRIVASTAVA, H. M. & MANOCHA, H. L. (1984) *A Treatise on Generating Functions.* John Wiley and Sons, New York.

[44] TODESCHINI, R. & CONSONNI, V. (2009) *Molecular Descriptors for Chemoinformatics*, 2nd ed., Wiley-VCH, Weinheim.

[45] TUTZAUER, F. (2007) Entropy as a measure of centrality in networks characterized by path-transfer flow. *Soc. Netw.* **29**(2), 249–265.

[46] VALENTE, T. W. & FOREMAN, R. K. (1998) Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Soc. Netw.* **20**(1), 89–105.

[47] VAN DEN BRINK, R. & GILLES, R. P. (2000) Measuring domination in directed networks. *Soc. Netw.* **22**(2), 141–157.

[48] WASSERMAN, S. & FAUST, K. (1994) *Social Network Aanalysis. Methods and Applications*, Cambridge University Press, Cambridge, UK.