




Properties and expressivity of linear geometric centralities ^{*}

Paolo Boldi  ^{*}, Flavio Furla  , Chiara Prezioso 

Dipartimento di Informatica, Università degli Studi di Milano, Italy

ARTICLE INFO

Keywords:

Network centrality
Centrality axioms
Linear transformation
Shortest paths
Farkas' lemma

ABSTRACT

Centrality indices are used to rank the nodes of a graph by importance: this is a common need in many concrete situations (social networks, citation networks, web graphs, for instance) and it was discussed many times in sociology, psychology, mathematics and computer science, giving rise to a whole zoo of definitions of centrality. Although they differ widely in nature, many centrality measures are based on shortest-path distances: such centralities are often referred to as *geometric*. Geometric centralities can use the shortest-path-length information in many different ways, but most of the existing geometric centralities can be defined as a linear transformation of the distance-count vector (that is, the vector containing, for every index t , the number of nodes at distance t).

In this paper we study this class of centralities, that we call *linear (geometric) centralities*, in their full generality. In particular, we look at them in the light of the axiomatic approach, and we study their expressivity: we show to what extent linear centralities can be used to distinguish between nodes in a graph, and how many different rankings of nodes can be induced by linear centralities on a given graph. The latter problem (which has a number of possible applications, especially in an adversarial setting) is solved by means of a linear programming formulation, which is based on Farkas' lemma, and is interesting in its own right.

1. Introduction and motivation

Network science is a discipline focused on analyzing network structures in order to obtain information that can be used to understand the underlying domain of the network itself. One key problem that network scientists often face is to determine the importance of nodes in a network: this is crucial for understanding the dynamics of the system and for making informed decisions based on the network's structure. Of course, the definition of importance is not unique, and it can vary depending on the context and the specific properties of the network being analyzed. Centralities have been defined and re-defined many times over the years, in different contexts and with different approaches, and they are used to rank nodes in a network by their importance [2].

In this work, we study indices that are often referred to as *geometric*, as they solely depend on how many nodes exist at every (shortest-path) distance. Notable examples of geometric centralities include in-degree, closeness [3,4], and harmonic [2] centrality. In particular, we propose and study a wide class of geometric measures that can be computed as linear combination of the number of nodes at every distance: these centralities will be referred to as *linear geometric centralities* (or just linear centralities, for short). The vast majority of the most used geometric centralities are linear (in particular, all centralities mentioned above are equivalent to some linear centrality).

^{*} This is an extended version of a paper presented at the 20th Workshop on Modelling and Mining Networks [1].

^{*} Corresponding author.

E-mail address: paolo.boldi@unimi.it (P. Boldi).

A special case of linear centralities are those assigning larger weights to shorter distances: the intuition here is that nearby connections are more relevant than long-range ones. These types of decaying centralities first appeared in Kishi [5] (albeit with the equivalent but opposite postulation that smaller centrality values denote a higher importance). In Kishi and Takeuchi [6] the same class was extended to directed graphs, and several of its properties were explored.

In more recent years, decaying centralities appeared in Skibski and Sosnowska [7], where they are called additive and studied through the so-called “axiomatic approach” (pioneered in Sabidussi [8]). Another example of this approach used to study these centralities can be found in Garg [9] and Boldi et al. [10]. In the latter, the impact of edge additions to the final scores and rankings yielded by these indices was investigated (obtaining, independently, results similar to those in Kishi [5]).

In the present paper, we relax the assumption of decaying weights, and study linear geometric centralities in their full generality: our aim is to analyze their properties, expressivity and to understand whether they capture a broader range of behaviors than decaying centralities. As it often happens, studying broader and more abstract classes of centralities allows us to understand better the properties of narrower classes. Our work focuses especially on expressivity, following the path opened, for instance, by Schoch and Brandes [11] in that it attempts to find general results for large sets of centralities.

In particular, the main goal of this work is to answer the following questions:

- what can we say about general properties of linear centralities as a function of the vector of coefficients? from an axiomatic perspective: what numerical properties of the vector are related to common axioms, e.g., those discussed in Boldi and Vigna [2]?
- given a pair of linear geometric centralities, is there always a graph on which they rank nodes differently? in other words: how different are linear centralities from one another?
- given a graph, what is the maximum number of rankings of its nodes that linear centralities can induce? how can we characterize precisely those rankings?

The latter problem is particularly interesting, as it has a number of possible applications, especially in an adversarial setting: for instance, if we want to design a graph so that all linear centralities will rank some nodes in a certain way (say, x is always ranked better than y , no matter what linear centrality is used), we can employ the results of this paper to determine whether it is possible to do so.

2. Notations and basic definitions

For every natural number n , we let $[n] = \{0, 1, \dots, n - 1\}$.

Graphs. A graph $G = (V_G, E_G)$ is given Bollobás [12] by a finite set of nodes V_G and¹ a set of arcs $E_G \subseteq V_G \times V_G$. Without loss of generality, we always assume that $V = [n]$ where n is the number of nodes.

We write $x \rightarrow y$ to mean that $x, y \in V$ and $(x, y) \in E$. A graph is *undirected* iff $x \rightarrow y$ implies $y \rightarrow x$. When drawing undirected graphs, pairs of opposite arcs are represented as a single undirected edge $x - y$.

A *path* of length k from $x \in V$ to $y \in V$ is a sequence (x_0, x_1, \dots, x_k) of nodes such that $x = x_0$, $y = x_k$ and $x_i \rightarrow x_{i+1}$ for all $i \in [k]$. The (*shortest path*) distance from x to y in G , denoted by $d_G(x, y)$, is the length of a shortest path from x to y , or ∞ if no path from x to y exists.

Note that in this paper, unless otherwise specified, we will consider *directed* graphs, where $d(x, y)$ and $d(y, x)$ can be different (it can even be the case that one is finite and the other is not!).

Graph homomorphisms. A graph homomorphism $\varphi : G \rightarrow H$ is a function $\varphi : V_G \rightarrow V_H$ such that $x \rightarrow_G y$ iff $\varphi(x) \rightarrow_H \varphi(y)$, for all $x, y \in V_G$. A bijective homomorphism $\varphi : G \rightarrow H$ is called an *isomorphism*; it is an *automorphism* if $G = H$. A graph with no non-trivial automorphism (i.e., its only automorphism is the identity) is called *rigid*.

Matrices and vectors. In this paper, we shall use column vectors; all vectors and matrices are conveniently indexed starting from 0. So, if $\mathbf{x} \in \mathbf{R}^n$ and $A \in \mathbf{R}^{m \times n}$ then

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,n-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-1,0} & a_{m-1,1} & \dots & a_{m-1,n-1} \end{bmatrix}$$

Infinite vectors and matrix product. If $\mathbf{a} \in \mathbf{R}^{\mathbb{N}}$, we view \mathbf{a} as an infinite vector of real numbers. If $A \in \mathbf{R}^{m \times \mathbb{N}}$, we write $A \cdot \mathbf{a}$ as a shortcut for

$$A \cdot \mathbf{a}[: n]$$

where $\mathbf{a}[: n]$ stands for the vector of the first n entries of \mathbf{a} .

¹ For this and similar notations, the subscript G is dropped whenever it is clear from the context.

Permutations. For every n , we let S_n be the symmetric group of all permutations (i.e., bijections) $\pi : [n] \rightarrow [n]$. The permutation matrix R_π is the binary $n \times n$ matrix such that

$$(R_\pi)_{ij} = 1 \text{ iff } \pi(i) = j.$$

Signum. We define $\text{sgn} : \mathbf{R} \rightarrow \mathbf{R}$ as the function defined by

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0. \end{cases}$$

3. Centralities, permutations and agreements

A (*graph*) *centrality* \mathfrak{f} [13] is a function associating with each graph G a map $\mathfrak{f}_G : V_G \rightarrow \mathbf{R}$ such that for any two graphs G, H , if $\varphi : G \rightarrow H$ is an isomorphism then

$$\mathfrak{f}_G(x) = \mathfrak{f}_H(\varphi(x))$$

for all $x \in V_G$.

While usually centralities are restricted to having non-negative values only, here we are going to relax this requirement and allow for negative centrality values. As usual, the value $\mathfrak{f}_G(x)$ is interpreted as a score indicating how central (i.e., important) node x is in G : larger values imply greater importance.

We say that \mathfrak{f} has *no ties* on G iff $\mathfrak{f}_G(x) \neq \mathfrak{f}_G(y)$ whenever $x, y \in V_G$ and $x \neq y$. Note that a centrality always has ties on non-rigid graphs: if $\phi : G \rightarrow G$ is an automorphism then by definition $\mathfrak{f}(\phi(x)) = \mathfrak{f}(x)$ for all x .

Here are some notable examples of centralities² (see, for instance, Boldi and Vigna [2] for further examples and taxonomy):

- in-degree: $\mathfrak{C}_G^{\text{in}}(x) = |\{y \mid y \rightarrow x\}|$;
- closeness [3]: $\mathfrak{C}_G^{\text{cl}}(x) = 1 / \sum_{y \in V, d(y,x) < \infty} d(y, x)$;
- Lin [14]: $\mathfrak{C}_G^{\text{Lin}}(x) = |\{y \in V \mid d(y, x) < \infty\}|^2 / \sum_{y \in V, d(y,x) < \infty} d(y, x)$;
- harmonic [2]: $\mathfrak{C}_G^{\text{harm}}(x) = \sum_{y \in V} 1/d(y, x)$, with the proviso that $1/\infty = 0$;
- Katz [15]: $\mathfrak{C}_G^{\text{Katz}}(x) = \sum_{\pi} \beta^{|\pi|}$, where the sum ranges over all paths π ending in x , and β is a parameter smaller than the reciprocal of the spectral radius of G ;
- betweenness [16]: $\mathfrak{C}_G^{\text{bet}}(x) = \sum_{y,z, \sigma_{yz} \neq 0} \sigma_{yz}(x) / \sigma_{yz}$, where the sum ranges over all pairs of nodes $y \neq z$, σ_{yz} is the number of shortest paths from y to z , and $\sigma_{yz}(x)$ is the number of such paths passing through x .

In most cases, more than the actual value a centrality assigns to each node, we are interested in how nodes are ranked with respect to a given centrality. This concept is captured formally by the following definition:

Definition 1 (Respecting a permutation). Given a graph G of n nodes, we say that \mathfrak{f} *respects* the permutation $\pi \in S_n$ on G iff $\mathfrak{f}_G(\pi(i)) \geq \mathfrak{f}_G(\pi(i+1))$ for all $i \in [n-1]$ (that is, π orders the nodes of G in non-increasing order of their centrality values with respect to \mathfrak{f}).

Note that if \mathfrak{f} has no ties on G then there is a unique permutation that \mathfrak{f} respects on G , called the *permutation induced* by \mathfrak{f} .

Definition 2 (Agreement between centralities). Two centralities \mathfrak{f} and \mathfrak{g} *agree* on a graph G iff they respect exactly the same permutations on G . (In particular, if they both have no ties on G , then they must induce the same permutation). Two centralities that agree on all graphs are called *equivalent*.

For example, the following two centralities:

$$\mathfrak{C}_G^{\text{cccl}}(x) = \frac{|V|}{\sum_{y \in V, d(y,x) < \infty} d(y, x)} \quad \text{“classical” closeness}$$

$$\mathfrak{C}_G^{\text{np}}(x) = - \sum_{y \in V, d(y,x) < \infty} d(y, x) \quad \text{negative peripherality}$$

are both equivalent to $\mathfrak{C}_G^{\text{cl}}$ (closeness). Classical closeness differs from closeness only by a positive multiplicative constant $|V|$, that has no impact on the ranking of nodes. For negative peripherality, instead of taking the reciprocal of the sum of distances to x , we take the *opposite* of the sum of distances: although the actual values are no doubt different, $\mathfrak{C}_G^{\text{np}}$ induces the same orders on the nodes as the one induced by $\mathfrak{C}_G^{\text{cl}}$.

² Some of these definitions were originally given for undirected graphs only. Here we are providing the adaptations for the general (i.e., directed) case, with the usual proviso that incoming paths are more interesting than outgoing paths.

4. Geometric and linear centralities

In the wide world of centralities, many depend only on the number of nodes at a(ny) given distance. As explained in Section 1, in this paper we focus on this class.

Definition 3 (Distance-count function). Given a graph G with n nodes, and a node $i \in V$, let the *distance-count function* of G at i be the function $c_{G,i} : \mathbf{N} \rightarrow \mathbf{N}$ defined by

$$c_{G,i}(k) = |\{j \in V : d(j, i) = k\}|,$$

that is, the number of nodes at distance k to i .

In Buckley and Harary [17], the first n values of the distance-count function at a node i are called *distance-degree sequence* of i : in fact, we do not need to take arbitrary distances into account, because $c_{G,i}(k) = 0$ for all $k \geq n$, because two nodes cannot have finite distance larger than $n - 1$. Moreover, note also that $c_{G,i}(0) = 1$ always, because there is exactly one node at distance 0 from i , that is, i itself. Differently from Kishi and Takeuchi [6], we shall not consider the distance-count function at infinity, i.e., we do not consider the number of nodes that cannot reach i .

A centrality is geometric if it is essentially dependent only on distance counts:

Definition 4 (Geometric centralities). A centrality f is *strictly geometric* iff for all graphs G and G' and all nodes $i \in V_G$ and $i' \in V_{G'}$ we have

$$c_{G,i} = c_{G',i'} \text{ implies } f_G(i) = f_{G'}(i').$$

A centrality f is *geometric* iff it is equivalent to one that is strictly geometric.

For instance, in-degree, closeness, Lin, negative peripheralty and harmonic are all strictly geometric: their value on a node x only depends on how many nodes are at each distance from x . Classical closeness ($\mathfrak{C}_G^{\text{cl}}$) is not strictly geometric: to compute its value on x you need to know also the number of nodes in the graph, which is not deducible only from the distance-count function unless the graph is strongly connected. Nonetheless, $\mathfrak{C}_G^{\text{cl}}$ is equivalent to closeness ($\mathfrak{C}_G^{\text{cl}}$), so $\mathfrak{C}_G^{\text{cl}}$ is geometric.

Conversely, Katz centrality is not geometric (because it does not depend on shortest paths, but on *all* paths), neither is betweenness (because it does not depend on the length of shortest paths, but on their number).

Geometric centralities depend only on distance counts, but they may do so in many different ways. We are now going to focus further our attention on the case where the dependence is linear. We begin by defining:

Definition 5 (Distance-count matrix). The *distance-count matrix* of a graph G of n nodes is the matrix $C = C_G \in \mathbf{R}^{n \times n}$ such that the i -th row of C is $c_{G,i}[\cdot : n]$. Said otherwise, $c_{i,k}$ is the number of nodes at distance k to i , for all $i, k \in [n]$.

Distance-degree sequences represented as matrices also appeared in Roginski et al. [18] with the name of *neighbor matrices*.

In Fig. 1, we show an example of distance-count matrix: node 0 (first row) has 1 node at distance 0 (itself), and $n - 1$ nodes at distance 1; node 1 (second row) has 1 node at distance 0, 1 node at distance 1 (node 0), and $n - 2$ nodes at distance 2; and so on.

Definition 6 (Linear centrality). Given $\mathbf{a} \in \mathbf{R}^N$ (the *coefficient vector*), the centrality $\mathfrak{Q}_G^{\mathbf{a}}$ is defined by

$$\mathfrak{Q}_G^{\mathbf{a}}(i) = (C_G \cdot \mathbf{a})_i.$$

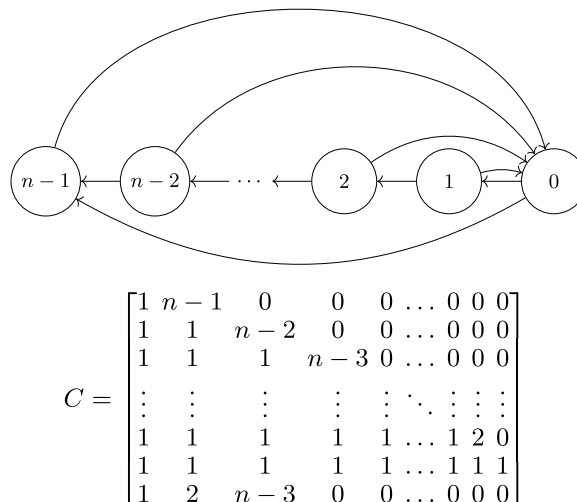


Fig. 1. A graph with n nodes, and its distance-count matrix C .

A centrality is *strictly linear (geometric)* if it is $\mathfrak{L}_G^{\mathbf{a}}$ for some $\mathbf{a} \in \mathbb{R}^N$. A centrality is *linear (geometric)* if it is equivalent to a strictly linear (geometric) centrality.

For instance,

- in-degree is strictly linear, with coefficients $(0, 1, 0, 0, \dots)^T$: its value on x is exactly the number of in-neighbors of x (i.e., nodes at distance 1 to x);
- negative peripherality is strictly linear, with coefficients $(0, -1, -2, -3, \dots)^T$;
- harmonic is strictly linear, with coefficients $(0, 1, 1/2, 1/3, 1/4, \dots)^T$;
- closeness and classic closeness are not strictly linear, but they are all linear, since they are all equivalent to negative peripherality.

Note that the value of a_0 is a constant additive factor that does not affect the ranking of nodes, so we can assume without loss of generality that $a_0 = 0$, unless otherwise stated.

A support for the computation of general linear centralities has been made available from within the WebGraph framework [19], a widely used library for the analysis of large graphs. The framework provides `LinearGeometricCentrality`, a class that can be used to compute any linear centrality, given the coefficient vector \mathbf{a} .

4.1. Lin is not linear

Note that a geometric centrality is not necessarily also linear. For example, Lin centrality is equivalent to negative peripherality only when the underlying graph is strongly connected, as demonstrated by the following counterexample.

Consider the disconnected undirected graph G in Fig. 2; if we look at the Lin centrality (refer to Section 3 for the formula) of nodes u, v, x , and y , we have that:

$$\mathfrak{L}_G^{\text{Lin}}(u) = \frac{8}{3} < \frac{9}{3} = \mathfrak{L}_G^{\text{Lin}}(v) \quad \text{and} \quad \mathfrak{L}_G^{\text{Lin}}(x) = \frac{50}{9} > \frac{49}{9} = \mathfrak{L}_G^{\text{Lin}}(y).$$

For Lin to be linear, we need to provide a strictly linear centrality to which it is equivalent, i.e., we need a coefficient vector \mathbf{a} such that $\mathfrak{L}_G^{\mathbf{a}}(u) < \mathfrak{L}_G^{\mathbf{a}}(v)$ and $\mathfrak{L}_G^{\mathbf{a}}(x) > \mathfrak{L}_G^{\mathbf{a}}(y)$. This means that we need a solution for the following system of inequalities:

$$\begin{cases} a_0 + a_1 + a_2 + a_3 < a_0 + a_1 + a_2 \\ a_0 + 3(a_1 + a_2 + a_3) > a_0 + 3(a_1 + a_2), \end{cases}$$

which obviously has no solution, as it requires both $a_3 < 0$ and $a_3 > 0$.

Therefore, there exists no linear centrality equivalent to Lin’s centrality on the graph in Fig. 2.

5. Centralities, ties and rigidity

By its very definition, every centrality (including, of course, all geometric centralities) has ties on all non-rigid graphs. We can say more:

Theorem 1. *All centrality measures have ties on graphs that are not rigid. On the other hand, there exists a centrality measure \mathfrak{f} such that, for all rigid graphs G , \mathfrak{f} has no ties on G .*

Proof. We exploit the lexicographic canonization process described in Babai and Luks [20]: let G be a rigid graph with n nodes, and A_π be the adjacency matrix of G when its nodes are permuted according to $\pi \in S_n$. Suppose that $A_\pi = A_\rho$ for some choice of permutations $\pi, \rho \in S_n$; then for all $x, y \in V_G$:

$$A_{\pi(x)\pi(y)} = A_{\rho(x)\rho(y)},$$

that is, $\pi(x) \rightarrow_G \pi(y)$ iff $\rho(x) \rightarrow_G \rho(y)$. But then $\rho^{-1} \circ \pi$ would be an automorphism of G , hence $\pi = \rho$. So, the matrices A_π are all distinct; hence, there is a unique π such that A_π is minimal in lexicographic order (we can order matrices lexicographically by considering the string obtained concatenating all of their rows). Now you can define \mathfrak{f}_G to be such a permutation $\pi : V_G \rightarrow [n]$: since π is a permutation, by definition it is injective. \square

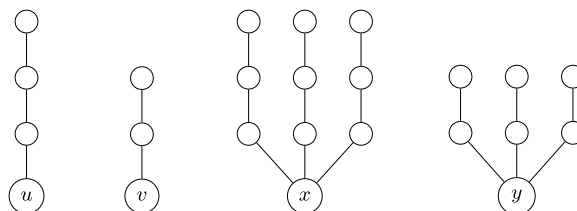


Fig. 2. A disconnected undirected graph showing that Lin centrality is not linear.

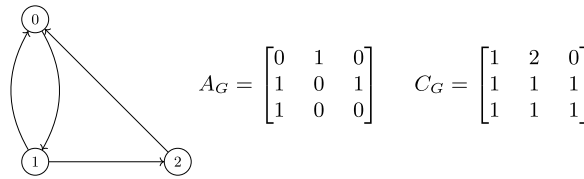


Fig. 3. A rigid graph that is not geometrically rigid, its adjacency matrix A_G and its distance-count matrix C_G .

Geometric centralities, though, are more limited and may give rise to unsolvable ties even on rigid graphs. For instance, the graph in Fig. 3 has no non-trivial automorphisms, so according to Theorem 1 there is a centrality that produces no ties on it; nonetheless, all geometric centralities give the same value to nodes 1 and 2, because the corresponding rows in its distance-count matrix C_G are the same. Let us provide a formal definition of this limitation:

Definition 7 (Geometrically rigid). A graph G is *geometrically rigid* if the rows of its distance-count matrix C_G are all distinct.

In Buckley and Harary [17], geometrically rigid graphs are also called *distance-degree injective*, in opposition to *distance-degree regular* graphs, i.e., those with all nodes having the same distance-degree sequence.

Clearly, every geometrically rigid graph is also rigid, but not the other way round, as the graph in Fig. 3 shows. Nonetheless, we can relate the two notions as follows:

Theorem 2. All geometric centrality measures have ties on graphs that are not geometrically rigid. On the other hand, there exists a geometrical centrality measure \dagger such that, for all geometrically rigid graphs G , \dagger has no ties on G .

Proof. Define $\dagger_G(x)$ to be the rank of x in the lexicographic order of the rows of C_G . Since by definition the rows of C_G are all distinct, the lexicographic order has no ties. On the other hand, if two rows of C_G are the same, the corresponding nodes will have the same value for all geometric centralities (because, by definition, geometric centralities are a function of the distance-count function of the node). \square

A similar statement can be proven also for linear centralities; in particular, we can show that there is always a linear centrality with no ties on geometrically rigid graphs:

Theorem 3. There exists a choice of coefficients $\mathbf{a} \in \mathbb{R}^N$ such that, for all graphs G , if G is geometrically rigid then $\mathfrak{Q}_G^{\mathbf{a}}$ has no ties on G .

Proof. Define $\mathbf{a} = (\log p_0, \log p_1, \log p_2, \dots)^T$, where p_i is the i -th prime number. Then,

$$\mathfrak{Q}_G^{\mathbf{a}}(i) = \sum_j c_{i,j} \log p_j = \log \left(\prod_j p_j^{c_{i,j}} \right).$$

Since $\log(x) = \log(y)$ iff $x = y$, $\mathfrak{Q}_G^{\mathbf{a}}$ produces a tie iff there are two distinct nodes i and i' such that the products on the right-hand side are equal; clearly, these products are prime factorizations, and this would imply that $c_{i,j} = c_{i',j}$ for all j , contradicting the initial hypothesis that G is geometrically rigid. \square

6. Axioms for linear centralities

While many special instances of linear centralities were studied before, adopting our broader view may be helpful to get a deeper understanding of the most general conditions underlying their properties. As an example in this direction, following a quite common approach, we will consider the axioms described in Boldi and Vigna [2] and determine under which general circumstances a linear centrality satisfies them.

6.1. Size and density axioms

The first two axioms that we are going to study are related to two specific (families of) graphs.

For given positive integers k, p , consider the graphs S and S_{xy} defined in Fig. 4. Recall that Boldi and Vigna [2]:

Definition 8 (Density axiom). We say that a centrality \dagger satisfies the *density axiom* iff for every fixed $k \geq 3$ and $p = k$, $\dagger_{S_{xy}}(x) > \dagger_{S_{xy}}(y)$.

Definition 9 (Size axiom). We say that a centrality \dagger satisfies the *size axiom* iff for every fixed p , there exists k_0 such that for all $k \geq k_0$, $\dagger_S(x) > \dagger_S(y)$, and for every fixed k , there exists p_0 such that for all $p \geq p_0$, $\dagger_S(x) < \dagger_S(y)$.

Given a coefficient vector \mathbf{a} , the geometric linear centralities of x and y in the two graphs S and S_{xy} are given by:

$$\begin{aligned} \mathfrak{Q}_{S_{xy}}^{\mathbf{a}}(x) &= a_0 + k \cdot a_1 + \sum_{i=2}^p a_i & \mathfrak{Q}_S^{\mathbf{a}}(x) &= a_0 + (k - 1) \cdot a_1 \\ \mathfrak{Q}_{S_{xy}}^{\mathbf{a}}(y) &= a_0 + 2 \cdot a_1 + (k - 1) \cdot a_2 + \sum_{i=2}^{p-1} a_i & \mathfrak{Q}_S^{\mathbf{a}}(y) &= \sum_{i=0}^{p-1} a_i. \end{aligned}$$

For the density axiom, we have that:

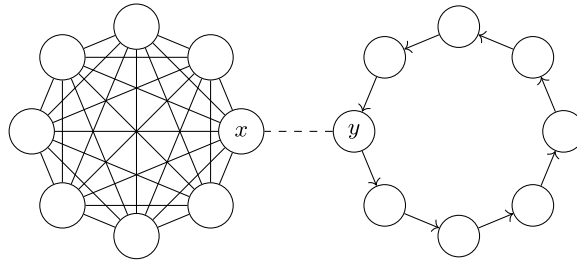


Fig. 4. A k -clique (left) and a directed p -cycle (right), possibly connected through a bridge $x - y$. We let S be the disconnected graph without the bridge and S_{xy} the strongly connected graph with the bridge.

Proposition 1. Ω^a satisfies the density axiom iff:

$$a_k > k(a_2 - a_1) + (2a_1 - a_2)$$

for all $k \geq 3$.

Proof. We start observing that

$$\Omega_{S_{xy}}^a(x) - \Omega_{S_{xy}}^a(y) = (k - 2)a_1 + (1 - k)a_2 + a_p = k(a_1 - a_2) + (a_2 - 2a_1) + a_p.$$

So, the condition

$$\Omega_{S_{xy}}^a(x) > \Omega_{S_{xy}}^a(y)$$

for $p = k$ is equivalent to

$$a_k > k(a_2 - a_1) + (2a_1 - a_2)$$

which ends the proof. \square

We now turn to the size axiom:

Proposition 2. Ω_G^a satisfies the size axiom iff $a_1 > 0$ and $\sum_i a_i$ diverges.

Proof. The size axiom requires that

$$\Omega_S^a(x) > \Omega_S^a(y) \tag{1}$$

holds ultimately (in k) for every fixed p , whereas

$$\Omega_S^a(x) < \Omega_S^a(y) \tag{2}$$

should hold ultimately (in p) for every fixed k . This time we have:

$$\Omega_S^a(x) - \Omega_S^a(y) = a_0 + (k - 1)a_1 - \sum_{i=0}^{p-1} a_i.$$

For (1) to hold, we must have that for every fixed p

$$(k - 1)a_1 > \sum_{i=1}^{p-1} a_i \tag{3}$$

holds for sufficiently large k . If $a_1 < 0$, there is an upper bound on the largest k for which (3) can hold:

$$k < \frac{\sum_{i=1}^{p-1} a_i}{a_1} + 1.$$

So we must have $a_1 \geq 0$. For (2) to hold, we must have that for every fixed k

$$(k - 1)a_1 < \sum_{i=1}^{p-1} a_i \tag{4}$$

holds for sufficiently large p . If $a_1 = 0$, this condition is incompatible with (3).

We conclude that $a_1 > 0$. On the other hand, when $a_1 > 0$, (4) implies the statement. \square

Power-law and exponential decay centralities [10] are special cases of linear centralities; the former have coefficients of the form $a_i = 1/i^\gamma$ for some $\gamma > 0$ and $i > 0$, with $\gamma = 1$ corresponding to harmonic centrality. The latter, instead, have coefficients of the form $a_i = \delta^i$ for some $\delta \in (0, 1)$ and $i > 0$; in both cases, as usual, we have $a_0 = 0$. From the properties above we can conclude that:

Corollary 1. *Power-law decay centralities always satisfy the density axiom, and they satisfy the size axiom iff $\gamma \leq 1$.*

Proof. The only non-trivial part of the proof is showing that the density axiom holds.

We have that $a_k = k^{-\gamma}$, for some $\gamma > 0$. Consider the inequality of [Proposition 1](#), that in this case becomes:

$$k^{-\gamma} > k(2^{-\gamma} - 1) + 2 - 2^{-\gamma}. \tag{5}$$

When $k = 3$, the right-hand side is $2^{1-\gamma} - 1$, which is negative or null for $\gamma \geq 1$ (and a fortiori for larger values of k , because the coefficient of k itself is negative), so (5) certainly holds for all $k \geq 3$. We are left to consider the case $\gamma \in (0, 1)$. Let us look at the function

$$f(k) = k^{-\gamma} - k(2^{-\gamma} - 1) - 2 + 2^{-\gamma}.$$

for $\gamma \in (0, 1)$. Since

$$f'(k) = -\gamma k^{-(\gamma+1)} - (2^{-\gamma} - 1),$$

the only zero of the derivative $f'(k)$ is

$$k_0 = \exp\left(-\frac{\log\left(\frac{1-2^{-\gamma}}{\gamma}\right)}{\gamma+1}\right) = \left(\frac{\gamma}{1-2^{-\gamma}}\right)^{\frac{1}{\gamma+1}}.$$

Since $f(1) = f(2) = 0$, by Rolle's theorem we have that $k_0 \in (1, 2)$.

As a result, f is monotonic for $k \geq 3$; its value at 3 is:

$$f(3) = 3^{-\gamma} - 3(2^{-\gamma} - 1) - 2 + 2^{-\gamma} = 3^{-\gamma} - 2^{1-\gamma} + 1 > 0.$$

On the other hand,

$$\lim_{k \rightarrow \infty} f(k) = \infty.$$

So the function $f(k)$ is increasing, hence positive, for $k \geq 3$. \square

Corollary 2. *Exponential decay centralities always satisfy the density axiom, and never satisfy the size axiom.*

Proof. Once again, the only non-trivial part of the proof is showing that the density axiom holds. The inequality of [Proposition 1](#) in this case is:

$$\delta^k > k(\delta^2 - \delta) + 2\delta - \delta^2;$$

thus, we want to show that

$$f(k) = \delta^k + k(\delta - \delta^2) + \delta^2 - 2\delta$$

is positive for all $\delta \in (0, 1)$ and $k \geq 3$. The derivative of the function is

$$f'(k) = \delta^k \log \delta + \delta - \delta^2;$$

let us call k_0 the value in which $f'(k)$ is null, that is,

$$k_0 = \frac{\log\left(\frac{\delta^2 - \delta}{\log(\delta)}\right)}{\log(\delta)}.$$

Noting that $f(1) = f(2) = 0$, we can state by Rolle's theorem that $k_0 \in (1, 2)$, since it is the only stationary point of the function f . Finally, we highlight that

$$f(3) = \delta^3 + 3(\delta - \delta^2) + \delta^2 - 2\delta = \delta^3 - 2\delta^2 + \delta = \delta(\delta - 1)^2 > 0,$$

and

$$\lim_{k \rightarrow \infty} f(k) = \infty.$$

Again, the function is increasing, hence positive, for $k \geq 3$. \square

In particular, among power-law and exponential decay centralities, harmonic has the property of satisfying both axioms, and it is the one doing so with the fastest possible decay.

6.2. Score monotonicity

We now consider the score monotonicity axiom [2], that is related to what happens to the centrality of node y when an arc $x \rightarrow y$ is added to an arbitrary graph.

Definition 10 (Score monotonicity). A centrality measure is *score monotone* if for every graph G and every pair of nodes x, y such that $x \not\rightarrow y$, when we add $x \rightarrow y$ to G the centrality of y increases.

To characterize score monotonicity, it is useful to introduce (as we did in Boldi et al. [10]) the *discrete derivative operator* Δ on sequences: given a sequence \mathbf{a} , we define $\Delta \mathbf{a}$ as the sequence

$$(\Delta \mathbf{a})_i = a_{i+1} - a_i$$

for all indices $i > 0$, and $(\Delta \mathbf{a})_0 = 0$. It is also useful to write $\mathbf{a} \geq 0$ ($\mathbf{a} \leq 0$, respectively) to mean that $a_i \geq 0$ ($a_i \leq 0$, resp.) for all i .

Score monotonicity for linear centrality can then be characterized by the following:

Proposition 3. Given a coefficient vector \mathbf{a} , the centrality $\mathfrak{L}^{\mathbf{a}}$ is score monotone iff $\mathbf{a} \geq 0$, $\Delta \mathbf{a} \leq 0$ and $(\Delta \mathbf{a})_1 < 0$.

Proof. The condition can be restated as follows: $a_1 > a_2$ and $a_i \geq a_{i+1} \geq 0$ for all i . Given a graph G and a pair of nodes x, y such that $x \not\rightarrow y$, let us call G' the graph obtained from G by adding the arc $x \rightarrow y$. Let us assume that $a_1 > a_i \geq a_{i+1} \geq 0$ for all $i \geq 2$, and call $d'(z) = d_{G'}(z, y)$ and $d(z) = d_G(z, y)$ for all $z \in V_G$. Adding an arc in a graph cannot increase distances, that is:

$$d'(z) \leq d(z) \text{ for all } z \in V_G,$$

which implies that $a_{d'(z)} \geq a_{d(z)}$, with the two inequalities being strict when $z = x$. Noting that linear centralities can be expressed also as summation over nodes, we can state that:

$$\mathfrak{L}_{G'}^{\mathbf{a}}(y) = \sum_{z \in V_G} a_{d'(z)} > \sum_{z \in V_G} a_{d(z)} = \mathfrak{L}_G^{\mathbf{a}}(y),$$

where, again, the inequality is strict because it is so when $z = x$.

For the “only if” part of the proof, we consider the following cases, and provide for each of them an example on which $\mathfrak{L}^{\mathbf{a}}$ fails to be score monotone.

(a) *Case* $a_1 \leq 0$. Suppose $a_1 \leq 0$ and consider the following graph G : We would then have

$$\mathfrak{L}_{G'}^{\mathbf{a}}(y) = a_1 \leq 0 = \mathfrak{L}_G^{\mathbf{a}}(y).$$

(b) *Case* $a_i < 0$. Suppose $a_1 > 0$ but $a_i < 0$ for some $i > 1$. Let i be the smallest such index, and k be defined by

$$k = \left\lceil \frac{a_1 + \dots + a_{i-1} + 1}{|a_i|} \right\rceil$$

and consider the following graph G : We have that:

$$\begin{aligned} \mathfrak{L}_{G'}^{\mathbf{a}}(y) &= a_1 + a_2 + \dots + a_{i-1} + k \cdot a_i = \\ &= a_1 + a_2 + \dots + a_{i-1} - k \cdot |a_i| \leq \\ &\leq a_1 + a_2 + \dots + a_{i-1} - (a_1 + \dots + a_{i-1} + 1) < 0 = \mathfrak{L}_G^{\mathbf{a}}(y). \end{aligned}$$

(c) *Case* $a_i < a_{i+1}$. To complete our proof, suppose that $a_1 > 0$ and $a_i \geq 0$ for all $i > 1$, but that $a_i < a_{i+1}$ for some $i \geq 1$; let i be the smallest such index i , and k be defined by:

$$k = \left\lceil \frac{a_1 + 1}{a_{i+1} - a_i} \right\rceil$$

Consider the following graph G . We have:

$$\begin{aligned} \mathfrak{L}_G^{\mathbf{a}}(y) &= a_1 + a_2 + \dots + k \cdot a_{i+1} \\ \mathfrak{L}_{G'}^{\mathbf{a}}(y) &= 2a_1 + a_2 + \dots + k \cdot a_i. \end{aligned}$$

Hence

$$\mathfrak{L}_G^{\mathbf{a}}(y) - \mathfrak{L}_{G'}^{\mathbf{a}}(y) = -a_1 + a_i + k(a_{i+1} - a_i) > 0,$$

that is, $\mathfrak{L}_{G'}^{\mathbf{a}}(y) < \mathfrak{L}_G^{\mathbf{a}}(y)$. \square

6.3. Rank monotonicity

The rank monotonicity axiom [21] considers the same scenario as for score monotonicity, but this time looking at what happens to the relation between the centrality of y and that of *other* nodes.

Definition 11 (Rank monotonicity). A centrality measure is *rank monotone*³ if for every graph G and every pair of nodes x, y such that $x \not\rightarrow y$, when we add $x \rightarrow y$ to G the following happens for all nodes u :

³ This is called “strict” rank monotonicity in Boldi et al. [21], but in this paper we are not going to treat the weak case.

- if the score of $w \neq y$ was smaller than or equal to the score of y , after adding $x \rightarrow y$ the score of w is smaller than the score of y .

The first property we prove is that score monotonicity is necessary for rank monotonicity:

Lemma 1. *Given a coefficient vector \mathbf{a} , if $\Omega^{\mathbf{a}}$ is rank monotone, then it is score monotone.*

Proof. Suppose, by contradiction, that $\Omega^{\mathbf{a}}$ is rank monotone but not score monotone, and let G be a graph for which score monotonicity fails. Let \bar{G} be an identical copy of G (the nodes of \bar{G} have the same names as those of G , with an $\bar{\cdot}$ to distinguish them). For the moment, assume that $a_1 > 0$, and consider the graph:

$$H = G \oplus \bar{G}$$

where \oplus denotes disjoint union. Of course

$$\Omega_H^{\mathbf{a}}(y) = \Omega_G^{\mathbf{a}}(y) = \Omega_{\bar{G}}^{\mathbf{a}}(\bar{y}) = \Omega_H^{\mathbf{a}}(\bar{y}).$$

If we add $x \rightarrow y$ to the graph H , without any other modification, and let H' be the new graph, we have (because score monotonicity does not hold):

$$\Omega_{H'}^{\mathbf{a}}(y) = \Omega_{G'}^{\mathbf{a}}(y) \leq \Omega_G^{\mathbf{a}}(y) = \Omega_H^{\mathbf{a}}(\bar{y}).$$

So the score of y in H used to be \geq (in fact: equal) to the score of \bar{y} , but it should become strictly higher once the arc $x \rightarrow y$ is added, which does not happen if score monotonicity fails. \square

Hence, if we want to look for vectors \mathbf{a} for which rank monotonicity holds, we can limit ourselves to those for which score monotonicity holds as well.

Like in Boldi et al. [10], we will show that rank monotonicity is related to the second-order discrete derivative $\Delta^2 \mathbf{a}$ of the coefficient vector \mathbf{a} (obtained from \mathbf{a} by applying the operator Δ twice). To characterize rank monotonicity, we will need the following lemma:

Lemma 2 (Lemma 5.1 of Boldi et al. [10]). *Given a coefficient vector \mathbf{a} , if $\Delta^2 \mathbf{a} \geq 0$ then for all i, j, k, ℓ with $i \leq j$, $k \geq 0$, $0 \leq \ell \leq j - i$ we have:*

$$a_{i+k} - a_{j+k-\ell} \leq a_i - a_j$$

with the inequality being strict if $i < j$, $k > 0$ and $(\Delta^2 \mathbf{a})_i > 0$.

We now present our conditions for rank monotonicity:

Proposition 4. *Suppose that the coefficient vector \mathbf{a} satisfies the following conditions:*

1. $\Omega^{\mathbf{a}}$ is score monotone, i.e., $\mathbf{a} \geq 0$, $\Delta \mathbf{a} \leq 0$ and $(\Delta \mathbf{a})_1 < 0$;
2. $\Delta^2 \mathbf{a} \geq 0$ and $(\Delta^2 \mathbf{a})_1 > 0$.

Then, the centrality $\Omega^{\mathbf{a}}$ is rank monotone.

Proof. Suppose that the two conditions hold, let G be a graph and G' the graph obtained after the addition of $x \rightarrow y$ to G , as usual. Let us use $d(-, -)$ and $d'(-, -)$ to denote distances in G and G' , respectively.

Consider any two nodes z and w : of course $d'(z, w) \leq d(z, w)$ (because adding an arc cannot make distances larger); suppose that $d'(z, w) < d(z, w)$. This means that there is a new shortest path from z to w using the new arc $x \rightarrow y$. Hence

$$d'(z, w) = d(z, x) + 1 + d(y, w) \geq d'(z, y) + d(y, w). \tag{6}$$

On the other hand, because of the triangular inequality

$$d(z, w) \leq d(z, y) + d(y, w); \tag{7}$$

hence, applying (6) first and then (7), we have

$$d'(z, y) \leq d'(z, w) - d(y, w) \leq d'(z, w) - d(z, w) + d(z, y). \tag{8}$$

Let

$$\begin{aligned} i &= d'(z, y) \\ j &= d(z, y) \\ k &= d'(z, w) - i = d'(z, w) - d'(z, y) \\ \ell &= j + k - d(z, w) = d'(z, w) - d(z, w) + d(z, y) - d'(z, y). \end{aligned}$$

Let us check that all conditions of Lemma 2 hold: $i \leq j$ holds because distances cannot become larger; $k \geq 0$ holds because of (6); finally $\ell \geq 0$ (because of (8)) and

$$\ell + i - j = i + k - d(z, w) = d'(z, w) - d(z, w) \leq 0;$$

therefore, $\ell \leq j - i$. Hence, applying the lemma we have:

$$a_{d'(z,w)} - a_{d(z,w)} \leq a_{d'(z,y)} - a_{d(z,y)}. \tag{9}$$

This relation holds for all z such that $d'(z, w) < d(z, w)$, but it is trivially true also when $d'(z, w) = d(z, w)$, because the sequence \mathbf{a} is non-increasing.

Moreover, when $z = x$ and $w \neq y$, if $d'(x, w) < d(x, w)$ we have $i = d'(x, y) = 1$, $j = d(x, y) > 1 = i$, $k = d'(x, w) - 1 > 0$. Hence, the inequality (9) holds strictly for $z = x$. If otherwise $d'(x, w) = d(x, w)$, then still the inequality holds strictly, because $a_1 > a_2 \geq a_j$, since $(\Delta \mathbf{a})_1 < 0$.

Recalling that

$$\mathfrak{Q}_G^{\mathbf{a}}(w) = \sum_z a_{d(z,w)} \quad \mathfrak{Q}_{G'}^{\mathbf{a}}(w) = \sum_z a_{d'(z,w)},$$

we can state the following for all $w \neq y$:

$$\mathfrak{Q}_{G'}^{\mathbf{a}}(y) - \mathfrak{Q}_G^{\mathbf{a}}(y) = \sum_z (a_{d'(z,y)} - a_{d(z,y)}) > \sum_z (a_{d'(z,w)} - a_{d(z,w)}) = \mathfrak{Q}_{G'}^{\mathbf{a}}(w) - \mathfrak{Q}_G^{\mathbf{a}}(w).$$

This means that the addition of the arc $x \rightarrow y$ increases the score of y more than the score of any other node w , and this is enough to conclude that rank monotonicity holds for $\mathfrak{Q}^{\mathbf{a}}$. \square

Observe that the conditions of Proposition 4 are sufficient for rank monotonicity, but not necessary (although score monotonicity is). We conjecture that they are also necessary: a clue in this direction is given in Boldi et al. [10], where convexity is proven to be equivalent in the undirected case to a property known as ‘‘basin dominance’’, which implies rank monotonicity. But also in that case we were unable to prove that convexity is necessary for rank monotonicity.

7. Distinguishable coefficients

A natural question we may want to ask is the following: how does $\mathfrak{Q}_G^{\mathbf{a}}$ change as a function of \mathbf{a} ? Is it possible that using different coefficients we end up defining equivalent linear centralities? The short answer is no, unless the coefficient vectors differ by a positive multiplicative constant. Let us provide a specific definition for this case.

Definition 12 (Proportionality). We say that $\mathbf{a}, \mathbf{b} \in \mathbf{R}^N$ are *proportional* iff there is a $\lambda > 0$ such that

$$b_k = \lambda \cdot a_k \text{ for all } k > 0. \tag{10}$$

Observe that in the definition the 0-th coefficient has no role. For this reason, and for what we have said before, in the rest of this section we can safely assume that the 0-th coefficient of all sequences is 0. Some observations are in order:

- if \mathbf{a} and \mathbf{b} are proportional then for all $k > 0$, either a_k and b_k are both zero, or they are both non-zero (and have the same sign);
- no sequence is proportional to $\mathbf{0} = (0, 0, 0, \dots)$, except $\mathbf{0}$ itself;
- if \mathbf{a} and \mathbf{b} are proportional, and $\mathbf{a} \neq \mathbf{0}$, then the constant $\lambda > 0$ of Definition 12 is unique;
- if \mathbf{a} and \mathbf{b} are not proportional, one of the following mutually exclusive statements holds:
 - (a) there is some $k > 0$ such that $a_k b_k = 0$ and $a_k \neq b_k$ (for example, $\mathbf{a} = (0, 1, 2, \dots)$ and $\mathbf{b} = (0, 1, 0, \dots)$, with $k = 2$);
 - (b) for all $i > 0$, $a_i = 0 \iff b_i = 0$ and the minimum index h such that $a_h \neq 0$ satisfies $b_h = \lambda \cdot a_h$ for some $\lambda < 0$ (for example, $\mathbf{a} = (0, 0, 2, 3, 0, \dots)$ and $\mathbf{b} = (0, 0, -4, -6, 0, \dots)$, with $h = 2$ and $\lambda = -2$);
 - (c) for all $i > 0$, $a_i = 0 \iff b_i = 0$ and there exists an index $k > 1$ and $\lambda > 0$ such that $b_i = \lambda \cdot a_i$ for all $0 < i < k$, but $b_k \neq \lambda \cdot a_k$, and moreover $a_h \neq 0$ for some $0 < h < k$ (for example, $\mathbf{a} = (0, 1, 2, 3, 0, \dots)$ and $\mathbf{b} = (0, 2, 4, 9, 0, \dots)$, with $k = 3$ and $\lambda = 2$).

The only non-trivial observation is the last one, which we are now going to prove. If there is any index for which either sequence is non-zero and the other is zero (case (a) above), there cannot exist a $\lambda > 0$ for Definition 12. Now, let us assume that the zeroes are aligned (i.e., that they appear in the same positions in both sequences): there must be an index at which the sequences are both non-zero (for otherwise both sequences would be $\mathbf{0}$ and proportional). Consider the minimum index h at which they are both non-zero, and let $\lambda = a_h/b_h$. If $\lambda < 0$, then case (b) of the last observation holds. If $\lambda > 0$, since the sequences are non-proportional there must be a larger index k such that $a_k \neq \lambda b_k$; the smaller such $k > h$ makes the statement of case (c) above true.

It is worth noting that case (c) above has some interesting implications:

Lemma 3. Given \mathbf{a} and \mathbf{b} , if case (c) of the previous observation holds, then the values a_h, a_k, b_h, b_k are all non-zero and $b_h a_k \neq b_k a_h$. Moreover, the inequality

$$(a_h - a_k)(b_{h+1} - b_{k-1}) \neq (a_{h+1} - a_{k-1})(b_h - b_k) \tag{11}$$

holds iff $a_{h+1} \neq a_{k-1}$ and $h + 1 < k$.

Proof. We start by proving the first part of the statement. By assumption $a_h \neq 0$ and $b_h = \lambda a_h \neq 0$. Moreover, either a_k and b_k are both zero, or none of them is; if they are, then $b_k = \lambda a_k$, contradicting the hypothesis. Finally,

$$b_h a_k = \lambda a_h a_k \neq a_h b_k.$$

For the second part of the statement, assume that $a_{h+1} \neq a_{k-1}$ and $h + 1 < k$, and, by contradiction, consider the following chain of equalities:

$$\begin{aligned} (a_h - a_k)(b_{h+1} - b_{k-1}) &= (a_{h+1} - a_{k-1})(b_h - b_k) \\ (a_h - a_k)(\lambda a_{h+1} - \lambda a_{k-1}) &= (a_{h+1} - a_{k-1})(\lambda a_h - b_k) \\ \lambda a_h a_{h+1} - \lambda a_h a_{k-1} - \lambda a_{h+1} a_k + \lambda a_{k-1} a_k &= \lambda a_h a_{h+1} - a_{h+1} b_k - \lambda a_h a_{k-1} + a_{k-1} b_k \\ -\lambda a_{h+1} a_k + \lambda a_{k-1} a_k &= -a_{h+1} b_k + a_{k-1} b_k \\ \lambda a_k (a_{k-1} - a_{h+1}) &= b_k (a_{k-1} - a_{h+1}) \\ \lambda a_k &= b_k, \end{aligned}$$

contradicting the assumptions. For the “only if” part, if $h + 1 = k$, we obtain

$$(a_h - a_k)(b_k - b_h) = (a_k - a_h)(b_h - b_k),$$

contradicting (11). On the other hand, if $h + 1 < k$ and $a_{h+1} = a_{k-1}$, then we have $a_{h+1} - a_{k-1} = 0$ but also $b_{h+1} - b_{k-1} = \lambda(a_{h+1} - a_{k-1}) = 0$, again contradicting (11). □

Based on these observations, we are ready to prove the following:

Theorem 4. For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$ the following holds:

1. if \mathbf{a} and \mathbf{b} are proportional, then $\mathfrak{Q}^{\mathbf{a}}$ and $\mathfrak{Q}^{\mathbf{b}}$ are equivalent centralities;
2. if \mathbf{a} and \mathbf{b} are not proportional, then there exists a graph G on which $\mathfrak{Q}_G^{\mathbf{a}}$ and $\mathfrak{Q}_G^{\mathbf{b}}$ do not agree; in other words, there is a graph G and two nodes $x, y \in V$ such that $\mathfrak{Q}_G^{\mathbf{a}}(x) \geq \mathfrak{Q}_G^{\mathbf{a}}(y)$ but $\mathfrak{Q}_G^{\mathbf{b}}(x) < \mathfrak{Q}_G^{\mathbf{b}}(y)$.

Proof. Given a graph $G = (V, E)$, if \mathbf{a} and \mathbf{b} are proportional, then $\mathfrak{Q}_G^{\mathbf{b}}(x) = \lambda \cdot \mathfrak{Q}_G^{\mathbf{a}}(x)$ for some $\lambda > 0$ and all $x \in V$, and it is easy to see that in such a case the two centralities are equivalent.

Let us now assume that \mathbf{a} and \mathbf{b} are not proportional, and look at the observations before Lemma 3. If we are in case (a), and k is the first index at which either sequence is zero and the other is not, let us consider a graph made of two bidirectional paths of length $k - 1$ and k , respectively. Assuming w.l.o.g. that $b_k \neq 0$, if $b_k > 0$ the statement is proven by letting x (y , respectively) be at the extreme end of the path of length $k - 1$ (k , respectively). If $b_k < 0$, just flip x and y , and everything holds thanks to the negative contribution of b_k to $\mathfrak{Q}_G^{\mathbf{b}}(x)$. If we are in case (b), instead, we just need a single bidirectional path of length h with x and y being the two consecutive nodes at one extreme end of the path, with their position depending on the sign of a_h and b_h .

We are left to show the statement for two non-proportional \mathbf{a} and \mathbf{b} both different from $\mathbf{0}$ that satisfy case (c) of the observation above. We can distinguish two subcases, depending on whether inequality (11) of Lemma 3 holds or not.

The first subcase we examine is when the inequality holds. Consider the undirected graphs $G_{h,k}$ in Fig. 5. The linear centralities of x and y with respect to \mathbf{a} and \mathbf{b} in this graph are

$$\begin{aligned} \mathfrak{Q}_{G_{h,k}}^{\mathbf{a}}(x) &= e_a + a_h s + a_{h+1} p + a_{k-1} q + a_k t \\ \mathfrak{Q}_{G_{h,k}}^{\mathbf{b}}(x) &= e_b + b_h s + b_{h+1} p + b_{k-1} q + b_k t \end{aligned}$$

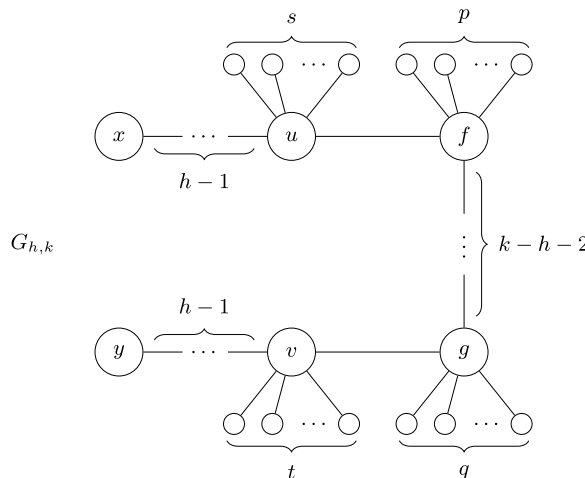


Fig. 5. The (family of) graphs $G_{h,k}$ used to distinguish non-proportional sequences when inequality (11) holds. The horizontal dotted line between x and u (resp., y and v) represents a path of length $h - 1$. Analogously, the vertical dotted line between f and g represents a path of length $k - h - 2$. Note that when $h = 1$, u gets to coincide with x , hence the s nodes in such a case are connected to x (everything holds symmetrically for y).

$$\begin{aligned} \mathfrak{Q}_{G_{h,k}}^a(y) &= e_a + a_h t + a_{h+1} q + a_{k-1} p + a_k s \\ \mathfrak{Q}_{G_{h,k}}^b(y) &= e_b + b_h t + b_{h+1} q + b_{k-1} p + b_k s, \end{aligned}$$

where $e_a = a_1 + a_2 + \dots + a_{k+h-2}$ and $e_b = b_1 + b_2 + \dots + b_{k+h-2}$. Then, the linear centralities $\mathfrak{Q}_{G_{h,k}}^a$ and $\mathfrak{Q}_{G_{h,k}}^b$ do not agree on $G_{h,k}$ if there is an integer nonnegative solution (s, p, q, t) to the following system of inequalities:

$$\begin{cases} e_a + a_h s + a_{h+1} p + a_{k-1} q + a_k t \geq e_a + a_h t + a_{h+1} q + a_{k-1} p + a_k s \\ e_b + b_h s + b_{h+1} p + b_{k-1} q + b_k t < e_b + b_h t + b_{h+1} q + b_{k-1} p + b_k s, \end{cases}$$

which is equivalent to

$$\begin{cases} \alpha(s - t) + \alpha'(p - q) \geq 0 \\ \beta(s - t) + \beta'(p - q) < 0, \end{cases} \tag{12}$$

where we have set $\alpha = a_h - a_k$, $\beta = b_h - b_k$, $\alpha' = a_{h+1} - a_{k-1}$ and $\beta' = b_{h+1} - b_{k-1}$. The two inequalities in (12) define two halfplanes (in the coordinate system $X = s - t$ and $Y = p - q$), whose intersection is unbounded and nonempty if $\alpha\beta' \neq \alpha'\beta$, which is exactly inequality (11) of Lemma 3. As a result, the system (12) has an integral solution (X, Y) , whence nonnegative s, t, p, q can be always found.

We are left to consider the subcase when inequality (11) does not hold. In such a case, we can apply the same line of reasoning as above to the graphs $\widehat{G}_{h,k}$ in Fig. 6, obtaining the following system of inequalities:

$$\begin{cases} e_a + a_h s + a_k p \geq e_a + a_h t + a_k q \\ e_b + b_h s + b_k p < e_b + b_h t + b_k q, \end{cases}$$

where $e_a = a_1 + a_2 + \dots + a_{k-1}$ and $e_b = b_1 + b_2 + \dots + b_{k-1}$. This is equivalent to

$$\begin{cases} a_h(s - t) - a_k(p - q) \geq 0 \\ b_h(s - t) - b_k(p - q) < 0. \end{cases}$$

This system has integer nonnegative solution (s, t, p, q) when $a_h b_k \neq b_h a_k$, which holds because of Lemma 3. \square

It is a noteworthy fact that all the graphs used to prove the second item of the statement above are undirected, and connected as long as $a_{h+1} \neq a_{k-1}$ and $h + 1 < k$ (except for case (a)).

As an example, consider the linear centralities \mathfrak{Q}^a and \mathfrak{Q}^b , with

$$\mathbf{a} = \left(0, 1, \frac{1}{2}, \frac{1}{3}, \dots\right)^T \quad \text{and} \quad \mathbf{b} = \left(0, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots\right)^T,$$

i.e., harmonic centrality and exponential decay centrality with $\delta = 1/2$. These two centralities can be distinguished on the graph $\widehat{G}_{1,3}$ with $s = q = 0$, $p = 6$, and $t = 2$. In fact, we have:

$$\begin{aligned} \mathfrak{Q}_{\widehat{G}_{1,3}}^a(x) &= \frac{7}{2} \geq \frac{7}{2} = \mathfrak{Q}_{\widehat{G}_{1,3}}^a(y) \\ \mathfrak{Q}_{\widehat{G}_{1,3}}^b(x) &= \frac{3}{2} < \frac{7}{4} = \mathfrak{Q}_{\widehat{G}_{1,3}}^b(y). \end{aligned}$$

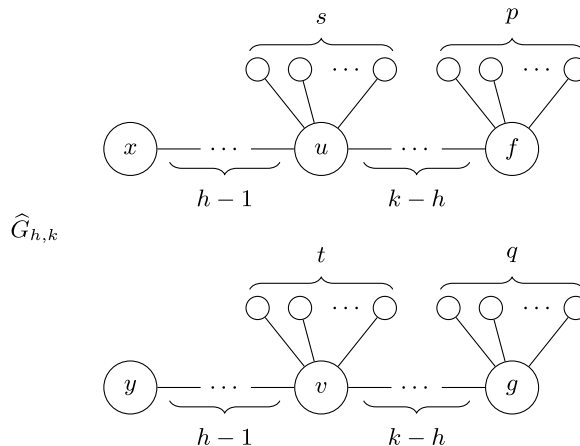


Fig. 6. The (family of) graphs $\widehat{G}_{h,k}$ used to distinguish non-proportional sequences when $a_{h+1} = a_{k-1}$ or $h + 1 = k$. The horizontal dotted lines at the bottom of the two connected components represent paths of length $h - 1$ (and $k - h$, respectively). Note that when $h = 1$, u gets to coincide with x , hence the s nodes in such a case are connected to x (everything holds symmetrically for y).

8. Graphs with many representable permutations

The previous section aimed at discussing when and on which graphs two sequences of coefficients yield different centralities. We now ask a dual question: given a graph G , which permutations of its nodes can be induced by linear centralities? In principle, we may have graphs where all linear centralities agree, regardless of how you choose the coefficients: the fact that there is a graph on which each pair of non-proportional centralities can be distinguished does not rule out the possibility that *on some graph* they may all agree.

On the other hand, it may be possible to have graphs that allow for many different permutations of their nodes, depending on the choice of coefficients. To avoid the complications introduced by nodes having the same centrality value, let us stick to the case with no ties.

Definition 13 (Representable permutation). We say that a permutation $\pi \in S_n$ is *representable* by a graph G with n nodes iff there is a choice of coefficients $\mathbf{a} \in \mathbf{R}^N$ such that $\Omega_G^{\mathbf{a}}$ has no ties and induces the permutation π on G .

Of course, we may classify graphs depending on how many permutations they can represent. More precisely, let us define the *representativeness* of a graph G with n nodes as follows:

$$R(G) = \frac{|\{\pi \in S_n \mid \pi \text{ is representable by } G\}|}{n!} \leq 1.$$

This ratio quantifies how expressive linear centralities are over a rigid graph G . If $R(G) = 1$, linear centralities cannot be less expressive than any other centrality: by choosing coefficients suitably, we can induce any permutation of the nodes, hence achieving what any other centrality can achieve. On the other hand, of course, if $R(G) < 1$ there are permutations that cannot be obtained using a linear centrality, although they may be obtained by some other centrality. In the rest of this section we show that at least in some cases linear centralities are maximally expressive (i.e., they have maximum representativeness); if we require the graph to be strongly connected, we will be able to obtain the same result but only asymptotically.

From now on, we will drop the assumption that $a_0 = 0$, as it will play a role in quantifying representativeness.

8.1. Rouché-Capelli theorem for the non-connected case

Given a graph G of n nodes, and a vector $\mathbf{v} \in \mathbf{R}^n$, determining if there exists a coefficient vector $\mathbf{a} \in \mathbf{R}^N$ such that $\Omega_G^{\mathbf{a}}(i) = v_i$ for all $i \in [n]$ can be seen as trying to find a solution \mathbf{a} for the following linear system of equations:

$$C_G \cdot \mathbf{a} = \mathbf{v}. \tag{13}$$

This is a system of n linear equalities with n unknowns, hence by the Rouché-Capelli theorem [22], it has a solution if and only if

$$\text{Rank}(C_G) = \text{Rank}(C_G|\mathbf{v}), \tag{14}$$

where $\text{Rank}(A)$ is the rank of matrix A , and $A|\mathbf{x}$ is the matrix A augmented with the column \mathbf{x} .

Now, if $\text{Rank}(C_G) = n$, Eq. (14) always holds, so if we find a graph G whose distance-count matrix has full rank, then the graph has representativeness equal to 1. Much more than this, in fact: not only can we find linear centralities that rank the nodes in any order we please, but we may even choose the *actual centrality values* for each single node. Such a graph exists:

Theorem 5. For every $n \geq 3$, the graph G_n in Fig. 7 has representativeness $R(G_n) = 1$.

Proof. The graph in Fig. 7 has the following distance-count matrix

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & n-1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & n-2 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & 1 & n-3 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & 1 & \dots & 1 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{bmatrix}$$

which is easily seen to have rank n (it is lower-triangular with no zeros on its diagonal). \square

8.2. Farkas' lemma for the general case

The main drawback of the solution in Fig. 7 is that the graph G_n is not strongly connected, albeit almost so. Therefore, we still wonder if a *strongly connected* graph can be permuted in all possible ways by linear centralities.

To tackle this problem (or, more in general, to have a tool to decide if a given permutation is representable by a graph), we shall use the following well-known result Farkas [23], Gale et al. [24]:

Lemma 4 (Farkas' lemma). Given $A \in \mathbf{R}^{m \times n}$ and $\mathbf{b} \in \mathbf{R}^m$, the following statements are equivalent:

1. there exists $\mathbf{x} \in \mathbf{R}^n$ such that $A\mathbf{x} \leq \mathbf{b}$;
2. there exists no $\mathbf{y} \in \mathbf{R}^m$ such that $A^T\mathbf{y} = \mathbf{0}$, $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^T\mathbf{y} < \mathbf{0}$.

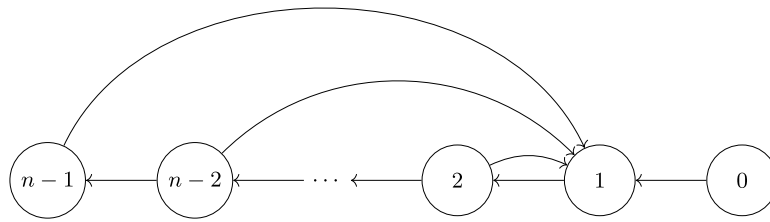


Fig. 7. A graph G_n with representativeness $R(G_n) = 1$.

A special version of Farkas' lemma is the following:

Corollary 3. Given $A \in \mathbf{R}^{m \times n}$, the following statements are equivalent:

1. there exists $\mathbf{x} \in \mathbf{R}^n$ such that $A\mathbf{x} < \mathbf{0}$;
2. there exists no $\mathbf{y} \in \mathbf{R}^m$ such that $A^T\mathbf{y} = \mathbf{0}$, $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$.

Proof. For every $\varepsilon > 0$, let (1_ε) and (2_ε) be the two equivalent statements of Lemma 4 when $\mathbf{b} = -\varepsilon\mathbf{1}$. Clearly, the first statement in the corollary is equivalent to $\exists \varepsilon > 0$ such that (1_ε) . By Farkas' lemma, the latter is equivalent to $\exists \varepsilon > 0$ such that (2_ε) holds, that is:

$$\exists \varepsilon > 0 \nexists \mathbf{y} \in \mathbf{R}^m \text{ such that } A^T\mathbf{y} = \mathbf{0}, \quad \mathbf{y} \geq \mathbf{0}, \quad -\varepsilon\mathbf{1}^T\mathbf{y} < 0.$$

Note that if $\varepsilon > 0$ and $\mathbf{y} \geq \mathbf{0}$ then $-\varepsilon\mathbf{1}^T\mathbf{y} < 0$ unless $\mathbf{y} = \mathbf{0}$. So we can express the above condition as:

$$\exists \varepsilon > 0 \nexists \mathbf{y} \in \mathbf{R}^m \text{ such that } A^T\mathbf{y} = \mathbf{0}, \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{y} \neq \mathbf{0},$$

or simply (since the latter condition does not depend on ε anymore)

$$\nexists \mathbf{y} \in \mathbf{R}^m \text{ such that } A^T\mathbf{y} = \mathbf{0}, \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{y} \neq \mathbf{0},$$

which ends the proof. \square

Define the matrix $\Xi \in \mathbf{R}^{(n-1) \times n}$ as follows

$$\Xi = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

Then, we can use Corollary 3 to determine if a given graph can represent a permutation as follows:

Theorem 6. A permutation π is representable by a graph G if and only if there does not exist $\mathbf{y} \in \mathbf{R}^{n-1}$ such that

$$(\Xi R_\pi C_G)^T \mathbf{y} = 0 \tag{15}$$

with $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$, where R_π is the permutation matrix of π (see Section 2).

Proof. For every $\mathbf{v} \in \mathbf{R}^n$, the vector is monotonically decreasing (i.e., $v_0 > v_1 > v_2 > \dots > v_{n-1}$) if and only if $\Xi\mathbf{v} < \mathbf{0}$, because the first inequality is $-v_0 + v_1 < 0$, that is, $v_0 > v_1$, and so on.

Recall that π is representable by G iff there exists $\mathbf{a} \in \mathbf{R}^n$ such that $\mathfrak{Q}_G^{\mathbf{a}}$ induces π on G . This means that

$$\mathfrak{Q}_G^{\mathbf{a}}(\pi(i)) > \mathfrak{Q}_G^{\mathbf{a}}(\pi(i+1))$$

for all $i \in [n-1]$. In other words, π is representable by G iff there exists $\mathbf{a} \in \mathbf{R}^n$ such that for all $i \in [n-1]$

$$(C_G \cdot \mathbf{a})_{\pi(i)} > (C_G \cdot \mathbf{a})_{\pi(i+1)},$$

or, equivalently,

$$(R_\pi C_G \cdot \mathbf{a})_i > (R_\pi C_G \cdot \mathbf{a})_{i+1}.$$

Applying the observation above to $\mathbf{v} = R_\pi C_G \cdot \mathbf{a}$, this is equivalent to saying that there exists $\mathbf{a} \in \mathbf{R}^n$ such that

$$\Xi R_\pi C_G \cdot \mathbf{a} < \mathbf{0}.$$

The result is obtained by applying Corollary 3 with $A = \Xi R_\pi C_G$. \square

In other words, to see if π is representable by G we can look at the system (15) of n equations with $n-1$ indeterminates: π is representable if and only if no non-negative non-trivial solution exists for the system.

More precisely, we will be using the following:

Corollary 4. For a given graph G and permutation π , let $X \in \mathbf{R}^{n \times k}$ be any matrix whose columns are linear combinations of the columns of C_G . Let $\rho = \pi^{-1}$. Then every $\mathbf{y} \in \mathbf{R}^{n-1}$ satisfying (15) also satisfies

$$\sum_{h: 0 < \rho(h)} x_{h,i} y_{\rho(h)-1} - \sum_{h: \rho(h) < n-1} x_{h,i} y_{\rho(h)} = 0 \quad \forall i \in [k]. \tag{16}$$

Proof. Let us rewrite the i -th element of the vector on the LHS of the system (15) in a more intelligible form:

$$((\exists R_\pi C)^T \mathbf{y})_i = \sum_{j < n-1} (\exists R_\pi C)_{i,j}^T y_j = \sum_{j < n-1} (\exists R_\pi C)_{j,i} y_j = \sum_{j < n-1} \sum_{h_1, h_2} \Xi_{j,h_1} (R_\pi)_{h_1, h_2} c_{h_2, i} y_j$$

where the variables h_1, h_2 range over $[n]$. Now, recalling the definition of Ξ and R_π , we can rewrite (15) as:

$$\sum_{j < n-1} (c_{\pi(j+1), i} - c_{\pi(j), i}) y_j = 0 \quad \forall i \in [n].$$

If we let $\rho = \pi^{-1}$, we can re-factor these equations as:

$$\sum_{h: 0 < \rho(h)} c_{h,i} y_{\rho(h)-1} - \sum_{h: \rho(h) < n-1} c_{h,i} y_{\rho(h)} = 0 \quad \forall i \in [n]. \tag{17}$$

So \mathbf{y} satisfies each of the n equalities (17). There is one such equality for every column of C_G . If all such equalities are satisfied, then *a fortiori* all their linear combinations are also satisfied. \square

A useful observation is that, if we look at a column i of a matrix X having only one single non-zero entry $x_{h,i}$, then condition (16) becomes

$$y_{\rho(h)} = y_{\rho(h)-1}$$

if $0 < \rho(h) < n - 1$; if, instead, $\rho(h) = 0$ ($\rho(h) = n - 1$, respectively) then the condition implies $y_0 = 0$ ($y_{n-2} = 0$, resp.).

The idea of representable permutations, besides being useful to estimate expressivity of linear centralities on a given graph, has another opposite application that we may call *robustness*. Suppose that you know that an adversary will rank the nodes of your graph using *some unknown* linear centrality (it can be closeness, harmonic, or anything else), and suppose that you want to be sure, say, that a certain node (or set of nodes) is ranked higher than another node (or set of nodes). If you can fiddle a bit with the arcs of the graph, you may want to try to impose a structure which guarantees that only the permutations that you like are representable, thus ruling out the possibility that the event you dislike happens, *whatever linear centrality will be applied by the adversary*. The usage of Eq. (16) is a very powerful tool to guarantee this kind of properties.

8.3. Representativeness of strongly connected graphs

As a first attempt at providing a sequence of strongly connected graphs with large representativeness, we have the following result:

Theorem 7. For $n > 3$, consider the graph G_n in Fig. 1, and let $\pi \in S_n$ and $\rho = \pi^{-1}$. The following two statements are equivalent:

1. $\rho(0) < \rho(n-1) < \rho(1)$ or $\rho(1) < \rho(n-1) < \rho(0)$;
2. the system (15) has no non-negative non-trivial solution.

Proof. The proof is rather long, hence we prefer to place it in the appendix. \square

The set of permutations satisfying the statement of Lemma 7 has cardinality $n!/3$, as shown by:

Lemma 5. If $n \geq 3$, there are $n!/3$ permutations $\rho \in S_n$, such that $\rho(0) < \rho(n-1) < \rho(1)$ or $\rho(1) < \rho(n-1) < \rho(0)$.

Proof. We prove this result by induction on n . For $n = 3$, there are only two permutations satisfying the constraint, and $3!/3 = 2$. For the inductive step, take any permutation $\rho' \in S_{n-1}$ satisfying the conditions in the statement, and build another permutation $\rho \in S_n$ so that:

- the relative order of $\rho(i)$ and $\rho(j)$ is the same as that of $\rho'(i)$ and $\rho'(j)$, for $i, j \in [n-2]$;
- the relative order of $\rho(i)$ and $\rho(n-1)$ is the same as that of $\rho'(i)$ and $\rho'(n-2)$, for $i \in [n-2]$.

Naturally $\rho(n-2)$ can be defined in n ways (because it is not involved in any of the above conditions). By inductive hypothesis, there are $(n-1)!/3$ permutations $\rho' \in S_{n-1}$ satisfying the hypothesis; we have just built the (only) possible $n(n-1)!/3 = n!/3$ permutations $\rho \in S_n$ with the same property. \square

From what we have shown so far, we can conclude that:

Corollary 5. For every $n > 3$, the graph G_n in Fig. 1 is strongly connected and $R(G_n) = 1/3$.

8.4. A graph that represents almost all permutations

In this section, we assume that $n \geq 4$, and we let \mathcal{P}_n be the set of permutations $\pi \in S_n$, with $\rho = \pi^{-1}$ such that one of the following (mutually exclusive) conditions holds:

1. $2 \leq \rho(n-1) \leq n-3$;
2. $\rho(n-1) = 1 \wedge \rho(n-2) \neq 0$;
3. $\rho(n-1) = n-2 \wedge \rho(n-2) \neq n-1$.

The following preliminary statement shows that almost all permutations belong to this set (in an asymptotic sense):

Lemma 6. *The set \mathcal{P}_n has cardinality $n!(n-3)/(n-1)$.*

Proof. Note that in \mathcal{P}_n there is no permutation such that $n-1$ is in the first (i.e., $\rho(n-1) = 0$) or last (i.e., $\rho(n-1) = n-1$) position.

We start with the set of permutations satisfying (1): it is easy to see that these are the permutations where $n-1$ is in one of the $(n-4)$ “middle” positions of the permutation, and the other $n-1$ elements can be permuted freely. Therefore, there are in total $(n-4)(n-1)!$ such permutations.

For the second constraint, we are considering the $(n-1)!$ permutations where $n-1$ is in the second position, and from these we have to cut off the $(n-2)!$ permutations of $(n-1)$ elements where $n-2$ comes first. Hence, there are $(n-1)! - (n-2)!$ permutations satisfying (2). The third constraint is analogous.

Hence, the total number of permutations in \mathcal{P}_n is:

$$\begin{aligned} |\mathcal{P}_n| &= (n-4)(n-1)! + 2((n-1)! - (n-2)!) = (n-1)! \left(n-2 - \frac{2}{n-1} \right) \\ &= (n-1)! \frac{n^2 - 3n}{n-1} = n! \frac{n-3}{n-1}, \end{aligned}$$

which ends the proof. \square

The next result connects \mathcal{P}_n to the graph G'_n of Fig. 8:

Theorem 8. *Consider the graph G'_n of Fig. 8 and let $\pi \in S_n$. The following two statements are equivalent:*

1. π belongs to \mathcal{P}_n ;
2. the permutation π is representable by the graph G'_n .

Proof. We start proving that (1) \implies (2). Let $\pi \in \mathcal{P}_n$ be fixed and $\mathbf{y} \in \mathbf{R}^{n-1}$ be a non-negative solution of the system (15): we will prove that $\mathbf{y} = \mathbf{0}$. Using Corollary 4, the vector \mathbf{y} must satisfy:

$$\sum_{h: 0 < \rho(h)} x_{h,i} y_{\rho(h)-1} - \sum_{h: \rho(h) < n-1} x_{h,i} y_{\rho(h)} = 0 \quad \forall i, \tag{18}$$

where X represents either the distance-count matrix of G'_n or, more generally, any matrix obtained by linearly combining the columns of the distance-count matrix. Index i ranges over all the columns of X , and there is one equation per column.

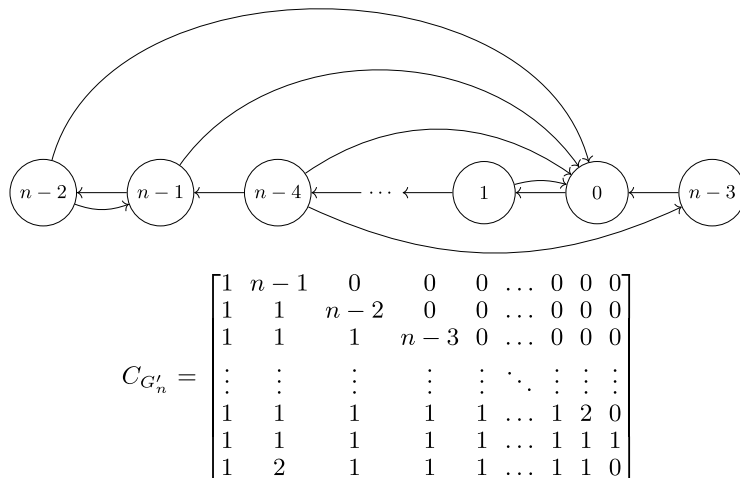


Fig. 8. The graph G'_n with $R(G'_n) = (n-3)/(n-1)$ and its distance-count matrix $C_{G'_n}$.

It is convenient to extend y adding one element to the left (y_{-1}) and one element to the right (y_{n-1}), with $y_{-1} = y_{n-1} = 0$. This way, we can rewrite (18) in the simpler form

$$\sum_{h=0}^{n-1} x_{h,i} (y_{\rho(h)-1} - y_{\rho(h)}) = 0 \quad \forall i. \tag{19}$$

Instead of using directly the matrix $X = C_{G'_n}$, we prefer to use the following:

$$X = \begin{pmatrix} n-1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 3-n & n-2 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 4-n & n-3 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 5-n & n-4 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

This matrix has $n - 1$ columns (i.e., it yields $n - 1$ equations), and its columns are obtained from $C_{G'_n}$ (Fig. 8) starting from the one of index 1, subtracting from each column the following one, except the last one that is left untouched.

Inner columns. All the columns of X except the first one and the last two contain exactly two non-zero entries, which are consecutive and with opposite sign. We call these $(n - 4)$ columns the “inner columns”, highlighted in grey in the matrix above. Eq. (19), when applied to these columns, becomes

$$(n - i - 1)(y_{\rho(i)-1} - y_{\rho(i)}) = (n - i - 3)(y_{\rho(i+1)-1} - y_{\rho(i+1)})$$

for $i = 1, \dots, n - 4$, which implies that:

$$\text{sgn}(y_{\rho(i)-1} - y_{\rho(i)}) = \text{sgn}(y_{\rho(i+1)-1} - y_{\rho(i+1)}).$$

This is, in fact, a chain of equalities between signs:

$$\text{sgn}(y_{\rho(1)-1} - y_{\rho(1)}) = \text{sgn}(y_{\rho(2)-1} - y_{\rho(2)}) = \text{sgn}(y_{\rho(3)-1} - y_{\rho(3)}) = \dots = \text{sgn}(y_{\rho(n-3)-1} - y_{\rho(n-3)}). \tag{20}$$

The last two columns. The very last column of X yields the equation:

$$y_{\rho(n-2)-1} = y_{\rho(n-2)}. \tag{21}$$

The next-to-last column, instead, gives

$$2(y_{\rho(n-3)-1} - y_{\rho(n-3)}) = -(y_{\rho(n-1)-1} - y_{\rho(n-1)})$$

which entails

$$\text{sgn}(y_{\rho(n-3)-1} - y_{\rho(n-3)}) = -\text{sgn}(y_{\rho(n-1)-1} - y_{\rho(n-1)}). \tag{22}$$

Observe that the left-hand side corresponds to the last element of the chain (20).

The first column. Finally, the first column of X yields:

$$(n - 1)(y_{\rho(0)-1} - y_{\rho(0)}) + (3 - n)(y_{\rho(1)-1} - y_{\rho(1)}) + (y_{\rho(n-1)-1} - y_{\rho(n-1)}) = 0.$$

Since $y_{\rho(1)-1} - y_{\rho(1)}$ is part of the chain (20) and has a negative coefficient, the second summand has the same sign as the last one because of (22). So the first addend must have the same sign as the elements of the chain (20):

$$\text{sgn}(y_{\rho(0)-1} - y_{\rho(0)}) = \text{sgn}(y_{\rho(1)-1} - y_{\rho(1)}). \tag{23}$$

The graph H. To consolidate all the knowledge gained so far, consider the undirected graph H whose n vertices are the pairs $(p - 1, p)$ for $p = 0, \dots, n - 1$, and with an edge between $(p - 1, p)$ and $(q - 1, q)$ whenever $p = \rho(i)$ and $q = \rho(i + 1)$ for some $i \in \{0, \dots, n - 4\}$. Each vertex $(p - 1, p)$ is interpreted as representing the difference $y_{p-1} - y_p$: when talking about a vertex, we may say “the sign of the vertex $(p - 1, p)$ ” to mean “the sign of the difference $y_{p-1} - y_p$ ”.

The presence of an edge between $(p - 1, p)$ and $(q - 1, q)$ represents the fact that

$$\text{sgn}(y_{p-1} - y_p) = \text{sgn}(y_{q-1} - y_q),$$

which we know by (20) and (23).

Looking at H , we see that it is in fact a chain starting at vertex $(\rho(0) - 1, \rho(0))$ and ending at $(\rho(n - 3) - 1, \rho(n - 3))$. This chain includes $n - 2$ vertices. All these vertices, hence, have the same sign. We refer to this sign as the “sign of the chain”.

There are 2 vertices left out of the chain (i.e., isolated in the graph), precisely the vertices in this set:

$$J = \{(\rho(n - 2) - 1, \rho(n - 2)), (\rho(n - 1) - 1, \rho(n - 1))\}.$$

We have some information about these vertices, though. The first one is in fact simple, because it has null sign (i.e., $\text{sgn}(y_{\rho(n-2)-1} - y_{\rho(n-2)}) = 0$), as in (21). The second one instead has a sign *opposite* to the sign of the chain, because of (22).

A graphical depiction of graph H and of all the information we have gained about its vertices is given in Fig. 9.

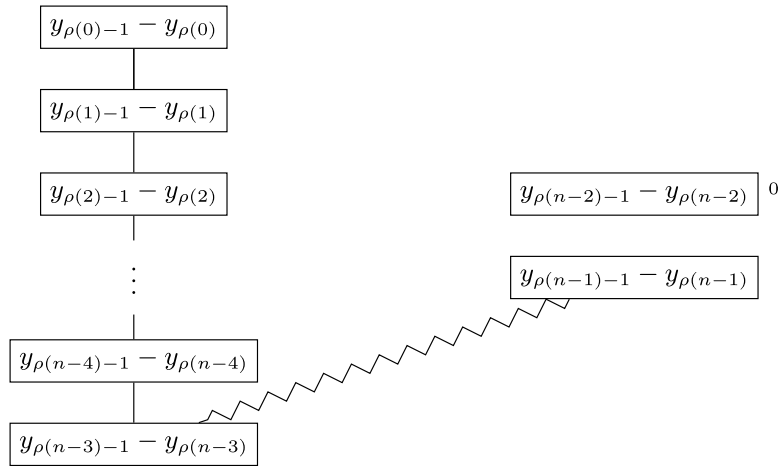


Fig. 9. The graph H . Node $(p - 1, \rho)$ is labelled explicitly as the difference it represents (i.e., $y_{p-1} - y_\rho$). Edges should be read “has-the-same-sign-as”. The zigzag line is read “has-opposite-sign-than”. The isolated node marked with 0 is known to have sign equal to zero (Eq. (21)).

The special vertices. Observe that all variables y_i (for $i = 0, \dots, n - 2$) appear in exactly two vertices, whereas the fake variables y_{-1} and y_{n-1} appear in exactly one vertex. We call the vertices containing y_{-1} and y_{n-1} “special” because the two variables y_{-1} and y_{n-1} are synthetic, they are not really part of y . So the sign of $(-1, 0)$ is ≤ 0 , whereas the sign of $(n - 2, n - 1)$ is ≥ 0 .

Case-by-case analysis. We end the proof with a case-by-case analysis, relying on the fact that $\pi \in \mathcal{P}_n$.

1. *Both special vertices belong to the chain.* In this case, necessarily the sign of the chain is 0 (all the vertices in the chain have the same sign, but there are two vertices of signs ≥ 0 and ≤ 0). So also the sign of node $(\rho(n - 1) - 1, \rho(n - 1))$ (because of (22)) is zero. In other words, all the differences appearing in all vertices are null. Also note that for the special vertices $y_{-1} - y_0 = 0$ implies $y_0 = 0$, and similarly $y_{n-2} = 0$. Using all the other equalities, we obtain $y = \mathbf{0}$.
2. *Both special vertices are out of the chain.* This fact would imply $\rho(n - 2) = 0$ and $\rho(n - 1) = n - 1$, or the other way round. But, by definition of \mathcal{P}_n , $\rho(n - 1)$ cannot be 0 or $n - 1$. So this case is impossible.
3. *Exactly one special vertex belongs to the chain.* Necessarily, it must be the case that $\rho(n - 2) = 0$ or $\rho(n - 2) = n - 1$, as the other vertex out of the chain cannot be special, as explained in the previous item. In the former case (the other one is analogous) we have that $y_{-1} - y_0 = 0$, hence $y_0 = 0$. Moreover, y_0 must appear also in $y_0 - y_1$, and either this vertex belongs to the chain, or it does not. If not, we have $\rho(n - 1) = 1$, which is impossible by hypothesis, since $\rho(n - 2) = 0$. Thus, the vertex belongs to the chain. Additionally, since y_0 is zero, vertex $y_0 - y_1$ must have sign ≤ 0 , but the other special vertex $(n - 2, n - 1)$ must be part of the chain too, and it has sign ≥ 0 . Again, these two facts together imply that the sign of the chain as well as the sign of the non-special vertices of H are all null, hence $y = \mathbf{0}$.

We now prove that (2) \implies (1), by contraposition: we assume that (1) does not hold (i.e., $\pi \notin \mathcal{P}_n$), and show that such a permutation is not representable by G'_n .

For π not to be in \mathcal{P}_n , one of the following must hold:

- a) $\rho(n - 1) = 0$;
- b) $\rho(n - 1) = 1 \wedge \rho(n - 2) = 0$;
- c) $\rho(n - 1) = n - 2 \wedge \rho(n - 2) = n - 1$;
- d) $\rho(n - 1) = n - 1$.

If (a) holds (the case ((d) is symmetrical), then there must be a coefficient vector \mathbf{a} such that $\mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 1) > \mathfrak{Q}_{G'_n}^{\mathbf{a}}(i)$ for all $i \in [n - 2]$.

More explicitly, we must have

$$\mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 1) = a_0 + 2a_1 + a_2 + \dots + a_{n-2} > \sum_{j \leq i} a_j + (n - i - 1)a_{i+1} = \mathfrak{Q}_{G'_n}^{\mathbf{a}}(i) \quad \text{for all } i \in [n - 2].$$

This yields the following system of inequalities:

$$\begin{aligned} \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 1) > \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 2) &\implies a_1 > a_{n-1} \\ \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 1) > \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 3) &\implies a_1 > a_{n-2} \\ \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 1) > \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 4) &\implies a_1 + a_{n-2} > 2a_{n-3} \\ \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 1) > \mathfrak{Q}_{G'_n}^{\mathbf{a}}(n - 5) &\implies a_1 + a_{n-3} + a_{n-2} > 3a_{n-4} \end{aligned}$$

$$\begin{aligned} \mathfrak{Q}_{G'_n}^a(n-1) > \mathfrak{Q}_{G'_n}^a(n-6) &\implies a_1 + a_{n-4} + a_{n-3} + a_{n-2} > 4a_{n-5} \\ &\dots \\ \mathfrak{Q}_{G'_n}^a(n-1) > \mathfrak{Q}_{G'_n}^a(i) &\implies a_1 + \sum_{j=i+2}^{n-2} a_j > (n-i-2)a_{i+1} \quad \text{for all } i \in [n-4]. \end{aligned}$$

Combining each inequality with the previous ones, we obtain:

$$\begin{aligned} a_1 &> a_{n-1} \\ a_1 &> a_{n-2} \\ 2a_1 &> a_1 + a_{n-2} > 2a_{n-3} \implies a_1 > a_{n-3} \\ 3a_1 &> 3a_{n-4} \implies a_1 > a_{n-4} \\ 4a_1 &> 4a_{n-5} \implies a_1 > a_{n-5} \\ &\dots \\ (n-i-2)a_1 &> (n-i-2)a_{i+1} \quad \text{for all } 0 \leq i \leq n-4. \end{aligned} \tag{24}$$

For $i = 0$ this leads to $a_1 > a_1$, which is clearly impossible. Thus, we conclude that π is not representable by G'_n if $\rho(n-1) = 0$ (or $\rho(n-1) = n-2$).

Finally, we are left to prove case (b) (again, case (c) is symmetrical). It is clear to see that (b) implies that $\mathfrak{Q}_{G'_n}^a(n-1) > \mathfrak{Q}_{G'_n}^a(i)$ for all $i \in [n-3]$. As before, this yields the following system of inequalities:

$$\begin{aligned} a_1 &> a_{n-2} \\ a_1 &> a_{n-3} \\ a_1 &> a_{n-4} \\ &\dots \\ (n-i-2)a_1 &> (n-i-2)a_{i+1} \quad \text{for all } 0 \leq i \leq n-4. \end{aligned}$$

Hence, for $i = 0$, this leads to the same contradiction $a_1 > a_1$. \square

As a consequence:

Corollary 6. *Almost all permutations are representable by the graph G'_n . More precisely*

$$\lim_{n \rightarrow \infty} R(G'_n) = 1.$$

9. Conclusion and open problems

This paper presented and studied for the first time the class of linear geometric centralities, providing general properties of expressivity and robustness. Here is a list of problems this paper suggests but that we were unable to solve:

- Is convexity necessary for a linear centrality to be rank monotone (Proposition 4)?
- Is it possible to find strongly connected graphs to distinguish pairs of non-proportional coefficients, in the cases of Theorem 4 where we had to rely on disconnected graphs?
- Is it possible to find some family G''_n of strongly connected graphs such that $R(G''_n) = 1$, at least for sufficiently large n ?

A path that might be worth investigating is the study of distance-count matrices, as they vehiculate more information than degree sequences [17], but have appeared much less in the literature. For instance, it would be interesting to figure out if there exists a polynomial-time algorithm that decides whether a given matrix is the distance-count matrix of some graph; as far as we know, this is still an open problem.

CRedit authorship contribution statement

Paolo Boldi: Writing – original draft; **Flavio Furia:** Writing – original draft; **Chiara Prezioso:** Writing – original draft.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgementst

This work was supported in part by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Italian MUR. Neither the European Union nor the Italian MUR can be held responsible for them.

Appendix A. Proof of Theorem 7

For the reader’s convenience we restate the theorem:

Theorem. For $n > 3$, consider the graph G_n in Fig. 1, and let $\pi \in S_n$ and $\rho = \pi^{-1}$. The following two statements are equivalent:

1. $\rho(0) < \rho(n - 1) < \rho(1)$ or $\rho(1) < \rho(n - 1) < \rho(0)$;
2. the system (15) has no non-negative non-trivial solution.

Proof. We start proving that (1) \implies (2). Let y be a non-negative solution to (16); we will prove that necessarily $y = 0$, using Corollary 4. The last $n - 2$ columns of C_{G_n} can be combined by subtracting from each column all the following ones, and then dividing by a suitable coefficient so that the topmost non-zero element is a 1, yielding

$$\begin{bmatrix} 1 & n-1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & \dots & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 & 1 \\ 1 & 2 & \frac{n-3}{n-2} & 0 & \dots & 0 & 0 & 0 \end{bmatrix}.$$

Now doing something similar for the first two columns (we subtract from each of them the last $n - 2$ columns, and multiply by a suitable constant), we obtain

$$\begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ \frac{1}{n-2} & \frac{1}{n-2} & \frac{n-3}{n-2} & 0 & \dots & 0 & 0 & 0 \end{bmatrix}.$$

Let $T = \{\rho(i) \mid i = 2, \dots, n - 2\}$ (i.e., looking at the left-hand side of condition (16), the indices of y involved in the last $n - 3$ columns). Let also $p = \rho(0)$, $q = \rho(n - 1)$, $r = \rho(1)$ (indices involved in the first three columns): by hypothesis, either $p < q < r$ or $r < q < p$. We assume $p < q < r$ (the other case is symmetrical). Note that necessarily $p < n - 1$, $0 < r$ and $0 < q < n - 1$.

Case 1. If $0 \in T$, the condition imposed by the last $n - 3$ columns imply

$$y_t = y_{t-1} \quad \forall t \in T, 0 < t < n - 1$$

$$y_0 = 0.$$

On the other hand, for the first two (identical) columns condition (16) yields:

$$y_{p-1} - y_p + \frac{1}{n-2}(y_{q-1} - y_q) = 0,$$

since $p, q \neq 0, n - 1$. Now start from $k = q - 1$: if $k \in T$ then $y_k = y_{k-1}$, so we can iterate with a smaller value of k , until eventually $k = p$. Hence $y_{q-1} = y_{q-2} = \dots = y_p$. Similarly, starting from $k = p - 1$, if $k \in T$ we have $y_k = y_{k-1}$, and we can iterate until $k = 0$. Hence $y_{p-1} = y_{p-2} = \dots = y_0 = 0$.

As a consequence, the last equation becomes

$$-y_p + \frac{1}{n-2}(y_p - y_q) = 0$$

$$-(n-2)y_p + y_p = y_q$$

$$-(n-3)y_p = y_q.$$

Since both y_p and y_q are non-negative, we must have $y_p = y_q = 0$.

The equation implied by the third column is (using Iverson’s bracket notation)

$$-y_r[r < n - 1] + y_{r-1} + \frac{n-3}{n-2}(y_{q-1} - y_q) = 0,$$

so plugging in $y_q = y_{q-1} = 0$ we distinguish two cases:

- if $r = n - 1$, then $y_{n-2} = 0$;
- if, instead $r < n - 1$, then $y_r = y_{r-1}$, and starting from $k = n - 2 \in T$ (hence $y_k = y_{k-1}$), we can iterate with a smaller k until reaching r , obtaining $y_{n-2} = y_{n-3} = \dots = y_r$.

In both cases, we can apply the same reasoning, starting from $k = r - 1$, down to q , yielding $y_{r-1} = y_{r-2} = \dots = y_q = 0$.

Finally, we notice that in all cases we have a set of equations involving all indices t , and having the form either $y_t = y_{t-1}$ or $y_t = 0$ (and in particular $y_0 = 0$). So $\mathbf{y} = \mathbf{0}$.

Case 2. If $0 \notin T$, then necessarily $p = 0$. The equation for the first two columns this time is

$$-y_0 + \frac{1}{n-2}(y_{q-1} - y_q) = 0,$$

and again starting from $k = q - 1$ and iterating we obtain $y_{q-1} = \dots = y_0$, yielding

$$-(n-2)y_0 + y_0 = y_q.$$

Once more $y_0 = y_q = 0$, and we can proceed as above.

We are now going to prove that (2) \implies (1), by contraposition. Let us assume that (1) does not hold, thus $\rho(n-1) > \rho(0)$ and $\rho(n-1) > \rho(1)$ (or, symmetrically, $\rho(0) > \rho(n-1)$ and $\rho(1) > \rho(n-1)$).

Under such conditions, given a coefficient vector \mathbf{a} , we have

$$\Omega_{G_n}^{\mathbf{a}}(0) = a_0 + (n-1)a_1$$

$$\Omega_{G_n}^{\mathbf{a}}(1) = a_0 + a_1 + (n-2)a_2$$

$$\Omega_{G_n}^{\mathbf{a}}(n-1) = a_0 + 2a_1 + (n-3)a_2,$$

and, by hypothesis,

$$a_0 + 2a_1 + (n-3)a_2 > a_0 + (n-1)a_1 \implies a_2 > a_1$$

$$a_0 + 2a_1 + (n-3)a_2 > a_0 + a_1 + (n-2)a_2 \implies a_1 > a_2,$$

which is impossible. Thus, π is not representable by G_n , and by Theorem 6 system (15) has a nonnegative non trivial solution \mathbf{y} , i.e., (2) cannot hold. \square

References

- [1] P. Boldi, F. Furla, C. Prezioso, Linear geometric centralities, in: M. Bloznelis, P. Drungilas, B. Kamiński, P. Prałat, M. Šileikis, F. Théberge, R. Vaiscekauskas (Eds.), *Modelling and Mining Networks*, Springer Nature Switzerland, Cham, 2025, pp. 1–16.
- [2] P. Boldi, S. Vigna, Axioms for centrality, *Internet Math.* 10 (3–4) (2014) 222–262.
- [3] A. Bavelas, A mathematical model for group structures, *Hum. Organ.* 7 (1948) 16–30.
- [4] A. Bavelas, Communication patterns in task-oriented groups, *J. Acoust. Soc. Am.* 22 (6) (1950) 725–730.
- [5] G. Kishi, On centrality functions of a graph, in: N. Saito, T. Nishizeki (Eds.), *Graph Theory and Algorithms*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1981, pp. 45–52.
- [6] G. Kishi, M. Takeuchi, A type of centrality functions for a directed graph, *Electron. Commun. Jpn. (Part I: Commun.)* 66 (5) (1983) 38–47.
- [7] O. Skibski, J. Sosnowska, Axioms for distance-based centralities, *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1) (2018).
- [8] G. Sabidussi, The centrality index of a graph, *Psychometrika* 31 (4) (1966) 581–603.
- [9] M. Garg, Axiomatic foundations of centrality in networks, *Soc. Sci. Res. Netw.* (2009).
- [10] P. Boldi, D. D’Ascenzo, F. Furla, S. Vigna, et al., Score and rank semi-monotonicity for closeness, betweenness, and distance–decay centralities, *Soc. Netw. Anal. Min.* 14 (1) (2024) 183.
- [11] D. Schoch, U. Brandes, Re-conceptualizing centrality in social networks, *Eur. J. Appl. Math.* 27 (6) (2016) 971–985.
- [12] B. Bollobás, *Modern graph theory*, 184 of *Graduate Texts in Mathematics*, Springer–Verlag, 1998.
- [13] U. Brandes, T. Erlebach, *Network Analysis: Methodological Foundations* (Lecture Notes in Computer Science), number 3418 in *Lecture Notes in Computer Science*, Springer-Verlag, 2005.
- [14] N. Lin, *Foundations of Social Research*, McGraw-Hill, New York, 1976.
- [15] L. Katz, A new status index derived from sociometric analysis, *Psychometrika* 18 (1) (1953) 39–43.
- [16] J.M. Anthonisse, The rush in a directed graph, Technical Report BN 9/71, Mathematical Centre, Amsterdam, 1971.
- [17] F. Buckley, F. Harary, *Distance in Graphs*, Addison-Wesley Pub. Co., Redwood City, Calif., 1990.
- [18] J.W. Roginski, R.M. Gera, E.C. Rye, *The Neighbor Matrix: Generalizing the Degree Distribution*, 2016, arXiv preprint arXiv:1510.06952.
- [19] P. Boldi, S. Vigna, The WebGraph framework I: compression techniques, in: *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*, ACM Press, Manhattan, USA, 2004, pp. 595–601.
- [20] L. Babai, E.M. Luks, Canonical labeling of graphs, in: *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, 1983, pp. 171–183.
- [21] P. Boldi, A. Luongo, S. Vigna, Rank monotonicity in centrality measures, *Netw. Sci.* 5 (4) (2017) 529–550.
- [22] I.R. Shafarevich, A.O. Remizov, *Linear Algebra and Geometry*, Springer Science & Business Media, 2012.
- [23] J. Farkas, Theorie der einfachen ungleichungen, *Journal für die reine und angewandte Mathematik (Crelles J.)* 1902 (124) (1902) 1–27.
- [24] D. Gale, H.W. Kuhn, A.W. Tucker, Linear programming and the theory of games, *Activ. Analy. Product. Alloc.* 13 (1951) 317–335.