Australian & New Zealand Journal of Statistics

WILEY

Statistical Society of Australia

The New Zealand Statistical Association

THEORY AND METHODS OPEN ACCESS

Distance Measures for Unweighted Undirected Networks: A Comparison Study

Anna Simonetto[1] | Matteo Ventura[2]

[1]Department of Civil, Environmental, Architectural Engineering and Mathematics, University of Brescia, Italy | [2]Department of Economics and Management, University of Brescia, Italy

Correspondence: Matteo Ventura (matteo.ventura@unibs.it)

Received: 12 April 2024 | **Revised:** 19 February 2025 | **Accepted:** 14 March 2025

Funding: Anna Simonetto has been partially supported by 'Fondazione Cariplo' (Italy) and 'Regione Lombardia' (Italy) under the project: 'Biodiversità, suolo e servizi ecosistemici. Strategie, metodi e tecniche per la realizzazione di food system robusti, resilienti e sostenibili'. Grant Emblematici Maggiori 2020-4135.

Keywords: Euclidean distance | Laplacian matrix | Minkowski distance | network comparison | spectra

ABSTRACT

Networks are mathematical structures that allow the representation of complex systems by jointly modelling the elements of the system and the relationships that exist among them. To analyse different contexts or systems, methodological tools are necessary to allow for the quantitative estimation of the differences existing between two or more networks. For this purpose, various tools have been proposed in the literature. This study is an exploratory analysis of the impacts that different methods (distances and spectral methods) have on the comparative evaluation of two networks. The analyses were conducted through a simulation study that considered three different perturbation schemes to investigate the behaviour of each method with increasing randomness in the perturbation scheme (i.e., edge removal). Results show that the distances between adjacency matrices are sensitive only to changes in the network density, while spectral methods are sensitive to changes in both the network density and the degree of the nodes.

1 | Introduction

Network science aims at understanding the structure of complex systems via the study of the relationships that characterise the interacting elements of the network (Coscia 2021). These relationships are described by structures composed of nodes, representing the actors of the system, and edges, representing the relationships between the actors of the system.

The study of networks is a well-established research area in mathematics and sociology (Albert and Barabási 2002; Barabási 2013; Crossley 2008; Lauritzen 1996; Strogatz 2001). Starting from the 2000s, there has been a strong proliferation of network analysis in several fields, such as medicine (Barabási et al. 2011; Butte et al. 2000; Szklarczyk et al. 2015), psychology (Bhushan et al. 2019; Borsboom et al. 2021; De Ron et al. 2021), transportation sciences (Ducruet and Lugo 2013; Guze 2014; Háznagy et al. 2015) and ecology (Anand et al. 2010; Creamer et al. 2016; Delmas et al. 2019).

A key challenge in network science is network comparison, a pervasive and complex task. This difficulty arises from the need to balance interpretability, result effectiveness and computational efficiency (Koutra et al. 2013; Tantardini et al. 2019). Methods for network comparison can be classified into two groups depending on node correspondence between networks (Tantardini et al. 2019). When two networks share the same node set or have a common subset of nodes with known pairwise correspondence between them, their distances fall under the category of known node-correspondence methods. Otherwise, they are categorised as unknown node-correspondence methods. Unknown node-correspondence methods ideally allow the comparison of any pair of networks by condensing their global structure into one or more statistics, which are then used to compute a distance.

Choice of the comparison method significantly influences the outcome of network comparison, making it essential to understand each method's characteristics to select the most suitable approach for a given problem. Tantardini et al. (2019) provided a review of methods for comparing networks, addressing both node-correspondence and unknown node-correspondence networks, and tested them on synthetic random networks considering different types of perturbations. In their study, Wills and Meyer (2020) investigated the performance of a set of distances on random synthetic graphs using a multi-scale lens. The authors categorised distances based on three scales, which measured the methods' capability to detect changes within a graph: the fine scale of the local connectivity, formed by the ego-net; the very large scale associated with communities; and a mesoscale bridging the scales from the local to the global scales. Wilson and Zhu (2008) examined the performances of the spectra of the adjacency, Laplacian and normalised Laplacian matrix. Specifically, the authors investigated the cospectrality of these matrix representations and compared the Euclidean distance between spectra and the edit distance between graphs. Martínez and Chavez (2019) compared the Euclidean distance with a dissimilarity measure defined by Schieber et al. (2017) and kernel-based measures, and highlighted that the Euclidean distance has a good trade-off between good and fast performances. The analysis of the state of the art reveals a lack of investigation on the relationship between perturbation intensity and network distance metrics, particularly in two widely used approaches for network comparisons: (i) distances based on adjacency matrices and (ii) distances based on spectral structures.

In this work, we analyse the effectiveness of Minkowski distances for adjacency matrices and Euclidean distance for spectral structures. These methods are computationally efficient and widely applied in network science. We decided to represent network structure by adjacency and spectral matrices, rather than specific network structures or estimation models, to enhance the generalisability of the results and their adaptability to various application contexts.

To evaluate the performance of selected comparison methods, we tested each method under three different perturbation schemes and four different network sizes. This study aims to highlight how different approaches respond to (i) an increasing number of perturbations, (ii) the complete disconnection of a node and (iii) the influence of the order in which edges are removed, ultimately allowing researchers to select the most suitable distance metric for their specific research questions.

2 | Undirected Unweighted Network

The key structure of a network is the graph. The graph can be identified by the pair $G = (V, E)$, where $V = \{1, \dots, n\}$ is the set of nodes and E is the set of edges. A graph is simple if it has no loops and if there are no multiple edges (Lauritzen 1996). The size of the graph is determined by the number of nodes and it is denoted by $N = |V|$. The node set can represent both quantitative and qualitative variables, and the edge set represents probabilistic relationships between variables. In this study, we focus on pure and undirected graphs, that is, nodes represent only one variable type, and the edge $\{i, j\}$ and its opposite $\{j, i\}$ are equal and both present in E. In the undirected graph, the edges are graphically represented by lines (Lauritzen 1996; Wills and Meyer 2020).

A graph can also be represented by the square $N \times N$ adjacency matrix ($\mathbf{A}$), where each row/column corresponds to a node and each cell represents an edge. In the case of an unweighted network, the cell a_{ij} of $\mathbf{A}$ is equal to 1 if the edge connecting the ith node and the jth node exists, or 0 otherwise. The adjacency matrix of an undirected graph is symmetric: $a_{ij} = a_{ji}$. If the graph has no self-loops, the diagonal of the adjacency matrix contains zeros and it divides the adjacency matrix of an undirected graph into two identical triangular halves (Coscia 2021).

3 | Comparing Networks

In this section, two node-correspondence methods for comparing networks are described: distances between adjacency matrices, and distances between the spectra of three different types of matrices (hereafter only spectral methods).

3.1 | Distances Between Adjacency Matrices

To compare two adjacency matrices ($\mathbf{A}_1$ and $\mathbf{A}_2$), we consider distances from the Minkowski family, as defined in Li et al. (2011), which are expressed as follows:

$$d_{MK}(\mathbf{A}_1, \mathbf{A}_2) = \left[\sum_{i=1}^{N} \sum_{j=1}^{N} |a_{1,ij} - a_{2,ij}|^k \right]^{\frac{1}{k}} \quad k \geq 1 \tag{1}$$

where $a_{1,ij}$ is the cell a_{ij} of $\mathbf{A}_1$, and $a_{2,ij}$ is the cell a_{ij} of $\mathbf{A}_2$.

By varying the value of k, three different notable distances can be obtained: the Manhattan distance, for $k = 1$; the Euclidean distance, for $k = 2$; and the Chebychev distance, for $k \to \infty$ (Coghetto 2016).

3.2 | Distances Between Spectral Methods

These approaches leverage spectral theory, a branch of mathematics dedicated to analysing and understanding the structural properties of networks through the eigenvectors and eigenvalues of associated matrices (Chung 1997). Spectral methods provide valuable insights into the organisation and behaviour of networks by transforming complex structural information into a more compact and mathematically tractable form. The spectrum $\mathbf{\Lambda}$ of a matrix is defined as the ordered sequence of its eigenvalues λ, which encapsulate critical information about the graph's topology and connectivity.

According to Wilson and Zhu (2008), the spectral distance between two graphs can be quantified as the Euclidean distance between the eigenvalues of the respective matrices representing the graphs, as shown in (2). This measure is particularly appealing because it provides a straightforward and computationally efficient way to compare networks, capturing structural differences directly through their spectral characteristics. Moreover, the spectral distance reflects variations in global properties, such as connectivity patterns and modularity, making it a powerful tool for identifying similarities and differences between complex networks.

$$d_{\lambda}(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2) = \sqrt{\sum_{i=1}^{N} \left(\lambda_{1,i} - \lambda_{2,i}\right)^2} \tag{2}$$

It is important to note that eigenvalues reflect the cohesiveness of node clusters and the underlying community structure within a network. Consequently, the choice of matrix representation significantly influences both the stability and the representational power of the eigenvalues. Furthermore, the spectrum of each matrix can emphasise distinct graph properties (Gera et al. 2018; Wilson and Zhu 2008).

In this study, we explored three spectral methods based on the Euclidean distance between the spectra of (i) adjacency matrices, (ii) Laplacian matrices and (iii) normalised Laplacian matrices.

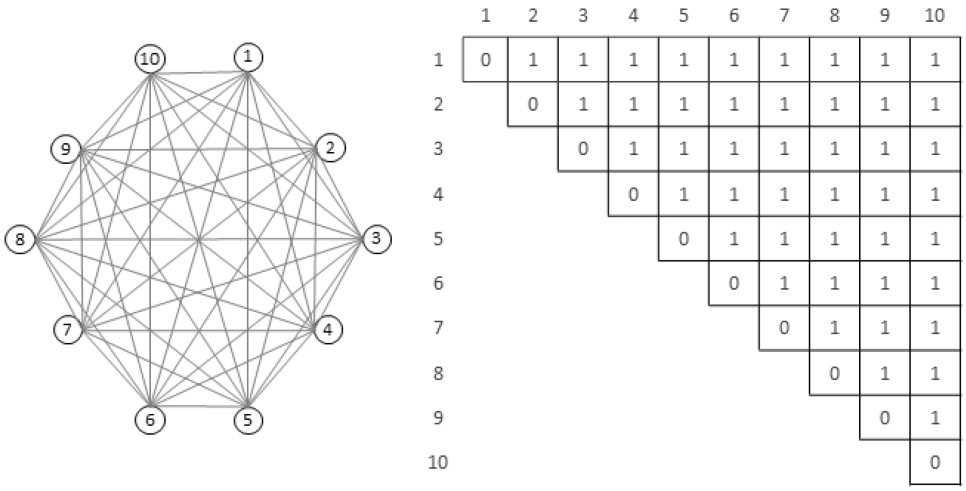
The Laplacian matrix $\boldsymbol{L}$ is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where A is the adjacency matrix of the graph and $\boldsymbol{D}$ is the diagonal degree matrix (Wilson and Zhu 2008). In $\boldsymbol{D}$, the diagonal entries are the degrees of the corresponding nodes, which describe each node's number of connections when the graph is simple (i.e., graph that does not have more than one edge between any two nodes and no edge starts and ends at the same node).

The normalised Laplacian matrix (Chen et al. 2004; Wilson and Zhu 2008) is defined as

$$\boldsymbol{\mathcal{L}} = \begin{cases} 1 & \text{if } i = j \text{ and } d_j \neq 0 \\ -\frac{1}{\sqrt{d_i d_j}} & \text{if } a_{ij} = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where d_i and d_j are, respectively, the degree of the node i and the degree of the node j, and a_{ij} is a cell of the adjacency matrix. The normalised Laplacian can be also written as $\boldsymbol{\mathcal{L}} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}$.

The Laplacian and its normalised form are both positive semidefinite, and therefore they have all the eigenvalues greater than or equal to 0; moreover, in the normalised Laplacian, all the eigenvalues are in the range $[0, 2]$ (Chen et al. 2004; Chung 1997).



	1	2	3	4	5	6	7	8	9	10
1	0	1	1	1	1	1	1	1	1	1
2		0	1	1	1	1	1	1	1	1
3			0	1	1	1	1	1	1	1
4				0	1	1	1	1	1	1
5					0	1	1	1	1	1
6						0	1	1	1	1
7							0	1	1	1
8								0	1	1
9									0	1
10										0

FIGURE 1 | Graphical (left) and matrix (right) representation of a full network (FullNet) with 10 nodes and 45 edges.

4 | Simulation Study

4.1 | Simulation Experiment Design

We designed a set of three simulation experiments based on distinct perturbation schemes: nested ordered removal of edges (NeOr), nested random removal (NeRa) and non-nested random removal (NoRa).

Each scheme follows a common experimental framework, beginning with a fully connected, undirected and unweighted network, referred to as FullNet (see Figure 1). Edges are systematically removed, gradually reducing the network's density. This process continues until all edges are eliminated, leaving an empty network.

After each perturbation step, we measured the distance between the perturbed network and the original FullNet. To quantify these differences, we evaluated seven distance metrics based on the adjacency matrix, including Manhattan, Euclidean and Minkowski distances with $k = \{5, 10, 20, 40\}$, as well as the limiting case $k \to \infty$. Additionally, we applied three spectral methods, computing Euclidean distances from the spectra of the adjacency matrix, Laplacian matrix and normalised Laplacian matrix.

To ensure comparability across different distance measures, we normalised each distance by dividing it by the distance between FullNet and the empty network. This normalisation constrained all distance values to a range between 0 and 1.

We conducted experiments on networks of 10, 25, 50 and 100 nodes to explore how network size affects the stability, sensitivity and behaviour of distance metrics under different perturbation schemes.

Since the nested ordered removal (NeOr) of edges follows a deterministic process, we performed a single simulation for this scheme. In contrast, for the NeRa and NoRa schemes, we conducted 100 repetitions, each time varying the order of edge removals to account for stochastic effects.

The following sections provide a detailed discussion of the methodology and results for each perturbation scheme. To support the reader in comparing different strategies and provide a useful tool for guiding the definition of their own analysis methodologies, Table 1 summarises the results of all simulations. The behaviour of the six distance measurement strategies is described based on the percentage of removed edges at which distance values of 0.25, 0.50 and 0.75 are observed.

4.2 | Nested Ordered Edge Removal

The first simulation scenario follows a strictly deterministic sequence of perturbations, ensuring that key events, such as node disconnections, occur at fixed, predictable stages. This approach guarantees consistency across simulations, making it easy to identify critical moments.

TABLE 1 | Percentage of removed edges at which distances of 0.25, 0.50 or 0.75 are observed.

		NeOr			NeRa			NoRa		
N nodes	**Distance**	**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**
10	Manhattan	28.26%	52.17%	76.09%	28.26%	52.17%	76.09%	28.26%	52.17%	76.09%
	Euclidean	8.70%	28.26%	58.70%	8.70%	28.26%	58.70%	8.70%	28.26%	58.70%
	Minkowski 40	4.35%	4.35%	4.35%	4.35%	4.35%	4.35%	4.35%	4.35%	4.35%
	Adjacency	26.09%	69.57%	91.30%	10.87%	32.61%	69.57%	10.87%	32.61%	69.57%
	Laplacian	17.39%	39.13%	71.74%	21.74%	47.83%	73.91%	21.74%	47.83%	73.91%
	Norm. Laplacian	21.74%	54.35%	80.43%	41.30%	73.91%	86.96%	41.30%	73.91%	86.96%
25	Manhattan	25.58%	50.50%	75.42%	25.58%	50.50%	75.42%	25.58%	50.50%	75.42%
	Euclidean	6.64%	25.58%	56.48%	6.64%	25.58%	56.48%	6.64%	25.58%	56.48%
	Minkowski 40	0.66%	0.66%	0.66%	0.66%	0.66%	0.66%	0.66%	0.66%	0.66%
	Adjacency	34.22%	71.43%	92.36%	7.31%	27.24%	61.13%	7.31%	27.57%	61.13%
	Laplacian	12.96%	38.87%	69.77%	22.59%	48.50%	74.42%	22.59%	48.50%	74.09%
	Norm. Laplacian	15.95%	43.19%	81.73%	61.46%	86.05%	93.69%	61.46%	86.05%	93.69%
50	Manhattan	25.12%	50.08%	75.04%	25.12%	50.08%	75.04%	25.12%	50.08%	75.04%
	Euclidean	6.36%	25.12%	56.36%	6.36%	25.12%	56.36%	6.36%	25.12%	56.36%
	Minkowski 40	0.16%	0.16%	0.16%	0.16%	0.16%	0.16%	0.16%	0.16%	0.16%
	Adjacency	40.29%	73.74%	92.99%	6.61%	26.10%	58.56%	6.61%	26.18%	58.56%
	Laplacian	11.66%	38.50%	69.90%	23.65%	49.10%	74.55%	23.65%	49.10%	74.55%
	Norm. Laplacian	15.58%	45.68%	81.16%	75.86%	92.41%	96.74%	75.86%	92.50%	96.66%
100	Manhattan	25.03%	50.03%	75.02%	25.03%	50.03%	75.02%	25.03%	50.03%	75.02%
	Euclidean	6.28%	25.03%	56.27%	6.28%	25.03%	56.27%	6.28%	25.03%	56.27%
	Minkowski 40	0.04%	0.04%	0.04%	0.04%	0.04%	0.04%	0.04%	0.04%	0.04%
	Adjacency	41.53%	74.35%	93.48%	6.40%	25.53%	57.38%	6.40%	25.51%	57.40%
	Laplacian	11.57%	38.34%	69.84%	24.28%	49.53%	74.77%	24.30%	49.53%	74.77%
	Norm. Laplacian	13.59%	43.95%	80.89%	86.23%	96.04%	98.28%	86.25%	96.04%	96.04%

Note: Different network sizes, distance metrics and perturbation schemes were considered.

Starting from the adjacency matrix of FullNet (Figure 1, right), edges are systematically removed in a structured, ordered manner. Specifically, the process eliminates edges by setting the first nonzero element a_{ij} in the matrix to zero, corresponding to the removal of the edge between nodes i and j. This is done row by row, proceeding sequentially from left to right within each row. A new row is processed only after all elements in the current row have been set to zero.

Visually, this method ensures that all edges connected to a node are removed before proceeding to the next node. The process systematically isolates one node at a time before moving forward, leading to a clear, step-by-step reduction in the network's connectivity (Figure 2).

The nested ordered removal generates a series of progressively perturbed networks, where the edges removed from a network with p perturbations are permanently removed in all subsequent networks with $p + q$ removed edges. The final perturbation involves removing the edge connecting nodes $N - 1$ and N, resulting in an empty network.

The results of this ordered removal are shown in Figure 3 for the case of a network with 10 nodes. However, similar considerations to the ones made in the following apply to networks of other sizes.

All distances computed from the adjacency matrices follow a consistently increasing pattern, and node disconnections do not affect this trend.

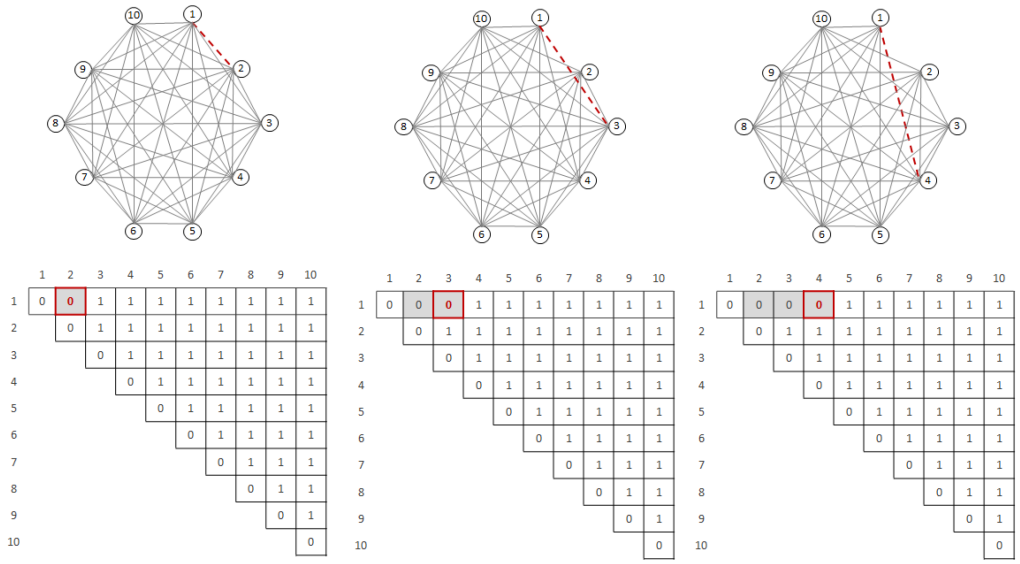


FIGURE 2 | Graphical (top) and matrix (bottom) representation of the first three steps of the NeOr perturbation procedure. The edge removed at each step and the corresponding entry in the adjacency matrix are highlighted in red.

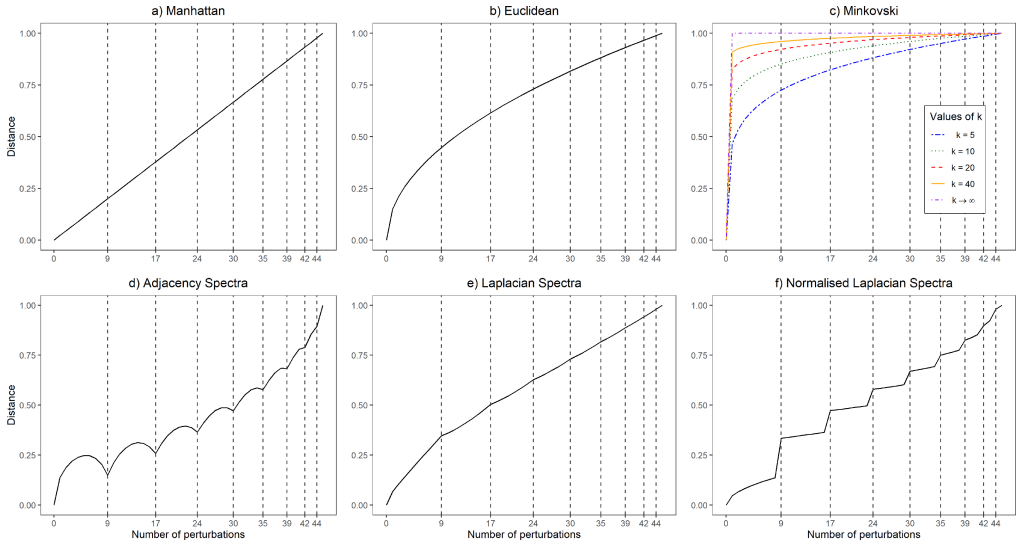


FIGURE 3 | Network size: 10 nodes—Results of the nested ordered edge removal (NeOr) applying the Manhattan (plot a), Euclidean (plot b) and Minkovski (plot c) distances on the adjacency matrix and the Euclidean distance on the adjacency spectra (plot d), the Laplacian spectra (plot e) and the normalised Laplacian Spectra (plot *f*). The perturbations corresponding to a complete node disconnection have been highlighted with a vertical grey dotted line. The vertical grey dotted line of the last perturbation (disconnecting node 9 and node 10) is not shown for the sake of clarity of the plots.

The Manhattan distance (Figure 3a) weighs equally all the perturbations. For each perturbation (i.e., each edge removed), the normalised distance between the FullNet and the perturbed network increases by $1/|E|$ ($|E|$ is the number of edges in the FullNet) independently by the type of edge removed or the density level of the perturbed network. In our simulation, the Manhattan distance between the FullNet and a perturbed network corresponds to the number of perturbations (i.e., the number of deleted edges) divided by $N(N-1)/2$ (45, in the case of the network with 10 nodes).

Applying the Euclidean distance (Figure 3b), the first perturbations have higher impacts on the normalised distance than the last ones. In the scenario with 10 nodes, the first three perturbations (i.e., from a network with 45 edges to a perturbed network with 42 edges) result in a normalised distance increase of 0.26. Conversely, the final three perturbations (i.e., from a perturbed network with three edges to an empty network) lead to a normalised distance increase of 0.03 points, which accounts for 13% of the impact of the initial three perturbations.

Focusing on the general Minkovski distances (Figure 3c), the weight of the first perturbations increases as k increases, up to the limiting case of k tending to infinity, whereby the first perturbation is associated with the maximum distance so that each further perturbation does not increase the distance between the FullNet and the perturbed networks.

The normalised distance estimation patterns obtained with spectral methods are relevantly different from those based on the adjacency matrix. The normalised distance computed on the adjacency matrix spectra has a non-monotonic pattern (Figure 3d). Downward peaks are observed when a node is completely disconnected (vertical lines). Since the removal is ordered, the pattern can be interpreted as follows. Focusing on a node, the first perturbations cause an increase in the normalised distance from the FUllNet till the node has lost about 75% of the edges it had at the beginning of the perturbation phase; then the distance starts to decrease. In the decreasing phase, the pth perturbed network has a lower normalised distance from the FullNet than the perturbed network at the $(p-1)$th step. For instance, the normalised distance of the network perturbed at step 10 (with one disconnected node and one edge removed from the second node) from FullNet is equal to 0.213, while the normalised distance of the network perturbed at step 5 (with no node disconnected and only four edges removed from node 1) is equal to 0.248.

The distance pattern based on the Laplacian spectra (Figure 3e) is monotonous, akin to the one based on the Euclidean distance computed on the adjacency matrix (Figure 3b). The first perturbations have a greater impact on the normalised distance than the last ones. Within the same node, the impact of each perturbation on the normalised distance is more homogeneous compared to case of the Euclidean distance on the adjacency matrix. Indeed, a quasi-linear effect within each node can be observed.

The distance pattern of the normalised Laplacian spectra is characterised by a stepwise monotonic pattern. Within the same node, perturbations lead to a relatively low increase in normalised distance. Strong increases in the distance are observed when a node is completely disconnected (corresponding to the vertical grey dotted line in Figure 3f). For instance, the perturbed network at step 8 showed a normalised distance of 0.1369 from the FullNet. The ninth perturbation, consisting of the disconnection of the first node, increases the normalised distance of the perturbed network at step 9 from the FullNet by 0.1966 (distance between network 9 and the FullNet is equal to 0.3335).

These observations, initially made for a network with 10 nodes, can be easily extended to networks of varying sizes. Focusing on the first column of Table 1, the following patterns emerge across simulations with networks of 10, 25, 50, and 100 nodes. For Manhattan, Euclidean and Minkowski 40 distances, the percentage of edge removals required to reach a given distance threshold decreases as the network size increases. However, the effect of network size becomes less pronounced as the threshold value increases.

In contrast, for the Euclidean distance between the spectra of the adjacency matrix, the percentage of removals required to reach a given threshold increases with the network size. The impact of network size is most pronounced at lower thresholds and becomes negligible as the threshold increases. For the Euclidean distance between the spectra of the Laplacian and normalised Laplacian matrices, similar patterns to those observed for Manhattan, Euclidean and Minkowski distances are seen. However, only at the 0.25 threshold is there a notable reduction in the percentage of removals required to reach it, while for higher thresholds this variation becomes negligible.

4.3 | Nested Random Edge Removal

The second perturbation scheme aims to relax the hypothesis of ordered edge removal of the NeOr scheme. The aim is to understand whether and to what extent the patterns identified in the previous scheme are influenced by the ordered structure of edge removal. The NeRa scheme starts from the FullNet. At each perturbation, a cell a_{ij} is randomly extracted from among those of the adjacency matrix, which are still equal to 1. The cell a_{ij} and the symmetric cell a_{ji} are set equal to 0, which corresponds to eliminating the edge between the node i and the node j. We proceeded with the perturbations until we reached the empty network. The scheme led to a nested set of perturbed networks because, if an edge was eliminated in the pth perturbation, it was surely eliminated also in all subsequent perturbations (Figure 4). The NeRa scheme was replicated 100 times, modifying the random extraction sequence of the edges to be removed.

The results of the distance patterns calculated on the adjacency matrices are equal in all NeRa replicates and to those of the NeOr scheme (Figure 3a–c). From these results, we can deduce that for all the distances considered and applied to the adjacency matrix, neither the position of the edge removed in the network nor the degree of the perturbed nodes is

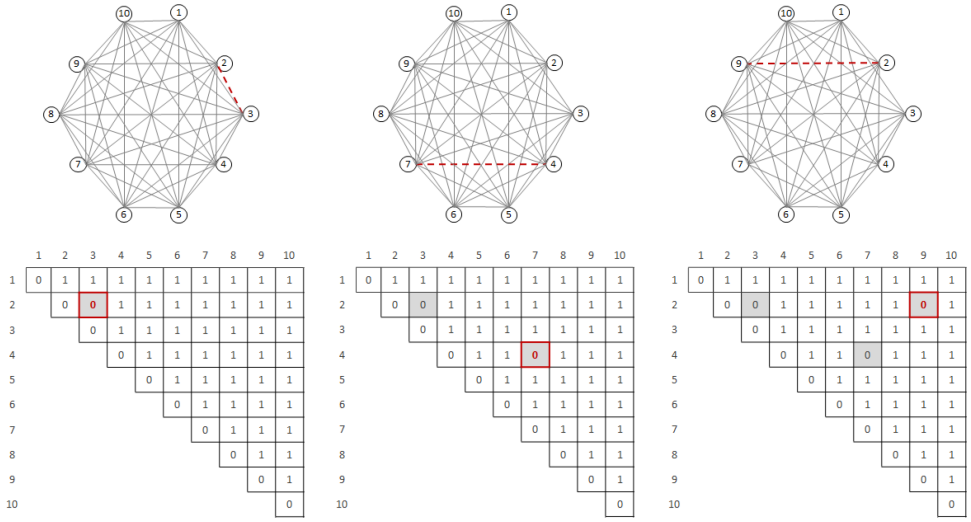


FIGURE 4 | Graphical (top) and matrix (bottom) representation of an example of the first three steps in the NeRa perturbation scheme. The edge removed at each step and the corresponding entry in adjacency matrix are highlighted in red.

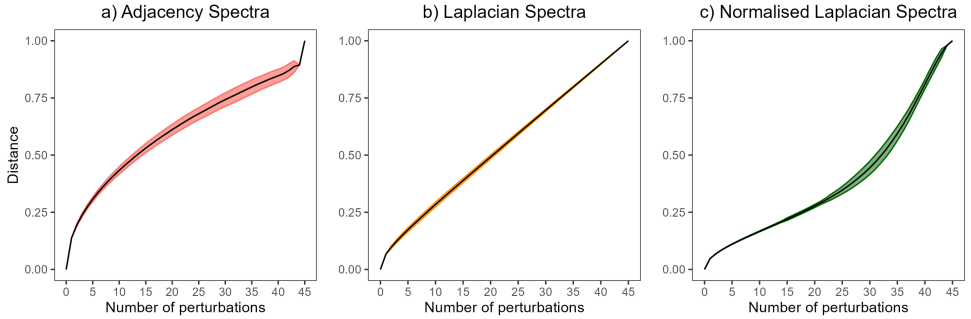


FIGURE 5 | Results of the nested random edge removal (NeRa scheme) applying the Euclidean distance on the adjacency spectra (plot a), the Laplacian spectra (plot b) and the normalised Laplacian spectra (plot c).

relevant in the definition of the distance. Only the number of edges removed with respect to the total of edges in the initial network (FullNet) has an impact on this set of distances.

In Figure 5, we report the mean and standard deviations of the normalised distances for the spectral methods measured in the 100 replicates of the NeRa scheme for networks with 10 nodes. All mean patterns are monotonous but differ according to the impact determined by the density of the disturbed network. The distance pattern measured on the adjacency spectra weighs more the first perturbations than the last ones. In the case with 10 nodes, half of the maximum observed normalised distance is reached between perturbations 13 and 14 (corresponding to 31% of the maximum number of removable edges). As the density of the disturbed network decreases, the absolute increase in normalised distance with each successive perturbation decreases. The variability around the mean pattern is contingent upon the diverse patterns of edge removal and node disconnection observed in each replicate.

The distance measured on the Laplacian spectra shows an almost linear pattern, with little variability around the mean value. No changes in the pattern are observed corresponding to the disconnection of nodes.

The trend of the mean normalised distance pattern calculated on the normalised Laplacian distance is quite different from the that of the adjacency spectra. The distance on the normalised Laplacian spectra weighs less perturbations carried out on the denser networks, while the contribution of the perturbations in the less dense networks is greater. In our study, half of the maximum observed normalised distance is reached between perturbations 32 and 33 (corresponding to the removal of 71% of the maximum number of removable edges). The variability concerning the mean value of the pattern is proportional to the number of disconnections made in that perturbation. In our case, we observed that the first

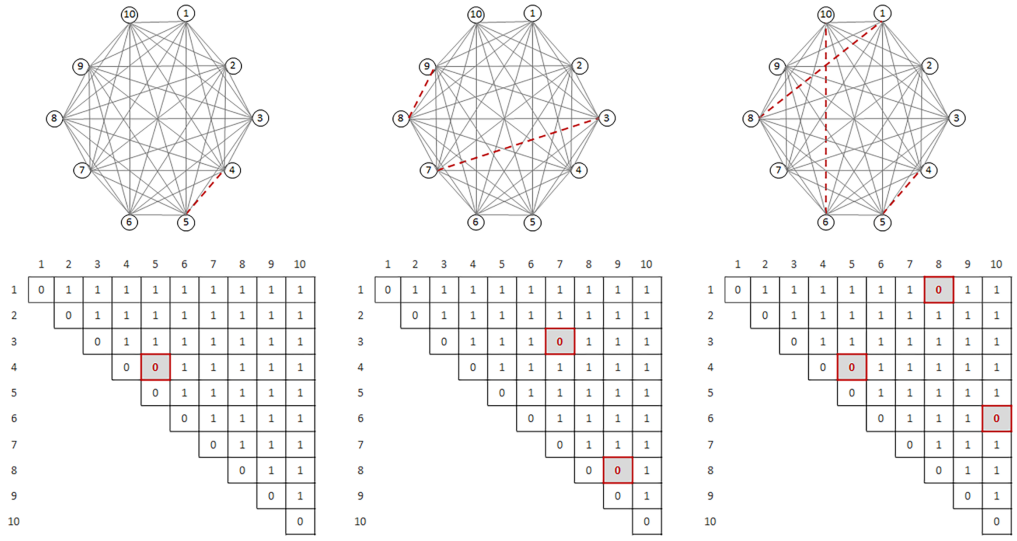


FIGURE 6 | Graphical (top) and matrix (bottom) representation of an example of the first three steps in the NoRa perturbation scheme. The edges removed at each step and the corresponding entries in adjacency matrix are highlighted in red. Note that the edge which connects node 4 and node 5 was randomly removed both in the first and third steps but was present in the second step.

disconnection over all the 100 replicates occurs at iteration 25, which corresponds to the point where the variability of normalised distance trends from the mean pattern increases.

To investigate the impact of network size on the considered measures, we refer to Table 1. The results indicate that the Manhattan, Euclidean and Minkowski distances are invariant to the type of edge removal, which corroborates the findings from the previous section. Additionally, the percentage of edges that must be removed to achieve a specific distance threshold decreases as network size increases.

Notably, the results for the spectral methods provide further insights. For the Euclidean distance applied to the adjacency matrix spectra, the percentage of edge removals required to reach a given threshold decreases as network size grows. In contrast, for the Euclidean distance applied to the spectra of the Laplacian and normalised Laplacian matrices, this percentage increases with network size. This effect is particularly pronounced for the normalised Laplacian, where the proportion of removed edges to reach a specific threshold is strongly influenced by network size, especially at lower threshold values.

These results confirm that the distance between adjacency matrix spectra behaves similarly to the Euclidean distance between adjacency matrices, while the Euclidean distance between Laplacian spectra follows a pattern resembling that of the Manhattan distance. On the other hand, the Euclidean distance between normalised Laplacian spectra exhibits more distinct characteristics, assigning greater weight to perturbations in denser networks and being more sensitive to network size, particularly for lower threshold values.

4.4 | Non-Nested Random Edge Removal

The third perturbation scheme relaxes the nesting condition. Starting from the FullNet, p edges are eliminated from the original network at the pth perturbation; the removed edges are chosen randomly in each perturbation. Therefore, an edge removed in the pth perturbation could be present in the $(p+1)$th perturbation. The last perturbation is the one in which all edges of the network are removed. The NoRa scheme was replicated 100 times, modifying the random extraction sequence of the edges to be removed (Figure 6).

Also for the NoRa scheme, the results of the distances on the adjacency matrices are the same as those of the NeOr scheme. The NoRa results on the measurements calculated on the spectra (see Figure 7 and Table 1) show patterns similar to those of the NeRa scheme (Figure 5); therefore, the relaxation of the nesting condition did not lead to a change in distance patterns.

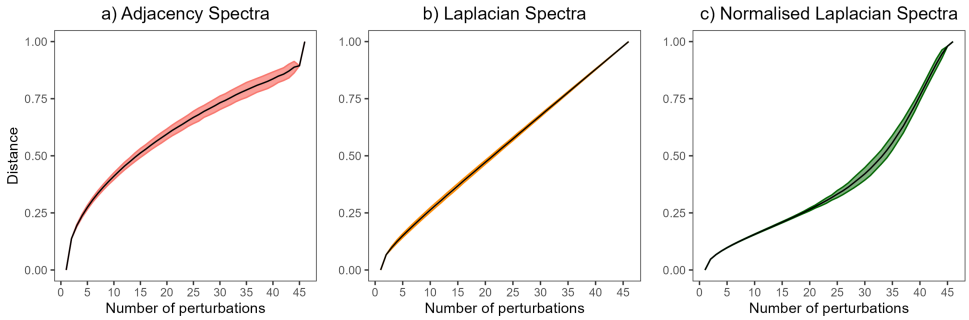


FIGURE 7 | Results of the non-nested random edge removal measured with the Euclidean distance on the adjacency spectra (plot a), the Laplacian spectra (plot b) and the normalised Laplacian spectra (plot c).

5 | Final Observations and Remarks

In this work, we employed two different node-correspondence approaches (based on adjacency matrix and spectral methods) to quantify differences between undirected unweighted networks with the same number of nodes. The performances of these methods were tested with a simulation experiment carried out considering three different perturbation schemes with increasing degrees of randomness. The simulations did not consider specific network structures or estimation models, as the observations were based on network representations (adjacency matrix or spectral transformations of the adjacency matrix) that underlie a variety of different models. The development of simulations on these objects allows for greater generalisability of the results and adaptability to different application contexts.

The simulation study revealed that the distance estimates based on adjacency matrices are primarily influenced by changes in network density and remain unaffected by variations in node degree. The behaviour of these metrics does not alter, even with the introduction of randomness in the second and third perturbation schemes.

The selection of a specific distance within the Minkowski family (i.e., determining which value of k to choose in (1)) depends on the objectives of the network comparison analysis. Our results demonstrate that the Manhattan distance emphasises the number of perturbations between the two networks. Larger values of k are better suited for placing greater weight on even small perturbations.

Spectral methods, instead, are influenced by changes both in the density of the network and in the node degree. For the nested ordered edge removal simulation, the adjacency spectral method exhibits a non-monotonic pattern, which complicates the interpretation of the obtained distance estimation. This is particularly challenging, as the estimated distances may appear smaller for more perturbed networks compared to less perturbed ones. The Laplacian and normalised Laplacian spectral methods, instead, present a monotonous pattern. In the random removal schemes, the impact of a perturbation estimated with the adjacency spectral method decreases as the network density decreases; in the normalised Laplacian spectral method, instead, the impact of a perturbation increases as the density of the network decreases.

From the previous observations and the results presented in Table 1, it is evident that the estimation of differences between networks is strongly influenced by the comparison tool employed. Understanding the behaviour of these tools is essential for accurately interpreting the phenomena being compared. This study provides valuable insights into the performance of methods based on adjacency matrices and three spectral methods, which are among the most commonly used in network analysis. No single tool emerges as optimal for all types of investigations, as the behaviour of the distances varies depending on the perturbation type, with more or less pronounced differences based on network size. Although it is not possible to define one 'optimal' tool, the results of this study can guide the reader in selecting the most appropriate method based on the research objectives, the context and the specific aspects they wish to emphasise in the comparison between two or more networks.

Considering that the most common investigation scenario is the NoRa, where comparisons between networks with random perturbations are allowed (i.e., not nested), the results show that both the Euclidean distance and the Euclidean distance between the spectra of the adjacency matrix yield similar outcomes. These measures are not significantly influenced by the network size and are suitable when the researcher wishes to emphasise edge removal, especially in cases where lower intensity perturbations are of greater interest. In contrast, the normalised Laplacian distance and the

Minkowski distance of order k — with k not excessively large — are more appropriate when the focus is on higher intensity perturbations. The normalised Laplacian offers the advantage of not requiring the selection of a specific k value, although it is more sensitive to network size. Finally, the Manhattan and Laplacian distances exhibit similar behaviour and can be used when the researcher wants to avoid having the distance overly influenced by network density.

Acknowledgement

Open access publishing facilitated by Universita degli Studi di Brescia, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Albert, R., and A. L. Barabási. 2002. "Statistical Mechanics of Complex Networks." *Reviews of Modern Physics* 74: 47–97.

Anand, M., A. Gonzalez, F. Guichard, J. Kolasa, and L. Parrott. 2010. "Ecological Systems as Complex Systems: Challenges for an Emerging Science." *Diversity* 2: 395–410.

Barabási, A. L. 2013. "Network Science." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371: 20120375.

Barabási, A. L., N. Gulbahce, and J. Loscalzo. 2011. "Network Medicine: A Network-Based Approach to Human Disease." *Nature Reviews Genetics* 12: 56–68.

Bhushan, N., F. Mohnert, D. Sloot, L. Jans, C. Albers, and L. Steg. 2019. "Using a Gaussian Graphical Model to Explore Relationships Between Items and Variables in Environmental Psychology Research." *Frontiers in Psychology* 10: 1050.

Borsboom, D., M. K. Deserno, M. Rhemtulla, et al. 2021. "Network Analysis of Multivariate Data in Psychological Science." *Nature Reviews Methods Primers* 1: 1–18.

Butte, A. J., P. Tamayo, D. Slonim, T. R. Golub, and I. S. Kohane. 2000. "Discovering Functional Relationships Between Rna Expression and Chemotherapeutic Susceptibility Using Relevance Networks." *Proceedings of the National Academy of Sciences* 97: 12182–12186.

Chen, G., G. Davis, F. Hall, Z. Li, K. Patel, and M. Stewart. 2004. "An Interlacing Result on Normalized Laplacians." *SIAM Journal on Discrete Mathematics* 18: 353–361.

Chung, F. R. 1997. *Spectral Graph Theory*. Vol. 92. American Mathematical Soc.

Coghetto, R. 2016. "Chebyshev Distance." *Formalized Mathematics* 24: 121–141.

Coscia, M. 2021. "The Atlas for the Aspiring Network Scientist." arXiv preprint arXiv:2101.00863.

Creamer, R., S. Hannula, J. Van Leeuwen, et al. 2016. "Ecological Network Analysis Reveals the Inter-Connection Between Soil Biodiversity and Ecosystem Function as Affected by Land Use Across Europe." *Applied Soil Ecology* 97: 112–124.

Crossley, N. 2008. "Small-World Networks, Complex Systems and Sociology." *Sociology* 42: 261–277.

De Ron, J., E. I. Fried, and S. Epskamp. 2021. "Psychological Networks in Clinical Populations: Investigating the Consequences of Berkson's Bias." *Psychological Medicine* 51: 168–176.

Delmas, E., M. Besson, M. H. Brice, et al. 2019. "Analysing Ecological Networks of Species Interactions." *Biological Reviews* 94: 16–36.

Ducruet, C., and I. Lugo. 2013. *Structure and Dynamics of Transportation Networks: Models, Methods and Applications*, 347–364. The SAGE handbook of transport studies.

Gera, R., L. Alonso, B. Crawford, et al. 2018. "Identifying Network Structure Similarity Using Spectral Graph Theory." *Applied Network Science* 3: 1–15.

Guze, S. 2014. "Graph Theory Approach to Transportation Systems Design and Optimization." *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation* 8: 571–578.

Háznagy, A., I. Fi, A. London, and T. Németh. 2015. "Complex Network Analysis of Public Transportation Networks: A Comprehensive Study." In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 371–378. IEEE.

Koutra, D., J. T. Vogelstein, and C. Faloutsos. 2013. "Deltacon: A Principled Massive-Graph Similarity Function." In *Proceedings of the 2013 SIAM International Conference on Data Mining*, 162–170. SIAM.

Lauritzen, S. L. 1996. *Graphical Models*. Vol. 17. Clarendon Press.

Li, Z., Q. Ding, and W. Zhang. 2011. "A Comparative Study of Different Distances for Similarity Estimation." In *International Conference on Intelligent Computing and Information Science*, 483–488. Springer.

Martínez, J. H., and M. Chavez. 2019. "Comparing Complex Networks: In Defence of the Simple." *New Journal of Physics* 21: 013033.

Schieber, T. A., L. Carpi, A. Díaz-Guilera, P. M. Pardalos, C. Masoller, and M. G. Ravetti. 2017. "Quantification of Network Structural Dissimilarities." *Nature Communications* 8: 1–10.

Strogatz, S. H. 2001. "Exploring Complex Networks." *Nature* 410: 268–276.

Szklarczyk, D., A. Franceschini, S. Wyder, et al. 2015. "String v10: Protein–Protein Interaction Networks, Integrated Over the Tree of Life." *Nucleic Acids Research* 43: D447–D452.

Tantardini, M., F. Ieva, L. Tajoli, and C. Piccardi. 2019. "Comparing Methods for Comparing Networks." *Scientific Reports* 9: 1–19.

Wills, P., and F. G. Meyer. 2020. "Metrics for Graph Comparison: A Practitioner's Guide." *PLoS One* 15: e0228728.

Wilson, R. C., and P. Zhu. 2008. "A Study of Graph Spectra for Comparing Graphs and Trees." *Pattern Recognition* 41: 2833–2841.