Centrality in complex networks under incomplete data

Shvydun, S.

DOI
10.1371/journal.pcsy.0000042

Publication date
2025

Document Version
Final published version

Published in
Complex Systems

PLOS COMPLEX SYSTEMS

RESEARCH ARTICLE

Centrality in complex networks under incomplete data

Sergey Shvydun*

Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

* s.shvydun@tudelft.nl

Abstract

The concept of centrality is one of the essential tools for analyzing complex systems. Over the years, a large number of centrality indices have been proposed that account for different aspects of a network. Unfortunately, most real networks are substantially incomplete, which affects the results of the centrality measures. This article aims to evaluate the sensitivity of 16 centrality measures to the presence of errors or incomplete information about the structure of a complex network. Our experiments are performed across 113 empirical networks. As a result, we identify centrality indices that are highly vulnerable to incomplete data.

Check for updates

OPEN ACCESS

Citation: Shvydun S (2025) Centrality in complex networks under incomplete data. PLoS Complex Sys 2(5): e0000042. https://doi.org/10.1371/journal.pcsy.0000042

Editor: Kosmas Kosmidis, Aristotle University of Thessaloniki: Aristoteleio Panepistemio Thessalonikes, GREECE

Received: May 26, 2024

Accepted: March 19, 2025

Published: May 13, 2025

Data availability statement: The data that support the findings of this study are publicly available from - Clauset A, Tucker E, Sainz M. The Colorado Index of Complex Networks., 2016: Available at https://icon.colorado.edu - https://doi.org/10.1007/978-3-030-36683-4_36

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Author summary

The robustness of centrality measures is a fundamental problem for the correct identification of important nodes in many real networks, which are partially observed in most cases. Existing studies do not fully address this issue because they are usually limited to a small number of both centrality measures and graphs, while the graph perturbations are performed at random. Our work investigates the robustness of 16 centrality measures by analyzing the variation in the relative ranking of nodes under a set of appropriately defined network perturbations. To draw meaningful and robust conclusions about the average sensitivity of a specific centrality measure, we perform our experiments on a large set of networks. Our findings demonstrate that certain centrality measures may be misinterpreted or misapplied when used on specific classes of networks, while the results of these measures require a cautious interpretation in the presence of missing or incorrect data.

Introduction

Many complex systems, including social, infrastructural, biological and economic systems, can be described as networks with nodes representing the entities and links representing interactions between them. A fundamental challenge in studying these networks is to understand the impact (or importance) of each node. The notion of importance can be defined in different ways depending on the nature of a network or features that a researcher wants to consider while ranking nodes. Therefore, many centrality measures have been introduced in the literature. The total number of proposed indices over the graph history is overwhelming [1], ranging from classical centralities [2] to measures that take into account specific features of a network [3–5]. These measures have shown great value in understanding many complex networks such as citation networks, computer networks and biological networks. In general, the choice of the most appropriate centrality measure depends on the type of network and the interpretation of important elements.

Unfortunately, most real networks are substantially incomplete or inaccurate due to the high cost of network measurements, data collection errors, the highly dynamic nature of networks, or data privacy issues. For instance, Ficara et al. [6] examine criminal networks that suffer from data incompleteness (due to the nature of the network), data incorrectness (unintentional data collection errors and intentional deception by criminals) and data inconsistency (misleading information from different sources). Aleskerov et al. [7] analyze the banking foreign claims network, which covers about 94% of total foreign claims as some countries do not report their statistics. Meshcheryakova [8] investigates the asymmetry in trade networks as many countries report their own versions of a trade flow between them due to different commodity classification systems, different costs calculations (including/excluding transportation and insurance costs) or a time delay. Hence, some centrality measures, which are very sensitive to small changes in the graph structure, can be misused and lead to the wrong interpretation of important within these networks.

The effects of missing or incorrect data on centrality measures have been extensively studied in the literature. Most studies examine the sensitivity of centrality in artificial networks such as Erdős–Rényi (ER) random graphs, Barabási–Albert (scale-free) graphs, Watts–Strogatz (small-world) graphs and other classical graph structures [9–15]. These studies are mainly limited to the sensitivity analysis of only 4-6 classical centrality measures to a small number of structural changes (link/node removal or addition) in a graph. Furthermore, all the perturbations in the structure of a network are performed at random, which might be meaningless for real-world networks.

Some studies examine the robustness of centrality measures in real-world networks. Bolland [16] examines the performance of 4 classical centrality measures under conditions of random variations in Chillicothe data. Herland et al. [17] consider 3 classical centrality measures and their robustness to random changes in 4 real networks. Niu et al. [18] evaluate the stability of 5 centrality measures on 9 real datasets towards random link addition/removal/rewiring and have evaluated the Spearman correlation between centrality rankings. Segarra and Ribeiro [11] investigate the effect of random changes on the air traffic network and the network of interactions between sectors of the US economy. Meshcheryakova and Shvydun [19] examine 13 centrality measures in criminal and food trade networks. Costenbader and Valente [20] determine how random sampling from 59 empirical directed networks affects the stability of 11 centrality measures. While previous studies have made significant contributions to the analysis of centrality measures, several limitations remain. First, many of these studies consider only a limited number of empirical networks and, thus, do not allow to generalize the results and draw meaningful conclusions about the class of centrality measures that are most sensitive to incomplete data. Second, most studies explore only classical centrality measures (e.g., degree, betweenness, closeness, and eigenvector centralities) and, consequently, do not provide a comparative analysis with other existing measures. Furthermore, the sensitivity of centrality measures is often assessed under the assumption that data are missing at random, which is not true for many empirical networks.

This article aims to evaluate the robustness of centrality measures on a large set of different benchmark network topologies. Fig 1 illustrates the methodology of our research. We consider 113 networks of varying sizes and domains. For each network, we sequentially perform multiple changes in the structure of the graph by removing existing or adding new links, thus, simulating the most common effects of incomplete or incorrect data. Next, we evaluate the ranking of nodes in initial and modified graphs using a particular centrality measure. Finally, we compare the rankings across different networks and evaluate the robustness of 16 centrality measures. We would like to emphasize that we focus on the problem of incomplete data. Hence, we do not investigate the targeted addition or deletion of links intended to affect the network's robustness and resilience [21,22].

Materials and methods

Preliminaries

We consider a graph $G = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} = \{1, ..., N\}$ is a set of nodes, $|\mathcal{N}| = N$, and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of links, $|\mathcal{L}| = L$. The graph G is described by an $N \times N$ adjacency matrix A whose elements a_{ij} are either one or zero depending on whether there is a link between nodes i and j or not and $a_{ii} = 0$ for $\forall i \in \mathcal{N}$. The graph is called undirected if A is symmetric ($a_{ij} = a_{ji}$ for $\forall i, j \in \mathcal{N}$) and directed, otherwise. Additionally, the graph G can be described by an $N \times N$ non-negative weight matrix W, where each element w_{ij} represents the intensity of a link between nodes i and j. Given a graph G, a centrality measure $c(\cdot)$ associates a real number $c(i)$ with each node $i \in \mathcal{N}$, which is usually interpreted as follows: the larger $c(i)$ is, the more central node i should be. We focus on undirected graphs and examine centrality measures that are applicable to these structures.

Centrality measures

In this subsection, we describe the centrality measures that we study in the paper.

1. **Degree centrality** [23]. For undirected graphs, the degree of node i is equivalent to the total number of adjacent links. For weighted graphs, the intensity of a link is considered. High values of the degree centrality indicate nodes with the most connections to other nodes, making it easier for them to access and influence other nodes locally.

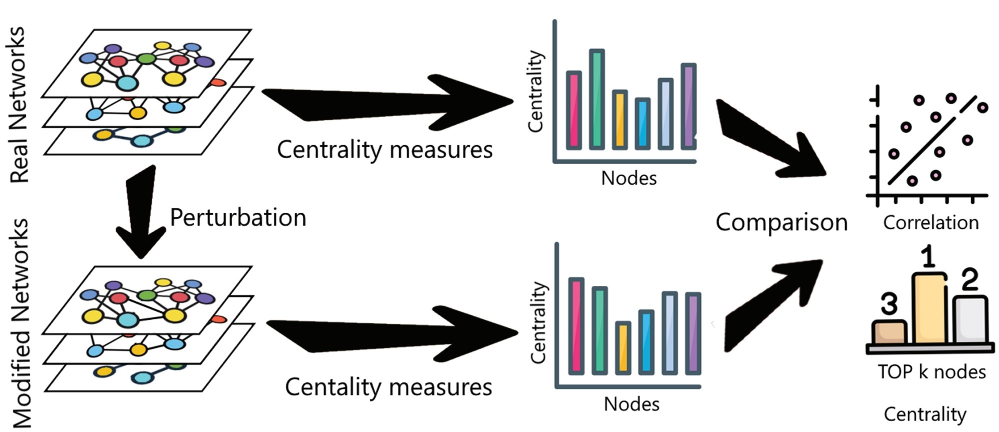

Fig 1. Perturbation analysis of centrality measures.

https://doi.org/10.1371/journal.pcsy.0000042.g001

2. ***k*-shell centrality** [24]. The centrality of node i assigns the highest order of a core that contains this node. A k-core is a maximal subgraph of G in which each node has at least degree k. Nodes with high k-shell levels tend to be located at the center of the network, giving them greater spreading capability.
3. **Collective Influence (CollInf)** [25]. The centrality of node i depends on its degree k_i and the degrees of its neighbours at a particular distance l [25], i.e.,

$$c(i) = (k_i - 1) \sum_{j:d_{ji} \leq l} (k_j - 1). \tag{1}$$

   Nodes with high collective influence have strong connections to their neighbors and are located within dense, well-connected network structures at radius l.
4. **Eigenvector centrality** [2,26]. The eigenvector centrality assigns the relative importance of node i by giving greater weight to more important neighbors, thereby contributing more to the centrality of node i than less important neighbors. The eigenvector centrality is defined by the principal eigenvector $\vec{c}$ of the adjacency matrix A with the leading eigenvalue λ_{max}, i.e.,

$$A \cdot \vec{c} = \lambda_{max} \cdot \vec{c}. \tag{2}$$

5. **Subgraph centrality** [27]. It computes the number of closed walks (walks where the first and the last nodes are the same) of different lengths in G, i.e.,

$$c(i) = \sum_{k=0}^{\infty} \frac{(A^k)_{ii}}{k!}. \tag{3}$$

   Nodes with high subgraph centrality scores are involved in many connected subgraphs and, consequently, contribute significantly to the overall connectivity and functionality of the network.
6. **PageRank** [28]. The importance of a node depends on the probability to be visited by a random walker, i.e.,

$$c(i) = \alpha \sum_{j=1}^{N} a_{ji} \cdot \frac{c(j)}{\sum_{k=1}^{N} a_{jk}} + \frac{1-\alpha}{N}, \tag{4}$$

   where $\alpha = 0.85$ is the probability to continue the walk. Nodes with high PageRank scores have the highest probability of being visited by the random walker and, consequently, play an important role in the network.
7. **Laplacian centrality** [29]. The Laplacian centrality of a node measures the network's ability to respond to the deactivation of that node. It measures the drop in the Laplacian energy after deleting a node from the graph, i.e.,

$$c(i) = \frac{E_L(G) - E_L(G_i)}{E_L(G)}, \tag{5}$$

   where G_i is the graph obtained by deleting node i from G and $E_L(G) = \sum_{k=1}^{N} \lambda_k^2$, λ_k are the eigenvalues of G's Laplacian matrix.

8. **Betweenness centrality** [30]. It measures how often a node lies on the shortest paths between other nodes, i.e.,

$$c(i) = \sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\sigma_{jk}(i)}{\sigma_{jk}}, \tag{6}$$

where σ_{jk} is the number of the shortest paths from node j to node k and $\sigma_{jk}(i)$ is the number of the shortest paths from node j to node k that pass through i. Nodes with high betweenness centrality play a vital role in facilitating the flow of information and resources throughout the network.
9. **Stress** [31]. A variant of the betweenness centrality where the centrality $c(i)$ of node i counts all of the shortest paths which pass through node i, i.e.,

$$c(i) = \sum_{j\neq k\neq i}\sigma_{jk}(i). \tag{7}$$

where $\sigma_{jk}(i)$ is the number of the shortest paths from j to k that pass through i.
10. **Current-flow betweenness** [32]. The extension of the betweenness centrality where all paths between nodes are considered. Nodes with high current-flow betweenness play a vital role in facilitating the flow of information that does not necessarily follow the shortest paths.
11. **Egocentric betweenness** [33]. The betweenness centrality of node i within its egocentric network (subgraph of G with node i and its 1-hop neighbors). Nodes with high egocentric betweenness play a vital role in facilitating the flow of information among their neighbors.
12. **Closeness centrality** [23]. It evaluates how close each node is to other nodes in the network, i.e.,

$$c(i) = \frac{N-1}{\sum_{j\neq i} d_{ij}}, \tag{8}$$

where d_{ij} is the length of the shortest path from node i to node j. Nodes with high closeness centrality are key players in quickly transmitting information along the shortest paths in the network.
13. **Harmonic centrality** [34]. A variant of the closeness centrality where the centrality of node i is computed as the sum of inverse distances to other nodes, i.e.,

$$c(i) = \sum_{j\neq i}\frac{1}{d_{ij}}. \tag{9}$$

14. **Decay** [35]. Decay centrality reflects the importance of nodes by accounting for the diminishing effects of distance. It is a variant of closeness centrality in which the distance d_{ij} between nodes i and j is subject to a decay parameter $\delta = 0.5$

$$c(i) = \sum_{j\neq i}\delta^{d_{ij}}. \tag{10}$$

15. **Current-flow Closeness** [32]. A variant of the closeness centrality, which utilizes the idea of electrical current in a network

$$c(i) = \frac{N-1}{\sum_{j \neq i} p_{ij}(i) - p_{ij}(j)}, \tag{11}$$

where $p_{ij}(i)$ is the absolute electrical potential of node *i* based on the electrical current supply from node *i* to node *j*. Nodes with high closeness centrality are key players in rapidly transmitting information that does not necessarily follow the shortest paths across the network.
16. **LRIC** [3]. LRIC is based on the concept that nodes are not homogeneous, with each node *i* having an individual threshold of influence q_i, while its neighbors may form groups to influence node *i*. Nodes with high LRIC scores have the greatest direct and indirect group influence on other nodes in the network.

For weighted networks, we consider only weighted degree, eigenvector, PageRank, Laplacian and LRIC centralities, because other betweenness-based and closeness-based measures require the definition of the shortest paths, which may differ across the datasets.

Datasets

The robustness of centrality measures is evaluated on 113 undirected networks. Table 1 provides a short description of the networks. We examine 113 research-quality networks of varying sizes and domains from the Index of Complex Networks (ICON) [36] and our previous study [19]. These networks describe various types of interactions, including social interactions between members of the US Congress (congress-Twitter [37]), members of a university karate club (karate [38]), university employees (CS-Aarhus [39]), movie characters [40], terrorists (the 9/11 terrorist network [41], Madrid2004-terrorists [42], Mali terrorist network [43], Noordin [44], RHODESBOMBING [45]), gangs (Italian gangs [46], London gang [47], Montreal street gangs [48]), criminals (e.g., Siren and Togo operations [49], cocaine smuggling [50], Sicilian Mafia [51]), households (Ugandan village networks [52]) and animals (dolphins [53] and zebras [54]). Additionally, we include one biological network (Fullerene C60 [55]), one economic network (Medieval Russian Trade [56]), 24 technological networks (Agis, Bbnplanet, Bics, Biznet, Cesnet, Chinanet, CrlNetworkServices, DeutscheTelekom, Dfn, Evolink, Forthnet, GtsCzechRepublic, GtsPoland, HurricaneElectric, Integra, Janetbackbone, Litnet, Niif, Psinet, Renater2010, Rnp, Sanet, Xeex, Xspedius [57]), two transportation networks (Amsterdam metro [58], PATH rail system [59]), and one word adjacency network (word-adj [60]). Most of the networks suffer from incompleteness. Therefore, the sensitivity analysis of centrality measures toward graph modifications is reasonable for these networks.

There are some limitations in our study related to the choice of datasets. Due to the high computational complexity of some classical centrality measures, we focus on networks with a relatively small number of nodes (most networks containing fewer than 100 nodes). Hence, the results of the study may not be generalizable to large networks.

Second, approximately 44% of the networks in our study are fictional, as they represent relationships between movie characters. Movie networks are the one-mode projections of bipartite networks, where character co-appearance in the same scene serves as a proxy for connectedness [40]. When *n* characters appear together in a scene, they form a fully connected clique of size *n* in the movie network. This results in an overlapping clique structure (a union of densely connected subgroups), which is often unnatural for real-world networks

Table 1. List of networks.

#	Dataset	*N*	*L*	#	Dataset	*N*	*L*	#	Dataset	*N*	*L*
Unweighted networks											
1	congress-Twitter	475	10222	2	dolphins	62	159	3	karate	34	78
4	mafia (calls)	100	124	5	mafia (meetings)	101	256	6	movie-11	66	211
7	movie-129	61	144	8	movie-146	63	260	9	movie-158	67	310
10	movie-178	109	326	11	movie-180	82	162	12	movie-209	69	135
13	movie-261	95	567	14	movie-271	76	185	15	movie-273	77	246
16	movie-274	60	246	17	movie-281	60	152	18	movie-298	68	214
19	movie-308	62	186	20	movie-316	94	271	21	movie-320	68	158
22	movie-337	76	200	23	movie-339	70	266	24	movie-34	95	378
25	movie-346	74	251	26	movie-356	78	219	27	movie-36	75	249
28	movie-370	64	251	29	movie-412	66	182	30	movie-426	66	209
31	movie-438	61	291	32	movie-442	70	339	33	movie-452	61	228
34	movie-456	101	293	35	movie-523	82	239	36	movie-526	76	174
37	movie-530	67	186	38	movie-541	77	318	39	movie-544	62	331
40	movie-573	61	331	41	movie-591	76	629	42	movie-632	65	258
43	movie-655	61	239	44	movie-659	99	317	45	movie-764	76	274
46	movie-803	71	264	47	movie-828	70	299	48	movie-832	75	217
49	movie-837	68	131	50	movie-840	63	591	51	movie-869	60	269
52	movie-877	76	201	53	movie-92	71	154	54	movie-93	62	284
55	terrorist-pairs	62	152	56	word-adj	112	425	57	9-11 terrorist network	60	126
58	Agis	25	30	59	Amsterdam metro	39	40	60	Bbnplanet	27	28
61	Bics	33	48	62	Biznet	29	33	63	CS-Aarhus (Facebook)	32	124
64	CS-Aarhus (coauthor)	20	21	65	CS-Aarhus (leisure)	47	88	66	CS-Aarhus (lunch)	60	193
67	CS-Aarhus (work)	60	194	68	Cesnet	44	51	69	Chinanet	42	66
70	CrlNetworkServices	33	38	71	DeutscheTelekom	30	55	72	Dfn	50	78
73	Evolink	35	43	74	Forthnet	62	62	75	Fullerene C60	60	90
76	GtsCzechRepublic	32	33	77	GtsPoland	33	37	78	HurricaneElectric	24	37
79	Integra	27	36	80	Italian gangs	65	113	81	Janetbackbone	29	45
82	Litnet	43	43	83	London gang	54	315	84	Madrid2004-terrorists	64	243
85	Mali terrorist network	36	67	86	Medieval Russian Trade	39	52	87	Montreal Street Gangs	29	75
88	Niif	36	41	89	Noordin: classmates	37	174	90	Noordin: communications	74	200
91	Noordin: friendship	59	90	92	Noordin: meetings	26	63	93	Noordin: operations	39	267
94	Noordin: training	28	130	95	PATH rail system	13	14	96	Psinet	24	25
97	RHODESBOMBING	22	66	98	Renater2010	43	56	99	Rnp	31	34
100	Sanet	43	45	101	Siren operation	44	103	102	Smuggling (ACERO)	25	37
103	Smuggling (JAKE)	38	50	104	Smuggling (JUANES)	51	93	105	Smuggling (MAMBO)	31	58
106	Togo operation	33	47	107	Ugandan village 17	65	257	108	Xeex	24	34
109	Xspedius	34	49	110	Zerbras	27	111				
Weighted networks											
1	congress-Twitter	475	10222	2	karate	34	78	3	les-miserables	77	254
4	mafia-calls	100	124	5	mafia-meetings	101	256				

https://doi.org/10.1371/journal.pcsy.0000042.t001

and can notably impact certain centrality measures. For instance, this overlapping clique structure leads to shorter average path lengths, resulting in high closeness centrality for many nodes. It also causes reduced betweenness scores for nodes within cliques, while nodes acting as bridges between cliques tend to have disproportionately high betweenness scores. A similar issue of overlapping clique structures occurs in certain terrorist and criminal networks, which are also one-mode projections derived from bipartite individual-event, individual-membership or individual-location networks [42–44,46].

Next, many movie networks are highly centralized, as primary actors tend to appear in most scenes. Fig 2 illustrates that only 16% of movie networks have low degree centralization

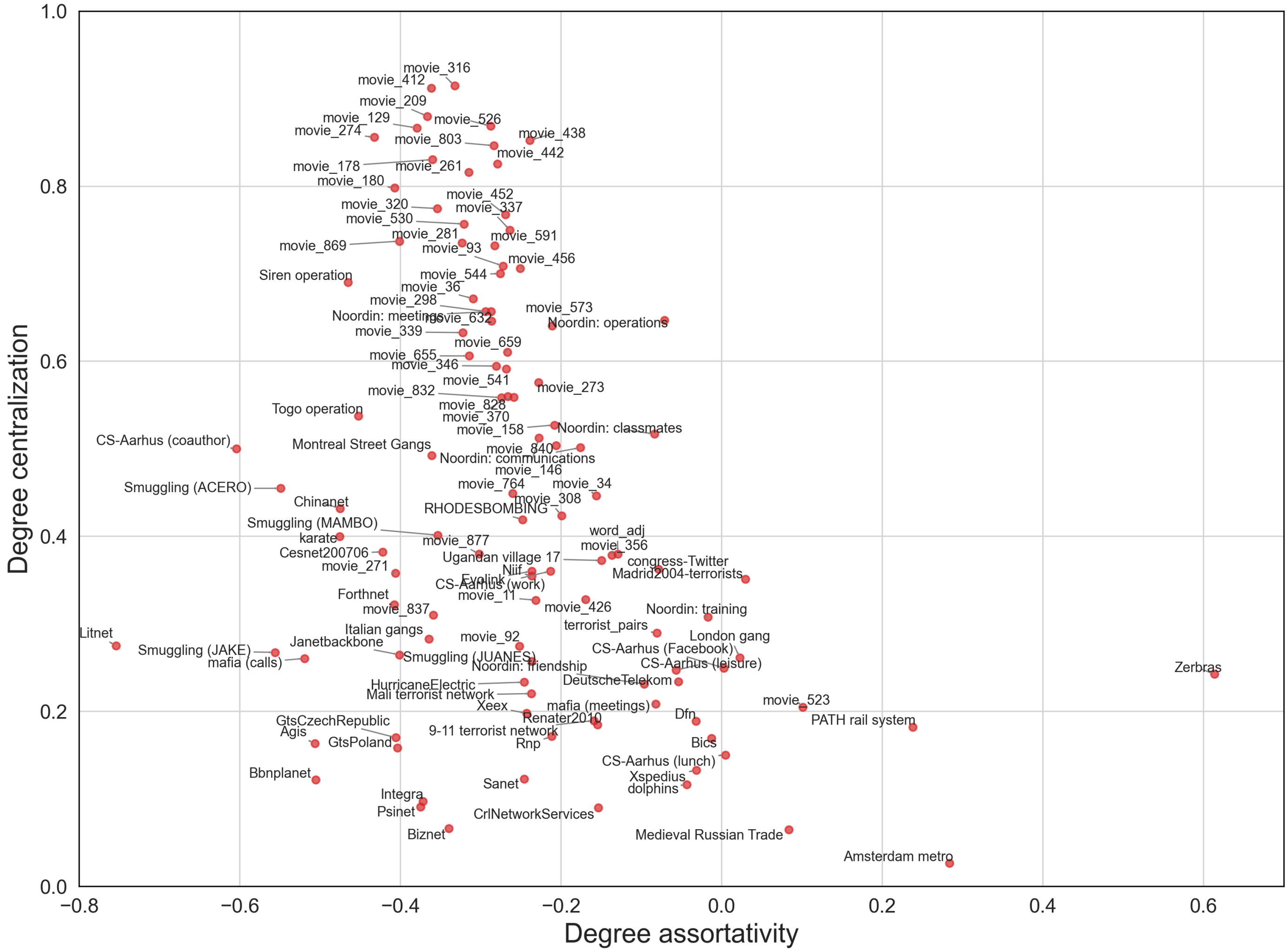

Fig 2. Degree assortativity and centralization across different datasets.

https://doi.org/10.1371/journal.pcsy.0000042.g002

[61] values (<0.4), while 59% exhibit degree centralization scores above 0.6. While movie networks serve as a valuable example of incomplete data—where many actor connections remain unobserved, similar to social network studies centered on specific individuals [62]—they may not fully capture the complexity of real-world social interactions.

Finally, our datasets primarily consist of movie (44%), social (27%) and technological (22%) networks, with other types accounting for approximately 7% of the total number of networks. Fig 2 also illustrates that approximately 91% of the networks exhibit negative degree assortativity values [63]. Therefore, our findings on the average sensitivity of classical centrality measures may not be generalizable to networks from other domains (financial, brain, biological, etc.) or to different network structures, such as those with positive degree assortativity.

The correlation between centrality measures across empirical datasets is discussed in S1 Appendix. For each of the networks, we apply various data imputation strategies and evaluate the sensitivity of centrality measures.

Data imputation

The effect of missing data on the results of centrality measures is hard to estimate as there are various causes of data incompleteness. For instance, trade networks suffer from data incompleteness because some countries do not report their trade flows (missing links), use distinct commodity classifications or evaluate bilateral trade costs differently (data inaccuracy). Protein networks have missing links as there are many undiscovered or unobserved interactions between proteins. Hence, there is no single best strategy how to modify the structure of a network and address all the issues of incomplete data. In other words, perturbation analysis should be performed with respect to the nature of a network of interest and the type of errors in the data.

Since nodes are often easier to observe in many empirical networks (e.g., social or international trade networks), we do not examine node removal or addition and instead focus solely on the relationships between nodes. Table 2 illustrates the actions that we apply to modify the structure of the networks. These actions take into account the two most common types of errors in the data: measurement errors and incomplete information about links. For unweighted networks, the RR scenario examines the robustness of centrality measures if some of the links have been identified incorrectly. For weighted networks, the RC scenario studies the effect of errors in the flow data. Other graph modifications examine the problem of missing links. The main difference between adding links strategies is the probability of choosing unobserved links between nodes. In RA, there is an equal probability of choosing the missing link. However, since missing links are not uniformly distributed in most empirical networks, we consider the DA, PA, and AA scenarios, which assume that the probability of missing links depends on the network's structure. In DA, the probability of choosing links between nodes is proportional to the product of their degrees. PA is driven by the idea of similarity between nodes, which can be estimated by the geodesic distance between them. Finally, the AA scenario is based on the Adamic-Adar index, which is often used to predict links in social networks [64]. The Adamic-Adar index $AA(i,j)$ between nodes i and j is

$$AA(i,j) = \sum_{k \in \mathcal{N}_i \cap \mathcal{N}_j} \frac{1}{log(|\mathcal{N}_k|)}, \tag{12}$$

where $\mathcal{N}_i$ denotes the set of neighbors of node i. The difference between RR, RA, DA, PA and AA scenarios is discussed in S2 Appendix.

For each network G and graph perturbation strategy, we generate 1,000 perturbed graphs $\widetilde{G}$ and then compare the similarity between centrality measures in G and $\widetilde{G}$. The parameter k

Table 2. List of modifications in the structure of a network.

			Network type	
#	**Name**	**Description**	**Unweighted**	**Weighted**
1	RR	Random removal of k% links	+	-
2	RC	Random change of link weights in the range of $[-k\%, k\%]$	-	+
3	RA	Random addition of k% new links	+	+
4	DA	Addition of k% new links with a probability that is proportional to the product of nodes degrees	+	+
5	PA	Addition of k% links with a probability that is inversely proportional to the shortest path distance between nodes	+	-
6	AA	Addition of k% links with a probability that is proportional to the Adamic–Adar index between nodes	+	-

https://doi.org/10.1371/journal.pcsy.0000042.t002

varies from 1% to 30% of the total links in the initial graph G. For weighted graphs, k is based on the total weight $w = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$ of links in G.

The list of graph changes, which is presented in Table 2, is not exhaustive. The choice of the appropriate imputation method depends on the nature of the missing data as well as the type of a network. A discussion of various data sampling and imputation techniques and the effect of missing data on network measurement are provided in [19,65–74].

Performance evaluation

In this subsection, we describe various metrics that are employed in the paper to assess the stability of the centrality measures. As illustrated in Fig 1, we first define the centrality of the nodes in the initial graph G and its modified version $\widetilde{G}$. For each centrality measure and graph modification, we rank all the nodes according to their centrality and then compare the rankings as well as the set of the most central elements. The higher the similarity between the rankings of the nodes in G and $\widetilde{G}$ is, the less sensitive the centrality is to a certain modification of the graph. The list of performance metrics is presented below.

1. **Correlation:** the Kendall rank correlation coefficient [75], which measures the similarity between two rankings of nodes as

$$\tau = \frac{N_c - N_d}{N(N-1)/2}, \tag{13}$$

where N_c and N_d are the numbers of concordant and discordant pairs.
2. **TOP1:** the Jaccard index, which measures the similarity between two sets of the most important nodes in G and $\widetilde{G}$ as

$$J(C, \widetilde{C}) = \frac{|C \cap \widetilde{C}|}{|C \cup \widetilde{C}|}, \tag{14}$$

where C and $\widetilde{C}$ denote sets of the most important nodes in G and $\widetilde{G}$. The most important nodes are the nodes with the highest centrality score.
3. **TOP3:** the Jaccard index between TOP-3 nodes in G and $\widetilde{G}$.
4. **TOP5:** the Jaccard index between TOP-5 nodes in G and $\widetilde{G}$.
5. **TOP10:** the Jaccard index between TOP-10 nodes in G and $\widetilde{G}$.

Fig 3 illustrates an example of the performed experiment for the betweenness centrality. Starting with the initial graph G, we construct a new network (one possible realization of $\widetilde{G}$) by randomly adding three links: (2,3), (5,7) and (6,8). We then evaluate and compare the betweenness centrality in graphs G and $\widetilde{G}$. The Kendall rank correlation coefficient between nodes rankings in G and $\widetilde{G}$ is 0.72, as node 2 in $\widetilde{G}$ has a higher betweenness score than node 7, node 5 ranks higher than node 1, and node 3 surpasses nodes 1 and 8. The Jaccard index for the TOP3 nodes is 1, indicating that nodes 6, 7, and 2 remain the most important in both graphs G and $\widetilde{G}$.

Since the robustness of centrality measures is evaluated in 1,000 realizations of graph $\widetilde{G}$, derived from the corresponding graph G, it is important to aggregate the result of our experiments across all the networks. We employ two methods: the average score and the Copeland score. For instance, the average correlation score of a centrality is the average correlation between graphs G and $\widetilde{G}$. The Copeland score is a social choice rule that measures the difference between the cardinality of dominating and dominated sets [76,77]. The *dominating* set of

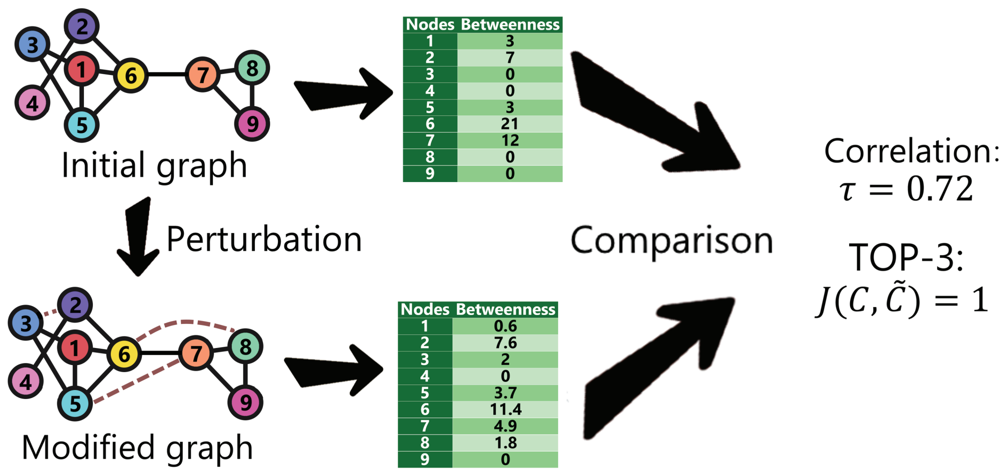

Fig 3. Perturbation analysis for the betweenness centrality: an example.

https://doi.org/10.1371/journal.pcsy.0000042.g003

centrality x includes centrality measures, which are more sensitive than x to the graph modification in most experiments. Similarly, the *dominated* set of x contains centralities that are less sensitive than x in most experiments. Therefore, the Copeland score performs a pairwise comparison of the centralities and identifies the measures that are less sensitive to perturbations in most of the experiments.

Fig 4 shows an example where 3 centrality measures are compared in 5 experiments by some performance metric X. Centralities A and B have the same average correlation coefficient, although A outperforms B in 80% of experiments. By contrast, A has the highest Copeland score as it is better than both B and C in most cases. Since our experiments explore possible realizations of the initial partially-observed network, the Copeland score provides additional insights into the ranking of centrality measures.

Results

In this subsection, we discuss the robustness of the centrality measures with respect to each graph modification. We present the results for the correlation coefficient as well as for the

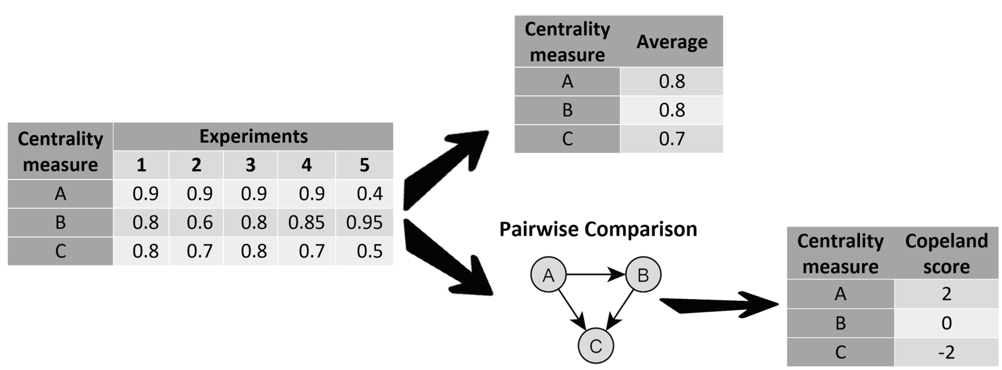

Fig 4. The difference between the average value and the Copeland score.

https://doi.org/10.1371/journal.pcsy.0000042.g004

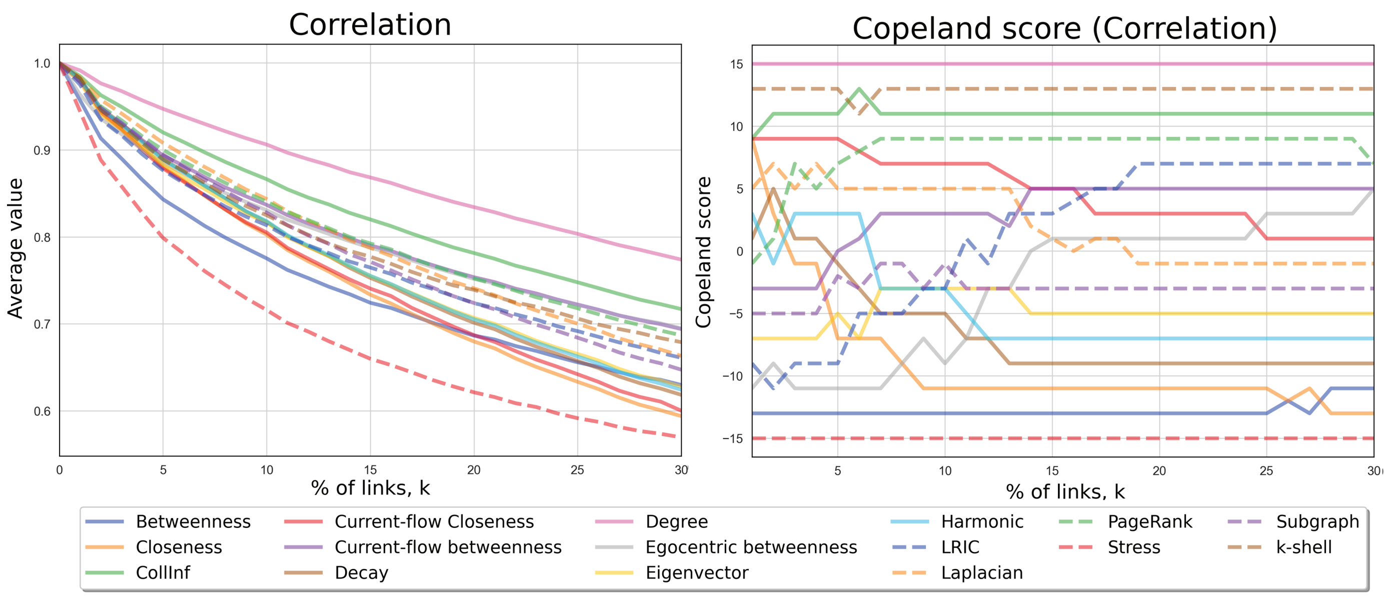

Fig 5. The average Kendal rank correlation and its Copeland score for RR.

https://doi.org/10.1371/journal.pcsy.0000042.g005

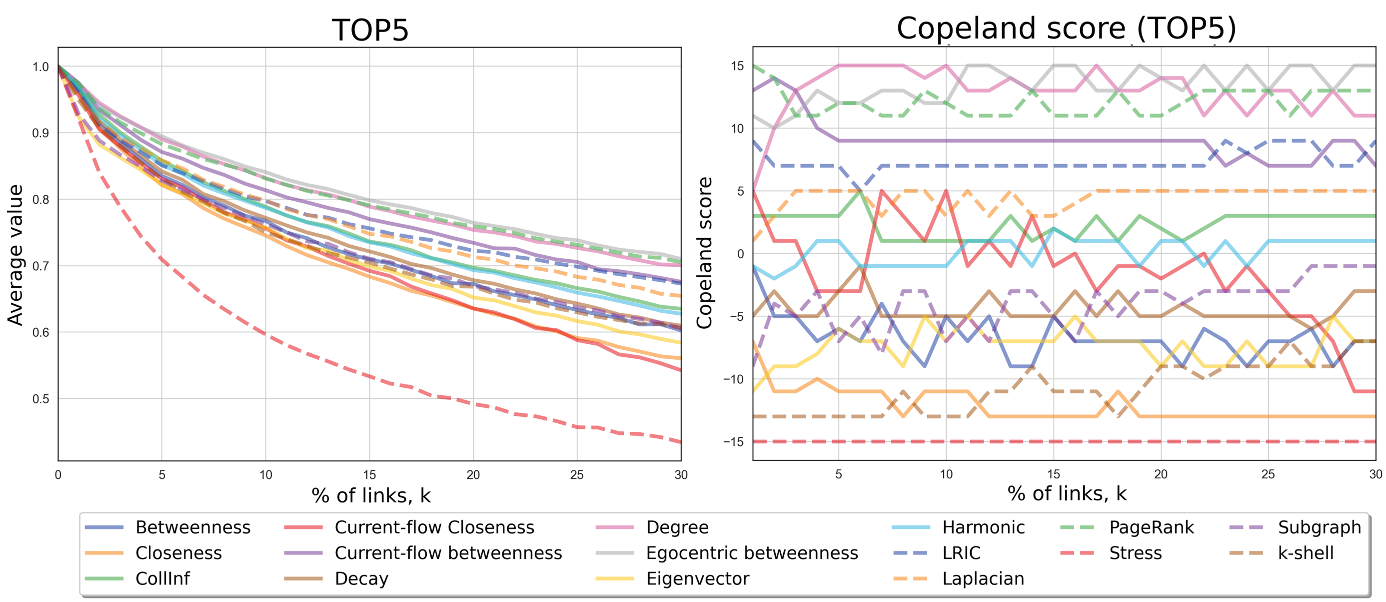

Fig 6. The average Jaccard index for TOP-5 nodes and its Copeland score for RR.

https://doi.org/10.1371/journal.pcsy.0000042.g006

TOP-5 most important nodes. The confidence intervals associated with the reported estimates are provided in S3 Appendix. Through our empirical analysis of 1,000 modified graphs, we observe that the robustness of centrality measures remains consistent across the sampled datasets. We also observe that the 95% confidence intervals are quite narrow, indicating a high level of precision in our estimates. The results for the TOP-1, TOP-3 and TOP-10 nodes agree with the results of this subsection and therefore are provided in S3-S4 Appendices.

Robustness of centrality measures in unweighted graphs

Random removal of links (RR) Fig 5 illustrates the robustness of the centrality measures with respect to the Kendal rank correlation if k% links are removed from the initial graph G. The degree centrality, which is the local measure, is the most robust measure that has a very strong correlation coefficient (≥0.78) with the observed graph G even if 30% of links are removed from the graph. By contrast, the majority of betweenness-based centralities are the most sensitive measures. The removal of links from G has a large impact on the shortest paths between the nodes, therefore, the ranking of nodes may change dramatically. At the same time, closeness, decay and harmonic centralities are more robust than the betweenness centrality because the removal of links has a higher impact on the shortest paths rather than on their lengths. Interestingly, some measures may outperform other centralities with the increase of k. For instance, the egocentric betweenness, which measures a local betweenness of nodes, becomes more robust than the closeness-based centralities for $k > 25$%. There

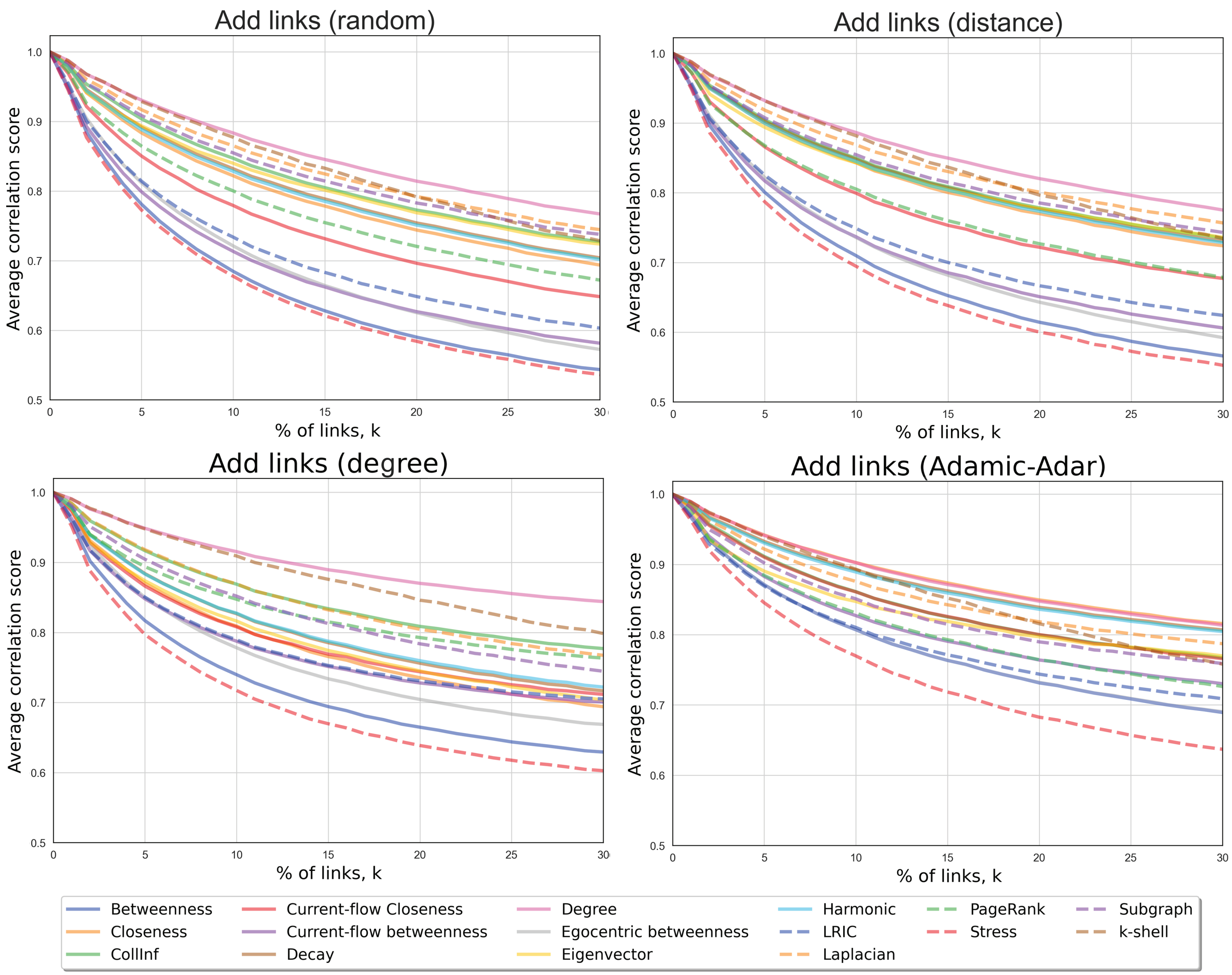

Fig 7. The average Kendal rank correlation of centrality measures for RA, DA, PA, and AA.

https://doi.org/10.1371/journal.pcsy.0000042.g007

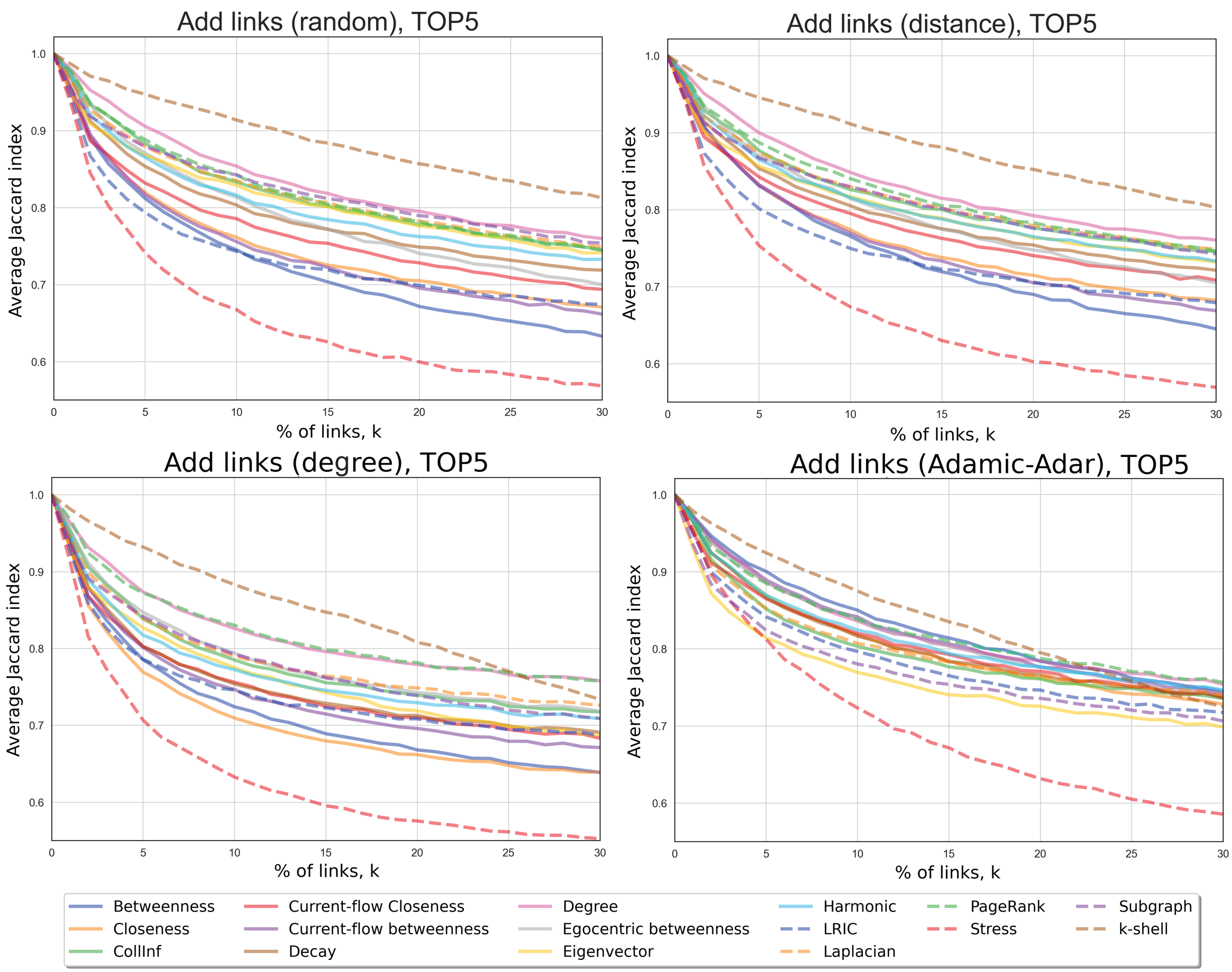

Fig 8. The comparison of TOP-5 nodes for RA, DA, PA and AA.

https://doi.org/10.1371/journal.pcsy.0000042.g008

are also centralities (LRIC and current-flow betweenness) that provide relatively more robust results for large k.

Next, we compare the sets of TOP-5 nodes with Jaccard index in Fig 6. Again, most betweenness-based centralities demonstrate the largest difference in the set of TOP-5 nodes. However, the k-shell centrality, which provides, on average, high Kendal rank correlation coefficient (see Fig 5), has a low Jaccard index for TOP-5 nodes. This observation may be caused by the fact that in many cases the set C of the most important nodes may be very large while the removal of links from G decreases the set C significantly. We can also conclude that the degree, the current-flow betweenness, PageRank, the egocentric betweenness and LRIC provide stable sets of the most important nodes.

Finally, we observe that degree, PageRank, LRIC and current-flow betweenness centralities are the least sensitive measures to the presence of k% random incorrect links in the initial undirected unweighted graph G. The current-flow betweenness centrality provides a robust

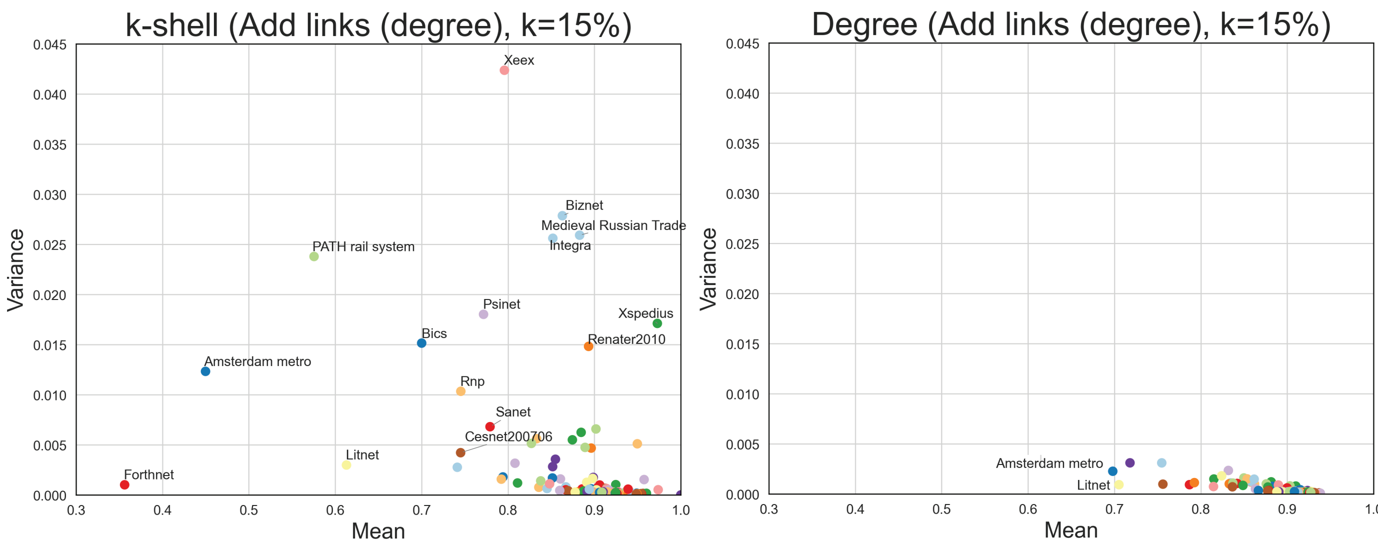

Fig 9. The mean and the variance of the Kendal rank correlation for degree and *k*-shell centralities across datasets.

https://doi.org/10.1371/journal.pcsy.0000042.g009

set of TOP-5 nodes for any observed k% while its correlation coefficient (see Fig 5) is more stable than most other measures only for $k > 10$%.

Addition of new links (RA, DA, PA and AA) Assume now that k% links are not present in a given graph G. Fig 7 illustrates the sensitivity of the centrality measures with respect to the average correlation coefficient and the link addition strategy.

First, we notice that the addition of k% links has almost the same impact on the robustness of centrality measures as the removal of k% links. However, the ranking of centrality measures is different for link addition and link removal strategies. The addition of links has the largest effect on all betweenness-based measures (stress, betweenness, egocentric betweenness and current-flow betweenness centralities) and the LRIC index. The k-shell and the degree centrality are the most robust measures for any link addition strategy. The ranking of the remaining centralities depends on the link addition strategy. Overall, we observe that the ranking of centrality measures does not differ significantly between RA, DA, PA, and AA scenarios.

Second, we compare the average Jaccard index between sets of TOP-5 nodes. Fig 8 shows that the stress centrality is the most sensitive index for all link addition scenarios. In contrast to the removal of links, the k-shell centrality provides the most robust results with respect to both the correlation score and the TOP-5 nodes for the link addition. We observe that the set of the most central nodes is relatively large and does not change significantly with the addition of new links. Among the other measures, we also observe a high Jaccard index (> 0.82) for the degree and the PageRank centralities. The Copeland score of the centrality measures is provided in S3 Appendix.

Next, we compare different link addition strategies. Fig 7 shows that the addition of random links (RA) to the graph G has the largest effect on the average correlation coefficient. The addition of links with a probability that is inversely proportional to the shortest path distance between nodes (PA) is similar to RA, which may be caused by the small diameter of real networks and, consequently, the fact that almost any link between nodes has almost the same probability. By contrast, degree-based (DA) and Adamic-Adar-based (AA) strategies have the lowest impact on the average correlation coefficient for most centralities. AA also provides the lowest sensitivity of centrality measures in terms of the TOP-5 nodes.

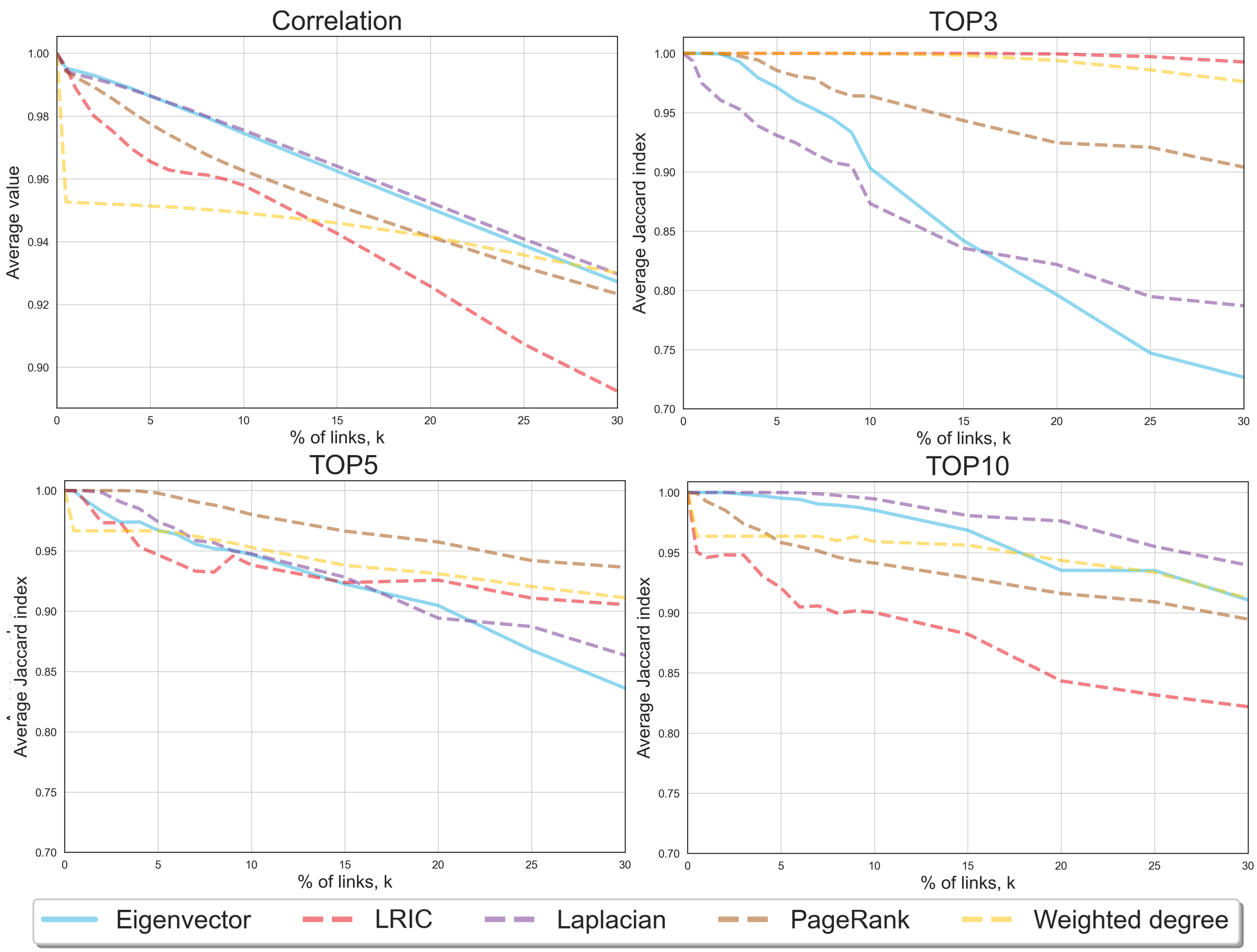

Fig 10. The average Kendal rank correlation and the average Jaccard index of centrality measures for RC.

https://doi.org/10.1371/journal.pcsy.0000042.g010

Finally, we emphasize that the presented results demonstrate the average sensitivity of centrality measures across 110 unweighted networks from Table 1. In principle, the robustness of centrality measures depends on the network and the graph perturbation strategy. For instance, Fig 9 illustrates how the mean and the variance of the Kendal rank correlation vary across datasets for DA (k = 15%). We consider degree (left) and k-shell (right) centralities, which are the most robust measures according to Fig 7. Each point corresponds to a particular dataset. Fig 9 proves that the robustness of the centrality measures depends on the dataset. For instance, the robustness of the k-shell centrality in some transportation and technological networks varies significantly, which can be attributed to their smaller size and lower density, where changes in the network structure have a greater impact on centrality values, depending on where links are added.

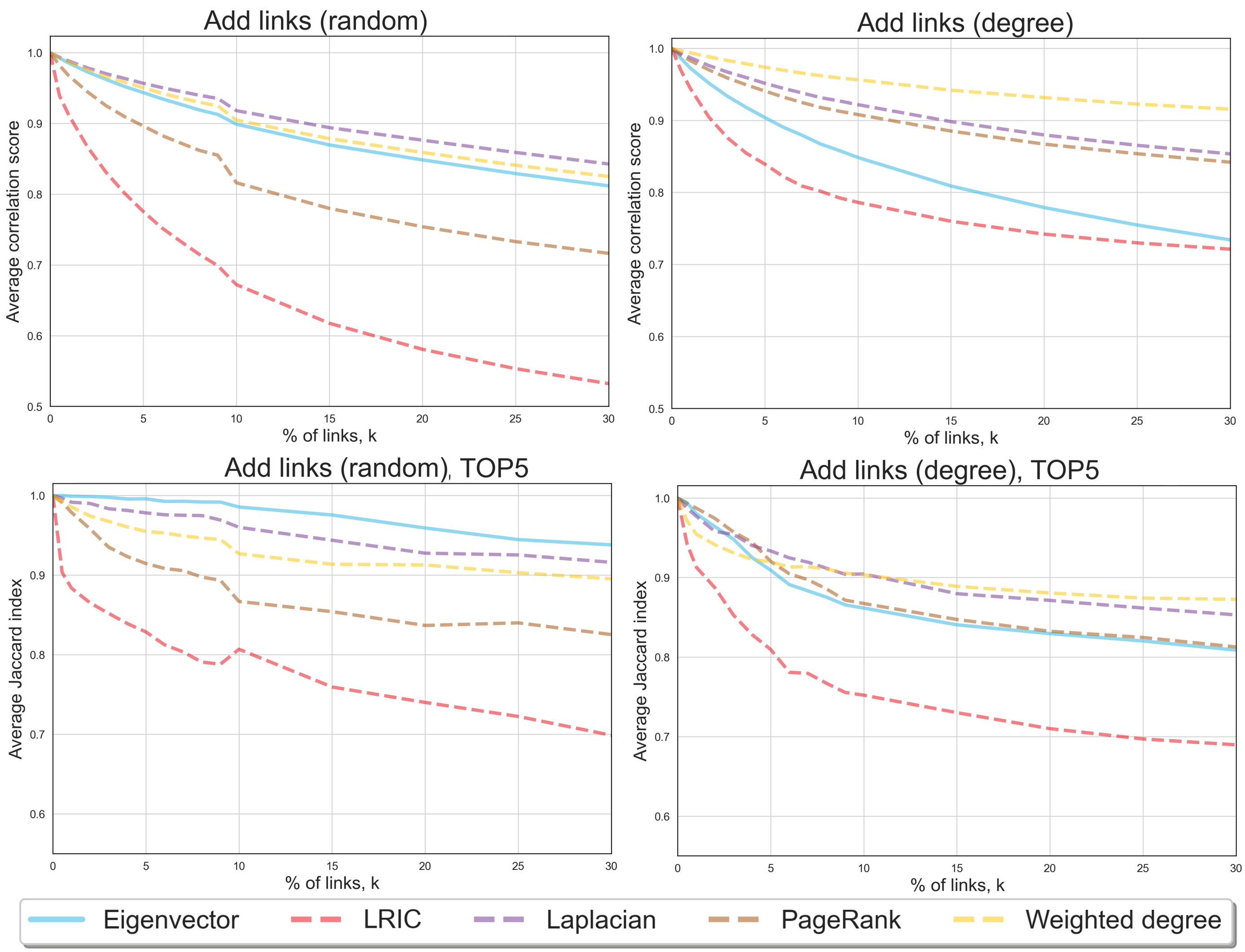

Fig 11. The average Kendal rank correlation and the average Jaccard index of centrality measures for RA and DA.

https://doi.org/10.1371/journal.pcsy.0000042.g011

Robustness of centrality measures in weighted graphs

Random change of link weights (RC) Fig 10 illustrates the robustness of centrality measures if the graph G contains errors in link weights. The confidence intervals associated with the reported estimates are provided in S4 Appendix.

Laplacian and eigenvector centralities demonstrate the highest robustness of nodes rankings. However, the sensitivity of centrality measures in terms of r most central elements varies with respect to r. Both LRIC and weighted degree provide quite robust sets of $r = 3$ most central nodes (TOP-3) but for $r = 10$ the most stable results are given by the Laplacian and eigenvector centralities. In other words, for LRIC and weighted degree, the perturbation of G has the largest effect on the ranking of TOP-10 nodes, which are not in TOP-3 and TOP-5. By contrast, for Laplacian and eigenvector centralities, the graph perturbation changes the ordering of TOP-10 nodes but not the list of these nodes. Overall, we observe that all centrality measures have a very strong average correlation coefficient (≥0.89) between rankings of nodes in graphs G and $\widetilde{G}$ even if there is up to 30% inaccuracy in the weights of the links.

Moreover, random changes in link weights (RC) do not significantly influence the centrality of nodes in the network.

Addition of new links (RA and DA) The effect of the link addition to weighted graphs is shown in Fig 11. The average Kendal rank correlation is higher in DA (degree-based link addition) than in RA (random addition of new links) for all centrality measures. LRIC and PageRank are the most sensitive to missing links in terms of the average correlation coefficient and the set of TOP-5 nodes under RA strategy. Other centrality measures are quite stable: the average Jaccard similarity between TOP5 nodes in graphs G and $\widetilde{G}$ is greater than 0.9 while the average Kendall rank correlation of nodes rankings is strong (> 0.8).

For DA strategy, weighted degree centrality provides the most stable results as DA favors high-degree nodes. On the contrary, the eigenvector centrality, which is among the most robust measures in RA, becomes more vulnerable in DA compared to other centrality measures. We also conclude that the LRIC index is the most sensitive measure for both RA and DA.

Discussion

The problem of data inaccuracy and incompleteness is a serious challenge for the analysis of complex systems. Since most real networks are partially observed, some centrality measures can be misused and lead to wrong interpretation. In this regard, the choice of the most appropriate centrality measure requires a careful examination.

We discussed the sensitivity of 16 centrality measures to different data imputation techniques. To draw meaningful and robust conclusions about the average sensitivity of different centrality measures, we have performed experiments on a large set of different benchmark network topologies.

Our main observation for unweighted networks is that the addition or the removal of new links has the largest impact on the betweenness-based centralities. Degree and PageRank centralities are among the least sensitive measures to link addition/removal. We have also identified some measures that are more robust to the link removal than to the link addition and vice versa. In general, there is also no evidence of considerable changes in the relative ranking of centrality measures in all the link addition strategies discussed.

For weighted networks, $k \leq 30\%$ inaccuracy in link weights does not significantly affect the centralities of the nodes (except LRIC). The existence of missing links has a greater influence on the robustness of the centrality measures than the presence of errors in the weights of the links. PageRank and Laplacian centrality are the most robust measures with respect to random link addition (RA) and degree-based link addition (DA) scenarios. Finally, we observe that the perturbations in weighted networks have a lower impact on node centrality than the perturbations in unweighted networks.

The results of our experiments provide an overall picture of the robustness of centrality measures. More precisely, we demonstrate how the centrality measures behave *on average* (in most experiments on real networks) to a specific type of graph modification. Our experiments do not aim to demonstrate that there exists the most robust measure for all networks because, in general, the perturbation analysis should be conducted with respect to the nature of errors in the network of interest. However, our study emphasizes the problem of the sensitivity of centrality measures in the presence of incomplete data.

As a limitation, our analysis is performed on small empirical networks due to the high computational complexity of some centrality measures. Hence, our findings from small networks may not fully generalize to larger networks, which typically exhibit more complex structures. Our experiments are conducted on a large set of networks, primarily consisting of social, technological, and movie networks. Many of these networks are fictional or represent one-mode projections derived from bipartite co-appearance, co-event, or co-location networks, and this structure affects certain centrality measures. As a result, the findings from the sensitivity analysis may not necessarily generalize to networks from other domains. Comparing centrality measures across larger networks, as well as networks from a wider variety of empirical domains, is an important next step in our future research.

Finally, our work is not intended to demonstrate the deficiency of some centrality indices, but rather to show that some centralities require a cautious interpretation in the presence of missing or incorrect links in the networks.

Supporting information

S1 Appendix. Correlation of centrality measures across empirical datasets.
(PDF)

S2 Appendix. Data imputation methods.
(PDF)

S3 Appendix. Comparison of TOP-k nodes and confidence intervals with respect to link addition/removal strategies in unweighted graphs.
(PDF)

S4 Appendix. Comparison of TOP-k nodes and confidence intervals with respect to link addition/removal strategies in weighted graphs.
(PDF)

Acknowledgments

Sergey Shvydun sincerely thank Natalia Meshcheryakova from the National Research University Higher School of Economics (HSE University) for her invaluable support, insightful comments and contributions, particularly in visualization and manuscript review.

References

1. Mahdi J. Centiserver: the most comprehensive centrality resource and web application for centrality measures calculation. 2024. Available from: https://www.centiserver.ir/centrality/list/
2. Newman MEJ. Networks: an introduction. Oxford, UK: Oxford University Press; 2010.
3. Aleskerov F, Shvydun S, Meshcheryakova N. New centrality measures in networks: how to take into account the parameters of the nodes and group influence of nodes to nodes. 1st edn. Chapman and Hall/CRC; 2021. https://doi.org/10.1201/9781003203421
4. Myerson RB. Graphs and cooperation in games. Math OR. 1977;2(3):225–9. https://doi.org/10.1287/moor.2.3.225
5. Chanhyun Kang, Molinaro C, Kraus S, Shavitt Y, Subrahmanian VS. Diffusion centrality in social networks. In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2012. p. 558–64. https://doi.org/10.1109/asonam.2012.95
6. Ficara A, Cavallaro L, Curreri F, Fiumara G, De Meo P, Bagdasar O, et al. Criminal networks analysis in missing data scenarios through graph distances. PLoS One. 2021;16(8):e0255067. https://doi.org/10.1371/journal.pone.0255067 PMID: 34379625
7. Aleskerov F, Andrievskaya I, Nikitina A, Shvydun S. Key borrowers detected by the intensities of their interactions. In: Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning (In 4 volumes), vol. 1, chap. 9. Singapore: World Scientific; 2020. p. 355–89. https://doi.org/10.1142/9789811202391_0009
8. Meshcheryakova N. Network analysis of bilateral trade data under asymmetry. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), The Hague, Netherlands; 2020. p. 379–83. https://doi.org/10.1109/ASONAM49781.2020.9381408
9. Borgatti SP, Carley KM, Krackhardt D. On the robustness of centrality measures under conditions of imperfect data. Social Networks. 2006;28(2):124–36. https://doi.org/10.1016/j.socnet.2005.05.001

10. Frantz TL, Cataldo M, Carley KM. Robustness of centrality measures under uncertainty: examining the role of network topology. Comput Math Organ Theory. 2009;15(4):303–28. https://doi.org/10.1007/s10588-009-9063-5
11. Segarra S, Ribeiro A. Stability and continuity of centrality measures in weighted graphs. IEEE Trans Signal Process. 2016;64(3):543–55. https://doi.org/10.1109/ICASSP.2015.7178599
12. Martin C, Niemeyer P. Influence of measurement errors on networks: estimating the robustness of centrality measures. Net Sci. 2019;7(2):180–95. https://doi.org/10.1017/nws.2019.12
13. Murai S, Yoshida Y. Sensitivity analysis of centralities on unweighted networks. In: The World Wide Web Conference. 2019. p. 1332–42. https://doi.org/10.1145/3308558.3313422
14. Meshcheryakova N, Shvydun S. Perturbation analysis of centrality measures. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining. IEEE; 2023. p. 407–14. https://doi.org/10.1145/3625007.3627590
15. Meshcheryakova N, Shvydun S. A comparative analysis of centrality measures in complex networks. ARC. 2024;85(8):773–84. https://doi.org/10.31857/s0005117924080031
16. Bolland JM. Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks. Soc Netw. 1988;10(3):233–53. https://doi.org/10.1016/0378-8733(88)90014-7
17. Herland M, Pastran P, Zhu X. An empirical study of robustness of network centrality scores in various networks and conditions. In: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence. 2013. p. 221–8. https://doi.org/10.1109/ictai.2013.42
18. Niu Q, Zeng A, Fan Y, Di Z. Robustness of centrality measures against network manipulation. Phys A: Statist Mech Appl. 2015;438:124–31. https://doi.org/10.1016/j.physa.2015.06.031
19. Meshcheryakova N, Shvydun S. Robustness of centrality measures under incomplete data. In: Cherifi H, Rocha L.M, Cherifi C, Donduran M (eds.) Complex networks & their applications XII. COMPLEX NETWORKS 2023. Studies in Computational Intelligence, vol. 1143. Cham: Springer; 2024. https://doi.org/10.1007/978-3-031-53472-0_27
20. Costenbader E, Valente TW. The stability of centrality measures when networks are sampled. Soc Netw. 2003;25(4):283–307. https://doi.org/10.1016/s0378-8733(03)00012-1
21. Albert R, Jeong H, Barabasi A. Error and attack tolerance of complex networks. Nature. 2000;406(6794):378–82. https://doi.org/10.1038/35019019 PMID: 10935628
22. Artime O, Grassia M, De Domenico M, Gleeson JP, Makse HA, Mangioni G, et al. Robustness and resilience of complex networks. Nat Rev Phys. 2024;6(2):114–31. https://doi.org/10.1038/s42254-023-00676-y
23. Freeman LC. Centrality in social networks conceptual clarification. Soc Netw. 1978;1(3):215–39. https://doi.org/10.1016/0378-8733(78)90021-7
24. Batagelj V, Zaversnik M. Generalized cores, arXiv:0202039 [Preprint]. 2002 [cited 2024 May 30]. Available from: https://arxiv.org/abs/cs/0202039
25. Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. Nature. 2015;524(7563):65–8. https://doi.org/10.1038/nature14604 PMID: 26131931
26. Bonacich P. Power and centrality: a family of measures. Am J Sociol. 1987;92(5):1170–82.
27. Estrada E, Rodríguez-Velázquez JA. Subgraph centrality in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2005;71(5 Pt 2):056103. https://doi.org/10.1103/PhysRevE.71.056103 PMID: 16089598
28. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Comput Netw ISDN Syst. 1998;30(1–7):107–17. https://doi.org/10.1016/s0169-7552(98)00110-x
29. Qi X, Fuller E, Wu Q, Wu Y, Zhang C-Q. Laplacian centrality: a new centrality measure for weighted networks. Inf Sci. 2012;194:240–53. https://doi.org/10.1016/j.ins.2011.12.027
30. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977;40(1):35–41.
31. Shimbel A. Structural parameters of communication networks. Bullet Math Biophys. 1953;15(4):501–7. https://doi.org/10.1007/bf02476438
32. Brandes U, Fleischer D. Centrality measures based on current flow. In: Diekert V, Durand B (eds.) STACS 2005. Lecture Notes in Computer Science, vol. 3404. Berlin, Heidelberg: Springer; 2005. https://doi.org/10.1007/978-3-540-31856-9_44
33. Freeman LC. Centered graphs and the structure of ego networks. Math Soc Sci. 1982;3(3):291–304. https://doi.org/10.1016/0165-4896(82)90076-2
34. Beauchamp MA. An improved index of centrality. Behav Sci. 1965;10:161–3. https://doi.org/10.1002/bs.3830100205 PMID: 14284290
35. Jackson M. Social and economic networks. Princeton: Princeton University Press; 2008.

36. Clauset A, Tucker E, Sainz M. The Colorado index of complex networks. 2016. Available from: https://icon.colorado.edu/
37. Fink CG, Omodt N, Zinnecker S, Sprint G. A Congressional Twitter network dataset quantifying pairwise probability of influence. Data Brief. 2023;50:109521. https://doi.org/10.1016/j.dib.2023.109521 PMID: 37701709
38. Zachary WW. An information flow model for conflict and fission in small groups. J Anthropol Res. 1977;33(4):452–73. https://doi.org/10.1086/jar.33.4.3629752
39. Magnani M, Micenkova B, Rossi L. Combinatorial analysis of multiple networks. arXiv:1303.4986 [Preprint]. 2013 [cited 2024 May 30]. Available from: https://arxiv.org/abs/1303.4986
40. Kaminski J, Schober M, Albaladejo R, Zastupailo O, Hidalgo C. Moviegalaxies-social networks in movies. 2012. Available from: http://moviegalaxies.com
41. Krebs VE. Mapping networks of terrorist cells. Connections. 2002;24(3):43–52.
42. Hayes B. Connecting the dots. Am Scientist. 2006;94(5):400–4.
43. Walther OJ, Christopoulos D. Islamic Terrorism and the Malian Rebellion. Terrorism Political Violence. 2014;27(3):497–519. https://doi.org/10.1080/09546553.2013.809340
44. Roberts N. Roberts and everton terrorist data: Noordin top terrorist network (subset). Machinereadable data file. 2011. Available from: https://www.thearda.com/data-archive?fid=TERRNET
45. Rhodes CJ, Jones P. Inferring missing links in partially observed social networks. J Oper Res Soc. 2009;60(10):1373–83. https://doi.org/10.1057/jors.2008.110
46. Coutinho J. Italian Gangs. 2016. Available from: https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/italiangangs
47. Grund TU, Densley JA. Ethnic homophily and triad closure. J Contemp Criminal Justice. 2014;31(3):354–70. https://doi.org/10.1177/1043986214553377
48. Descormiers K, Morselli C. Alliances, conflicts, and contradictions in Montreal's street gang landscape. Int Criminal Justice Rev. 2011;21(3):297–314. https://doi.org/10.1177/1057567711418501
49. Morselli C. Inside criminal networks. New York: Springer; 2009.
50. Giménez-Salinas A, Framis A. Illegal networks or criminal organizations: power, roles and facilitators in four cocaine trafficking structures. In: Morselli C, editors. Crime and Networks. Routledge; 2011. p. 131–47.
51. Ficara A, Cavallaro L, De Meo P, Fiumara G, Catanese S, Bagdasar O et al. Social network analysis of Sicilian Mafia interconnections. In: Cherifi H, Gaito S, Mendes J, Moro E, Rocha L, editors.Complex networks and their applications VIII: Volume 2 Proceedings of the Eighth International Conference on Complex Networks and Their Applications; 2020: p. 440–50. https://doi.org/10.1007/978-3-030-36683-4_36
52. Chami GF, Ahnert SE, Kabatereine NB, Tukahebwa EM. Social network fragmentation and community health. Proc Natl Acad Sci U S A. 2017;114(36):E7425–31. https://doi.org/10.1073/pnas.1700166114 PMID: 28739906
53. Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Behav Ecol Sociobiol. 2003;54(4):396–405. https://doi.org/10.1007/s00265-003-0651-y
54. Sundaresan SR, Fischhoff IR, Dushoff J, Rubenstein DI. Network metrics reveal differences in social organization between two fission-fusion species, Grevy's zebra and onager. Oecologia. 2007;151(1):140–9. https://doi.org/10.1007/s00442-006-0553-6 PMID: 16964497
55. Aref S, Wilson MC. Balance and frustration in signed networks. J Complex Netw. 2018;7(2):163–89. https://doi.org/10.1093/comnet/cny015
56. Pitts FR. The medieval river trade network of Russia revisited. Soc Netw. 1978;1(3):285–92. https://doi.org/10.1016/0378-8733(78)90025-4
57. Knight S, Nguyen HX, Falkner N, Bowden R, Roughan M. The internet topology zoo. IEEE J Select Areas Commun. 2011;29(9):1765–75. https://doi.org/10.1109/jsac.2011.111002
58. Gemeente Amsterdam Klaas-Bindert de Haan. (n.d.). Maps Amsterdam. https://maps.amsterdam.nl/
59. PATH Rail System: Schedules and Maps. [Internet]. Available from: https://www.panynj.gov/path/en/schedules-maps.html
60. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E Stat Nonlin Soft Matter Phys. 2006;74(3 Pt 2):036104. https://doi.org/10.1103/PhysRevE.74.036104 PMID: 17025705
61. Freeman LC. Centrality in social networks conceptual clarification. Soc Netw. 1978;1(3):215–39. https://doi.org/10.1016/0378-8733(78)90021-7

62. Mapping the Trump World. Trump World. [Internet]. Available from: https://sharmila-velamur.github.io/trumpworld/home.html

63. Newman MEJ. Mixing patterns in networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2003;67(2 Pt 2):026126. https://doi.org/10.1103/PhysRevE.67.026126 PMID: 12636767

64. Liben-Nowell D, Kleinberg J. The link prediction problem for social networks. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management. 2003. p. 556–9. https://doi.org/10.1145/956863.956972

65. Galaskiewicz J. Estimating point centrality using different network sampling techniques. Soc Netw. 1991;13(4):347–86. https://doi.org/10.1016/0378-8733(91)90002-b

66. Huisman M. Imputation of missing network data: some simple procedures. J Soc Struct. 2009;10(1):1–29.

67. Huisman M, Krause RW. Imputation of missing network data. In: Alhajj R, Rokne J, editors. Encyclopedia of social network analysis and mining. Springer, New York, NY; 2017. pp. 1-10. https://doi.org/10.1007/978-1-4614-7163-9_394-1

68. Smith JA, Moody J. Structural effects of network sampling coverage I: nodes missing at random. Soc Netw. 2013;35(4):652–68. https://doi.org/10.1016/j.socnet.2013.09.003 PMID: 24311893

69. Smith JA, Moody J, Morgan J. Network sampling coverage II: the effect of non-random missing data on network measurement. Soc Netw. 2017;48:78–99. https://doi.org/10.1016/j.socnet.2016.04.005 PMID: 27867254

70. Žnidaršič A, Ferligoj A, Doreian P. Stability of centrality measures in valued networks regarding different actor non-response treatments and macro-network structures. Net Sci. 2017;6(1):1–33. https://doi.org/10.1017/nws.2017.29

71. Krause RW, Huisman M, Steglich C, Snijders TAB. Missing network data a comparison of different imputation methods. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE; 2018. p. 159–63. https://doi.org/10.1109/asonam.2018.8508716

72. Krause RW, Huisman M, Steglich C, Snijders T. Missing data in cross-sectional networks – an extensive comparison of missing data treatment methods. Soc Netw. 2020;62:99–112. https://doi.org/10.1016/j.socnet.2020.02.004

73. Kossinets G. Effects of missing data in social networks. Soc Netw. 2006;28(3):247–68. https://doi.org/10.1016/j.socnet.2005.07.002

74. Smith J, Morgan J, Moody J. Network sampling coverage III: imputation of missing network data under different network and missing data conditions. Soc Netw. 2022;68:148–78. https://doi.org/10.1016/j.socnet.2021.05.0025

75. Kendall M. Rank correlation methods, 4th edn. Griffin, London; 1970.

76. Saari DG, Merlin VR. The Copeland method. Econ Theory. 1996;8(1):51–76. https://doi.org/10.1007/bf01212012

77. Shvydun S. Normative properties of multi-criteria choice procedures and their superpositions: I. Working paper WP7/2015/07 (Part 1). Moscow: HSE Publishing House. arXiv:1611.00524 [Preprint]. 2015 [cite 2024 May 30]. Available from: https://arxiv.org/abs/1611.00521