Influence of measurement errors on networks

Martin, Christoph; Niemeyer, Peter

Published in:
Network Science

DOI:
10.1017/nws.2019.12

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Network Science **7** (2): 180–195, 2019
doi:10.1017/nws.2019.12

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Influence of measurement errors on networks: Estimating the robustness of centrality measures

Christoph Martin* and Peter Niemeyer

Institute of Information Systems, Leuphana University of Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany
*Corresponding author. Emails: cmartin@uni.leuphana.de, niemeyer@uni.leuphana.de

Action Editor: Ulrik Brandes

Abstract

Most network studies rely on a measured network that differs from the underlying network which is obfuscated by measurement errors. It is well known that such errors can have a severe impact on the reliability of network metrics, especially on centrality measures: a more central node in the observed network might be less central in the underlying network. Previous studies have dealt either with the general effects of measurement errors on centrality measures or with the treatment of erroneous network data. In this paper, we propose a method for estimating the impact of measurement errors on the reliability of a centrality measure, given the measured network and assumptions about the type and intensity of the measurement error. This method allows researchers to estimate the robustness of a centrality measure in a specific network and can, therefore, be used as a basis for decision-making. In our experiments, we apply this method to random graphs and real-world networks. We observe that our estimation is, in the vast majority of cases, a good approximation for the robustness of centrality measures. Beyond this, we propose a heuristic to decide whether the estimation procedure should be used. We analyze, for certain networks, why the eigenvector centrality is less robust than, among others, the pagerank. Finally, we give recommendations on how our findings can be applied to future network studies.

Keywords: centrality measures, measurement error, missing data, robustness

1. Introduction

Measurement errors in network data are a central problem in the field of network analysis, as virtually all empirical network data are affected by some kind of measurement error. Previous research has shown that these errors often have a major impact on the results of network measures, especially on centrality measures (Costenbader & Valente, 2003; Smith & Moody, 2013). For example, a more central node in the measured (erroneous) network might be less central in the hidden (unobserved, error-free) network.

Currently, most applied network studies only report that measurement errors might have affected the data collection (e.g. due to the absence of actors on the day of the survey (Wang et al., 2016), due to the study design, or due to the quality of some external data source (Fischer et al., 2018)). Most of the time, however, the impact of these measurement errors on centrality measures is not quantified. This might be due to the fact that there is currently no established way to estimate the impact of these of measurement errors on centrality measures. In this paper, we present a method to approximate this impact, given a measured network and some hypotheses about the underlying error mechanism.

Current research can be divided into two main categories: impact studies and treatment studies.

In impact studies, researchers have investigated the impact that different types of measurement errors have on the reliability of centrality measures in the case of random graphs (Borgatti et al., 2006; Frantz et al., 2009; Wang et al., 2012) and real-world networks (Costenbader & Valente, 2003; Kim & Jeong, 2007; Wang et al., 2012; Smith & Moody, 2013; Platig et al., 2013; Silk et al., 2015; Niu et al., 2015; Lee & Pfeffer, 2015), cf. Smith et al. (2017) for an extensive survey.

These studies provide guidelines for researchers on how to design future studies (e.g., what kind of measurement error might be especially harmful in a given scenario) and suggestions on which centrality measure might be more reliable in a given scenario (Smith & Moody, 2013). Unfortunately, it is difficult to identify general patterns for the reliability of centrality measures based on network metrics. As common sense suggests, centrality measures become less reliable with an increasing level of error, but the particular relationship between error level and reliability is highly dependent on the type of measurement error, the centrality measure, and the network structure (Wang et al., 2012).

There are also studies which address how to treat erroneous network data, in order to reconstruct the unknown true network. Such treatments can, for example, be used to reconstruct partially observed networks or to estimate the network statistics of the underlying network (Butts, 2003; Huisman, 2009; Handcock & Gile, 2010; Kim & Leskovec, 2011; Frantz & Carley, 2017; Wang et al., 2016; Newman, 2018; Krause et al., 2018; Žnidaršič et al., 2018).

Our contribution connects these two areas. We propose a method for estimating the impact of measurement errors on the reliability of a centrality measure, given the measured network and assumptions about the type and intensity of the measurement error. This method allows researchers to measure the robustness of a centrality measure in a specific network and can, therefore, be used as a basis for decision-making — for example, to decide whether the centrality values are reliable enough for the purposes of the study or whether one of the aforementioned treatment procedures should be applied. One of the strengths of this method is that the estimates are easy to calculate. Simply explained, we apply the assumed error mechanism (e.g., random removal of 10% of the edges) several times, independently, to the measured network and suggest the mean impact of this procedure as the estimate for the measurement error between the unknown true network and the measured network.

We test this method in various simulation scenarios based on random graphs and real-world networks as well. We find that the estimation works in many cases, especially at lower error levels (e.g., 10% missing edges or vertices). At higher error levels (e.g. 30% missing edges or vertices), the estimation still works for degree centrality and pagerank. For sparse or small networks the situation is more challenging, especially in the case of eigenvector centrality.

The rest of this paper is organized as follows: we formalize the concepts of robustness and error mechanisms in Section 2. The estimation method is presented in Section 3, and the experiments are described and discussed in Section 4. A summary and concluding recommendations can be found in Section 5.

2. Basic concepts

Let G be an undirected, unweighted, finite graph with vertex set $V(G)$ and edge set $E(G)$. A centrality measure c is a real-valued function that assigns centrality values to all nodes in a graph and is invariant to structure-preserving mappings, that is, centrality values depend solely on the structure of a graph. External information (e.g., node or edge attributes) has no influence on the centrality values (Koschützki et al., 2005). We denote the centrality value for node $u \in V(G)$ by $c_G(u)$ and the centrality values for all nodes in G $(u_1, u_2, \ldots, u_n)$ by the vector $c(G) := (c(u_1), \ldots, c(u_n))$.

The following centrality measures are used in this study: closeness centrality, betweenness centrality (Freeman, 1978), degree centrality, eigenvector centrality (Bonacich, 1987), and the

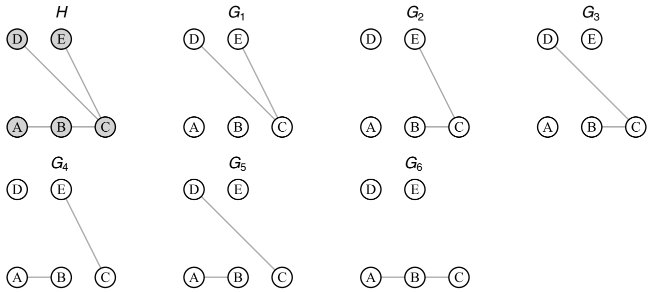

Figure 1. In this example, φ is defined as the error mechanism "50% of all edges are missing uniformly at random". Hence, $\varphi(H)$ is a random graph with possible outcomes $\Omega = \{G_1, G_2, \ldots, G_6\}$ and $P(G_i) = \frac{1}{6}$.

pagerank (damping factor 0.85) (Brin & Page, 1998). All centrality measures are calculated using the igraph library (version 0.7.1, Csardi & Nepusz (2006)).

Let G and G' be two graphs and c a centrality measure. A pair of nodes $u, v \in V(G) \cap V(G')$ and $u \neq v$ is called concordant w.r.t. c if both nodes have distinct centrality values and the order of u and v is the same in $c(G)$ and $c(G')$, that is, either $c_G(u) < c_G(v)$ and $c_{G'}(u) < c_{G'}(v)$ or $c_G(u) > c_G(v)$ and $c_{G'}(u) > c_{G'}(v)$. A pair of nodes is called discordant if both nodes have distinct centrality values and the order of u and v in $c(G)$ differs from the order of u and v in c(G'), that is, either $c_G(u) < c_G(v)$ and $c_{G'}(u) > c_{G'}(v)$ or $c_G(u) > c_G(v)$ and $c_{G'}(u) < c_{G'}(v)$. Ties are neither concordant nor discordant.

A random graph consists of a finite set of graphs Ω equipped with a function P that assigns a probability to every graph in this set (Bollobás & Riordan, 2002).

Network data can be influenced by a variety of different measurement errors. Wang et al. (2012) categorized measurement errors into six groups: false-negative nodes and edges, false-positive nodes and edges, and false aggregation and disaggregation. For example, when 10% of the edges are missing in the measured network data, the graph constructed from this observed data suffers from false-negative edges.

To describe measurement errors, we introduce the notion of an error mechanism. An error mechanism φ is a procedure that describes measurement errors thay may occur during the data collection (e.g., 50% of the edges are missing, at random). For a given graph G, the error mechanism $\varphi(G)$ is defined as a random graph. The outcomes of $\varphi(G)$ are the graphs that result from G by applying the given error procedure, and each of these graphs is equipped with the probability of occurrence. To illustrate this concept, consider the graphs shown in Figure 1. The initial graph is denoted by H (drawn in the upper-left corner). We assume that we know the error mechanism that compromises the data collection. For this example, we assume that the error mechanism φ is edges missing uniformly at random with an error level of 50%. All graphs in the set of possible outcomes for this random graph $\Omega = \{G_1, G_2, \ldots, G_6\}$ are also shown in Figure 1. In this example, the probability function is $P(G_i) = \frac{1}{6}$; all graphs in Ω occur with the same probability. However, this concept is not limited to a uniform distribution.

In general, error mechanisms can rely on node or edge attributes. In this study, we focus on four common error mechanisms that do not depend on external attributes:

1. Nodes missing uniformly at random (rm nodes): A fraction of nodes (and all edges connected to these nodes) is missing in the measured network. All nodes have the same probability to be missing in the measured network.
2. Edges missing uniformly at random (rm edges unif.): A fraction of edges is missing in the measured network. All edges have the same probability to be missing in the measured network.
3. Edges missing proportionally (rm edges prop.): A fraction of edges is missing in the measured network. The probability that an edge is missing in the measured network is proportional to the sum of the degree values of the endpoints.

4. Spurious edges (add edges): The measured network contains too many edges. Every non-existing edge has the same probability to be erroneously observed.

Let G and G' denote graphs on the same vertex set and c a centrality measure. To measure the robustness of c w.r.t. these two graphs, we use Kendall's tau ("tau-b") rank correlation coefficient (Kendall, 1945). Correlations are commonly used to measure the robustness of centrality measures. Like existing studies, we also used rank correlations to minimize the influence of outliers (Kim & Jeong, 2007; Lee & Pfeffer, 2015; Wang et al., 2012).

We calculate the robustness ρ for a centrality measure c with respect to G and G' as follows:

$$\rho_c(G, G') = \frac{n_c - n_d}{\sqrt{(n_c + n_d + n_t) * (n_c + n_d + n_{t'})}} \tag{1}$$

with n_c as the number of concordant pairs and n_d as the number of discordant pairs w.r.t. the order given by $c(G)$ and $c(G')$. Ties only in $c(G)$ are denoted by n_t, ties only in $c(G')$ by $n_{t'}$.[1] It follows straightforwardly that the values of $\rho_c(G, G')$ are in the interval $[-1, 1]$.

Let us apply this concept to a graph illustrated in Figure 1. Assume that we have measured the graph labeled as G_6 and that we are interested in the robustness of the degree centrality. Then, the degree centrality values are $deg(H) = (1, 2, 3, 1, 1)$ and $deg(G_6) = (1, 2, 1, 0, 0)$. Based on the degree values, we can calculate the robustness of the degree centrality with respect to G_6 and H: $\rho_{deg}(G_6, H) = 0.53$.

If there are no ties in $c(G)$ and $c(G')$, then $\rho_c(G, G') = \frac{n_c}{n_c + n_d}$ which is Goodman and Kruskal's rank correlation coefficient γ (Goodman & Kruskal, 1954). In this special case, $\frac{\rho_c(G,G')+1}{2}$ is the probability that two nodes with distinct centrality values, randomly chosen from the common vertex set of G and G', have the same order in $c(G)$ and $c(G')$, that is, they are concordant.

3. How to estimate the robustness of centrality measures

In network studies, the measured network data often contain sampling errors (Leecaster et al., 2016; Schulz, 2016; Wang et al., 2016). But in general, the authors of such studies have no tools to describe the impact of sampling errors on the network measures (e.g., centrality measures) they apply. In general, the assumptions made about sampling errors are mentioned in the limitations, but they are not considered as part of the network model.

The robustness concept as introduced in Section 2 helps researchers to describe this impact: given the measured network M, the (unknown) hidden network H, and a centrality measure c, the robustness $\rho_c(H, M)$ measures the impact of the sampling error on the centrality values of the nodes in the measured network. Thus, the robustness can be used to measure the reliability of a centrality measure with respect to sampling errors. We call $\rho_c(H, M)$ the "true robustness".

As the hidden network H is not known, the true robustness cannot be computed explicitly. In this section, we propose a method for the estimation of the true robustness based on the measured network M. Moreover, we provide an example for the application and demonstrate how the estimation results can be evaluated.

Our estimation approach is based on the observation that, given a graph G, a centrality measure, and some error procedure, in many experiments the robustness is nearly proportional to the error intensity. That is, removing 20% of the edges, randomly, has about twice as much impact on the centrality measure as removing 10% of the edges (Borgatti et al., 2006; Frantz et al., 2009; Wang et al., 2012). Now let G' denote the graph resulting from G by removing 10% of the edges, and let G'' denote the graph resulting from G' by removing 10% of the remaining edges. One possible explanation for the observed linearity could be that the robustness with respect to G and G' is close to the robustness with respect to G' and G'', which is $\rho_c(G, G') \sim \rho_c(G', G'')$.

If we apply this idea to our definition of the true robustness and take into account our notion of error mechanism (as random graphs), we yield the following estimation:

$$\hat{\rho}_c(M, H) := E(\rho_c(M, \varphi(M))). \tag{2}$$

Since $\varphi(M)$ is a random graph, $\rho_c(M, \varphi(M))$ is a random variable; hence we use the expected value of this expression as the estimate for the robustness. In practice, this value is computed by sampling.

4. Experiments

In this chapter, the efficiency of the estimation method is analyzed under different conditions using two types of simulation experiments. First, experiments based on random graphs are conducted, followed by experiments based on real-world networks. To control for the true robustness, in all experiments we start with a given *hidden* network H and construct the *measured* network M by applying the error mechanism to the hidden network. As part of these experiments, we calculate the following values for each run:

True robustness: For every single experiment we compute the true robustness $\rho_c(M, H)$.
Estimated robustness: For every single experiment we compute $\hat{\rho}_c(M, H)$, as introduced in Section 3.
Mean and standard deviation (SD) of true robustness: For every series of experiments (i.e., fixed centrality measure, fixed error mechanism, fixed random type or initial real-work network), we report the mean value and the SD of all corresponding true robustness values. Note that large values indicate that the true robustness very much depends on the specific choice of removed/added vertices/edges. In such cases, estimating the true robustness given just the measured network and the error mechanism is hardly possible. In this sense, a large SD of true robustness is a good indicator for ill-conditioned estimation problems. Note that these values cannot be computed without knowing the hidden network.
Mean SD of estimation: By definition, the estimated robustness is an expected value. Here we compute the mean of all corresponding SDs. Note that this value by construction is very closely related to the SD of true robustness, but it can be computed without knowing the hidden network.
Mean estimation error: For every series of experiments (i.e., fixed centrality measure, fixed error mechanism, fixed random type, or initial real-work network), we compute the mean absolute difference between true robustness and estimated robustness.
Implicit error: As a kind of benchmark for the estimation error, we compute the robustness error that would occur if we ignored the impact of the measurement error on the centrality measure. In this case we would consider the correlation between the centralities of hidden and measured network to be 1. That is, for every series of experiments we define the implicit error as (1—mean true robustness).

4.1 Experiments based on random graphs

As a first step to validate whether the proposed methods yield useful results, we apply the four error mechanisms (node missing uniformly, edges missing uniformly, edges missing proportional, and spurious edges) to Erdös–Rényi graphs (ER graph) (Erdös & Rényi, 1959) and Barabasi-Albert graphs (BA graph) (Barabási & Albert, 1999) and estimate the corresponding robustness. For every error mechanism, we consider two cases: a moderate scenario of 10% error level and a more intense scenario with 30% error level. For all combinations of centrality measures and error mechanisms, we perform the experiment described below 1,000 times:

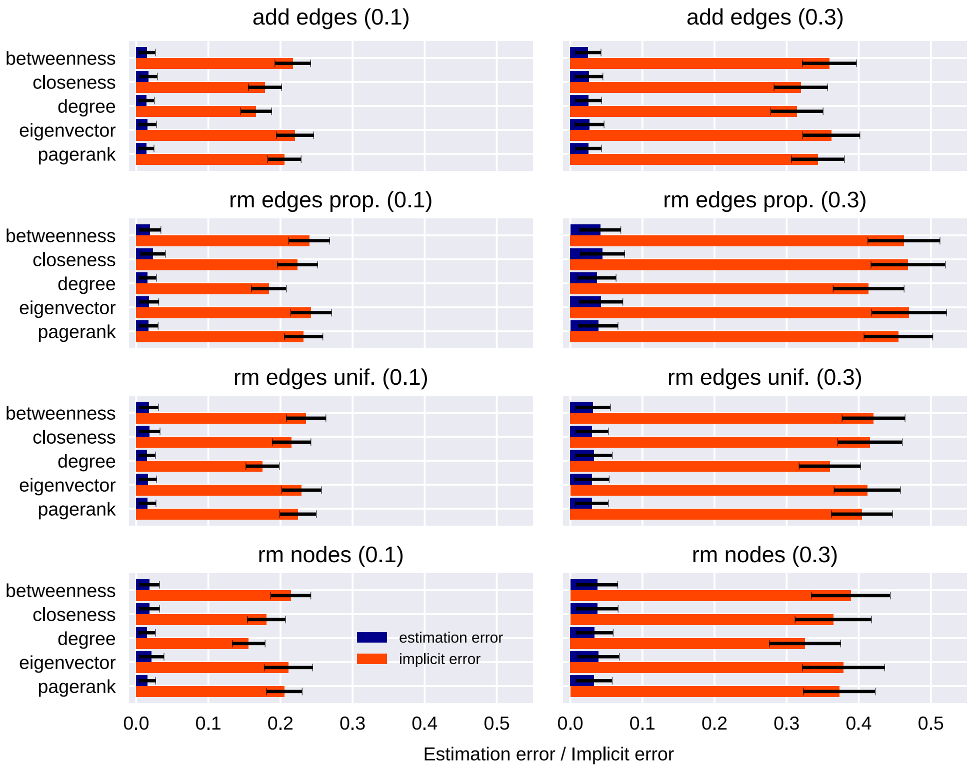

Figure 2. (Color online) The results for **ER graphs**. The blue bar indicates the mean absolute error of the estimation ($|\rho - \hat{\rho}|$) for 1,000 simulation runs. We call this error the estimation error. The red bar indicates the error that is caused by the flawed data collection. This error would be accepted if the influence of the measurement error was to be ignored when analyzing the network ($1 - \rho$). We call this value the implicit error. The length of the bars indicates the mean values and the error bars the SDs. In the case of ER graphs the behavior is homogeneous: the true robustness depends primarily on the error intensity. The estimation errors are consistently low.

(1) We generate a random graph and denote it by H. This graph represents the (error-free) hidden network. We use two types of random graphs:
   a. an ER graph with 100 nodes and edge probability 0.2 and[2]
   b. a BA graph with 100 nodes (parameter $m = 11$, undirected).
(2) We choose a graph from $\varphi(H)$ and denote it by M. This graph represents the measured network that is affected by measurement errors. For evaluation purposes, the true robustness $\rho_c(H, M)$ is calculated and denoted by ρ.
(3) Based on the measured network M, we estimate the true robustness $\hat{\rho}_c(M, H)$.

The results for the random graphs are shown in Figures 2 and 3. Every panel shows the results for the five centrality measures under the influence of one of the four error mechanisms with either 10% or 30% intensity.

The blue bar indicates the mean estimation error of the 1,000 simulation runs. The red bar indicates the implicit error as defined at the beginning of this section. Note again that the implicit error is closely related to the mean true robustness since it is defined as (1—implicit error). That is, a long red bar indicates a weak correlation between the centrality vector in measured and hidden networks and vice versa. The length of the bars indicates the mean values and the error bars the corresponding SDs.

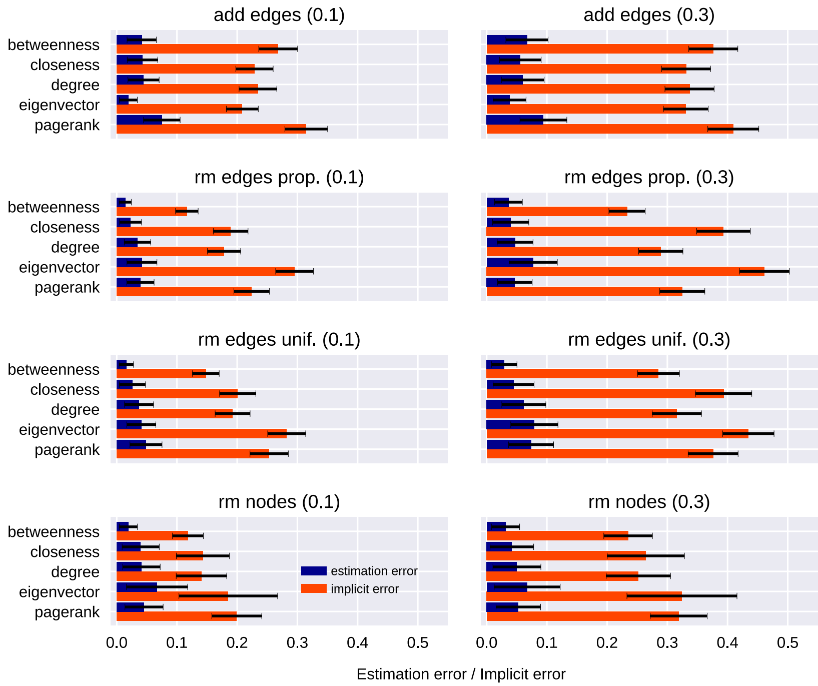

Figure 3. (Color online) The results for **BA graphs**. The centrality measures have varying reactions to the different error mechanisms. The eigenvector centrality variation is noticeably high when nodes are removed.

Now let us first focus on the impact of different error mechanisms on the true robustness (1—red bar). For example, the true robustness of the betweenness in an ER graph under the influence of 10% spurious edges (1st panel in Figure 2) is 0.78 with an SD of 0.03.

For ER graphs (Figure 2), within the two error levels, there are only small differences regarding the influence of the error mechanisms on the centrality measures. The degree is the most robust measure in this setting.

With an error level of 30%, the robustness is always lower than in the respective cases with 10%. The SD is also higher. In contrast to the cases with 10%, at 30% the absence of edges depending on the edge degree leads to lower robustness when compared to the other error mechanisms. These findings are conclusive with (Borgatti et al., 2006; Frantz et al., 2009). The homogeneity of the results regarding the different centrality measures is not suprising given the high correlation between the centrality measures in the case of ER random graphs (Valente et al., 2008).

For BA graphs, we make similar observations. Higher error levels lead to lower robustness. In contrast to ER graphs, however, degree centrality is not always the most robust here. We notice that the SD is not as homogeneous as in the ER experiments (e.g., eigenvector centrality).

Regarding estimation errors, the pattern is the same for both graph types. The estimation error is always small compared to the implicit error. With fixed intensity the difference between the error types and the centrality measures is small. The estimates at 10% error level have a smaller error than at 30%. The SD is homogeneous. The estimates are most accurate for spurious edges, worst for missing nodes and missing edges proportional to the edge degree.

For BA graphs (Figure 3) the results are more heterogeneous. Although the true robustness is at about the same level as for ER graphs, the estimation errors vary strongly depending on the choice of centrality measure and error mechanism. If nodes are missing, the estimates are usually poorest and the variance highest.

Table 1. Statistics of real-world networks.

Network	Nodes	Edges	Clustering	Density	Diameter	Source
Dolphin	62	159	0.3029	0.0841	8	Lusseau et al. (2003)
Jazz	198	2,742	0.6334	0.1406	6	Gleiser & Danon (2003)
Protein	1,458	1,948	0.1403	0.0018	19	Jeong et al. (2001)
Hamsterster	1,788	12,476	0.1655	0.0078	14	Kunegis (2013)

Notes: If the original network is not connected, we consider only the largest connected component.

4.2 Experiments based on real-world networks

Next, we apply our methods from Section 3 to real-world networks to investigate the suitability of these methods for practical application. We choose four networks from different domains and thus different structural properties to get an impression of how these methods perform on real data (see Table 1 for descriptive statistics). As before we use our proposed methods to estimate the robustness of five centrality measures under the influence of four error mechanisms and two error intensities. For every combination of network, centrality measure, and error mechanism, the experimental setup is as follows:

(1) Due to the very nature of the hidden networks, we cannot access them. Hence, for the sake of our experiments, we treat the real-world network as the error-free hidden network H. (This is a common approach used in existing studies about the robustness of centrality measures (Wang et al., 2012)).
(2) To simulate erroneous data collection, we choose a graph from $\varphi(H)$ and denote it by M. This graph represents the measured network that is affected by measurement errors. For evaluation purposes, the true robustness $\rho_c(H, M)$ is calculated and denoted by ρ.
(3) Based on the measured network M, we estimate the true robustness $\hat{\rho}_c(M, H)$.

For every combination, we perform this experiment 1,000 times.

The results for real-world networks are shown in Figures 4–8. In summary, the results are promising. The estimation error is always below, in most cases far below, the implicit error. With a lower error level (10%), the error of estimation is often very low (mean estimate error values below 0.03).

The robustness of the centrality values is usually strongly dependent on the respective network structure and the type of error, where a higher error level always reduces the robustness. Regardless of the error type, degree centrality and pagerank are most robust, while eigenvector centrality is most sensitive to measurement errors.

Degree centrality (Figure 6) is robust in the case of all studied networks, most values are in regions below 0.1. However, in the Protein and Dolphin networks, the values are considerably lower (visible in the figure by the higher implicit error) than in the other two networks, in all error types and intensities. The estimation errors are also low (mean values at 0.1 error level mostly below 0.02). The results for pagerank (Figure 8) are similar; it is usually robust but somewhat more sensitive than the degree centrality. The difference between the implicit error and the estimation error is equally large. Also, for betweenness (Figure 4) and closeness (Figure 5), the implicit error is greater than the estimation error, but the difference between them is smaller compared to degree and pagerank. In addition, the SD is higher for both the estimation error and the robustness, especially for the error types and missing nodes and additional edges. The latter has a particularly strong influence on the betweenness. This might be due to the fact that by adding edges (randomly), many new abbreviations are created between the nodes and thus the high diameters (especially in the Hamsterster and Protein networks) are reduced.

The overall impression of the results is most heterogeneous for the eigenvector centrality (Figure 7). The difference between the Jazz and Hamsterster networks, on one, and the Protein

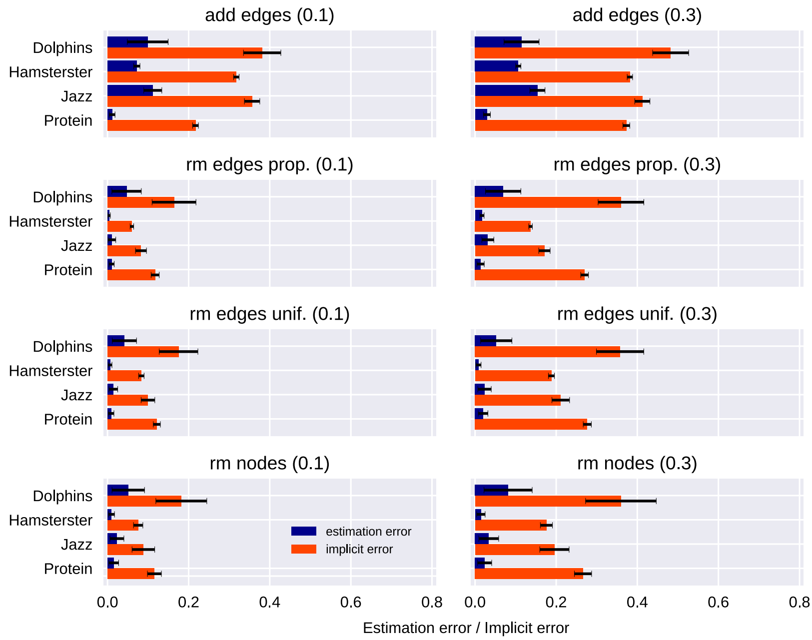

Figure 4. (Color online) The results for the **betweenness centrality** in real-world networks. The true robustness depends on the network, type, and intensity of the error. The estimates are good in most cases, but the error-type spurious edges lead to increased estimation errors.

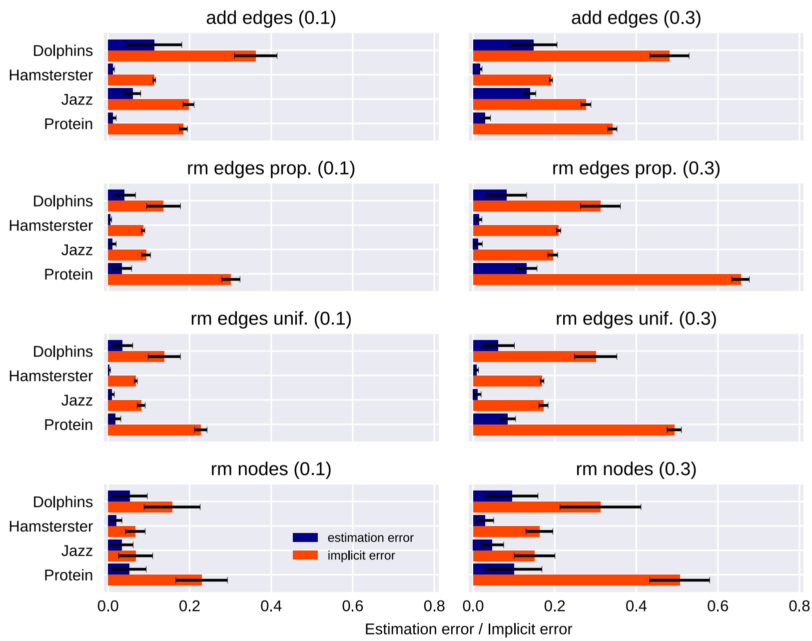

Figure 5. (Color online) The results for the **closeness centrality** in real-world networks. The true robustness depends on the network, type, and intensity of the error. The estimates are good in most cases, especially for low error intensities. For higher intensities the errors are higher, especially in case of the Dolphin and Protein networks.

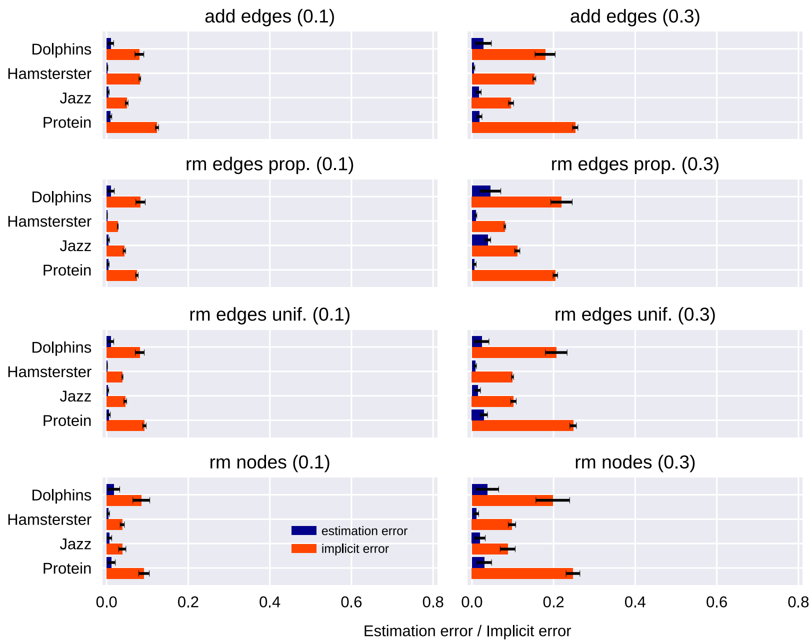

Figure 6. (Color online) The results for the **degree centrality** in real-world networks. As expected, the robustness in this case is the highest and the estimation error the smallest.

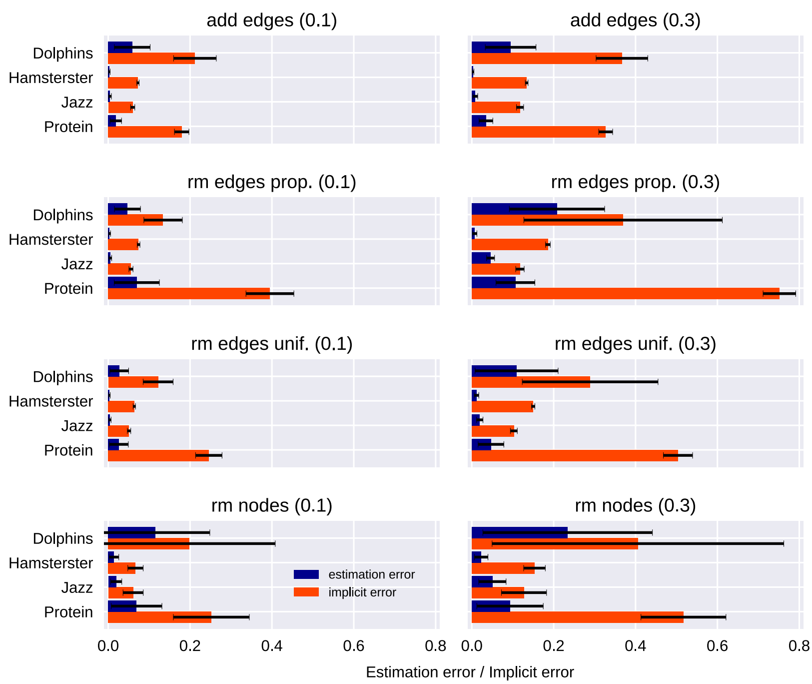

Figure 7. (Color online) The results for the **eigenvector centrality** in real-world networks. The results can be divided into two groups. In the Jazz and Hamsterster networks the robustness is high and the estimation error low. In the Protein and Dolphin networks both values are considerably worse, and the fluctuation of both values is higher. This effect is discussed in detail in Section 4.

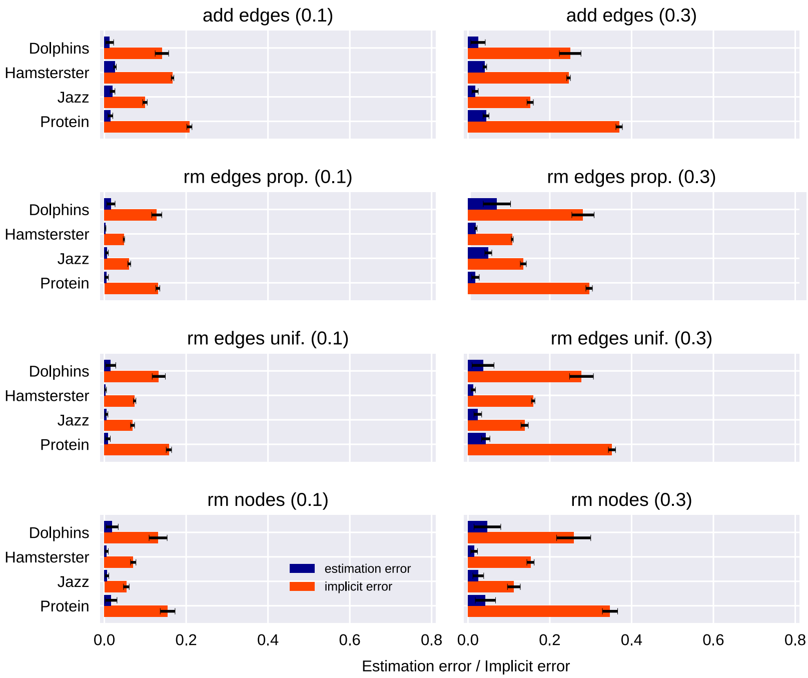

Figure 8. (Color online) The results for the **pagerank** in real-world networks. The true robustness depends on the network, type, and intensity of the error. Although robustness is reduced by increasing the error level, estimation errors are relatively low for both intensities.

Figure 9. The results for the additional experiment regarding the Dolphin network (robustness of eigenvector centrality when 30% of the nodes are removed randomly). These results demonstrate that the eigenvector centrality of the removed nodes has a strong influence on the robustness. Especially if the removed nodes have a large share of the total eigenvector centrality (>30%), the order of the nodes (based on the eigenvector centrality) has little to do with the order of the nodes in the original network.

and Dolphin networks, on the other hand, is most noticeable in this case. While implicit error for the first mentioned networks is comparable to the other centrality measures (except the error mechanism missing nodes), for the other two networks the robustness is low and the SD high (very high for the Protein network). This effect is particularly strong when nodes are missing or edges are missing proportionally to the edge degree. The estimation error varies particularly strongly, especially in cases where the robustness has a high SD.

In our experiments with real-world networks, we have observed that in some cases concerning eigenvector centrality very high implicit errors as well as estimation errors occur. Therefore we want to take a more detailed look at one of these cases. For this purpose we have again performed

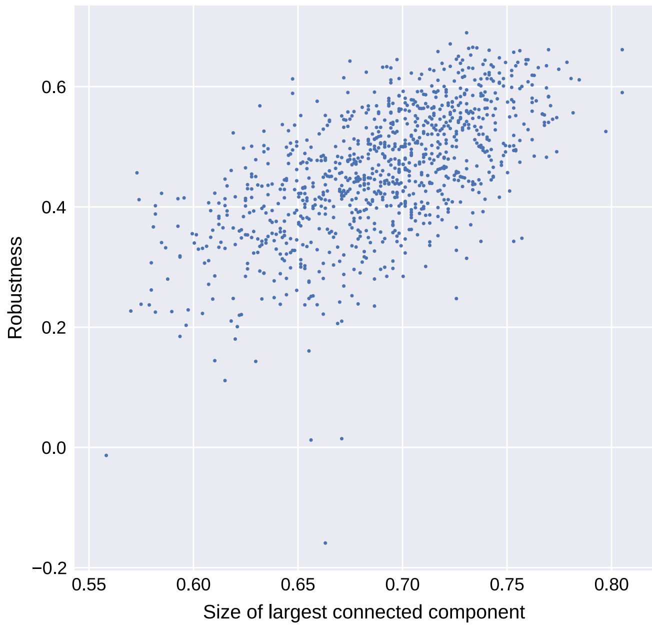

Figure 10. The results for the additional experiment regarding the Protein network (robustness of eigenvector centrality when 30% of the nodes are removed randomly). We observe that the lower the size of the largest connected component in the modified network, the lower the robustness (recall, the number of nodes removed is constant in this experiment).

an experiment (Dolphin network, 30% nodes missing randomly, eigenvector centrality). We track the robustness and the percentage of the total eigenvector centrality that is associated with the removed nodes. The results of this experiment are shown in Figure 9. These results demonstrate that the eigenvector centrality of the removed nodes has a strong influence on the robustness. Especially if the removed nodes have a large share of the total eigenvector centrality (>30%), the order of the nodes (based on the eigenvector centrality) has little to do with the order of the nodes in the original network.

There are two main reasons for this observation. First, the eigenvector values in the Dolphin network are unequally distributed among the nodes. Of the 62 nodes, 12 (approx. 19%) hold more than 50% of the eigenvector centrality (we call these nodes "high evc nodes").

Furthermore, removing a node with high eigenvector centrality has a high impact on all adjacent nodes in contrast, for example, to the degree centrality and pagerank, where the removal of a hub has a small effect on all adjacent nodes.[3]

The Dolphin network also exhibits another effect. With 62 nodes this network is relatively small. The number of "high evc nodes" removed by the error mechanism is binomial distributed ($n = 12, p = 0.3$). Due to the low number of trials in this distribution, the number of possible values of the random variable has a relatively broad range. For example, the events that only one "high evc node" is removed and six "high evc nodes" are removed have approximately the same probability of about 7%. However, the effects of these two events on the robustness differ drastically. In this case, the robustness of the centrality measure depends primarily on the outcome of the random experiment.

In the Protein network, the eigenvector centrality is also unevenly distributed. Here 41 of 1,458 nodes (2.8%) hold 50% of the eigenvector centrality. Furthermore, this network is relatively sparse (1,948 edges). Therefore, we will carry out the aforementioned experiment again with this network. This time we track the size of the largest connected component of the modified network in addition to the robustness of the eigenvector centrality. The results are shown in Figure 10. From this, we can see that the lower the size of the largest connected component, the lower the robustness (recall, the number of nodes removed is constant in this experiment).

Table 2. The mean SD of estimation (*e*) and the SD of true robustness (*t*) for the experiments with real-world networks.

			Betw		Clos		Deg		Evc		Page	
Network	Error	Level	*t*	*e*	*t*	*e*	*t*	*e*	*t*	*e*	*t*	*e*
Dolphin	add edges	10%	4.61	3.66	5.24	3.71	1.10	1.18	5.21	3.62	1.65	1.61
		30%	4.45	3.90	4.79	3.76	2.47	2.69	6.32	4.21	2.65	2.89
	rm edges prop.	10%	5.41	4.73	4.15	4.32	1.16	1.18	4.71	10.24	1.24	1.60
		30%	5.64	6.66	4.91	7.17	2.63	3.31	24.25	24.94	2.73	3.47
	rm edges unif.	10%	4.77	4.76	3.95	4.27	1.11	1.14	3.66	5.47	1.61	1.71
		30%	5.84	6.18	5.20	6.43	2.66	2.73	16.59	16.37	2.92	3.07
	rm nodes	10%	6.28	7.23	6.88	7.81	2.07	2.15	20.94	21.56	2.21	2.51
		30%	8.71	10.39	9.93	14.07	4.14	5.42	35.64	30.70	4.17	5.70
Hamst	add edges	10%	0.69	0.57	0.33	0.28	0.22	0.22	0.28	0.27	0.32	0.29
		30%	0.67	0.55	0.40	0.37	0.37	0.38	0.34	0.33	0.45	0.42
	rm edges prop.	10%	0.44	0.43	0.42	0.38	0.11	0.11	0.33	0.34	0.17	0.18
		30%	0.44	0.62	0.53	0.47	0.20	0.23	0.58	0.58	0.24	0.29
	rm edges unif.	10%	0.67	0.64	0.36	0.36	0.14	0.14	0.30	0.31	0.31	0.33
		30%	0.75	0.69	0.46	0.47	0.25	0.28	0.41	0.46	0.42	0.50
	rm nodes	10%	1.15	1.18	2.42	2.47	0.52	0.54	1.88	1.94	0.66	0.69
		30%	1.49	1.78	3.30	3.77	0.89	1.10	2.65	3.30	0.89	1.17
Jazz	add edges	10%	1.90	1.43	1.36	0.94	0.35	0.38	0.52	0.50	0.55	0.47
		30%	1.87	1.48	1.23	0.82	0.63	0.74	0.87	0.86	0.77	0.79
	rm edges prop.	10%	1.36	1.10	1.07	1.02	0.31	0.36	0.50	0.54	0.36	0.43
		30%	1.41	1.85	1.22	1.30	0.65	1.04	1.04	2.56	0.73	1.15
	rm edges unif.	10%	1.69	1.73	0.97	0.97	0.35	0.35	0.42	0.44	0.49	0.56
		30%	2.18	2.31	1.13	1.14	0.69	0.79	0.85	0.99	0.89	1.11
	rm nodes	10%	2.80	2.88	4.18	4.25	0.90	0.95	2.47	2.76	0.70	0.81
		30%	3.62	4.32	4.99	5.77	1.85	2.31	5.51	8.02	1.57	2.26
Protein	add edges	10%	0.72	0.66	0.98	0.82	0.40	0.37	1.77	1.48	0.63	0.53
		30%	0.85	0.79	1.13	0.86	0.67	0.68	1.70	1.05	0.79	0.69
	rm edges prop.	10%	1.00	0.85	2.23	1.67	0.33	0.32	5.85	5.74	0.49	0.52
		30%	0.96	1.10	2.13	1.94	0.57	0.57	3.99	10.87	0.80	0.69
	rm edges unif.	10%	0.88	0.90	1.54	1.49	0.42	0.39	3.23	3.53	0.66	0.63
		30%	1.00	1.09	1.76	1.45	0.80	0.64	3.57	10.24	0.88	0.87
	rm nodes	10%	1.72	1.78	6.33	6.19	1.31	1.33	9.28	9.42	1.81	1.80
		30%	2.12	2.60	7.35	6.65	1.70	2.01	10.39	11.70	1.86	1.93

The values have been multiplied by 100 for better readability.

4.3 Discussion of SDs

As discussed at the beginning of this section, the problem of guessing the true robustness based on some measured network and knowledge about the underlying error mechanism is ill-conditioned if the true robustness depends very much on the specific choice of added/removed vertices/edges—that is, the SD of true robustness is large. As illustrated in Tables 2 and 3 large values (SD true robustness > 0.15) are observed only for the combinations of eigenvector centrality and Dolphin networks. The same tables indicate that the mean SD of the estimated robustness is a very good indicator for a large SD of true robustness. In fact, the SD of the estimated robustness has large values (SD estimated robustness > 0.15) for exactly the same combinations. While the SD

Table 3. The mean SD of estimation (*e*) and the SD of true robustness (*t*) for the experiments with random graphs.

Network		Barabasi				Erdös-Rényi			
		10%		30%		10%		30%	
Centrality	Error	*t*	*e*	*t*	*e*	*t*	*e*	*t*	*e*
Betweenness	add edges	3.23	2.17	4.07	2.92	2.48	1.96	3.76	3.23
	rm edges prop.	1.87	1.63	3.01	2.76	2.84	2.05	5.00	4.02
	rm edges unif.	2.22	1.99	3.49	3.13	2.75	1.95	4.37	3.41
	rm nodes	2.56	2.59	4.04	4.86	2.79	2.36	5.47	5.56
Closeness	add edges	3.10	2.27	4.07	3.03	2.31	1.79	3.72	3.10
	rm edges prop.	2.88	2.45	4.46	3.59	2.80	2.29	5.16	4.44
	rm edges unif.	3.01	2.40	4.68	3.40	2.68	2.14	4.47	3.68
	rm nodes	4.42	4.57	6.42	8.60	2.65	2.47	5.30	5.70
Degree	add edges	3.14	2.28	4.08	3.01	2.15	1.75	3.63	3.12
	rm edges prop.	2.76	1.84	3.67	2.75	2.43	1.91	4.92	4.19
	rm edges unif.	2.90	1.99	4.09	2.92	2.31	1.76	4.26	3.40
	rm nodes	4.20	4.04	5.34	5.88	2.27	1.96	4.94	5.18
Eigenvector	add edges	2.65	1.85	3.72	2.80	2.59	1.91	3.97	3.21
	rm edges prop.	3.14	2.18	4.14	3.33	2.83	2.24	5.19	4.77
	rm edges unif.	3.16	2.08	4.27	3.12	2.75	2.05	4.61	3.88
	rm nodes	8.20	7.66	9.14	9.29	3.36	3.08	5.71	6.55
Pagerank	add edges	3.55	2.33	4.24	2.99	2.31	1.76	3.67	3.03
	rm edges prop.	2.95	1.89	3.74	2.61	2.67	1.91	4.77	3.90
	rm edges unif.	3.18	2.09	4.15	2.86	2.53	1.79	4.24	3.25
	rm nodes	4.15	4.15	4.73	5.50	2.47	2.03	4.98	5.10

The values have been multiplied by 100 for better readability.

of true robustness is not accessible without knowledge of the hidden network (and hence cannot serve as a heuristic to decide whether to apply our method), the SD of the estimated robustness can be computed, given the measured network and the error mechanism. Therefore, we recommend using the suggested estimation for true robustness only if the SD of estimated robustness is small.

5. Summary and recommendations

Errors in network data are a ubiquitous problem in network analysis. Even though the reliability of centrality measures has been studied extensively in the literature, there is no method available that allows researchers to estimate the reliability of centrality measures in the case of imperfect measured data.

In the first part of this study, we proposed such a method for estimating the impact of measurement errors on the reliability of a centrality measure, given the measured network and assumptions about the type and intensity of the measurement error. To check the applicability of this method we have conducted a series of simulation experiments based on random graphs and real-world networks as well.

Regarding the robustness of random graphs and real-world networks, the results are conclusive with existing studies (Borgatti et al., 2006; Frantz et al., 2009). Moreover, we have observed that the (measurable) SD of the estimated robustness is a good indicator for the (not measurable) SD of the true robustness. Our results provide compelling evidence that our proposed estimation method is a suitable technique for the estimation of the robustness of centrality measures.

Based on these findings, we would like to offer the following recommendations for network studies, where centrality measures are analyzed and hypotheses about underlying measurement errors are available:

- Researchers should compute the estimated robustness and the corresponding SD for all relevant centrality measures (Python code[4] can be downloaded and easily extended for specific centrality measures and error mechanisms).
- For those centrality measures where the computed SD is large (>0.15), the concept of true robustness is not appropriate as it is expected to depend very much on the specific added/removed edges/vertices. In all other cases, we recommend to report the estimated robustness in order to contribute to a better assessability of the conclusions of the study, which are based on centrality measures.
- If there are different centrality measures to choose from that are equally suitable for the study, the estimated robustness can be used as a further selection criterion.
- The estimated robustness can also be used as a basis for deciding whether to apply treatment procedures, such as imputation (Huisman, 2009; Wang et al., 2016; Krause et al., 2018).

Future studies should analyze the stability of the presented estimation approach with respect to the assumptions on the underlying error mechanism. If 20% missing edges (randomly) are assumed, while actually 15% of the vertices are missing proportional to vertex degree, what impact does this imprecise error-assumption have on the estimated true robustness? To extend the approach to directed networks, suitable error mechanisms have to be analyzed. While the current study focuses on robustness with respect to certain node-level metrics, it might be interesting to see corresponding results for other metrics, such as the estimation of the most central node (Frantz & Carley, 2017).

Conflict of interest. The authors have nothing to disclose.

Notes

1 It may occur that $V(G') \neq V(G)$. In these cases, we only consider entries in $c(G)$ and $c(G')$ that correspond to nodes that are in both graphs (G and G'). This is a common approach for the comparison of graphs on different vertex sets (Wang et al., 2012).

2 Our experiments have shown that the choice of p has little influence on the main results associated with this section. Hence we will only consider the case of $p = 0.2$.

3 Removing a high evc node reduces the unnormalized evc-value of all neighbors by the high evc-value of the removed node. In the case of the degree centrality, the centrality of all neighbors is reduced by 1. Similarly, in the case of pagerank, the unnormalized centrality of each neighbor is reduced by the pagerank of the hub divided by its (high) degree.

4 `https://github.com/crsqq/EstimatingCentralityRobustness`

References

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(October), 509–512.

Bollobás, B., & Riordan, O. (2002). Mathematical results on scale-free random graphs. *Handbook of graphs and networks: From the genome to the internet*, 1–38.

Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, *92*(5), 1170–1182.

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

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998).*

Butts, C. T. (2003). Network inference, error, and informant (in)accuracy: A Bayesian approach. *Social Networks*, *25*(2), 103–140.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks*, *25*(4), 283–307.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Interjournal, Complex Systems*, *1695*, 1–9.

Erdös, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, *6*, 290–297.
Fischer, M., Parkins, K., Maizels, K., Sutherland, D. R., Allan, B. M., Coulson, G., & Di Stefano, J. (2018). Biotelemetry marches on: A cost-effective GPS device for monitoring terrestrial wildlife. *Plos One*, *13*(7), e0199617.
Frantz, T. L., & Carley, K. M. (2017). Reporting a network's most-central actor with a confidence level. *Computational and Mathematical Organization Theory*, *23*(2), 301–312.
Frantz, T. L., Cataldo, M., & Carley, K. M. (2009). Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory*, *15*(4), 303–328.
Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.
Gleiser, P. M., & Danon, L. (2003). Community structure in jazz. *Advances in Complex Systems*, *6*(4), 565–573.
Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764.
Handcock, M. S., & Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics*, *4*(1), 5–25.
Huisman, M. (2009). Imputation of missing network data: Some simple procedures. *Journal of Social Structure*, **10**(1), 1–29.
Jeong, H., Mason, S. P., Barabasi, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42.
Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, **33**(3), 239–251.
Kim, M., & Leskovec, J. (2011). The network completion problem: Inferring missing nodes and edges in networks. In *Siam international conference on data mining*, 47–58.
Kim, P. J., & Jeong, H. (2007). Reliability of rank order in sampled networks. *European Physical Journal b*, *55*(1), 109–114.
Koschützki, D., Lehmann, K., & Peeters, L. (2005). Centrality Indices. In U. Brandes, & T. Erlebach (eds.), *Network Analysis: Methodological Foundations*, 16–61. Springer Berlin Heidelberg.
Krause, R. W., Huisman, M., Steglich, C., & Sniiders, T. A. (2018, August). Missing Network Data A Comparison of Different Imputation Methods. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 159–163.
Kunegis, J. (2013). KONECT - The Koblenz network collection. In *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*.
Lee, J.-S., & Pfeffer, J. (2015). Robustness of network centrality metrics in the context of digital communication data. In *Proceedings of the 48th Hawaii international conference on system sciences*.
Leecaster, M., Toth, D. J. A., Pettey, W. B. P., Rainey, J. J., Gao, H., Uzicanin, A., & Samore, M. (2016). Estimates of social contact in a middle school based on self-report and wireless sensor data. *Plos One*, *11*(4).
Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology*, *54*(4), 396–405.
Newman, M. E. J. (2018). Network structure from rich but noisy data. *Nature Physics*, *14*, 542–545.
Niu, Q., Zeng, A., Fan, Y., & Di, Z. (2015). Robustness of centrality measures against network manipulation. *Physica A: Statistical Mechanics and Its Applications*, *438*, 124–131.
Platig, J., Ott, E., & Girvan, M. (2013). Robustness of network measures to link errors. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *88*(6).
Schulz, J. (2016). Using Monte Carlo simulations to assess the impact of author name disambiguation quality on different bibliometric analyses. *Scientometrics*, *107*(3), 1283–1298.
Silk, M. J., Jackson, A. L., Croft, D. P., Colhoun, K., & Bearhop, S. (2015). The consequences of unidentifiable individuals for the analysis of an animal social network. *Animal Behaviour*, *104*, 1–11.
Smith, J. A., & Moody, J. (2013). Structural Effects of Network Sampling Coverage I: Nodes Missing at Random. *Social Networks*, *35*(4).
Smith, J. A., Moody, J., & Morgan, J. H. (2017). Network sampling coverage II: The effect of non-random missing data on network measurement. *Social Networks*, *48*, 78–99.
Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections*, *28*(1), 16–26.
Wang, C., Butts, C. T., Hipp, J. R., Jose, R., & Lakon, C. M. (2016). Multiple imputation for missing edge data: A predictive evaluation method with application to Add Health. *Social Networks*, *45*, 89–98.
Wang, D. J., Shi, X., McFarland, D. A., & Leskovec, J. (2012). Measurement error in network data: A re-classification. *Social Networks*, *34*(4), 396–409.
Žnidaršič, A., Ferligoj, A., & Doreian, P. (2018). Stability of centrality measures in valued networks regarding different actor non-response treatments and macro-network structures. *Network Science*, *6*(01), 1–33.

Cite this article: Martin C., and Niemeyer P. (2019). Influence of measurement errors on networks: Estimating the robustness of centrality measures. *Network Science* **7**, 180–195. https://doi.org/10.1017/nws.2019.12