Sensitivity Analysis of Centralities on Unweighted Networks

Shogo Murai*
The University of Tokyo
Bunkyo-ku, Tokyo, Japan
s.murai@is.s.u-tokyo.ac.jp

Yuichi Yoshida†
National Institute of Informatics
Chiyoda-ku, Tokyo, Japan
yyoshida@nii.ac.jp

ABSTRACT

Revealing important vertices is a fundamental task in network analysis. As such, many indicators have been proposed for doing so, which are collectively called centralities. However, the abundance of studies on centralities blurs their differences.

In this work, we compare centralities based on their sensitivity to modifications in the graph. Specifically, we introduce a quantitative measure called (average-case) edge sensitivity, which measures how much the centrality value of a uniformly chosen vertex (or an edge) changes when we remove a uniformly chosen edge. Edge sensitivity is applicable to unweighted graphs, regarding which, to our knowledge, there has been no theoretical analysis of the centralities. We conducted a theoretical analysis of the edge sensitivities of six major centralities: the closeness centrality, harmonic centrality, betweenness centrality, endpoint betweenness centrality, PageRank, and spanning tree centrality. Our experimental results on synthetic and real graphs confirm the tendency predicted by the theoretical analysis. We also discuss an extension of edge sensitivity to the setting that we remove a uniformly chosen set of edges of size k for an integer $k \geq 1$.

CCS CONCEPTS

• **Information systems** → **Web mining**; • **Theory of computation** → *Graph algorithms analysis.*

KEYWORDS

Centrality, sensitivity, edge deletion

ACM Reference Format:
Shogo Murai and Yuichi Yoshida. 2019. Sensitivity Analysis of Centralities on Unweighted Networks. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3308558.3313422

1 INTRODUCTION

One of the most important problems in network analysis is revealing important vertices in a given network. To solve this problem, several measures of the importance of a vertex have been proposed, and these are collectively called *(network) centralities*. For example, the closeness centrality [4, 5, 27] and harmonic centrality [21] measure the importance of a vertex based on its distance to other vertices. The betweenness centrality [14] considers the number of shortest paths that pass through the vertex. PageRank [25] regards a vertex as important if a random walk on the network has a high staying probability at the vertex.

Table 1: Average-case edge (k-)sensitivity of centralities. Here, n is the number of vertices, m is the number of edges, D is the diameter of the graph, α is the decaying factor for PageRank, b is the number of bridges, and χ_k is the probability that the graph becomes disconnected by removing a uniformly chosen set of k edges.

Centrality	Upper bound ($k = 1$)	Lower bound ($k = 1$)	Upper Bound (general)
Closeness	$O\left(\frac{b+D}{m}\right)$	$\Omega\left(\frac{b+D}{m}\right)$	$O\left(\frac{k^2D}{m} + \chi_k\right)$
Harmonic	$O\left(\frac{D}{m}\right)$	$\Omega\left(\frac{D}{m}\right)$	$O\left(\frac{kD}{m}\right)$
Betweenness	—	∞	—
Endpoint betweenness	$O\left(\frac{D^2}{m}\right)$	$\Omega\left(\frac{D^2}{m}\right)$	$O\left(\frac{k^2D^2}{m}\right)$
PageRank	$O\left(\frac{\alpha}{m(1-\alpha)^2}\right)$	$\Omega\left(\frac{\alpha}{m(1-\alpha)}\right)$	$O\left(\frac{k\alpha}{m(1-\alpha)^2}\right)$
Spanning tree	$O\left(\frac{\sqrt{n}}{m}\right)$	$\Omega\left(\frac{1}{m}\right)$	$O\left(k\sqrt{\frac{kn}{m(m-k)}}\right)$

However, the abundance of studies on centralities raises another question: *Which centrality should we use?* One consideration in response to this question is that centralities have been used to analyze real-world networks and we may miss some edges when constructing network datasets. This motivates us to compare centralities based on their sensitivity to changes in the network.

Although some previous works have analyzed the sensitivity of centralities, it remains difficult to distinguish between centralities because only empirical results have been presented with respect to particular networks [7, 10, 13, 24, 34]. There are only a few theoretical works on the sensitivity of centralities, and most of them considered the sensitivity of a specific centrality and do not compare different centralities in terms of sensitivity [15, 26, 31]. To the best of our knowledge, the only theoretical work comparing the sensitivity of several centralities is by Segarra and Ribeiro [29], who analyzed the sensitivity of centralities to small changes in the weight of edges. However, many real-world networks are essentially unweighted. For such networks, their measure may not characterize the sensitivity well, because theirs necessarily considers *continuous* changes of the edge weights, which are not apparent for such networks.

Our contributions. In this work, we propose a novel quantitative measure for the sensitivity of centralities to edge removals, called

*Supported by JST ERATO Grant Number JPMJER1305
†Supported by JSPS KAKENHI Grant Number JP17H04676.

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313422

(average-case) edge sensitivity. Intuitively speaking, it measures how the centrality of a uniformly chosen vertex changes by removing a uniformly chosen edge. Not only is the proposed measure intuitive, as we will see soon, it also provides for a finer comparison of centralities than the sensitivity measure studied in [29].

We conducted a theoretical analysis of the edge sensitivities of six major centralities: the closeness centrality [4, 5, 27], harmonic centrality [21], betweenness centrality [14], endpoint betweenness centrality [9], PageRank [25], and spanning tree centrality [30]. An overview of our results is shown in Table 1, where the lower bounds indicate sensitivities of worst-case graphs. Our upper bounds show that, for a typical case of $D \approx 1/(1-\alpha) \ll b \approx n$, the harmonic centrality is the most stable (i.e., the least sensitive). Although the closeness centrality is very sensitive, it can be as stable as the harmonic centrality when the graph has only a small number of bridges. Our lower bounds show that our upper bounds are almost tight except for spanning tree centrality.

In practice, it is likely that we miss more than one edge when constructing network datasets and the edge sensitivity introduced above may not appropriately reflect such situations. To address this issue, for an integer $k \geq 1$, we introduce *(average-case) k-edge sensitivity*, which matches the previous edge sensitivity when $k = 1$. Intuitively speaking, it measures how the centrality of a uniformly chosen vertex changes by removing a uniformly chosen set of edges of size k. An overview of our results is shown in the last column of Table 1, which shows a similar tendency to the case of edge sensitivity. Note that $\chi_1 = b/m$. While Table 1 contains the lower bounds for sensitivity (in the worst case graph), we omitted the proof of these bounds from this paper due to the limitation of space.

Finally, we computed the (k-)edge sensitivities of synthetic and real-world graphs and obtained experimental results that confirm our theoretical bounds—expect for the spanning tree centrality, which was significantly less sensitive than predicted by our theoretical upper bound. Also, our experimental results suggest that edge sensitivity can be used as a good proxy for k-edge sensitivity when comparing centralities in terms of sensitivity against removals of multiple edges, which is desirable because computing the former is computationally less expensive than computing the latter.

Organization. The rest of this paper is organized as follows. In Section 2, we review previous works on the sensitivity of centralities. In Section 3, we introduce the notations used throughout this paper and definitions for the centralities we analyze. We define edge sensitivity in Section 4 and theoretically bound the edge sensitivities of those centralities in Section 5. In Section 6, we introduce and study k-edge sensitivity. Finally, Section 7 shows our experimental results, which confirm our theoretical analysis.

2 RELATED WORK

In this section, we review previous works pertaining to the sensitivity of centralities.

Most of the following works are empirical, however. In [7], the sensitivity of various centralities was experimentally analyzed using synthetic graphs, several robustness measures (such as the recall of vertices with the highest centralities), and several types of modifications to the graph (such as edge removal). Costenbader and Valente [10] studied the sensitivity of various centralities on graphs constructed from a larger graph by sampling vertices. Frantz *et al.* [13] conducted an empirical investigation of the relationship between the network topology and the sensitivity of centralities.

Sensitivity has been also studied from theoretical aspects. Ghosal and Ghosal and Barabási [15] theoretically analyzed the sensitivity of the ranking by PageRank on networks of several models. Platig et al. [26] and Tsugawa and Ohsaki [31] evaluated the sensitivity of the degree centrality. However, there are only a few theoretical works that compare sensitivities of centralities. Ng et al. [23] compares sensitivity of two eigenvector-based centralities (HITS [18] and PageRank [25]). Ufimtsev et al. [32] qualitatively disucussed how the centrality ranking is affected by the graph structure. Recently, Segarra and Ribeiro [29] introduced a sensitivity measure of a centrality for weighted networks. Their measure classifies centralities into three categories: *stable*, *continuous*, and *non-continuous*. As such, their measure is not useful for comparing centralities quantitatively.

3 PRELIMINARIES

In what follows, n and m denote the number of vertices and edges, respectively, of the graph we are concerned with.

Let $G = (V, E)$ be an undirected or directed graph. For an edge $e \in E$, we denote by $G \setminus e$ the subgraph $(V, E \setminus \{e\})$. Similarly, for a subset of edges $F \subseteq F$, $G \setminus F$ denotes the subgraph $(V, E \setminus F)$. For vertices $u, v \in V$, let $d_G(u, v)$ be the (geodesic) distance from u to v, that is, the length of a shortest path from u to v. If v is unreachable from u, we define $d_G(u, v) = \infty$. The diameter of G, denoted by D_G, is defined as $\max_{u,v \in V} d_G(u, v)$. Let $A_G \in \{0, 1\}^{n \times n}$ be the adjacency matrix of G. That is, $A_G(u, v) = 1$ if and only if $uv \in E$.

Let $G = (V, E)$ be a directed graph on n vertices. We denote the out-degree of a vertex v in G by $\delta_G(v)$. Let $\Delta_G \in \mathbb{R}^{n \times n}$ be a diagonal matrix whose diagonal element $\Delta_G(v, v)$ equals $\delta_G(v)$. Then, the row-normalized adjacency matrix $\bar{A}_G \in \mathbb{R}^{n \times n}$ is defined as $\bar{A}_G = {\Delta_G}^{-1} A_G$. If a vertex $u \in V$ has no outgoing edges, we let $\bar{A}_G(u, v) = 0$ for every $v \in V$.

If the graph G is clear from the context, we may omit G in the above-mentioned notations.

3.1 Centralities

In this paper, we discuss four vertex centralities, namely, the closeness, harmonic, and betweenness centralities, along with PageRank. Furthermore, we discuss an edge centrality, namely, the spanning tree centrality. Here, we present the definitions for these centralities.

Definition 3.1 (Closeness centrality [4, 5, 27]). *For a graph $G = (V, E)$, the* closeness centrality *of a vertex $u \in V$ is defined as*

$$\mathrm{CC}_G(u) = \frac{n-1}{\sum_{v \in V} d_G(u, v)}.$$

If G is disconnected, $\sum_{v \in V} d_G(u, v) = \infty$ and thus we let $\mathrm{CC}_G(u) = 0$.

Definition 3.2 (Harmonic centrality [21]). *For a graph $G = (V, E)$, the* harmonic centrality *of a vertex $u \in V$ is defined as*

$$\mathrm{HC}_G(u) = \sum_{v \in V, v \neq u} \frac{1}{d_G(u, v)}.$$

It should be noted that $\mathrm{HC}_G(u)$ *is well-defined even if* G *is not connected, because we have* $\frac{1}{d_G(u,v)} = 0$ *if* v *is unreachable from* u.

Definition 3.3 (Betweenness centrality [14]). *For a graph* $G = (V, E)$*, the* betweenness centrality *of a vertex* $u \in V$ *is defined as*

$$\mathrm{BT}_G(u) = \sum_{(s,t)\in V^2, s\neq u\neq t} \frac{\sigma_{G,st}(u)}{\sigma_{G,st}},$$

where $\sigma_{G,st}$ *is the number of shortest paths from* s *to* t *in* G *and* $\sigma_{G,st}(u)$ *is the number of such paths that pass through* u.

In this definition, $\sigma_{G,st}$ can be 0 if s and t are disconnected in G (then, $\sigma_{G,st}(u)$ is also 0 for every $u \in V$). We simply let $\frac{\sigma_{G,st}(u)}{\sigma_{G,st}} = 0$ if there is no path from s to t in G. This corresponds to taking the sum over pairs (s, t) such that s and t are connected.

There is a variant of betweenness centrality that counts the *shortest paths having* v *as an endpoint.*

Definition 3.4 (Endpoint betweenness [9]). *For a graph* $G = (V, E)$*, the* endpoint betweenness centrality *of a vertex* $u \in V$ *is defined as*

$$\mathrm{BT}'_G(u) = \sum_{(s,t)\in V^2} \frac{\sigma_{G,st}(u)}{\sigma_{G,st}}.$$

We can see that BT and BT′ are closely tied:

Proposition 3.5. *When* G *is connected,* $\mathrm{BT}'_G(v) = \mathrm{BT}_G(v) + 2n - 1$ *holds for every* $v \in V$.

Definition 3.6 (PageRank [25]). *Let* $G = (V, E)$ *be a directed graph. Let* $0 < \alpha < 1$ *be a decaying factor and* $\boldsymbol{\pi} \in \mathbb{R}^n_{\geq 0}$ *be an initial distribution such that* $\|\boldsymbol{\pi}\|_1 = 1$*. Then, the PageRank vector* $\mathrm{PR}_G \in \mathbb{R}^n_{\geq 0}$ *of* G *is defined as*

$$\mathrm{PR}_G = (1-\alpha)(I - \alpha \bar{A}_G^\top)^{-1}\boldsymbol{\pi},$$

and the PageRank of a vertex u *is* $\mathrm{PR}_G(u)$.

As in [25], we use the uniform distribution as the initial distribution. That is, $\boldsymbol{\pi} = \frac{1}{n}\mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^n$ is the all-one vector.

The four centralities above are all defined for vertices. We next present the definition for the spanning tree centrality, which is defined for edges.

Definition 3.7 (Spanning tree centrality [30]). *For a connected graph* $G = (V, E)$*, the* spanning tree centrality *of an edge* $e \in E$ *is defined as*

$$\mathrm{ST}_G(e) = \frac{\tau_G(e)}{\tau_G}$$

where τ_G *is the number of spanning trees in* G *and* $\tau_G(e)$ *is the number of spanning trees including* e.

Note that this definition can be extended to disconnected graphs by considering spanning forests instead of spanning trees.

4 AVERAGE-CASE EDGE SENSITIVITY

In this section, we introduce sensitivity notions regarding centralities.

4.1 Sensitivity of vertex centralities

We first consider the sensitivity of vertex centralities such as closeness and harmonic centralities.

Before providing the definition, we outline some requirements that the sensitivity notion must satisfy. First, it should be invariant under the scaling of centralities. Note that different centralities may have different scales: the closeness centrality of a vertex takes is value within $[0, 1]$, whereas the harmonic centrality of a vertex is at least 1 (unless the vertex is isolated). Moreover, there are some variations to the definition of a centrality. For example, the definition of the closeness centrality in [6] differs from ours by a factor of $n - 1$. Because we want to compare different centralities, their sensitivities should be invariant under scaling.

Second, we want to understand how the structure of the underlying network affects the sensitivity. Hence, the notion of sensitivity should be parameterized by the underlying network.

To describe our sensitivity measure, we first state some definitions. A vertex centrality can be regarded as a function that, given a graph $G = (V, E)$ and a vertex $v \in V$, computes the centrality value of v in G. With this observation, we say that a family of functions $C = \{C_G : V(G) \to \mathbb{R}_{\geq 0}\}_G$ indexed by a graph G is a *vertex centrality function*. For example, CC, HC, BT, and PR introduced in Section 3 are all centrality functions.

Definition 4.1 (Average-case sensitivity for vertex centralities). *For a graph* $G = (V, E)$ *and a vertex centrality function* $C = \{C_G\}_G$*, the* (average-case) edge sensitivity *of* C *on* G *is defined as*

$$S^v_C[G] = \frac{1}{|V||E|}\sum_{e\in E}\sum_{v\in V}\frac{\left|C_{G\setminus e}(v) - C_G(v)\right|}{C_G(v)}$$

In other words, the edge sensitivity is the expected value of the relative change of the centrality value of a uniformly chosen vertex when a uniformly chosen edge is removed. The smaller the edge sensitivity of a centrality is, the less sensitive its centrality is to changes in the network.

As explained above, the edge sensitivity is invariant under scaling:

Theorem 4.2. *Let* C *be a centrality function and* C' *be another centrality function, such that* $C'_G(u) = \lambda(G)C_G(u)$*, where* $\lambda(G) > 0$ *is a scaling factor that may depend on* G*. Then, we have* $S^v_C[G] = S^v_{C'}[G]$ *for any graph* G.

4.2 Sensitivity of edge centralities

Next, we extend the proposed edge sensitivity from vertex centralities to edge centralities, such as the spanning tree centrality.

As in the case of vertex centralities, we define the general notion of an edge centrality: We say that a family of functions $C = \{C_G : E(G) \to \mathbb{R}_{\geq 0}\}_G$ indexed by a graph G is an *edge centrality function.*

Definition 4.3 (Average-case sensitivity for edge centralities). *For a graph* $G = (V, E)$ *and an edge centrality function* $C = \{C_G\}_G$*, the edge sensitivity of* C *on* G *is defined as*

$$S^e_C[G] = \frac{1}{|E|(|E|-1)}\sum_{f\in E}\sum_{e\in E\setminus\{e\}}\frac{\left|C_{G\setminus f}(e) - C_G(e)\right|}{C_G(e)}$$

This sensitivity is also invariant under scaling per Theorem 4.2.

The distinction between $S_C^v[G]$ and $S_C^e[G]$ is not essential, because they are defined for different families of centrality functions. Thus, $S_C^v[G]$ and $S_C^e[G]$ are hereafter written as $S_C[G]$.

5 THEORETICAL ANALYSIS

In this section, we theoretically analyze the edge sensitivities of the centralities introduced in Section 3. The proofs of our lower bounds are deferred to the full version due to space limitation. Then, we compare those centralities in terms of edge sensitivity in Section 5.6.

5.1 Closeness centrality

As mentioned above, the closeness centrality of a vertex in a disconnected graph is 0, and hence its edge sensitivity is not well defined against disconnected graphs. For connected graphs, we have the following bound:

Theorem 5.1. *Let G be a connected undirected graph with at least 2 vertices and b bridges. Then, we have $\frac{b}{m} \le S_{\mathrm{CC}}[G] \le \frac{2D_G + b}{m}$.*

To prove this theorem, we use the following lemma:

Lemma 5.2 ([28]). *Let $G = (V, E)$ be an undirected graph and $F \subseteq E$. If $G \setminus F$ is connected, then $D_{G \setminus F} \le (|F| + 1)D_G$ holds.*

Proof of Theorem 5.1. Let $B \subseteq E$ be the set of bridges in $G = (V, E)$. Since removing a bridge makes the graph disconnected, we have $\frac{|C_{G\setminus e}(v) - C_G(v)|}{C_G(v)} = 1$ if $e \in B$. Then, we have

$$S_{\mathrm{CC}}[G] \ge \frac{1}{|V||E|} \sum_{e \in B} \sum_{v \in V} \frac{\left|\mathrm{CC}_{G\setminus e}(v) - \mathrm{CC}_G(v)\right|}{\mathrm{CC}_G(v)} = \frac{b}{m}.$$

Let $W_G(u) = \sum_{v \in V} d(u, v)$. Then, $\mathrm{CC}_G(u) = \frac{n-1}{W_G(u)}$ holds. It is also easy to see that $W_G(u) \le W_{G\setminus e}(u)$ for any edge $e \in E$. Therefore, we have

$$\frac{\left|\mathrm{CC}_{G\setminus e}(v) - \mathrm{CC}_G(v)\right|}{\mathrm{CC}_G(v)} = \frac{W_{G\setminus e}(v) - W_G(v)}{W_{G\setminus e}(v)} \le \frac{W_{G\setminus e}(v) - W_G(v)}{W_G(v)}.$$

Using this, we obtain

$$\sum_{e \in E\setminus B} \frac{\left|\mathrm{CC}_{G\setminus e}(v) - \mathrm{CC}_G(v)\right|}{\mathrm{CC}_G(v)} \le \frac{1}{W_G(v)} \sum_{e \in E\setminus B} (W_{G\setminus e}(v) - W_G(v))$$
$$\le \frac{1}{W_G(v)} \sum_{w \in V} \sum_{e \in E\setminus B} (d_{G\setminus e}(v, w) - d_G(v, w)). \quad (1)$$

For $v, w \in V$, we consider a shortest path $p(v, w)$ connecting v and w in G. Note that $p(v, w)$ contains $d_G(v, w)$ edges. If an edge $e \in E\setminus B$ is not a member of $p(v, w)$, we have $d_{G\setminus e}(v, w) = d_G(v, w)$. Otherwise, we have $d_{G\setminus e}(v, w) \le 2D_G$ according to Lemma 5.2. Therefore,

$$\sum_{w \in V} \sum_{e \in E\setminus B} (d_{G\setminus e}(v, w) - d_G(v, w))$$
$$= \sum_{w \in V} \sum_{e \in (E\setminus B) \cap p(v,w)} (d_{G\setminus e}(v, w) - d_G(v, w))$$
$$\le \sum_{w \in V} 2D_G d_G(v, w) = 2W_G(v)D_G. \quad (2)$$

Using (1), (2), and the fact that $\frac{|C_{G\setminus e}(v) - C_G(v)|}{C_G(v)} = 1$ for $e \in B$, we obtain $S_{\mathrm{CC}}[G] \le \frac{2D_G + b}{m}$, as desired. □

5.2 Harmonic centrality

Theorem 5.3. *Let G be an undirected or directed graph. Then, we have $S_{\mathrm{HC}}[G] \le \frac{n-1}{nm} \sum_{v \in V} \frac{1}{\mathrm{HC}[G](v)}$.*

Proof. Analogous to the proof for the closeness centrality, we have for any vertex $v \in V(G)$,

$$\sum_{e \in E} \frac{\left|\mathrm{HC}_{G\setminus e}(v) - \mathrm{HC}_G(v)\right|}{\mathrm{HC}_G(v)} = \sum_{e \in E} \frac{\mathrm{HC}_G(v) - \mathrm{HC}_{G\setminus e}(v)}{\mathrm{HC}_G(v)}$$
$$= \frac{1}{\mathrm{HC}_G(v)} \sum_{e \in E} \sum_{w \in V, w \ne v} \left(\frac{1}{d_G(v, w)} - \frac{1}{d_{G\setminus e}(v, w)} \right)$$
$$= \frac{1}{\mathrm{HC}_G(v)} \sum_{w \in V, w \ne v} \sum_{e \in p(v,w)} \frac{1}{d_G(v, w)} \le \frac{n-1}{\mathrm{HC}_G(v)}.$$

Note that the inequality holds for both undirected and directed graphs. It follows that $S_{\mathrm{HC}}[G] \le \frac{n-1}{nm} \sum_{v \in V} \frac{1}{\mathrm{HC}[G](v)}$ holds. □

Corollary 5.4. *Let G be a (strongly) connected graph. Then, we have $S_{\mathrm{HC}}[G] \le \frac{\bar{d}_G}{m} \le \frac{D_G}{m}$, where $\bar{d}_G$ is the average distance between vertices: $\bar{d}_G = \frac{1}{n(n-1)} \sum_{u, v \in V, u \ne v} d(u, v)$.*

Proof. We have $\frac{n-1}{\mathrm{HC}[G](v)} \le \frac{1}{n-1} \sum_{w \ne v} d(v, w)$ by the HM-AM inequality. It follows that $\frac{n-1}{n} \sum_{v \in V} \frac{1}{\mathrm{HC}[G](v)} \le \bar{d}_G$. Combining this with Theorem 5.3, we obtain $S_{\mathrm{HC}}[G] \le \frac{\bar{d}_G}{m}$. □

5.3 Betweenness centrality

Our definition of edge sensitivity cannot be directly applied to the betweeness centrality because $\mathrm{BT}_G(v) = 0$ and $\mathrm{BT}_{G\setminus e}(v) > 0$ could hold at the same time, resulting in a zero-division problem. More specifically, consider a triangle graph consisting of three vertices $\{v, x, y\}$ and let $e = (x, y)$. Then, $\mathrm{BT}_G(v) = 0$ holds, whereas $\mathrm{BT}_{G\setminus e}(v) > 0$ because the only path between x and y in $G \setminus e$ passes through v.

Another issue with the betweenness centrality is its instability. Consider a complete graph K_n on n vertices. It is clear that $\mathrm{BT}_{K_n}(v) = 0$ for every v and $\mathrm{BT}_{K_n\setminus xy}(v) > 0$ for $v \ne x, y$. That is, removing an edge renders the centralities of $n - 2$ vertices nonzero even though they were all zero before its removal.

However, we can give non-trivial bounds on the edge sensitivity of endpoint betweeness centrality:

Theorem 5.5. *Let G be a connected undirected graph. Then, we have $S_{\mathrm{BT}'}[G] \le \frac{{D_G}^2 + 3D_G}{m}$.*

Although we consider endpoint betweenness centrality rather than (normal) betweenness centrality, their values are relatively close for highly important vertices, since the average betweenness centrality is $\Omega(n)$. Therefore, in practice, we can suppose that analysis on endpoint betweenness centrality will provide useful results for betweenness centrality as well.

For a graph G, we denote the number of paths of length d between two vertices s and t in G by $\sigma_{G,st}^d$. Similarly, the number of such paths passing through a vertex v is denoted by $\sigma^d_{G,st}(v)$. Before proving this theorem, we need the following technical lemma:

Lemma 5.6. *If s and t are connected in G, we have*

(1) $\sum_{v\in V}\sigma_{G,st}(v) = (d_G(s,t)+1)\sigma_{G,st}$,

(2) $\sum_{e\in E}(\sigma^{d_G(s,t)}_{G,st}(v) - \sigma^{d_G(s,t)}_{G\setminus e,st}(v)) = d_G(s,t)\sigma^{d_G(s,t)}_{G,st}(v)$.

Proof. (1) Both sides of the equation count the number of pairs (p,v), where p is a shortest path between s and t, and v is a vertex that p passes through.

(2) Note that $\sigma^{d_G(s,t)}_{G,st}(v) - \sigma^{d_G(s,t)}_{G\setminus e,st}(v)$ is the number of paths of length $d_G(s,t)$ between s and t passing through both v and e. Both sides of the equation count the number of pairs (p,e), where p is a path of length $d_G(s,t)$ between s and t, and e is an edge that p passes through. □

Here we define $d'_G(s,t)$ by $d'_G(s,t) = \begin{cases} d_G(s,t) & d_G(s,t) < \infty \\ -1 & \text{otherwise}\end{cases}$. Then, (1) of Lemma 5.6 can be generalized to graphs G in which s and t may not be connected: $\sum_{v\in V}\frac{\sigma_{G,st}(v)}{\sigma_{G,st}} = d'_G(s,t)+1$.

Proof of Theorem 5.5. For every pair of vertices $(s,t)\in V^2$, we let $P(s,t) = \left\{e\in E \mid d_G(s,t) = d_{G\setminus e}(s,t)\right\}$. To bound $S_{\mathrm{BT}'}[G]$, we first consider the *sum of absolute differences*:

$$A = \frac{1}{nm}\sum_{v\in V}\sum_{e\in E}\left|\mathrm{BT}'_{G\setminus e}(v) - \mathrm{BT}'_G(v)\right|.$$

Using the triangle inequality, we have

$$A \le \frac{1}{nm}\sum_{(s,t)\in V^2}\sum_{e\in E}\sum_{v\in V}\left|\frac{\sigma_{G\setminus e,st}(v)}{\sigma_{G\setminus e,st}} - \frac{\sigma_{G,st}(v)}{\sigma_{G,st}}\right|.$$

From Lemma 5.6, we have

$$\begin{aligned}
&\sum_{e\in E}\sum_{v\in V}\left|\frac{\sigma_{G\setminus e,st}(v)}{\sigma_{G\setminus e,st}} - \frac{\sigma_{G,st}(v)}{\sigma_{G,st}}\right| \\
=&\sum_{e\in E}\sum_{v\in V}\left|\frac{\sigma_{G\setminus e,st}(v)}{\sigma_{G\setminus e,st}} - \frac{\sigma^{d_G(s,t)}_{G\setminus e,st}(v)}{\sigma_{G,st}} + \frac{\sigma^{d_G(s,t)}_{G\setminus e,st}(v)}{\sigma_{G,st}} - \frac{\sigma_{G,st}(v)}{\sigma_{G,st}}\right| \\
\le& 2\cdot\sum_{e\in E}\sum_{v\in V}\left(\frac{\sigma_{G,st}(v)}{\sigma_{G,st}} - \frac{\sigma^{d_G(s,t)}_{G\setminus e,st}(v)}{\sigma_{G,st}}\right) \\
&+\sum_{e\in P(s,t)}\sum_{v\in V}\left|\frac{\sigma_{G\setminus e,st}(v)}{\sigma_{G\setminus e,st}} - \frac{\sigma_{G,st}(v)}{\sigma_{G,st}}\right| \\
\le& 2D_G(D_G+1) + \sum_{e\in P(s,t)}(d'_{G\setminus e}(s,t) - d'_G(s,t)) \le 4{D_G}^2 + 2D_G.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
A &= \frac{1}{nm}\sum_{(s,t)\in V^2, s\neq t}\sum_{e\in E}\sum_{v\in V}\left|\frac{\sigma_{G\setminus e,st}(v)}{\sigma_{G\setminus e,st}} - \frac{\sigma_{G,st}(v)}{\sigma_{G,st}}\right| \\
&\le \frac{n-1}{m}\cdot(4{D_G}^2 + 2D_G).
\end{aligned}$$

Furthermore, $\mathrm{BT}'_G(u) \ge 2n-1$ follows from Proposition 3.5. Accordingly, we have $S_{\mathrm{BT}'}[G] \le \frac{A}{2n-1} \le \frac{2{D_G}^2 + D_G}{m}$. □

5.4 PageRank

To analyze the edge sensitivity of PageRank, we use a random-walk-based representation of PageRank [3]. For a walk $w = w_0w_1\ldots w_{\ell(w)}$ of length $\ell(w)$ in a graph G of n vertices and a matrix $M\in\mathbb{R}^{n\times n}$, we define the *weight of w with respect to M* by

$$W_M(w) = \prod_{i=0}^{\ell-1} M(w_i, w_{i+1}).$$

Lemma 5.7. *The Pagerank of a vertex $v\in V(G)$ satisfies*

$$\mathrm{PR}_G(v) = \sum_{w\in\mathrm{walks}_G(\star,v)}\frac{1-\alpha}{n}W_{\alpha\bar{A}_G}(w),$$

where $\mathrm{walks}_G(\star,v)$ *is the set of walks in G terminating at v.*

The following is almost immediate from the definition and we omit the proof:

Lemma 5.8. *Let* walks_G *be the (infinite) set of walks in G. Then, we have $\sum_{w\in\mathrm{walks}_G,\ell(w)=l}\frac{1-\alpha}{n}W_{\alpha\bar{A}_G}(w) \le (1-\alpha)\alpha^l$. Equality holds if every vertex in G has at least one out-going edge.*

We first bound $S_{\mathrm{PR}}[G]$ exclusively for special graphs.

Lemma 5.9. *Suppose that every vertex $v\in V$ satisfies $\delta_G(v)\ge 1$. Then, we have $S_{\mathrm{PR}}[G] \le \frac{2\alpha}{m(1-\alpha)^2}$.*

Proof. Let $e = st\in E$ be an edge to be removed, and let $M = \alpha\bar{A}_G$, $M'' = \alpha\bar{A}_{G\setminus e}$ and $M'\in\mathbb{R}^{n\times n}$ be the matrix obtained from M by replacing $M(s,t)$ with 0. Then, it is clear that $M\ge M'\le M''$ and thus $W_M(w)\ge W_{M'}(w)\le W_{M''}(w)$ holds for every walk w. Additionally, since every vertex in G has at least one out-going edge, we have $\sum_{v\in V}\mathrm{PR}_G(v) = 1 \ge \sum_{v\in V}\mathrm{PR}_{G\setminus e}(v)$ from Lemma 5.8. Therefore, we have

$$\begin{aligned}
&\sum_{v\in V}\left|\mathrm{PR}_G(v) - \mathrm{PR}_{G\setminus e}(v)\right| \le \frac{1-\alpha}{n}\sum_{w\in\mathrm{walks}_G(\star,v)}|W_M(w) - W_{M''}(w)| \\
\le&\frac{1-\alpha}{n}\sum_{w\in\mathrm{walks}_G(\star,v)}((W_M(w) - W_{M'}(w)) + (W_{M''}(w) - W_{M'}(w))) \\
\le&\frac{2(1-\alpha)}{n}\sum_{w\in\mathrm{walks}_G(\star,v)}(W_M(w) - W_{M'}(w)) \\
\le&\frac{2(1-\alpha)}{n}\sum_{w\in\mathrm{walks}_G(\star,v), e\in w}W_M(w).
\end{aligned}$$

From Lemma 5.8, we obtain

$$\begin{aligned}
&\sum_{e\in E}\sum_{v\in V}\left|\mathrm{PR}_G(v) - \mathrm{PR}_{G\setminus e}(v)\right| \le \frac{2(1-\alpha)}{n}\sum_{e\in E}\sum_{\substack{w\in\mathrm{walks}_G:\\ e\in w}}W_M(w) \\
\le&\frac{2(1-\alpha)}{n}\sum_{w\in\mathrm{walks}_G}W_M(w)\ell(w) \le 2\sum_{\ell=0}^{\infty}\ell(1-\alpha)\alpha^\ell = \frac{2\alpha}{1-\alpha}.
\end{aligned}$$

Here note that $\mathrm{PR}(v)\ge\frac{1-\alpha}{m}$ holds because there is a walk of length 0 starting and ending at v for every $v\in V$. Finally we have

$$\begin{aligned}
S_{\mathrm{PR}}[G] &= \frac{1}{nm}\sum_{e\in E}\sum_{v\in V}\frac{\left|\mathrm{PR}_{G\setminus e}(v) - \mathrm{PR}_G(v)\right|}{\mathrm{PR}_G(v)} \\
&\le \frac{1}{nm}\cdot\frac{n}{1-\alpha}\sum_{e\in E}\sum_{v\in V}\left|\mathrm{PR}_{G\setminus e}(v) - \mathrm{PR}_G(v)\right| \le \frac{2\alpha}{m(1-\alpha)^2}. \quad \square
\end{aligned}$$

From Lemma 5.9, we prove the same inequality for a general G:

Theorem 5.10. *We have* $S_{\mathrm{PR}}[G] \leq \frac{2\alpha}{m(1-\alpha)^2}$.

Proof. Let $T = \{v \in V \mid \delta_G(v) = 0\}$, $F = \{(t,t) \mid t \in T\}$ and $G' = (V, E \cup F)$. We can see that $\mathrm{PR}_{G'}(v) = \mathrm{PR}_G(v)$ if $\delta_G(v) \geq 1$ and $\mathrm{PR}_{G'}(v) = \frac{\mathrm{PR}_G(v)}{1-\alpha}$ if $\delta_G(v) = 0$. Therefore, we have $\frac{\left|\mathrm{PR}_{G'\setminus e}(v) - \mathrm{PR}_{G'}(v)\right|}{\mathrm{PR}_{G'}(v)} = \frac{\left|\mathrm{PR}_{G\setminus e}(v) - \mathrm{PR}_G(v)\right|}{\mathrm{PR}_G(v)}$ for every $e \in E$ and $v \in V$. Accordingly from Lemma 5.9, we obtain

$$\begin{aligned} S_{\mathrm{PR}}[G] &= \frac{1}{nm} \sum_{e \in E} \sum_{v \in V} \frac{\left|\mathrm{PR}_{G\setminus e}(v) - \mathrm{PR}_G(v)\right|}{\mathrm{PR}_G(v)} \\ &= \frac{1}{nm} \sum_{e \in E} \sum_{v \in V} \frac{\left|\mathrm{PR}_{G'\setminus e}(v) - \mathrm{PR}_{G'}(v)\right|}{\mathrm{PR}_{G'}(v)} \\ &\leq \frac{|E \cup F|}{|E|} S_{\mathrm{PR}}[G'] \leq \frac{2\alpha}{m(1-\alpha)^2}. \end{aligned}$$ □

5.5 Spanning tree centrality

To analyze the edge sensitivity of the spanning tree centrality, we use the relationship between the spanning tree centrality and *effective resistance*, which is defined as follows:

Definition 5.11 (Effective resistance). *For an undirected graph $G = (V, E)$, we imagine an electric circuit obtained by replacing every edge in G with a resistor of a unit resistance. For every pair of vertices $(u, v) \in V^2$, the* effective resistance *between u and v is defined as the difference in their electric potential when we inject a unit electricity from u to v.*

Theorem 5.12 (see, e.g., [22]). *Let $G = (V, E)$ be an undirected graph. For every edge $e = uv \in E$, $\mathrm{ST}_G(e)$ is the effective resistance between u and v in G.*

The following pertains to effective resistance:

Theorem 5.13 (Rayleigh's monotonicity law [11]). *If the resistance of a resistor in a circuit increases, for every $(u, v) \in V^2$, the effective resistance between u and v does not decrease.*

Remark. *Theorem 5.13 applies to the removal of resistors, because the removal of a resistor corresponds to increasing its resistance to ∞.*

Theorem 5.14 (Dirichlet principle [11]). *For $v \in \mathbb{R}^V$, the energy consumption when the electrical potential of w is set to $v(w)$ for every $w \in V$ is $C(v) := \sum_{st \in E} (v(s) - v(t))^2$. Under the constraints of $v(x) = 1$ and $v(y) = 0$ ($x, y \in V$), the energy $C(v)$ is minimized when the electrical flow determined by v satisfies Kirchoff's law. In other words, $C(v)$ is at least the reciprocal of the effective resistance between x and y.*

Based on these facts, we prove some technical lemmas that are useful for our analysis.

Lemma 5.15. *Let $G = (V, E)$ be an undirected graph. For every $e, f \in E$ ($e \neq f$), we have* $\mathrm{ST}_G(e) \leq \mathrm{ST}_{G\setminus f}(e) \leq \frac{\mathrm{ST}_G(e)}{1-\mathrm{ST}_G(e)}$.

Proof. With Theorem 5.12, we can identify the spanning tree centrality of e with the effective resistance of $e \in E$. The first inequality immediately holds from Theorem 5.13.

We let $e = st$. Let $v \in \mathbb{R}^V$ be the electrical potential of every vertex when we apply a unit voltage between s and t in $G \setminus f$. From Theorem 5.14, the energy consumed when we apply a unit voltage between s and t in G is upper-bounded by the energy consumption when the electrical potential is given by v (note that this does not necessarily satisfy Kirchoff's law). Since this energy consumption is equivalent to the reciprocal of the effective resistance, we have $\frac{1}{\mathrm{ST}_G(e)} \leq \frac{1}{\mathrm{ST}_{G\setminus f}(e)} + 1$. Then, it is clear that $\mathrm{ST}_{G\setminus f}(e) \leq \frac{\mathrm{ST}_G(e)}{1-\mathrm{ST}_G(e)}$. □

Lemma 5.16. *Let $G = (V, E)$ be an undirected graph and $e, f \in E$ be distinct edges. Then,*

$$0 \leq \frac{\mathrm{ST}_{G\setminus f}(e) - \mathrm{ST}_G(e)}{\mathrm{ST}_G(e)} \leq \sqrt{2\left(\mathrm{ST}_{G\setminus f}(e) - \mathrm{ST}_G(e)\right)}.$$

Proof. The first inequality is immediate from Lemma 5.15.

Let $x = \mathrm{ST}_G(e)$ and $y = \mathrm{ST}_{G\setminus f}(e)$. To show the second inequality, it suffices to show $\frac{\sqrt{y-x}}{x} \leq \sqrt{2}$. If $x \geq \frac{1}{2}$, we have $\frac{\sqrt{y-x}}{x} \leq \sqrt{2}$ since $y \leq 1$ (from the definition of a spanning tree centrality). If $x < \frac{1}{2}$, using Lemma 5.15, we have $\frac{\sqrt{y-x}}{x} \leq \frac{1}{\sqrt{1-x}} \leq \sqrt{2}$. In both cases, we obtain the desired inequality. □

Lemma 5.17. *Let $G = (V, E)$ be a connected undirected graph. Then, $\sum_{e \in E} \mathrm{ST}_G(e) = |V| - 1$.*

Proof. This follows immediately from $\sum_{e \in E} \tau_G(e) = (|V| - 1)\tau_G$, which holds because every spanning tree in G has $|V| - 1$ edges. □

Theorem 5.18. *Let G be a two-connected undirected graph. Then, we have* $S_{\mathrm{ST}}[G] \leq \sqrt{\frac{2(n-1)}{m(m-1)}}$.

Proof. For $e, f \in E$ ($e \neq f$), we let $\delta_G(e, f) = \mathrm{ST}_{G\setminus f}(e) - \mathrm{ST}_G(e)$. From Lemma 5.17, we have

$$\begin{aligned} \sum_{e, f \in E, e \neq f} \delta_G(e, f) &= \sum_{f \in E} \sum_{e \in E \setminus \{f\}} (\mathrm{ST}_{G\setminus f}(e) - \mathrm{ST}_G(e)) \\ &= \sum_{f \in E} \left(\sum_{e \in E \setminus \{f\}} \mathrm{ST}_{G\setminus f}(e) - \sum_{e \in E} \mathrm{ST}_G(e) + \mathrm{ST}_G(f) \right) \\ &= \sum_{f \in E} \mathrm{ST}_G(f) = n - 1. \end{aligned}$$

Additionally, from Lemma 5.16,

$$\sum_{e, f \in E, e \neq f} \frac{\mathrm{ST}_{G\setminus f}(e) - \mathrm{ST}_G(e)}{\mathrm{ST}_G(e)} \leq \sum_{e, f \in E, e \neq f} \sqrt{2\delta_G(e, f)}.$$

Therefore, we have $\sum_{e, f \in E, e \neq f} \sqrt{2\delta_G(e, f)} \leq \sqrt{2(n-1)m(m-1)}$ by the Cauchy-Schwarz inequality. Thus we have

$$S_{\mathrm{ST}}[G] = \frac{1}{m(m-1)} \sum_{e, f \in E, e \neq f} \frac{\mathrm{ST}_{G\setminus f}(e) - \mathrm{ST}_G(e)}{\mathrm{ST}_G(e)} \leq \sqrt{\frac{2(n-1)}{m(m-1)}}$$

as desired. □

5.6 Comparison of centralities in terms of sensitivity

We turn now to a comparison of the centralities we have analyzed.

First, the edge sensitivities of the closeness and harmonic centralities are similar with regard to the diameter D_G and the number of edges m. However, the former degrades considerably when there are many bridges in the graph. Hence, the harmonic centrality is preferable to the closeness centrality unless the graph is highly connected.

In contrast, the edge sensitivity of endpoint betweenness centrality is $\Theta(D_G^2/m)$, which is conspiciously worse than that of the harmonic centrality.

The sensitivity of PageRank depends on the decaying factor α: the higher α is, the larger the sensitivity is. PageRank with larger value of α gives much weight on longer walks. Since longer walks are more likely to contain the randomly removed edge, PageRank can be more sensitive with larger α.

The edge sensitivities of the closeness, harmonic, and betweenness centralities depend on the diameter D_G, which is unavoidable according to our lower bounds (Table 1). By contrast, the edge sensitivity of PageRank depends on α (and m) rather than D_G. Although this implies that PageRank may be more stable than other centralities when the graph has a large diameter, the harmonic centrality is preferable for typical real small-world networks that have diameters of $O(\log n)$ [33].

Although our current bound on the edge sensitivity of spanning tree centrality is $O(\sqrt{n}/m)$, we cannot conclude that it is worse than other centralities, insofar as we do not have a matching lower bound.

6 AVERAGE-CASE k-EDGE SENSITIVITY

In this section, we discuss k-edge sensitivity, which is an extension of edge sensitivity for removing k edges.

6.1 Definition

Average-case edge sensitivity (Definition 4.1) considers the case that a (random) edge is missing in the observed graph. In practice, however, it is more likely that multiple edges are missing, and this motivates us to define the following *(average-case) k-edge sensitivity*:

Definition 6.1 (Average-case k-edge sensitivity for vertex centralities). *For a graph $G = (V, E)$ and a vertex centrality function $C = \{C_G\}_G$, the* k-edge sensitivity *of C on G is defined as*

$$S^v_{C,k}[G] = \frac{1}{|V|\binom{|E|}{k}} \sum_{F\subseteq E,\,|F|=k} \sum_{v\in V} \frac{\left|C_{G\setminus F}(v) - C_G(v)\right|}{C_G(v)}.$$

The k-edge sensitivity of an edge centrality $S^e_{C,k}[G]$ is defined analogously (note that the sum of the relative differences is taken over $v \in V \setminus F$). Again, we simply write $S^v_{C,k}[G]$ and $S^e_{C,k}[G]$ as $S_{C,k}[G]$. Note that $S_C[G] = S_{C,1}[G]$ holds and hence k-edge sensitivity is a natural extension of edge sensitivity. Also, it is notable that a k-edge sensitivity can be probabilistically defined:

$$S^v_{C,k}[G] = \mathrm{E}_F\left[\frac{1}{|V|}\sum_{v\in V}\frac{\left|C_{G\setminus F}(v) - C_G(v)\right|}{C_G(v)}\right],$$

where F is a uniformly chosen subset of E with cardinality k.

6.2 Theoretical analysis

In this section, we extend results in Section 5 to k-edge sensitivity.

Theorem 6.2. *Let G be a connected undirected graph. Then, $S_{\mathrm{CC},k}[G] \le \frac{k(k+1)D_G}{m} + \chi_k[G]$, where $\chi_k[G]$ is the probability that G becomes disconnected by removing a uniformly chosen set of k edges in G.*

Proof. We let $c(G, x) = x$ if G is connected and $c(G, x) = 0$ otherwise. For a fixed vertex $v \in V$,

$$\begin{aligned}
&E\left[c\left(G\setminus F, \sum_{w\in V}(d_{G\setminus F}(v,w) - d_G(v,w))\right)\right]\\
\le &E\left[c\left(G\setminus F, \sum_{w\in V,\,F\cap p(v,w)\neq\emptyset}(d_{G\setminus F}(v,w) - d_G(v,w))\right)\right]\\
\le &\sum_{w\in V} k(k+1)D_G d_G(v,w) = k(k+1)W_G(v)D_G
\end{aligned}$$

because F and $p(v, w)$ intersects with probability at most $\frac{kd_G(v,w)}{m}$, and if v and w are connected in $G \setminus F$, $d(v, w) \le (k + 1)D_G$ follows from Lemma 5.2. The rest of the proof is similar to that of Theorem 5.1. □

Theorem 6.3. *Let G be a (strongly) connected graph. Then, we have $S_{\mathrm{HC},k}[G] \le \frac{kD_G}{m}$.*

Proof. Let π_{uv} be a shortest path from u to v in G. Suppose a subset of edges $F \subseteq E$ is removed from G. Then, the distance from u to v decreases only if F and π_{uv} shares at least one edge. If F is chosen uniformly at random from every subset of E with cardinality k, this occurs with probability at most $\frac{kd_G(u,v)}{m}$. Thus we have $S_{\mathrm{HC},k}[G] \le \frac{kD_G}{m}$ as in the proof of Theorem 5.3 and Corollary 5.4. □

Theorem 6.4. *We have $S_{\mathrm{PR},k}[G] \le \frac{2\alpha k}{m(1-\alpha)^2}$.*

Proof. The probability that a walk w of length $\ell(w)$ shares at least one edge with a uniformly random subset F of E with cardinality k is at most $\frac{k\ell(w)}{m}$. Therefore, as in the proof of Lemma 5.9, we have

$$\mathrm{E}\left[\sum_{v\in V}\left|\mathrm{PR}_G(v) - \mathrm{PR}_{G\setminus F}(v)\right|\right] \le \frac{2k\alpha}{1-\alpha}.$$

The rest of the proof is similar to that of Lemma 5.9 and Theorem 5.10. □

Theorem 6.5. *Let G be a undirected connected graph. Then, we have $S_{\mathrm{BT},k}[G] \le \frac{(k^2+3k){D_G}^2/2+kD_G}{m}$.*

To prove this, we introduce an extension of Lemma 5.6:

Lemma 6.6. *If s and t are connected in G, we have*

$$\mathrm{E}\left[\sigma^{d_G(s,t)}_{G,st}(v) - \sigma^{d_G(s,t)}_{G\setminus F,st}(v)\right] \le \frac{k}{m}\cdot d_G(s,t)\sigma^{d_G(s,t)}_{G,st}(v).$$

Table 2: Dataset for experiments on real graphs

Name	Directed?	n	m
ca-CondMat [20]	No	21363	91286
ca-GrQc [20]	No	4158	13422
email-EuAll [20]	Yes	34203	151132
facebook_combined [20]	No	4039	88234
petster-hamster [19]	No	2000	16098

Proof of Theorem 6.5. From Lemma 6.6, we have

$$\mathrm{E}\left[\sum_{v\in V}\left(\frac{\sigma_{G,st}(v)}{\sigma_{G,st}}-\frac{\sigma^{d_G(s,t)}_{G\setminus F,st}(v)}{\sigma_{G,st}}\right)\right]\le\frac{kd_G(s,t)(D_G+1)}{m}.$$

Also, we have $E\left[d'_{G\setminus F}(s,t)-d'_G(s,t)\right]\le k(k+1){D_G}^2$. Using them, the proof is analogous to that of Theorem 5.5. □

Theorem 6.7. *Let G be a undirected graph. Then, we have $S_{\mathrm{ST},k}[G]\le\frac{k\sqrt{(k+1)n}}{m-k}$.*

Proof. Lemma 5.15 can be generalized as $\frac{1}{\mathrm{ST}_G(e)}\le\frac{1}{\mathrm{ST}_{G\setminus F}(e)}+|F|$. Using this, we have

$$0\le\frac{\mathrm{ST}_{G\setminus F}(e)-\mathrm{ST}_G(e)}{\mathrm{ST}_G(e)}\le\sqrt{k(k+1)\left(\mathrm{ST}_{G\setminus F}(e)-\mathrm{ST}_G(e)\right)}.$$

which is an extension of Corollary 5.16. Additionally, we can see $\mathrm{E}\left[\sum_{e\in E\setminus F}\left(\mathrm{ST}_{G\setminus F}(e)-\mathrm{ST}_G(e)\right)\right]\le\frac{k(n-1)}{m}$ as in the proof of Theorem 5.18. Therefore, by the Cauchy–Schwarz inequality, we obtain $S_{\mathrm{ST},k}\le k\sqrt{\frac{(k+1)(n-1)}{m(m-k)}}$. □

6.3 Comparison of centralities in terms of k-edge sensitivity

We first compare the upper bounds on k-edge sensitivity with the lower bounds on (1-)edge sensitivity, as a function of k. For the harmonic centrality and PageRank, the ratio is $O(k)$. This means that the harmonic centrality and PageRank are stable against removal of multiple edges as much as against removal of a single edge because, roughly speaking, removal of k edges will cause k times more changes in the centrality values than removal of one edge. However, this ratio is $O(k^2)$ for the closeness centrality and the endpoint betweenness centrality, and is $O(k^{1.5})$ for the spanning tree centrality. Recall that, in Section 5, we argued that the harmonic centrality is preferable for small-world networks. Our results for k-edge sensitivity reinforces this argument.

7 EXPERIMENTS

In Section 5, we analyzed the theoretical bounds with regard to the sensitivity of various centralities. In this section, we confirm that these bounds reflect actual tendencies by conducting experiments on synthetic and real graphs to answer this question. We also experimentally show that edge sensitivity is a good proxy for k-edge sensitivity for large k when comparing various centralities in terms of sensitivity against removals of multiple edges.

7.1 Method

We performed experiments on synthetic and real-world graphs. The former were intended to clarify the tendency of sensitivity on graphs of varying size. The latter were used to determine how our sensitivity metric applies to real-world networks.

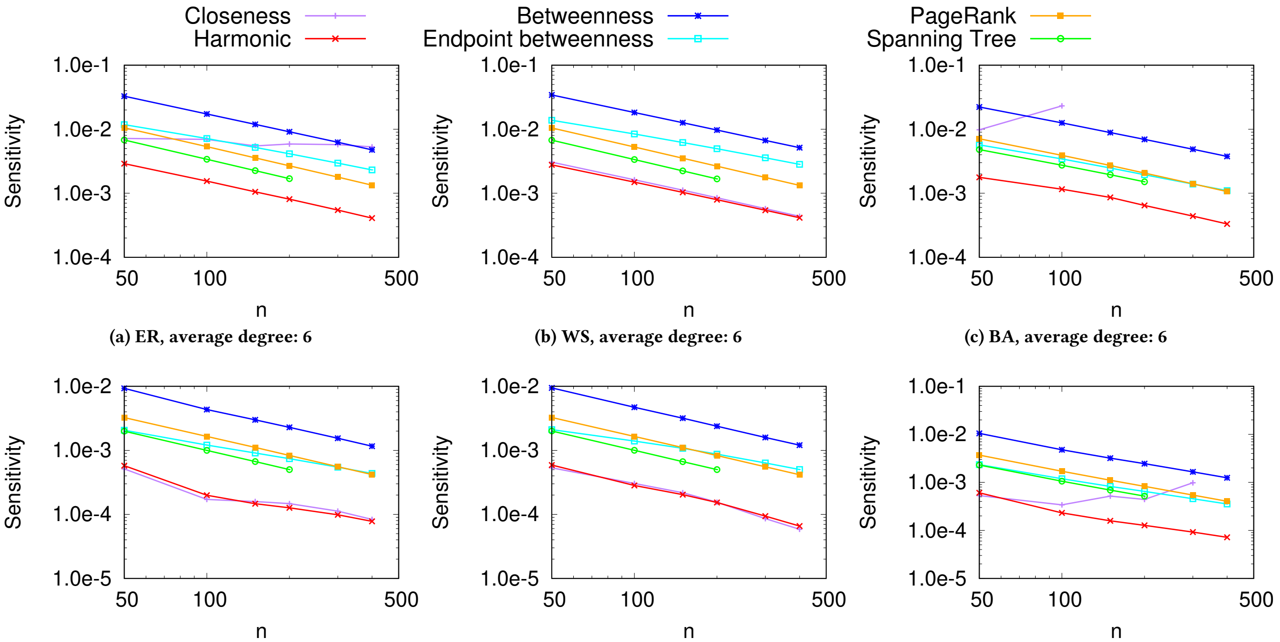
In the experiments on synthetic graphs (Section 7.2), we first generated three kinds of undirected graphs with varying parameters and scales. We used the following three methods for generating graphs:

- Erdős-Rényi (ER) for uniformly random graphs [12]. Average degree: 6, 20.
- Watts-Strogatz (WS) for small-world graphs [33] w diameter of $O(\log n)$. Average degree: 6, 20. Rewiring probability: 0.2.
- Barabasi-Albert (BA) for scale-free graphs [1, 2] with a power-law degree distribution. Core size: 20. Edge probability in the core: 0.5. Average degree of non-core vertices: 6, 20.

We tested graphs of a different number of vertices: 50, 100, 150, 200, 300 and 400. For each setting, we generated 100 graphs. Then we exactly computed the sensitivity of each network for six centralities: the closeness centrality, harmonic centrality, betweeness centrality, endpoint betweenness centrality, PageRank, and spanning tree centrality. The (endpoint) betweenness centrality was computed using the Brandes' algorithm [8] and PageRank was computed using the power iteration method. The spanning tree centrality was computed with the formula provided in [17]. This formula requires computing the pseudo-inverse of matrices. We used the Eigen library [16] for this computation. Since the spanning tree centrality is computationally expensive, we computed it only for graphs of at most 200 vertices.

In the experiments on real-world graphs (Section 7.3), we used directed and undirected graphs provided in the Stanford Large Network Dataset Collection (SNAP) [20] and KONECT [19]. Details for these graphs are shown in Table 2. For each graph, we extracted the largest strongly connected component and removed multiple edges, leaving only single and self edges; n and m in Table 2 refer to the preprocessed graph. As it is too costly to compute sensitivity values for these networks, we estimated them with random sampling. Specifically, we randomly picked 1000 edges, and for each sampled edge e, we computed the *partial sensitivity* $\frac{1}{|V|}\sum_{v\in V}\frac{|C_{G\setminus e}(v)-C_G(v)|}{C_G(v)}$. The average partial sensitivity over all sampled edges was then used as an estimation of the sensitivity value. Furthermore, we derived the standard deviation of the partial sensitivities. We computed sensitivity of spanning tree using a sample of 112 edges (instead of 1000 edges) exclusively for small graphs. For the closeness centrality of undirected graphs, the partial sensitivity will be 1 when e is a bridge. Therefore, we first computed the set of bridges $B\subseteq E$ and then took a sample of 112 edges in $E\setminus B$. The sensitivity of the closeness centrality was estimated by first computing the partial sensitivity of the sampled edges and then augmenting the contribution of bridges to the sensitivity.

Using synthetic graphs, we also computed the k-edge sensitivity for $k>1$ (Section 7.4). The settings used to generate synthetic graphs are identical to the experiments for single-edge sensitivity on synthetic graphs, expect that we fixed the number of vertices to 400. After generating graphs, we estimated the k-edge sensitivity with different ratio of removed edges ρ: 0.01, 0.02, 0.03, 0.05, 0.07, 0.1, 0.15 and 0.2 (the graphs are different among experiments for different ρ). More specifically, for a graph with m edges and the ratio of removed edges ρ, the actual number k of the removed edges



(a) ER, average degree: 6

(b) WS, average degree: 6

(c) BA, average degree: 6

(d) ER, average degree: 20

(e) WS, average degree: 20

(f) BA, average degree: 20

Figure 1: Sensitivity on synthetic graphs of different number of vertices

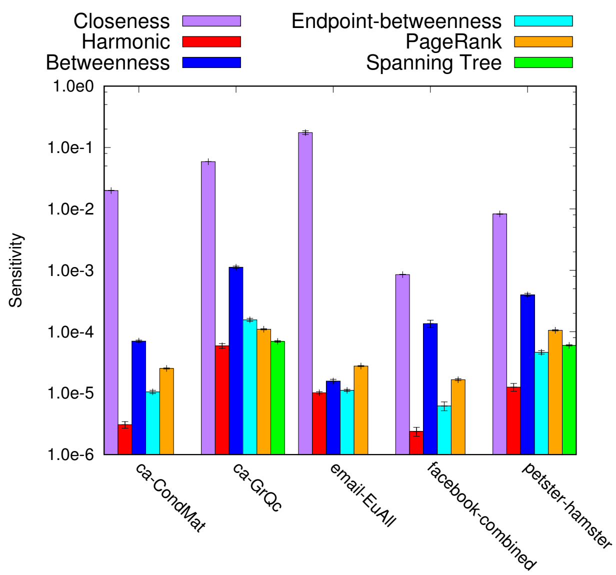


Figure 2: (Estimated) sensitivity of centralities on real-world graphs. The error bars show 1σ ranges.

is ρm, rounded to the nearest integer. We estimated the k-edge sensitivity by random sampling with 500 trials (for the spanning tree centrality, 100 trials) of edge removal. The detailed procedure for random sampling is the same as that for estimating sensitivities of real-world graphs.

All the experiments were conducted on a Windows 10 desktop with an Intel Core i9-7960X (2.8GHz) processor and 32 GiB of RAM. Experiment codes were implemented in C++11 and compiled with gcc 5.4.0 on Ubuntu 16.04 on Windows Subsystem for Linux.

7.2 Synthetic graphs

Figure 1 shows the edge sensitivities of the six centralities on synthetic graphs. Some of the generated graphs were disconnected, and we report the average over connected graphs for the closeness centrality because it was not well defined on disconnected graphs. Additionally, as explained in Section 5, the original betweenness centrality suffers from the zero-division problem. Thus, we altered the definition of the sensitivity of the betweenness centrality to

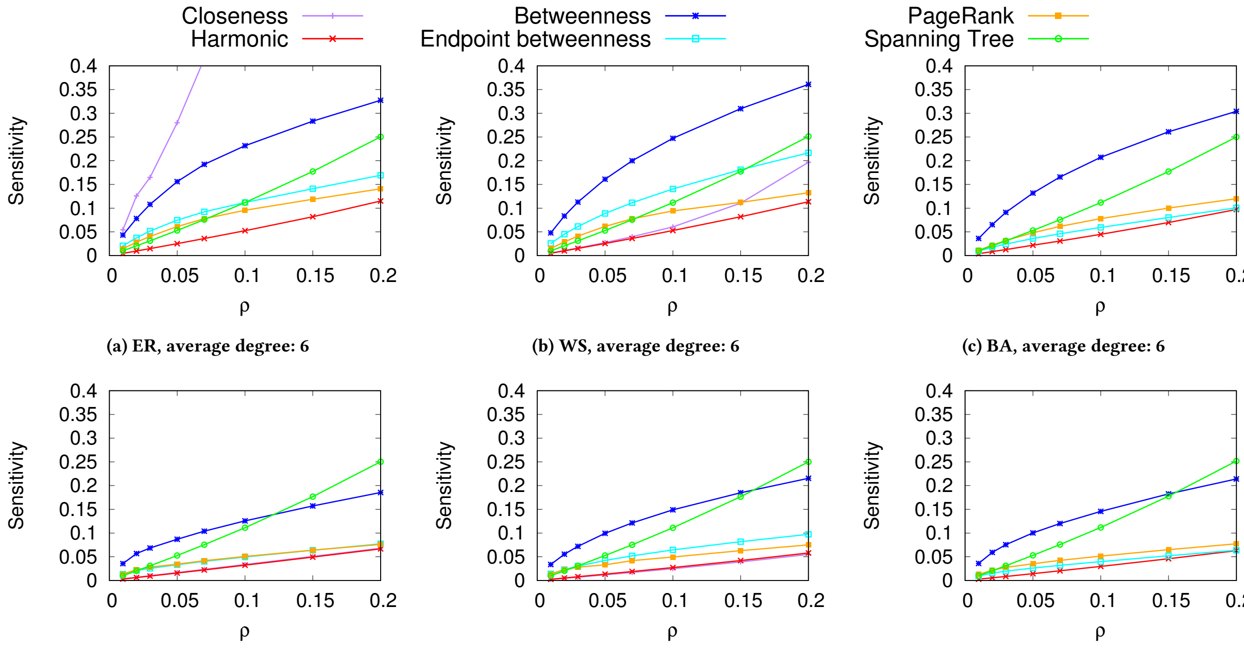
$$\frac{1}{|V||E|}\sum_{e\in E}\sum_{v\in V}\frac{\left|C_{G\setminus e}(v)-C_G(v)\right|}{\max\{C_G(v),C_{G\setminus e}\}}.$$

We let $\frac{|C_{G\setminus e}(v)-C_G(v)|}{\max\{C_G(v),C_{G\setminus e}\}}=0$ when $C_G(v)=C_{G\setminus e}(v)=0$.

We observe that the sensitivity values decrease as the number of vertices increases, except in the case of the closeness centrality in Figure 1a, 1c, and 1f. As explained in Section 5, the sensitivity of the closeness centrality depends to a considerable extent on the number of bridges in the graph. Indeed, the graphs generated by ER with an average degree 6 may have some vertices of degree 1, making the edges adjacent to them bridges. Moreover, BA generates scale-free graphs, in which the degree distribution is highly biased. Thus, it is natural for graphs generated by BA to contain many vertices of degree 1.

On the other hand, in Figure 1b, 1d and 1e, the closeness and harmonic centralities have almost the same stability values. This suggests that the difference in sensitivity between the closeness and harmonic centralities is mainly due to the presence of bridges.

Apart from the inferiority of the closeness centrality on graphs with many bridges, the closeness and harmonic centralities were among the least sensitive, followed by the endpoint betweenness



(a) ER, average degree: 6
(b) WS, average degree: 6
(c) BA, average degree: 6
(d) ER, average degree: 20
(e) WS, average degree: 20
(f) BA, average degree: 20

Figure 3: Multi-edge sensitivity on synthetic graphs with different ratio of removed edges ($n = 400$)

centrality and PageRank. This result is consistent with our theoretical analysis, in which the closeness centrality (without the presence of bridges) and the harmonic centrality have smaller sensitivities than the endpoint betweenness and PageRank.

However, we observe that the sensitivity of spanning tree centrality is consistently less than the endpoint betweenness centrality and PageRank. Although our analysis in Theorem 5.18 states $S_{\mathrm{SP}}(G) = O(\sqrt{n}/m)$, our experimental results suggest that $S_{\mathrm{SP}}(G)$ is almost proportional to $\frac{1}{m}$ on synthetic graphs. Narrowing this gap will be pursued in future work.

7.3 Real-world graphs

In Figure 2, we show approximations to the edge sensitivities of the fix centralities on real-world graphs. We see tendencies that are similar to the results with synthetic graphs: the harmonic centrality was the least sensitive, followed by the spanning-tree centrality, endpoint-betweenness centrality and PageRank, and finally the betweenness centrality. Notably, the sensitivity of the closeness centrality varied significantly depending on the graph, which may caused by different ratios of bridges.

7.4 k-edge sensitivity

In Figure 3, we show the k-edge sensitivities of the six centralities on synthetic graphs. First, we can observe that the ranking among the six centralities with respect to k-edge centrality is almost the same as that with respect to edge sensitivity.

Next, we investigate how sensitivity values change as ρ increases. For all settings and all centralities, the k-edge sensitivity is almost proportional to ρ when ρ is small (specifically, $\rho < 0.05$). For the harmonic centrality and PageRank, this result is consistent with our analysis in Section 6, as our upper bounds for those centralities are linear in k. However, there are gaps between the empirical results and theoretical analysis for the closeness centrality and the betweenness centrality, for which our upper bounds are merely $O(k^2)$. Explaining these gaps is a future work. Also, we can observe that the sensitivity of the spanning tree centrality steadily grows as ρ increases, while others except for closeness tend to slow down. Investigating its reason is also a future work.

8 CONCLUSION

In this work, to compare centralities based on their sensitivity to modifications in the graph, we introduced a quantitative measure called (average-case) k-edge sensitivity, which measures how much the centrality value of a uniformly chosen vertex (or an edge) changes when we remove a uniformly chosen set of k edges. We theoretically gave upper bounds on the k-edge sensitivities of six major centralities. For $k = 1$, we also gave almost matching lower bounds except for the spanning tree centrality. Our experimental results on synthetic and real graphs confirmed the tendency predicted by the theoretical analysis and confirmed that single edge sensitivity can be used as a good proxy for k-edge sensitivity when comparing various centralities in terms of sensitivity against removals of multiple edges.

In this work, we considered edge removal on unweighted graphs as the error model. We suppose we can extend the analysis to other settings, such as edge addition, vertex removal, or graphs with different edge lengths. Analyzing the sensitivity of centralities on such settings will be a subject of future work.

ACKNOWLEDGMENTS

S. Murai would like to thank Hiroshi Imai for giving useful comments and advices.

REFERENCES

[1] Edoardo M Airoldi and Kathleen M Carley. 2005. Sampling algorithms for pure network topologies: a study on the stability and the separability of metric embeddings. *ACM SIGKDD Explorations Newsletter* 7, 2 (2005), 13–22.

[2] Réka Albert and Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of modern physics* 74, 1 (2002), 47.

[3] Konstantin Avrachenkov, Nelly Litvak, Danil Nemirovsky, and Natalia Osipova. 2007. Monte Carlo methods in PageRank computation: When one iteration is sufficient. *SIAM J. Numer. Anal.* 45, 2 (2007), 890–904.

[4] Alex Bavelas. 1950. Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America* 22, 6 (1950), 725–730.

[5] Murray A. Beauchamp. 1965. An improved index of centrality. *Behavioral Science* 10, 2 (1965), 161–163.

[6] Paolo Boldi and Sebastiano Vigna. 2014. Axioms for centrality. *Internet Mathematics* 10, 3-4 (2014), 222–262.

[7] Stephen P Borgatti, Kathleen M Carley, and David Krackhardt. 2006. On the robustness of centrality measures under conditions of imperfect data. *Social networks* 28, 2 (2006), 124–136.

[8] Ulrik Brandes. 2001. A faster algorithm for betweenness centrality. *Journal of mathematical sociology* 25, 2 (2001), 163–177.

[9] Ulrik Brandes. 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* 30, 2 (2008), 136–145.

[10] Elizabeth Costenbader and Thomas W Valente. 2003. The stability of centrality measures when networks are sampled. *Social networks* 25, 4 (2003), 283–307.

[11] Peter G Doyle and J Laurie Snell. 1984. Random walks and electric networks. (1984).

[12] P. Erdős and A. Rényi. 1959. On random graphs. I. *Publ. Math. Debrecen* 6 (1959), 290–297.

[13] Terrill L Frantz, Marcelo Cataldo, and Kathleen M Carley. 2009. Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory* 15, 4 (2009), 303.

[14] Linton C Freeman. 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry* 40, 1 (1977), 35–41.

[15] Gourab Ghoshal and Albert-László Barabási. 2011. Ranking stability and super-stable nodes in complex networks. *Nature communications* 2 (2011), 394.

[16] Gaël Guennebaud, Benoît Jacob, et al. 2010. Eigen v3. http://eigen.tuxfamily.org.

[17] Douglas J Klein and Milan Randić. 1993. Resistance distance. *Journal of mathematical chemistry* 12, 1 (1993), 81–95.

[18] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.

[19] Jérôme Kunegis. 2013. KONECT - The Koblenz Network Collection. In *Proc. Int. Web Observatory Workshop*. 1343–1350.

[20] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[21] Massimo Marchiori and Vito Latora. 2000. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications* 285, 3-4 (2000), 539–546.

[22] Charalampos Mavroforakis, Richard Garcia-Lebron, Ioannis Koutis, and Evimaria Terzi. 2015. Spanning edge centrality: Large-scale computation and applications. In *Proceedings of the 24th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 732–742.

[23] Andrew Y Ng, Alice X Zheng, and Michael I Jordan. 2001. Link analysis, eigenvectors and stability. In *International Joint Conference on Artificial Intelligence*, Vol. 17. Lawrence Erlbaum Associates Ltd, 903–910.

[24] Qikai Niu, An Zeng, Ying Fan, and Zengru Di. 2015. Robustness of centrality measures against network manipulation. *Physica A: Statistical Mechanics and its Applications* 438 (2015), 124–131.

[25] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report. Stanford InfoLab.

[26] John Platig, Edward Ott, and Michelle Girvan. 2013. Robustness of network measures to link errors. *Physical Review E* 88, 6 (2013), 062812.

[27] Gert Sabidussi. 1966. The centrality index of a graph. *Psychometrika* 31, 4 (1966), 581–603. https://doi.org/10.1007/BF02289527

[28] A. A. Schoone, H. L. Bodlaender, and J. van Leeuwen. 1987. Improved diameter bounds for altered graphs. In *Graph-Theoretic Concepts in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 227–236.

[29] S. Segarra and A. Ribeiro. 2016. Stability and Continuity of Centrality Measures in Weighted Graphs. *IEEE Transactions on Signal Processing* 64, 3 (Feb 2016), 543–555.

[30] Andreia Sofia Teixeira, Pedro T Monteiro, João A Carriço, Mário Ramirez, and Alexandre P Francisco. 2013. Spanning edge betweenness. In *Workshop on mining and learning with graphs*, Vol. 24. 27–31.

[31] Sho Tsugawa and Hiroyuki Ohsaki. 2015. Analysis of the robustness of degree centrality against random errors in graphs. In *Complex Networks VI*. Springer, 25–36.

[32] Vladimir Ufimtsev, Soumya Sarkar, Animesh Mukherjee, and Sanjukta Bhowmick. 2016. Understanding stability of noisy networks through centrality measures and local connections. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2347–2352.

[33] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world' networks. *nature* 393, 6684 (1998), 440.

[34] Barbara Zemljič and Valentina Hlebec. 2005. Reliability of measures of centrality and prominence. *Social Networks* 27, 1 (2005), 73–88.