Random Walk Decay Centrality

Tomasz Wąs,[1] Talal Rahwan,[2] Oskar Skibski[1]
[1]Institute of Informatics, University of Warsaw, Poland
[2]Computer Science, New York University, Abu Dhabi, UAE
{t.was, o.skibski}@mimuw.edu.pl, talal.rahwan@nyu.edu

Abstract

We propose a new centrality measure, called the *Random Walk Decay centrality*. While most centralities in the literature are based on the notion of shortest paths, this new centrality measure stems from the random walk on the network. We provide an axiomatic characterization and show that the new centrality is closely related to PageRank. More in detail, we show that replacing only one axiom, called *Lack of Self-Impact*, with another one, called *Edge Swap*, results in the new axiomatization of PageRank. Finally, we argue that *Lack of Self-Impact* is desirable in various settings and explain why violating *Edge Swap* may be beneficial and may contribute to promoting diversity in the centrality measure.

Introduction

Centrality measures—methods for identifying the most important nodes in a network based solely on its topology—have been extensively studied in the literature on graph theory and network analysis for over 50 years (Newman 2010). Fueled by the ever-growing availability of relational data, as well as the access to unprecedented computational power, centrality measures have become an essential part of every network analysis toolkit over the past two decades (Freeman 2008). These measures are increasingly being applied in numerous subareas of computer science, including the world-wide web (Page et al. 1999), viral marketing (Hinz et al. 2011), and energy saving in communication networks (Bianzino et al. 2011), just to name a few.

Most standard centrality measures are based on the notion of shortest paths (Freeman 1979). One of the most fundamental such measures is *Closeness centrality*, whereby the importance of a node is determined based on the inverse of the sum of distances from all other nodes in the graph (Sabidussi 1966). Since this centrality measure is well-defined only for strongly connected graphs, Jackson (2008) proposed an alternative, named *Decay centrality*, which works also for disconnected graphs.

These standard centrality measures are based on two simplifying assumptions. Firstly, they assume that the nodes of the network do not have any weights—an assumption that does not always hold in practical applications. Secondly, they assume that information always travels in a network through the fastest route(s)—an assumption that requires all nodes to know the entire topology of the network, which is rarely the case in real-world networks. In fact, whether one is modeling the spread of gossip through a social group, the propagation of viruses through a computer network, or the way users surf the Internet, such processes tend to spread more chaotically in practice, e.g., through somewhat random paths as opposed to shortest paths (Lerman and Ghosh 2010; Borgatti 2005; Huberman et al. 1998).

To address these limitations, a number of centrality measures have been developed based on *random walks* in the graph. According to this approach, the source of information is a node chosen randomly according to a given distribution of node weights, and the information then propagates through the network by moving along random outgoing edges (Lovász 1993). Perhaps the first and most influential such centrality measure is *PageRank* (Page et al. 1999). Its success has lead to the development of random-walk versions of various centrality measures, such as *Random Walk Closeness* centrality (White and Smyth 2003) and *Random Walk Betweenness* centrality (Newman 2005).

Against this background, in this paper we propose a new centrality measure, called *Random Walk Decay*, which is a random-walk version of Decay centrality. To highlight the similarities and differences between both centralities, we use an axiomatic approach. Specifically, we show that Random Walk Decay centrality and Decay centrality both satisfy five basic properties (called *axioms*): *Locality*, *Sink Merging*, *Directed Leaf Proportionality*, *One-Node Graph* and *Lack of Self-Impact*. In addition to those five axioms, if the centrality measure is based on shortest paths (i.e., satisfies the *Shortest Paths Property*), we obtain Decay centrality; on the other hand, if it is based on random walks (i.e., satisfies the *Random Walk Property*), we obtain Random Walk Decay.

Furthermore, using the axiomatic approach, we compare the Random Walk Decay centrality to PageRank. We show that replacing only one axiom (*Lack of Self-Impact*) with another one (*Edge Swap*) gives us a new axiomatization of PageRank. The former axiom, Lack of Self-Impact, states that “removing your own outgoing edges does not affect your centrality”; thus, it corresponds to a certain strategy-proof property satisfied by Random Walk Decay centrality. The latter axiom, Edge Swap, states that “if two nodes have the same centralities and the same number of outgoing edges, then we can swap one of their edges without affecting their centralities"; thus, violating this axiom allows Random Walk Decay centrality to promote diversity of incoming edges. In result, we argue that Random Walk Decay centrality has properties violated by PageRank which can be found desirable in various setting.

Preliminaries

In this section, we introduce basic notations and definitions.

Graphs: In this paper, we consider directed multigraphs.[1] A *(multi)graph* is an ordered pair, $G = (V, E)$, where V is the set of nodes and $E \sqsubseteq V \times V$ is the multiset of directed edges. We will use $\sqcup$ and $-$ to denote multiset union and difference, respectively. For graph G, the sets of nodes and edges are denoted by $V(G)$ and $E(G)$, respectively.

Furthermore, we associate with each node v a nonnegative weight, denoted by $\beta(v)$. For $v \in V(G)$ by δ_v we denote the particular vector of node weights β such that $\delta_v(v) = 1$ and $\delta_v(u) = 0$ for every $u \in V(G) \setminus \{v\}$. The sum of weights of all nodes in a graph G will be denoted by $\beta(G)$.

An edge (u, v) is an outgoing edge for the node u and an incoming edge for node v. If $u = v$, this edge is called a *loop*. The multiset of outgoing edges for v is denoted by $\Gamma^+_G(v)$. Analogously, the multiset of incoming edges for v is denoted by $\Gamma^-_G(v)$. Moreover, we define $\Gamma^\pm_G(v) = \Gamma^+_G(v) \sqcup \Gamma^-_G(v)$. A node without outgoing edges is called a *sink*. A node without incoming and outgoing edges is *isolated*.

A *walk*, $p = (v_1, \ldots, v_k)$ is an ordered sequence of nodes such that $(v_i, v_{i+1}) \in E$ for every $i \in \{1, \ldots, k-1\}$. A (simple) *path* is a walk in which all nodes (except possibly the first and the last one) are distinct. A (simple) *cycle* is a path such that $v_1 = v_k$. The length of a *walk* is the length of the sequence minus one. If there exists a walk that starts in u and ends in v, then u is called a *predecessor* of node v. The set of all predecessors of node v is denoted by $P_G(v)$.

A graph is *strongly connected* if there exists at least one path between any two nodes. A graph is *(weakly) connected* if there exists at least one path between any two nodes if we treat it as an undirected graph.

The graph obtained from G with node weights β by merging node u with node v is denoted by $M_{u\to v}(G, \beta)$. Formally:

$$M_{u\to v}(G, \beta) = (V(G) \setminus \{u\}, E(G) - \Gamma^\pm_G(u) \sqcup E', \beta'),$$

where $E' = \bigsqcup_{(u,w)\in E(G)}\{(v, w)\} \cup \bigsqcup_{(w,u)\in E(G)}\{(w, v)\}$ and $\beta'(v) = \beta(v) + \beta(u)$, and $\beta'(w) = \beta(w)$ for $w \in V \setminus \{u, v\}$. Also, for graph G and multiset of edges E', we define $G + E' = (V(G), E(G) \sqcup E')$ and $G - E' = (V(G), E(G) - E')$.

We say that two graphs, G, G', overlap on S if $V(G) \cap V(G') = S$. If $V(G) \cap V(G') = \emptyset$, then G and G' do not overlap, and are said to be *disjoint*. The sum of two graphs along with their node weights is defined as:

$$(G, \beta) + (G', \beta') = ((V(G) \cup V(G'), E(G) \sqcup E(G')), \beta''),$$

where $\beta''(v) = \beta(v) + \beta'(v)$ for $v \in V(G) \cap V(G')$, $\beta''(v) = \beta(v)$ for $v \in V(G) \setminus V(G')$ and $\beta''(v) = \beta'(v)$ for $v \in V(G') \setminus V(G)$.

Centrality measures: A *centrality measure*, F, is a function that assigns to every node, v, in every graph, G, a real value reflecting the importance of v in G.

Freeman (1979) in his seminal work identified three centrality measures that capture different aspects of a node in the graph. The most basic one, called the *Degree centrality*, assesses a node by the number of its edges. For directed graphs, both In- and Out-Degree centralities are considered. The other two centrality measures focus on the shortest paths in the graph. Specifically, the *Betweenness centrality* evaluates a node, v, based on the proportion of shortest paths (between any two other nodes) to which v belongs. In contrast, the *Closeness centrality*, originally proposed by Sabidussi (1966), identifies the nodes that are closest to all other nodes, and that is by computing the inverse of the sum of distances to other nodes in the graph:

$$C_v(G) = \frac{1}{\sum_{u\in V(G)\setminus\{v\}} dist(u, v)},$$

where $dist(u, v)$ is the distance from u to v defined as the length of a shortest path from u to v.

The Closeness centrality is well-defined only for strongly connected graphs. To address this shortcoming, Jackson (2008) proposed an alternative, called *Decay centrality*:

$$Y_v(G) = \sum_{u\in V(G)\setminus\{v\}} a^{dist(u,v)},$$

for a *decay parameter* $a \in (0, 1)$. Here, if we treat a as the probability of a successful move from one node to another via an edge, then Decay centrality can be interpreted as the expected number of nodes that can reach v via shortest paths.

Personalized centrality measures: Most standard centrality measures were proposed for graphs without weights of nodes. However, they can usually be easily adapted to this richer setting (Koschützki et al. 2005). In this context, we define the *personalized Decay centrality* as follows:

$$Y_v(G, \beta) = \sum_{u\in V(G)} \beta(u) \cdot a^{dist(u,v)}. \qquad (1)$$

The personalized Decay centrality introduces two modifications to the original definition. Firstly, the contribution of a node u to the centrality of v (i.e., $a^{dist(u,v)}$) is now multiplied by the weight of u. Secondly, we now sum over all nodes ($\sum_{u\in V}$), rather than over all nodes other than v ($\sum_{u\in V\setminus\{v\}}$). To understand the rationale behind this latter modification, consider an extreme scenario in which only a single node, say v, has a positive weight. Here, if we sum over all nodes other than v, then any node with a connection to v would have a positive centrality, whereas v itself would have a centrality equal to zero, as all nodes not connected to v—a rather unintuitive outcome in most interpretations of node weights.

An important personalized centrality measure is *PageRank* (Page et al. 1999). This measure is defined by the following recursive formula:

$$PR_v(G, \beta) = a \cdot \left(\sum_{(u,v)\in\Gamma^-_G(v)} \frac{PR_u(G, \beta)}{|\Gamma^+_G(u)|} \right) + \beta(v), \qquad (2)$$

[1]Multigraphs can be interpreted as edge-weighted graphs where weights of edges are natural numbers. The results of this paper easily translate to arbitrary edge-weighted graphs. However, for clarity of presentation, we limit ourselves to multigraphs.

for a parameter $a \in (0, 1)$. Since we assume multiple edges between two nodes, then node u may appear multiple times on the right-hand side of the equation. It has been proven that, for a fixed a, this formula uniquely characterizes a centrality measure (see, e.g., Bianchini, Gori, and Scarselli 2005).

Random Walk Decay Centrality

In this paper, we propose a new centrality measure that is based on the notion of a *random walk* on a graph. The *random walk* is defined in the following way (Lovász 1993):

- at the beginning (at moment $t = 0$), we choose one node according to the distribution of node weights;
- in the k-th step (at moment $t = k$, for $k \geq 1$), while in node u, we choose one of the outgoing edges of u, say $(u, v) \in \Gamma^+_G(u)$, uniformly at random, and move along this edge to node v.

Formally, the random walk is a sequence of random nodes $w = (w(0), w(1), \dots)$ that is a Markov chain, defined through its initial distribution, i.e.,

$$\mathbb{P}_{G,\beta}(w(0) = v) = \beta(v)/\beta(G),$$

and a transition matrix $M = (p_G(u, v))_{u,v \in V}$ where the probability of moving from node u to node v, denoted by $p_G(u, v)$, is the number of edges from u to v divided by the number of all outgoing edges from u:

$$p_G(u, v) = |\{(u, v) \in \Gamma^+_G(u)\}|/|\Gamma^+_G(u)|. \quad (3)$$

To deal with the fact that sinks would break the infinite walk, we assume that—besides the nodes of the graph—there exists one additional "terminal" absorbing state e to which we move from all sinks in the graph. Formally, we have $p_G(u, e) = 1$ if u is a sink, $p_G(u, e) = 0$ otherwise, and $p_G(e, e) = 1$. In result, we can think of the random walk as the set of all possible infinite walks on the graph, each associated with its probability.

Example 1. *Consider the random walk on the graph from Figure 1. The random walk starts in node u or node w, because these are the only nodes with non-zero weights. From node u, the walk moves to v with probability* $2/3$*, and stays in u with probability* $1/3$*. From node w, the walk moves either to v or to t, both with probability* $1/2$*. From node v, the walk always moves to w. Node t is a sink, so from t the walk moves to the absorbing state e and loops therein. Consequently, the probabilities of the different combinations of the first four nodes in the walk are as follows:*

$1/54$	$(u, u, u, u, \dots)$	$1/8$	$(w, v, w, v, \dots)$
$1/27$	$(u, u, u, v, \dots)$	$1/8$	$(w, v, w, t, \dots)$
$1/9$	$(u, u, v, w, \dots)$	$1/4$	$(w, t, e, e, \dots)$
$1/6$	$(u, v, w, v, \dots)$		
$1/6$	$(u, v, w, t \dots)$		

Let us introduce some additional terminology. For node $v \in V(G)$, we will consider the probability that it will be visited for the k-th time in moment t. We will call it k-th *visiting probability* in moment t and denote it by $VP^v_{G,\beta}(t, k)$:

$$VP^v_{G,\beta}(t, k) = \mathbb{P}_{G,\beta}(w(t) = v, |\{s \leq t : w(s) = v\}| = k). \quad (4)$$

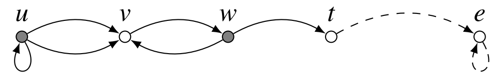

Figure 1: A sample graph. The weight of every grey nodes is 1, while the weight of every white node is 0.

Now, we say that a centrality measure is a *random walk centrality* if it depends solely on the node's *visiting probabilities*.

Random Walk Property (RWP): *For every two graphs G, G' with node weights β, β' such that $\beta(G) = \beta'(G')$ and node $v \in V(G) \cap V(G')$, if $VP^v_{G,\beta}(t, k) = VP^v_{G',\beta'}(t, k)$ for every $t, k \in \mathbb{N}$, then $F_v(G, \beta) = F_v(G', \beta')$.*

Example 2. *Consider again the random walk on the graph from Figure 1. Here,* $VP^v_{G,\beta}(t, k)$ *equals:*

$k \setminus t$	0	1	2	3	4	5	6
1	0	7/12	1/9	1/27	1/81	1/243	1/729
2	0	0	0	7/24	1/18	1/54	1/162
3	0	0	0	0	0	7/48	1/36

In more detail, we clearly have $VP^v_{G,\beta}(t, k) = 0$ *if* $t < k$. *Furthermore,* $VP^v_{G,\beta}(1, 1) = 1/2 \cdot 2/3 + 1/2 \cdot 1/2$ *which corresponds to walks* $(u, v, \dots)$ *and* $(w, v, \dots)$. *Additionally,* $VP^v_{G,\beta}(t, 1) = 1/2 \cdot 1/3^{t-1} \cdot 2/3$ *for* $t > 1$ *because we have to start at u and loop for* $t - 1$ *times there in order to enter node v for the first time at moment* $t > 1$. *Finally, we have* $VP^v_{G,\beta}(t, k) = 1/2 \cdot VP^v_{G,\beta}(t - 2, k - 1)$ *for* $t, k > 1$ *because we may only return back to v in two steps, which happens with probability* $1/2$.

Random walk centrality measures evaluate the nodes in a given graph by analyzing different properties of the random walk. For instance, *Random Walk Closeness centrality* is defined as the inverse of the expected time needed for the random walk to reach a specific node for the first time (White and Smyth 2003). Formally, for every graph G and every node $v \in V(G)$:

$$RWC_v(G, \beta) = 1/\left(\sum_{t=0}^{\infty} t \cdot VP^v_{G,\beta}(t, 1)\right). \quad (5)$$

The Random Walk Closeness centrality suffers for several problems of its original—the Closeness centrality. In particular, if there exists a node with non-zero weight from which v cannot be reached, then the centrality of v equals zero.

In this paper, we propose the following centrality measure, which is a translation of the Decay centrality (Jackson 2008) to the random walk model.

Definition 1. Random Walk Decay centrality *is a centrality measure defined for every graph, G, and every node, $v \in V(G)$, as:*

$$RWD_v(G, \beta) = \beta(G) \cdot \sum_{t=0}^{\infty} a^t \cdot VP^v_{G,\beta}(t, 1), \quad (6)$$

for a decay parameter $a \in (0, 1)$.

Let us explain this formula. For node v and moment t, $VP^v_{G,\beta}(t,1)$ is the probability that node v will be reached by the random walk at moment t for the first time. If we assume, as in the interpretation of the Decay centrality, that each step succeeds with probability a, then $a^t \cdot VP^v_{G,\beta}(t,1)$ is the probability of reaching the node v successfully. This expression is summed over all possible moments, $t \geq 0$. Finally, the whole expression is multiplied by $\beta(G)$. This is because each node u, is a starting point with probability $\beta(u)/\beta(G)$. Thus, by multiplying the sum by $\beta(G)$, the random walk that starts from node u is considered with the weight $\beta(u)$, just as in the personalized Decay centrality (Eq. (1)). To put it differently, the Random Walk Decay measures the probability of reaching a node by the random walk assuming a constant probability, $(1-a)$, of breaking the walk.

Example 3. *Let us compute the Random Walk Decay centrality for node v in the graph from Figure 1. Based on Example 2 we get:*

$$RWD_v(G,\beta) = 2\left(\frac{7a}{12} + \frac{a^2}{9} + \frac{a^3}{27} + \dots\right) = \frac{21a - 7a^2 + 18}{6(3-a)}.$$

For $a = 1/2$ we get $RWD_v(G,\beta) = 214/105$. Similar calculations show that $RWD_u(G,\beta) = 1/2$, $RWD_w(G,\beta) = 3/5$ and $RWD_t(G,\beta) = 3/10$.

We end this section by discussing PageRank. Recall that Page et al. (1999) proposed PageRank along with a random walk interpretation named *the random surfer model*. The Markov chain defined by the random surfer model differs from that of the random walk. More in details, in the random surfer model, at each step the *surfer* stops moving along edges with some probability, and instead jumps to a randomly selected node. Nevertheless, in the following proposition we show that PageRank also satisfies the Random Walk Property (RWP), i.e., it can be expressed in terms of the (standard) random walk on a graph.

Proposition 1. *PageRank is equal to*

$$PR_v(G,\beta) = \beta(G) \cdot \sum_{t=0}^{\infty} \left(a^t \cdot \sum_{k=1}^{\infty} VP^v_{G,\beta}(t,k) \right). \quad (7)$$

In result, PageRank satisfies Random Walk Property (RWP).

Proof. It suffices to prove that the centrality measure defined in (7) satisfies the recursive formula from (2). By considering the nodes from which the random walk can move to node v, we get:

$$\sum_{k=1}^{\infty} VP^v_{G,\beta}(t,k) = \sum_{u \in V} \sum_{k=1}^{\infty} p_G(u,v) \cdot VP^u_{G,\beta}(t-1,k), \quad \text{for } t > 0,$$

and $\sum_{k=1}^{\infty} VP^v_{G,\beta}(t,k) = \beta(v)/\beta(G)$ for $t = 0$. Combining it with Eq. (7) leads to (2). □

Formula (7) is similar to the formula for the Random Walk Decay. In a nutshell, the Random Walk Decay takes into account only the first time the random walk visits a node, while PageRank takes into account also further times. Thus, PageRank measures the expected number of times a node is reached by the random walk assuming a constant probability, $(1-a)$, of breaking the walk.

Axiomatic Characterization

In this section, we axiomatically characterize our new centrality measure—the Random Walk Decay centrality. The characterization is built in a close relation to the Decay centrality and PageRank.

We begin with axioms satisfied by all three centralities—the Decay centrality, the Random Walk Decay centrality and PageRank. Our first axiom, *Locality*, states that the centrality of a node depends solely on nodes connected to it. To put it differently, the centrality of a node does not change if we add to the graph a second, disjoint graph.

Locality (LOC): *For every two disjoint graphs G, G', node weights β, β' and node $v \in V(G)$*

$$F_v((G,\beta) + (G',\beta')) = F_v(G,\beta).$$

This basic axiom was proposed for graphs without weights by Skibski et al. (2016).

Our second axiom is called *Sink Merging*. It states that if we merge two nodes without outgoing edges, i.e., sinks, without joint predecessors, then the centrality of the resulting node will be the sum of the centralities of both sinks; moreover, the centralities of other nodes will not change.

Sink Merging (SM): *For every graph G, node weights β and sinks $u, v \in V(G)$ such that $P_G(v) \cap P_G(u) = \emptyset$*

$$F_v(M_{u \to v}(G,\beta)) = F_u(G,\beta) + F_v(G,\beta),$$

and $F_w(M_{u \to v}(G,\beta)) = F_w(G,\beta)$ for any $w \in V(G) \setminus \{u,v\}$.

This axiom is a much weaker version of *Merging*, proposed for PageRank by Wąs and Skibski (2018), that considered merging arbitrary nodes, possibly with outgoing edges.

The third axiom is *Directed Leaf Proportionality*, which requires that, if we add an edge from a sink u to an isolated node v, then the gain in the centrality of v is proportional to the centrality of u.

Directed Leaf Proportionality (DLP): *There exists a constant, $a > 0$, such that for every graph G, node weights β, sink $u \in V(G)$ and isolated node $v \in V(G)$:*

$$F_v(G + \{(u,v)\},\beta) - F_v(G,\beta) = a \cdot F_u(G,\beta).$$

This axiom is a directed and weighted version of Leaf Proportionality, proposed by Skibski and Sosnowska (2018).

Our fourth axiom, *One-Node Graph*, is a simple normalization property: if there is only one node in the graph and its weight equals 1, then its centrality also equals 1.

One-Node Graph (1-NG): *For every node v*

$$F_v((\{v\},\emptyset),\delta_v) = 1.$$

We note that without 1-NG, the remaining axioms implies that centrality measure is unique up to a scalar multiplication.

Now, let us introduce the next axiom, *Lack of Self-Impact*, which is one of two axioms that distinguishes the Random Walk Decay centrality and PageRank. This axiom states that the centrality of a node does not depend on its outgoing edges. In the next section, we show that this axiom can be considered as a strategy-proofness condition if nodes were to choose their own outgoing edges.

Lack of Self-Impact (LSI): *For every graph G, node weights β and $(v,u) \in E(G)$*
$$F_v(G,\beta) = F_v(G - \{(v,u)\},\beta).$$

In the following theorem, we show that the Random Walk Decay centrality is the only centrality measure that satisfies the Random Walk Property and the above five axioms: Locality, Sink Merging, Directed Leaf Proportionality, One-Node Graph and Lack of Self-Impact.

Theorem 2. *The Random Walk Decay centrality is the unique centrality measure that satisfies LOC, SM, DLP, 1-NG, LSI and RWP.*

Proof (Sketch). Due to space restrictions, we present only the sketches of the proofs in the paper.

The main idea of the proof relies on the class of broken cactus graphs. A graph G is called a *(directed) cactus* if it is strongly connected and each edge is part of exactly one cycle (Palbom 2005). A graph G is called a *broken cactus* if there exist two nodes, s, t, called *start* and *end* nodes, such that t is a sink and $M_{t\to s}(G,\beta)$ is a cactus graph.

Assume that a centrality measure F satisfies LOC, SM, DLP, 1-NG, LSI and RWP. We begin our proof by showing that if the graph is a broken cactus such that only its start has non-zero weight, then the centrality F of its end equals the Random Walk Decay centrality (up to scalar multiplication).

Claim 1: *If centrality measure F satisfies LOC, SM, DLP and RWP, then there exists $\alpha \geq 0$ such that for every broken cactus G that begins in s and ends in t it holds that $F_t(G,\delta_s) = \alpha RWD_t(G,\delta_s)$.*

This claim is proved by observing that every broken cactus can be obtained from a path by adding cactus graphs that overlap with the path on a single node.

Next, we show that for every sink we can construct a collection of broken cactus graphs such that the weighted sum of visiting probabilities of the ends of these graphs equals the visiting probability of a sink in the original graph.

Claim 2: *For every graph G, node weights β and sink $v \in V(G)$, there exists a collection of broken cactus graphs $G_1,\dots,G_n$, that start in $s_1,\dots,s_n$, respectively, end in v and pairwise overlap on $\{v\}$ such that*
$$VP^v_{G,\beta}(t,k) = \sum_{i=1}^{n} c_i \cdot VP^v_{G_i,\delta_{s_i}}(t,k)$$
for some constants $c_1,\dots,c_n \geq 0$, for every $t \geq 0$ and $k \geq 1$.

We prove this claim by induction on the number of incoming edges of nodes that are not sinks and considering graphs with only one node having a non-zero weight.

Note that **Claim 2** is a general property of the random walk and visiting probabilities, and does not depend on the axiom nor the centrality measure definitions. However, by combining it with RWP we get that the centrality of a sink can be determined from the centralities of the ends of broken cactus graphs. Hence, **Claim 1** and **Claim 2** imply that the centrality of a sink is equal to the Random Walk Decay centrality (up to scalar multiplication).

Claim 3: *If a centrality measure F satisfies LOC, SM, DLP and RWP, then there exists $\alpha \geq 0$ such that for every graph G, node weights β and sink $v \in V(G)$ it holds that $F_v(G,\beta) = \alpha RWD_v(G,\beta)$.*

Now, consider an arbitrary graph, G, and a node, $v \in V(G)$. From LSI we know that the centrality of node v in graph G is the same as in the graph $G - \Gamma^+_G(v)$ obtained by removing all outgoing edges of v. In the latter graph, node v is a sink, hence from **Claim 3** and LSI we know that there exists a constant α such that $F_v(G,\beta) = F_v(G - \Gamma^+_G(v)) = \alpha RWD_v(G,\beta)$. Finally, 1-NG implies that $\alpha = 1$; this concludes the proof of Theorem 2. ◻

The personalized Decay centrality based on the shortest paths also satisfies the five axioms stated above—LOC, SM, DLP, 1-NG and LSI—but violates the Random Walk Property. To obtain a unique axiomatic characterization, we introduce the *Shortest Paths Property*—an axiom that captures the fact that the centrality is based on distance, i.e., shortest paths, from other nodes in the graph. Our axiom is a direct translation of the definition of the class of *distance based centralities* by Skibski and Sosnowska (2018) to weighted and directed graphs.

Shortest Paths Property (SPP): *For every two graphs G, G', node weights β, β' such that $\beta(G) = \beta'(G')$ and node $v \in V(G) \cap V(G')$, if*
$$|\{u \in V(G) : dist(u,v) = k, \beta(u) = \alpha\}| = |\{u \in V(G') : dist(u,v) = k, \beta'(u) = \alpha\}|,$$
for every $k \in \mathbb{N}$ and $\alpha \in \mathbb{R}$, then $F_v(G,\beta) = F_v(G',\beta')$.

The following theorem shows that replacing the Random Walk Property with the Shortest Paths Property in the axiomatization of the Random Walk Decay centrality results in an axiomatization of the personalized Decay centrality. It is easy to observe that Lack of Self-Impact is implied by the Shortest Paths Property, so we omit the former axiom from the axiomatic characterization.

Theorem 3. *The personalised Decay centrality is the unique centrality measure that satisfies LOC, SM, DLP, 1-NG and SPP.*

Proof (Sketch). We will use induction on the number of edges in graph G. Based on LOC, it suffices to consider only connected graphs—if graph is not connected, then the centrality of every node is the same as in a connected graph with the same or less number of edges.

If a connected graph, G, has no edges, then it must have only one node, and from 1-NG, LOC and SM it can be shown that $F_v(G,\beta) = \beta(v) = Y_v(G,\beta)$. Now assume that G has at least one edge and for every graph G' with less edges it holds that $F_v(G',\beta) = Y_v(G',\beta)$ for every $v \in V(G')$ and weights β. Fix node $v \in V(G)$. We will show that the centrality of v in G can be computed based on centralities in graphs with a smaller number of edges; hence it is unique.

First, observe that if there exists a node, u, with more than one outgoing edge, then at least one of these edges, say (u,w), can be removed without changing the distance from u

	LOC	SM	DLP	1-NG	SPP	RWP	LSI	ES
Decay	+	+	+	+	+	−	+	−
RW-Decay	**+**	**+**	**+**	**+**	**−**	**+**	**+**	**−**
PageRank	+	+	+	+	−	+	−	+

Table 1: Summary of our axiomatic characterizations. The plus sign (+) indicates that the centrality measure satisfies the axiom, whereas the minus sign (-) indicates that the centrality measure violates it.

to v; hence, from SSP we get $F_v(G,\beta) = F_v(G - \{(u,w)\},\beta)$. Analogously, if v has any outgoing edge, (v,w), then from SSP deleting it does not affect the centrality of v and we get: $F_v(G,\beta) = F_v(G - \{(v,w)\},\beta)$.

It remains to consider a connected graph in which (1) every node has at most one outgoing edge; (2) v is a sink. Since G is connected, v must have at least one incoming edge. We consider two cases separately:

- If v has exactly one incoming edge, (u,v), then from LOC, DLP and the fact that u is a sink in graph $G - \{(u,v)\}$ we get that $F_v(G,\beta) = a \cdot F_u(G - \{(u,v)\},\beta) + \beta(v)$.
- If v has at least $k \geq 2$ incoming edges, then—since every node has at most one outgoing edge—these edges must be incident to k different nodes. In such a case, graph G can be decomposed into k graphs, $(G_1,\beta_1),\dots,(G_k,\beta_k)$, that overlap only on node v, each with exactly one incoming edge to v. These graphs have fewer edges than G, and in all these graphs, v is a sink. Thus, using SM we get that $F_v(G,\beta) = F_v(G_1,\beta_1) + \dots + F_v(G_k,\beta_k)$.

This concludes the proof of Theorem 3. □

Finally, let us introduce our last axiom, called *Edge Swap*. This axiom states that if two nodes, u, v, have the same centrality and the same number of outgoing edges, then their edges are interchangeable. More specifically, if we replace edges (u,u') and (v,v') with edges (u,v') and (v,u'), then the centralities of all nodes will not change. As we will discuss in the next section, this axiom forbids the centrality from promoting the diversity of incoming edges.

Edge Swap (ES): *For every graph G node weight β and nodes $u, v \in V(G)$ such that $F_u(G,\beta) = F_v(G,\beta)$ and $|\Gamma^+_G(u)| = |\Gamma^+_G(v)|$, if $(u,u'),(v,v') \in E(G)$, then the following holds for every $w \in V(G)$:*

$$F_w(G - \{(u,u'),(v,v')\} + \{(u,v'),(v,u')\},\beta) = F_w(G,\beta).$$

The following theorem states that replacing Lack of Self-Impact with Edge Swap in the axiomatization of the Random Walk Decay centrality results in an axiomatization of PageRank.

Theorem 4. *PageRank is the unique centrality measure that satisfies LOC, SM, DLP, 1-NG, ES and RWP.*

Proof (Sketch). We use induction on the number of cycles in G. If there are no cycles in G, then the random walk visits each node once. Hence, from RWP, the centrality of any node, v, is the same as in the graph without outgoing edges of v and the thesis follows from **Claim 3** from the proof of Theorem 2 and the fact that for every sink the Random Walk Decay centrality and PageRank are equal.

Now, consider an arbitrary graph G with k cycles and node weights β. Fix $v \in V(G)$ that belongs to some cycle. Let us construct a new graph $G' = (\{v', w\}, \{(v', w), \dots, (v', w)\})$, disjoint with G, where $\beta(v') = F_v(G)$ and $|E(G')| = |\Gamma^+_G(v)|$. From LOC we know that $F_u(G) = F_u(G + G')$ for every $u \in V(G)$. Moreover, it can be shown from 1-NG, LOC and SM that $F_{v'}(G') = \beta(v') = F_v(G)$. Thus, nodes v and v' have the same centralities in $G + G'$ and the same number of outgoing edges. Now, from ES replacing all outgoing edges of v with all outgoing edges of v' does not affect the centralities in the graph. Observe that this operation breaks the cycles that v belongs to in G. Hence, the obtained graph has less cycles than G and the thesis follows from the inductive assumption. □

Table 1 summarizes our axiomatic results. Since SPP implies LSI, based on Theorem 3 we know that there exists no centrality that satisfies LOC, SM, DLP, 1-NG, SPP and ES.

Comparison with PageRank

Our axiomatic characterizations highlight two differences between the Random Walk Decay centrality and PageRank. In this section, we focus on these two differences and show how they affect the behaviour of these centrality measures.

Strategy-proofness (with respect to outgoing edges)

In many settings, outgoing edges are subject to the node's decision or manipulations. Examples include the Twitter social network (where outgoing edges represent the accounts that are followed by a user) and the World Wide Web (where outgoing edges represent the links to other websites). Consequently, Lack of Self-Impact can be considered a property of *strategy-proofness* for centrality measures—if outgoing edges do not affect the centrality of a node, then the node has no incentive to manipulate its outgoing connections.

Interestingly, PageRank does not satisfy Lack of Self-Impact. In the following example we show how, by adding outgoing edges, a node can increase its centrality and position in the ranking according to PageRank, but not according to the Random Walk Decay centrality.

Example 4. *Consider graph G from Figure 2. Graph G consists of two 4-cycles, (u_1,u_2,u_3,u_4,u_1), (v_1,v_2,v_3,v_4,v_1). The two cycles are connected via 3 edges: (v_4,u_4), (u_3,v_3), and (u_2,v_2). Due to the edges connecting both cycles, the nodes v_2, v_3, and v_4 are visited more often (and earlier) by the random walk, and are thus ranked first by both PageRank and the Random Walk Decay centrality. Node u_1, that will be of our interest, is ranked 5th according to both measures.*

Figure 2 also depicts G', which is obtained from G by adding the edge (u_1,u_4). Since this is an outgoing edge for u_1, adding it does not affect the Random Walk Decay centrality of u_1. In contrast, this edge has a significant impact on PageRank of u_1. The reason lies in the fact that the random walk will now visit u_1 much more often—whenever the random walk reaches node u_1, with probability $1/2$ it will go back to u_4 from which the only outgoing edge goes to u_1. In

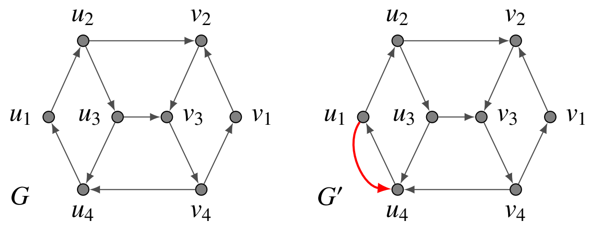

Figure 2: Two graphs considered in Example 4, namely $G = (V, E)$ and $G' = (V, E')$ with unit weights: $\beta(v) = 1$ for every $v \in V$. Graph G' is obtained from graph G by adding (u_1, u_4).

result, both u_1 *and* u_4 *top the ranking according to PageRank. The centralities of all nodes for* $a = 0.8$ *are as follows:*

node v	$PR_v(G,\beta)$	$PR_v(G',\beta)$	$RWD_v(G,\beta)$	$RWD_v(G',\beta)$
v_3	6.82 *(1st)*	6.08 *(3rd)*	4.67 *(1st)*	4.36 *(1st)*
v_4	6.45 *(2nd)*	5.86 *(4th)*	4.43 *(2nd)*	4.21 *(2nd)*
v_2	5.80 *(3rd)*	5.13 *(5th)*	4.13 *(3rd)*	3.77 *(4th)*
u_2	4.84 *(4th)*	3.62	3.75 *(4th)*	3.02
$\boldsymbol{u_1}$	**4.81 *(5th)***	**6.56 *(2nd)***	**3.72 *(5th)***	**3.72 *(5th)***
u_4	4.76	6.95 *(1st)*	3.68	3.94 *(3rd)*
v_1	3.58	3.35	2.76	2.60
u_3	2.94	2.45	2.48	2.17

In Example 4, a node improved its PageRank by adding an edge to its direct predecessor. In the next section, we will discuss how incoming edges affect both centrality measures.

Diversity (of incoming edges)

One of the characteristic properties of PageRank is its recursive formula (see (2)), which states that PageRank of a node depends solely on PageRank of its direct predecessors (or, more precisely, nodes incident to its incoming edges). Intuitively speaking, PageRank of a node does not depend on the position in the network of its predecessors, but only on their centrality. In our axiomatic characterization, this property is captured by Edge Swap, which implies that an incoming edge from a node with the lowest centrality in a densely connected part of the graph could be as profitable as an incoming edge from a node with the highest centrality in a different, less densely connected part.

The Random Walk Decay centrality does not satisfy Edge Swap. In fact, the node can achieve higher centrality if it has incoming edges from a diverse set of nodes, i.e., coming from different parts of the network. We demonstrate this point with the following example.

Example 5. *Consider graph G from Figure 3. This graph consists of three more densely connected parts, so called* communities*:* $\{u_1, u_2, u_3\}$*,* $\{v_1, v_2, v_3\}$ *and* $\{w_1, w_2, w_3, w_4\}$*. These communities are connected through nodes* u_1, v_1, w_1 *which form a 3-clique. Since* w_1 *belongs to the biggest community, both its Random Walk Decay centrality and its PageRank are the highest. The nodes* u_1 *and* v_1 *have the second highest values, with symmetrical positions in the graph.*

Figure 3 also depicts the graph G'*, which is obtained from G by rewiring the two highlighted (red) edges. Specifically, the edges* (u_2, u_1) *and* (v_2, v_1) *are replaced by* (u_2, v_1) *and* (v_2, u_1)*; in result, the two new edges connect two communities. Since* u_2 *and* v_2 *both have two edges and clearly the same centralities in graph G, from Edge Swap we know that PageRank of every node in* G' *is the same as in G. In contrast, the centralities of both nodes* u_1 *and* v_1 *increase according to the Random Walk Decay centrality. This is because, according to this centrality, an edge from a different community is more profitable than an edge from your own community. In our example, the random walk that starts from nodes* v_2 *and* v_3 *reaches node* u_1 *faster in graph* G'*; the same holds for node* v_1*. In result, in* G'*, the Random Walk Decay centralities of* u_1 *and* v_1 *are higher than the Random Walk Decay centrality of node* w_1*. The centralities of all nodes for* $a = 0.8$ *are:*

node v	$PR_v(G,\beta)$	$PR_v(G',\beta)$	$RWD_v(G,\beta)$	$RWD_v(G',\beta)$
w_1	7.15 *(1st)*	7.15 *(1st)*	4.40 *(1st)*	4.40 *(3rd)*
$\boldsymbol{u_1}$	**7.06 *(2nd)***	**7.06 *(2nd)***	**4.15 *(2nd)***	**4.46 *(1st)***
$\boldsymbol{v_1}$	**7.06 *(2nd)***	**7.06 *(2nd)***	**4.15 *(2nd)***	**4.46 *(1st)***
w_3	4.33 *(4th)*	4.33 *(4th)*	2.62	2.62
w_2, w_4	4.16 *(5th)*	4.16 *(5th)*	2.74 *(4th)*	2.74
u_2, v_2	4.02	4.02	2.56	2.85 *(4th)*
u_3, v_3	4.02	4.02	2.56	2.70

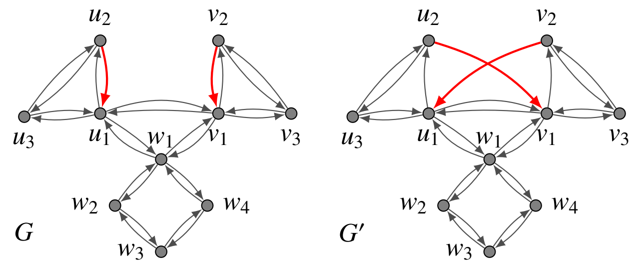

Figure 3: Two graphs considered in Example 5, namely $G = (V, E)$ and $G' = (V, E')$ with unit weights: $\beta(v) = 1$ for every $v \in V$. Graph G' is obtained from graph G by replacing edges (u_2, u_1) and (v_2, v_1) with edges (u_2, v_1) and (v_2, u_1).

Example 5 shows that the Random Walk Decay centrality increases when incoming edges become more diverse. As such, it avoids putting at the top of the ranking several nodes from the same community, which often happens in PageRank (Avrachenkov and Litvak 2006; Zhirov, Zhirov, and Shepelyansky 2010).

Related Work

Our paper belongs to a line of papers that study the axiomatic properties of centrality measures (Boldi and Vigna 2014; Bloch, Jackson, and Tebaldi 2016; Skibski, Michalak, and Rahwan 2018). In particular, our axiomatization of the Decay centrality relies on a recent axiomatization for undirected graphs proposed by Skibski and Sosnowska (2018).

To date, the only axiomatized centrality measure based on the random walk was PageRank. Palacios-Huerta and Volij (2004) proposed an axiomatization of the simplified version of PageRank. Altman and Tennenholtz (2005) also focused on a simplified version of PageRank, but axiomatized the ranking, rather than the numerical values. Recently, Wąs and Skibski (2018) proposed the first axiomatization of

PageRank in its general form. Our new axiomatization significantly differs from all of these axiomatizations.

Conclusions

In this paper, we proposed Random Walk Decay centrality—a new centrality measure based on the random walk. We provided an axiomatic characterization using six axioms, and proved that replacing only one property leads to an axiomatization of the Decay centrality. Furthermore, we showed that Random Walk Decay works similarly to PageRank, but has certain properties that can be more desirable in various settings. In our future work, we plan to perform a comparative experimental analysis of Random Walk Decay and PageRank using real-world networks.

Acknowledgements

This work was supported by the Foundation for Polish Science within the Homing programme (Project title: "Centrality Measures: from Theory to Applications") and the Polish National Science Center grant 2016/23/B/ST6/03599.

References

Altman, A., and Tennenholtz, M. 2005. Ranking systems: the PageRank axioms. In *Proceedings of the 6th ACM Conference on Electronic Commerce (ACM-EC)*, 1–8.

Avrachenkov, K., and Litvak, N. 2006. The effect of new links on google pagerank. *Stochastic Models* 22(2):319–331.

Bianchini, M.; Gori, M.; and Scarselli, F. 2005. Inside pagerank. *ACM Transactions on Internet Technology (TOIT)* 5(1):92–128.

Bianzino, A. P.; Chaudet, C.; Rossi, D.; Rougier, J.-L.; and Moretti, S. 2011. The green-game: Striking a balance between qos and energy saving. In *Teletraffic Congress (ITC), 2011 23rd International*, 262–269.

Bloch, F.; Jackson, M. O.; and Tebaldi, P. 2016. Centrality measures in networks. *Available at SSRN 2749124*.

Boldi, P., and Vigna, S. 2014. Axioms for centrality. *Internet Mathematics* 10(3-4):222–262.

Borgatti, S. P. 2005. Centrality and network flow. *Social networks* 27(1):55–71.

Freeman, L. C. 1979. Centrality in social networks: Conceptual clarification. *Social Networks* 1(3):215–239.

Freeman, L. C. 2008. Going the wrong way on a one-way street: Centrality in physics and biology. *Journal of Social Structure* 9(2):1–15.

Hinz, O.; Skiera, B.; Barrot, C.; and Becker, J. U. 2011. Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing* 75(6):55–71.

Huberman, B. A.; Pirolli, P. L.; Pitkow, J. E.; and Lukose, R. M. 1998. Strong regularities in world wide web surfing. *Science* 280(5360):95–97.

Jackson, M. O. 2008. *Social and economic networks*, volume 3. Princeton university press.

Koschützki, D.; Lehmann, K. A.; Tenfelde-Podehl, D.; and Zlotowski, O. 2005. Advanced centrality concepts. In *Network Analysis*, volume 3418 of *Lecture Notes in Computer Science*. Springer. 83–111.

Lerman, K., and Ghosh, R. 2010. Information contagion: An empirical study of the spread of news on digg and twitter social networks. *Icwsm* 10:90–97.

Lovász, L. 1993. Random walks on graphs: A survey. *Combinatorics, Paul Erdos is eighty* 2(1):1–46.

Newman, M. E. 2005. A measure of betweenness centrality based on random walks. *Social networks* 27(1):39–54.

Newman, M. 2010. *Networks: an introduction*. Oxford university press.

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The PageRank citation ranking: bringing order to the web. *Stanford InfoLab*.

Palacios-Huerta, I., and Volij, O. 2004. The measurement of intellectual influence. *Econometrica* 72(3):963–977.

Palbom, A. 2005. Complexity of the directed spanning cactus problem. *Discrete applied mathematics* 146(1):81–91.

Sabidussi, G. 1966. The centrality index of a graph. *Psychometrika* 31(4):581–603.

Skibski, O., and Sosnowska, J. 2018. Axioms for distance-based centralities. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 1218–1225.

Skibski, O.; Rahwan, T.; Michalak, T. P.; and Yokoo, M. 2016. Attachment centrality: An axiomatic approach to connectivity in networks. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 168–176.

Skibski, O.; Michalak, T. P.; and Rahwan, T. 2018. Axiomatic characterization of game-theoretic centrality. *Journal of Artificial Intelligence Research* 62:33–68.

White, S., and Smyth, P. 2003. Algorithms for estimating relative importance in networks. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 266–275. ACM.

Wąs, T., and Skibski, O. 2018. Axiomatization of the PageRank centrality. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 3898–3904.

Zhirov, A.; Zhirov, O.; and Shepelyansky, D. 2010. Two-dimensional ranking of wikipedia articles. *The European Physical Journal B* 77(4):523–531.