Axiomatization of the PageRank Centrality

Tomasz Wąs, Oskar Skibski
University of Warsaw, Poland
{t.was,o.skibski}@mimuw.edu.pl

Abstract

We propose an axiomatization of PageRank. Specifically, we introduce five simple axioms—*Foreseeability*, *Outgoing Homogeneity*, *Monotonicity*, *Merging*, and *Dummy Node*—and show that PageRank is the only centrality measure that satisfies all of them. Our axioms give new conceptual and theoretical underpinnings of PageRank and show how it differs from other centralities.

1 Introduction

PageRank, introduced almost 20 years ago by Page *et al.* [1999], had a huge contribution to the way the current Internet looks like. While originally proposed to assess the importance of web pages in Google search engine, PageRank has quickly became one of the standard method of assessing the importance of nodes in complex networks—it is used to highlight cancer genes in proteomic data [Iván and Grolmusz, 2010], find the most influential users in the Twitter social network [Weng *et al.*, 2010], and evaluate the prestige of journals based on the citation network [Bollen *et al.*, 2006].

In the network science terminology, PageRank is a *centrality measure*—given a network (graph), it assigns to every node a number reflecting its importance [Brandes and Erlebach, 2005; Costa *et al.*, 2007; Grando *et al.*, 2016]. The standard centrality measures include the *degree centrality* that counts the number of links of a node, *closeness centrality* that assesses a node based on the average distance from other nodes, and *betweenness centrality* that focus on the number of shortest paths that goes through a given node [Freeman, 1979]. These measures significantly rely on the number of links that a given node has. Clearly, such an approach is undesirable in the context of web pages, where the cost of creating a link is marginal. As noted by Page *et al.* [1999], *"any evaluation strategy which counts replicable features of web pages is prone to manipulation."*

That is why PageRank was built upon the *eigenvector centrality* [Bonacich, 1972]. This centrality is defined recursively as the solution of the following equality: "centrality of a node is proportional to the sum of centralities of nodes pointing toward it" [Newman, 2010]. Intuitively, the eigenvector centrality is similar to the degree centrality with the difference that the degree centrality only counts the number of neighbours, while the eigenvector centrality sums their centralities.

PageRank introduces two important changes to the definition of the eigenvector centrality. First of all, the centrality of every neighbour, before being summed, is divided by its out-degree, i.e., the number of its edges. Thus, a centrality of a particular node is spread among successors, and not added to each node independently. In that way, the role of a node is limited by its importance. Second of all, each node is assigned an additional weight, regardless of incoming edges. These weights can represent, for example, how well a given web page matches the desirable criteria [Haveliwala, 2002]. Node weights greatly affect the behaviour of the value; in fact, a version of PageRank without node weights, called *"a slightly simplified version of the PageRank"* by Page *et al.* [1999], has worse computational properties and is hard to define for some graphs [Langville and Meyer, 2011].

Although PageRank has been proven to be an effective tool in various settings, the theoretical analysis of its properties is lacking. Given the plethora of alternative centrality measures that have been proposed in the literature [Koschützki *et al.*, 2005], it is important to answer the question: how does PageRank differ from other similar measures? In other words, what properties characterize PageRank?

We answer this question by proposing an axiomatization of PageRank based on five simple axioms: (1) *Foreseeability*, states that the centrality depends solely on predecessors, and not successors of a node. (2) *Outgoing Homogeneity* states that multiplying weights of outgoing edges of a particular node does not affect centralities in the graph. (3) *Monotonicity* states that adding an edge to the graph should increase a centrality of its end-node. (4) *Merging* states that if edges of two nodes are equally important, then merging both nodes does not affect centralities of others. Finally, (5) *Dummy Node* implies that a node without any edges should have a centrality equal to its weight. We prove that PageRank is the only centrality measure that satisfies all the above axioms.

To date, the original, general form of PageRank has not been axiomatized. There exist two axiomatizations of the simplified version of PageRank without node weights. Palacios-Huerta and Volij [2004] studied four properties of the *Invariant method*, which is this version of PageRank adapted to the setting of scientific journal citation network. Altman and Tennenholtz [2005] characterized PageRank for

graphs with unweighted edges and nodes. Moreover, only the ranking of nodes is axiomatized, and not the absolute values. Since weights of nodes significantly affect the value, many of the properties proposed in these papers do not work for the original, general version.[1]

2 Preliminaries

In this section, we introduce basic definitions and notations.

Graphs: In this paper, we consider directed graphs with possible loops with weighted edges and nodes. Formally, a *graph* is an ordered 4-tuple, $G = (V, E, \beta, \omega)$, where V is the set of its nodes with the weight function $\beta : V \to \mathbb{R}_{\geq 0}$ and $E \subseteq V \times V$ is the set of its edges with the weight function $\omega : E \to \mathbb{R}_{>0}$. For convenience, we will assume that $\omega(e) = 0$ if $e \notin E$. If $\beta(v) = 1$ for every $v \in V$, we will denote it by **1**. Analogously, if $\omega(e) = 1$ for every $e \in E$, we will also denote it by **1**. The set of all graphs for a given set of nodes V is denoted by $\mathcal{G}^V$ and the set of all possible graphs is denoted by $\mathcal{G}$.

An edge (u, v) is an *outgoing* edge for the node u which is the start of this edge and an *incoming* edge for its end, node v. If $u = v$, the edge is called a *loop*. The set of outgoing edges from v is denoted by $\Gamma_G^+(v)$, and from U by $\Gamma_G^+(U)$:

$$\Gamma_G^+(v) = \{(v, u) \in E : u \in V\}, \quad \Gamma_G^+(U) = \bigcup_{v \in U} \Gamma_G^+(v).$$

Analogously, the set of incoming edges to v is denoted by $\Gamma_G^-(v)$, and to U by $\Gamma_G^-(U)$. Also, we define $\Gamma_G^\pm(v) = \Gamma_G^+(v) \cup \Gamma_G^-(v)$. A node without any outgoing edges is called a *sink*.

The *out-degree* of a node is the sum of weights of its outgoing edges, i.e., $\deg_G^+(v) = \sum_{e \in \Gamma_G^+(v)} \omega(e)$. Analogously, the *in-degree* is the sum of weights of its incoming edges.

A *walk* is an ordered set of $k > 1$ nodes, $p = (v_1, \ldots, v_k)$, such that every pair of consequent nodes form an edge, i.e., $(v_i, v_{i+1}) \in E$ for every $i \in \{1, \ldots, k-1\}$. If additionally, the last and the first nodes form an edge, i.e., $(v_n, v_1) \in E$, then a walk is a *cycle*. The *length of a walk* is a number of nodes in p minus 1. By $\Omega_G(u, v)$ we will denote the set of all walks in G between u and v, i.e., such that u is its first node, and v is the last one.

If there exists a walk from u to v, then v is called a *successor* of u and u is a *predecessor* of v. The set of all predecessors of v is denoted by $P_G(v)$:

$$P_G(v) = \{u \in V : \Omega_G(u, v) \neq \emptyset\}.$$

Consider graph $G = (V, E, \beta, \omega)$. For a subset of nodes, $U \subseteq V$, we define $\beta[U] : U \to \mathbb{R}_{\geq 0}$ as $\beta[U](v) = \beta(v), \forall v \in U$. Analogously, for a subset of edges, $M \subseteq E$, we define $\omega[M] : M \to \mathbb{R}_{>0}$ as $\omega[M](e) = \omega(e), \forall e \in M$.

For a subset of nodes, $U \subseteq V$, a *subgraph induced by U* is a graph obtained from G by removing nodes $V \setminus U$ and their edges:

$$G[U] = (U, E[U], \beta[U], \omega[E[U]]),$$

where $E[U] = \{(u, v) \in E : u, v \in U\}$. Similarly, for a subset of edges, $M \subseteq E$, a *subgraph induced by M*, denoted by $G[M]$, is a graph obtained from G by removing edges from $E \setminus M$ and nodes without edges:

$$G[M] = (V[M], M, \beta[V[M]], \omega[M]),$$

where $V[M] = \{v \in V : M \cap \Gamma_G^\pm(v) \neq \emptyset\}$. Observe that if $A \subseteq B \subseteq V$ or $A \subseteq B \subseteq E$, then $G[A] = (G[B])[A]$. For a subset of edges, $E' \subseteq E$, we also use a shorthand notation: $G - E' = (V, E \setminus E', \beta, \omega[E \setminus E'])$.

Graph $G = (V, E, \beta, \omega)$ is called a *cycle graph* if $V = \{v_1, \ldots, v_n\}$ and $E = \{(v_1, v_2), \ldots, (v_{n-1}, v_n), (v_n, v_1)\}$.

An *in-tree* is a tree with edges directed towards its root. Formally, $G = (V, E, \beta, \omega)$ is an in-tree with the *root* $v \in V$, if $|\Gamma_G^+(v)| = 0$, $|\Gamma_G^+(u)| = 1$ for every $V \setminus \{u\}$, and there exists no cycle in G. In an in-tree, for every $u \in V$ there exists a unique walk from u to v; we denote its length by $h(u)$. Also, we assume that $h(v) = 0$.

Graph operations: Consider two graphs $G = (V, E, \beta, \omega)$ and $G' = (V', E', \beta', \omega')$.

We say that G' is *isomorphic* to G if there exist a bijection, $f : V \to V'$, such that $E' = \{(f(s), f(t)) : (s, t) \in E\}$, $\beta'(s) = \beta(f^{-1}(s))$ for every $s \in V'$, and $\omega'(s, t) = \omega(f^{-1}(s), f^{-1}(t))$ for every $(s, t) \in E'$.

The sum of G and G' is defined as follows:

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

where $(\omega + \omega')(e) = \omega(e) + \omega'(e)$ and $(\beta + \beta')(v) = \beta(v) + \beta'(v)$, assuming $\beta(v) = 0$ for $v \notin V$ and $\beta'(v) = 0$ for $v \notin V'$.

By *multiplying* graph G by a scalar, $c \in \mathbb{R}_{>0}$, we get

$$c \cdot G = (V, E, c \cdot \beta, c \cdot \omega).$$

Merging node u with v results in a graph where $\beta(v) = \beta(v) + \beta(u)$ and edges of u are "added" to node v. Formally, consider a function $f : V \to V \setminus \{u\}$ that maps both u and v into v: $f(u) = f(v) = v$, and $f(w) = w$ for every $w \in V \setminus \{u, v\}$. Consequently, from merging u with v we obtain the graph $G' = (V \setminus \{u\}, E', \beta', \omega')$, where:

- $E' = \{(f(s), f(t)) : (s, t) \in E\}$,
- $\beta'(s) = \sum_{q \in f^{-1}(s)} \beta(q)$ for every $s \in V \setminus \{u\}$, and
- $\omega'(s, t) = \sum_{p \in f^{-1}(s), q \in f^{-1}(t)} \omega(p, q)$ for every $(s, t) \in E'$.

An example can be seen in Figure 1, where graph G_4 is obtained from G_5 by merging node u with v.

PageRank: A *centrality measure*, $F : \mathcal{G}^V \to \mathbb{R}_{\geq 0}^V$, is a function which assigns to every graph G and its node v a real value, denoted by $F_v(G)$.

PageRank [Page *et al.*, 1999], is a centrality measure, parametrised by $a \in (0, 1)$, defined by the following recursive formula:

$$PR_v(G) = a \left(\sum_{(u,v) \in \Gamma_G^-(v)} \frac{\omega(u, v)}{\deg_G^+(u)} PR_u(G) \right) + \beta(v) \quad (1)$$

for every $v \in V$. It can be easily proved that for a given value of a there exist exactly one vector $PR(G)$ for any graph G [Bianchini *et al.*, 2005].
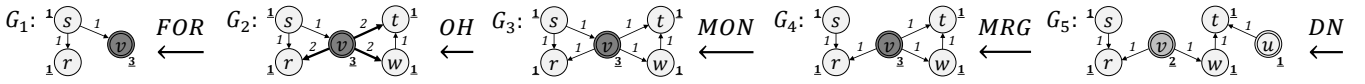
---

[1]More in detail, only one of our axioms (*Merging*) is similar to axiom 3.4 (*collapsing*) from Altman and Tennenholtz [2005]. Other axioms, most of which rely on linearity of the simplified version of PageRank, could not be easily extended to the general version.

Figure 1: An illustration of our five axioms. DN implies $F_v(G_5) = 2$ and $F_u(G_5) = 1$. Since equation (2) is satisfied, MRG implies $F_v(G_4) = F_v(G_5) + F_u(G_5)$. MON implies $F_v(G_3) > F_v(G_4)$. OH implies $F_v(G_2) = F_v(G_3)$. Finally, FOR implies $F_v(G_1) = F_v(G_2)$. Edge weights are in *italic* font and node weights are **bolded**.

The PageRank centrality was proposed to determine the importance of web pages on the Internet. In this setting, nodes are referred to as *pages*, outgoing edges are named *links*, and incoming edges—*backlinks*.

To illustrate the way the PageRank centrality works, the authors proposed the model of a *random surfer*. In this model, the importance of each page is the stationary probability of this page being visited by a random surfer in an infinite walk on the entire network. In each step of this walk, the surfer clicks on a random link on the page he is; thus, while being on page v, he moves to the page u with the probability $\omega(v, u)/\deg_G^+(v)$. However, with some probability, $1 - a$, the surfer "gets bored", breaks his walk and start again from an indeterminate node. In such a case, he chooses each node, u, with the probability $\beta(u)$.[2]

As can be seen from the random surfer model, weights β can represent known preferences of the surfer over pages in the network. In lack of additional information, the uniform distribution can be chosen.

3 Our Axioms

In this section, we present our five axioms. In explanations, we will often use the web page terminology.

Our first axiom states that the importance of a page should depend not on its links (outgoing edges), but on the backlinks (incoming ones). Analogously, the importance of pages linking to our page should depend on their backlinks. In result, *Foreseeability* states that we can "foresee" a centrality of a page based on all predecessors and their links.

Foreseeability (FOR): For $G \in \mathcal{G}^V$ and $v \in V$:
$$F_v(G) = F_v(G[\Gamma_G^+(P_G(v))]),$$
if $P_G(v) \neq \emptyset$, and $F_v(G) = F_v(G[\{v\}])$, otherwise.

Note that Forseeeability takes into account also links of the predecessors, as the importance of a specific link of a predecessor (in particular, the one linking to our page) depends on the existence of other links.

Our second axiom, *Outgoing Homogeneity*, states that an importance of a given page should not depend on the absolute number of its links. This important property laid at the foundation of PageRank. In the graph setting, it means that multiplying the weights of the outgoing edges of any node by a constant should not affect anyone's centrality.[3]

---

[2]This interpretation is valid if $\sum_{v \in V} \beta(v) = 1 - a$. However, if the equation does not hold, the values can be easily scaled.

[3]This axiom is similar to *invariance with respect to reference intensity*—a property defined in an axiomatic approach to scientific journal positioning introduced by Palacios-Huerta and Volij [2004].

Outgoing Homogeneity (OH): For $G = (V, E, \beta, \omega)$, $v \in V$ and $s > 0$ if
$$\omega'(e) = \begin{cases} s \cdot \omega(e) & \text{if } e \in \Gamma_G^+(v), \\ \omega(e) & \text{otherwise,} \end{cases}$$
then $F(V, E, \beta, \omega') = F(V, E, \beta, \omega)$.

Monotonicity states that if a page receives an additional backlink from a site that has non-zero popularity (which means it is sometimes visited) its importance increases.[4]

Monotonicity (MON): For $G = (V, E, \beta, \omega)$ and $u, v \in V$ such that $(u, v) \notin E$ and $F_u(G) > 0$:
$$F_v(V, E \cup \{(u, v)\}, \beta, \omega') > F_v(V, E, \beta, \omega),$$
where $\omega'(e) = \omega(e)$ for every $e \in E$.

For our next axiom, *Merging*, consider a situation in which two pages are being merged into one, preserving links and backlinks of both pages. The axiom states that as long as the proportion of importance to the number of links on both pages is the same (i.e., links on both pages are equally important), merging both pages does not change their combined importance or the importance of other pages.

Merging (MRG): For $G \in \mathcal{G}^V$ and $u, v \in V$, if
$$\deg_G^+(u) \cdot F_v(G) = \deg_G^+(v) \cdot F_u(G), \qquad (2)$$
then for the graph, G', obtained from G by merging node u with v, $F_v(G') = F_v(G) + F_u(G)$ and $F_s(G') = F_s(G)$ for every $s \in V \setminus \{u, v\}$.

For nodes u, v with outgoing edges and positive centralities condition (2) simply means that the ratio of their out-degree is equal to the ratio of their centrality, i.e.: $\deg_G^+(u)/\deg_G^+(v) = F_u(G)/F_v(G)$.

Our last axiom, *Dummy Node*, considers a simple borderline case—if there exists a page without any links nor backlinks, then its centrality is equal to its weight.

Dummy Node (DN): For $G \in \mathcal{G}^V$ and $v \in V$:
$$\Gamma_G^\pm(v) = \emptyset \Rightarrow F_v(G) = \beta(v).$$

We conclude this section with the theorem that states the PageRank centrality satisfies all above axioms.

Theorem 1. *The PageRank centrality for any constant $a \in (0, 1)$ satisfies FOR, OH, MON, MRG and DN.*

Proof. Due to space constraints, we will omit several proofs in the paper. □

---

[4]A similar axiom under the same name was proposed by both Sabidussi [1966] and Skibski *et al.* [2016].

4 Uniqueness of the PageRank Centrality

This section is devoted to proving Theorem 12, which states that each centrality measure that satisfies our five axioms is a PageRank centrality (for some constant a).

We begin with simple properties implied by our axioms.

Proposition 2. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG and DN. For an arbitrary graph, $G = (V, E, \beta, \omega)$, the following statements are true:*

a. *(Locality[5]): For every $G' \in \mathcal{G}^{V'}$ such that $V \cap V' = \emptyset$ and $v \in V$ it holds that $F_v(G + G') = F_v(G)$.*

b. *(Anonymity[6]): If G' is isomorphic to G with isomorphism f, then $F_v(G) = F_{f(v)}(G')$ for every $v \in V$.*

c. *(Zero Nodes): $F_v(G) = 0$ if and only if $\beta(v) = 0$ and $\beta(u) = 0$ for every $u \in P_G(v)$.*

d. *(Removing Zero Nodes): If $V_0 = \{v \in V : F_v(G) = 0\}$ and $v \in V \setminus V_0$, then $F_v(G[V \setminus V_0]) = F_v(G)$.*

e. *(Homogeneity): $F(c \cdot G) = c \cdot F(G)$ for every $c \in \mathbb{R}_{>0}$.*

As we will show in Theorem 12, our axioms imply that a centrality measure is the PageRank centrality parametrised by an arbitrary constant $a \in (0, 1)$ (equation (1)). The following lemma characterizes constant a.

Lemma 3. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. There exists a constant, $a \in (0, 1)$, such that for every graph, $G = (V, E, \beta, \omega)$, if node $v \in V$ forms an isolated loop, i.e., $\Gamma_G^{\pm}(v) = \{(v, v)\}$, then:*

$$F_v(G) = \frac{\beta(v)}{1 - a}.$$

Proof. Consider graph $\hat{G} = (\{u\}, \{(u, u)\}, \mathbf{1}, \mathbf{1})$. Let us define $a = 1 - 1/F_u(\hat{G})$: in this way, we get that $F_u(\hat{G}) = 1/(1 - a)$. From Homogeneity (Proposition 2e) and OH:

$$F_u(\{u\}, \{(u, u)\}, \beta, \omega) = \beta(u) \cdot F(\hat{G}) = \beta(u)/(1 - a),$$

for an arbitrary β and ω. Now, from Locality (Proposition 2a) and Anonymity (Proposition 2b) we get that for every graph, $G = (V, E, \beta, \omega)$, and every node, $v \in V$, such that $\Gamma_G^{\pm}(v) = \{(v, v)\}$ it holds that $F_v(G) = \beta(v)/(1 - a)$.

It remains to prove that $a \in (0, 1)$. Since $1/F_u(\hat{G}) > 0$, we have $a < 1$. From MON and DN:

$$F_u(\hat{G}) > F_u(\{u\}, \emptyset, \mathbf{1}, \emptyset) = 1,$$

which implies that $a > 0$. This concludes the proof. $\square$

We will say that F satisfies axioms *with constant a* if it satisfies $F_u(\{u\}, \{(u, u)\}, \mathbf{1}, \mathbf{1}) = 1/(1 - a)$.

In our proof, we will focus on the following class of graphs.

Definition 1. *(F-normalised graphs) For a centrality measure, F, graph $G \in \mathcal{G}^V$ is F-normalised if for every $v \in V$, its out-degree equals its centrality, i.e., $\deg_G^+(v) = F_v(G)$.*

---

[5]Skibski *et al.* [2016] proposed this property as one of the axioms for the attachment and degree centralities.

[6]It was one of the axioms proposed by Sabidussi [1966].

Observe that in a F-normalised graph each pair of nodes satisfies equation (2). Moreover, if F satisfies MRG, then merging any two nodes results in an F-normalised graph.

In what follows, we show that if F satisfies all our axioms, then every graph, $G \in \mathcal{G}^V$, can be transformed into an F-normalised graph, $\tilde{G}$, in which centralities of nodes from V are the same as in G, i.e.: $F_v(G) = F_v(\tilde{G})$ for every $v \in V$.

We will construct this graph in three steps. (1) First, if needed, we add a new node and incoming edges to it to make sure that each node with a positive centrality has at least one outgoing edge (function f^+). (2) Second, we remove all of the outgoing edges of the nodes with a zero centrality (function f^-). (3) Finally, we adjust the weights of the edges to level out-degree with centrality (function f^*).

Definition 2. *(F-normalisation) For a centrality measure, F, and a graph, G, F-normalisation of G is graph $\tilde{G} = (f^* \circ f^- \circ f^+)(G)$, where $f^+, f^-, f^* : \mathcal{G} \rightarrow \mathcal{G}$ are functions defined for every graph $G = (V, E, \beta, \omega)$ as follows:*

- $f^+(G) = (V \cup \{\tilde{v}\}, E \cup E', \beta', \omega')$, *where*

  - $E' = \{(u, \tilde{v}) : \Gamma_G^+(u) = \emptyset, F_u(G) > 0\} \cup \{(\tilde{v}, \tilde{v})\}$,
  - $\beta'(\tilde{v}) = 0$ *and* $\beta'(v) = \beta(v)$ *for* $v \in V$, *and*
  - $\omega'(e) = 1$ *for* $e \in E'$ *and* $\omega'(e) = \omega(e)$ *for* $e \in E$.

  if $|E'| > 1$, *and* $f^+(G) = G$, *otherwise.*

- $f^-(G) = G - \{(u, v) \in E : F_u(G) = 0\}$; *and*

- $f^*(G) = (V, E, \beta, \omega')$, *where* $\omega'(u, v) = \omega(u, v) \frac{F_u(G)}{\deg_G^+(u)}$

  for every $(u, v) \in E$.

The following lemma states that F-normalisation is indeed F-normalised and it preserves the centralities of the nodes.

Lemma 4. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. For every graph $G \in \mathcal{G}^V$, its F-normalisation $\tilde{G}$ is F-normalised and $F_v(\tilde{G}) = F_v(G)$ for every $v \in V$.*

Proof. Let us consider an arbitrary graph $G = (V, E, \beta, \omega)$. Fix $v \in V$. First, we will show that functions f^+, f^-, and f^* do not modify the centrality of v.

Consider function f^+. Let $V' = \{u \in V : \Gamma_G^+(u) = \emptyset, F_u(G) > 0\}$ denotes the set of sinks with positive centrality in G. If $V' = \emptyset$, then $f^+(G) = G$. Assume otherwise. Clearly, nodes from V' are not among the predecessors of v in G. In result, adding a new node, $\tilde{v}$, and edges between V' and $\tilde{v}$ does not affect the set of predecessors and the set of their edges. Thus, FOR implies $F_v(G) = F_v(f^+(G))$.

Consider function f^-. Observe that removing all outgoing edges from nodes with zero centrality is equivalent to removing those nodes and adding them back again as isolated nodes. Therefore, from Zero Nodes and Removing Zero Nodes (Propositions 2c and 2d) we get $F_v(G) = F_v(f^-(G))$.

Consider function f^*. Immediately from OH we get that $F_v(G) = F_v(f^*(G))$.

Finally, we argue that graph $\tilde{G} = f^*(f^-(f^+(G)))$ is F-normalised. Consider graph $G' = f^-(f^+(G))$. Since we proved that $F_v(G') = F_v(f^+(G)) = F_v(G)$, we know that
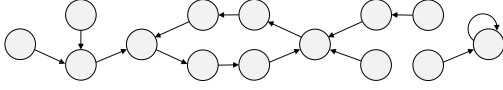
Figure 2: An example of an out-unity graph.

Figure 3: An illustration to the proof of Lemma 9.

$F_v(G') = 0$ if and only if $\Gamma^+_{G'}(v) = \emptyset$. In result, after applying function f^*, we get that $\deg^+_{\tilde{G}}(u) = F_u(G') = F_u(\tilde{G})$. Thus, $\tilde{G}$ is F-normalised. $\square$

To make use of the class of F-normalised graphs we show that it forms a vector space, i.e., it is closed under addition and scalar multiplication.

Lemma 5. *(Linearity) Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. For every two F-normalised graphs $G, G' \in \mathcal{G}^V$ and $c \in \mathbb{R}_{>0}$, graphs $G + G'$ and $c \cdot G$ are F-normalised and*

$$F(G + G') = F(G) + F(G').$$

Proof (Sketch). In a nutshell, we construct the graph $G + G'$ by creating an independent copy of graph G', denoted $\hat{G}'$, and merging all pairs of corresponding nodes in $G + \hat{G}'$. From Locality and Anonymity (Proposition 2a and 2b), we get that $F_v(G) = F_v(G + \hat{G}')$ and $F_v(G') = F_{f(v)}(G + \hat{G}')$ for an isomorphism f between G' and $\hat{G}'$. Now, MRG implies $F_v(G + G') = F_v(G + \hat{G}') + F_{f(v)}(G + \hat{G}')$. Also, by looking at the out-degree in graph $G + G'$, we get that $G + G'$ is F-normalised.

For $c \cdot G$, the result simply follows from Homogeneity (Proposition 2e). $\square$

In the following definition, we propose a new class of graphs, called *out-unity graphs*. As we will show in Lemma 11, F-normalised out-unity graphs form a basis of all F-normalised graphs without sinks.

Definition 3. *(Out-Unity Graphs[7]) A graph $G \in \mathcal{G}^V$ is an* out-unity graph *if every node $v \in V$ has exactly one outgoing edge. The set of all such graphs will be denoted by:*

$$\mathcal{OUG} = \big\{ G = (V, E, \beta, \omega) \in \mathcal{G} : \forall_{v \in V} |\Gamma^+_G(v)| = 1 \big\}.$$

Observe that each out-unity graph consists of one or more disconnected cycles with in-trees attached to their nodes. See Figure 2 for an example. Furthermore, we note that if centrality measure satisfies OH, then the edge-weight function does not affect the centrality of any node.

Corollary 6. *Let F be a centrality measure that satisfies OH. Then, for every out-unity graph $G = (V, E, \beta, \omega) \in \mathcal{OUG}$: $F(G) = F(V, E, \beta, \mathbf{1})$.*

In the remainder of this section, we will prove that if F satisfies our axioms, then it is the PageRank centrality. Specifically:

---

[7]Wąs and Skibski [2018] studied similar *Unity Graphs* with single *incoming* (not outgoing) edge for every node when discussing other feedback centralities—eigenvector and Katz centralities.
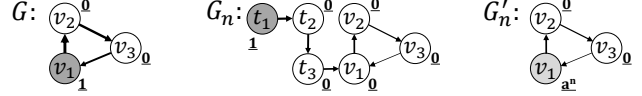
- In Theorem 10 we show that if G is an out-unity graph, then $F(G) = PR(G)$. We do it by first showing that:
  - if G contains an in-tree, then we can change the weight of its root and remove its other nodes, preserving the centralities of the remaining nodes in G (Lemma 7).
  - if G is an in-tree, $F(G) = PR(G)$ (Lemma 8).
  - if G is a cycle graph with a positive weight in only one of its nodes, then $F(G) = PR(G)$ (Lemma 9).
- In Lemma 11, we show that every F-normalised graph without sinks can be deconstructed into several F-normalised out-unity graphs.
- In Theorem 12, by combining Theorem 10 and Lemma 11, we show that $F(G) = PR(G)$ for every graph $G \in \mathcal{G}$.

In the following two lemmas we consider in-trees.

Lemma 7. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. For a graph, $G \in \mathcal{G}^V$, if $G' \in \mathcal{G}^{V'}$ is an in-tree with root v such that $V \cap V' = \{v\}$, then:*

$$F_u(G+G') = F_u\big(G + (\{v\}, \emptyset, F_v(G') \cdot \mathbf{1}, \emptyset)\big) \text{ for every } u \in V.$$

Lemma 8. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN with constant a. If graph $G = \mathcal{G}^V$ is an in-tree, then for every $u \in V$, it holds that:*

$$F_u(G) = \beta(u) + \sum_{w \in P_G(u)} a^{h(w) - h(u)} \beta(w) = PR_u(G).$$

Now, we move to simple cycle graphs.

Lemma 9. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. For every cycle graph $G = (V, E, \beta, \omega)$ such that $\beta(v_1) = 1$ and $\beta(v_i) = 0$ for every $i \in \{2, \dots, n\}$, it holds:*

$$F_{v_i}(G) = \frac{a^{i-1}}{1 - a^n} = PR_{v_i}(G), \ \forall i \in \{1, \dots, n\}.$$

Proof. It is easy to check the equality for PageRank. Thus, we will focus on proving that $F_{v_i}(G) = a^{i-1}/(1 - a^n)$.

Roughly speaking, we will consider graphs consisted of a cycle with a simple in-tree attached to one of its nodes (as graph G_n in Figure 3). Then, using two different methods we will obtain a cycle graph: (1) by removing an in-tree and (2) by merging respective nodes. From Lemma 7 and MRG, this will imply our thesis.

To simplify the notation, we assume $v_{n+1} = v_1$ and will use both symbols interchangeably. For every $k \in \{1, \dots, n\}$ and nodes $T_k = \{t_1, \dots, t_k\}$ we construct a graph G_k in which nodes V form the same cycle and nodes $T_k \cup \{v_{k+1}\}$ compose an in-tree with v_{k+1} as its root, i.e.:

$$G_k = (V \cup T_k, E \cup \{(t_1, t_2), \dots, (t_{k-1}, t_k), (t_k, v_{k+1})\}, \beta_i, \omega_i),$$

where $\beta_k(t_1) = 1$ and $\beta_k(u) = 0$ for any $u \in V \cup T_k \setminus \{t_1\}$ and ω_k is such that G_k is F-normalised (for an example with $k = n = 3$ see Figure 3). From Lemma 8 we know that the centrality of in-tree formed by $T_k \cup \{v_{k+1}\}$ is equal to a^k.

Now, to use Lemma 7, let us consider graph G'_k (see Figure 3), defined as

$$G'_k = (V, E, \beta'_k, \omega'_k) = G_k[V] + (\{v_{k+1}\}, \emptyset, a^k \cdot \mathbf{1}, \emptyset).$$

Then, $F_u(G'_k) = F_u(G_k)$ for every $u \in V$. Now, notice that both G'_k and G are cycle graphs with n nodes among which only one have positive weight. Thus taking graph $(1/a^k)G'_k$ and adjusting its edge weights we can obtain a graph isomorphic to G. Therefore, from Corollary 6, Anonymity (Proposition 2b) and Homogeneity (Proposition 2e) we have

$$F_{v_{k+1}}(G_k) = F_{v_{k+1}}(G'_k) = a^k F_{v_1}(G). \tag{3}$$

On the other hand, consider sequentially merging nodes $t_k, \ldots, t_1$ with $v_k, \ldots, v_1$ in the graph G_k. This way, we obtain graph $\hat{G} = (V, E, \beta, \hat{\omega})$ for some $\hat{\omega}$, i.e., a graph that differs (or may differ) from G only in weights of edges. Thus, from MRG and Corollary 6 we get:

$$F_{v_{k+1}}(G) = \begin{cases} F_{v_1}(G_n) + F_{t_1}(G_n) & \text{if } k = n, \\ F_{v_{k+1}}(G_k) & \text{otherwise.} \end{cases}$$

As $F_{t_1}(G_n) = 1$, from FOR and DN, equation (3) yields $F_{v_{k+1}}(G) = a^k F_{v_1}(G)$ for $k \in \{1, \ldots, n-1\}$ and $F_{v_1}(G) = 1 + a^n F_{v_1}(G)$ which concludes the proof. $\square$

As every out-unity graph is consisted of one or more cycles with possible in-trees attached to their nodes, Lemmas 7-9 are enough to characterize centrality measure on out-unity graphs.

Theorem 10. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. For every F-normalised out-unity graph $G \in \mathcal{OUG}$, $F(G) = PR(G)$.*

Proof (Sketch). Let us fix graph $G = (V, E, \beta, \omega)$ and a node, $v \in V$. We will show that $F_v(G) = PR_v(G)$. If v does not belong to any cycle, then it is a part of an in-tree. Thus, from FOR and Lemma 8 the thesis follows.

Assume otherwise, and let $C \subseteq V$ be a subset of nodes such that $v \in C$ and $G[C]$ is a cycle graph. By using Lemma 7 and Lemma 8 we can construct graph $G' = (C, E[C], \beta', \omega[C])$ such that $F_v(G) = F_v(G')$ for every $v \in C$. From Corollary 6 and Linearity (Lemma 5) we can split G' into $|C|$ graphs, $\beta'_u \cdot G'_u$, where graph G'_u for every $u \in C$ is a cycle graph where only node u has a positive weight equal to 1. From Lemma 9 we know that $F_v(G'_u) = PR_v(G'_u)$ and as Linearity (Lemma 5) also holds for PageRank centrality we obtain the thesis. $\square$

In the following lemma we show that every F-normalised graph can be deconstructed into out-unity graphs.

Lemma 11. *Let F be a centrality measure that satisfies FOR, OH, MON, MRG, and DN. For every F-normalised graph $G \in \mathcal{G}^V$ without sinks, there exist F-normalised out-unity graphs $G_1, \ldots, G_k \in \mathcal{G}^V$ such that $G = G_1 + \cdots + G_k$.*

Proof. Let $G = (V, E, \beta, \omega)$. We will prove this lemma by induction on $|E| - |V|$. If $|E| - |V| = 0$, then G is already an out-unity graph. Assume $|E| - |V| > 0$. Let v be a node with more than one outgoing edge, and let $e \in \Gamma_G^+(v)$ be its arbitrary edge. Consider graphs:

$$G_e = G - (\Gamma_G^+(v) \setminus \{e\}), \quad G_r = G - \{e\}.$$

Now, consider F-normalisations of G_e and G_r: $\tilde{G}_e$ and $\tilde{G}_r$. Since all nodes have outgoing edges in G, then they also have in G_e and G_r. Thus, $f^+(G_e) = G_e$ and $f^+(G_r) = G_r$, which means $f^* \circ f^- \circ f^+$ does not add new nodes or edges to both graphs. Hence, both graphs have less edges than G. Now, consider graph G':

$$G' = c_e \tilde{G}_e + c_r \tilde{G}_r, \quad c_e = \frac{\deg_{G_e}^+(v)}{F_v(G_e)}, c_r = \frac{\deg_{G_r}^+(v)}{F_v(G_r)}.$$

In the remainder of this proof we will show that $G' = G$; since both $c_e \tilde{G}_e$ and $c_r \tilde{G}_r$ are F-normalised and have less edges than G, this will conclude the inductive step.

Let $G' = (V, E', \beta', \omega')$. We begin by showing that $E' = E$. We already argued that f^+ applied to G_e and G_r does not add any edges. Function f^* does not modify the set of edges. Thus, it remains to consider function f^-, which removes outgoing edges of nodes which have zero centrality. From Zero Nodes (Proposition 2c) we know that every node in G either has a predecessor with a positive weight or has the positive weight itself. Since the same is true in at least one of the graphs G_e and G_r, f^- will not remove the same edge from both graphs.

Clearly, $\beta' = (c_e + c_r)\beta$. Consider ω'. First, we focus on outgoing edges of node v: $\Gamma_{G'}^+(v)$. Weights c_e and c_r are defined in a way that $\omega'(v, u) = \omega(v, u)$ for every $(v, u) \in \Gamma_G^+(v)$ (note that $\deg_{\tilde{G}_e}^+(v) = \deg_{G_e}^+(v)/c_e$). Since both G and G' are F-normalised, this implies:

$$F_v(G) = \deg_G^+(v) = \deg_{G'}^+(v) = F_v(G').$$

Now, consider outgoing edges of node $u \in V \setminus \{v\}$: $\Gamma_{G'}^+(u)$. We see that in each graph $G_e, G_r, \tilde{G}_e, \tilde{G}_r$, and G' the relation between weights of any two edges is the same as in G. Thus, from OH, we can adjust them so that $\deg_{G'}^+(u) = (c_e + c_r) \cdot \deg_G^+(u)$. By doing so, for every $u \in V \setminus \{v\}$ and v we get:

$$F(G') \stackrel{\text{OH}}{=} F(V, E, (c_e + c_r)\beta, (c_e + c_r)\omega) \stackrel{\text{P2e}}{=} (c_e + c_r)F(G).$$

But we have already proved that $F_v(G) = F_v(G')$; therefore, $c_e + c_r = 1$, $\beta' = \beta$, and—since $\deg_{G'}^+(u) = F_u(G') = F_u(G) = \deg_G^+(u)$—also $\omega' = \omega$. Hence, $G = G'$. $\square$

We are now ready to state our main theorem.

Theorem 12. *If centrality measure satisfies FOR, OH, MON, MRG, and DN, then it is a PageRank centrality.*

Proof. Assume that F satisfies FOR, OH, MON, MRG, and DN. Let $G \in \mathcal{G}^V$ be an arbitrary graph and $\tilde{G}$ its F-normalisation. On behalf of Lemma 4, we know that $F_v(\tilde{G}) = F_v(G)$ for every $v \in V$. If there exists a sink, denoted by v, in $\tilde{G}$, then from the fact that $\tilde{G}$ is F-normalised we

know that v has zero centrality. From Zero Nodes (Proposition 2c) and Removing Zero Nodes (Proposition 2d) we know that $F_v(\tilde{G}) = 0$ and $F_u(\tilde{G}) = F_u(\tilde{G}[V \setminus \{v\}])$ for every $u \in V \setminus \{v\}$. Thus, we can assume that all nodes have outgoing edges. Then, by using Lemma 11, we take out-unity graphs $G_1, \ldots, G_k$ such that $\tilde{G} = G_1 + \cdots + G_k$. From Linearity (Lemma 5) we have $F(\tilde{G}) = F(G_1) + \cdots + F(G_k)$.

Now, for each $i \in \{1, \ldots, k\}$ we know that $F(G_i) = PR(G_i)$ on behalf of Theorem 10. Observe that graph G_i is also PR-normalised. Since the PageRank centrality also satisfies Linearity we get that $PR(\tilde{G}) = PR(G_1) + \cdots + PR(G_k)$. Moreover, none of the functions f^*, f^-, f^+ influence the PageRank centrality of nodes in a graph. Thus, $PR_v(\tilde{G}) = PR_v(G)$ for every $v \in V$ and

$$F_v(G) = F_v(\tilde{G}) = PR_v(\tilde{G}) = PR_v(G).$$

This concludes the proof of Theorem 12. $\qquad \square$

We note that all five axioms are independent from each other, i.e., there exist a centrality measure that satisfies any four of them and does not satisfy the fifth one.

5 Conclusions

In this paper, we proposed an axiomatization of the PageRank centrality. Specifically, we proved that if a centrality measure satisfies five simple axioms—Foreseeability, Outgoing Homogeneity, Monotonicity, Merging and Dummy Node—then it is the PageRank centrality. Our characterization is the first one of the PageRank in its original, general form.

In our future work, we plan to study other centrality measures based on random walks in a graph, such as the random walk closeness centrality.

Acknowledgements

References

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

[Bianchini et al., 2005] Monica Bianchini, Marco Gori, and Franco Scarselli. Inside pagerank. *ACM Transactions on Internet Technology (TOIT)*, 5(1):92–128, 2005.

[Bollen et al., 2006] Johan Bollen, Marko A Rodriquez, and Herbert Van de Sompel. Journal status. *Scientometrics*, 69(3):669–687, 2006.

[Bonacich, 1972] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.

[Brandes and Erlebach, 2005] Ulrik Brandes and Thomas Erlebach. *Network analysis: Methodological foundations*. Springer-Verlag, 2005.

[Costa et al., 2007] L da F Costa, Francisco A Rodrigues, Gonzalo Travieso, and Paulino Ribeiro Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1):167–242, 2007.

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

[Grando et al., 2016] Felipe Grando, Diego Noble, and Luis C Lamb. An analysis of centrality measures for complex and social networks. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2016.

[Haveliwala, 2002] Taher H Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.

[Iván and Grolmusz, 2010] Gábor Iván and Vince Grolmusz. When the web meets the cell: using personalized PageRank for analyzing protein interaction networks. *Bioinformatics*, 27(3):405–407, 2010.

[Koschützki et al., 2005] Dirk Koschützki, Katharina Anna Lehmann, Dagmar Tenfelde-Podehl, and Oliver Zlotowski. Advanced centrality concepts. In *Network Analysis*, volume 3418 of *Lecture Notes in Computer Science*, pages 83–111. Springer, 2005.

[Langville and Meyer, 2011] Amy N Langville and Carl D Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.

[Newman, 2010] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.

[Page et al., 1999] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: bringing order to the web. *Stanford InfoLab*, 1999.

[Palacios-Huerta and Volij, 2004] Ignacio Palacios-Huerta and Oscar Volij. The measurement of intellectual influence. *Econometrica*, 72(3):963–977, 2004.

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

[Skibski et al., 2016] Oskar Skibski, Talal Rahwan, Tomasz P. Michalak, and Makoto Yokoo. Attachment centrality: An axiomatic approach to connectivity in networks. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 168–176, 2016.

[Weng et al., 2010] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM)*, pages 261–270, 2010.

[Wąs and Skibski, 2018] Tomasz Wąs and Oskar Skibski. An axiomatization of the eigenvector and Katz centralities. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1258–1265, 2018.