Discrete Event Dyn Syst
DOI 10.1007/s10626-017-0248-7

Ranking nodes in general networks: a Markov multi-chain approach

Joost Berkhout[1] · Bernd F. Heidergott[2]

Received: 28 September 2016 / Accepted: 31 March 2017

Abstract The basis of Google's acclaimed PageRank is an artificial mixing of the Markov chain representing the connectivity structure of the network under study with a maximally connected network where every node is connected to every other node. The rate with which the original network is mixed with the strongly connected one is called damping factor. The choice of the damping factor can influence the ranking of the nodes. As we show in this paper, the ranks of transient nodes, i.e., nodes not belonging to a strongly connected component without outgoing links in the original network, tend to zero as the damping factor increases. In this paper we develop a new methodology for obtaining a meaningful ranking of nodes without having to resort to mixing the network with an artificial one. Our new ranking relies on an adjusted definition of the ergodic projector of the Markov chain representing the original network. We will show how the new ergodic projector leads to a more structural way of ranking (transient) nodes. Numerical examples are provided to illustrate the impact of this new ranking approach.

Keywords Ranking nodes in networks · Markov multi-chains · Random surfer · Google's PageRank

---

This article is part of the Topical Collection: *Special Issue on Performance Analysis and Optimization of Discrete Event Systems*
Guest Editors: Christos G. Cassandras and Alessandro Giua

Part of this work was done while Joost Berkhout was with Vrije Universiteit Amsterdam

✉ Joost Berkhout
joost.berkhout@cwi.nl

Bernd F. Heidergott
b.f.heidergott@vu.nl

1 Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands

2 Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Published online: 27 May 2017

1 Introduction

Google's PageRank was developed in 1998 (Brin and Page 1998) to provide a ranking of webpages in terms of importance based on the underlying hyperlink-network of the web. The success of the PageRank algorithm in Google's search engine, its simplicity, the generality of the mathematics, its uniqueness and guaranteed existence are features that ensured that PageRank is nowadays widely used in a variety of other application areas, including in social and behavioral science, economics, marketing, biology and security. For a recent review concerning PageRank and its applications see Gleich (2015), other reviews are (Berkhin 2005; Franceschet 2011; Langville and Meyer 2004; Bianchini et al. 2005). Google's PageRank is an eigenvector method, for a survey of this type of methods in the light of the web see (Langville and Meyer 2005). From an even broader perspective, Google's PageRank falls into the class of centrality measures for networks, for a recent mathematical classification see Boldi and Vigna (2014).

PageRank on a general network can be explained via the so-called random surfer model. In this model a random surfer randomly chooses an available edge at her current location and moves to the connected node. At arrival she again chooses a new node based on the available edges and continues moving towards the new node, and so forth. Most real-life networks are not connected, e.g., the internet web as was shown in Broder et al. (2000). In order to make a ranking over all nodes possible for reducible networks, PageRank has been extended so as to model a *bored* random surfer on the network: before choosing a new node to move to, she flips a d-coin; with probability d she randomly moves to a new node based on the network structure, and with probability $1 - d$ she jumps (possibly random) to a page independent of her current location. The probability d is an input parameter of the PageRank algorithm and is called *damping factor*. The resetting of the random search as modeled by the damping factor can be interpreted as model for the behavior of a typical human internet-surfer who "gets bored" and abandons her search.

A well known problem of Google's PageRank is that the choice of d may highly effect the ranking and computational performance, see, for example, (Langville and Meyer 2011). A choice of d near 1 places much greater emphasis on the network structure and therefore, perhaps, gives a "truer" ranking. On the other hand, a value of d near 1 gives in to the convergence speed of the power method often used for computation as observed in (Avrachenkov et al. 2008; Langville and Meyer 2003). Moreover, sensitivity issues may arise when d is large, which means that the PageRank scores becomes more and more sensitive to small changes in P, for more details see Section 4.4 in Langville and Meyer (2003). In general, the choice of d affects the resulting ranking, (Bressan and Peserico 2010; Fu et al. 2006), especially on the subtop of the ranked nodes which by the implementation of PageRank in Google's search engine may have significant consequences, see also (Langville and Meyer 2003). As comes forward in Gleich (2015), choosing appropriate values for d allows for tailoring Google's PageRank to particular application areas. However, how to choose the factor in general remains an open question.

In this paper, the focus is on the unclear score trade-off made between certain type of nodes based on the network structure when choosing d in Google's PageRank. Arguable important nodes of transient nature can get little score when choosing d inaccurate as was also observed in, e.g., (Avrachenkov and Litvak 2006; Boldi et al. 2005, 2009). Nodes of transient nature are common in real-life data, e.g., in the web network as shown in Broder et al. (2000) and voting networks such as for the Wikipedia example from Berkhout (2016). To the best of our knowledge, this structural dependency is underexposed in Google's PageRank literature.

Focusing on the network structure, this paper proposes a random-surfer-based ranking that does not depend on a user-dependent non-transparent factor that is critical. Our approach can be described in a nutshell as follows. First, rankings are obtained for transient nodes (transient states in Markov chain terms), and non-transient strongly connected components (in short s.c.c. or ergodic classes in Markov chain terms), separately. These rankings are based on the scores obtained from the newly introduced *extended ergodic projector*. Then, via a majority principle, these rankings can be combined to obtain a ranking over the whole network. Alternatively, personalized rankings can be efficiently computed via the extended ergodic projector.

This paper is the extended version of the conference paper (Berkhout 2016), and results are proved in this paper more rigorously and statements are richer. In addition, new results on the impact of the damping factor in Google's PageRank are provided and the concept of extended ergodic projector from Berkhout (2016) is generalized, which allows a more careful score transfer from transient nodes to non-transient nodes (in Markov chain terms 'transient states' and 'ergodic states', respectively). Furthermore, the paper reports on applications of our new ranking concept to a real-life network.

This paper is organized as follows. The mathematical details of Google's PageRank and Markov multi-chains are discussed in Section 2. After analyzing the impact of the damping factor in Google's PageRank, the Generalized Ranking is introduced in Section 3. In that same section the Generalized Ranking is analyzed. A real-life numerical example is presented in Section 4 which illustrates the practical impact. Section 5 concludes the paper.

2 Preliminaries on Google's PageRank and Markov multi-chains

General networks can typically be modeled as directed graphs with weighted edges. In particular, let $G = (V, E)$ describe a directed graph with finite vertex set V and edge set $E \subseteq V \times V$. By convention, $|V| = n$ and $|E| = m$ (where $|A|$ indicates the cardinality of a set A). All edges are weighted by a positive function w based on the underlying network dynamics[1], i.e., $w : E \to (0, \infty)$. An edge $e = (i, j)$ means that there is a directed relation from i to j and the relation-strength described by weight $w(e)$. E.g., in a social network $w(i, j) > w(i, k)$ might imply that social agent i is more engaged with agent j than with agent k. In the definition of PageRank in Berkhout (2016) it holds that $w(e) = 1$ for all $e \in E$, meaning that the analysis here is more general.

The location of a random surfer[2] on the graph-modeled-network will be described by a discrete-time Markov chain $(X_t : t = 0, 1, 2, \dots)$ on state space V. Particularly, $X_t \in V$ is the location of the random surfer at time t. Write P as the transition matrix of the Markov chain of which the (i, j)-th element is given by

$$P(i, j) = \frac{w(i, j)}{\sum_{(i,k)\in E} w(i, k)}, \quad \text{for all } (i, j) \in E,$$

and $P(i, j) = 0$ for all $(i, j) \notin E$. In case $\sum_{j\in V} P(i, j) = 0$, i.e., the so-called 'dangling' node i has no directed edges towards other nodes, we set $P(i, i) = 1$ as the authors believe

[1] Positivity of w is needed for the row normalization later on. Possibly a scaling is required to make the weights positive in practical applications.

[2] The term *random walk* also applies, but since the ranking is inspired by Google's PageRank we will use their concept of random surfer.

that this adjustment, to ensure that P is a stochastic matrix, leads to an interpretation most close to that of a random surfer on the original network structure. Alternatively one might replace the i-th row of P with $v^\top$, where v represents some preference of the random surfer on the nodes, or a uniform distribution. Another option is to direct all dangling nodes towards an artificial node, i.e., making all dangling nodes transient of nature. For an analysis of the different approaches of dealing with dangling nodes see (Bianchini et al. 2005). We would like to emphasize that the analysis in this paper does not depend on the particular choice of how to deal with dangling nodes. In any case, networks remain reducible and all arguments regarding dangling nodes still apply to connected groups of nodes without any outgoing edges. For example, two connected nodes without any outgoing links suffer from the same problem as dangling nodes; a random surfer might get stuck inside this component in the long-run.

The t-step transition probabilities are given by $P^t = P^{t-1}P$, for $t \in \mathbb{N}$, in which

$$(P^t)(i,j) = \Pr(X_t = j \mid X_0 = i), \quad \text{for all } (i,j) \in V \times V.$$

Let us assume for simplicity that Markov chain $(X_t : t = 0, 1, 2, \dots)$ is aperiodic, else, without affecting the results too much, mix P with a small proportion of the identify matrix I. The so-called ergodic projector Π_P is obtained via

$$\Pi_P = \lim_{t \to \infty} P^t, \tag{1}$$

and its (i,j)-th element, i.e., $\Pi_P(i,j)$, describes the long-term probability that the random surfer can be found in node $j \in V$ when starting in node $i \in V$. It is known, see for example (Aldous and Fill 2002), that irreducibility together with aperiodicty of the Markov chain implies that all rows of Π_P are equal to the unique stationary distribution $\pi_P^\top$ that can be solved from

$$\pi_P^\top P = \pi_P^\top, \quad \text{and } \pi_P^\top \bar{1} = 1, \tag{2}$$

where $\bar{1}$ is a column vector with n ones.

Google's PageRank (Brin and Page 1998) modifies the transition matrix P so that the Markov chain is guaranteed to be irreducible and thus is known to have a unique stationary distribution. To that end, PageRank considers a *bored* random surfer that at each discrete time step gets bored with probability $1 - d$, with $d \in [0, 1)$. Once bored the random surfer randomly selects a new node from the network according to some personalization vector v where we assume that v is stochastic, i.e., v is a probability vector on V. Specifically, PageRank constructs this irreducible transition matrix P_d out of P via

$$P_d := P_d(v) = dP + (1-d)\bar{1}v^\top, \tag{3}$$

where $d \in [0, 1)$, and v is a stochastic vector. By construction, P_d for $d \in [0, 1)$ is irreducible and has a unique stationary distribution $\pi_{P_d}^\top$. Initially, the founders of Google's PageRank suggested a uniform stochastic vector for v. Later this was relaxed to general stochastic vectors to allow for more flexibility. As noted in (Langville and Meyer 2005), the choice of v does not affect the mathematical nor computational aspects, but it does alter the ranking in a predictable way. In particular, via v the user can discriminate between nodes in the network. This can be used, e.g., to control spamming on the internet which is the result of users setting up dummy websites to fool the ranking mechanisms and purposefully increase the ranking of certain websites, see also (Langville and Meyer 2003). Once the dummy websites are uncovered, the corresponding values in v can be made small to diminish their impact. Moreover, based on the interest of a particular user, the weights in v can be shifted to ensure a personalized ranking.

In contrast to PageRank we do not consider a *bored* random surfer, i.e., the transition matrix P is not modified so that the Markov chain may fail to be irreducible and has no unique stationary distribution satisfying (2). An aperiodic reducible Markov chain with multiple closed sets of communicating states, in short ergodic classes, and possible transient states will be called a Markov multi-chain. Without losing on generality, after relabeling the states, P and Π_P of a Markov multi-chain can be rewritten in the following canonical forms, respectively

$$P = \begin{bmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & P_{\mathcal{E}} & 0 \\ P_{T1} & P_{T2} & \cdots & P_{T\mathcal{E}} & P_{TT} \end{bmatrix} \quad \text{and} \quad \Pi_P = \begin{bmatrix} \Pi_1 & 0 & 0 & \cdots & 0 \\ 0 & \Pi_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \Pi_{\mathcal{E}} & 0 \\ R_1 & R_2 & \cdots & R_{\mathcal{E}} & 0 \end{bmatrix}, \tag{4}$$

where $\mathcal{E}$ is the number of ergodic classes. For the i-th ergodic class, P_i gives the one step transition probabilities between ergodic states from the i-th ergodic class, P_{Ti} gives the one step probabilities from transient states towards ergodic states from ergodic class i, and Π_i gives a square matrix of which all rows equal the unique stationary distribution of the chain inside the i-th ergodic class, i.e., all rows in Π_{P_i} equal $\pi_{P_i}^\top$ which is the unique distribution satisfying $\pi_{P_i}^\top P_i = \pi_{P_i}^\top$. In the canonical form of Π_P, $R_i(j,k)$ gives the equilibrium probability of visiting ergodic state k (which is part of the i-th ergodic class) when starting in transient state j. It holds, see, e.g., (Berkhout and Heidergott 2014), that

$$R_i = (I - P_{TT})^{-1} P_{Ti} \bar{1} \pi_{P_i}^\top,$$

where I is an appropriate sized identity matrix, and $\bar{1}$ is an appropriate sized vector of ones (of size equal to the number of states in ergodic class i). If a Markov multi-chain P is known to start with a fixed initial distribution $\mu^\top$, then the unique stationary distribution of the system becomes $\mu^\top \Pi_P$.

Via the following diagonal criterion from (Berkhout and Heidergott 2015) one can classify a state via the diagonal elements of Π_P: When $\Pi_P(i,i) > 0$ state i is an ergodic state else state i is transient. Based on the nature of the corresponding states, i.e., transient or ergodic, we subdivide all nodes from V into two sets. Set V_T denotes the set of all 'transient' nodes, i.e.,

$$V_T = \{i \in V : \Pi_P(i,i) = 0\},$$

and similarly, V_E denotes the set of all 'ergodic' nodes, i.e., $V_E = V \setminus V_T$. Using this notation the number of transient (ergodic) states based on graph G is $|V_T|$ ($|V_E|$).

The alternative page ranking builds on the deviation matrix, where the deviation matrix D_P of P is given by

$$D_P = \sum_{t=0}^{\infty} (P^t - \Pi_P), \tag{5}$$

or alternatively

$$D_P = (I - P + \Pi_P)^{-1} - \Pi_P,$$

Existence of D_P is guaranteed for finite-state aperiodic Markov chains, see (Heidergott et al. 2007) for a proof. It holds that $D_P + \Pi_P$ equals the fundamental matrix, (Kemeny and Snell 1976). As was observed in (Meyer 1975), wherever the fundamental matrix appears it can be directly replaced by D_P. Furthermore, the deviation matrix is known as the group inverse of $I - P$ in the literature (Meyer 1975; 1982). In (Meyer 1975) (Section 5) it is shown that the computation of D_P can be done efficiently without knowing Π_P in advance in case of aperiodic and irreducible Markov chains; see also (Xia and Glynn 2016) for a recent paper.

Denote the transient part of D_P by $D_{P_{TT}}$. Then it follows from Eq. 5 that

$$D_{P_{TT}} = \sum_{t=0}^{\infty}((P_{TT})^t - \underbrace{\Pi_{P_{TT}}}_{=0}) = \sum_{t=0}^{\infty}(P_{TT})^t = (I - P_{TT})^{-1}, \tag{6}$$

where we used that $\Pi_{P_{TT}} = 0$ by the transient nature, see also Eq. 4. Note that all elements of $D_{P_{TT}}$ are non-negative since all elements of $(P_{TT})^t$ for $t = 0, 1, 2, \ldots$ are non-negative. A probabilistic interpretation for $D_{P_{TT}}$ is as follows: for transient states i and j, $D_{P_{TT}}(i, j)$ is the expected total number of times that the process visits state j when starting in state i. For details see (Kemeny and Snell 1976).

The goal of analyzing the long-term behavior of a random surfer is to score the nodes such that a ranking can be obtained. We will now formalize how - based on a general score vector π - one can obtain a ranking. Let $\pi_{[i]}$ denote the i-th order statistic (in descending order) of score vector π. Then the j-th element of ranking $r(\pi)$, denoted by $r_j(\pi)$, is given through π by

$$r_j(\pi) = i \;\Leftrightarrow\; \pi(i) = \pi_{[j]}, \;\; j \in V,$$

where we assume that ties are arbitrarily broken. Specifically for Google's PageRank (GP) it holds

$$r_{\mathrm{GP}} := r(\pi_{P_d(v)}),$$

with $P_d(v)$ as in Eq. 3, i.e., r_{GP} is a function of P, d and v.

3 Generalized ranking

Before introducing the extension of the Generalized Ranking from (Berkhout 2016) we will first discuss the shortcomings of Google's PageRank in more detail by analyzing the impact of damping factor d.

3.1 Analysis of Google's PageRank

The following proposition provides a new perspective on Google's PageRank. It essentially follows from Theorem 3.1 in (Berkhout 2016) but with a different focus. It will be used to uncover some properties of Google's PageRank and to relate it later-on to the new ranking. Moreover, in contrast to what was stated in (Berkhout 2016) it is not necessary that all elements in the personalization vector v are greater than zero.

Proposition 1 *For Google's PageRank's $P_d(v)$ as defined in Eq. 3, with v a stochastic vector and $d \in [0, 1)$, it holds that the ergodic projector is equal to*

$$\Pi_{P_d(v)} = \bar{1}v^\top \mathbb{E}[P^{X_{1-d}}] = \bar{1}v^\top (1-d)\sum_{t=0}^{\infty} d^t P^t, \tag{7}$$

where X_{1-d} is a geometrically distributed random variable with success parameter $1 - d$.

Proof Using that the rows of $P_d(v)$ sum up to one shows that

$$(dP + (1-d)\bar{1}v^\top)\bar{1}v^\top = \bar{1}v^\top,$$

and thus it holds for finite $t \in \mathbb{N}$ that

$$(P_d)^t = (dP)^t + (1-d)\bar{1}v^\top \sum_{i=0}^{t-1}(dP)^i.$$

Letting in the previous equation $t \to \infty$ gives for $d \in [0, 1)$

$$\Pi_{P_d} = \lim_{t\to\infty}(P_d)^t = \bar{1}v^\top(1-d)\sum_{i=0}^{\infty}(dP)^i,$$

which can be rewritten using the definition of X_{1-d}. □

Equation 7 shows that Π_{P_d} has equal rows. In words: the stationary distribution is equal for every initial state, i.e., ultimately the chain will reach the unique ergodic class with a unique stationary distribution. Equality of rows of an ergodic projector is a fundamental property of a Markov uni-chain which leads to the following corollary.

Corollary 1 *The Markov chain with P_d has only one ergodic class and possibly transient states for all $d \in [0, 1)$.*

It follows from Eq. 7 that Google's PageRank can be written in terms of the so-called resolvent from (Kartashov 1996) with discounting factor d. Particularly, the resolvent of P as defined in (Kartashov 1996) is

$$(1-d)\sum_{k=0}^{\infty}(dP)^k.$$

The resolvent of P can be rewritten as $E[P^{X_{1-d}}]$, which via (1) intuitively explains the following limit property

$$\lim_{d\uparrow 1}(1-d)\sum_{k=0}^{\infty}(dP)^k = \Pi_P,$$

for a formal proof see (Kartashov 1996). Inserting this limit result into Eq. 7 yields

$$\lim_{d\uparrow 1}\Pi_{P_d} = \bar{1}v^\top\Pi_P, \tag{8}$$

which shows by the canonical form of Π_P for Markov multi-chains, see Eq. 4, that transient states do not receive any score in this extreme case with respect to $d \in [0, 1)$. For the other extreme case regarding $d \in [0, 1)$ it directly follows from Eq. 7 that

$$\text{for } d = 0: \ \Pi_{P_d} = \bar{1}v^\top,$$

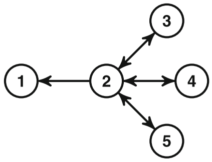
i.e., the scores are completely based on personalization vector v (and independent of P).

A value for d commonly recommended in the literature is $d = 0.85$, which originates from (Brin and Page 1998). As noticed in (Langville and Meyer 2011), this seems to be a "workable compromise between efficiency and effectiveness"; taking a large fraction of the underlying network into account as well as providing a reasonable convergence rate for computation with the power method. However, (Gleich 2015) shows that different instances require different values for d to ensure an effective analysis with Google's PageRank. In (Avrachenkov et al. 2008) perturbation analysis was used to study the choice for d which satisfies certain fairness criterion. In particular, backed by a numerical study, the authors argue that a value of d approximately equal to 1/2 can mitigate the score trade-off issues for certain instances, such as for the web and other large networks. A similar value of $d =$

1/2 was recommended in (Bressan and Peserico 2010) for certain type of applications. Moreover, in (Bressan and Peserico 2010) it is shown that a graph exists for which the top k nodes assume all possible $k!$ orderings as the damping factor varies in an arbitrarily small interval, (e.g., $[0.85 - \epsilon, 0.85 + \epsilon]$ for a $\epsilon > 0$). Although being an academic theoretical graph, it shows that d can have significant impact.

The following example illustrates that a fixed choice of d may lead to unsatisfactory rankings and, moreover, it illustrates the impact of the choice for d.

Example 1 Consider the following graph $G = (V, E)$ consisting of $n = 5$ nodes:



and for which $w(e) = 1$ for all $e \in E$. Since node 1 is a dangling node, i.e., $\sum_{j\in V} P(1, j) = 0$, we set $P(1, 1) = 1$. For personalization vector

$$v = [1\ 1\ 1\ 1\ 1]^\top/5$$

this leads to

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \text{ and } P_d(v) = \begin{bmatrix} \frac{1+4d}{5} & \frac{1-d}{5} & \frac{1-d}{5} & \frac{1-d}{5} & \frac{1-d}{5} \\ \frac{4+d}{20} & \frac{1-d}{5} & \frac{4+d}{20} & \frac{4+d}{20} & \frac{4+d}{20} \\ \frac{1-d}{5} & \frac{1+4d}{5} & \frac{1-d}{5} & \frac{1-d}{5} & \frac{1-d}{5} \\ \frac{1-d}{5} & \frac{1+4d}{5} & \frac{1-d}{5} & \frac{1-d}{5} & \frac{1-d}{5} \\ \frac{1-d}{5} & \frac{1+4d}{5} & \frac{1-d}{5} & \frac{1-d}{5} & \frac{1-d}{5} \end{bmatrix}.$$

It holds for $d \in [0, 1)$ that

$$\pi^\top_{P_d(v)} = \left[\frac{4+d}{5(4-3d^2)} \ \ \frac{4+8d-12d^2}{5(4-3d^2)} \ \ \frac{(1-d)(4+d)}{5(4-3d^2)} \ \ \frac{(1-d)(4+d)}{5(4-3d^2)} \ \ \frac{(1-d)(4+d)}{5(4-3d^2)} \right],$$

which is plotted in Fig. 1.

Figure 1 shows that increasing d ultimately leads to less mass for the transient states compared to the ergodic state. Furthermore, it illustrates that different values of d lead to different rankings. In particular, for $d \in (0.6, 1]$ state 1 is ranked first while intuitively looking at the in-degrees, state 2 is the authority and should be ranked first.

Although Example 1 is simple, the principle findings can be generalized to more realistic graphs. The key observation is that in choosing d a trade-off is made between the relative importance of the different classes of the reducible Markov multi-chain, a feature that best to the author's knowledge is underexposed in Google's PageRank literature. Even though no perfect ranking exists as was proven by Arrow's impossibility theorem, (Langville and Meyer 2012), we believe that taking the Markov chain structure into account leads to intuitively better rankings.

For now let us focus on the fraction of the total score awarded by Google's PageRank to transient states. Denote this d-dependent fraction (Fr) of the total score by $\mathrm{Fr}_{\mathrm{GP}}(d)$ and assume that $v = \bar{1}/n$, i.e., v is a uniform stochastic vector. It follows from Proposition 1 that

$$\mathrm{Fr}_{\mathrm{GP}}(d) = \frac{1-d}{n}\bar{1}^\top \sum_{t=0}^{\infty} d^t (P_{TT})^t \bar{1}, \tag{9}$$

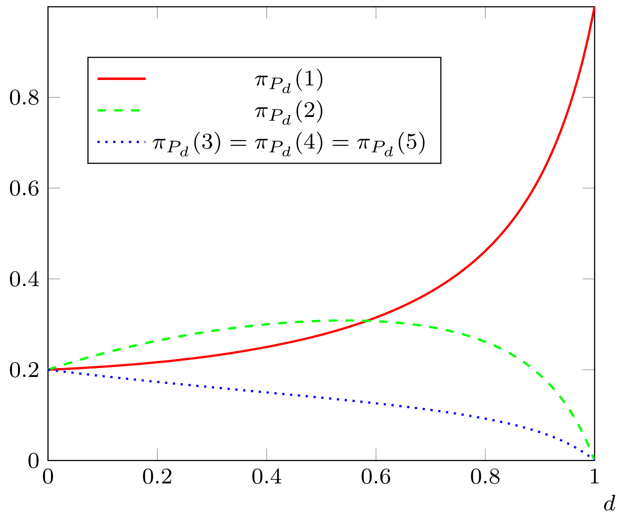


Fig. 1 Plot of $\pi^\top_{P_d(v)}$, $d \in [0, 1]$, for Example 1 in case v is a uniform stochastic vector

where $\bar{1}$ denotes an appropriate sized column vector of ones, i.e., in this case of size $|V_T|$, and recall that P_{TT} denotes the transient part of P (see also (4)). The following proposition explores properties regarding $\text{Fr}_{\text{GP}}(d)$ from Eq. 9.

Proposition 2 *For $v = \bar{1}/n$ it holds that*

(i) $\text{Fr}_{\text{GP}}(0) = \frac{|V_T|}{n}$
(ii) $\lim_{d\uparrow 1} \text{Fr}_{GP}(d) = 0$
(iii) $\forall d \in [0, 1): \quad \text{Fr}_{\text{GP}}{}'(d) < 0$
(iv) $\forall d \in [0, 1): \quad \text{Fr}_{\text{GP}}{}''(d) \leq 0$
(v) $\forall d \in [0, 1): \quad \text{Fr}_{\text{GP}}(d) \geq \frac{(1-d)|V_T|}{n}$,

where $Fr'_{GP}(d)$ and $Fr''_{GP}(d)$ denote the first and second order derivatives of $Fr_{GP}(d)$ w.r.t. d, respectively.

Proof Part (i) follows immediately from the definition in Eq. 9 by noting that

$$\bar{1}^\top (P_{TT})^0 \bar{1} = \bar{1}^\top \bar{1} = |V_T|,$$

and (ii) is a consequence of the analysis regarding (8) (recall that $\Pi_{P_{TT}} = 0$).

To prove part (iii) one shows that

$$\text{Fr}'_{\text{GP}}(d) = \frac{1-d}{n} \sum_{t=1}^{\infty} \left(t - \frac{d}{1-d} \right) d^{t-1} a_t - \frac{|V_T|}{n}, \tag{10}$$

where we defined the sequence $(a_n)_{n=0}^{\infty}$ given by

$$a_t = \bar{1}^\top (P_{TT})^t \bar{1}, \quad t \geq 0.$$

Note that $a_0 = |V_T|$ and by the transient nature it holds that $a_0 > a_t$ for all $t \in \mathbb{N}$. Moreover, it holds that

$$a_{t+1} \leq a_t, \quad \forall t \in \mathbb{N} \tag{11}$$

which follows by rewriting $a_{t+1} = \bar{1}^\top (P_{TT})^t P_{TT} \bar{1}$ and observing that all elements of non-negative vector $P_{TT}\bar{1}$ are ≤ 1 and that there exists at least on element < 1, else it would violate its transient nature.

Using these properties we will prove (iii) by showing that the first term on the right hand side of Eq. 10 can be bounded from above by $|V_T|/n$. In particular, splitting the infinity sum at

$$t^\star = \left\lfloor \frac{d}{1-d} \right\rfloor,$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x (note that $t^\star \geq 0$ for $d \in [0, 1)$), gives for the first term on the right hand side of (10)

$$\begin{aligned} &\frac{1-d}{n} \sum_{t=1}^{\infty} \left(t - \frac{d}{1-d}\right) d^{t-1} a_t \\ &= \frac{1-d}{n} \left(\sum_{t_1=1}^{t^\star} \left(t_1 - \frac{d}{1-d}\right) d^{t_1-1} a_{t_1} + \sum_{t_2=t^\star+1}^{\infty} \left(t_2 - \frac{d}{1-d}\right) d^{t_2-1} a_{t_2} \right) \end{aligned} \tag{12}$$

of which all terms of the first summation are non-positive and those of the second summation non-negative. Using property (11), this enables us to bound the terms of the first summation as

$$\left(t_1 - \frac{d}{1-d}\right) d^{t_1-1} a_{t_1} \leq \left(t_1 - \frac{d}{1-d}\right) d^{t_1-1} a_{t^\star+1}, \quad \text{for } t_1 = 1, 2, \ldots, t^\star,$$

and similar for the second summation terms

$$\left(t_2 - \frac{d}{1-d}\right) d^{t_2-1} a_{t_2} \leq \left(t_2 - \frac{d}{1-d}\right) d^{t_2-1} a_{t^\star+1},$$

for $t_2 = t^\star + 1, t^\star + 2, \ldots, \infty$. Applying these upperbounds for the summation terms in Eq. 12 leads to

$$\begin{aligned} \frac{1-d}{n} \sum_{t=1}^{\infty} \left(t - \frac{d}{1-d}\right) d^{t-1} a_t &\leq \frac{(1-d) a_{t^\star+1}}{n} \sum_{t=1}^{\infty} \left(t - \frac{d}{1-d}\right) d^{t-1} \\ &= \frac{(1-d) a_{t^\star+1}}{n} \left(\sum_{t_1=1}^{\infty} t_1 d^{t_1-1} - \sum_{t_2=1}^{\infty} \frac{d}{1-d} d^{t_2-1} \right) \\ &= \frac{(1-d) a_{t^\star+1}}{n} \left(\frac{1}{(1-d)^2} - \frac{d}{(1-d)^2} \right) \\ &= \frac{a_{t^\star+1}}{n} \\ &< \frac{|V_T|}{n}, \end{aligned}$$

where we used after some standard calculus arguments in the last bound that $a_{t^\star+1} < a_0 = |V_T|$, which proves via (10) part (iii).

Regarding part (iv) it follows from Eq. 10 that the second order derivative $\mathrm{Fr}''_{\mathrm{GP}}(d)$ equals

$$\begin{aligned}\mathrm{Fr}''_{\mathrm{GP}}(d) &= \frac{-1}{n}\sum_{t=1}^{\infty}\left(t-\frac{d}{1-d}\right)d^{t-1}a_t \\ &\qquad +\frac{1-d}{n}\sum_{t=1}^{\infty}\left(t(t-1)-\frac{d^2}{(1-d)^2}-\frac{td}{1-d}\right)d^{t-2}a_t \\ &= \frac{1-d}{n}\sum_{t=1}^{\infty}d^{t-2}a_t\left(t(t-1)-\frac{2dt}{1-d}\right),\end{aligned}$$

similar as for the first order derivative, splitting the last summation at $t^\star = \left\lfloor \frac{1+d}{1-d}\right\rfloor$ reveals that we can bound $\mathrm{Fr}''_{\mathrm{GP}}(d)$ via

$$\mathrm{Fr}''_{\mathrm{GP}}(d) \leq \frac{(1-d)a_{t^\star+1}}{n}\sum_{t=1}^{\infty}d^{t-2}\left(t(t-1)-\frac{2dt}{1-d}\right),$$

which, using standard calculus arguments, can be rewritten as

$$\begin{aligned}\mathrm{Fr}''_{\mathrm{GP}}(d) &\leq \frac{(1-d)a_{t^\star+1}}{n}\left(\sum_{t_1=1}^{\infty}d^{t_1-2}t_1(t_1-1)-\frac{2}{1-d}\sum_{t_2=1}^{\infty}t_2d^{t_2-1}\right) \\ &= \frac{(1-d)a_{t^\star+1}}{n}\left(\frac{2}{(1-d)^3}-\frac{2}{1-d}\frac{1}{(1-d)^2}\right) \\ &= 0\end{aligned}$$

which shows that $\mathrm{Fr}''_{\mathrm{GP}}(d) \leq 0$.

For part (v) note that $P_{TT} \geq 0$ (where 0 indicates a $|V_T| \times |V_T|$ zeros matrix) implies via (9) that

$$\mathrm{Fr}_{\mathrm{GP}}(d) \geq \frac{1-d}{n}a_0 = \frac{(1-d)|V_T|}{n}.$$

□

Remark 1 Since all elements of P_{TT} are ≥ 0 it can be concluded from (9) that

$$d \in (0,1)\ :\ \mathrm{Fr}_{\mathrm{GP}}(d) = \frac{(1-d)|V_T|}{n} \qquad \text{if and only if} \qquad P_{TT} = 0,$$

with 0 an appropriate sized matrix of zeros.

As a consequence of Proposition 2, it holds in general for any $d \in [0, 1)$ that

$$\frac{(1-d)|V_T|}{n} \leq \mathrm{Fr}_{\mathrm{GP}}(d) \leq \frac{|V_T|}{n}, \tag{13}$$

which also means that $\mathrm{Fr}_{\mathrm{GP}}(d) \in (0, |V_T|/n]$ for all $d \in [0, 1)$. These insights regarding $\mathrm{Fr}_{\mathrm{GP}}(d)$ are illustrated in Fig. 2 specifically. In Fig. 2, for different instances with same n and $|V_T|$ the values for $\mathrm{Fr}_{\mathrm{GP}}(d)$ are plotted for $d \in [0, 1)$. The range for $\mathrm{Fr}_{\mathrm{GP}}(d)$ dictated by (13) is marked with the gray area, and furthermore, all graphs of $\mathrm{Fr}_{\mathrm{GP}}(d)$ are decreasing throughout interval $d \in [0, 1)$. Moreover, due to Proposition 2 (iv), the derivative of $\mathrm{Fr}_{\mathrm{GP}}(d)$ does not increase for increasing $d \in [0, 1)$.

Figure 2 illustrates that the fraction of score awarded to transient states strongly depends on the choice of $d \in [0, 1)$, i.e., a user implicitly has to trade-off the score distribution over

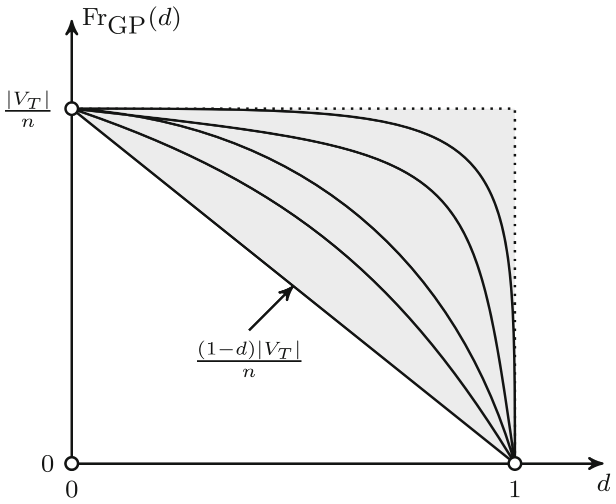
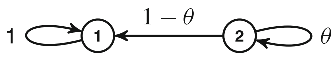


Fig. 2 For all instances with n nodes of which $|V_T|$ are of transient nature, the graph for $\mathrm{Fr}_{\mathrm{GP}}(d)$ lies in the gray area and is decreasing for $d \in [0, 1)$

transient and ergodic states. In particular, Fig. 2 uncovers that certain instances in combination with a relative large value of d tend to be very sensitive with respect to the specific choice of d, i.e., a small shift in d may lead to a completely different score distribution over the transient and ergodic states. This is basically in line with the observations in (Bressan and Peserico 2010) regarding the volatility of the rankings for certain instances. So in conclusion, as the type of instance is generally unknown beforehand, any general advice for a fixed d is unsatisfactory regarding the score distribution between transient and ergodic states.

A minimal example illustrating Fig. 2 is the following graph where the edges are labeled with the transition probabilities of the random surfer.



So the corresponding θ-dependent transition matrix, denoted by P_θ, equals

$$P_\theta = \begin{bmatrix} 1 & 0 \\ 1-\theta & \theta \end{bmatrix}. \tag{14}$$

In Fig. 3, $\mathrm{Fr}_{\mathrm{GP}}(d)$ is plotted for P_θ with varying values for θ. Note that $\mathrm{Fr}_{\mathrm{GP}}(d)$ equals the second element in vector $\pi^\top_{P_d(u)}$ with $P = P_\theta$ in this simple example of size $n = 2$. For each choice of θ, P_θ represents a different 'instance'.

The example from Fig. 3 shows that when θ becomes larger, line $\mathrm{Fr}_{\mathrm{GP}}(d)$ for $d \in [0, 1)$ approaches the right angle starting in coordinate (0,0.5) then to (1,0.5) and ending in (1,0). In other words, for relative large θ, the fraction of score awarded to transient states approximately equals $|V_T|/n = 1/2$ for $d \in [0, d_0]$, where d_0 approaches 1 for increasing θ. For $d > d_0$ the fraction falls rapidly and thus leads to extra sensitivity issues.

3.2 Motivation and developments of revised generalized ranking

The problem of the choice of damping factor d in Google's PageRank was addressed in (Berkhout 2016) by focusing on the network structure. Via the introduction of the so-called

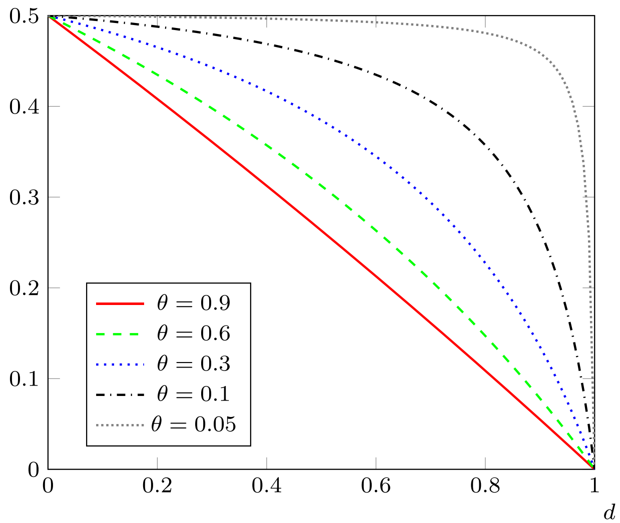


Fig. 3 Plots of $\mathrm{Fr}_{\mathrm{GP}}(d)$ with $P = P_\theta$ as given in Eq. 14 for varying θ

extended ergodic projector this lead to the Generalized Ranking (GR) in (Berkhout 2016). We will first discuss the philosophies behind GR and sketch how these are incorporated into GR. Subsequently a shortcoming of GR is identified, and eventually a generalized version of the extended ergodic projector is introduced that leads to a refined version of GR that addresses this shortcoming.

The key philosophy of GR is the following:

Keep the underlying network structure in tact.

This means that in contrast to Google's PageRank $P_d(v)$ for $d \in (0, 1)$, no modifications are made to P. Generally, this leads to a Markov multi-chain described by P as in Eq. 4.

The idea behind GR can be explained as follows. Each state gets initially a share of the total score (which equals 1) according to v. Based on the network structure, all these 'initial scores' are redistributed over reachable states. E.g., state i can redistribute $v(i)$ score.

The initial scores of all ergodic states inside an ergodic class are redistributed according to the location of a random surfer in the long-term. This is in line with the following philosophy:

The most intuitive way to obtain rankings inside an ergodic class is to analyze the long-term location of a random surfer.

Regarding transient states a different random surfer philosophy is adopted since by the very nature of transient states, the unmodified random surfer approach is inadequate, see also the zeros matrix in the transient part of Π_P in Eq. 4. Particularly, per transient state the following, random surfer inspired, philosophy is adopted to obtain scores for transient states:

A transient state is more important when the random surfer visits this particular state relatively more often (compared to other transient states) before leaving the transient part.

Using the information of deviation matrix D_P this philosophy can be implemented. Particularly, per transient state i, element (i, j) of $D_{P_{TT}}$ gives the expected number of visits to transient state j when starting in i. Normalizing this vector gives a ranking per transient state. For more details regarding the deviation matrix see Section 2.

As transient states lead to ergodic states a score transfer from the transient part to the ergodic part is appropriate. Regarding this the following philosophy was adopted in (Berkhout 2016):

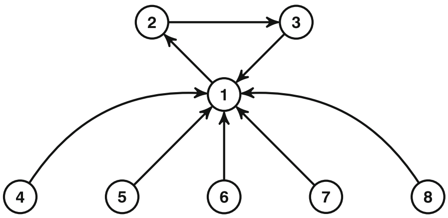
> *Score should be transfered from the transient states to directly reachable ergodic states. Moreover, the amount of score that is transfered should be negatively correlated with the stay duration inside the transient part before moving to the ergodic part.*

The implementation of this philosophy in the extended ergodic projector circumvents the opaque choice of $d \in [0, 1)$. In other words, the score trade-off between transient and ergodic states is solely based on the network structure. Note that $D_{P_{TT}}$ contains information about the stay duration of the random surfer inside the transient part.

After all states have redistributed their initial scores, the final ranking can be based on the end scores. In case $v = \bar{1}/n$, i.e., a uniform stochastic vector, a democracy principle applies as each state gets an equal ‘vote’ (initial score) that can be divided over different states. This way larger connected components have more ‘voting power’/authority in the score division that leads to the ranking.

In this paper, GR from (Berkhout 2016), with the principles as discussed above, is improved to ensure a more subtle ranking. In particular, the improvement ensures that score-mass not only flows from the transient part to particular ergodic states (that can be reached directly from transient states), but to the complete ergodic class to which that ergodic state belongs to. This way, the increase in importance of an ergodic state due to its ‘popularity’ among transient states has also an effect for other states in the particular ergodic class. The following example illustrates this shortcoming of GR from (Berkhout 2016) that the improved version of GR from this paper overcomes.

Example 2 Consider the following example represented by graph $G = (V, E)$ consisting of $n = 8$ nodes of which nodes 1, 2 and 3 constitute a single ergodic class and nodes 4, 5, 6, 7 and 8 are transient (i.e., $\mathcal{E} = 1$, $|V_E| = 3$ and $|V_T| = 5$).



We assume that $w(e) = 1$ for all $e \in E$ so that the corresponding transition matrix equals

$$P = \begin{bmatrix} 0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \end{bmatrix}. \tag{15}$$

Calculating the extended ergodic projector $\widetilde{\Pi}_P$ from (Berkhout 2016) (more details will follow later on) and averaging over the rows give the following ranking (with most important node first)

$$(1;\ 0.4375),\ (\{2,3\};\ 0.125),\ (\{4,5,\ldots,8\};\ 0.0625),$$

where in $(i;\ x)$, i denotes the node and x its score, where nodes in curly brackets $\{\cdots\}$ obtained the same score. Obviously node 1 is the 'star' in this network, so this justifies to rank node 2 above node 3 as node 2 is directly pointed to by 'star'-node 1. This effect does not come forward in the extended ergodic projector from (Berkhout 2016).

In order to ensure a more subtle score transfer from transient states to ergodic states we adopt the following philosophy:

Score from the transient part received by an ergodic state should be shared on a discounted way with other states of the corresponding ergodic class according to the network structure.

This is reasonable since the ergodic states to which the ergodic state points to are part of the reason why she is 'popular' among transient states. Moreover, an ergodic state can in turn also obtain extra mass in a similar way. However, a significant discounting should apply to the score division where 'further away' states should receive a lower share.

In the following, we redefine the extended ergodic projector so that the above philosophy is adopted. Letting

$$Q = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_{\mathcal{E}} \end{bmatrix},$$

i.e., Q is the 'ergodic' part of transition matrix P in Eq. 4, define for $k \in \mathbb{N}$

$$W_k = D_{P_{TT}}[P_{T1}\ P_{T2}\ \cdots\ P_{T\mathcal{E}}]Q^k,$$

or alternatively after rewriting $D_{P_{TT}}$

$$W_k = \sum_{t=0}^{\infty}(P_{TT})^t[P_{T1}\ P_{T2}\ \cdots\ P_{T\mathcal{E}}]Q^k,$$

where the (i, j)-th element reads as:

$W_k(i, j)$ = *the probability that a random surfer, when starting from transient state* i*, reaches ergodic state* j *if it is* k *discrete-time steps ago that the random surfer left the transient part*,

and where $[P_{T1}\ P_{T2}\ \cdots\ P_{T\mathcal{E}}]$ denotes the $|V_T| \times |V_E|$ matrix as given in the canonical form of P in Eq. 4. For $\gamma \in [0, 1)$, let

$$W(\gamma) = (1-\gamma)\sum_{k=0}^{\infty}\gamma^k W_k,$$

or, after inserting W_k,

$$W(\gamma) = D_{P_{TT}}[P_{T1}\ P_{T2}\ \cdots\ P_{T\mathcal{E}}](1-\gamma)\sum_{k=0}^{\infty}\gamma^k Q^k. \tag{16}$$

Remark 2 Note that by definition, W from (Berkhout 2016) equals $W(0) = W_0$. Furthermore, since W_k is a stochastic matrix for $k = 0, 1, 2, \ldots$ this implies that $W(\gamma)$ is a stochastic matrix.

Using the above definitions, we can now formally introduce the Generalized Ranking (GR) that satisfies all previously mentioned principles. In the following $\text{dg}(\beta)$ denotes a zeros matrix with on the diagonal vector β.

Generalized ranking (GR)

Step 1: Uncover the Markov chain structure (i.e., identify ergodic classes and transient states), and calculate Π_Q (the ergodic part of Π_P) and $D_{P_{TT}}$.

Step 2: Choose $\gamma \in [0, 1)$ small and compute $W(\gamma)$.

Step 3: Compute the extended ergodic projector

$$\widetilde{\Pi}_P(\gamma) = \left[\begin{array}{cccc|c} \Pi_1 & 0 & \cdots & 0 & \\ 0 & \Pi_2 & \ddots & \vdots & \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & \Pi_{\mathcal{E}} & \\ \hline \multicolumn{4}{c|}{\text{dg}(\beta)W(\gamma)} & (I - \text{dg}(\beta))Y \end{array}\right],$$

where β is a vector with elements

$$\beta(i) = (D_{P_{TT}}(i, \cdot)\bar{1} + 1)^{-1}, \quad \text{for } i = 1, 2, \ldots, |V_T|,$$
$$Y = \text{dg}(D_{P_{TT}}\bar{1})^{-1} D_{P_{TT}}$$

and

$$W(\gamma) = D_{P_{TT}}[P_{T1} \; P_{T2} \; \cdots \; P_{T\mathcal{E}}](1-\gamma)\sum_{k=0}^{\infty} \gamma^k Q^k.$$

Step 4: Choose personalization vector v and return ranking

$$r_{\text{GR}} := r(v^\top \widetilde{\Pi}_P(\gamma)).$$

Remark 3 Note that we may write

$$\Pi_Q = \begin{bmatrix} \Pi_1 & 0 & \cdots & 0 \\ 0 & \Pi_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Pi_{\mathcal{E}} \end{bmatrix},$$

so that

$$\widetilde{\Pi}_P(\gamma) = \left[\begin{array}{c|c} \Pi_Q & 0 \\ \hline \text{dg}(\beta)W(\gamma) & (I - \text{dg}(\beta))Y \end{array}\right]$$

Choosing $v = \bar{1}/n$ (a uniform stochastic vector) gives the unprejudiced ranking $r(\bar{1}\widetilde{\Pi}_P/n)$. Alternatively, in case of personalization vector v (entries not necessarily > 0), ranking $r(v^\top \widetilde{\Pi}_P)$ prioritizes states from the preferred communicating components of the network.

Recall that all elements of $D_{P_{TT}}$ are non-negative, and moreover, all diagonal elements of $D_{P_{TT}}$ are ≥ 1. This ensures that β and Y exist. Furthermore, it shows that Y is a stochastic matrix. As discussed previously, each row of matrix Y provides scores for a ranking over the transient states.

Vector β contains the information about the stay duration inside the transient part. In particular, from the interpretation of $D_{P_{TT}}$ it follows that value $1/\beta(i) - 1 = D_{P_{TT}}\bar{1}$ equals the expected number of steps in the transient part before entering an ergodic state when starting in transient state i. Since $1 \leq D_{P_{TT}}\bar{1} < \infty$, note that

$$0 < \beta(i) \leq \frac{1}{2}, \quad \text{for } i = 1, 2, \ldots, |V_T|. \tag{17}$$

Vector β determines how much score is transfered from the transient part to the ergodic part. Transient state i awards a fraction $\beta(i)$ of its initial score $v(i)$ to ergodic states. The division of score-share $\beta(i)v(i)$ over the ergodic states is proportional to the i-th row of $W(\gamma)$ denoted by $W(\gamma)(i, \cdot)$. The remaining score of $(1 - \beta)v(i)$ is divided over the transient states according to $Y(i, \cdot)$.

One might argue that again a user-dependent value is introduced with $\gamma \in [0, 1)$, bringing us to similar problems as with the choice of damping factor $d \in [0, 1)$. However, the choice for γ has far less consequences than the choice of d. Particularly, the score fraction appointed to the different ergodic classes and the transient part is fixed in advanced (as we will formalize in Eq. 18 later on). The effect of γ only applies to the redistribution of score from the transient part over the ergodic class, to enable more subtle scoring inside ergodic classes. The main contribution of scores inside ergodic classes however, comes from other states inside the ergodic class and thus the effect of γ is limited regarding the global scoring. In practice, our advice is to take γ not too small so that only nearby nodes benefit from the extra popularity of an ergodic state.

In the following we revisit Example 1 to illustrate GR.

Example 1 (Continued) It holds that

$$D_{P_{TT}} = (I - P_{TT})^{-1} = \begin{bmatrix} 4 & 1 & 1 & 1 \\ 4 & 2 & 1 & 1 \\ 4 & 1 & 2 & 1 \\ 4 & 1 & 1 & 2 \end{bmatrix}.$$

We note that $D_{P_TT}(2, 3) = 1$ means that the random surfer is expected to visits transient state 4 one time before leaving the transient part when starting in transient state 3. Using the result for $D_{P_{TT}}$ gives, respectively,

$$\beta = \begin{bmatrix} \frac{1}{8} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}^\top,$$

$$Y = \begin{bmatrix} \frac{4}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \end{bmatrix},$$

and

$$W(\gamma) = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^\top,$$

i.e., $W(\gamma)$ is in this simple example independent of γ as the ergodic class is of size 1. Combining the above leads to

$$\widetilde{\Pi}_P(\gamma) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{8} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{9} & \frac{4}{9} & \frac{2}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{9} & \frac{1}{9} & \frac{2}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{9} & \frac{1}{9} & \frac{1}{9} & \frac{2}{9} \end{bmatrix}$$

which leads to the unprejudiced scoring (i.e., with a uniform stochastic vector as v)

$$\bar{1}\widetilde{\Pi}_P(\gamma)/4 \approx \begin{bmatrix} 0.29 & 0.37 & 0.11 & 0.11 & 0.11 \end{bmatrix},$$

which leads to ranking 2, 1, {3, 4, 5}, i.e., state 2 is ranked first.

The following illustrates the effect that γ has on the extended ergodic projector in Example 2.

Example 2 (Continued) Rewriting P as given in (15) in canonical block-form gives

$$P = \left[\begin{array}{c|c} Q & 0 \\ \hline P_{T1} & P_{TT} \end{array}\right]$$

where Q is a 3×3 matrix, 0 represents a 3×5 zeros matrix, P_{T1} a 5×3 matrix, and P_{TT} a 5×5 matrix. Since $P_{TT} = 0$ it holds that

$$D_{P_{TT}} = (I - P_{TT})^{-1} = I$$

with I an identify matrix of size 5, which gives $\beta = \bar{1}/2$, $Y = I$ and $W = (1-\gamma)P_{T1}(I - \gamma Q)^{-1} = \bar{1}\left[\frac{1}{\gamma^2+\gamma+1} \;\; \frac{\gamma}{\gamma^2+\gamma+1} \;\; \frac{\gamma^2}{\gamma^2+\gamma+1}\right]$. This leads to

$$\widetilde{\Pi}_P(\gamma) = \left[\begin{array}{ccc|cccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & & & & \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & & 0 & & \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & & & & \\ \hline \frac{1}{2(\gamma^2+\gamma+1)} & \frac{\gamma}{2(\gamma^2+\gamma+1)} & \frac{\gamma^2}{2(\gamma^2+\gamma+1)} & \frac{1}{2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \frac{1}{2} & \cdots & 0 \\ & & & \vdots & \ddots & \ddots & 0 \\ \frac{1}{2(\gamma^2+\gamma+1)} & \frac{\gamma}{2(\gamma^2+\gamma+1)} & \frac{\gamma^2}{2(\gamma^2+\gamma+1)} & 0 & \cdots & 0 & \frac{1}{2} \end{array}\right].$$

The ranking, with the most important state in front, that follows from averaging the rows of $\widetilde{\Pi}_P(\gamma)$ (i.e., v is a uniform stochastic vector) for $\gamma \in [0, 1)$ is

$$1, \; 2, \; 3, \; \{4, 5, \ldots, 8\},$$

i.e., it ranks 2 above 3 in contrast to the ranking based on $\widetilde{\Pi}_P = \widetilde{\Pi}_P(0)$ as shown previously. In particular, the scores for states 1, 2 and 3 are, respectively,

$$\frac{1}{8} + \frac{3}{16} \cdot \frac{1}{\gamma^2+\gamma+1}, \quad \frac{1}{8} + \frac{3}{16} \cdot \frac{\gamma}{\gamma^2+\gamma+1}, \quad \frac{1}{8} + \frac{3}{16} \cdot \frac{\gamma^2}{\gamma^2+\gamma+1},$$

where the scores for transient states $4, 5, \ldots, 8$ are still 1/16. Note that inserting $\gamma = 0$ leads to the same results as previously.

We conclude this subsection with an investigation of properties of the extended ergodic projector $\widetilde{\Pi}_P(\gamma)$.

In light of Proposition 2, it is interesting to analyze fraction Fr_{GR} of the total score that is awarded by GR to transient states. It holds for a general instance with n states of which $|V_T|$ are transient, that this fraction equals

$$\text{Fr}_{\text{GR}} = \frac{\sum_{i=1}^{|V_T|}(1-\beta_i)}{n} \tag{18}$$

which shows that it is a fixed fraction independent of γ. Moreover, for each instance with n nodes of which $|V_T|$ of them are transient it holds that

$$\frac{|V_T|}{2n} \leq \text{Fr}_{\text{GR}} < \frac{|V_T|}{n},$$

where $\text{Fr}_{GR} = \frac{|V_T|}{2n}$ if and only if $P_{TT} = 0$ (where 0 represents an appropriate sized matrix of zeros). It can be concluded that Fr_{GR} is solely based on the dynamics inside the transient part of the network.

Remark 4 We may rewrite $\widetilde{\Pi}_P(\gamma)$ as

$$\widetilde{\Pi}_P(\gamma) = \left[\begin{array}{c|c} \Pi_Q & 0 \\ \hline \text{dg}(\beta)\begin{bmatrix} \pi^\top_{Q_\gamma(W_0(1,\cdot))} \\ \pi^\top_{Q_\gamma(W_0(2,\cdot))} \\ \vdots \\ \pi^\top_{Q_\gamma(W_0(|V_T|,\cdot))} \end{bmatrix} & (I-\text{dg}(\beta))Y \end{array}\right],$$

where $\pi^\top_{Q_\gamma(W_0(i,\cdot))}$, for $i = 1, 2, \ldots, |V_T|$, is the unique stationary distribution of the Markov chain

$$Q_\gamma(W_0(i,\cdot)) = \gamma Q + (1-\gamma)\bar{1}W_0(i,\cdot).$$

Note that the above Markov chain is equal to Google's PageRank transition matrix for damping factor γ, personalization vector $W_0(i,\cdot)$ and original transition matrix Q. In words, the division of score (that originates from the transient part) inside the ergodic class is based on the stationary distribution of the Google's PageRank transition matrix with different personalization vectors per transient state and a common damping factor γ.

The following proposition investigates the behavior of $\widetilde{\Pi}_P(\gamma)$ regarding the extreme values of $\gamma \in [0, 1)$. In particular, it uncovers the connection of $\widetilde{\Pi}_P(\gamma)$ with the extended ergodic projector $\widetilde{\Pi}_P$ from (Berkhout 2016).

Proposition 3 *Inserting* $\gamma = 0$ *into*

$$\widetilde{\Pi}_P(\gamma) = \left[\begin{array}{c|c} \Pi_Q & 0 \\ \hline dg(\beta)W(\gamma) & (I-dg(\beta))Y \end{array}\right] \tag{19}$$

gives

$$\widetilde{\Pi}_P(0) = \widetilde{\Pi}_P, \tag{20}$$

with $\widetilde{\Pi}_P$ *as in* (Berkhout 2016), *and*

$$\lim_{\gamma\uparrow 1} \widetilde{\Pi}_P(\gamma) = \Pi_P + \left[\frac{0}{(I - dg(\beta))\,[-R_1 \; -R_2 \; \cdots \; -R_{\mathcal{E}} \; Y]}\right]. \tag{21}$$

Proof Expression (20) follows directly from Eq. 19 by observing that $W(0) = W_0 = W$, with W as defined in (Berkhout 2016).

Regarding (21), it follows from Eq. 16 that

$$\lim_{\gamma\uparrow 1} W(\gamma) = D_{P_{TT}}[P_{T1} \; P_{T2} \; \cdots \; P_{T\mathcal{E}}] \lim_{\gamma\uparrow 1}(1-\gamma)\sum_{k=0}^{\infty}\gamma^k Q^k$$

using the limit result for the modified resolvent from (Kartashov 1996)

$$= D_{P_{TT}}[P_{T1} \; P_{T2} \; \cdots \; P_{T\mathcal{E}}]\Pi_Q$$

inserting the canonical block form of Π_Q

$$\begin{aligned} &= D_{P_{TT}}[P_{T1}\Pi_{P_1} \; P_{T2}\Pi_{P_2} \; \cdots \; P_{T\mathcal{E}}\Pi_{P_{T\mathcal{E}}}] \\ &= [R_1 \; R_2 \; \cdots \; R_{\mathcal{E}}], \end{aligned}$$

where we used in the last equation that $R_i = D_{P_{TT}} P_{Ti} \Pi_{P_i}$, for $i = 1, 2, \ldots, E$, see also Eq. 4. Inserting the above expression for $\lim_{\gamma\uparrow 1} W(\gamma)$ into $\lim_{\gamma\uparrow 1} \widetilde{\Pi}_P(\gamma)$ gives

$$\lim_{\gamma\uparrow 1} \widetilde{\Pi}_P(\gamma) = \left[\begin{array}{cccc|c} \Pi_1 & 0 & \cdots & 0 & \\ 0 & \Pi_2 & \ddots & \vdots & \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & \Pi_{\mathcal{E}} & \\ \hline \multicolumn{4}{c|}{\mathrm{dg}(\beta)[R_1 \; R_2 \; \cdots \; R_{\mathcal{E}}]} & (I - \mathrm{dg}(\beta))Y \end{array}\right],$$

which can be rewritten to Eq. 21 using the canonical form of Π_P from Eq. 4. □

Remark 5 Expression (21) shows that for $\gamma \uparrow 1$, $\widetilde{\Pi}_P(\gamma)$ has ergodic projector Π_P as basis. Then, based on the stay duration inside the transient part (i.e., dg(β)) score is transferred from the ergodic part (indicated by the minuses) to the transient part. Particularly, the score transfer from the transient part to an ergodic state is redivided over the entire corresponding ergodic class proportionally to the unique stationary distribution inside this ergodic class.

3.3 Analysis of the generalized ranking

In the following we provide a detailed analysis of GR that leads to an insightful analytical interpretation of GR. As a first step we define per transient state j Markov chain $P_j^\star$ which describes the location of a random surfer with the following surfing behavior. On the transient part, the random surfer chooses her next location according to the transient part from P. Once the random surfer moves out of the transient part, say she wants to move to ergodic state k, she flips a $(1-\gamma)$-coin (head with chance $1-\gamma$). When the coin shows head her next location will be indeed ergodic state k. Else she randomnly chooses a next potential location according to $P(k, \cdot)$, say ergodic state l, and again flips a coin. In case the outcome is head, the next location of the random surfer after leaving the transient part is ergodic state l (so not the initial state k). She repeats this procedure until she throws head. Once the location of the random surfer is an ergodic state, she directly jumps back to transient state j afterwards. The precise definition of $P_j^\star$ that describes the location of this random surfer is provided below.

Definition 1 For transient state j let

$$P_j^\star = \left[\begin{array}{c|c} 0 & \bar{1} I(j,\cdot) \\ \hline [P_{T1}\ P_{T2}\ \cdots\ P_{T\mathcal{E}}]\,(1-\gamma)\sum_{k=0}^{\infty}(\gamma Q)^k & P_{TT} \end{array}\right],$$

where $\bar{1}$ is a column-vector of ones of size $|V_E|$, and $I(j,\cdot)$ is the j-th row of an identity matrix of size $|V_T|$.

Lemma 1 establishes ergodicity of $P_j^\star$.

Lemma 1 *Given transient state j, the Markov chain $P_j^\star$ has a unique stationary distribution.*

Proof By construction, the set of transient and ergodic states that can be reached from j forms a communicating class, and this class has a unique stationary distribution according to the Perron-Frobenius theorem. Further, all remaining transient states will eventually in finite time lead to an ergodic state for $\gamma \in (0, 1)$ which in turn will lead to the communicating class, i.e., these states in the artificial Markov chain become transient. □

Elaborating on Lemma 1, Theorem 1 provides an interpretation of $\widetilde{\Pi}_P(\gamma)$ in terms of "local" stationary distributions.

Theorem 1 *It holds for $\widetilde{\Pi}_P(\gamma)$ for any $\gamma \in [0, 1)$ that:*

(i) *$\widetilde{\Pi}_P(\gamma)$ is a stochastic matrix.*
(ii) *For every ergodic state i: the i-th row of $\widetilde{\Pi}_P(\gamma)$ is the unique stationary distribution inside the corresponding ergodic class.*
(iii) *For every transient state j: the j-th row of $\widetilde{\Pi}_P(\gamma)$ is the unique stationary distribution of $P_j^\star$.*

Proof It holds that Y is a stochastic matrix (see also the discussion after the introduction of GR). Also matrix $W(\gamma)$ is stochastic as was argued in Remark 2. Since $0 < \beta(i) \le 1/2$ for all $i = 1, 2, \ldots, |V_T|$ (see also Eq. 17), this proves that all rows of $\widetilde{\Pi}_P(\gamma)$ corresponding to transient states are stochastic. As the rows of ergodic states are clearly stochastic, this proves Statement (i). Moreover, the first quadrant of $\widetilde{\Pi}_P(\gamma)$ coincides with that of Π_P so that Statement (ii) readily follows.

To prove Statement (iii), we will denote $\widetilde{\Pi}_P(\gamma)$ by $\widetilde{\Pi}_P^\gamma$ for notational easiness, e.g., the j-th row of $\widetilde{\Pi}_P(\gamma)$ can be written as $\widetilde{\Pi}_P^\gamma(j,\cdot)$. Similar, we will denote $W(\gamma)$ with W^γ. Because of Statement (i) and Lemma 1, proving that for transient state j

$$\widetilde{\Pi}_P^\gamma(j,\cdot)(P_j^\star) = \widetilde{\Pi}_P^\gamma(j,\cdot) \tag{22}$$

shows that $\widetilde{\Pi}_P^\gamma(j,\cdot)$ is the unique stationary distribution of $P_j^\star$ (third statement).

In particular, multiplying the left hand side of Eq. 22 with $\frac{1}{1-\beta_j}$ (to make the proof more clear) and rewriting gives in case of transient state j

$$
\begin{aligned}
&\frac{1}{1-\beta_j}\widetilde{\Pi}^{\gamma}_{P}(j,\cdot)(P^{\star}_{j})\\
&=\left[\frac{\beta_j}{1-\beta_j}W^{\gamma}(j,\cdot)\;\middle|\;Y(j,\cdot)\right](P^{\star}_{j})\\
&=\left[\underbrace{Y(j,\cdot)\,[P_{T1}\;P_{T2}\;\cdots\;P_{T\mathcal{E}}]\,(1-\gamma)\sum_{k=0}^{\infty}(\gamma Q)^k}_{(\star)}\;\middle|\;\underbrace{Y(j,\cdot)P^{j-}_{TT}}_{(\star\star)}\right.\\
&\qquad\left.\underbrace{\frac{\beta_j}{1-\beta_j}W^{\gamma}(j,\cdot)\bar{1}+Y(j,\cdot)P_{TT}(\cdot,j)}_{(\star\star\star)}\quad\underbrace{Y(j,\cdot)P^{j+}_{TT}}_{(\star\star)}\right],
\end{aligned}
$$

where $\bar{1}$ denotes a ones vector of size $|V_E|$, P^{j-}_{TT} denotes matrix P_{TT} left of column j and, vice versa, P^{j+}_{TT} denotes matrix P_{TT} right of column j. Specifically,

$$P^{j-}_{TT}=[P_{TT}(\cdot,1)\;P_{TT}(\cdot,2)\;\cdots\;P_{TT}(\cdot,j-1)]$$

and

$$P^{j+}_{TT}=[P_{TT}(\cdot,j+1)\;P_{TT}(\cdot,j+2)\;\cdots\;P_{TT}(\cdot,|V_T|)]\,.$$

We will prove (22) stepwise by showing that the three parts ($\star$), ($\star\star$), ($\star\star\star$) indicated above are equal to

$$\frac{1}{1-\beta_j}\widetilde{\Pi}^{\gamma}_{P}(j,\cdot)=\left[\frac{\beta_j}{1-\beta_j}W^{\gamma}(j,\cdot)\;\middle|\;Y(j,\cdot)\right]. \tag{23}$$

Note that in case $j=1$ the term P^{j-}_{TT} from part ($\star\star$) does not apply, just as P^{j+}_{TT} from part ($\star\star$) does not apply when $j=|V_T|$.

Since by definition

$$Y(j,\cdot)=\frac{1}{D_{P_{TT}}(j,\cdot)\bar{1}}=\frac{1}{1/\beta_j-1}D_{P_{TT}}(j,\cdot) \tag{24}$$

it holds for ($\star$) that

$$
\begin{aligned}
&Y(j,\cdot)\,[P_{T1}\;P_{T2}\;\cdots\;P_{T\mathcal{E}}]\,(1-\gamma)\sum_{k=0}^{\infty}(\gamma Q)^k\\
&\qquad=\frac{1}{1/\beta_j-1}D_{P_{TT}}(j,\cdot)\,[P_{T1}\;P_{T2}\;\cdots\;P_{T\mathcal{E}}]\,(1-\gamma)\sum_{k=0}^{\infty}(\gamma Q)^k\\
&\qquad=\frac{1}{1/\beta_j-1}W^{\gamma}(j,\cdot)\\
&\qquad=\frac{\beta_j}{1-\beta_j}W^{\gamma}(j,\cdot),
\end{aligned}
$$

which shows that ($\star$) equals the corresponding part in Eq. 23.

For an arbitrary transient state $k \neq j$, we will show that $Y(j, \cdot)P_{TT}(\cdot, k) = Y(j, k)$ (i.e., parts $(\star\star)$ equal the corresponding parts in Eq. 23) which proves (22) for both parts $(\star\star)$. It holds that

$$\begin{aligned} Y(j, \cdot)P_{TT}(\cdot, k) &= \frac{1}{1/\beta_j - 1} D_{P_{TT}}(j, \cdot)P_{TT}(\cdot, k) \\ &= \frac{1}{1/\beta_j - 1} \sum_{t=1}^{\infty} (P_{TT})^t(j, k) \\ &= \frac{1}{1/\beta_j - 1} D_{P_{TT}}(j, k) \\ &= Y(j, k) \end{aligned}$$

where we used in the second equation that element (j, k) of $(P_{TT})^t$ can be written as

$$(P_{TT})^t(j, k) = (P_{TT})^{t-1}(j, \cdot)P_{TT}(\cdot, k),$$

for $t \in \mathbb{N}$, and the third equation follows because

$$(P_{TT})^0(j, k) = I(j, k) = 0, \qquad \text{for } j \neq k.$$

Lastly, for part $(\star\star\star)$ it should be proven that it equals $Y(j, j)$, see also Eq. 23. In particular,

$$(\star\star\star) = \frac{\beta_j}{1-\beta_j} W^{\gamma}(j, \cdot)\bar{1} + Y(j, \cdot)P_{TT}(\cdot, j)$$

using that $[P_{T1}\ P_{T2}\ \cdots\ P_{T\mathcal{E}}]\,\bar{1} = \bar{1} - P_{TT}\bar{1}$ (rows P sum up to 1)

$$= \frac{\beta_j}{1-\beta_j} D_{P_{TT}}(j, \cdot)(I - P_{TT})\bar{1} + Y(j, \cdot)P_{TT}(\cdot, j)$$

since $D_{P_{TT}} = (I - P_{TT})^{-1}$ it holds that $D_{P_{TT}}(j, \cdot)(I - P_{TT}) = I(j, \cdot)$ and furthermore using Eq. 24 gives

$$= \frac{\beta_j}{1-\beta_j} + \frac{1}{1/\beta_j - 1} D_{P_{TT}}(j, \cdot)P_{TT}(\cdot, j)$$

because $D_{P_{TT}}(j, \cdot)P_{TT}(\cdot, j) = \sum_{t=1}^{\infty}(P_{TT})^t(j, j) = D_{P_{TT}}(j, j) - 1$

$$= \frac{\beta_j}{1-\beta_j} + \frac{1}{1/\beta_j - 1}(D_{P_{TT}}(j, j) - 1)$$

again using Eq. 24

$$\begin{aligned} &= \frac{\beta_j}{1-\beta_j} + Y(j, j) - \frac{1}{1/\beta_j - 1} \\ &= Y(j, j). \end{aligned}$$

Since we showed that parts $(\star)$, $(\star\star)$ and $(\star\star\star)$ equal the corresponding parts of Eq. 23, this ends the proof of Statement (iii). □

The extended ergodic projector has the following interpretation. The stochastic matrix $\widetilde{\Pi}_P(\gamma)$ provides the stationary distributions for the ergodic classes and that of the Markov chains constructed for the transient states, which model a random surfer on the transient states that restarts at the specific transient state after an ergodic state is reached. Eventually, all scores can be uniformly combined to obtain an unprejudiced ranking, or a personalization vector v can be used for a personalized ranking.

We would like to close this section by commenting on the computational aspects of GR. First of all, one has to uncover the Markov chain structure in step 1. This can be done in linear time ($O(n+m)$) by applying Tarjan's algorithm (Tarjan 1972) from graph theory for finding the strongly connected components of a graph. Then for each ergodic class, one has to calculate the stationary distribution inside that ergodic class. This requires solving $\mathcal{E}$ times a system of linear equations of varying sizes corresponding to the ergodic class sizes. Alternatively, one might apply the power method per ergodic class for an approximation. Recall that all rows of the ergodic projector corresponding to states from an ergodic class are equal. Regarding $D_{P_{TT}}$, we have to invert matrix $I - P_{TT}$ (of size $|V_T| \times |V_T|$) as shown in Eq. 6 (recall that Π_P on the transient part is zero). In step 2 of GR, one has to compute $W(\gamma)$ which requires a matrix inverse for each ergodic class. However, when choosing γ small (which is advised), it may be sufficient to replace inside $W(\gamma)$

$$\sum_{k=0}^{\infty} \gamma^k Q^k \qquad \text{by} \qquad \sum_{k=0}^{K} \gamma^k Q^k$$

for K a small integer. This way the unwanted scenario as sketched in Example 2 is addressed with least amount of computational effort. So in conclusion, GR is surely computationally more involved than Google's PageRank. However, the reward is a more structural way of ranking and, as the following section illustrates, GR is still applicable to large real-life instances.

4 Real-life numerical experiment

In this section the new ranking approach GR is applied to the Wikipedia vote network dataset as given by Leskovec and Krevl (2014). In the words of Leskovec and Krevl (2014), Wikipedia is an online free encyclopedia written collaboratively by users around the world. A small part of Wikipedia contributors are administrators, who are users with access to additional technical features that aid in maintenance. In order for a user to become an administrator a request for adminship is issued and the Wikipedia community decides via an election who to promote to adminship. The voting history until January 3 2008 was collected by Leskovec and Krevl (2014) consisting of 2794 elections captured in a vote network given by $G = (V, E)$.

Nodes V in the vote network represent Wikipedia users and a directed edge $e = (i, j) \in E$ indicates that user i voted for user j to become admin. There are $n = 7115$ nodes (users) and $|E| = 103689$ directed edges (votes). For the construction of a Markov chain described by P we take $w(e) = 1$ for all $e \in E$. There are $|V_E| = 1005$ ergodic states and all of them are dangling nodes since they did not vote on other nodes (i.e., they are absorbing of nature). In total, $|V_T| = 6110$ of all states are transient, and the major strongly connected component consists of 1300 transient states. The remaining 4810 transient states did not get any votes but did vote on others. To give an impression of the instance, a plot of all edges in the vote network can be found in Fig. 4 in which each dot corresponds to a vote.

In the following two ranking methods are compared, i.e., Google's PageRank (GP) with $d = 0.85$ (which is the most common value for d in literature) and GR with $\gamma = 0$ (as all ergodic states are dangling nodes the choice of γ does not matter in this instance), both for a uniform personalization vector $v = \bar{1}/n$. For notational easiness we define $u := \bar{1}/n$.

As $\|\pi_{P_{0.85}(u)} - u^\top \Pi_P(0)\|_\infty \approx 0.64$ ($\|\cdot\|_\infty$ is the sum of absolute values which is a value ranging from 0 to 2) it follows that GP and GR lead to different scores. Specifically, numerical findings on the top 5 rankings of ergodic and transient states, respectively, are

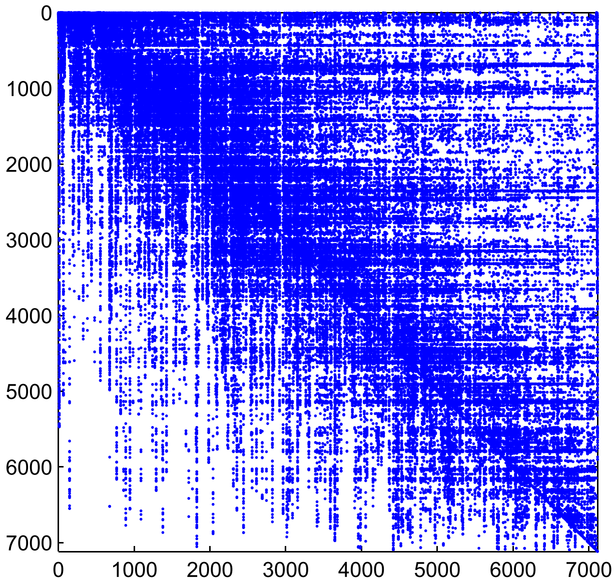


Fig. 4 Plot of all edges E, i.e., votes, in the Wikipedia vote network dataset from (Leskovec and Krevl 2014).

presented in Tables 1 and 2. Per method the top 5 indexes are provided in logical order. The two numbers in brackets provide the score of the node and its overall ranking.

The main observation is that GP significantly favors ergodic states above transient states compared to GR. E.g., the highest GP ranked transient state, i.e., state 3650, is ranked 56 while GR ranks the same transient state as third. The in-degree (i.e., number of votes received) of this transient state 3650 is 457 with $\bar{1}^\top P(\cdot, 3650) \approx 68$, while ergodic state 2270, i.e., the 2nd ranked state by GP, has in-degree 150 and $\bar{1}^\top P(\cdot, 2270) \approx 44$. Based on these first-order dynamics, it is thus questionable whether this ranking gap is reasonable. Furthermore, the preference of ergodic states by GP compared to GR follows from Fig. 5 which plots the score vectors of GP versus GR. Particularly, a distinction is made between transient and ergodic states in Fig. 5 by using dots and crosses, respectively. Placing your

Table 1 Top 5 rankings of ergodic states for the Wikipedia vote network dataset

Ranking Approach	Top 5: index (score, overall ranking)
GP ($d = 0.85$ and $v = u$)	2411 (9.14E−3, #1)
	2270 (7.03E−3, #2)
	6538 (6.04E−3, #3)
	1087 (5.67E−3, #4)
	6585 (5.38E−3, #5)
GR ($\gamma = 0$ and $v = u$)	2411 (3.89E−3, #1)
	2270 (3.41E−3, #4)
	1087 (2.79E−3, #6)
	6538 (2.52E−3, #8)
	4366 (2.33E−3, #9)

Table 2 Top 5 rankings of transient states for the Wikipedia vote network dataset

Ranking Approach	Top 5: index (score, overall ranking)
GP ($d = 0.85$ & $v = u$)	3650 (1.92E−3, #56)
	13 (1.54E−3, #85)
	5807 (1.50E−3, #92)
	2205 (1.09E−3, #174)
	2057 (1.04E−3, #188)
GR ($\gamma = 0$ and $v = u$)	5807 (3.87E−3, #2)
	3650 (3.82E−3, #3)
	13 (3.24E−3, #5)
	2205 (2.52E−3, #7)
	6059 (2.03E−3, #15)

hand on Fig. 5 and moving the hand from right to left reveals that only at the end the transient crosses come into sight in case of GP with $d = 0.85$, whereas moving the hand from top to bottom shows that GR mixes transient and ergodic states in the ranking, i.e., dots and crosses appear together. In particular, the top 10 ranking of GR consists of 4 transient states and 6 ergodic states and it holds that:

– 97% of the top 100 of GP are ergodic states.
– 44% of the top 100 of GR are ergodic states.

In the following, the preference of GP with $d = 0.85$ towards transient states are made more apparent by looking at the vote network in which the node sizes are based on the given scores. Particularly, more score means larger nodes and vice versa. Moreover, the ergodic states are indicated by green nodes and placed on the left side of the network and

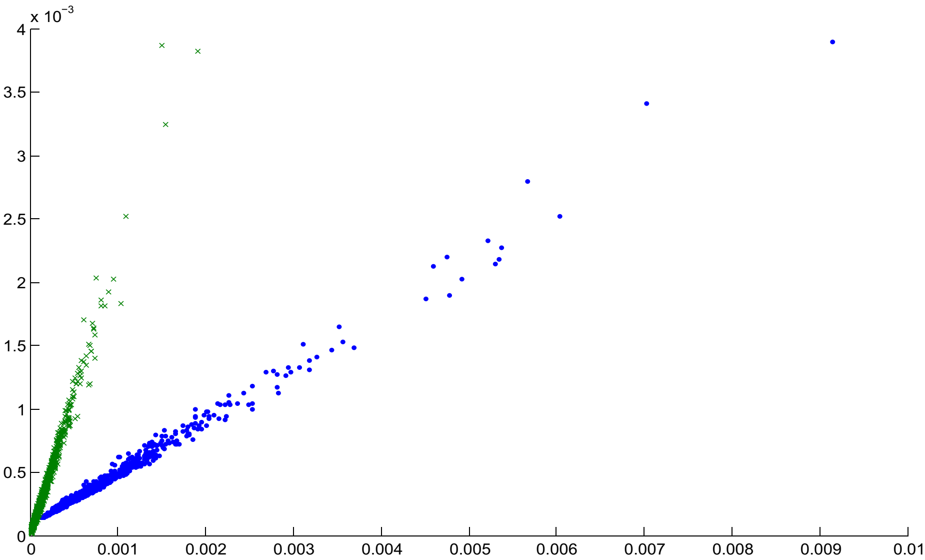


Fig. 5 For every ergodic state i: • plots $(\pi_{P_{0.85}(u)}(i), (u^\top \widetilde{\Pi}_P(0))(i))$. For every transient state j: × plots $(\pi_{P_{0.85}(u)}(j), (u^\top \widetilde{\Pi}_P(0))(j))$

Fig. 6 Wikipedia vote network where node sizes are based on GP scores ($d = 0.85$ & $v = u$)

the transient states are red nodes which are placed on the right side of the network. These voting networks in which node sizes are based on the scorings from GP and GR can be found in Figs. 6 and 7, respectively. The node sizes clearly show that GP gives significantly more score to ergodic states than to transient states compared to GR.

In order to get more insight into the question whether this is reasonable or not the vote network with node sizes based on in-degrees (i.e., the number of votes a node gets) is shown in Fig. 8. Note that as each vote counts equally in case the scoring is based on in-degrees this is a democracy. Figure 8 uncovers that there is a lot of voting going on among transient

Fig. 7 Wikipedia vote network where node sizes are based on GR scores ($\gamma = 0$ & $v = u$)

Fig. 8 Wikipedia vote network where node sizes are based on in-degrees

states and in fact the Wikipedia users with the most votes are transient of nature. This clearly indicates that intuitively GP tends to undervalue transient states, while the score distribution provided by GR seems to give a better score-balance between ergodic and transient states while also taking long-term dynamics into account. So overall for this instance, GR seems to give more appropriate importance score to the transient states compared to the GP for the fixed $d = 0.85$, which seems intuitively appealing.

GP with an alternative value for d provides most likely a more appropriate score distribution between transient and ergodic states. It holds that $\mathrm{Fr}_{\mathrm{GP}}(0.85) \approx 0.315$ whereas $\mathrm{Fr}_{\mathrm{GR}} \approx 0.635$ and the fraction of transient states compared to total number of states is $|V_T|/n \approx 0.859$. Figure 9 shows numerically what $\mathrm{Fr}_{\mathrm{GP}}(d)$ for $d \in [0, 1)$ in this instance

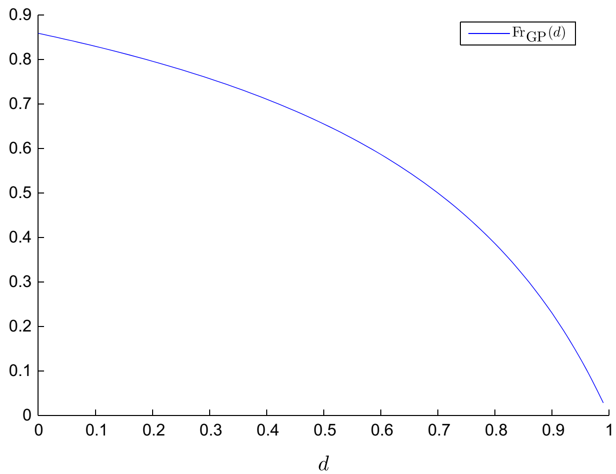


Fig. 9 Development of $\mathrm{Fr}_{\mathrm{Fr}}(d)$ for $d \in [0, 1)$ in the Wikipedia vote network

looks like. It shows that GP with $d \approx 0.55$ gives approximately the same fraction of score to transient states as GR.

5 Conclusion

The choice of the damping factor of Google's PageRank might have a large impact on the valuation of nodes that are of transient nature, a common feature in real-life networks. By keeping the underlying network intact, an approach is proposed that uses structural network dynamics to determine an appropriate score-distribution between transient and ergodic nodes. This way it takes away the unclear score trade-off in Google's PageRank between transient and ergodic nodes that is nestled in the choice of the damping factor. The newly introduced Generalized Ranking values transient states more appropriately than Google's PageRank (for a fixed damping factor) while maintaining the random walk interpretation, which is one of the strong points of Google's PageRank.

Ranking methodologies are mostly applied to large-scale data, computational aspects are thus of importance. As this article does not focus on computational issues, a direction of further research is to investigate how the Generalized Ranking can be implemented efficiently. Another topic of further research, is to analyze the extent to which the Generalized Ranking can be used for analyzing different type of instances in practice.

References

Aldous D, Fill J (2002) Reversible Markov Chains and Random Walks on Graphs. online version available at http://www.stat.berkeley.edu/users/aldous/RWG/book.html. Unfinished monograph, recompiled in 2014

Avrachenkov K, Litvak N (2006) The effect of new links on Google Pagerank. Stoch Model 22(2):319–331

Avrachenkov K, Litvak N, Pham KS (2008) A singular perturbation approach for choosing the PageRank damping factor. Internet Math 5(1-2):47–69

Berkhin P (2005) A survey on PageRank computing. Internet Math 2(1):73–120. doi:10.1080/15427951.2005.10129098

Berkhout J (2016) Google's PageRank algorithm for ranking nodes in general networks. In: Proceedings of the 13th international workshop on discrete event systems (WODES 2016), pp 153–158

Berkhout J, Heidergott BF (2014) A series expansion approach to risk analysis of an inventory system with sourcing. In: Proceedings of the 12th international workshop on discrete event systems (WODES 2014), vol 12, pp 510–515

Berkhout J, Heidergott BF (2015) Efficient algorithm for computing the ergodic projector of Markov multi-chains. Procedia Comput Sci 51:1818–1827. doi:10.1016/j.procs.2015.05.403. http://www.sciencedirect.com/science/article/pii/S1877050915012119

Bianchini M, Gori M, Scarselli F (2005) Inside PageRank. ACM Trans Internet Technol 5(1):92–128. doi:10.1145/1052934.1052938. http://delivery.acm.org/10.1145/1060000/1052938/p92-bianchini.pdf?ip=208.165.55.137&id=1052938&acc=ACTIVE SERVICE&key=842BC1E250410AEB.FDB887D4E02C11F2.4D4702B0C3E38B35.4D4702B0C3E38B35&CFID=710709363&CFTOKEN=16287877&__acm__=1441563203_9363b1527fecd54000

Boldi P, Santini M, Vigna S (2005) PageRank as a function of the damping factor. In: Proceedings of the 14th international conference on World Wide Web - WWW '05, p 557. doi:10.1145/1060745.1060827. http://portal.acm.org/citation.cfm?doid=1060745.1060827

Boldi P, Santini M, Vigna S (2009) PageRank: functional dependencies. ACM Trans Inf Syst 27(4):1–23. doi:10.1145/1629096.1629097

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

Bressan M, Peserico E (2010) Choose the damping, choose the ranking? J Discret Algorithms 8(2):199–213. doi:10.1016/j.jda.2009.11.001

Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30(1-7):107–117. http://www.sciencedirect.com/science/article/pii/S016975529800110X

Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A, Wiener J (2000) Graph structure in the Web. Comput Netw 33(1):309–320. doi:10.1016/S1389-1286(00)00083-9
Franceschet M (2011) PageRank: standing on the shoulders of giants. Commun ACM 54(6):92–101. http://dl.acm.org/citation.cfm?id=1953146
Fu HH, Lin DKJ, Tsai HT (2006) Damping factor in Google page ranking. Appl Stoch Model Bus Ind 22(5-6):431–444. doi:10.1002/asmb.656
Gleich DF (2015) PageRank beyond the web. SIAM Rev 57(3):321–363. doi:10.1137/140976649. arXiv:1407.5107
Heidergott BF, Hordijk A, van Uitert M (2007) Series expansions for finite-state Markov chains. Probab Eng Informational Sci 21(3):381–400. doi:10.1017/S0269964807000034. http://dl.acm.org/citation.cfm?id=1291117.1291120
Kartashov NV (1996) Strong stable markov chains. TBIMC Scientific Publishers, VSP
Kemeny JG, Snell JL (1976) Finite Markov chains: With a New Appendix "Generalization of a Fundamental Matrix. Springer, New York. http://books.google.nl/books/about/Finite_markov_chains.html?id=WORLAAAAMAAJ&pgis=1
Langville AN, Meyer CD (2003) Fiddling with PageRank. CRSC- TR03-34. North Carolina State University, Raleigh, NC. http://www.ncsu.edu/crsc/reports/reports03.html Tech. rep.
Langville AN, Meyer CD (2004) Deeper inside PageRank. Internet Math 1(3):335–380. doi:10.1080/15427951.2004.10129091
Langville AN, Meyer CD (2005) A survey of eigenvector methods for web information retrieval. SIAM Rev 47(1):135–161. doi:10.1137/S0036144503424786
Langville AN, Meyer CD (2011) Google's pagerank and beyond: The science of search engine rankings. Princeton University Press
Langville AN, Meyer CD (2012) Do the math: why no ranking system is no. 1. Sci Am 307(1)
Leskovec J, Krevl A (2014). SNAP Datasets: Stanford Large Network Dataset Collection, Webpage: https://snap.stanford.edu/data/wiki-Vote.html
Meyer CD (1975) The role of the group generalized inverse in the theory of finite Markov chains. SIAM Rev 17(3):443–464. doi:10.1137/1017044
Meyer CD (1982) Analysis of finite Markov chains by group inversion techniques. In: Campbell SL (ed) Recent applications of generalized inverses, research notes in mathematics, vol 66. Pitman, Boston, London, pp 50–81
Tarjan RE (1972) Depth-first search and linear graph algorithms. SIAM J Comput 1(2):146–160
Xia L, Glynn PW (2016) A generalized fundamental matrix for computing fundamental quantities of Markov systems. arXiv:1604.04343v2 [math.OC]. http://arxiv.org/abs/1604.04343

Joost Berkhout was born in Alkmaar, the Netherlands, in 1990. He received the B.Sc., M.Sc. and Ph.D. degrees in the field of Econometrics and Operations Research from the Vrije Universiteit Amsterdam in 2011, 2012 and 2016, respectively. During his Ph.D. trajectory, Joost was awarded with the Student Best Paper Award at the International Workshop on Discrete Event Systems (WODES) 2016 in Xi'an (China). In 2016, he joined the Centrum Wiskunde & Informatica (CWI) in Amsterdam as a post-doctoral researcher on a project about scheduling optimization in the biomaterial industry. His current research interests include Markov chain theory and applications, optimization of simulated systems and scheduling optimization.

Bernd F. Heidergott is professor of Stochastic Optimization at the Department of Econometrics and Operations Research at the Vrije Universiteit Amsterdam, the Netherlands. He received his Ph.D. degree from the University of Hamburg, Germany, in 1996, and held postdoc positions at various universities before joining the Vrije Universiteit. He is programme director of the B.S.c and M.Sc. Econometrics and Operations Research, and research fellow of the Tinbergen Institute and of EURANDOM. His research interests are optimization and control of discrete event systems, perturbation analysis of Markov chains and max-plus algebra.