A Structural Result for Personalized PageRank and its Algorithmic Consequences

Daniel Vial
University of Michigan
Ann Arbor, MI, USA
dvial@umich.edu

Vijay Subramanian
University of Michigan
Ann Arbor, MI, USA
vgsubram@umich.edu

ACM Reference Format:

Daniel Vial and Vijay Subramanian. 2019. A Structural Result for Personalized PageRank and its Algorithmic Consequences. In *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '19 Abstracts)*, June 24–28, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3309697.3331481>

1 INTRODUCTION

Many natural and man-made systems can be represented as graphs, sets of objects (called nodes) and pairwise relations between these objects (called edges). These include the brain, which contains neurons (nodes) that exchange signals through chemical pathways (edges), the Internet, which contains websites (nodes) that are connected via hyperlinks (edges), etc. To study graphs, researchers in diverse domains have used Personalized PageRank (PPR) [6]. Informally, PPR assigns to each node v a vector π_v , where $\pi_v(w)$ describes the importance or relevance of node w from the perspective of v . PPR has proven useful in many applications, both practical and graph-theoretic. Examples include recommending who a user should follow on Twitter [7] (v may wish to follow w if $\pi_v(w)$ is large) and local graph partitioning [2] (the set of nodes w with large $\pi_v(w)$ can be viewed as a community surrounding v).

In this work, we study the matrix containing the set of PPR vectors, $\Pi_n = \{\pi_v\}_{v \in V_n}$, where V_n is the node set of an underlying graph of n nodes. Our focus is both structural and algorithmic. From a structural perspective, we consider the dimensionality of this matrix. Our interest in this quantity arises due to an apparent paradox: on the one hand, Π_n is full rank, and is thus an n^2 -dimensional object; on the other hand, many graphs appearing in practice satisfy $|E_n| = O(n)$ (i.e. they are sparse), and are thus n -dimensional objects. Hence, our first goal is to understand whether this gap is truly present, or if it disappears under a different definition of dimensionality. From an algorithmic perspective, we consider schemes for fast estimation of Π_n from the given graph G_n . Such schemes are necessary because computing Π_n has complexity $O(n^3)$, which is infeasible for massive graphs arising in modern domains.

Our primary contributions are twofold. First, we resolve the apparent dimensionality paradox by showing that, for a notion of dimensionality similar to rank, Π_n has dimensionality $O(n^{c_1})$ for some $c_1 < 2$. In other words, we show the dimensionality

gap between the PPR matrix and the underlying graph is much smaller than it appears at first glance. Second, we leverage this low dimensionality to design an algorithm to estimate Π_n in a manner analogous to (online) nonnegative matrix factorization (NMF). We establish an l_1 accuracy guarantee for each PPR vector and bound the algorithm's complexity as $O(n^{c_2})$ for some $c_2 < 2$. This improves upon all estimation schemes in the literature, the most competitive of which have complexity $O(n^2 \log n)$ in our setting. Our theoretical results hold assuming the underlying graph follows a particular generative model; roughly, the sparse directed configuration model with heavy-tailed in-degrees. In the full paper, we show empirically that similar conclusions hold for other random graph models and for real graphs (including social networks and web crawls).

In the remainder of this extended abstract, we define more precisely the problems and results described above.

2 BACKGROUND

We begin by defining our main object of study, the PPR matrix. For this, first let P denote the row-normalized adjacency of a given directed graph $G_n = (V_n, E_n)$ of n nodes. Additionally, let $\alpha_n \in (0, 1)$, let $\mathbf{1}_n$ denote the ones vector of length n , let $v \in V_n$, and let $e_v^\top$ denote the length- n vector with 1 in the v -th coordinate and 0 elsewhere. We then define PPR as follows:

DEFINITION 1. For $v \in V_n$, let $P_v = (1 - \alpha_n)P + \alpha_n \mathbf{1}_n e_v^\top$. Then the PPR vector corresponding to v , denoted π_v , is the stationary distribution of the Markov chain with transition matrix P_v . Additionally, the PPR matrix, denoted Π_n , is the matrix with rows $\{\pi_v\}_{v \in V_n}$.

As mentioned previously, one can view $\pi_v(w)$ as a measure of the importance or relevance of w from the perspective of v . To see this, let $\{X_i\}_{i=0}^\infty$ denote the Markov chain with transition matrix P and let $L \sim \text{Geometric}(\alpha_n)$. Then, as shown in the full paper,

$$\pi_v(w) = \alpha_n \mathbb{E} \left[\sum_{i=0}^{L-1} \mathbf{1}(X_i = w) \mid X_0 = v \right],$$

which implies $\pi_v(w)$ is large when w is frequently visited (a notion of importance or relevance) on $\text{Geometric}(\alpha_n)$ -length walks starting at v (a notion of v 's perspective).

We also note that α_n in Definition 1 is typically assumed to be a constant; in contrast, we let $\alpha_n = \Theta(1/\log n)$. Using recent results from the mixing times literature [3], we show that, if instead $\alpha_n = o(1/\log n)$, Π_n is arbitrarily close to a rank one matrix for n large. Hence, the choice $\alpha_n = \Theta(1/\log n)$ occurs just before a threshold, after which PPR dimensionality becomes trivial.

3 STRUCTURAL RESULT

Our main structural focus is the dimensionality of Π_n . A standard measure of this dimension is $\text{rank}(\Pi_n)$, but one can show

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS '19 Abstracts, June 24–28, 2019, Phoenix, AZ, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6678-6/19/06.

<https://doi.org/10.1145/3309697.3331481>

$\text{rank}(\Pi_n) = n \forall n \in \mathbb{N}$, so we will instead consider a different notion of dimensionality. This notion is motivated by the following observation: the rank of the matrix with rows $\{x_i\}_{i \in I}$ (where I is some finite set) can be bounded by $|X' \cup X''|$, where $X' \subset \{x_i\}_{i \in I}$ and $X'' = \{x_i \notin X' : x_i \text{ is not a linear combination of } X'\}$. We will relax this by only including in X'' those $x_i \notin X'$ that are not “close” to a linear combination of X' . Mathematically, for $\epsilon > 0$ we define

$$\min_{K_n \subset V_n} \Delta(K_n, \epsilon) = |K_n| + |\{v \in V_n \setminus K_n : B_v(K_n, \epsilon) \text{ holds}\}|, \quad (1)$$

where $B_v(K_n, \epsilon)$ is the event

$$\left\{ \inf_{\{\beta_v(k)\}_{k \in K_n} \subset \mathbb{R}} \left\| \pi_v - \left(\sum_{k \in K_n} \beta_v(k) \pi_k + \alpha_n e_v^T \right) \right\|_1 \geq \epsilon \right\}. \quad (2)$$

As discussed in the full paper, analyzing the dimensionality measure (1) for an arbitrary graph is quite difficult, in part because the graph’s local structure plays a prominent role. For this reason, we instead analyze (1) assuming G_n follows a particular generative model, which guarantees a well-behaved local structure. The specific model considered is the directed configuration model (DCM) [5]. In this model, one realizes a degree sequence $\{N_v, D_v\}_{v \in V_n}$, connects N_v incoming and D_v outgoing half-edges to each $v \in V_n$, and randomly pairs incoming half-edges with outgoing half-edges to form edges. In addition to assuming this generative model, we require the degree sequence to satisfy certain assumptions. First, we assume the graph is sparse, meaning that several empirical moments (such as the average out-degree $\sum_{v \in V_n} D_v/n$) converge to finite limits with high probability (w.h.p.) as $n \rightarrow \infty$. Second, we assume that a vanishing fraction of nodes contains a non-vanishing fraction of incoming edges; mathematically, we require

$$\exists K_n \subset V_n \text{ s.t. } \mathbb{E}|K_n| = o(n), \frac{\sum_{k \in K_n} N_k}{\sum_{v \in V_n} N_v} \xrightarrow{n \rightarrow \infty} p \in (0, 1) \text{ w.h.p.}$$

Taken together, these assumptions suggest a graph with constant out-degrees but heavy-tailed in-degrees. This in turn suggests a graph like Twitter, where users do not follow a sizable portion of all users (constant out-degrees) and a small number of celebrities have huge numbers of followers (heavy-tailed in-degrees).

We now state our first result, which bounds the dimensionality measure (1) under the assumptions described above. In particular, we show that (1) scales sublinearly w.h.p. as $n \rightarrow \infty$:

THEOREM 1. *Given the random graph model and assumptions described above, we have for any $\epsilon > 0$ independent of n ,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\Delta(K_n, \epsilon) \geq n^{c_1} \right] = 0,$$

where $c_1 \in (0, 1)$ depends on ϵ , the choice of α_n , and certain parameters of the degree sequence.

The proof of Theorem 1 is one of the main theoretical contributions of this work and (at a high level) proceeds as follows. First, we show that (1) relates closely to the probability that a certain random walk hits K_n . Second, we approximate the graph construction with a certain branching process (using techniques from [4]). Third, we analyze the relevant hitting probabilities on the tree resulting from this branching process (the comparatively simple structure of the tree makes this analysis feasible).

4 ALGORITHMIC CONSEQUENCES

Our second objective is to leverage the low dimensionality guaranteed by Theorem 1 algorithmically. The basic algorithm we consider is motivated by (1)-(2) and proceeds as follows:

- (1) Choose K_n and estimate $\{\pi_k\}_{k \in K_n}$ directly.
- (2) For $v \notin K_n$ s.t. $B_v(K_n, \epsilon)$ holds, estimate π_v directly.
- (3) For $v \notin K_n$ s.t. $B_v(K_n, \epsilon)$ fails, estimate π_v as a linear combination of the $\{\pi_k\}_{k \in K_n}$ estimates.

Under this scheme, the number of nodes v for which π_v is estimated directly is (by definition) the dimensionality measure (1). Hence, Theorem 1 guarantees that direct estimation is necessary for only a vanishing fraction of all nodes. Additionally, (2) suggests that the resulting estimates will be accurate in the l_1 norm. However, this basic scheme is simply a proof-of-concept; it is not immediately clear how to directly estimate π_v , nor is it clear how to determine if $B_v(K_n, \epsilon)$ holds or fails. Our main algorithmic contributions are to overcome these issues, and can be roughly described as follows:

- The algorithm from [2] can be used for the direct estimation in Steps 1 and 2 above and guarantees l_1 accuracy.
- A certain choice of $\{\beta_v(k)\}_{k \in K_n}$ approximates the infimum in (2); moreover, this $\{\beta_v(k)\}_{k \in K_n}$ can be estimated using a modified version of the algorithm from [1].
- The errors incurred by estimating this $\{\beta_v(k)\}_{k \in K_n}$ and $\{\pi_k\}_{k \in K_n}$ can be controlled by careful parameter choices, ensuring l_1 accuracy for Step 3 above.
- The matrices $\{\beta_v(k)\}_{v \notin K_n, k \in K_n}$ and $\{\pi_k\}_{k \in K_n}$ can be carefully “sparsified” (i.e. certain elements can be set to zero), ensuring that the matrix multiplication in Step 3 is efficient.

Ultimately, these insights allow us to prove the following:

THEOREM 2. *Under the assumptions of Theorem 1, our algorithm outputs estimates $\{\hat{\pi}_v\}_{v \in V_n}$ of $\{\pi_v\}_{v \in V_n}$ satisfying $\|\hat{\pi}_v - \pi_v\|_1 < \epsilon \forall v \in V_n$, where $\epsilon > 0$ is an input to the algorithm. With high probability, its complexity is $O(n^{c_2})$ for some $c_2 < 2$ that depends on ϵ , the choice of α_n , and certain parameters of the degree sequence.*

This complexity guarantee is strictly better than any in the literature, the most competitive of which (those from [1, 2]) have complexity $O(n^2 \log n)$ in our setting. We also note that the basic framework of our algorithm is similar to that found in [8]. However, the authors of [8] provide no analytical guarantees for their scheme.

REFERENCES

- [1] Reid Andersen, Christian Borgs, Jennifer Chayes, John Hopcroft, Vahab Mirrokni, and Shang-Hua Teng. 2008. Local computation of PageRank contributions. *Internet Mathematics* 5, 1-2 (2008), 23–45.
- [2] Reid Andersen, Fan Chung, and Kevin Lang. 2006. Local graph partitioning using PageRank vectors. In *2006 IEEE Symposium on Foundations of Computer Science*. IEEE, 475–486.
- [3] Charles Bordenave, Pietro Caputo, and Justin Salez. 2018. Random walk on sparse random digraphs. *Probability Theory and Related Fields* 170, 3-4 (2018), 933–960.
- [4] Ningyuan Chen, Nelly Litvak, and Mariana Olvera-Cravioto. 2017. Generalized PageRank on directed configuration networks. *Random Structures & Algorithms* 51, 2 (2017), 237–274.
- [5] Ningyuan Chen and Mariana Olvera-Cravioto. 2013. Directed random graphs with given degree distributions. *Stochastic Systems* 3, 1 (2013), 147–186.
- [6] David F Gleich. 2015. PageRank beyond the Web. *SIAM Rev.* 57, 3 (2015), 321–363.
- [7] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. 2013. WTF: The who to follow service at Twitter. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 505–514.
- [8] Glen Jeh and Jennifer Widom. 2003. Scaling personalized web search. In *Proceedings of the 12th International Conference on World Wide Web*. ACM, 271–279.