HAL
open science

PageRank in undirected random graphs

Konstantin Avrachenkov, Arun Kadavankandy, Liudmila Ostroumova, Andrei Raigorodskii

To cite this version:

Konstantin Avrachenkov, Arun Kadavankandy, Liudmila Ostroumova, Andrei Raigorodskii. PageRank in undirected random graphs. [Research Report] RR-8983, Inria Sophia Antipolis. 2015, pp.26. hal-01227383v2

HAL Id: hal-01227383

https://inria.hal.science/hal-01227383v2

Submitted on 25 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PageRank in Undirected Random Graphs

K. Avrachenkov, A. Kadavankandy,
L. Ostroumova Prokhorenkova, and A. Raigorodskii.

RESEARCH
REPORT
N° 8983
Nov 2016
Project-Team Maestro

ISSN 0249-6399 ISRN INRIA/RR--8983--FR+ENG

PageRank in Undirected Random Graphs

K. Avrachenkov, A. Kadavankandy*,
L. Ostroumova Prokhorenkova, and A. Raigorodskii.

Project-Team Maestro

Research Report n° 8983 — Nov 2016 — 26 pages

Abstract: PageRank has numerous applications in information retrieval, reputation systems, machine learning, and graph partitioning. In this paper, we study PageRank in undirected random graphs with an expansion property. The Chung-Lu random graph is an example of such a graph. We show that in the limit, as the size of the graph goes to infinity, PageRank can be approximated by a mixture of the restart distribution and the vertex degree distribution. We also extend the result to Stochastic Block Model (SBM) graphs, where we show that there is a correction term that depends on the community partitioning.

Key-words: PageRank, undirected random graphs, expander graphs, Chung-Lu random graphs, Stochastic Block Model

* Corresponding author, arun.kadavankandy@inria.fr

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Le Pagerank sur les Graphes Aléatoires Non-oriéntés

Résumé : Le PageRank connait de nombreuses applications sur les domaines de la recuperation d'informations, des systemes de reputations, de l'apprentissage automatique, et du partitionnement des graphes. Dans ce travail, nous étudions le PageRank sur les graphes aléatoires non-oriéntés avec une propriete de dilatation, par exemple les graphes de Chung-Lu. Nous montrons que dans la limite, lorsque la taille du graphe tend vers l'infini, le PageRank peut etre representé par un mélange de la distribution de redemmarage et la distribution de dégrées des sommets du graphe. Nous considérons aussi le Stochastic Block Model (SBM), où on découvre qu'il existe dans l'expression asymptotique du PageRank une terme qui est en fonctiion de la structure de communauté du graphe.

Mots-clés : PageRank, Graphes Aléatoires non-oriéntés, Graphes expanseurs, Chung-Lu, Stochastic Block Model

1 Introduction

PageRank has numerous applications in information retrieval [22, 32, 38], reputation systems [21, 26], machine learning [4, 5], and graph partitioning [1, 12]. A large complex network can often be conveniently modeled by a random graph. It is surprising that not many analytic studies are available for PageRank in random graph models. We mention the work [6] where PageRank was analysed in preferential attachment models and the more recent works [10, 11], where PageRank was analysed in directed configuration models. According to several studies [18, 20, 29, 37], PageRank and in-degree are strongly correlated in directed networks such as the Web graph. Apart from some empirical studies [9, 33], to the best of our knowledge, there is no rigorous analysis of PageRank on basic undirected random graph models such as the Erdős-Rényi graph [19] or the Chung-Lu graph [14]. In this paper, we attempt to fill this gap and show that under certain conditions on the preference vector and the spectrum of the graphs, PageRank in these models can be approximated by a mixture of the preference vector and the vertex degree distribution when the size of the graph goes to infinity. First, we show the convergence in total variation norm for a general family of random graphs with expansion property. Then, we specialize the results for the Chung-Lu random graph model proving the element-wise convergence. We also analyse the asymptotics of PageRank on Stochastic Block Model (SBM) graphs, which are random graph models used to benchmark community detection algorithms. In these graphs the asymptotic expression for PageRank contains an additional correction term that depends on the community partitioning. This demonstrates that PageRank captures properties of the graph not visible in the stationary distribution of a simple random walk.We conclude the paper with numerical experiments and several future research directions.

2 Definitions

Let $G^{(n)} = (V^{(n)}, E^{(n)})$ denote a family of random graphs, where $V^{(n)}$ is a vertex set, $|V^{(n)}|= n$, and $E^{(n)}$ is an edge set, $|E^{(n)}|= m$. Matrices and vectors related to the graph are denoted by bold letters, while their components are denoted by non-bold letters. We denote by $\mathbf{A}^{(n)}$ the associated adjacency matrix with elements

$$A_{ij}^{(n)} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are connected,} \\ 0, & \text{otherwise,} \end{cases}$$

In the interest of compactness of notation, the superscript n is dropped when it is not likely to cause confusion. In this work, since we analyze PageRank on undirected graphs, we have $\mathbf{A}^T = \mathbf{A}$. The personalized PageRank is denoted by $\boldsymbol{\pi}$. We consider unweighted graphs; however our analysis easily extends to some families of weighted undirected graphs. Let $\mathbf{1}$ be a column vector of n ones and let $\mathbf{d} = \mathbf{A1}$ be the vector of degrees. It is helpful to define $\mathbf{D} = \text{diag}(\mathbf{d})$, a diagonal matrix with the degree sequence on its diagonal.

Let $\mathbf{P} = \mathbf{AD}^{-1}$ be column-stochastic Markov transition matrix corresponding to the standard random walk on the graph and let $\mathbf{Q} = \mathbf{D}^{-1/2}\mathbf{AD}^{-1/2}$ be the symmetrized transition matrix, whose eigenvalues are the same as those of $\mathbf{P}$. Note that the symmetrized transition matrix is closely related to the normalized Laplacian $\boldsymbol{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{AD}^{-1/2} = \mathbf{I} - \mathbf{Q}$ [13], where $\mathbf{I}$ is the identity matrix. Further we will also use the resolvent matrix $\mathbf{R} = [\mathbf{I} - \alpha\mathbf{P}]^{-1}$ and the symmetrized resolvent matrix $\mathbf{S} = [\mathbf{I} - \alpha\mathbf{Q}]^{-1}$.

Note that since $\mathbf{Q}$ is a symmetric matrix, its eigenvalues λ_i, $i = 1, ..., n$ are real and can be arranged in decreasing order, i.e., $\lambda_1 \geq \lambda_2 \geq ...$. In particular, we have $\lambda_1 = 1$. The value $\delta = 1 - \max\{|\lambda_2|, |\lambda_n|\}$ is called the spectral gap.

In what follows, let K, C be arbitrary constants independent of graph size n, which may change from one line to the next (of course, not causing any inconsistencies).

For two functions $f(n), g(n)$, $g(n) = O(f(n))$ if $\exists C, N$ such that $\left|\frac{g(n)}{f(n)}\right| \leq C$, $\forall n > N$ and $g(n) = o(f(n))$ if $\limsup_{n\to\infty} \left|\frac{g(n)}{f(n)}\right| = 0$. Also $f(n) = !(g(n))$ or $f(n) \gg g(n)$ if $g(n) = o(f(n))$.

We use $\mathbb{P}, \mathbb{E}$ to denote probability and expectation respectively. An event E is said to hold with high probability (w.h.p.) if $\exists N$ such that (s.t.) $\mathbb{P}(E) \geq 1 - O(n^{-c})$ for some $c > 0$, $\forall n > N$. Recall that if a finite number of events hold true w.h.p., then so does their intersection. Furthermore, we say that a sequence of random variables $X_n = o(1)$ w.h.p. if there exists a function $\psi(n) = o(1)$ such that the event $\{X_n \leq \psi(n)\}$ holds w.h.p.

In the first part of the paper, we study the asymptotics of PageRank for a family of random graphs with the following two properties:
Property 1: $d_{max}^{(n)}/d_{min}^{(n)} \leq K$, where $d_{max}^{(n)}$ and $d_{min}^{(n)}$ are the maximum and minimum degrees, respectively.

Property 2: W.h.p., $\max\{|\lambda_2^{(n)}|, |\lambda_n^{(n)}|\} = o(1)$. The above two properties can be regarded as a variation of the expansion property. In the standard case of an expander family, one requires the graphs to be regular and the spectral gap $\delta = 1 - \max\{|\lambda_2|, |\lambda_n|\}$ to be bounded away from zero (see, e.g., [36]). Property 1 is a relaxation of the regularity condition, whereas Property 2 is stronger than the requirement for the spectral gap to be bounded away from zero. These two properties allow us to consider several standard families of random graphs such as ER graphs, regular random graphs with increasing average degrees, and Chung-Lu graphs. For Chung-Lu graphs Property 1 imposes some restriction on the degree spread of the graph.

Remark: Property 2 implies that the graph is connected w.h.p., since the spectral gap is strictly greater than zero.

Later, we study the asymptotics of PageRank for specific classes of random graphs namely the Chung-Lu graphs, and the Stochastic Block Model. Recall that the Personalized PageRank vector with preference vector $\mathbf{v}$ is defined as the stationary distribution of a modified Markov chain with transition matrix

$$\widetilde{\mathbf{P}} = \alpha\mathbf{P} + (1-\alpha)\mathbf{v}\mathbf{1}^T, \tag{1}$$

where α is the so-called damping factor [22]. In other words, $\boldsymbol{\pi}$ satisfies

$$\boldsymbol{\pi} = \widetilde{\mathbf{P}}\boldsymbol{\pi}, \tag{2}$$

or,

$$\boldsymbol{\pi} = (1-\alpha)[\mathbf{I} - \alpha\mathbf{P}]^{-1}\mathbf{v} = (1-\alpha)\mathbf{R}\mathbf{v}, \tag{3}$$

where (3) holds when $\alpha < 1$.

3 Convergence in total variation

We recall that for two discrete probability distributions u and v, the total variation distance $d_{\mathrm{TV}}(u, v)$ is defined as $d_{\mathrm{TV}}(u, v) = \frac{1}{2}\sum_i |u_i - v_i|$. This can also be thought of as the L^1-norm distance measure in the space of probability vectors, wherein for $\mathbf{x} \in \mathbb{R}^n$, the L^1-norm is defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$. Since for any probability vector $\boldsymbol{\pi}$, $\|\boldsymbol{\pi}\|_1 = 1$ $\forall n$, it makes sense to talk about convergence in 1-norm or TV-distance. Also recall that for a vector $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_2 = \sqrt{\sum_i |x_i|^2}$ is the L^2-norm. Now we are in a position to formulate our first result.

Proposition 1 *Let a family of graphs $G^{(n)}$ satisfy Properties 1 and 2. If, in addition, $\|\mathbf{v}\|_2 = O(1/\sqrt{n})$, PageRank can be asymptotically approximated in total variation norm by a mixture of the restart distribution $\mathbf{v}$ and the vertex degree distribution. Namely, w.h.p.,*

$$d_{TV}(\boldsymbol{\pi}^{(n)}, \overline{\boldsymbol{\pi}}^{(n)}) = o(1) \quad \text{as } n \to \infty,$$

where

$$\overline{\boldsymbol{\pi}}^{(n)} = \frac{\alpha \mathbf{d}^{(n)}}{\mathrm{vol}(G^{(n)})} + (1-\alpha)\mathbf{v}, \tag{4}$$

with $\mathrm{vol}(G^{(n)}) = \sum_i d_i^{(n)}$.

Observations:

1. This result says that PageRank vector asymptotically behaves like a convex combination of the preference vector and the stationary vector of a standard random walk with transition matrix $\mathbf{P}$; with the weight being α, and that it starts to resemble the random walk stationary vector as α gets close to 1.

2. One of the possible intuitive explanations of the result of Theorem 1 is based on the observation that when Properties 1 & 2 hold, as $n \to \infty$, the random walk mixes approximately in one step and so for any probability vector $\mathbf{x}$ $\mathbf{P}\mathbf{x}$ is roughly equal to $\mathbf{d}/\mathrm{vol}(G)$, the stationary distribution of the simple random walk. The proposed asymptotic approximation for PageRank can then be seen to follow from the series representation of PageRank if we replace $\mathbf{P}\mathbf{v}$ by $\mathbf{d}/\mathrm{vol}(G)$. Note that since $\mathbf{d}/\mathrm{vol}(G)$ is the stationary vector of the simple random walk, if $\mathbf{P}\mathbf{v} = \mathbf{d}/\mathrm{vol}(G)$, it also holds that $\mathbf{P}^k\mathbf{v} = \mathbf{d}/\mathrm{vol}(G), \forall k \geq 2$. Making these substitutions in the series representation of PageRank, namely

$$\boldsymbol{\pi} = (1-\alpha)\left(\mathbf{I} + \alpha\mathbf{P} + \alpha^2\mathbf{P}^2 + \ldots\right)\mathbf{v}, \tag{5}$$

   we obtain

$$\begin{aligned}\boldsymbol{\pi} &= (1-\alpha)\mathbf{v} + (1-\alpha)\alpha(1+\alpha+\alpha^2+\ldots)\frac{\mathbf{d}}{\mathrm{vol}(G)} \\ &= (1-\alpha)\mathbf{v} + \alpha\frac{\mathbf{d}}{\mathrm{vol}(G)}.\end{aligned}$$

3. The condition on the 2-norm of the preference vector $\mathbf{v}$ can be viewed as a constraint on its allowed localization.

Proof of Theorem 4: First observe from (1) that when $\alpha = 0$, we have $\widetilde{\mathbf{P}} = \mathbf{v}\mathbf{1}^T$, hence from (2) we obtain $\boldsymbol{\pi} = \mathbf{v}$, since $\mathbf{1}^T\boldsymbol{\pi} = 1$. Similarly for the case $\alpha = 1$, $\widetilde{\mathbf{P}} = \mathbf{P}$ and so $\boldsymbol{\pi}$ in this case is just the stationary distribution of the original random walk, which is well-defined and equals $\frac{\mathbf{d}}{\mathrm{vol}(G)}$ since by Property 2 the graph is connected. Examining (4) for these two cases we can see that the statement of the theorem holds trivially for both $\alpha = 0$ and $\alpha = 1$. In what follows, we consider the case $0 < \alpha < 1$. We first note that the matrix $\mathbf{Q} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ can be written as follows by Spectral Decomposition Theorem [7]:

$$\mathbf{Q} = \mathbf{u}_1\mathbf{u}_1^T + \sum_{i=2}^{n} \lambda_i \mathbf{u}_i\mathbf{u}_i^T, \tag{6}$$

where $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues and $\{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n\}$ with $\mathbf{u}_i \in \mathbf{R}^n$ and $\|\mathbf{u}_i\|_2 = 1$ are the corresponding orthogonal eigenvectors of $\mathbf{Q}$. Recall that $\mathbf{u}_1 = \mathbf{D}^{1/2}\mathbf{1}/\sqrt{\mathbf{1}^T\mathbf{D}\mathbf{1}}$ is the Perron–Frobenius eigenvector. Next, we rewrite (3) in terms of the matrix $\mathbf{Q}$ as follows

$$\boldsymbol{\pi} = (1-\alpha)\mathbf{D}^{1/2}[\mathbf{I} - \alpha\mathbf{Q}]^{-1}\mathbf{D}^{-1/2}\mathbf{v}. \tag{7}$$

Substituting (6) into (7), we obtain

$$\begin{aligned}\boldsymbol{\pi} &= (1-\alpha)\mathbf{D}^{1/2}\left(\frac{1}{1-\alpha}\mathbf{u}_1\mathbf{u}_1^T + \sum_{i=2}^{n}\frac{1}{1-\alpha\lambda_i}\mathbf{u}_i\mathbf{u}_i^T\right)\mathbf{D}^{-1/2}\mathbf{v} \\ &= \mathbf{D}^{1/2}\mathbf{u}_1\mathbf{u}_1^T\mathbf{D}^{-1/2}\mathbf{v} + (1-\alpha)\mathbf{D}^{1/2}\left(\sum_{i\neq 1}\frac{1}{1-\alpha\lambda_i}\mathbf{u}_i\mathbf{u}_i^T\right)\mathbf{D}^{-1/2}\mathbf{v}.\end{aligned}$$

Let us denote the error vector by $\boldsymbol{\epsilon} = \boldsymbol{\pi} - \overline{\boldsymbol{\pi}}$. Note that since $\mathbf{u}_1 = \frac{\mathbf{D}^{1/2}\mathbf{1}}{\sqrt{\mathrm{vol}(G)}}$, we can write $\overline{\boldsymbol{\pi}}$ as

$$\begin{aligned}\overline{\boldsymbol{\pi}} &= \alpha\frac{\mathbf{d}}{\mathrm{vol}(G)} + (1-\alpha)\mathbf{v} \\ &\overset{(a)}{=} \alpha\frac{\mathbf{D}\mathbf{1}\mathbf{1}^T\mathbf{v}}{\mathrm{vol}(G)} + (1-\alpha)\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\mathbf{v} \\ &= \alpha\mathbf{D}^{1/2}\frac{\mathbf{D}^{1/2}\mathbf{1}}{\sqrt{\mathrm{vol}(G)}}\frac{\mathbf{1}^T\mathbf{D}^{1/2}}{\sqrt{\mathrm{vol}(G)}}\mathbf{D}^{-1/2}\mathbf{v} + (1-\alpha)\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\mathbf{v} \\ &= \alpha\mathbf{D}^{1/2}\mathbf{u}_1\mathbf{u}_1^T\mathbf{D}^{-1/2}\mathbf{v} + (1-\alpha)\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\mathbf{v},\end{aligned}$$

where in (a) above we used the fact that $\mathbf{1}^T\mathbf{v} = 1$, since $\mathbf{v}$ is a probability vector. Then, we can write $\boldsymbol{\epsilon}$ as

$$\begin{aligned}\boldsymbol{\epsilon} &= \boldsymbol{\pi} - \alpha\mathbf{D}^{1/2}\mathbf{u}_1\mathbf{u}_1^T\mathbf{D}^{-1/2}\mathbf{v} - (1-\alpha)\mathbf{D}^{1/2}\mathbf{I}\mathbf{D}^{-1/2}\mathbf{v} \\ &= (1-\alpha)\mathbf{D}^{1/2}\left(\sum_{i\neq 1}\frac{\mathbf{u}_i\mathbf{u}_i^T}{1-\alpha\lambda_i} - (\mathbf{I} - \mathbf{u}_1\mathbf{u}_1^T)\right)\mathbf{D}^{-1/2}\mathbf{v} \\ &= (1-\alpha)\mathbf{D}^{1/2}\left(\sum_{i\neq 1}\mathbf{u}_i\mathbf{u}_i^T\frac{\alpha\lambda_i}{1-\alpha\lambda_i}\right)\mathbf{D}^{-1/2}\mathbf{v}.\end{aligned} \tag{8}$$

Now let us bound the L^1-norm $\|\boldsymbol{\epsilon}\|_1$ of the error:

$$\begin{aligned}\|\boldsymbol{\epsilon}\|_1/(1-\alpha) &\overset{(a)}{\leq} \sqrt{n}\|\boldsymbol{\epsilon}\|_2/(1-\alpha) \\ &\overset{(b)}{\leq} \sqrt{n}\|\mathbf{D}^{1/2}\|_2\left\|\sum_{i\neq 1}\mathbf{u}_i\mathbf{u}_i^T\frac{\alpha\lambda_i}{1-\alpha\lambda_i}\right\|_2\|\mathbf{D}^{-1/2}\|_2\|\mathbf{v}\|_2 \\ &\overset{(c)}{\leq} \sqrt{d_{max}/d_{min}}\sqrt{n}\max_{i>1}\left|\frac{\alpha\lambda_i}{1-\alpha\lambda_i}\right|\|\mathbf{v}\|_2 \\ &\leq C\sqrt{d_{max}/d_{min}}\max(|\lambda_2|, |\lambda_n|)\end{aligned} \tag{9}$$

where in (a) we used the fact that for any vector $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$ by Cauchy-Schwartz inequality. In (b) we used the submultiplicative property of matrix norms, i.e., $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$. We obtain (c) by noting that the norm of a diagonal matrix is the leading diagonal value and the fact that for a symmetric matrix the 2-norm is the largest eigenvalue in magnitude. The last inequality is obtained by noting that the assumption $\lambda_i = o(1)$ w.h.p. $\forall i > 1$ implies that $\exists N$ s.t. $\forall n > N$, $|1 - \alpha\lambda_i| > C$ for some constant C and the fact that $\|\mathbf{v}\|_2 = O(1/\sqrt{n})$.
Observing that d_{max}/d_{min} is bounded w.h.p. by Property 1 and $\max(|\lambda_2|, |\lambda_n|) = o(1)$ w.h.p. by Property 2 we obtain our result. □

Note that in the case of standard PageRank, $v_i = 1/n, 1 \leq i \leq n$, and hence $\|\mathbf{v}\|_2 = O(1/\sqrt{n})$, but Theorem 1 also admits more general preference vectors than the uniform one.

Corollary 1 *The statement of Theorem 1 also holds with respect to the weak convergence, i.e., for any function f on V such that* $\max_{x \in V} |f(x)| \leq 1$,

$$\sup \left\{ \sum_v f(v)\pi_v - \sum_v f(v)\overline{\pi}_v \right\} = o(1) \quad \textit{w.h.p.}$$

Proof: This follows from Theorem 1 and the fact that the left-hand side of the above equation is upper bounded by $2\, d_{\mathrm{TV}}(\boldsymbol{\pi}_n, \overline{\boldsymbol{\pi}}_n)$ [30]. □

4 Chung-Lu random graphs

In this section, we study the PageRank for the Chung-Lu model [14] of random graphs. These results naturally hold for w.h.p. graphs also. The spectral properties of Chung-Lu graphs have been studied extensively in a series of papers by Fan Chung et al [15, 16].

4.1 Chung-Lu Random Graph Model

Let us first provide a definition of the Chung-Lu random graph model.

Definition 1 **Chung-Lu Random Graph Model** *A Chung-Lu graph $\mathcal{G}(w)$ with an expected degree vector $\mathbf{w} = (w_1, w_2, \ldots w_n)$, where w_i are positive real numbers, is generated by drawing an edge between any two vertices v_i and v_j independently of all other pairs, with probability $p_{ij} = \frac{w_i w_j}{\sum_k w_k}$. To ensure that the probabilities p_{ij} are well-defined, we need* $\max_i w_i^2 \leq \sum_k w_k$.

In the following, let $w_{\max} = \max_i w_i$ and $w_{\min} = \min_i w_i$. Below we specify a corollary of Theorem 1 as applied to these graphs. But before that we need the following lemmas about Chung-Lu graphs mainly taken from [15, 16].

Lemma 1 *If the expected degrees $w_1, w_2, \ldots w_n$ satisfy $w_{\min} \gg \log(n)$, then in $\mathcal{G}(w)$ we have, w.h.p.,* $\max_i |\frac{d_i}{w_i} - 1| = o(1)$.

In the proof we use Bernstein Concentration Lemma [8]:

Lemma 2 *(Bernstein Concentration Lemma [8]) If $Y_n = X_1 + X_2 + \ldots X_n$, where X_i are independent random variables such that $|X_i| \leq b$ and if $B_n^2 = \mathbb{E}(Y_n - \mathbb{E}(Y_n))^2$ then*

$$\mathbb{P}\{|Y_n - \mathbb{E}(Y_n)| \geq \epsilon\} \leq 2 \exp \frac{-\epsilon^2}{2(B_n^2 + b\epsilon/3)},$$

for any $\epsilon > 0$.

Proof of Lemma 1: This result is shown in the sense of convergence in probability in the proof of [16, Theorem 2]; using Lemma 2 we show the result holds w.h.p. By a straight forward application of Lemma 2 to the degrees d_i of the Chung-Lu graph we obtain

$$\mathbb{P}\left(\max_{1\le i\le n}\left|\frac{d_i}{w_i}-1\right|\ge\beta\right)\le\frac{2}{n^{c/4-1}},\quad\text{if}\quad\beta\ge\sqrt{\frac{c\log(n)}{w_{\min}}}=o(1)$$

if $w_{\min}\gg\log(n)$. □We present below a perturbation result for the eigenvalues of Hermitian matrices, called Weyl's inequalities, which we will need for our proofs.

LEMMA 3 *[25, Theorem 4.3.1] Let $\mathbf{A},\mathbf{B}\in\mathbb{R}^{n\times n}$ be Hermitian and let the eigenvalues $\lambda_i(\mathbf{A})$, $\lambda_i(\mathbf{B})$ and $\lambda_i(\mathbf{A}+\mathbf{B})$ be arranged in decreasing order. For each $k=1,2,\ldots n$ we have*

$$|\lambda_k(\mathbf{A}+\mathbf{B})-\lambda_k(\mathbf{A})|\le\|\mathbf{B}\|_2,$$

where $\|\mathbf{B}\|_2$ is the induced 2-norm or the spectral norm of $\mathbf{B}$.

The following lemma is an application of Theorem 5 in [15].

LEMMA 4 *If $w_{\max}\le Kw_{\min}$, for some $K>0$ and $\overline{w}=\sum_k w_k/n\gg\log^6(n)$, then for $\mathcal{G}(w)$ we have almost surely (a.s.)*

$$\|\mathbf{C}\|_2=\frac{2}{\sqrt{\overline{w}}}(1+o(1)),$$

where $\mathbf{C}=\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}-\boldsymbol{\chi}^T\boldsymbol{\chi}$, $\mathbf{W}=diag(\mathbf{w})$, and $\boldsymbol{\emptyset}_i=\sqrt{w_i/\sum_k w_k}$ is a row vector.

Proof: It can be verified that when $w_{\max}\le Kw_{\min}$ and $\overline{w}\gg\log^6(n)$, the condition in [15, Theorem 5], namely, $w_{\min}\gg\sqrt{\overline{w}}\log^3(n)$, is satisfied and hence the result follows. □

LEMMA 5 *For $\mathcal{G}(w)$ with $w_{\max}\le Kw_{\min}$, and $\overline{w}\gg\log^6(n)$,*

$$\max(\lambda_2(\mathbf{P}),-\lambda_n(\mathbf{P}))=o(1)\quad w.h.p.,$$

where $\mathbf{P}$ is Markov matrix.

Proof: Recall that $\mathbf{Q}=\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ is the normalized adjacency matrix. We want to be able to bound the eigenvalues $\lambda_i, i\ge 2$ of $\mathbf{Q}$. We do this in two steps. Using Lemmas 1 and 3 we first show that if we replace the degree matrix $\mathbf{D}$ in the expression for $\mathbf{Q}$ by the expected degree matrix $\mathbf{W}=\mathbb{E}(\mathbf{D})$, the eigenvalues of the resulting matrix are close to those of $\mathbf{Q}$. Then, using Lemma 4 we show that the eigenvalues of $\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}$ roughly coincide with those of $\boldsymbol{\chi}^T\boldsymbol{\chi}$, which is a unit rank matrix and hence only has a single non-zero eigenvalue. Thus we arrive at the result of Lemma 5. Now we give the detailed proof.

The first step, $\|\mathbf{Q}-\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\|_2=o(1)$ w.h.p. follows from Lemma 1 and the same argument as in the last part of the proof of Theorem 2 in [16]. We present the steps in the derivation here for the sake of completeness.

Since the 2-norm of a diagonal matrix is the maximum diagonal in absolute value, we have

$$\|\mathbf{W}^{-1/2}\mathbf{D}^{1/2}-\mathbf{I}\|_2=\max_{\{i=1,2,\ldots\}}\left|\sqrt{\frac{d_i}{w_i}}-1\right|\le\max_{\{i=1,2,\ldots\}}\left|\frac{d_i}{w_i}-1\right|=o(1),\tag{10}$$

by Lemma 1. Also observe that

$$\|\mathbf{Q}\|_2=\max_{\{i=1,2,\ldots n\}}|\lambda_i(\mathbf{Q})|=\max_{\{i=1,2,\ldots n\}}|\lambda_i(\mathbf{P})|=1.\tag{11}$$

We now proceed to bound the norm of the difference $\|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\|$ as follows

$$\begin{aligned}
&\|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\|_2 \\
&= \|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\|_2 \\
&= \|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q}\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\|_2 \\
&= \|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q} + \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q}\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\|_2 \\
&\overset{(a)}{=} \|(\mathbf{I} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2})\mathbf{Q}\|_2 + \|\mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q}(\mathbf{I} - \mathbf{D}^{1/2}\mathbf{W}^{-1/2})\|_2 \\
&\overset{(b)}{\leq} \|(\mathbf{I} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2})\|_2\|\mathbf{Q}\|_2 + \|\mathbf{W}^{-1/2}\mathbf{D}^{1/2}\|_2\|\mathbf{Q}\|_2\|\mathbf{I} - \mathbf{D}^{1/2}\mathbf{W}^{-1/2}\|_2 \\
&\overset{(c)}{=} o(1) + (1 + o(1))o(1) = o(1) \quad w.h.p., \qquad (12)
\end{aligned}$$

where (a) follows from triangular inequality of norms, in (b) we used submultiplicativity of matrix norms, and (c) follows from (10), (11) and the fact that $\|\mathbf{W}^{-1/2}\mathbf{D}^{1/2}\|_2 \leq \|\mathbf{I}\|_2 + \|\mathbf{W}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}\|_2 = (1 + o(1))$.

By Lemma 3 we have for any i,

$$|\lambda_i(\mathbf{Q}) - \lambda_i(\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2})| \leq \|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\|_2 = o(1), \qquad (13)$$

by (12). Furthermore, using Lemma 3 and the fact that $\lambda_i(\boldsymbol{\chi}^T\boldsymbol{\chi}) = 0$ for $i > 1$, we have for $i \geq 2$,

$$\begin{aligned}
&|\lambda_i(\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2})| \\
&= |\lambda_i(\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}) - \lambda_i(\boldsymbol{\chi}^T\boldsymbol{\chi})| \leq \|\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2} - \boldsymbol{\chi}^T\boldsymbol{\chi}\|_2 \\
&= o(1), \qquad (14)
\end{aligned}$$

where the last inequality follows from Lemma 4.
Now recall that $\max(\lambda_2(\mathbf{P}), -\lambda_n(\mathbf{P})) = \max_{\{i\geq 2\}}|\lambda_i(\mathbf{Q})|$. We have for any i,

$$|\lambda_i(\mathbf{Q})| \leq |\lambda_i(\mathbf{Q}) - \lambda_i(\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2})| + ||\lambda_i(\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2})|, \qquad (15)$$

which implies from (13) and (14):

$$\max_{\{i\geq 2\}}|\lambda_i(\mathbf{Q})| = o(1).$$

□Armed with these lemmas we now present the following corollary of Theorem 1 in the case of Chung-Lu graphs.

Corollary 2 *Let $\|\mathbf{v}\|_2 = O(1/\sqrt{n})$, and $\alpha \in (0,1)$. Then PageRank $\boldsymbol{\pi}$ of the Chung-Lu graph $\mathcal{G}(w)$ can asymptotically be approximated in TV distance by $\overline{\boldsymbol{\pi}}$, defined in Theorem 1, if $\overline{w} \gg \log^6(n)$ and $w_{\max} \leq K w_{\min}$ for some K that does not depend on n.*

Proof: Using Lemma 1 and the condition that $w_{\max} \leq K w_{\min}$, one can show that $\exists K'$ s.t. $\frac{d_{max}}{d_{min}} \leq K'$ w.h.p. Then the result is a direct consequence of Lemma 5 and the inequality from (9). □

We further note that this result also holds for ER graphs $\mathcal{G}(n, p_n)$ with n nodes and edge probability p_n such that $np_n \gg \log^6(n)$, where we have $(w_1, w_2, \ldots w_n) = (np_n, np_n, \ldots np_n)$.

4.2 Element-wise Convergence

In Corollary 2 we proved the convergence of PageRank in TV distance for Chung-Lu random graphs. Note that since each component of PageRank could decay to zero as the graph size grows to infinity, this does not necessarily guarantee convergence in an element-wise sense. In this section, we provide a proof for our convergence conjecture to include the element-wise convergence of the PageRank vector. Here we deviate slightly from the spectral decomposition technique and eigenvalue bounds used hitherto, and instead rely on well-known concentration bounds to bound the error in convergence.

Let $\overline{\mathbf{\Pi}} = \text{diag}\{\overline{\pi}_1, \overline{\pi}_2, \ldots \overline{\pi}_n\}$ be a diagonal matrix whose diagonal elements are made of the components of the approximated PageRank vector and $\widetilde{\boldsymbol{\delta}} = \overline{\mathbf{\Pi}}^{-1}(\boldsymbol{\pi} - \overline{\boldsymbol{\pi}})$, i.e., $\widetilde{\delta}_i = (\pi_i - \overline{\pi}_i)/\overline{\pi}_i = \epsilon_i/\overline{\pi}_i$, where $\boldsymbol{\epsilon}$ is the unnormalized error defined in Section 3. Then using (8) we obtain

$$\widetilde{\delta}_i = \left((1-\alpha)v_i + \alpha\frac{d_i}{\text{vol(G)}}\right)^{-1}\left[\mathbf{D}^{1/2}\sum_{j\neq 1}\frac{\alpha\lambda_j}{1-\alpha\lambda_j}\mathbf{u}_j\mathbf{u}_j^T\mathbf{D}^{-1/2}\mathbf{v}\right]_i.$$

Therefore, using $\mathbf{v}'$ to denote $n\mathbf{D}^{-1/2}\mathbf{v}$ we can bound $\left\|\widetilde{\boldsymbol{\delta}}\right\|_\infty = \max_i|\widetilde{\delta}_i|$ as follows

$$\left\|\widetilde{\boldsymbol{\delta}}\right\|_\infty \leq \frac{1}{\min_i\left((1-\alpha)v_i + \alpha\frac{d_i}{\text{vol}(G)}\right)}\left\|\mathbf{D}^{1/2}\sum_{j\neq 1}\frac{\alpha\lambda_j}{1-\alpha\lambda_j}\mathbf{u}_j\mathbf{u}_j^T\mathbf{D}^{-1/2}\mathbf{v}\right\|_\infty \tag{16}$$

$$\leq \frac{\sum_i d_i/n}{\alpha d_{\min}}\sqrt{d_{max}}\left\|\sum_{j\neq 1}\frac{\alpha\lambda_j}{1-\alpha\lambda_j}\mathbf{u}_j\mathbf{u}_j^T\mathbf{v}'\right\|_\infty. \tag{17}$$

Here $d_{\min}$ denotes $\min_i d_i$. To obtain (17) we used the submultiplicativity property of matrix norms, the fact that $\|\mathbf{D}^{1/2}\|_\infty = \sqrt{\max_i d_i} = \sqrt{d_{\max}}$ and the fact that $v_i \geq 0, \forall i \in V$.

Define $\widetilde{\mathbf{Q}} = \mathbf{Q} - \mathbf{u}_1\mathbf{u}_1^T$, the restriction of the matrix $\mathbf{Q}$ to the orthogonal subspace of $\mathbf{u}_1$.

Lemma 6 *For a Chung-Lu random graph $\mathcal{G}(w)$ with expected degrees $w_1, \ldots w_n$, where $w_{\max} \leq Kw_{\min}$ and $w_{\min} \gg \log(n)$, we have w.h.p.,*

$$\left\|\widetilde{\mathbf{Q}}\mathbf{v}'\right\|_\infty = o(1/\sqrt{w_{\min}}),$$

when $v_i = O(1/n)$ $\forall i$.

This lemma can be proven by a few applications of Lemma 1 and Bernstein's concentration inequality. To keep the train of thought intact, please refer to Appendix A for a detailed proof of this lemma.

In the next lemma we prove an upper bound on the infinity norm of the matrix $\mathbf{S} = (\mathbf{I} - \alpha\mathbf{Q})^{-1}$.

Lemma 7 *Under the conditions of Lemma 6, $\|\mathbf{S}\|_\infty \leq C$ w.h.p., where $\mathbf{C}$ is a number independent of n that depends only on α and K.*

Proof: Note that $\mathbf{S} = (\mathbf{I} - \alpha\mathbf{Q})^{-1} = \mathbf{D}^{-1/2}(\mathbf{I} - \alpha\mathbf{P})^{-1}\mathbf{D}^{1/2}$. Therefore, $\|\mathbf{S}\|_\infty \leq \sqrt{\frac{d_{max}}{d_{min}}}\left\|(\mathbf{I} - \alpha\mathbf{P})^{-1}\right\|_\infty$ and the result follows since $\left\|(\mathbf{I} - \alpha\mathbf{P})^{-1}\right\|_\infty \leq \frac{1}{1-\alpha}$ [28] and using Lemma 1. □Now we are in a position to present our main result in this section.

Proposition 2 *Let $v_i = O(1/n)$ $\forall i$, and $\alpha < 1$. PageRank $\boldsymbol{\pi}$ converges element-wise to $\overline{\boldsymbol{\pi}} = (1-\alpha)\mathbf{v} + \alpha\mathbf{d}/\mathrm{vol}(G)$, in the sense that $\max_i\ (\pi_i - \overline{\pi}_i)/\overline{\pi}_i = o(1)$ w.h.p., on the Chung-Lu graph $\mathcal{G}(w)$ with expected degrees $\{w_1, w_2, \ldots w_n\}$ such that $w_{\min} > \log^c(n)$ for some $c > 1$ and $w_{\max} \leq K w_{\min}$, for some K, a constant independent of n.*

Proof: Define $\mathbf{Z} = \sum_{i\neq 1} \frac{\alpha\lambda_i}{1-\alpha\lambda_i}\mathbf{u}_i\mathbf{u}_i^T$. We then have:

$$\begin{aligned}
\mathbf{Z} &= \sum_{i=1}^{n} \frac{\alpha\lambda_i}{1-\alpha\lambda_i}\mathbf{u}_i\mathbf{u}_i^T - \frac{\alpha}{1-\alpha}\mathbf{u}_1\mathbf{u}_1^T \\
&= (\mathbf{I}-\alpha\mathbf{Q})^{-1}\alpha\mathbf{Q} - \frac{\alpha}{1-\alpha}\mathbf{u}_1\mathbf{u}_1^T \\
&= \mathbf{S}\left[\alpha\mathbf{Q} - \frac{\alpha}{1-\alpha}(\mathbf{I}-\alpha\mathbf{Q})\mathbf{u}_1\mathbf{u}_1^T\right] \\
&= \alpha\mathbf{S}\widetilde{\mathbf{Q}} \qquad (18)
\end{aligned}$$

Now from (17) we have

$$\begin{aligned}
\left\|\widetilde{\boldsymbol{\delta}}\right\|_\infty &\leq C\frac{\sum_i d_i/n}{d_{\min}}\sqrt{d_{max}}\|\mathbf{S}\widetilde{\mathbf{Q}}\mathbf{v}'\|_\infty \\
&\overset{(a)}{\leq} C\frac{\sum_i d_i/n}{d_{\min}}\sqrt{d_{max}}o(1/\sqrt{w_{\min}}) \\
&\leq C\frac{d_{\max}}{d_{\min}}\sqrt{w_{\max}(1+o(1))}\frac{1}{\sqrt{w_{\min}}}o(1) \\
&= C\frac{w_{\max}}{w_{\min}}\sqrt{\frac{w_{\max}}{w_{\min}}}(1+o(1))o(1) \\
&= C\left(\frac{w_{\max}}{w_{\min}}\right)^{\frac{3}{2}}o(1) \\
&\leq Co(1) \quad \text{w.h.p.},
\end{aligned}$$

where in (a) we used (18) and Lemmas 6 and 7. The rest of the inequalities are obtained by repeatedly using the fact that $d_{\max} = w_{\max}(1+o(1))$ and $d_{min} = w_{\min}(1+o(1))$, from Lemma 1. The last step follows from the assumption that $w_{\max} \leq K w_{\min}$ for some constant K. □

Corollary 1 (ER Graphs) *For an ER graph $\mathcal{G}(n, p_n)$ such that $np_n \gg \log(n)$, we have that asymptotically the personalized PageRank $\boldsymbol{\pi}$ converges pointwise to $\overline{\boldsymbol{\pi}}$ for $\mathbf{v}$ such that $v_i = O(1/n)$.*

5 Asymptotic PageRank for the Stochastic Block Model

In this section, we extend the analysis of PageRank to Stochastic Block Models (SBM) with constraints on average degrees. The SBM is a random graph model that reflects the community structure prevalent in many online social networks. It was first introduced in [24] and has been analyzed subsequently in several works, specifically in the community detection literature, including [17], [27], [34] and several extensions thereof as in [23] and [39], and the references therein.

For the sake of simplicity we focus on an SBM graph with two communities, but the idea of the proof extends easily to generalizations of this simple model.

Definition 1 *[Stochastic Block Model (SBM) with two communities]: An SBM graph $\mathcal{G}(m, n-m, p, q)$ with two communities is an undirected graph on a set of disjoint vertices C_1, C_2 such that $C_1 \cup C_2 = V$, and let $|C_1| = m$ and $|C_2| = n-m$. Furthermore, if two vertices $i, j \in C_k, k = 1, 2$, then $\mathbb{P}((i,j) \in E) = p$, if $i \in C_1$ and $j \in C_2$, then $\mathbb{P}((i,j) \in E) = q$. The probabilities p, q may scale with n and we assume that $m > n/2$ and $p > q$; this last assumption is necessary for modeling the community structure of a network.*

Remark: For the sake of simplicity, we assume that the edge probabilities within both communities are equal to p, but this is a minor assumption and can be generalised so that community 1 has a different edge probability to community 2.

For an SBM graph we use $w_{\max}$ and $w_{\min}$ to denote the maximum and the minimum expected degrees of the nodes respectively. From Definition 1, by our assumption on m, p and q, we have $w_{\max} = mp + (n-m)q$ and $w_{\min} = (n-m)p + mq$. Note that our results only depend on these two parameters. We present our main result on SBM graphs in the following theorem.

Proposition 3 *For a Stochastic Block Model with $w_{\min} = !(\log^3(n))$ and $\frac{w_{\max}}{w_{\min}} \leq C$, PageRank with preference vector $\mathbf{v}$ such that $\|\mathbf{v}\|_2 = O(\frac{1}{\sqrt{n}})$ satisfies*

$$\|\boldsymbol{\pi} - \overline{\boldsymbol{\pi}}_{\mathrm{SBM}}\|_{TV} = o(1)$$

w.h.p., where

$$\overline{\boldsymbol{\pi}}_{\mathrm{SBM}} = (1-\alpha)\left(\mathbf{I} - \alpha\overline{\mathbf{P}}\right)^{-1}\mathbf{v}. \tag{19}$$

Here $\overline{\mathbf{P}}$ represents the "average" Markov matrix given as $\overline{\mathbf{P}} = \overline{\mathbf{A}}\mathbf{W}^{-1}$ where $\mathbf{W} = \mathbb{E}(\mathbf{D})$ and $\overline{\mathbf{A}} = \mathbb{E}(\mathbf{A})$.

Discussion: Let us look at the permissible values of m, p, q under the assumptions in the above theorem. Recall that we have $w_{\min} = (n-m)p + mq > nq$. Therefore the condition on the growth of minimum expected degree is met, for example, if $q = \omega(\log^3(n)/n)$. On the other hand we have

$$\frac{w_{\max}}{w_{\min}} = \frac{mp + (n-m)q}{(n-m)p + mq} = \frac{\frac{m}{n-m}\frac{p}{q} + 1}{\frac{m}{n-m} + \frac{p}{q}} \quad ,$$

which remains bounded if either $m/(n-m)$ or p/q tends to infinity, but not both.

The following corollary of Theorem 3 gives an interesting expression for PageRank for an SBM graph with two equal-sized communities.

Corollary 2 *For an SBM graph as in Definition 1, with $m = n/2$, (n assumed to be even) such that $p + q \gg \log^3(n)/n$ the PageRank vector $\boldsymbol{\pi}$ with preference vector $\mathbf{v}$ such that $\|\mathbf{v}\|_2 = O(\frac{1}{\sqrt{n}})$ satisfies*

$$\|\boldsymbol{\pi} - \overline{\boldsymbol{\pi}}_{\mathrm{SBM}}\|_{TV} \to 0$$

w.h.p as $n \to \infty$ where

$$\overline{\boldsymbol{\pi}}_{\mathrm{SBM}} = \alpha\frac{1}{n}\mathbf{1} + (1-\alpha)\left(\mathbf{v} + \frac{\alpha\beta}{1-\alpha\beta}(\mathbf{v}^T\mathbf{u})\mathbf{u}\right), \tag{20}$$

where $\beta := \frac{p-q}{p+q}$, and $\mathbf{u} \in \mathbb{R}^n$ is a unit vector such that $u_i = \frac{1}{\sqrt{n}}$, for $i \in C_1$ and $u_i = -\frac{1}{\sqrt{n}}$ for $i \in C_2$.

Proof: With equal-sized communities, i.e., $m = n/2$, we have $w_{\max} = w_{\min} = \frac{n}{2}(p+q)$. Therefore the conditions of Theorem 3 are satisfied if $p + q \gg \log^3(n)/n$. Observe that the expected adjacency matrix can be written as $\overline{\mathbf{A}} = \frac{p+q}{2}\mathbf{1}\mathbf{1}^T + \frac{n}{2}(p-q)\mathbf{u}\mathbf{u}^T$. Furthermore, $\mathbf{W} = \frac{n}{2}(p+q)\mathbf{I}$. Therefore $\overline{\mathbf{P}} = \overline{\mathbf{A}}\mathbf{W}^{-1} = \frac{1}{n}\mathbf{1}\mathbf{1}^T + \frac{p-q}{p+q}\mathbf{u}\mathbf{u}^T$. From (19), the asymptotic PageRank $\overline{\boldsymbol{\pi}}_{\text{sbm}}$ is therefore given as

$$\overline{\boldsymbol{\pi}}_{\text{sbm}} = \alpha\overline{\mathbf{P}}\overline{\boldsymbol{\pi}}_{\text{sbm}} + (1-\alpha)\mathbf{v}.$$

Consequently, $\overline{\boldsymbol{\pi}}_{\text{sbm}} = \frac{\alpha}{n}\mathbf{1} + \alpha\beta\mathbf{u}\mathbf{u}^T\overline{\boldsymbol{\pi}}_{\text{sbm}} + (1-\alpha)\mathbf{v}$, or $\left[\mathbf{I} - \alpha\beta\mathbf{u}\mathbf{u}^T\right]\overline{\boldsymbol{\pi}}_{\text{sbm}} = \frac{\alpha}{n}\mathbf{1} + (1-\alpha)\mathbf{v}$. By Woodbury Matrix Inversion Lemma in [25], $\left[\mathbf{I} - \alpha\beta\mathbf{u}\mathbf{u}^T\right]^{-1} = \mathbf{I} + \frac{\alpha\beta}{1-\alpha\beta}\mathbf{u}\mathbf{u}^T$. Hence we obtain $\overline{\boldsymbol{\pi}}_{\text{sbm}} = \frac{\alpha}{n}\mathbf{1} + (1-\alpha)\left(\mathbf{v} + \frac{\alpha\beta}{1-\alpha\beta}(\mathbf{u}^T\mathbf{v})\mathbf{u}\right)$, using the fact that $\mathbf{u}$ and $\mathbf{1}$ are orthogonal vectors. □The above corollary asserts that on an SBM matrix the PageRank is well approximated in the asymptotic regime of large graph size by the convex combination of the uniform probability vector $\frac{1}{n}\mathbf{1}$, which is the asymptotic stationary distribution of a simple random walk on the SBM graph, and a linear combination of the preference vector $\mathbf{v}$ and the projection of the preference vector onto the community partitioning vector $\mathbf{u}$. Thus in this simple scenario of SBM graphs with equally sized communities, we observe that PageRank incorporates information about the community structure, in the form of a term correlated with the partition vector $\mathbf{u}$, as opposed to the usual random walk, which misses this information. It can also be inferred from (20) that if the correlation between the preference vector $\mathbf{v}$ and $\mathbf{u}$ is large, e.g., when the seed set of PageRank is chosen to be in one of the communities, the resulting PageRank will display a clear delineation of the communities. This provides a mathematical rationale for why PageRank works for semi-supervised graph partitioning [5], at least in the asymptotic regime.

To prove Theorem 3 we need the following Lemmas, whose proofs are given in Appendix B.

Lemma 8 *For an SBM graph $\mathcal{G}(m, n-m, p, q)$, when $w_{\min} = !(\log^3(n))$ it can be shown that for some C,*

$$\max_{1\le i\le n}\left|\frac{D_i}{\mathbb{E}(D_i)} - 1\right| \le C\sqrt{\frac{\log(n)}{w_{\min}}} \ w.h.p.$$

The proof of this lemma follows from applying Bernstein's concentration lemma to the degrees of SBM graph. The proof is given in Appendix B.1.

For ease of notation, let $\overline{\mathbf{Q}} = \mathbf{W}^{-1/2}\mathbb{E}(\mathbf{A})\mathbf{W}^{-1/2}$, where $\mathbf{W} = \mathbb{E}(\mathbf{D})$. As before $\mathbf{Q} = \mathbf{D}^{1/2}\mathbf{A}\mathbf{D}^{1/2}$. We need the following concentration result on $\mathbf{Q}$.

Lemma 9 *For an SBM graph for which $w_{\min} = !(\log^3(n))$, and $\frac{w_{\max}}{w_{\min}} \le C$ for some C, it can be shown that*

$$\|\mathbf{Q} - \overline{\mathbf{Q}}\|_2 = C\frac{\sqrt{\log(n)w_{\max}}}{w_{\min}} = o(1)$$

w.h.p.

We prove this lemma in Appendix B.2.

Proof of Theorem 3: We write the error between $\boldsymbol{\pi}$ and $\overline{\boldsymbol{\pi}}$ as follows

$$\begin{aligned}\boldsymbol{\delta} &= \boldsymbol{\pi} - \overline{\boldsymbol{\pi}} \\ &= (1-\alpha)\left[\mathbf{D}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{D}^{-1/2} - \mathbf{W}^{1/2}(\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}\mathbf{W}^{-1/2}\right]\mathbf{v} \\ &= (1-\alpha)\left[\mathbf{W}^{1/2}\left((\mathbf{I}-\alpha\mathbf{Q})^{-1} - (\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}\right)\mathbf{W}^{-1/2}\right]\mathbf{v}+ \\ &(1-\alpha)\left[\mathbf{D}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{D}^{-1/2} - \mathbf{W}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{W}^{-1/2}\right]\mathbf{v}, \end{aligned} \tag{21}$$

where in the last equality we added and subtracted $\mathbf{W}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{W}^{-1/2}$ and reordered terms. Now we analyse the two terms in square brackets in the last equality in (21), which we denote T_1 and T_2, respectively. Notice that we have $\|\boldsymbol{\delta}\|_1 \leq \|T_1\|_1 + \|T_2\|_1$. Next we show that as $n \to \infty$, $\|T_1\|_1$ and $\|T_2\|_1$ are $o(1)$ separately and consequently we obtain the result of the theorem.

Let us first consider T_1. We have

$$\begin{aligned}T_1 &= (1-\alpha)\left[\mathbf{W}^{1/2}\left((\mathbf{I}-\alpha\mathbf{Q})^{-1} - (\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}\right)\mathbf{W}^{-1/2}\right]\mathbf{v} \\ &= (1-\alpha)\mathbf{W}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\left(\overline{\mathbf{Q}}-\mathbf{Q}\right)(\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}\mathbf{W}^{-1/2}\mathbf{v},\end{aligned}$$

which we obtained by factoring out $(\mathbf{I}-\alpha\mathbf{Q})^{-1}$ and $(\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}$ on the left and right sides of the square brackets. Next we focus on the 2-norm of T_1.

$$\begin{aligned}\|T_1\|_2 &\overset{(a)}{\leq} (1-\alpha)\sqrt{w_{\max}}\|(\mathbf{I}-\alpha\mathbf{Q})^{-1}\|_2\|\overline{\mathbf{Q}}-\mathbf{Q}\|_2\|(\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}\|_2\frac{1}{\sqrt{w_{\min}}}\|\mathbf{v}\|_2 \\ &\overset{(b)}{\leq} \frac{1}{1-\alpha}\sqrt{\frac{w_{\max}}{w_{\min}}}\|\mathbf{Q}-\overline{\mathbf{Q}}\|_2\|\mathbf{v}\|_2 \\ &\overset{(c)}{\leq} C\frac{\sqrt{\log(n)w_{\max}}}{w_{\min}\sqrt{n}} \\ &= C\sqrt{\frac{\log(n)}{nw_{\max}}}\frac{w_{\max}}{w_{\min}}.\end{aligned}$$

This proves $\|T_1\|_1 \leq \sqrt{n}\|T_1\|_2$[1] $\leq C\sqrt{\frac{\log(n)}{w_{\max}}}\frac{w_{\max}}{w_{\min}} = o(1)$, from the assumptions of the theorem. Here in (a) we used the submultiplicative property of matrix norms and the fact that 2-norm of diagonal matrices is the maximum diagonal element in magnitude. The inequality (b) follows because $\|(\mathbf{I}-\alpha\mathbf{Q})^{-1}\|_2 \leq \frac{1}{1-\alpha}$ and $\|(\mathbf{I}-\alpha\overline{\mathbf{Q}})^{-1}\|_2 \leq \frac{1}{1-\alpha}$ and step (c) follows from Lemma 9 and the assumption that $\|\mathbf{v}\|_2 = O(1/\sqrt{n})$.

Next we analyse the second term T_2. For ease of notation we denote $\widetilde{\mathbf{R}} = \mathbf{W}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{W}^{-1/2}$. Then by simple algebraic manipulations

$$\begin{aligned}T_2 &= (1-\alpha)\left[\mathbf{D}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{D}^{-1/2} - \mathbf{W}^{1/2}(\mathbf{I}-\alpha\mathbf{Q})^{-1}\mathbf{W}^{-1/2}\right]\mathbf{v} \\ &= (1-\alpha)\left(\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\widetilde{\mathbf{R}}\mathbf{W}^{1/2}\mathbf{D}^{-1/2} - \widetilde{\mathbf{R}}\right)\mathbf{v} \\ &= (1-\alpha)\left(\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\widetilde{\mathbf{R}}\left(\mathbf{W}^{1/2}\mathbf{D}^{-1/2} - \mathbf{I}\right) + \left(\mathbf{D}^{1/2}\mathbf{W}^{-1/2} - \mathbf{I}\right)\widetilde{\mathbf{R}}\right)\mathbf{v},\end{aligned}$$

[1] By Cauchy Schwartz inequality on norms.

where the last step is obtained by adding and subtracting $\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\widetilde{\mathbf{R}}$.
Now we have $\|\mathbf{D}^{1/2}\mathbf{W}^{-1/2}-\mathbf{I}\|_2 = \max_i \left|\sqrt{\frac{d_i}{w_i}}-1\right| \leq \max_i \left|\frac{d_i}{w_i}-1\right| \leq C\sqrt{\frac{\log(n)}{w_{\min}}}$ w.h.p. by Lemma 8 and similarly $\|\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\|_2 \leq \|\mathbf{D}^{1/2}\mathbf{W}^{-1/2}-\mathbf{I}\|_2 + \|\mathbf{I}\|_2 \leq C\sqrt{\frac{\log(n)}{w_{\min}}}+1$. In addition $\|\mathbf{W}^{1/2}\mathbf{D}^{-1/2}-\mathbf{I}\|_2 = \max_i \left|\sqrt{\frac{w_i}{d_i}}-1\right| \leq \max_i \left|\frac{w_i}{d_i}-1\right|$. It can be shown that since $\max_i \left|\frac{d_i}{w_i}-1\right| \leq C\sqrt{\frac{\log(n)}{w_{\min}}}$ w.h.p. (by Lemma 8), then $\max_i \left|\frac{w_i}{d_i}-1\right| \leq C\sqrt{\frac{\log(n)}{w_{\min}}}$ w.h.p.[2] Therefore $\|\mathbf{W}^{1/2}\mathbf{D}^{-1/2}\|_2 \leq \|\mathbf{W}^{1/2}\mathbf{D}^{-1/2}-\mathbf{I}\|_2 + \|\mathbf{I}\|_2 \leq C\sqrt{\frac{\log(n)}{w_{\min}}}+1$ w.h.p. Using the above facts and denoting $\delta = C\sqrt{\frac{\log(n)}{w_{\min}}}$ we obtain

$$\begin{aligned}
\|T_2\|_2 &\leq \left(\|\mathbf{D}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}\|_2\|\widetilde{\mathbf{R}}\|_2\|\mathbf{W}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}-\mathbf{I}\|_2 + \|\mathbf{D}^{\frac{1}{2}}\mathbf{W}^{-\frac{1}{2}}-\mathbf{I}\|_2\|\widetilde{\mathbf{R}}\|_2\right)\|\mathbf{v}\|_2 \\
&\leq C(\delta(\delta+1)\frac{1}{1-\alpha}+\delta)\frac{1}{1-\alpha}\sqrt{\frac{w_{\max}}{nw_{\min}}} \qquad (22)\\
&\leq C\delta\sqrt{\frac{w_{\max}}{nw_{\min}}}\text{w.h.p.} \qquad (23)
\end{aligned}$$

Hence we have $\|T_2\|_1 \leq \sqrt{n}\|T_2\|_2 \leq C\delta\sqrt{\frac{w_{\max}}{w_{\min}}}$ w.h.p., which from our assumptions is $o(1)$. Here in (22) we used the fact that

$$\|\widetilde{\mathbf{R}}\|_2 = \|\mathbf{W}^{1/2}\left(\mathbf{I}-\alpha\mathbf{Q}\right)^{-1}\mathbf{W}^{-1/2}\|_2 \leq \sqrt{\frac{w_{\max}}{w_{\min}}}\|\mathbf{I}-\alpha\mathbf{Q}\|_2 \leq \frac{1}{1-\alpha}\sqrt{\frac{w_{\max}}{w_{\min}}} \leq C,$$

and that $\|\mathbf{v}\|_2 \leq C/\sqrt{n}$, for some C. □ *Remark:* This method of proof can be extended to similar models like the Stochastic Block Model with multiple communities and their generalizations, e.g., Random Dot Product Graphs [2].

6 Experimental Results

In this section, we provide experimental evidence to further illustrate the analytic results obtained in the previous sections. In particular, we simulated ER graphs with $p_n = C\frac{\log^7(n)}{n}$ and Chung-Lu graphs with the degree vector w sampled from a geometric distribution so that the average degree $\overline{w} = cn^{1/3}$, clipped such that $w_{\max} = 7w_{\min}$, for various values of graph size, and plotted the maximum of normalized error $\widetilde{\delta}$ and TV distance error $\|\delta\|_1$, respectively, in Figures 1 and 2. As expected, both these errors decay as functions of n, which illustrates that the PageRank vector does converge to the asymptotic value.

In the spirit of further exploration, we have also conducted simulations on power-law graphs with exponent $\beta = 4$ using the Chung-Lu graph model with $w_i = ci^{-1/(\beta-1)}$, for $i_0 \leq i \leq n+i_0$ with

$$c = \frac{\beta-2}{\beta-1}dn^{1/(\beta-1)},$$

$$i_0 = n\left[\frac{d(\beta-1}{m(\beta-2)}\right]$$

[2]This follows since we can write $\frac{d_i}{w_i} = 1+\eta_i$, with $\max_i|\eta_i| = O\left(\sqrt{\frac{\log(n)}{w_{\min}}}\right) = o(1)$ w.h.p., then $\frac{w_i}{d_i} = \frac{1}{1+\eta_i} = 1-\eta_i+O(\eta_i^2)$, hence $\max_i|\frac{w_i}{d_i}-1| = O(\max_i|\eta_i|) = O\left(\sqrt{\frac{\log(n)}{w_{\min}}}\right) = o(1)$ w.h.p.

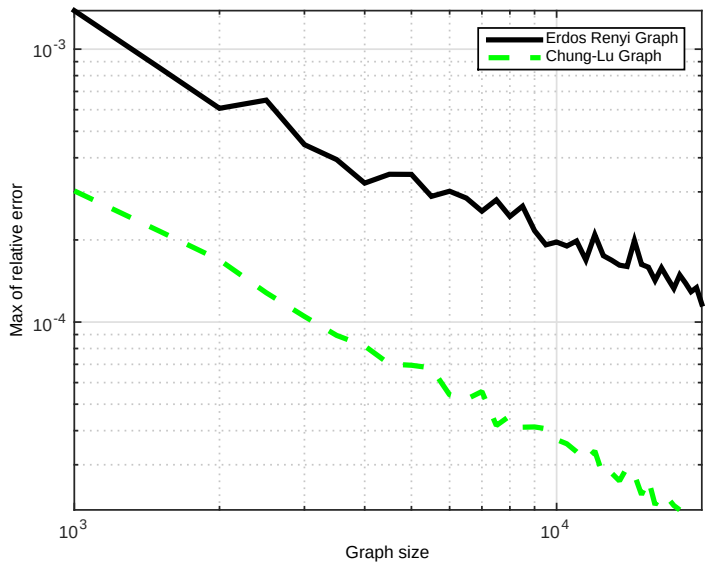

Figure 1: Log-log plot of maximum normalized error for ER and Chung-Lu graphs

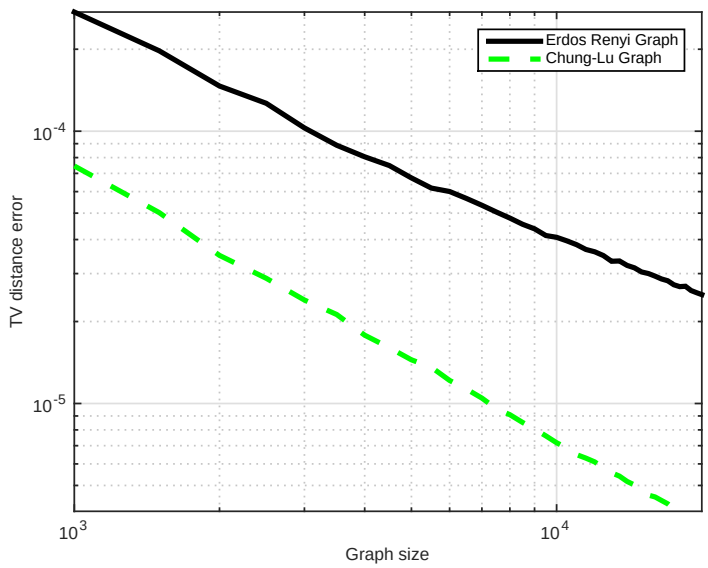

Figure 2: Log-log plot of TV distance error for ER and Chung-Lu graphs

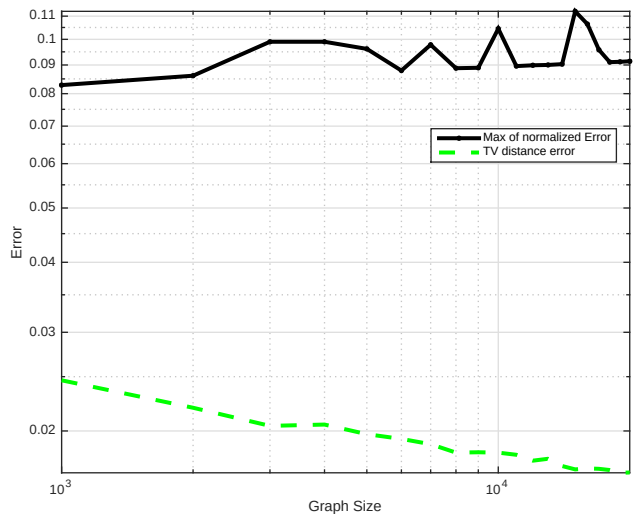

Figure 3: Log-log plot of TV distance and maximum error for power-law graphs

Please refer to [15] for details. We set max degree $m = n^{1/3}$ and average degree $d = n^{1/6}$. In Figure 3 we observe that for this graph the max-norm of the relative error does not converge to zero. On the other hand the TV-norm seems to converge to zero with graph size, albeit very slowly. Note that these graphs satisfy Property **??** [15], but they do not satisfy Property **??**. Therefore at this point, it is not possible to conclude whether the assumption of bounded variation of degrees is necessary for the convergence to hold. It might be interesting to investigate in detail the asymptotic behavior of PageRank in undirected power-law graphs.

Furthermore, we also see that in the case $\mathbf{v} = \mathbf{e}_i$, the standard unit vector, for some i we do not have the conjectured convergence, as can be seen on Figure 4 in the case of ER graphs. It can also be seen from our analysis that if $v_k = 1$ for some k, the quantity $\left\|\widetilde{Q}D^{-1/2}v\right\|_\infty$, becomes:

$$\max_i \left| \sum_j \left(\frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_l d_l} \right) v_j / \sqrt{d_j} \right| = \max_i \frac{1}{\sqrt{d_i} d_k} \left| A_{ik} - \frac{d_i d_k}{\sum_l d_l} \right|,$$

which is $O\left(\frac{1}{\sqrt{w_{\min}} w_k}\right)$ and does not fall sufficiently fast. We simulated an SBM matrix with two communities of equal size, with $p = 0.1$ and $q = 0.01$. In Figure 5 we plot the maximum normalized error and the TV-distance error against graph size on a log-log plot. As expected both errors go to zero for large graph sizes.

7 Conclusions

In this work, we have shown that when the size of a graph tends to infinity, the PageRank vector lends itself to be approximated by a mixture of the preference vector and the degree distribution, for a class of undirected random graphs including the Chung-Lu graph. We expect that these findings will shed more light on the behaviour of PageRank on undirected graphs, and possibly help to optimize the PageRank operation, or suggest further modifications to better capture local

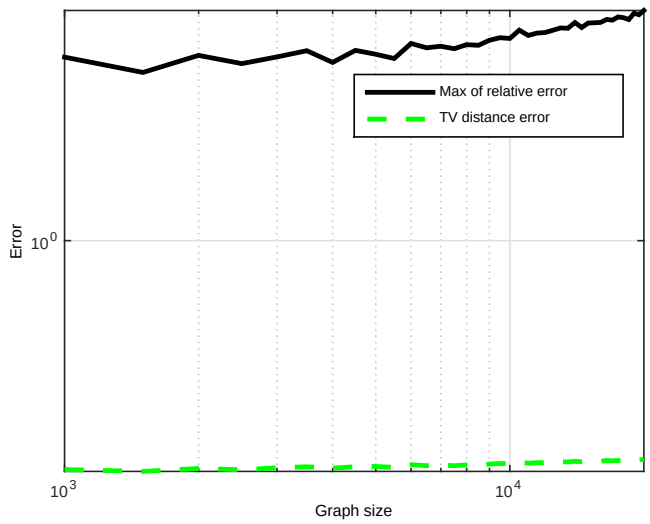

Figure 4: Log-log plot of TV distance and maximum relative error for ER-graph when $v = e_1$

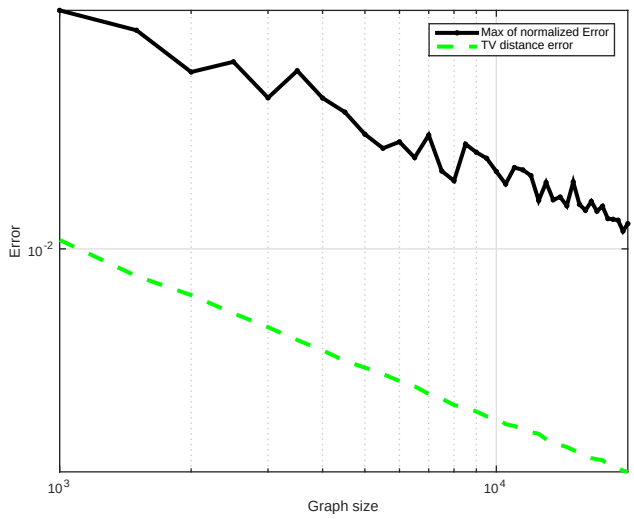

Figure 5: Log-log plot of maximum normalized error and TV-distance error for an SBM graph

graph properties. We also obtain an asymptotic expression for the PageRank on SBM graphs. It is seen that this asymptotic expression contains information about community partitioning in the simple case of SBM with equal-sized communities. It would be interesting to study the implications of our results for community detection algorithms.

Acknowledgements

We would like to thank Nelly Litvak for stimulating discussions on the topic of the paper. The work of K. Avrachenkov and A. Kadavankandy was partly funded by the French Government (National Research Agency, ANR) through the "Investments for the Future" Program reference #ANR-11-LABX-0031-01 and the work of L. Ostroumova Prokhorenkova and A. Raigorodskii was supported by Russian Science Foundation (# 16-11-10014).

References

[1] Andersen, R., Chung, F., and Lang, K. Local graph partitioning using pagerank vectors. In Proceedings of IEEE FOCS, 2006.

[2] Athreya, A., Priebe, C. E., Tang, M., Lyzinski, V., Marchette, D. J., and Sussman, D. L. (2013). A limit theorem for scaled eigenvectors of random dot product graphs. Sankhya A, 1-18.

[3] Avrachenkov, K., Cottatellucci, L., and Kadavankandy, A. Spectral properties of random matrices for stochastic block model. In Proceedings of WiOpt Workshop PhysComNet, 2015.

[4] Avrachenkov, K., Dobrynin, V., Nemirovsky, D., Pham, S.K., and Smirnova, E. Pagerank based clustering of hypertext document collections. In *Proceedings of ACM SIGIR*, pp. 873–874, 2008.

[5] Avrachenkov, K., Gonçalves, P., Mishenin, A., and Sokol, M. Generalized optimization framework for graph-based semi-supervised learning. In *Proceedings of SIAM Conference on Data Mining*, vol. 9, 2012.

[6] Avrachenkov, K., and Lebedev, D. PageRank of scale-free growing networks. Internet Mathematics, 3(2), 207-231, 2006.

[7] Bhatia, R. Matrix analysis. Springer Science and Business Media, vol. 169, 2013.

[8] Billingsley, P. Probability and measure. John Wiley & Sons, 2008.

[9] Boudin, F. A comparison of centrality measures for graph-based keyphrase extraction. In Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), 2013.

[10] Chen, N., Litvak, N., and Olvera-Cravioto, M. PageRank in scale-free random graphs. In Proceedings of WAW, 2014.

[11] Chen, N., Litvak, N., and Olvera-Cravioto, M. Generalized PageRank on directed configuration networks. Random Structures & Algorithms 2016.

[12] Chung, F. A local graph partitioning algorithm using heat kernel PageRank. In Proceedings of WAW 2009.

[13] Chung, F. Spectral graph theory. American Mathematical Soc, v.92, 1997.

[14] Chung, F., and Lu, L. The average distances in random graphs with given expected degrees. PNAS, v.99(25), 15879-15882, 2002.

[15] Chung, F., Lu, L., and Vu, V. Spectra of random graphs with given expected degrees. PNAS, v.100(11), 6313-6318, 2003.

[16] Chung, F., and Radcliffe, M. On the spectra of general random graphs. The electronic journal of combinatorics, 18(1), 2011.

[17] Condon A. and Karp R. Algorithms for graph partitioning on the planted partition model, Random Structures and Algorithms, vol. 18, 2001, pp.116–140.

[18] Ding, C., He, X., Husbands, P., Zha, H., and Simon, H.D. PageRank, HITS and a unified framework for link analysis. In Proceedings of ACM SIGIR, 2002.

[19] Erdős, P., and Rényi, A. On random graphs. Publicationes Mathematicae Debrecen, v.6, 290-297, 1959.

[20] Fortunato, S., Boguna, M., Flammini, A., and Menczer, F. Approximating PageRank from in-degree. In Proceedings of WAW, 2008.

[21] Gkorou, D., Vinko, T., Pouwelse, J., and Epema, D. Leveraging node properties in random walks for robust reputations in decentralized networks. In Proceedings of IEEE Peer-to-Peer Computing (P2P), 2013.

[22] Haveliwala, T.H. Topic-sensitive pagerank. In Proceedings of WWW, 517-526, 2002.

[23] Heimlicher, S., Lelarge M., and Massoulié, L. Community detection in the labelled stochastic block model. arXiv preprint arXiv:1209.2910 (2012).

[24] Holland, Paul W., Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. Social networks 5.2 (1983): 109-137.

[25] Horn, R.A., and Johnson, C.R. Matrix analysis. Cambridge university press, 2012.

[26] Kamvar, S.D., Schlosser, M.T., and H. Garcia-Molina, H. The eigentrust algorithm for reputation management in p2p networks. In Proceedings of WWW, 2003.

[27] Karrer, B. and Newman M.E.J. Stochastic blockmodels and community structure in networks. Physical Review E 83.1 (2011): 016107.

[28] Langville, A.N., and Meyer, C.D. Deeper inside pagerank. Internet Mathematics, 1(3), 335-380, 2004.

[29] Litvak, N., Scheinhardt, W. R., and Volkovich, Y. In-degree and PageRank: Why do they follow similar power laws? *Internet mathematics*, 4(2-3), 175-198, 2007.

[30] Levin, D.A., Peres, Y., and Wilmer, E.L. Markov chains and mixing times. American Mathematical Soc., 2009.

[31] Moler, C., Moler, K. Numerical Computing with MATLAB. SIAM, 2003.

[32] Page, L., Brin, S., Motwani, R., and Winograd, T. PageRank: Bringing order to the web. Stanford Digital Libraries Working Paper, v.72, 1997.

[33] Perra, N. and Fortunato, S. Spectral centrality measures in complex networks. Phys. Rev. E 78, 036107, 2008.

[34] Rohe, K., Chatterjee, S. and Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. The Annals of Statistics (2011): 1878-1915.

[35] Tropp, Joel A. User-friendly tail bounds for sums of random matrices. Foundations of computational mathematics 12.4 (2012): 389-434.

[36] Vadhan, S. Pseudorandomness. Foundations and Trends in Theoretical Computer Science, v.7(1-3), pp.1-336, Now publishers, 2012.

[37] Volkovich, Y., and Litvak, N. Asymptotic analysis for personalized web search. *Adv. Appl. Prob.*, 42(2), 577-604, 2010.

[38] Yeh, E., Ramage, D., Manning, C. D., Agirre, E., and Soroa, A. WikiWalk: random walks on Wikipedia for semantic relatedness. In Proceedings of the Workshop on Graph-based Methods for Natural Language Processing, 2009.

[39] Zhao, Y., Levina E., and Zhu J. Consistency of community detection in networks under degree-corrected stochastic block models. The Annals of Statistics (2012): 2266-2292.

A Proof of Lemma 6

From Lemma 1, we have for Chung-Lu graphs that: $d_i = w_i(1+\epsilon_i)$, where $\eta \equiv \max_i \epsilon_i = o(1)$ with high probability. In the proof we assume explicitly that $v_i = 1/n$, but the results hold in the slightly more general case where $v_i = O(1/n)$ uniformly $\forall i$, i.e., $\exists K$ such that $\max_i nv_i \leq K$. It can be verified easily that all the bounds that follow hold in this more general setting. The event $\{\eta = o(1)\}$, holds w.h.p. asymptotically from Lemma 1. In this case, we have

$$\sum_j \left(\frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_i d_i} \right) \frac{v_j}{\sqrt{d_j}} = \sum_j \left(\frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_k d_k} \right) \frac{v_j}{\sqrt{w_j}} (1+\varepsilon_j)$$

where ε_j is the error of convergence, and we have $\max_j \varepsilon_j = O(\eta)$. Therefore,

$$\begin{aligned} \left\| \widetilde{\mathbf{Q}} \mathbf{v}' \right\|_\infty &\leq \left\| \widetilde{\mathbf{Q}} \mathbf{q} \right\|_\infty + \max_i \varepsilon_i \left\| \widetilde{\mathbf{Q}} \mathbf{q} \right\|_\infty \\ &\leq \left\| \widetilde{\mathbf{Q}} \mathbf{q} \right\|_\infty (1+o(1)) \quad \text{w.h.p.,} \end{aligned} \tag{24}$$

where $\mathbf{q}$ is a vector such that $q_i = \frac{nv_i}{\sqrt{w_i}}$. Furthermore, we have w.h.p.

$$\begin{aligned} \frac{A_{ij}}{\sqrt{d_i d_j}} - \frac{\sqrt{d_i d_j}}{\sum_k d_k} &= \frac{A_{ij}}{\sqrt{w_i(1+\epsilon_i) w_j (1+\epsilon_j)}} - \frac{\sqrt{w_i(1+\epsilon_i) w_j(1+\epsilon_j)}}{\sum_k w_k (1+\epsilon_k)} \\ &= \frac{A_{ij}}{\sqrt{w_i w_j}} \left(1 + O(\epsilon_i) + O(\epsilon_j)\right) - \frac{\sqrt{w_i w_j}}{\sum_k w_k} \left(\frac{1 + O(\epsilon_i) + O(\epsilon_j)}{1 + O(\eta)} \right) \\ &= \left(\frac{A_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k} \right) (1 + \delta_{ij}), \end{aligned}$$

where δ_{ij} is the error in the ij^{th} term of the matrix and $\delta_{ij} = O(\eta)$ uniformly, so that $\max_{ij} \delta_{ij} = o(1)$ w.h.p. Consequently, defining $\widetilde{\overline{Q}}_{ij} = \frac{A_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k}$ we have:

$$\begin{aligned}
\left\|\widetilde{\mathbf{Q}}\mathbf{q}\right\|_\infty &\leq \left\|\widetilde{\overline{\mathbf{Q}}}\mathbf{q}\right\|_\infty + \max_i |\sum_j \widetilde{\overline{Q}}_{ij}\delta_{ij}q_j| \\
&\leq \left\|\widetilde{\overline{\mathbf{Q}}}\mathbf{q}\right\|_\infty + O(\eta)\max_i \frac{1}{\sqrt{w_{\min}}}\sum_j |\widetilde{\overline{\mathbf{Q}}}_{ij}| \\
&\leq \left\|\widetilde{\overline{\mathbf{Q}}}\mathbf{q}\right\|_\infty + o(1)\frac{1}{\sqrt{w_{\min}}}\left(C\sqrt{\frac{w_{\max}}{w_{\min}}} + \frac{w_{\max}}{w_{\min}}\right) \qquad (25) \\
&\leq \left\|\widetilde{\overline{\mathbf{Q}}}\mathbf{q}\right\|_\infty + o(1/\sqrt{w_{\min}}) \qquad (26)
\end{aligned}$$

where in (25) we used the fact the $O(\eta)$ is a uniform bound on the error and it is $o(1)$ w.h.p. and $\max_j q_j \leq \frac{1}{\sqrt{w_{\min}}}$. In (25) we also used the fact that

$$\begin{aligned}
\max_i \sum_j |\widetilde{\overline{Q}}_{ij}| &\leq \max_i \sum_j \frac{A_{ij}}{\sqrt{w_i w_j}} + \sum_j \frac{\sqrt{w_i w_j}}{\sum_k w_k} \\
&\leq \max_i \frac{1}{\sqrt{w_{\min}}}\frac{d_i}{\sqrt{w_i}} + \max_i \frac{\sqrt{w_i w_{\max}}}{w_{\min}} \\
&\overset{(a)}{\leq} C\sqrt{\frac{w_i}{w_{\min}}} + \frac{w_{\max}}{w_{\min}} \\
&\leq C\sqrt{\frac{w_{\max}}{w_{\min}}} + \frac{w_{\max}}{w_{\min}},
\end{aligned}$$

where C is some constant. In (a) above we used the fact that w.h.p. $d_i = w_i(1 + o(1))$, by Lemma 1, hence $\exists C$ such that $\forall n$ large enough $d_i \leq Cw_i$.

Now we proceed to bound $\left\|\widetilde{\overline{\mathbf{Q}}}\mathbf{q}\right\|_\infty$. Substituting for $q_i = \frac{1}{\sqrt{w_i}}$, we get

$$\begin{aligned}
\sum_j \frac{1}{\sqrt{w_j}}\left(\frac{A_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k}\right) &= \sum_j \frac{1}{w_j\sqrt{w_i}}\left(A_{ij} - \frac{w_i w_j}{\sum_i w_i}\right) \\
&\equiv \frac{1}{\sqrt{w_i}}X_i. \qquad (27)
\end{aligned}$$

We seek to bound $\max_i |X_i|$:

$$X_i = \sum_j \frac{1}{w_j}\left(A_{ij} - \frac{w_i w_j}{\sum_i w_i}\right).$$

Furthermore, $\mathbb{E}(X_i^2) = \sum_j \frac{1}{w_j^2}\mathbb{E}(A_{ij} - p_{ij})^2$, with $p_{ij} = \frac{w_i w_j}{\sum w_i}$. So, $\mathbb{E}(X_i^2) = \sum_j \frac{1}{w_j^2}p_{ij}(1 - p_{ij}) \leq \frac{w_i}{\sum_i w_i}\sum_j \frac{1}{w_j} \leq n\frac{p_i}{w_{\min}}$, where $p_i = \frac{w_i}{\sum_i w_i}$, and $\frac{A_{ij}}{w_j} \leq 1/w_{\min}$. Therefore using Bernstein Concentration Lemma for $\epsilon < n\max_i p_i$:

$$\begin{aligned}
\mathbb{P}\left(\max_i |\sum_j (A_{ij} - p_{ij})/w_j| \geq \epsilon\right) &\leq n\max_i \exp\left(-\frac{\epsilon^2}{2(p_i n/w_{\min}) + \epsilon/w_{\min}}\right) \\
&\leq n\max_i \exp\left(-\frac{w_{\min}\epsilon^2}{2(np_i + \epsilon)}\right)
\end{aligned}$$

$$\leq n\exp(-\epsilon^2 w_{\min}/(4n\max_i p_i))$$
$$\leq n\exp(\frac{-\epsilon^2 \mathrm{vol} w_{\min}}{4w_{\max}n}), \tag{28}$$

where $\frac{\mathrm{vol}}{n} = \frac{\sum_i w_i}{n} \geq w_{\min}$. It can be verified that when $\epsilon = \frac{1}{(\overline{w})^\alpha}$, for some $\alpha > 0$, the RHS of (28) can be upper bounded by $n^{-(\gamma K-1)}$, if $\overline{w} \geq (\gamma\log(n))^{\frac{1}{1-2\alpha}}$, for some large enough γ, which can be easily satisfied if $w_{\min} \gg O(\log^c(n))$, for some $c > 1$, where K is a constant such that $w_{\max} \leq Kw_{\min}$. Thus, finally, from (27) and (26) we have $\left\|\widetilde{\mathbf{Q}}\mathbf{q}\right\|_\infty = o(1/\sqrt{w_{\min}})$, w.h.p., and therefore from (24), we get $\left\|\widetilde{\mathbf{Q}}\mathbf{v}'\right\|_\infty = o(1/\sqrt{w_{\min}})$.

□

B Proof of Lemmas in Section 5

B.1 Proof of Lemma 8

The proof is an application of Bernstein's Concentration Lemma. Note that for $1 \leq i \leq m$, $D_i = \sum_j A_{ij}$. Here the mean degree $\mathbb{E}(D_i) = mp + (n-m)q = t_1$, and the variance $B_n^2 = mp(1-p)+(n-m)q(1-q) \leq t_1$ for $i \leq m$. Similarly for $i > m$, $\mathbb{E}(D_i) = mq+(n-m)p = t_2$ is and variance $\mathrm{Var}[D_i] \leq t_2$. Then, the minimum average degree $w_{\min} = \min(t_1, t_2)$. By Bernstein's Lemma, for $\epsilon = C\sqrt{\frac{\log(n)}{w_{\min}}}$,

$$\mathbb{P}\left(\max_{1\leq i\leq m}|D_i - t_1| \geq \epsilon t_1\right) \leq 2m\exp\left(\frac{-\epsilon^2 t_1^2}{2(t_1\epsilon/3 + t_1)}\right)$$
$$= 2m\exp\left(\frac{-\epsilon^2 t_1}{1+\epsilon/3}\right)$$
$$= O(n^{-c}),$$

for some c. Hence $\max_{1\leq i\leq m}\left|\frac{D_i - t_1}{t_1}\right| \leq C\sqrt{\frac{\log(n)}{w_{\min}}}$ w.h.p. Similarly

$$\max_{1+m\leq i\leq n/2}\left|\frac{D_i - t_2}{t_2}\right| \leq C\sqrt{\frac{\log(n)}{w_{\min}}}, \text{ w.h.p.}$$

Combining the two bounds above we get,

$$\max_{1\leq i\leq n}\left|\frac{D_i}{\mathbb{E}(D_i)} - 1\right| \leq C\sqrt{\frac{\log(n)}{w_{\min}}}, \text{ w.h.p.} \tag{29}$$

□

B.2 Proof of Lemma 9

To prove Lemma 9 we need the following lemma on the spectral norm of the difference between the adjacency matrix and its mean.

Lemma 10 *For an SBM matrix $G(m, n-m, p, q)$ with adjacency matrix $\mathbf{A}$ and $\overline{\mathbf{A}} = \mathbb{E}(\mathbf{A})$, there exists a constant K s.t.*

$$\|\mathbf{A} - \overline{\mathbf{A}}\|_2 \leq K\sqrt{\log(n) w_{max}}, \ \ w.h.p.,$$

where $w_{max} = \max(m, n-m)p + \min(m, n-m)q$ is the maximum average degree, if $w_{\max} = !(\log^3(n))$.

To prove this Lemma we need the Matrix Bernstein Concentration result, which we state below for the sake of completeness:

Lemma 11 *[35, Theorem 1.4]. Let $\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_t$ be independent random matrices with common dimension $d_1 \times d_2$. Assume that each matrix has bounded deviation from its mean, i.e.,*

$$\|\mathbf{S}_k - \mathbb{E}(\mathbf{S}_k)\| \leq R, \ \textit{for each } k = 1, \ldots n.$$

Let $\mathbf{Z} = \sum_{k=1}^{t} \mathbf{S}_k$ and introduce a variance parameter

$$\sigma_{\mathbf{Z}}^2 = \max\left\{\|\mathbb{E}\left((\mathbf{Z} - \mathbb{E}(\mathbf{Z}))(\mathbf{Z} - \mathbb{E}(\mathbf{Z}))^H\right)\|, \|\mathbb{E}\left((\mathbf{Z} - \mathbb{E}(\mathbf{Z}))^H(\mathbf{Z} - \mathbb{E}(\mathbf{Z}))\right)\|\right\}.$$

Then

$$\mathbb{P}\{\|\mathbf{Z} - \mathbb{E}(\mathbf{Z})\| > t\} \leq (d_1 + d_2).\exp\left(\frac{-t^2/2}{\sigma_{\mathbf{Z}}^2 + Rt/3}\right), \tag{30}$$

for all $t \geq 0$.

Proof of Lemma 10: With $\mathbf{Z} = \mathbf{A}$, in Lemma 11, we can decompose $\mathbf{Z}$ as sums of Hermitian matrices $\mathbf{S}_{i'j'}$, $\mathbf{Z} = \sum_{1 \leq i' < j' \leq n} \mathbf{S}_{i'j'}$ such that:

$$(\mathbf{S}_{i'j'})_{ij} = \begin{cases} A_{i'j'} \text{ if } i = i', j = j', \\ A_{i'j'} \text{ if } i = j', j = i', \\ 0 \text{ otherwise.} \end{cases} \tag{31}$$

Notice that if $\mathbf{x} \neq 0$, $\|(\mathbf{S}_{i'j'} - \mathbb{E}(\mathbf{S}_{i'j'}))\mathbf{x}\|_2 = |2x_{i'}x_{j'}(A_{i'j'} - \mathbb{E}(A_{i'j'}))| < |x_{i'}^2 + x_{j'}^2|$. Consequently $\|\mathbf{S}_{i'j'} - \mathbb{E}(\mathbf{S}_{i'j'})\|_2 < 1$, giving $R = 1$ in the statement of Lemma 30. Let $\mathbf{Y} = \mathbb{E}\left((\mathbf{Z} - \mathbb{E}\mathbf{Z})^H(\mathbf{Z} - \mathbb{E}\mathbf{Z})\right)$, then

$$Y_{ij} = \begin{cases} v_1 & \text{if } i = j, i \leq m, \\ v_2 & \text{if } i = j, i > m, \\ 0 & \text{otherwise,} \end{cases} \tag{32}$$

where $v_1 = mp(1-p) + q(1-q)(n-m), v_2 = (n-m)p(1-p) + mq(1-q)$. Therefore $\sigma_{\mathbf{Z}}^2 = \max(v_1, v_2) = \max(n-m, m)p + \min(n-m, m)q = \sigma^2$. By our assumptions on the probabilities, $\sigma^2 = !(\log^3(n))$. Thus it follows that

$$\begin{aligned} \mathbb{P}(\|\mathbf{A} - \overline{\mathbf{A}}\| \geq t\sigma) &\leq 2n\exp\left(\frac{-t^2\sigma^2}{2\sigma^2 + t\sigma/3}\right) \\ &\leq 2n\exp(-t^2/3), \end{aligned}$$

if $\sigma > t$. The RHS is $O(n^{-c})$ if $t > \sqrt{r\log(n)}$, for some r. □Finally we are in a position to prove Lemma 9

Proof of Lemma 9: We prove this result in two steps. First we show that

$$\|\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} - \mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\|_2 = C\sqrt{\frac{\log(n)}{w_{\min}}} = o(1). \tag{33}$$

Observe that

$$\begin{aligned}
&\|\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} - \mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2}\|_2 = \|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q}\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\| \\
&= \|\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q} + \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q}\mathbf{D}^{1/2}\mathbf{W}^{-1/2}\|_2 \\
&= \|(\mathbf{I} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2})\mathbf{Q} + \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\mathbf{Q}(\mathbf{I} - \mathbf{D}^{1/2}\mathbf{W}^{-1/2})\|_2 \\
&\leq \delta + (1+\delta)\delta,
\end{aligned}$$

where $\delta = \max_i \left|\frac{d_i}{w_i} - 1\right|$. In the last line we used the fact that $\|\mathbf{Q}\|_2 = 1$, $\|\mathbf{I} - \mathbf{W}^{-1/2}\mathbf{D}^{1/2}\|_2 = \max_i \left|\sqrt{\frac{d_i}{w_i}} - 1\right| \leq \max_i \left|\frac{d_i}{w_i} - 1\right|$ and

$$\|\mathbf{W}^{-1/2}\mathbf{D}^{1/2}\|_2 \leq \|\mathbf{W}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}\|_2 + \|\mathbf{I}\|_2 \leq \delta + 1.$$

By Lemma 8, $\delta \leq C\sqrt{\frac{\log(n)}{w_{\min}}} = o(1)$ w.h.p. Next we show that

$$\|\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2} - \mathbf{W}^{-1/2}\overline{\mathbf{A}}\mathbf{W}^{-1/2}\|_2 \leq \frac{C\sqrt{\log(n)w_{\max}}}{w_{\min}} = o(1). \tag{34}$$

Now using Lemma 10 we have

$$\begin{aligned}
\|\mathbf{W}^{-1/2}\mathbf{A}\mathbf{W}^{-1/2} - \mathbf{W}^{-1/2}\overline{\mathbf{A}}\mathbf{W}^{-1/2}\| &\leq \frac{\|\mathbf{A} - \overline{\mathbf{A}}\|_2}{w_{\min}} \\
&\leq \frac{c\sqrt{\log(n)w_{\max}}}{w_{\min}} \\
&= o(1), \text{ w.h.p.},
\end{aligned}$$

if $w_{\min} = !(\sqrt{\log(n)w_{\max}})$, which is satisfied when $w_{\max} \leq Cw_{\min}$ for some C, and $w_{\max} = !(\log^3(n))$. The result of Lemma 9 then follows from (33) and (34) by applying the triangular inequality. □

Contents

1 Introduction **3**

2 Definitions **3**

3 Convergence in total variation **4**

4 Chung-Lu random graphs **7**
4.1 Chung-Lu Random Graph Model 7
4.2 Element-wise Convergence 10

5 Asymptotic PageRank for the Stochastic Block Model **11**

6 Experimental Results **15**

7 Conclusions **17**

A Proof of Lemma 6 **21**

B Proof of Lemmas in Section 5 **23**
B.1 Proof of Lemma 8 23
B.2 Proof of Lemma 9 23

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399