Network Science **9** (S1): S4–S22, 2021
doi:10.1017/nws.2020.34

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Gradient and Harnack-type estimates for PageRank

Paul Horn[1] and Lauren M. Nelsen[2,*]

[1]University of Denver, Denver, CO 80208, USA (*e-mail:* paul.horn@du.edu) and [2]University of Indianapolis, Indianapolis, IN 46227, USA
*Corresponding author. Email: nelsenl@uindy.edu

Special Issue Editor: Hocine Cherifi

Abstract

Personalized PageRank has found many uses in not only the ranking of webpages, but also algorithmic design, due to its ability to capture certain geometric properties of networks. In this paper, we study the diffusion of PageRank: how varying the jumping (or teleportation) constant affects PageRank values. To this end, we prove a gradient estimate for PageRank, akin to the Li–Yau inequality for positive solutions to the heat equation (for manifolds, with later versions adapted to graphs).

Keywords: PageRank; discrete curvature; random walks; gradient estimate

1. Introduction/Background

Personalized PageRank, developed by Brin & Page (1998), was developed to rank the importance of webpages; the personalized version was intended to rank the importance of webpages with respect to a specified seed. The novelty at the time of introduction was that it uses the geometric aspects of the network—in particular the link structure—to rank the webpages.

While there are a variety of interpretations of PageRank available, one of the most fruitful ones is that it gives the distribution of a random walk allowed to diffuse for a geometrically distributed number of steps. The parameter controlling the (expected) length of the random walk involved is the "jumping" or "teleportation" constant. As the jumping constant controls the length, it controls locality—that is, how far from the seed the random walk is (likely) willing to stray.

When the jumping constant is small, the involved walks are (on average) short, and the mass of the distribution will remain concentrated near the seed. As the jumping constant increases, then the involved walk will (likely) be much longer. This allows the random walk to mix, and the involved distribution tends toward the stationary distribution of the random walk. As the PageRank of individual vertices (for a fixed jumping constant) can be thought of as a measure of importance to the seed, then as the jumping constant increases this importance diffuses.

In this paper, we are interested in how this importance diffuses as the jumping constant increases. This diffusion is related to the network's geometry; in particular, the importance can get "caught" by small cuts. This partially accounts for PageRank's importance in web search but has other uses as well—for instance, Andersen, Chung, and Lang use PageRank to implement local graph partitioning algorithms in Andersen et al. (2008), and in Chung et al. (2009), the authors investigate the evolution of the contact process, a continuous time model of disease, using PageRank.

This paper seeks to understand the diffusion of influence (as the jumping constant changes) in analogy to the diffusion of heat. The study of solutions to the heat equation $\Delta u = \frac{\partial}{\partial t} u$ on both graphs and manifolds has a long history, motivated by its close ties to geometric properties of graphs. On graphs, the relationship between heat flow and PageRank has been exploited several times. For instance, Chung (2007) introduced the notion of heat kernel PageRank and used it to improve the algorithm of Anderson, Chung, and Lang for graph partitioning.

A particularly useful way of understanding positive solutions to the heat equation is through curvature lower bounds, which can be used to prove "gradient estimates", which bound how heat diffuses locally in space and time and which can be integrated to obtain Harnack inequalities. Most classical of these is the Li–Yau inequality (Li & Yau, 1986), which (in its simplest form) states that if u is a positive solution on a non-negatively curved n-dimensional compact manifold, then u satisfies

$$\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \leq \frac{n}{2t} \tag{1}$$

This inequality shows how, in a sense, heat "smooths out" as time goes to infinity in such a way that at any point where the gradient remains large the time derivative (which represents the heating/cooling at a point) is large enough to ensure the entire quantity is bounded. In the graph setting, Bauer, et al. proved a gradient estimate for the heat kernel on graphs in Bauer et al. (2015). In this paper, we aim to prove a similar inequality for PageRank. Our gradient estimate, which is formally stated as Theorem 1 below, is proved using the exponential curvature dimension inequality *CDE*, introduced by Bauer et al. It, like the Li–Yau inequality shows how heat smooths out as time gets larger, shows that the PageRank vector smooths out as the jumping constant changes.

We remark that, in some ways, our inequality is more closely related to another inequality of Hamilton (1993) which bounds merely $\frac{|\nabla u|^2}{u^2}$ and was established for graphs by Horn in (2019). Indeed, Theorem 1 shows a bound on a function of this form where u is replaced by a PageRank vector (and ∇ is appropriately defined for graphs). This shows that not only the gradient of PageRank is decreasing, but also actually a normalized version of such a bound is decreasing.

Other related works establish gradient estimates for eigenfunctions for the Laplace matrix; these include Chung et al. (2014).

This paper is organized as follows: in the next section, we introduce definitions for both PageRank and the graph curvature notions used. We further establish a useful "time parameterization" for PageRank, which allows us to think of increasing the jumping constant as increasing a time parameter and makes our statements and proofs cleaner. In Section 3, we prove a gradient estimate for PageRank. In Section 4, we use this gradient estimate to prove a Harnack-type inequality that allows us to compare PageRank at two vertices in a graph.

In the conference version of this paper, Horn & Nelsen (2019), this result was announced and a sketched proof of the gradient estimate was given. The present paper expands on this by giving a full proof but more critically by providing a full statement and proof of the Harnack inequality which was not fully described in the conference version.

2. Preliminaries

2.1 Spectral graph theory and graph Laplacians

Spectral graph theory involves associating a matrix (or operator) with a graph and investigating how eigenvalues of the associated matrix reflect graph properties. The most familiar such matrix is the adjacency matrix A, whose rows and columns are indexed with vertices and

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \sim v_j \\ 0 & \text{else} \end{cases}$$

where $v_i \sim v_j$ if and only if the vertices v_i and v_j are adjacent.

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

The adjacency matrix is a real, symmetric matrix and has n eigenvalues, $\lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Eigenvalues of A capture many graph properties but not all. For instance, they tell us how many edges are in the graph, but they do not tell us if the graph is connected in general.

Another matrix that is often studied is the Laplacian matrix (which is sometimes called the *combinatorial Laplacian matrix*). The Laplacian matrix is defined to be $L = D - A$, where D is the diagonal matrix of vertex degrees.

In this work, the principal matrix that we will consider is the normalized Laplace operator

$$\Delta = D^{-1}A - I$$

where D is the diagonal matrix of vertex degrees and $D^{-1}A$ is the transition probability matrix for a simple random walk.

Thought of as an operator acting on functions $f : V \to \mathbb{R}$, the Laplacian acts as

$$\Delta f(x) = \frac{1}{\deg(x)} \sum_{y \sim x} (f(y) - f(x))$$

As a quick observation, note that Δ is non-positive semidefinite. This is contrary to usual sign conventions in graph theory but is the proper sign convention for the Laplace–Beltrami operator in Riemannian manifolds, the analogy we emphasize in this paper. Also note that this matrix is (up to sign) the unsymmetrized version of the normalized Laplacian popularized by Chung (see Chung, 1997), $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$.

We remark briefly that the machinery utilized in this paper—using curvature lower bounds as introduced in Section 2.3 below—can be adapted to much more general Laplace operators essentially without change: arbitrary edge and vertex weights (w_{xy} and $\mu(x)$, respectively) can be introduced to obtain a Laplace operator so that

$$\Delta f(x) = \frac{1}{\mu(x)} \sum_{y \sim x} w_{xy}(f(y) - f(x))$$

Setting $\mu(x) := \deg(x)$ yields the normalized Laplacian, while setting $\mu(x) := 1$ yields the combinatorial Laplacian. For the purpose of the study of PageRank, however, the normalized Laplace operator is the most natural.

One of the most important bits of geometric information certified by the spectrum is expansion: the number of edges leaving subsets. For instance, when looking for disjoint spanning structures in graphs, a sparse cut limits the number we can hope to find. It turns out that sparse cuts in our graph are related to an eigenvalue of the normalized Laplacian matrix through Cheeger's inequality. Before discussing Cheeger's inequality, we introduce some necessary notation.

For a graph G, the volume of a subset S, of $V(G)$, denoted by $\mathrm{Vol}(S)$, is defined as follows:

$$\mathrm{Vol}(S) = \sum_{v \in S} \deg(v)$$

We define $E(S, \overline{S})$ to be the set of edges with one end in S and the other outside of S and let $h_G(S) = \dfrac{|E(S, \overline{S})|}{\min\{\mathrm{Vol}(S), \mathrm{Vol}(\overline{S})\}}$. The *Cheeger constant* (or *isoperimetric constant*), h_G is then defined by

$$h_G = \min_S h_G(S)$$

Determining h_G is computationally difficult but is related to an eigenvalue of the normalized Laplacian matrix through the following result known as *Cheeger's inequality*.

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

Proposition 1 (**Cheeger's inequality, Chung, 1997**). *If G is a connected graph and λ_1 is the second smallest eigenvalue of the normalized Laplacian of G, then*

$$\frac{h_G^2}{2} < \lambda_1 \leq 2h_G$$

Proposition 1 reveals that a sparse (normalized) cut in G is equivalent to λ_1 being small. This equivalence shows up in the close relationship between spectral properties of Δ and the mixing of random walks and is really key in the diffusion of PageRank as explored below.

We remark that the lower bound $|\lambda_1| \geq h_G^2/2$ of this inequality is analogous to Cheeger's inequality in Riemannian geometry (Cheeger, 1970). The first graph theoretical result was by Dodziuk (1984), for infinite graphs. Later versions were proved for (finite) regular graphs (Alon, 1986). It is interesting to note that not just are the statements of the Cheeger inequality similar in Riemannian geometry and graph theory, but even the proofs are similar.

2.2 PageRank

Personalized PageRank was introduced as a ranking mechanism (Jeh & Widom, 2003) to rank the importance of webpages with respect to a seed. To define personalized PageRank, we introduce the following operator which we call the PageRank operator. This operator, $P(\alpha)$, is defined as follows:

$$P(\alpha) = (1-\alpha)\sum_{k=1}^{\infty} \alpha^k W^k$$

where $W = D^{-1}A$ is the transition probability matrix for a simple random walk. Here, the parameter α is known as the jumping or teleportation constant. For a finite n-vertex graph, $P(\alpha)$ is a square matrix; the *personalized PageRank vector* of a vector $u : V \to \mathbb{R}$ is

$$u^T P(\alpha) = (1-\alpha)\sum_{k=1}^{\infty} \alpha^k u^T W^k$$

PageRank can then be viewed as the distribution of a geometric sum of the distribution of simple random walks, that is, the expected distribution of a simple random walk of length geometrically distributed with parameter $1-\alpha$ starting at initial distribution u. As $\alpha \to 1$, the expected length of this geometric random walk tends to infinity, and the resulting distribution tends to the limiting distribution of a simple random walk (which, for a finite graph is proportional to the degree).

It has been noticed (Chung, 2007, 2011) that PageRank has many similarities to the heat kernel $e^{t\Delta}$. Chung defined the notion of "Heat Kernel PageRank" to exploit these similarities. In this work, we take inspiration in the opposite direction: we are interested in understanding the action of the PageRank operator in analogy to solutions of the heat equation.

In order to emphasize our point of view, we note that graph theorists view the heat kernel operator in two different ways: for a vector $u : V \to \mathbb{R}$ studying the evolution of

$$u^T e^{t\Delta}$$

as $t \to \infty$ is really studying the evolution of the *continuous time random walk* while studying the evolution of

$$e^{t\Delta} u$$

as $t \to \infty$ is studying the solutions to the heat equation

$$\Delta u = u_t$$

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

The differing behavior of these two evolutions comes from the fact that (for irregular graphs) the left and right eigenvectors of $\Delta = W - I$ are different: the left Perron–Frobenius eigenvector of Δ is proportional to the degrees of a graph (as it captures the stationary distribution of the random walk), while the right Perron–Frobenius eigenvector is the constant vector. In particular, as $t \to \infty$, the vector $e^{t\Delta}u$ tends to a constant. Physically, this represents the "heat" on a graph evening out, and this regularization (and the rate of regularization) is related to a number of geometric features of a graph.

A similar feature holds for PageRank. As $\alpha \to 1$, $u^T P(\alpha)$ tends to a vector proportional to degrees, but $P(\alpha)u$ regularizes. Here, we study this regularization. Note that the real novelty of what we do here is that we are understanding the evolution of PageRank *locally* using a notion of graph curvature defined at the end of this section.

To see this regularization, notice that

$$\begin{aligned} W &= D^{-1}A \\ &= D^{-1/2}\left(D^{-1/2}AD^{-1/2}\right)D^{1/2} \\ &= D^{-1/2}\left(\sum_{i=0}^{n-1} \lambda_i \varphi_i \varphi_i^T\right) D^{1/2} \end{aligned}$$

where $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1}$ are the eigenvalues of $D^{-1/2}AD^{-1/2}$ and $\varphi_0, \cdots, \varphi_{n-1}$ are the corresponding orthonormal eigenvectors. Let $\mathbf{d}$ be the n-dimensional vector where $\mathbf{d}_i = \deg(v_i)$. Notice that $\varphi_0 = \frac{\mathbf{d}^{1/2}}{\sqrt{\mathrm{Vol}(G)}}$ is the eigenvector corresponding to $\lambda_0 = 1$, since

$$\begin{aligned} D^{-1/2}AD^{-1/2}\mathbf{d}^{1/2} &= D^{-1/2}A\mathbb{1} \\ &= D^{-1/2}\mathbf{d} \\ &= \mathbf{d}^{1/2} \end{aligned}$$

So

$$W^k = D^{-1/2}\left(\sum_{i=0}^{n-1} \lambda_i^k \varphi_i \varphi_i^T\right) D^{1/2}$$

Thus,

$$\begin{aligned} P(\alpha) &= (1-\alpha)\sum_{k=0}^{\infty} \alpha^k W^k \\ &= (1-\alpha)\sum_{k=0}^{\infty} D^{-1/2}\left(\sum_{i=0}^{n-1} (\alpha\lambda_i)^k \varphi_i\varphi_i^T\right) D^{1/2} \\ &= (1-\alpha)\sum_{i=0}^{n-1} D^{-1/2}\left(\sum_{k=0}^{\infty} (\alpha\lambda_i)^k \varphi_i\varphi_i^T\right) D^{1/2} \text{ by Fubini's theorem} \\ &= (1-\alpha)\sum_{i=0}^{n-1} D^{-1/2}\left(\frac{1}{1-\alpha\lambda_i}\varphi_i\varphi_i^T\right) D^{1/2} \\ &= (1-\alpha)\left(\frac{1}{1-\alpha}D^{-1/2}\frac{\mathbf{d}^{1/2}\left(\mathbf{d}^{1/2}\right)^T}{\mathrm{Vol}(G)}D^{1/2} + \sum_{i=1}^{n-1} D^{-1/2}\left(\frac{1}{1-\alpha\lambda_i}\varphi_i\varphi_i^T\right)D^{1/2}\right) \\ &= \frac{\mathbb{1}\mathbf{d}^T}{\mathrm{Vol}(G)} + \sum_{i=1}^{n-1}\frac{1-\alpha}{1-\alpha\lambda_i}D^{-1/2}\varphi_i\varphi_i^T D^{1/2} \end{aligned}$$

Consider

$$\frac{1-\alpha}{1-\alpha\lambda_i} \tag{2}$$

Notice that if G is connected and not bipartite, then $\lambda_i < 1$ for $1 \le i \le n-1$. So $\dfrac{1-\alpha}{1-\alpha\lambda_i} \to 0$ as $\alpha \to 1$. As $\alpha \to 1$, the dominant term in $P(\alpha)$ becomes $\dfrac{\mathbb{1}\mathbf{d}^T}{\mathrm{Vol}(G)}$. Note that the smaller the λ_i are, the faster this tends to zero. This is Cheeger's inequality in action—if $\lambda_1 \approx 1$, then there is a sparse cut (since $D^{-1/2}AD^{-1/2} = I - \mathcal{L}$), so diffusion will take longer. If the λ_i are far from 1, then there will not be a sparse cut, so the diffusion will happen more quickly.

If φ is a probability distribution, then

$$\varphi^T \frac{\mathbb{1}\mathbf{d}^T}{\mathrm{Vol}(G)} = \frac{\mathbf{d}^T}{\mathrm{Vol}(G)}$$

which is the stationary distribution for a random walk.

If, instead, we consider $P(\alpha)u$ for a vector u, then $\dfrac{\mathbb{1}\mathbf{d}^T}{\mathrm{Vol}(G)}u = c\mathbb{1}$, for a constant c. So $P(\alpha)$ regularizes or "smooths out" as $\alpha \to 1$.

We note that the left and right actions of the PageRank operator are closely related, and we study the left action versus the right action. For an undirected graph,

$$u^T P(\alpha) = (P(\alpha)^T u)^T = (DP(\alpha)D^{-1}u)^T$$

so that the regularization of $D^{-1}u$ can be translated into information on the "mixing" of the personalized PageRank vector seeded at u.

To complete the analogy between $P(\alpha)u$ and $e^{t\Delta}u$, it is helpful to come up with a time parameterization $t = t(\alpha)$ so we can view the regularization as a function of "time", in analogy to the heat equation. To do this in the best way, it is useful to think of $\alpha = \alpha(t)$ and compute $\dfrac{\partial}{\partial t}P_\alpha$.

Lemma 1

$$\frac{\partial}{\partial t}P_\alpha = \frac{\alpha'}{(1-\alpha)^2}\Delta P_\alpha^2$$

where $\Delta P_\alpha^2 = \Delta P_\alpha(P_\alpha)$.

Proof. From the chain rule and algebra, we have

$$\begin{aligned}\frac{\partial}{\partial t}P_\alpha &= \frac{\partial}{\partial t}(1-\alpha)(I-\alpha W)^{-1}\\ &= \alpha'((\alpha W - I)(I-\alpha W)^{-2} + (1-\alpha)W(I-\alpha W)^{-2}) = \frac{\alpha'}{(1-\alpha)^2}\Delta P_\alpha^2\end{aligned}$$

□

This is remarkably close to the heat equation if $\alpha'(t) = (1-\alpha)^2$; solving this separable differential equation yields that $\alpha = \alpha(t) = 1 - \frac{1}{t+C}$. Since we desire a parameterization so that $\alpha(0) = 0$ and $\alpha \to 1$ as $t \to \infty$, this gives us that $C = 1$ from whence we obtain

$$\boxed{\begin{aligned}\alpha(t) &= 1 - \frac{1}{t+1} &(3)\\ t &= \frac{\alpha}{1-\alpha} &(4)\end{aligned}}$$

Given the time parameterization in Equation (3), we get the following Corollary to Lemma 1.

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

Corollary 2

$$\frac{\partial}{\partial t}P_\alpha = \Delta P_\alpha^2$$

where $\Delta P_\alpha^2 = \Delta P_\alpha(P_\alpha)$.

Proof. From Lemma 1 and our choice of parameterization, we see that

$$\frac{\partial}{\partial t}P_\alpha = \frac{\alpha'}{(1-\alpha)^2}\Delta P_\alpha^2 = \frac{\frac{1}{(t+1)^2}}{\left(\frac{1}{t+1}\right)^2}\Delta P_\alpha^2 = \Delta P_\alpha^2$$

□

Fix a vector $u : V \to \mathbb{R}$. From now on, we let

$$\boxed{f = P_\alpha u} \tag{5}$$

Lemma 2. *For* $f = P_\alpha u$ *and* $t = \dfrac{\alpha}{1-\alpha}$, *we have that* $\Delta f = \dfrac{f-u}{t}$.

Proof. We know that $W = D^{-1}A$ and $\Delta = W - I$, so

$$\begin{aligned}\Delta P_\alpha &= (W-I)(1-\alpha)(I-\alpha W)^{-1}\\ &= -\frac{1}{\alpha}(I-\alpha W)(1-\alpha)(I-\alpha W)^{-1} + \frac{1-\alpha}{\alpha}\cdot(1-\alpha)(I-\alpha W)^{-1}\\ &= \frac{1-\alpha}{\alpha}(P_\alpha - I)\end{aligned}$$

Hence,

$$\Delta f = \Delta P_\alpha u = \frac{(1-\alpha)}{\alpha}(P_\alpha - I)u = \frac{f-u}{t}$$

□

2.3 Graph curvature

In this paper, we study the regularization of $P(\alpha)u$ for an initial seed u as $\alpha \to 1$. On one hand, as seen above, the information about this regularization is contained in the spectral decomposition of the random walk matrix W. As $\alpha \to 1$, all eigenvalues of P_α tend to zero except for the eigenvalue, 1, of W, and this is what causes the regularization. Thus, the difference between $P_\alpha u$ and the constant vector can be bounded in terms of (say) the infinity norms of eigenvectors of P_α and α itself.

On the other hand, curvature lower bounds (in graphs and manifolds) have proven to be important ways to understand the local evolution of solutions to the heat equation. As we have already noted important similarities between heat solutions and PageRank, we seek similar understanding in the present case. Curvature, for graphs and manifolds, gives a way of understanding the local geometry of the object. A manifold (or graph) satisfying a curvature lower bound at every point has a locally constrained geometry which allows a *local* understanding of heat flow through which a "gradient estimate" can be proved. These gradient estimates can then be "integrated" over space-time to yield Harnack inequalities which compare the "heat" of different points at different times.

While a direct analog of the Ricci curvature is not defined in a graph setting, a number of graph theoretical analogs have been developed recently in an attempt to apply geometrical ideas in the graph setting. In the context of proving gradient estimates of heat solutions, a new notion

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

of curvature known as the *exponential curvature dimension inequality* was introduced in Bauer et al. (2015). In order to discuss the exponential curvature dimension inequality, we first need to introduce some notation. As seen above, the normalized Laplace operator, Δ, on a graph G is defined at a vertex x by

$$\Delta f(x) = \frac{1}{\deg(x)} \sum_{y \sim x} (f(y) - f(x))$$

Definition 3. *For $x \in V(G)$,*

$$\widetilde{\sum}_{y\sim x} h(x, y) = \frac{1}{\deg(x)} \sum_{y \sim x} h(x, y)$$

Definition 4. *The gradient form Γ is defined by*

$$\begin{aligned} 2\Gamma(f, g)(x) &= (\Delta(f \cdot g) - f \cdot \Delta(g) - \Delta(f) \cdot g)(x) \\ &= \widetilde{\sum}_{y\sim x} (f(y) - f(x))(g(y) - g(x)) \end{aligned}$$

and we write $\Gamma(f) = \Gamma(f, f)$.

Note that $\Gamma(f, g)$ plays the role of $\langle \nabla f, \nabla g\rangle$ (and $\Gamma(f)$ of $|\nabla f|^2$) in the graph theoretical setting and that it is not uncommon notation in the literature to use the gradient (i.e. ∇) instead of Γ.

In general, there is no "chain rule" that holds for the Laplacian on graphs. However, the following formula from Bauer et al. (2015) does hold for the Laplacian and will be useful to us:

$$\Delta f = 2\sqrt{f}\Delta\sqrt{f} + 2\Gamma\left(\sqrt{f}\right) \tag{6}$$

At the heart of the exponential curvature dimension inequality is an idea that had been used previously based on the Bochner formula. The Bochner formula reveals a connection between solutions to the heat equation and the curvature of a manifold. The Bochner formula implies that for an n-dimensional manifold with Ricci curvature at least K, we have

$$\frac{1}{2}\Delta|\nabla f|^2 \geq \langle \nabla f, \nabla \Delta f\rangle + \frac{1}{n}(\Delta f)^2 + K|\nabla f|^2 \tag{7}$$

An important insight of Bakry and Emery was that an object satisfying an inequality like (7) could be used as a *definition* of a curvature lower bound even when curvature could not be directly defined. Such an inequality became known as a *curvature dimension inequality*, or the CD inequality. Bauer et al. introduced a modification of the CD inequality that defines a new notion of curvature on graphs that we will use here Bauer et al. (2015), the exponential curvature dimension inequality.

Definition 5. *A graph is said to satisfy the exponential curvature dimension inequality $CDE(n, K)$ if, for all positive $f : V \to \mathbb{R}$ and at all vertices $x \in V(G)$ satisfying $(\Delta f)(x) < 0$*

$$\Delta\Gamma(f) - 2\Gamma\left(f, \frac{\Delta f^2}{2f}\right) \geq \frac{2}{n}(\Delta f)^2 + 2K\Gamma(f) \tag{8}$$

where the inequality in (8) is taken pointwise.

While the inequality (8) may seem somewhat unwieldy, as shown in Bauer et al. (2015), it arises from "baking in" the chain rule and is actually equivalent to the standard curvature dimension inequality (7) in the setting of diffusive semigroups (where the Laplace operator satisfies the chain rule). Additionally, in Bauer et al. (2015), it is shown that some graphs, including the Ricci flat graphs of Chung and Yau, satisfy $CDE(n, 0)$ (and hence are non-negatively curved for this curvature notion), and some general curvature lower bounds for graphs are given.

We remark that the parameters n and K of CDE can be interpreted as an effective dimension (upper bound) and curvature lower bound of the graph: if $CDE(n, K)$ hold for some graph G so does $CDE(n', K')$ for $n' > n$ or $K' < K$. Thus, this gives a notion of "dimension" to the (inherently combinatorial) graph. There is often a bit of a trade-off between curvature and dimension—but as the curvature is generally the most important part in applications of the CDE inequality, when showing a graph satisfies $CDE(n, K)$ one typically tries to first take K as large as possible, then n as small as possible. Graphs are known to exist, for instance, satisfying $CDE(\infty, 0)$, which satisfy $CDE(n, 0)$ for no finite n, while the standard Cayley graphs $\mathbb{Z}_d$ are known to satisfy $CDE(2d, 0)$. In gradient estimates like what we prove, the dimension tends to control the constants.

An important observation is that this notion of curvature only requires looking at the second neighborhood of a graph, and hence this kind of curvature is truly a local property (and hence a curvature lower bound can be certified by only inspecting second neighborhoods of vertices).

3. Gradient estimate for PageRank

Our main result will make use of the following lemma from Bauer et al. (2015), and we include its simple proof for completeness.

Lemma 3 (**Bauer et al., 2015**). *Let $G(V, E)$ be a (finite or infinite) graph, and let $f, H : V \times \{t^\star\} \to \mathbb{R}$ be functions. If $f \geq 0$ and H has a local maximum at $(x^\star, t^\star) \in V \times \{t^\star\}$, then*

$$\Delta(fH)(x^\star, t^\star) \leq (\Delta f)H(x^\star, t^\star)$$

Proof. Observe that

$$\begin{aligned}\Delta(fH)(x^\star, t^\star) &= \widetilde{\sum}_{y\sim x^\star}(f(y, t^\star)H(y, t^\star) - f(x^\star, t^\star)H(x^\star, t^\star))\\ &\leq \widetilde{\sum}_{y\sim x^\star}(f(y, t^\star)H(x^\star, t^\star) - f(x^\star, t^\star)H(x^\star, t^\star))\\ &= (\Delta f)H(x^\star, t^\star)\end{aligned}$$

□

Our goal is to show that $\frac{\Gamma(\sqrt{f})}{\sqrt{f\cdot M}} \leq \frac{C(t)}{t}$ for some function $C(t)$, where $0 \leq f(x) \leq M$ for all $x \in V(G)$ and $t \in [0, \infty)$. However, $\frac{C(t)}{t}$ is badly behaved as $t \to 0$. The way that we handle this is by showing that $H := t \cdot \frac{\Gamma(\sqrt{f})}{\sqrt{f\cdot M}} \leq C(t)$. If H is a function from $V \times [0, \infty) \to \mathbb{R}$, then instead consider H as a function from $V \times [0, T] \to \mathbb{R}$ for some $T > 0$. Then, by compactness, there is a point $(x^\star, t^\star)$ in $V \times [0, T]$ at which $H(x, t)$ is maximized. Using the CDE inequality, along with some other lemmas and an identity, we are able to relate H^2 with H. This allows us to find an upper bound for H, and thus for $\frac{\Gamma(\sqrt{f})}{\sqrt{f\cdot M}}$. Our situation is a little easier because we consider a fixed t.

Lemma 4. *Let G be a graph, and let $f = P_\alpha u$ for some seed u. Suppose $0 \leq f(x) \leq M$ for all $x \in V(G)$ and $t \in [0, \infty)$ and let $H = \frac{t\Gamma(\sqrt{f})}{\sqrt{f\cdot M}}$. Then*

$$\Delta\sqrt{f} = \frac{f - u}{2t\sqrt{f}} - \frac{\sqrt{M}H}{t}$$

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

Proof. Using (6), we get that $\Delta\sqrt{f} = \dfrac{\Delta f - 2\Gamma\left(\sqrt{f}\right)}{2\sqrt{f}}$. Thus,

$$\begin{aligned}\Delta\sqrt{f} &= \frac{\Delta f}{2\sqrt{f}} - \frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f}} \\ &= \frac{f-u}{2t\sqrt{f}} - \frac{\sqrt{M}H}{t} \text{ by Lemma 2.}\end{aligned}$$

□

At the heart of the proof of the Li–Yau inequality on manifolds is the identity

$$\Delta \log u = \frac{\Delta u}{u} - |\nabla \log u|^2 = \frac{\Delta u}{u} - \frac{|\nabla u|^2}{u^2}$$

The Li–Yau inequality on graphs (Bauer et al., 2015) uses the identity

$$\frac{\Delta\sqrt{u}}{\sqrt{u}} = \frac{\Delta u}{u} - \frac{\Gamma(\sqrt{u})}{u}$$

Lemma 4 is similar to these other identities and the CDE inequality allows us to exploit this relationship.

Theorem 1. *Let G be a graph satisfying CDE$(n, 0)$ and let $f = P_\alpha u$ for some seed u. Suppose $0 \leq f \leq M$ for all $x \in V(G)$ and $t \in (0, \infty)$. Then*

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f \cdot M}} \leq \frac{n+4}{n+2} \cdot \frac{1}{t} + \sqrt{\frac{n}{n+2}} \cdot \frac{1}{\sqrt{t}}$$

Notice that a true Li–Yau-type inequality would have a time derivative. (Note the u_t in (1).) However, proving this in space is just as strong as it would be with the time derivative.

Proof. Let $H = \dfrac{t\Gamma(\sqrt{f})}{\sqrt{f \cdot M}}$. Fix $t > 0$. Let $(x^\star, t)$ be a point in $V \times \{t\}$ such that $H(x, t)$ is maximized. All of the following computations are made at the point $(x^\star, t)$. In order to apply the exponential curvature dimension inequality to $\sqrt{f}$, we must have that $\Delta\sqrt{f} < 0$.

If $\Delta\sqrt{f} \geq 0$, then by Lemma 4, we get that $\dfrac{f-u}{2t\sqrt{f}} - \dfrac{\sqrt{M}H}{t} \geq 0$. Thus,

$$\frac{\sqrt{M}H}{t} \leq \frac{f-u}{2t\sqrt{f}} \leq \frac{\sqrt{f}}{2t}$$

which implies that

$$H \leq \frac{\sqrt{f}}{2\sqrt{M}} \leq \frac{1}{2}$$

So we can assume $\Delta\sqrt{f} < 0$, which allows us to use the inequality in (8).
Then, we have that

$$\begin{aligned}(\Delta\sqrt{f})H &\geq \Delta(\sqrt{f}H) \text{ by Lemma 3} \\ &\geq \frac{t}{\sqrt{M}}\left(\frac{2}{n}\left(\Delta\sqrt{f}\right)^2 + 2\Gamma\left(\sqrt{f}, \frac{\Delta f}{2\sqrt{f}}\right)\right) \text{ by (8)}\end{aligned}$$

Thus,

$$\left(\Delta\sqrt{f}\right)H \geq \frac{t}{\sqrt{M}}\left(\frac{2}{n}\left(\frac{f-u}{2t\sqrt{f}} - \frac{\sqrt{M}H}{t}\right)^2 + \frac{1}{t}\Gamma\left(\sqrt{f}\right) - \frac{1}{t}\Gamma\left(\sqrt{f}, \frac{u}{\sqrt{f}}\right)\right)$$

using Lemmas 2, 4, and the fact that Γ is bilinear. Notice that

$$\begin{aligned}
&\frac{1}{t}\Gamma\left(\sqrt{f}\right) - \frac{1}{t}\Gamma\left(\sqrt{f}, \frac{u}{\sqrt{f}}\right) \\
&= \frac{1}{t}\left(\frac{1}{2}\widetilde{\sum}_{y\sim x}\left(\sqrt{f(y)} - \sqrt{f(x)}\right)^2 - \frac{1}{2}\widetilde{\sum}_{y\sim x}\left(\sqrt{f(y)} - \sqrt{f(x)}\right)\left(\frac{u(y)}{\sqrt{f(y)}} - \frac{u(x)}{\sqrt{f(x)}}\right)\right) \\
&\geq -\frac{1}{2t}\widetilde{\sum}_{y\sim x}\left(\sqrt{f(y)} - \sqrt{f(x)}\right)\left(\frac{u(y)}{\sqrt{f(y)}} - \frac{u(x)}{\sqrt{f(x)}}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left(\Delta\sqrt{f}\right)H &\geq \frac{2}{\sqrt{M}nt}\left(\frac{(f-u)^2}{2\sqrt{f}} - \frac{(f-u)\sqrt{M}H}{\sqrt{f}} + MH^2\right) \\
&\qquad + \frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{M}} - \frac{1}{\sqrt{M}}\Gamma\left(\sqrt{f}, \frac{u}{\sqrt{f}}\right) \\
&\geq \frac{2\left(-\sqrt{f}\sqrt{M}H + MH^2\right)}{\sqrt{M}nt} \\
&\qquad - \frac{1}{2\sqrt{M}}\widetilde{\sum_{y\sim x}}\left(\sqrt{f(y)} - \sqrt{f(x)}\right)\left(\frac{u(y)}{\sqrt{f(y)}} - \frac{u(x)}{\sqrt{f(x)}}\right) \\
&\geq \frac{2\left(MH^2 - \sqrt{f}\sqrt{M}H\right)}{\sqrt{M}nt} \\
&\qquad - \frac{1}{2\sqrt{M}}\widetilde{\sum_{y\sim x}}\left(u(x)\left(1 - \sqrt{\frac{f(y)}{f(x)}}\right) + u(y)\left(1 - \sqrt{\frac{f(x)}{f(y)}}\right)\right) \\
&\geq \frac{2\left(MH^2 - \sqrt{f}\sqrt{M}H\right)}{\sqrt{M}nt} - \frac{\sqrt{M}}{2}
\end{aligned}$$

Here, the last inequality—replacing the large averaged sum by $\sqrt{M}$ follows first by noting that one of $\sqrt{f(y)/f(x)}$ or $\sqrt{f(x)/f(y)}$ is at least one. This means that one of the interior terms is non-positive; this subtracted term is non-negative and can be dropped. In the other case, we use the fact that $u \leq M$; this implies that each term in the interior of the averaged sum is at most M. Thus, the average is at most M, and the inequality follows.

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

By Lemma 4, we have that

$$\begin{aligned}\Delta\sqrt{f} &= \frac{f-u}{2t\sqrt{f}} - \frac{\sqrt{M}H}{t}\\ &\le \frac{\sqrt{f}}{2t} - \frac{\sqrt{M}H}{t}\end{aligned}$$

So we have that $(\Delta\sqrt{f})H \le \frac{\sqrt{f}H}{2t} - \frac{\sqrt{M}H^2}{t}$.
This implies that

$$\frac{\sqrt{f}H}{2t} - \frac{\sqrt{M}H^2}{t} \ge \frac{2\sqrt{M}H^2}{nt} - \frac{2\sqrt{f}H}{nt} - \frac{\sqrt{M}}{2}$$

Combining terms, we get

$$\left(\frac{\sqrt{f}}{2t} + \frac{2\sqrt{f}}{nt}\right)H + \frac{\sqrt{M}}{2} \ge \left(\frac{2\sqrt{M}}{nt} + \frac{\sqrt{M}}{t}\right)H^2$$

Multiplying by $t/\sqrt{M}$ yields the inequality

$$\left(\frac{\sqrt{f}}{2\sqrt{M}} + \frac{2\sqrt{f}}{n\sqrt{M}}\right)H + \frac{t}{2} \ge \left(\frac{2}{n} + 1\right)H^2$$

which implies

$$\left(\frac{1}{2} + \frac{2}{n}\right)H + \frac{t}{2} \ge \left(\frac{2}{n} + 1\right)H^2$$

since $\frac{\sqrt{f}}{\sqrt{M}} \le 1$. Thus,

$$\begin{aligned}H^2 &\le \frac{\left(\frac{1}{2} + \frac{2}{n}\right)H}{\left(1 + \frac{2}{n}\right)} + \frac{t}{2\left(1 + \frac{2}{n}\right)}\\ &= C_1 \cdot H + C_2 \cdot t \qquad (9)\end{aligned}$$

for constants $C_1 = C_1(n)$ and $C_2 = C_2(n)$.
If $C_1 H \ge C_2 \cdot t$, then $H^2 \le 2C_1 H$, which implies that $H \le 2C_1$. Thus,

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f \cdot M}} \le \frac{2C_1}{t}$$

If $C_2 \cdot t > C_1 H$, then $H^2 \le 2C_2 t$, so $H \le \sqrt{2C_2 t}$. Therefore,

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f \cdot M}} \le \frac{\sqrt{2C_2 t}}{t} = \frac{\sqrt{2C_2}}{\sqrt{t}}$$

Since

$$C_1 = \frac{\frac{1}{2} + \frac{2}{n}}{1 + \frac{2}{n}}$$

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

and

$$C_2 = \frac{1}{2\left(1 + \frac{2}{n}\right)}$$

we have that

$$\begin{aligned}
\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f} \cdot M} &\leq \frac{2C_1}{t} + \frac{\sqrt{2C_2}}{\sqrt{t}} \\
&= 2 \cdot \left(\frac{\frac{1}{2} + \frac{2}{n}}{1 + \frac{2}{n}}\right) \cdot \frac{1}{t} + \sqrt{\frac{2}{2\left(1 + \frac{2}{n}\right)}} \cdot \frac{1}{\sqrt{t}} \\
&= \frac{1 + \frac{4}{n}}{1 + \frac{2}{n}} \cdot \frac{1}{t} + \sqrt{\frac{1}{1 + \frac{2}{n}}} \cdot \frac{1}{\sqrt{t}} \\
&= \frac{n+4}{n+2} \cdot \frac{1}{t} + \sqrt{\frac{n}{n+2}} \cdot \frac{1}{\sqrt{t}}
\end{aligned}$$

□

Remark. In a typical application of the maximum principle, we maximize over $[0, T]$ and then use information from the time derivative. Here, we do not do this. This is important because of the form of the inequality (9). Because of the dependence of this inequality on the time where the maximum occurs, if the $t^\star$ maximizing the function over all $[0, T]$ is considered, then the result will depend on $t^\star$, giving a bound like $H \leq \frac{\sqrt{2C_2 t^\star}}{t}$. However, since we are able to do the computation at t, this problem does not arise.

Note that the result of Theorem 1 gives us a bound for any time t, but we have two regimes: if t is small, then the first term dominates, and if t is large, then the second term dominates.

Corollary 6

- *For* $0 < t \leq 1 + \frac{6n+16}{n^2+2n}$,

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f} \cdot M} \leq \frac{2(n+4)}{(n+2)} \cdot \frac{1}{t}$$

- *For* $t \geq 1 + \frac{6n+16}{n^2+2n}$,

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f} \cdot M} \leq 2\sqrt{\frac{n}{n+2}} \cdot \frac{1}{\sqrt{t}}$$

Proof. We want to show that if t is small, then $\frac{2C_1}{t}$ is the dominating term in the bound from Theorem 1, and if t is large, then $\frac{\sqrt{2C_2}}{\sqrt{t}}$ dominates. Let $A = 2C_1$ and $B = \sqrt{2C_2}$. We are interested in knowing when $\frac{A}{t} = \frac{B}{\sqrt{t}}$. This is equivalent to $\sqrt{t} = \frac{A}{B}$, so

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

$$t = \frac{A^2}{B^2} = \frac{(2C_1)^2}{(\sqrt{2C_2})^2} = \frac{4C_1^2}{2C_2} = \frac{2\left(\dfrac{\dfrac{1}{2}+\dfrac{2}{n}}{1+\dfrac{2}{n}}\right)^2}{\left(\dfrac{1}{2\left(1+\dfrac{2}{n}\right)}\right)} = \frac{4\left(\dfrac{1}{2}+\dfrac{2}{n}\right)^2}{1+\dfrac{2}{n}}$$

$$= \frac{4\left(\dfrac{1}{4}+\dfrac{2}{n}+\dfrac{4}{n^2}\right)}{1+\dfrac{2}{n}}$$

$$= \frac{n^2+8n+16}{n^2+2n}$$

$$= \frac{n^2+2n+6n+16}{n^2+2n}$$

$$= 1 + \frac{6n+16}{n^2+2n}$$

If $t \leq 1 + \dfrac{6n+16}{n^2+2n}$, then $\dfrac{A}{t} \geq \dfrac{B}{\sqrt{t}}$, so

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f \cdot M}} \leq 2 \cdot \frac{A}{t} = \frac{2(n+4)}{(n+2)} \cdot \frac{1}{t}$$

If $t \geq 1 + \dfrac{6n+16}{n^2+2n}$, then $\dfrac{A}{t} \leq \dfrac{B}{\sqrt{t}}$, so

$$\frac{\Gamma\left(\sqrt{f}\right)}{\sqrt{f \cdot M}} \leq 2 \cdot \frac{B}{\sqrt{t}} = 2\sqrt{\frac{n}{n+2}} \cdot \frac{1}{\sqrt{t}}$$

□

4. Harnack-type inequality

One of the most classical uses of gradient estimates is to prove "Harnack inequalities"; comparison inequalities that multiplicatively compare functions at different points. In this vein, Theorem 3 can be used to compare the PageRank of two vertices in a graph, depending on the distance between them.

The classical form of a Harnack inequality one obtains from the Li–Yau inequality for heat diffusion is the following.

Proposition 7 (**Bauer et al., 2015**). *Suppose G is a graph satisfying $CDE(n, 0)$. Let $T_1 < T_2$ be real numbers, and let $d(x, y)$ denote the distance between $x, y \in V(G)$. If u is a positive solution to the heat equation on G, then*

$$u(x, T_1) \leq u(y, T_2) \left(\frac{T_2}{T_1}\right)^n \exp\left(\frac{4Dd(x, y)^2}{T_2 - T_1}\right)$$

where $D = \max_{v \in V(G)} \deg(v)$.

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

This allows a comparison of heat at different points and different times. In turn, this can be used to detect geometric features of a graph. Delmotte (1999) showed the equivalence of a sufficiently strong Harnack inequality for heat flow to several other conditions, such as volume doubling and satisfying a Poincaré inequality. Horn et al. (2019) complemented the work of Delmotte by showing that curvature lower bounds on graphs can be used to prove these equivalent conditions, along with deriving other geometric consequences from curvature lower bounds.

Using Theorem 1, we are able to relate PageRank at different vertices, but our result is not quite of the right form to be a classical Harnack inequality. The fundamental reason for this is the scaling in Theorem 1. In particular, the dependence on the maximum function value weakens our estimate, and it would be better if we had an f instead of $\sqrt{f \cdot M}$ in the denominator. Since we do not, this makes proving a "Harnack-type" inequality more difficult.

The key step to is to compare the PageRank at adjacent vertices. From now on, we will consider t fixed and write $f(x)$ instead of $f(x, t)$. If a vertex, w, is adjacent to a vertex, z, then we want to lower bound $\sqrt{f(z)}$ by a function only involving $f(w)$. The trick to this is to rewrite $\sqrt{\frac{f(w)}{f(z)}}$ so that we can use Theorem 1 in order to get rid of the "$\sqrt{f(z)}$" in the denominator; in such a way, we get a proper multiplicative comparison between values.

Lemma 5. *Let G be a graph satisfying $CDE(n, 0)$ and let $f = P_\alpha u$ for some seed u. Suppose $0 \le f \le M$ for all $x \in V(G)$. If $w \sim z$ and $t \in [1 + \frac{6n+16}{n^2+2n}, \infty)$, then*

$$\sqrt{\frac{f(w)}{f(z)}} \le \frac{4CD\sqrt{M}}{\sqrt{t}} \cdot \frac{1}{\sqrt{f(w)}} + 2$$

where $D = \max_{v \in V(G)} \deg(v)$ and $C = \sqrt{\frac{n}{n+2}}$.

If $t \in (0, 1 + \frac{6n+16}{n^2+2n})$, then

$$\sqrt{\frac{f(w)}{f(z)}} \le \frac{4CD\sqrt{M}}{t} \cdot \frac{1}{\sqrt{f(w)}} + 2$$

where $C = \frac{2(n+4)}{(n+2)}$.

Proof. We prove this in the case where $t \in [1 + \frac{6n+16}{n^2+2n}, \infty)$; in the case where t is smaller, the proof is identical except for the result of applying the gradient estimate through Corollary 6.

If $\sqrt{f(z)} \ge \frac{1}{2}\sqrt{f(w)}$, then $\sqrt{\frac{f(w)}{f(z)}} \le 2 \le \frac{4CD\sqrt{M}}{\sqrt{t}} \cdot \frac{1}{\sqrt{f(w)}} + 2$

If $\sqrt{f(z)} < \frac{1}{2}\sqrt{f(w)}$, then

$$\begin{aligned}
\sqrt{\frac{f(w)}{f(z)}} &= \frac{\sqrt{f(w)} - \sqrt{f(z)} + \sqrt{f(z)}}{\sqrt{f(z)}} \\
&= \frac{\sqrt{f(w)} - \sqrt{f(z)}}{\sqrt{f(z)}} + 1 \\
&= \frac{D\left(\sqrt{f(w)} - \sqrt{f(z)}\right)^2}{D\sqrt{f(z)}\left(\sqrt{f(w)} - \sqrt{f(z)}\right)} + 1
\end{aligned}$$

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

$$\leq \frac{2CD\sqrt{M}}{\sqrt{t}} \cdot \frac{1}{\sqrt{f(w)} - \sqrt{f(z)}} + 1$$
$$\leq \frac{2CD\sqrt{M}}{\sqrt{t}} \cdot \frac{2}{\sqrt{f(w)}} + 1 \text{ since } \sqrt{f(z)} < \frac{1}{2}\sqrt{f(w)}$$
$$\leq \frac{4CD\sqrt{M}}{\sqrt{t}} \cdot \frac{1}{\sqrt{f(w)}} + 2 \qquad \square$$

We remark that a slight variant of this proof can prove a one-step inequality which looks more akin to the standard inductive one-step inequality for proving a Harnack inequality but is in general weaker when iterated due to the form of the exponent.

Corollary 8. *Let G be a graph satisfying CDE$(n, 0)$ and let $f = P_\alpha u$ for some seed u. Suppose $0 \leq f \leq M$ for all $x \in V(G)$. If $w \sim z$ and $t \in [1 + \frac{6n+16}{n^2+2n}, \infty)$, then*

$$f(z) \geq f(w) \exp\left(-\frac{8CD\sqrt{M}}{\sqrt{t \cdot f(w)}} - 2\right)$$

If $w \sim z$ and $t \in (0, 1 + \frac{6n+16}{n^2+2n})$, then

$$f(z) \geq f(w) \exp\left(-\frac{8CD\sqrt{M}}{t\sqrt{f(w)}} - 2\right)$$

This follows from Lemma 5 using the inequality

$$1 + x \leq e^x$$

Using Lemma 5, we can prove Theorem 2.

Theorem 2. *Let G be a graph satisfying CDE$(n, 0)$ and let $f = P_\alpha u$ for some seed u. Suppose $0 \leq f \leq M$ for all $x \in V(G)$ and $t \in (0, \infty)$. If dist$(x, y) = d$, where $d \geq 2$, then*

$$\frac{1}{\sqrt{f(y)}} \leq 4^{2^d - 2} \cdot \max_{k=0,d} \left\{ \frac{A^{2^k - 1}}{(\sqrt{f(x)})^{2^k}} \right\}$$

where $A = \frac{8CD\sqrt{M}}{\sqrt{t}}$ if $t \in [1 + \frac{6n+16}{n^2+2n}, \infty)$ and $A = \frac{8CD\sqrt{M}}{t}$ if $t \in (0, 1 + \frac{6n+16}{n^2+2n})$

Proof. We proceed by induction on d.

If $d = 2$, then let $z \in V(G)$ such that $x \sim z$ and $z \sim y$. Then by Lemma 5,

$$\frac{1}{\sqrt{f(y)}} \leq \frac{4CD\sqrt{M}}{\sqrt{tf(z)}} + \frac{2}{\sqrt{f(z)}}$$
$$\leq \max\left\{ \frac{A}{f(z)}, \frac{4}{\sqrt{f(z)}} \right\}$$
$$\leq \max\left\{ A \cdot \max\left\{ \frac{A^2}{(f(x))^2}, \frac{4^2}{f(x)} \right\}, 4 \cdot \max\left\{ \frac{A}{f(x)}, \frac{4}{\sqrt{f(x)}} \right\} \right\}$$

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

$$= \max\left\{\frac{A^3}{(f(x))^2}, \frac{4^2 A}{f(x)}, \frac{4^2}{\sqrt{f(x)}}\right\}$$

$$\leq 4^2 \cdot \max\left\{\frac{A^3}{(f(x))^2}, \frac{A}{f(x)}, \frac{1}{\sqrt{f(x)}}\right\}$$

Assume the result holds for $d \geq 2$. Let $x, y \in V(G)$ with $\mathrm{dist}(x, y) = d+1$. Let $z \in V(G)$ such that $z \sim y$ and $\mathrm{dist}(x, z) = d$. Then by Lemma 5, we have that

$$\begin{aligned}
\frac{1}{\sqrt{f(y)}} &\leq \frac{4CD\sqrt{M}}{\sqrt{t}f(z)} + \frac{2}{\sqrt{f(z)}} \\
&\leq \max\left\{\frac{A}{f(z)}, \frac{4}{\sqrt{f(z)}}\right\} \\
&\leq \max\left\{A \cdot 4^{2^{d+1}-2^2} \cdot \max_{k=0,d}\left\{\frac{A^{2^{k+1}-2}}{(\sqrt{f(x)})^{2^{k+1}}}\right\}, 4 \cdot 4^{4^{2^d}-2} \cdot \max_{k=0,d}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\}\right\} \\
&\qquad\qquad \text{by the inductive hypothesis} \\
&= \max\left\{4^{2^{d+1}-2^2} \cdot \max_{k=0,d}\left\{\frac{A^{2^{k+1}-1}}{(\sqrt{f(x)})^{2^{k+1}}}\right\}, 4^{2^d-1} \cdot \max_{k=0,d}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\}\right\} \\
&\leq 4^{2^{d+1}-2} \cdot \max\left\{\max_{k=0,d+1}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\}, \max_{k=0,d}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\}\right\} \\
&= 4^{2^{d+1}-2} \cdot \max_{k=0,d+1}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\}
\end{aligned}$$

where the last step follows since

$$\max_{k=0,d}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\} \leq \max_{k=0,d+1}\left\{\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}}\right\}$$

To see this, notice that $\frac{A^{2^k-1}}{(\sqrt{f(x)})^{2^k}} = \left(\frac{A}{\sqrt{f(x)}}\right)^{2k} \cdot \frac{1}{A}$. This function of k is either increasing or decreasing, so the maximum over the interval $0 \leq k \leq d+1$ is achieved at either $k = 0$ or $k = d+1$. □

5. Conclusions, applications, and future work

In this paper, we investigated PageRank as a diffusion, using recently developed notions of discrete curvature. These results, while theoretical (and in some cases not as strong as would be desired due to the dependence on the maximum value "M" in the gradient estimate), show that curvature aspects of graphs can be used to understand relative importance in networks—at least when ranking is based on random walk-based diffusions.

Regarding these points, we highlight the following:

- Curvature is a local property—based only on second neighborhood conditions. An upshot of this is that it can be certified quickly. While the work here focuses on situations where the

https://doi.org/10.1017/nws.2020.34 Published online by Cambridge University Press

entire graph is non-negatively curved for simplicity, work in Bauer et al. (2015) and Horn (2019) shows that these methods can be used when only parts of the graph satisfy such a radius using cutoff functions. In principle, these yield algorithms that are linear, either in the size of the graph or even in a considered portion of the graph, verifying curvature conditions and elucidating PageRank's diffusion in bounded degree graphs.

- The influence of the jumping constant on PageRank has been important for certain algorithms or in the analysis of certain stochastic processes (such as in Andersen et al., 2008; Chung et al., 2014, 2009). On the other hand, for its use in web search, the jumping constant was originally picked rather arbitrarily, or at least experimentally (see, e.g. Page et al., 1998). Due to the importance of the PageRank vector (and particularly, how the PageRank vector "falls off") in the application papers cited above, a study of this phenomenon seems important for the analysis of complex networks and stochastic processes on them. In particular, these applications produce the best results when there is a sharp drop in the PageRank vector, and whose existence depends on a lack of regularity. This paper, then, should be seen as part of this thrust by investigating local conditions that imply a well-controlled smoothing of the PageRank vector.
- There are several interesting areas for improvement here: the non-ideal scaling in Theorem 1 leads to a weaker than ideal result in Lemma 5. While Lemma 5 seems a reasonable result, when iterated it quickly loses power (unlike the Harnack inequality from a "properly scaled" gradient estimate like in Proposition 7). While a "properly scaled" Theorem 1 may not even be true, we suspect the scaling can be improved. An interesting question is whether a true "Hamilton type" gradient estimate is true: Is $\Gamma(\sqrt{f})/f \leq C\log(M/f)t^{-1}$? Note that the addition of the logarithmic term damages a Harnack inequality, but the results obtainable from this are far better than we obtain. Also, a version including the time derivative term is desirable as well.

Acknowledgments. Horn's work was partially supported by Simons Collaboration grant #525309. Portions of the work were completed while Horn was enjoying the hospitality of the Yau Mathematical Sciences Center, Tsinghua University, Beijing China.

Conflict of interest. The authors have nothing to disclose.

References

Alon, N. (1986). Eigenvalues and expanders. *Combinatorica*, *6*(2), 83–96. Theory of computing (Singer Island, Fla., 1984).

Andersen, R., Chung, F., & Lang, K. (2008). Local partitioning for directed graphs using PageRank. *Internet Mathematics*, *5*(1–2), 3–22.

Bauer, F., Horn, P., Lin, Y., Lippner, G., Mangoubi, D., & Yau, S.-T. (2015). Li-Yau inequality on graphs. *Journal of Differential Geometry*, *99*(3), 359–405.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*(1), 107–117. Proceedings of the Seventh International World Wide Web Conference.

Cheeger, J. (1970). A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)* (pp. 195–199). Princeton, NJ: Princeton University Press.

Chung, F. (2007). The heat kernel as the PageRank of a graph. *Proceedings of the National Academy of Sciences*, *104*(50), 19735–19740.

Chung, F. (2011). PageRank as a discrete Green's function. In *Geometry and analysis. No. 1*, vol. 17, Advanced Lectures in Mathematics (ALM) (pp. 285–302). Somerville, MA: International Press

Chung, F., Horn, P., & Hughes, J. (2014). Multicommodity allocation for dynamic demands using PageRank vectors. *Internet Mathematics*, *10*(1–2), 49–65.

Chung, F., Horn, P., & Tsiatas, A. (2009). Distributing antidote using PageRank vectors. *Internet Mathematics*, *6*(2), 237–254 (2010).

Chung, F., Lin, Y., & Yau, S.-T. (2014). Harnack inequalities for graphs with non-negative Ricci curvature. *Journal of Mathematical Analysis and Applications*, *415*(1), 25–32.

Chung, F. R. K. (1997). *Spectral graph theory*, vol. 92, CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI.

Delmotte, T. (1999). Parabolic Harnack inequality and estimates of Markov chains on graphs. *Revista Matemática Iberoamericana*, *15*(1), 181–232.

Dodziuk, J. (1984). Difference equations, isoperimetric inequality and transience of certain random walks. *Transactions of the American Mathematical Society*, *284*(2), 787–794.

Hamilton, R. S. (1993). A matrix Harnack estimate for the heat equation. *Communications in Analysis and Geometry*, *1*(1), 113–126.

Horn, P. (2019) A spacial gradient estimate for solutions to the heat equation on graphs. *SIAM Journal on Discrete Mathematics*, *33*(2), 958–975.

Horn, P., Lin, Y., Liu, S., & Yau, S.-T. (2019). Volume doubling, Poincaré inequality and Gaussian heat kernel estimate for non-negatively curved graphs. *Journal f́ur die Reine und Angewandte Mathematik*, *757*, 89–130.

Horn, P., & Nelsen, L. M. (2019). A gradient estimate for PageRank. In H. Cherifi, S. Gaito, J. Mendes, E. Moro, & L. Rocha (Eds.) *Complex networks and their applications VIII*, vol. 881, Studies in Computational Intelligence (pp. 15–26). Cham: Springer.

Jeh, G., & Widom, J. (2003). Scaling personalized web search. In *Proceedings of the 12th World Wide Web Conference (WWW)* (pp. 271–279).

Li, P. & Yau, S.-T. (1986). On the parabolic kernel of the Schrödinger operator. *Acta Mathematica*, *156*(3–4), 153–201.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. In *Proceedings of the 7th international World wide web conference, Brisbane, Australia* (pp. 161–172).

Cite this article: Horn P. and Nelsen L. M. (2021). Gradient and Harnack-type estimates for PageRank. *Network Science* **9**, S4–S22. https://doi.org/10.1017/nws.2020.34