PageRank and Random Walks

Ian Chen

June 18, 2026

1 Introduction

Last week, we showed that conductance optimization is closely linked to spectral clustering via Cheeger's inequality. There, we found that good clusters are formed by sweeps of the Fiedler vector. Here, we suppose that we are given a seed, i.e. a vertex within the target cluster, and our goal is to grow the seed into the target cluster. Towards this, the influence measure we discuss is known as Personalized PageRank, and we analyze its properties.

2 Nibbling and Partitioning

The problem of interest last week was to find a set S that minimizing $\Phi(S)$, where Φ denotes the conductance

$$\Phi(S) = \frac{|\partial S|}{\min(\mathrm{vol}(S), \mathrm{vol}(V - S))}$$

We will describe a near-linear time approximation built solely on *local computation*, i.e. small subgraphs at a time. This has many practical advantages, e.g. memory locality and horizontal scaling to distributed systems.
Theorem. *If* $\mathrm{vol}(C) \leq \mathrm{vol}(V)/2$*, then there is some* $C' \subset C$ *with* $\mathrm{vol}(C') \geq \mathrm{vol}(C)/2$*, where for seeds* $v \in C'$*, then* PR-NIBBLE(v, α, ϵ) *produces a cut of conductance* $\tilde{O}(\sqrt{\alpha})$ *in time* $\tilde{O}(1/\epsilon\alpha)$ *for* $\alpha = \tilde{O}(\Phi(C)), \epsilon = \tilde{O}(\frac{1}{\mathrm{vol}(C)})$.

Given this, it now is a matter of binary search in order to find a cut of a given threshold with high probability:

1. Binary search on the size 2^b of S
2. Sample $\tilde{O}(m/2^b)$ seeds with probability proportional to its degree
3. Run PR-NIBBLE; stop if found a cut of $\tilde{O}(\sqrt{\varphi})$

3 Personalized PageRank

On an undirected graph, the walk matrix W is given by AD^{-1}, i.e. each vertex picks uniformly from its neighbors. Lazy walk matrices, $\tilde{W} = \frac{1}{2}(I + W)$ guarantees a-periodicity. To form an ergodic chain, the canonical PageRank introduces a teleportation probability α, so that it considers the matrix $\hat{W} = \alpha \mathbf{1}_n \mathbf{1}_n^T + (1 - \alpha)\tilde{W}$.

Personalized PageRank (PPR) can be seen as a generalization: $\hat{W}_s = \alpha \boldsymbol{s} \mathbf{1}_n^T + (1 - \alpha)\tilde{W}$. In particular, we will focus on $\boldsymbol{s} = \mathcal{X}_v$, teleporting back to a starting vertex, so that the stationary distribution ${\pi_\alpha}^{(\infty)}$ satisfies

$${\pi_\alpha}^{(\infty)} = \alpha \mathcal{X}_v + (1 - \alpha)\tilde{W}{\pi_\alpha}^{(\infty)} = \alpha(I - (1 - \alpha)\tilde{W})^{-1}\mathcal{X}_v = \alpha \sum_{t=0}^{\infty} (1 - \alpha)^t \tilde{W}^t \mathcal{X}_v$$

We denote $\pi = {\pi_\alpha}^{(\infty)}$ as the PPR vector (wrt v).

4 Analysis

We will interchange back and forth between mass being on nodes and being on edges, with $\pi(e) = \frac{\pi(u)}{\deg(u)}$. as our goal will be to perform a sweep-cut, we study the Lovasz-Simonovitz curve, $\boldsymbol{p}\{k\}$ the sum of the k-largest entries in $\boldsymbol{p}$, extending to $[0, 2m]$ by linear interpolation. As $\boldsymbol{p}\{k\}$ is concave, it lies above the line from $(0,0)$ to $(2m,1)$.

Recall that the stationary distribution $\tilde{W}$ is $\psi_V = \frac{\boldsymbol{d}}{\mathrm{vol}(V)}$, since $AD^{-1}\frac{\boldsymbol{d}}{\mathrm{vol}(V)} = A\frac{\mathbf{1}}{\mathrm{vol}(V)} = \frac{\boldsymbol{d}}{\mathrm{vol}(V)}$. In particular, $\psi\{x\} = \frac{x}{2m}$. Whenever $\pi(S)$ deviates from $\psi(S)$, we expect a small conductance cut around S. For example, given a $(c>1)$-edge-expander, i.e. every $|N(S)| \geq c|S|$, then $\pi(S)$ will hug $\psi(S)$ closely for all S. In particular,

Theorem 1. *If $\pi(S) - \psi(S) > \alpha t + \sqrt{\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))}(1 - \frac{\varphi^2}{8})^t$, then $\Phi(S_j) \leq \varphi$ for some sweep cut S_j on π.*

Proof. Considering how the probability mass moves after a step of W,

$$(\tilde{W}\pi)(S) = \frac{1}{2}p(E(S,V)) + \frac{1}{2}p(E(V,S)) = \frac{1}{2}p(E(S)) + \frac{1}{2}p(E(S,V) \cup E(V,S))$$

Now, as π is the unique stationary distribution of the lazy random walk with teleportation,

$$\pi(S) = \alpha + \frac{1-\alpha}{2}(\pi(E(S)) + \pi(E(S,V) \cup E(V,S)))$$

With the Lovasz-Simonovitz curve being concave by construction, and moreover $\pi\{\mathrm{vol}(S)\} \geq \pi(S)$, then

$$\begin{aligned}
\pi\{\mathrm{vol}(S)\} &\leq \alpha + \frac{1}{2}(\pi\{\mathrm{vol}(S) - |\partial S|\} + \pi\{\mathrm{vol}(S) + |\partial S|\}) \\
&= \alpha + \frac{1}{2}(\pi\{\mathrm{vol}(S)(1-\varphi)\} + \pi\{\mathrm{vol}(S)(1+\varphi)\})
\end{aligned}$$

for any $\varphi \leq \Phi(S)$ by concavity. Let $f_t(x) = \alpha t + \sqrt{\min(x, 2m-x)}(1 - \frac{\varphi^2}{8})^t$. We will prove that if there is no sweep cut S_j with $\Phi(S_j) < \varphi$ and $\pi(S_j) - \psi(S_j) > f_t(\mathrm{vol}(S_j))$, then $\pi\{x\} - \frac{x}{2m} \leq f_t(x)$.

Indeed, for $t = 0$, then as the slope of $p\{x\}$ is bounded by 1, then

$$\pi\{x\} - \frac{x}{2m} \leq \min(1, \min(x, 2m-x)) \leq \sqrt{\min(x, 2m-x)}$$

Now, for $t > 1$, we need to show $\pi\{x_j\} - \frac{x_j}{2m} \leq f_t(x_j)$ for all $x_j = \mathrm{vol}(S_j)$ (as piecewise linear and concave). Fix j, and assume that $\Phi(S_j) \geq \varphi$. Then, for the case $x \leq m$,

$$\begin{aligned}
p\{x_j\} &\leq \alpha + \frac{1}{2}(\pi\{x_j(1-\varphi)\} + \pi\{x_j(1+\varphi)\}) \\
&\leq \alpha + \frac{1}{2}\left(f_{t-1}(x_j(1-\varphi)) + f_{t-1}(x_j(1+\varphi)) + \frac{x_j(1-\varphi) + x_j(1+\varphi)}{2m}\right) \\
&\leq \frac{x_j}{2m} + \alpha + \alpha(t-1) + \frac{1}{2}\left(\sqrt{x_j(1-\varphi)} + \sqrt{x_j(1+\varphi)}\right)\left(1 - \frac{\varphi^2}{8}\right)^{t-1}
\end{aligned}$$

and we win by Taylor expanding $\frac{1}{2}\left(\sqrt{x(1-\varphi)} + \sqrt{x(1+\varphi)}\right) \leq \sqrt{x}\left(1 - \frac{\varphi^2}{8}\right)$. The case $x \geq m$ is similar. $\square$

What's left is design an algorithm to exploit the above, namely addressing approximations of π, as well as arguing that good seeds are plentiful, producing our desired output size. $\pi_{\mathcal{X}_v - \boldsymbol{r}}$ is an ϵ-approximate PageRank vector if $\boldsymbol{r}$ is non-negative and satisfies $r(u \leq \epsilon \deg(u)$ for all u. Indeed, given such approximate vectors,

$$\pi_{\mathcal{X}_v}\{\mathrm{vol}(S)\} \geq \pi_{\mathcal{X}_v - \boldsymbol{r}}\{\mathrm{vol}(S)\} \geq \pi_{\mathcal{X}_v}\{\mathrm{vol}\, S\} - \boldsymbol{r}\{\mathrm{vol}(S)\} \geq \pi_{\mathcal{X}_v}\{\mathrm{vol}(S)\} - \epsilon\,\mathrm{vol}(S)$$

Our approximation algorithm starts with $\boldsymbol{r}^{(0)} = \mathcal{X}_v$, initializes $\boldsymbol{p}^{(0)} = 0$, and iteratively pushes probability out of $\boldsymbol{r}$. Namely, in each step we find a vertex u with $r(u) > \epsilon d(u)$, then update

$$\boldsymbol{p}^{(t+1)} = \boldsymbol{p}^{(t)} + \alpha r^{(t)}(u)\mathcal{X}_u, \qquad \boldsymbol{r}^{(t+1)} = \boldsymbol{r}^{(t)} - r^{(t)}(u)\mathcal{X}_u + (1-\alpha)r^{(t)}(u)\tilde{W}\mathcal{X}_u$$

Indeed, by the triangle inequality, we decrease the L1 mass on $\boldsymbol{r}$ by at least $\alpha r(u)$, and thus converges in at most $\frac{1}{\epsilon\alpha}$ iterations. Moreover, that also bounds the mass pushed onto $\boldsymbol{p}$, i.e. $\mathrm{vol}(\mathrm{support}(\boldsymbol{p})) \leq \frac{1}{\epsilon\alpha}$.

Finally, we show the existence of good seeds, i.e. those with highly concentrated PPR scores.

Theorem. *For $C \subset V$, there exists $C' \subset C$ with $\mathrm{vol}(C') \geq \frac{\mathrm{vol}(C)}{2}$, such that when $v \in C'$, then $\pi_{\mathcal{X}_v}(C) \geq 1 - \frac{\Phi(C)}{\alpha}$.*

Proof. Via linearity of the PageRank operator, it is sufficient to show that $\pi_{\psi_C}(\bar{C}) \leq \Phi(C)\frac{1-\alpha}{2\alpha}$, since this bounds the expected number of bad vertices from sampling according to degree, by Markov's inequality. Indeed, by considering the movement of probability from π_{ψ_C} to $\tilde{W}\pi_{\psi_C}$,

$$\begin{aligned}\pi_{\psi_C}(\bar{C}) &= (\alpha\psi_C + (1-\alpha)W\pi_{\psi_C})(\bar{C}) = (1-\alpha)(W\pi_{\psi_C})(\bar{C}) \\ &\leq (1-\alpha)(\pi_{\psi_C})(\bar{C}) + \frac{1-\alpha}{2}\pi_{\psi_C}\{|\partial C|\}\end{aligned}$$

which is what we needed to show. □

Putting our results together, for $\tilde{\pi} = \pi_{\mathcal{X}_v - \boldsymbol{r}}$ an ϵ-approximate PageRank vector, then

$$\tilde{\pi}(C) \geq 1 - \frac{\Phi(C)}{\alpha} - \epsilon\,\mathrm{vol}(C) \implies \Phi(\tilde{p}) = \tilde{O}(\sqrt{\alpha})$$

An additional check of that $\tilde{\pi}\left\{2^b\right\} - \tilde{\pi}\left\{2^{b-1}\right\} > O(\frac{1}{\log m})$ leads to the promised output size constraints.

5 A Spectral View

$$\begin{aligned}\frac{\boldsymbol{\pi}}{\beta} &= \left(\frac{\beta}{\alpha}(I - (1-\alpha)\tilde{W})\right)^{-1}\boldsymbol{s} = \left(\beta I + (I + D^{-1}A)\right)^{-1}\boldsymbol{s} \\ &= \left(D^{-1/2}(\beta I + I + D^{-1/2}AD^{1/2})D^{-1/2}\right)^{-1}\boldsymbol{s} = D^{-1/2}\mathcal{G}_\beta D^{1/2}\boldsymbol{s}\end{aligned}$$

Thus, we are looking at the symmetrized Green's function acting on $\boldsymbol{s}$. The spectral decomposition then shows

$$\mathcal{G}_\beta = \sum_{i=1}^{n-1}\frac{1}{\lambda_i + \beta}\boldsymbol{v}_i\boldsymbol{v}_i^T$$

then we see that by setting $-\beta$ close to the second eigenvalue, then this amplifies the signal of our Fiedler vector. Indeed, this gives us an algebraic proof of the approximation ratio bypassing the Lovasz-Simonovitz curve. Moreover, a proof of Cheeger's inequality. TODO

6 Bibliographic Notes

This material is primarily derived from Andersen, Chung, and Lang [1], using the Lovasz-Simonovitz curve [3]. The interpretation of the PageRank operator as a low-pass filter is discussed in Chung and Zhao [2], as well as other applications of the β-normalized discrete Green's function. Finally, the spectral proof of the same result is derived from Spieman [4], with the notation for the L-S curve borrowed there.

References

[1] Reid Andersen, Fan Chung, and Kevin Lang. "Using PageRank to Locally Partition a Graph". In: *Internet Mathematics* 4.1 (Jan. 2007), pp. 35–64. ISSN: 1944-9488. DOI: 10.1080/15427951.2007.10129139. URL: http://dx.doi.org/10.1080/15427951.2007.10129139.

[2] Fan Chung and Wenbo Zhao. "PageRank and Random Walks on Graphs". In: *Fete of Combinatorics and Computer Science*. Springer Berlin Heidelberg, 2010, pp. 43–62. ISBN: 9783642135804. DOI: 10.1007/978-3-642-13580-4_3. URL: http://dx.doi.org/10.1007/978-3-642-13580-4_3.

[3] L. Lovasz and M. Simonovits. "Random walks in a convex body and an improved volume algorithm". In: *Random Structures & Algorithms* 4.4 (Jan. 1993), pp. 359–412. ISSN: 1098-2418. DOI: 10.1002/rsa.3240040402. URL: http://dx.doi.org/10.1002/rsa.3240040402.

[4] Daniel Spieman. *Spectral and Algebraic Graph Theory*. Incomplete draft. 2025. URL: http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf.