PageRank and Random Walks

Ian Chen

University of Illinois Urbana-Champaign

June 18, 2026

Nibbling towards a Partition

- Our goal is to find a set S, $2\,\mathrm{vol}(S) \leq \mathrm{vol}(V)$, minimizing $\Phi(S) = \frac{|\partial S|}{\mathrm{vol}(S)}$.
- Spectral partitioning finds an $O(\sqrt{\Phi_{\mathrm{OPT}}})$ approximation (Cheeger)
- Let $v \in C$ be a seed, $\alpha = \tilde{O}(\Phi(C))$, $\epsilon = \tilde{O}(\frac{1}{\mathrm{vol}(C)})$
- PR-Nibble(v, α, ϵ) finds an $\tilde{O}(\sqrt{\alpha})$-cut in time $\tilde{O}(1/\epsilon\alpha)$
  - Linear in the size of the output (for fixed conductance)
  - Sampling seeds, binary searching on ϵ and α, finds approximately optimal global cut

Personalized PageRank

- On an undirected graph, the walk matrix $W = AD^{-1}$
- Lazy walk matrix $\tilde{W} = \frac{1}{2}(I + AD^{-1})$ is aperiodic
- Stationary distribution is $\psi = \frac{\mathbf{d}}{\mathrm{vol}(V)}$
- Teleportation α with starting $\mathbf{s}$, $\hat{W} = \alpha \mathbf{s} \mathbf{1}^T + (1-\alpha)\tilde{W}$ is ergodic
- Let $\mathbf{s} = \mathcal{X}_v$; stationary distribution measures locality of u wrt v

$$\begin{aligned}\pi_{\alpha,\mathbf{s}} &= \alpha \mathbf{s} + (1-\alpha)\tilde{W}\pi_{\alpha,\mathbf{s}} = \alpha(I - (1-\alpha)\tilde{W})^{-1}\mathbf{s} \\ &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \tilde{W}^t \mathbf{s}\end{aligned}$$

Random Walks and Cuts

- Low conductance sets have small boundaries
- Low conductance sets have higher than expected PPR mass

Theorem

If $\pi(S) - \psi(S) > \alpha t + \sqrt{\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))}(1 - \frac{\phi^2}{8})^t$*, then* $\exists S_j^\pi$ *sweep s.t.*

$$\Phi(S_j^\pi) \leq \phi$$

- Sample vertices v, estimate PPR, and compare conductance for sweep cuts.

Analysis

- Associate probability on the edges, $\mathbf{p}(uv) = \frac{\mathbf{p}(u)}{\deg(u)}$, e.g. $\psi(e) \equiv \frac{1}{2m}$
- Let $\mathbf{p}\{\,k\,\}$ be the mass on the k largest entries of $\mathbf{p}$, e.g. $\psi\,\{\,k\,\} = \frac{k}{2m}$
- Lovasz-Simonovitz curve extends $\mathbf{p}\,\{\,k\,\}$ to $[0, 2m]$ by linear interpolation
  - Concave, piecewise linear
  - $\mathbf{p}\,\{\,0\,\} = 0$, $\mathbf{p}\{\,2m\,\} = 1$
  - $\mathbf{p}\,\{\,x\,\} \geq \psi\,\{\,x\,\}$
- We are bounding the convergence of a random walk by a *curve*

Analysis (cont.)

- Denote $f_t(x)$ as $\alpha t + \sqrt{\min(x, 2m - x)}(1 - \frac{\phi^2}{8})^t$

Theorem

If $\nexists S_j$ s.t. $\Phi(S_j) < \phi$ and $\pi(S_j) - \psi(S_j) > f_t(\mathrm{vol}(S_j))$, then $\pi\{x\} - \frac{x}{2m} \leq f_t(x)$.

- We prove this for $x \leq m$; other case is similar
- Base case: $t = 0$, follows by having slope bounded by 1

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

Analysis (cont.)

- Study the distribution of mass after one step of $\tilde{W}$

$$
\begin{aligned}
(\tilde{W}\pi)(S) &= \frac{1}{2}\pi(E(S,V)) + \frac{1}{2}\pi(E(V,S)) \\
&= \frac{1}{2}\pi(E(S,S)) + \frac{1}{2}\pi(E(S,V) \cup E(V,S))
\end{aligned}
$$

- $|E(S)| = \mathrm{vol}(S) - |\partial S| = \mathrm{vol}(S)(1-\phi)$
- $|E(S,V) \cup E(V,S)| = \mathrm{vol}(S) + |\partial S| = \mathrm{vol}(S)(1+\phi)$
- We constructed π to be stationary wrt the PageRank walk:

$$
\pi\left\{\,\mathrm{vol}(S)\,\right\} \leq \alpha + \frac{1}{2}\left(\pi\left\{\,\mathrm{vol}(S)(1-\phi)\,\right\} + \pi\left\{\,\mathrm{vol}(S)(1+\phi)\,\right\}\right)
$$

Analysis (cont.)

- Plugging in the inductive hypothesis, then

$$\begin{aligned}\pi\left\{\,\mathrm{vol}(S)\,\right\} &\le \alpha + (t-1)\alpha + \frac{\sqrt{\mathrm{vol}(S)}}{2}(\sqrt{1-\phi} + \sqrt{1+\phi})(1-\frac{\phi^2}{8})^{t-1} \\ &\le \alpha t + \sqrt{\mathrm{vol}(S)}(1-\frac{\phi^2}{8})^{t}\end{aligned}$$

Remarks

- Lovasz-Simonovitz curve used for mixing of $\tilde{W}$ (convex volume est.)
- We showed that $\pi\left\{\,\mathrm{vol}(S)\,\right\}$ is locally smoothed, by the conductance
- If $\mathcal{L} = D^{-1/2} L D^{-1/2} = \sum_{i=0}^{n-1} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$,
  - The Green's function $\mathcal{G} = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^T$
  - Power iteration amplifies the signals with small eigenvalues
  - PPR normalizes $\beta I + \mathcal{L}$ (shift-inverse method) so that

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

- The PPR vector we compute is given by (for $\beta = (2\alpha)/(1-\alpha)$):

$$\frac{\pi_{\alpha,\mathbf{s}}}{\beta} = D^{-1/2} \mathcal{G}_\beta \mathcal{D}^{1/2} \mathbf{s}$$