Dirichlet Problems, Random walks, and Page Rank: From Graphs to Complexes

P. Horn* A. Jadbabaie** G. Lippner* M. Zargham†

*Department of Mathematics, Harvard University

**Electrical and Systems Engineering, Computer and Information Science Operations and Information Management, University of Pennsylvania

†Electrical and Systems Engineering, University of Pennsylvania

Overview

Overview

Motivation: Discrete potential theory

Discrete potential theory has connections to

- random walks (a harmonic function's expected value is preserved during a random walk)
- electric networks (harmonic functions describe voltages in a network)
- PageRank algorithm for web ranking
- consensus problems
- spectral clustering
- spreading processes (e.g. disease spreading)
- network tomography (through boundary value problems)

A key ingredient of this theory is the discrete Laplace operator, or the graph Laplacian

The Laplacian on a graph

- $G(V,E)$ simple graph.
- The Laplace operator Δ acts on functions $f : V \to \mathbb{R}$ by the formula
$$\Delta f(x) = \sum_{y \sim x} f(x) - f(y)$$
- Discrete analogue of the classical Laplace operator $\partial_{xx} + \partial_{yy}$ in $\mathbb{R}^2$.

This motivates the definition of harmonic functions

Definition

f is called harmonic on the set $W \subset V$ if $\Delta f(x) = 0$ for every $x \in W$. The set $S = V \setminus W$ is the boundary.

Simplicial complexes

- Graphs encode binary relations, but why stop there? higher order relations can be included as well
- Such relations are best modeled on simplicial complexes, the higher dimensional analogues of graphs.
- Edges are generalized to triangles, tetrahedra, and general k simplicies
- We get a more faithful discretization, almost like including higher order terms in a "combinatorial Taylor expansion"
- Simplicial complexes are collection of k-simplicies that are closed under inclusion of faces,
- Let X be a simplicial complex and let F_k denote the collection of k-dimensional faces.
- A k-dimensional face $\sigma \in F_k$ has $k+1$ vertices and generalizes the notion of a vertex and edge

$$\sigma = [v_0, v_1, \ldots, v_k].$$

The higher order Laplace operator I

What is the correct analogue of the Laplace operator for simplicial complexes?

- The operator acts on "alternating" functions defined on oriented k-faces. A function is alternating if $f(\bar{\sigma}) = -f(\sigma)$.
- $C^k(X)$ denotes the space of such functions.
- There is a natural boundary map $\partial_k : C^k(X) \to C^{k-1}(X)$ coming from algebraic topology.
- It is defined by

$$\partial_k f([v_0, \dots, v_k] = \sum_{i=0}^{k+1} (-1)^i f([v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_k]),$$

  and called the boundary map.
- generalizes the node-edge incidence matrix of a graph

The higher order Laplace operator II

- It's adjoint is $\partial_k^* : C^{k-1}(X) \to C^k(X)$.
- Easy to check: $\partial_k \circ \partial_{k+1} = 0$.
- The k-Laplacian acting on $C^k(X)$ is defined by

$$\Delta_k = \partial_{k+1}\partial_{k+1}^* + \partial_k^*\partial_k$$

.

Remark

For $k = 0$ we get back the original graph Laplacian. For $k = 1$ we get the Helmholtzian

Definition

f is called harmonic on the set $W \subset X$ if $\Delta f(x) = 0$ for every $x \in W$. The set $S = X \setminus W$ is the boundary.

- On graphs, harmonic functions are piece-wise constant (on each connected component), representing the zeroth

Hodge decomposition

$$C^{k-1}(X) \xrightarrow{\partial_k^*} C^k(X) \xrightarrow{\partial_{k+1}^*} C^{k+1}(X)$$

$$C^{k-1}(X) \xleftarrow{\partial_k} C^k(X) \xleftarrow{\partial_{k+1}} C^{k+1}(X)$$

The following are well-known facts about the Laplace operator:

- $\Delta_k f = 0 \iff \partial_k f = 0$ and $\partial_{k+1}^* f = 0$
- $C^k(X) = \ker \Delta_k \oplus \mathrm{Im}(\partial_{k+1}) \oplus \mathrm{Im}(\partial_k^*)$ is an orthogonal decomposition called the Hodge decomposition.
- $\mathrm{Im}(\partial_k^*)$ is called the gradient part, $\mathrm{Im}(\partial_{k+1})$ is called the curl part, and $\ker \Delta_k$ is called the harmonic part.
- $\dim \ker \Delta_k = \dim H^k(X)$ is the kth Betti-number of X and corresponds to the (co)homology.
- This is the discrete analog of the Calculus result that says curl-free vector fields are gradient of a scalar field (no harmonic part).

Overview

Dirichlet problems I

- Applications of the Laplace operator often involve finding harmonic functions with certain boundary conditions (Inverse problems, tomography, leader-follower flocking, localization,...)
- Reminder: denote F_k the set of k-faces. Fix $S \subset F_k$ that will be called the *boundary*, and $F_k \setminus S$ the *interior*.

Definition (Dirichlet problem)

Given f_0 on the boundary, find extension f that is harmonic in the interior.

Definition (Inhomogeneous Dirichlet problem)

Given f_0 on the boundary and g_0 in the interior. Find extension f such that $\Delta_k(f) = g_0$ in the interior.

- On a graph (i.e. $k = 0$) these always have a unique solution.

Dirichlet problems II

- This is not the case for higher dimensions.
- On the other hand there is a new type of problem that has solutions only for $k > 0$.

Definition (Strong Dirichlet problem)

Given f_0 on the boundary, find extension f that is harmonic on the whole space.

- It is important to understand, how does the boundary set S affect the solutions to these Dirichlet problems.
- As a first step we can prove:

Theorem

The Dirichlet problem can be solved for any boundary set S.

Types of boundary sets

- We classify possible boundary sets according to the solvability of the Dirichlet problems:
- The boundary $S \subset F_k$ is
  - ample: if the solution to the Dirichlet problem is unique.
  - modest: if the Strong Dirichlet problem has a solution.

Remark

Bigger boundaries are more likely to be ample, smaller boundaries are more likely to be modest.

Results

Theorem

The Inhomogeneous Dirichlet problem can be solved exactly when S is ample.

Theorem

The smallest possible ample sets are also modest. For such an S both the inhomogeneous and the strong Dirichlet problems can be solved, and the solution is unique.

Ample edge sets

- It is important to characterize ample sets for applications.
- Imagine we can only observe a set S. We know that a harmonic function f is zero on S.
- How can we be sure that f is identically zero?
- The answer is: we can be sure exactly if S is ample!

Definition

X is surface like if every edge is incident to at most 2 faces.

Theorem

Suppose X is surface like and $n = \dim H^1(X)$. If S is not ample then S can be separated from X by cutting at most $6n$ edges.

Random Walks and Harmonic functions

- Let $X_0, X_1, \ldots$ be a simple random walk on a graph.
- Harmonic functions are fixed in expectation:

$$f(X_i) = E(f(X_{i+1})).$$

- This is also true for harmonic functions with boundary conditions - as long as the walk is in the interior.

Solving Dirichlet problems on graphs:

Start a random walk in $X_0 = v$ and wait until it hits the boundary. Let T denote the time of hitting. Then define

$$f(v) = E(f(X_T)) = E(f_0(X_T)).$$

This is the unique solution of Dirichlet problem with boundary condition f_0. $\square$

Random walks on complexes?

- Define a random walk $W_0, W_1, W_2, \ldots$ on the space of oriented k-faces.
- The transition from W_i to W_{i+1} can be of three types:
  1. reverse orientation
  2. k-face $\to$ $(k+1)$-face $\to$ k-face
  3. k-face $\to$ $(k-1)$-face $\to$ k-face
- Carefully choosing transition probabilities of (1), (2), and (3) we can get

$$f(W_i) = k \cdot E(f(W_{i+1})).$$

- Then the solution to the Dirichlet problem is

$$f(W_0) := E(k^T f_0(W_T))$$

where T is the first time the walk hits the boundary.

Overview

PageRank Operator on Graphs

- Personalized PageRank: Geometric sum of random walks starting at seed $\mathbf{s}$

$$\mathrm{pr}(\alpha, \mathbf{s}) = \mathbf{s}(1-\alpha)\sum_{t=0}^{\infty}(\alpha W)^t \qquad W = D^{-1}A$$

- $\mathrm{pr}(\alpha, \mathbf{s})$ entries: Gives measure of *relative importance* of vertices with respect to seed.
- Classical PageRank: $\mathbf{s} = \frac{1}{n}\mathbf{1}$. Entries give measure of *global importance*.

Question: How to rank importance of *edges/faces* in simplicial complex?
Find analogue of PageRank?

Rethinking PageRank: Green's functions

- Green's functions: Inverse of Laplacian Δ or normalized Laplacian $\mathcal{L} = I - D^{-1}A$ on space orthogonal to null space of Δ:

$$\Delta = \sum_{i=0}^{n-1} \lambda_i \phi_i^* \phi_i \qquad \mathcal{G} = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \phi_i^* \phi_i$$

- PageRank operator: 'β'-Green's function

$$\mathrm{pr}(\alpha, \cdot) = (1-\alpha) \sum_{t=0}^{\infty} (\alpha W)^t = \frac{\beta}{\beta I + \mathcal{L}}$$

Inverse to 'shifted Laplacian' - gives way to generalize PageRank!

Simplicial Complex PageRank Operator

Definition (Normalized Simplicial Laplacian)

$$\mathcal{L}_k = \partial_k^* D_{(k)}^{-1} \partial_k + D_{(k+1)}^{-1} \partial_{k+1} \partial_{k+1}^*$$

$D_{(k+1)}$ = diagonal degree matrix of k-faces.
$D_{(k+1)}(f, f) = \#\ (k+1)$ faces the k-face f lies in.

Definition (Simplicial Complex PageRank Operator)

$$\mathrm{pr}^{(k)}(\beta, \cdot) = (\beta I + \mathcal{L}_k)^{-1}$$

β-Green's function for the $\mathcal{L}_k$ Laplacian.

Problem: $\mathrm{pr}^{(k)}(\beta, \cdot)$ not stochastic
Question: How to rank with $\mathrm{pr}^{(k)}(\beta, \cdot)$?

Rethinking personalized (graph) PageRank

For a vertex v:

- $\mathrm{pr}(\alpha, \chi_v)$: vector entries measure relative rankings of vertices with respect to seed.
- $||\mathrm{pr}(\alpha, \chi_v)||_2$: measures 'spread' of $\mathrm{pr}(\alpha, \chi_v)$.
- $||\mathrm{pr}(\alpha, \chi_v)||_2$ small:
  - Entries of $\mathrm{pr}(\alpha, \chi_v)$ small
  - $\Rightarrow$ short random walks starting at v mix
  - $\Rightarrow$ v 'not important' to 'bottlenecks'
- $||\mathrm{pr}(\alpha, \chi_v)||_2$ large:
  - Some entries of $\mathrm{pr}(\alpha, \chi_v)$ large
  - $\Rightarrow$ short random walks starting at v don't mix
  - $\Rightarrow$ v 'important' to 'bottlenecks'.

Idea: $||\mathrm{pr}(\alpha, \chi_v)||_2$ measures significance of v to certain geometric graph features (bottlenecks)

Geometry through Hodge Decomposition

Idea

Use $||\mathrm{pr}^{(2)}(\beta, \chi_e)||_2$ to give ranking of edges.

- More generally: Different rankings via Hodge Decomposition
- Write $\chi_e = h_e + g_e + c_e$
  - $g_e = \mathrm{proj}_{\mathrm{Im}\partial_1^*}(\chi_e)$: Gradient flow
  - $c_e = \mathrm{proj}_{\mathrm{Im}\partial_2}(\chi_e)$: Curl flow
  - $h_e = \mathrm{proj}_{\ker \partial_2^* / \mathrm{Im}\partial_1^*}(\chi_e)$: Harmonic flow
- $\mathrm{pr}^{(2)}(\beta, \cdot)$ acts independently on each component.
- Gives four rankings:

$$||\mathrm{pr}^{(2)}(\beta, \chi_2)||_2 \qquad ||\mathrm{pr}^{(2)}(\beta, g_e)||_2 \qquad ||\mathrm{pr}^{(2)}(\beta, c_e)||_2$$
$$||\mathrm{pr}^{(2)}(\beta, h_e)||_2$$

Geometry of Rankings

Hodge Decomposition:

$$\chi_e = h_e + g_e + c_e$$

- $||\mathrm{pr}^{(2)}(\beta, h_e)||_2$: Measures importance of e to ‘holes’.
- $||\mathrm{pr}^{(2)}(\beta, g_e)||_2$: Measures importance of e to ‘sparse cuts’.
- $||\mathrm{pr}^{(2)}(\beta, c_e)||_2$:
  - High for edges central to dense areas where inconsistency arises
  - Less well understood
- $||\mathrm{pr}^{(2)}(\beta, \chi_e)||_2$: Combination of three influences.

Question: Influence of β?

Influence of β

What role does β play?

- Eigenvalue λ of $\mathcal{L}_k \to$ eigenvalue of $\frac{1}{\beta+\lambda} \to 0$.
- β large: Eigenvalues similar in magnitude, all faces ranked the same.
- β small: $\frac{1}{\beta+\lambda} \approx \frac{1}{\lambda}$ - projections onto eigenvectors with small λ dominate.

Edge Rankings

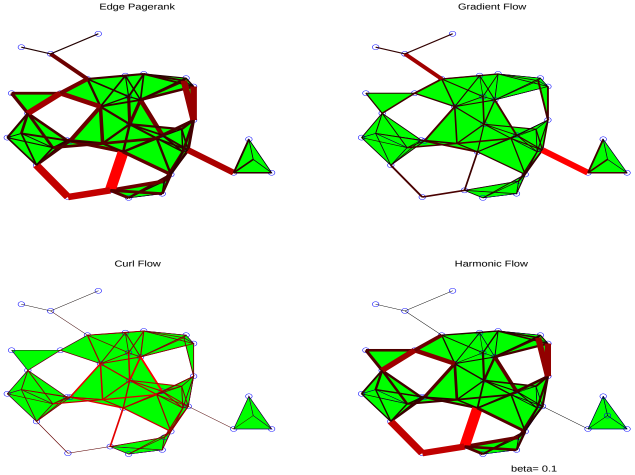
Video: Combined PageRank of edges in complex as $\beta \to 0$.

As $\beta \to 0$, edges in sparse cuts (eg. 59) and crucial to holes most important.

Edge Rankings

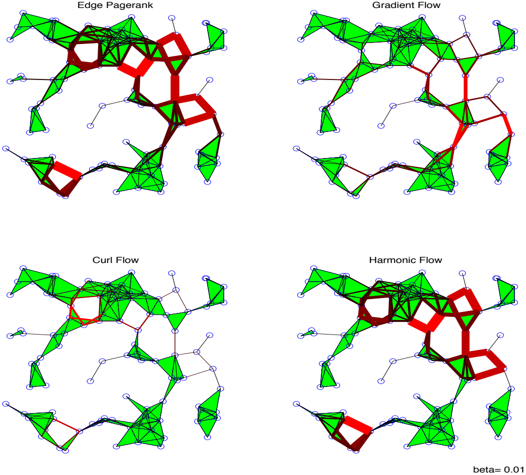
Video: PageRank of component parts as $\beta \to 0$

Edge rankings and flow decompositions



PageRank of component parts for $\beta = 0.1$

Edge rankings and flow decompositions



PageRank of component parts for $\beta = 0.01$

Conclusions

- Dirichlet problems on complexes are more subtle than on graphs: existence of unique solutions is not guaranteed. Ample sets give us the right framework
- Page Rank can be generalized to complexes to measure importance of edges and higher order simplices.
- Edge importance is measured by the "spread" of personalized PageRank in the curl, harmonic and gradient subspaces.
- changing the teleportation constant β allows for measuring both importance and *persistence*.