Nonlinear Laplacian for Digraphs and Its Applications to Network Analysis

Yuichi Yoshida
National Institute of Informatics &
Preferred Infrastructure, Inc.

@WSDM 2016

Question

Can we develop spectral graph theory for digraphs?

- Spectral graph theory analyzes graph properties via eigenpairs of associated matrices.
  - Adjacency matrix, incidence matrix, Laplacian
- Applications
  - Approximation to graph parameters (e.g, chromatic number), community detection, visualization, etc.
- Well established for undirected graphs.

Question

Can we develop spectral graph theory for digraphs?

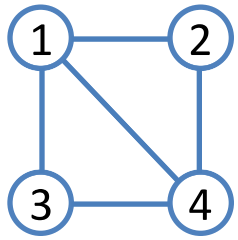
- Many real-world networks are directed!
  - Web graph, Twitter followers, phone calls, paper citations, food web, metabolic network.

- Extensions for digraphs are largely unexplored and unsatisfying.

Laplacian

- Graph G = (V, E)
- Adjacency matrix: A_G
- Degree matrix: D_G
- Laplacian $L_G := D_G - A_G$

G



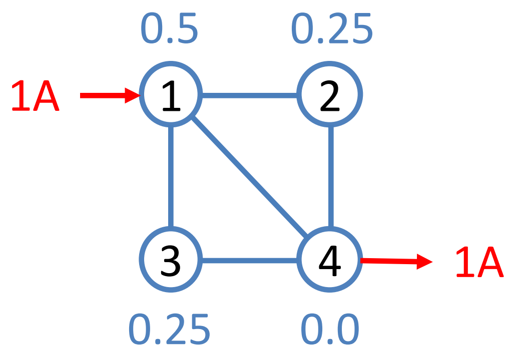
$$\underset{D_G}{\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}} - \underset{A_G}{\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}} = \underset{L_G}{\begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}}$$

- Normalized Laplacian $\mathcal{L}_G := D_G^{-1/2} L_G D_G^{-1/2} = I - D_G^{-1/2} A_G D_G^{-1/2}$

Interpretation of Laplacian

- Regard G as an electric circuit.
- An edge = a resistance of 1Ω.
- Flow a current of **b**(u) ampere to each vertex u $\in$ V.

The voltages of vertices can be computed by solving $L_G\mathbf{x} = \mathbf{b}$



Extensions for Digraphs

Existing extensions of Laplacians for digraphs:

1. $L_G = D_G^+ - A_G$
   – Asymmetric and hence eigenpairs are complex-valued.
2. Chung's Laplacian
   – Assume strong connectivity. Need random walks to interpret its eigenpairs.

Our contributions

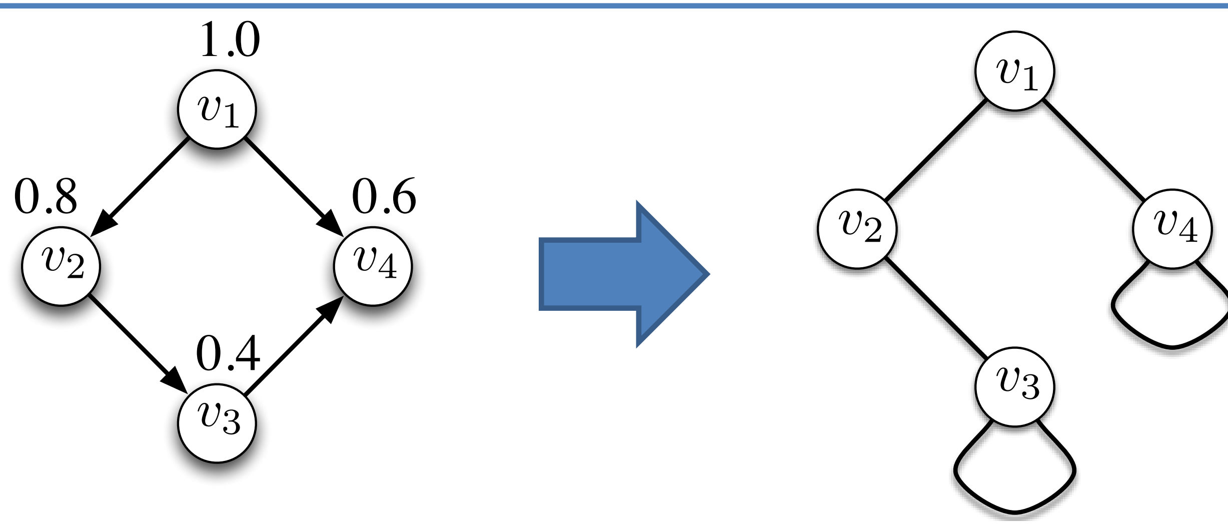
1. Laplacian for digraphs whose eigenpairs can be interpreted more combinatorially.
2. Algorithm that computes a small eigenvalue.
3. Applications to visualization and community detection.

Nonlinear Laplacian

Nonlinear Laplacian L_G: $\mathbb{R}^n \to \mathbb{R}^n$ for a digraph G:

From a vector $\mathbf{x} \in \mathbb{R}^n$, we compute $L_G(\mathbf{x})$ as follows

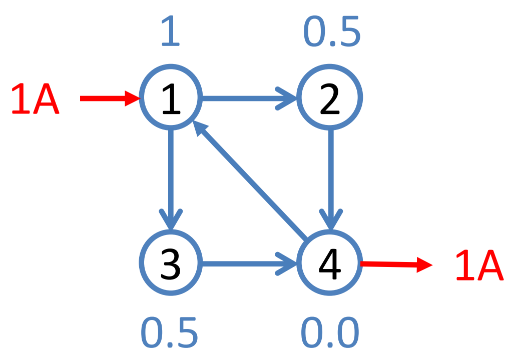
1. Define an *undirected* graph as follows: for each arc u → v
   - If $\mathbf{x}(u) \geq \mathbf{x}(v)$, add an (undirected) edge {u, v}.
   - Otherwise, add self-loops.
2. Let L_H be the Laplacian of H.
3. Output $L_H\mathbf{x}$.



Interpretation

- Regard G = (V, E) as an electric circuit.
- An edge = a diode of 1Ω (current flows only one way).
- Flow a current of **b**(u) ampere to each vertex u ∈ V.

The voltages of vertices can be computed by solving $L_G(\mathbf{x}) = \mathbf{b}$.



Eigenpair of Nonlinear Laplacian

- Normalized Laplacian $\mathcal{L}_G : \mathbf{x} \longmapsto D_G^{-1/2} L_G (D_G^{-1/2}\mathbf{x})$

- $(\lambda, \mathbf{v})$ is an eigenpair of $\mathcal{L}_G$ if $\mathcal{L}_G(\mathbf{v}) = \lambda\mathbf{v}$
  - Trivial eigenpair: $(\lambda_1 = 0, \mathbf{v}_1)$.

> For any subspace U of positive dimension, $\Pi_U\mathcal{L}_G$ has an eigenpair. (Π_U = Projection matrix to U)

$\Rightarrow$ Nontrivial eigenpair of $\mathcal{L}_G$ exists by choosing $U = \mathbf{v}_1^{\perp}$.

Let λ_2 be the smallest eigenvalue orthogonal to $\mathbf{v}_1$.

Algorithm

Computing λ_2 is (likely to be) NP-hard.

Suppose we start the diffusion process

$$d\boldsymbol{x} = -\Pi_U \mathcal{L}_G(\boldsymbol{x})dt$$

from a vector in the subspace U = $\mathbf{v}_1^{\perp}$.

- **x** converges to an eigenvector orthogonal to $\mathbf{v}_1$.
- Rayleigh quotient $\mathcal{R}_G(\boldsymbol{x}) := \frac{\boldsymbol{x}^T \Pi_U \mathcal{L}_G(\boldsymbol{x})}{\boldsymbol{x}^T \boldsymbol{x}}$ never increases during the process.

$\Rightarrow$ We can get a eigenvector of a small eigenvalue.

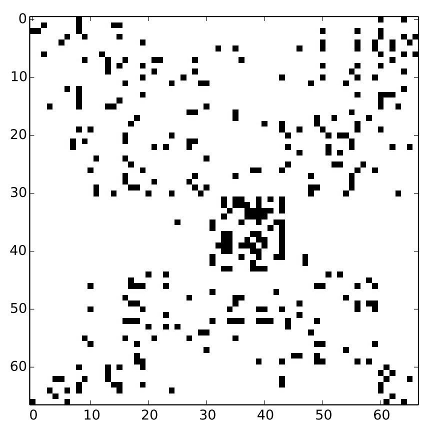
Visualization: Chung's & Nonlinear Laplacian

Friendship network at a high school in Illinois

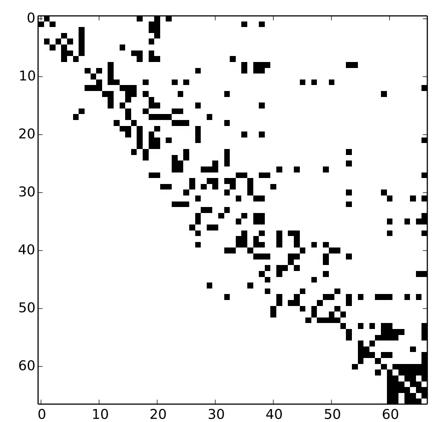
- u → v: u regards v as a friend.

Reorder vertices according to the eigenvector computed by the diffusion process.

Chung's Laplacian



Nonlinear Laplacian



Our method shows the directivity of the network more clearly.

Visualization: Interpretation

Laplacian for undirected graphs

$\lambda_2 = \min \sum_{\{u,v\} \in E} \left(\boldsymbol{x}(u) - \boldsymbol{x}(v)\right)^{\mathbf{2}}$ s.t. $\|\mathbf{x}\| = 1$, $x \perp \mathbf{v}_1$

- Adjacent vertices are placed near.

Chung's Laplacian

$\lambda_2 = \min \sum_{u \to v \in E} \left(\boldsymbol{x}(u) - \boldsymbol{x}(v)\right)^{\mathbf{2}} \pi_u / d_u^+$ s.t. $\|\mathbf{x}\| = 1$, $x \perp \mathbf{v}_1$.

- Important vertices (w.r.t. RW) are placed in the middle.

Nonlinear Laplacian

$\lambda_2 = \min \sum_{u \to v \in E} \max(\boldsymbol{x}(u) - \boldsymbol{x}(v), 0)^{\mathbf{2}}$ s.t. $\|\mathbf{x}\| = 1$, $x \perp \mathbf{v}_1$.

- If $\mathbf{x}(u) \leq \mathbf{x}(v)$, then we get no penalty.
- In particular, $\lambda_2 = 0$ when G is a DAG.

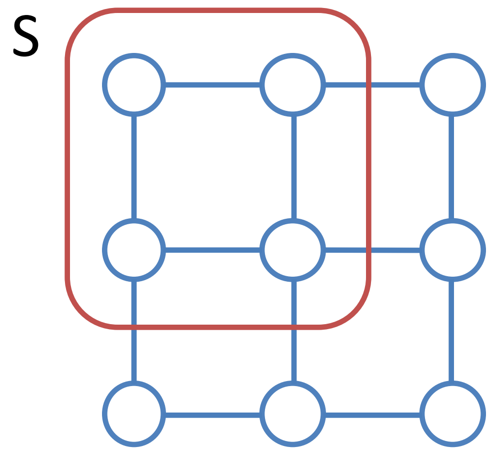
Community Detection: Undirected Graphs

S: Vertex set

vol(S): Total degree of vertices in S

cut(S): # of edges between S and V-S

The conductance $\phi(S)$ of S is $\dfrac{\mathrm{cut}(S)}{\min(\mathrm{vol}(S),\mathrm{vol}(V-S))}$



$\phi(S) = 4/12 = 1/3$

Small conductance

→ Good community

Community Detection: Undirected Graphs

Cheeger's inequality ('70)

$$\lambda_2/2 \leq \min_S \phi(S) \leq \sqrt{(2\lambda_2)}$$

- We can efficiently compute S with $\phi(S) \leq \sqrt{(2\lambda_2)}$ from $\mathbf{v}_2$.
- Still widely used.

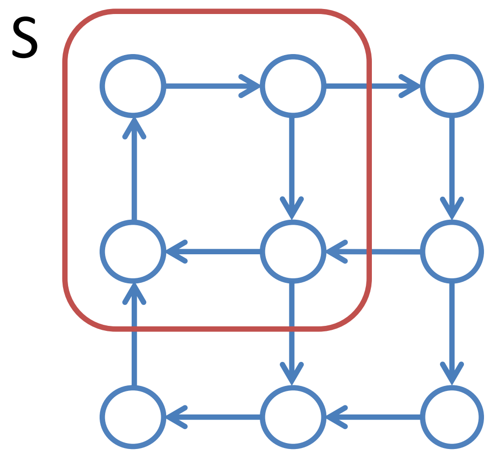
Community Detection: Digraphs

S: Vertex set

vol(S): Total indegrees + outdegrees of vertices in S

$cut^+(S)$: # of arcs from S to V-S

(Directed) conductance ϕ(S) of S is $\frac{\min(\mathrm{cut}^+(S), \mathrm{cut}^+(V-S))}{\min(\mathrm{vol}(S), \mathrm{vol}(V-S))}$



$\phi^+(S) = 2/12 = 1/6$

Community Detection: Digraphs

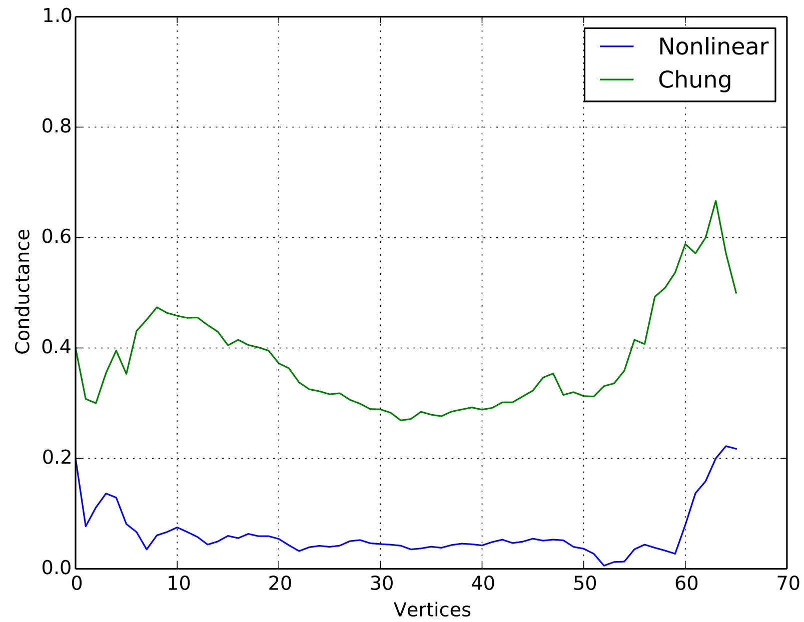
Cheeger's inequality for digraphs

$$\lambda_2/2 \leq \min_S \phi(S) \leq 2\sqrt{\lambda_2}$$

- We can efficiently compute S with $\phi(S) \leq 2\sqrt{\mathcal{R}_G(\mathbf{x})}$ from **x**.

Community Detection: Digraphs

Reorder vertices according to the obtained eigenvector in the high school network, and plot ϕ of each prefix set.



- ϕ is low everywhere = directivity
- ϕ rapidly increases = community

Summary

Nonlinear Laplacian for digraphs

- Strong connectivity is not needed.
- Eigenpairs are combinatorially interpretable.
- Applications to visualization and community detection.

Future Work

- Approximation of λ_2.
- Finding a community in time proportional to its size.
- Other applications.