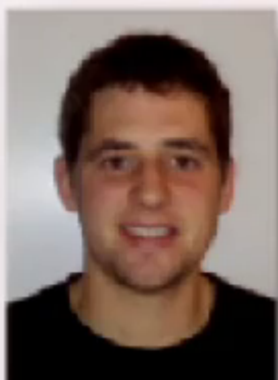
Spacey Random Walks on Higher-Order Markov Chains



Joint work with Austin Benson, Lek-Heng Lim, supported by NSF CAREER CCF-1149756 IIS-1422918

David F. Gleich
Purdue University

Spacey walk
on Google Images
From Film.com

WARNING!

This talk presents the “forward” explicit derivation (i.e. lots of little steps)

rather than the implicit “backwards” derivation (i.e. big intuitive leaps)

PageRank: The initial condition

My dissertation

Models & Algorithms for PageRank Sensitivity

The essence of PageRank

PageRank
beyond
the Web
arXiv:1407.5107

chain **P**, PageRank

chain with great "utility"

nary distribution

ence $(\mathbf{I} - \alpha \mathbf{P})\mathbf{x} = (1 - \alpha)\mathbf{v}$

bility

PageRank: The initial condition

My dissertation

Models & Alg

How Google's PageRank Quantifies Things (Like History's Best Tennis Player) Beyond The Web

by Jessica Leber
Fast Magazine

PageRank can be used for everything from analyzing the world's most important books to predicting traffic flow to ending sports arguments.

Be careful about what you discuss after a talk…

I gave a talk

at the Univ. of Chicago and visited Lek-heng Lim

He told me about a new idea

in Markov chains analysis and tensor eigenvalues

Approximate stationary distributions of higher-order Markov chains

Due to Michael Ng and collaborators

A higher order Markov chain

depends on the last *few* states.

$$P(X_{t+1} = i \mid \text{history}) = P(X_{t+1} = i \mid X_t = j, X_{t-1} = k)$$

These become Markov chains on the product state space.
But that's usually too large for stationary distributions.

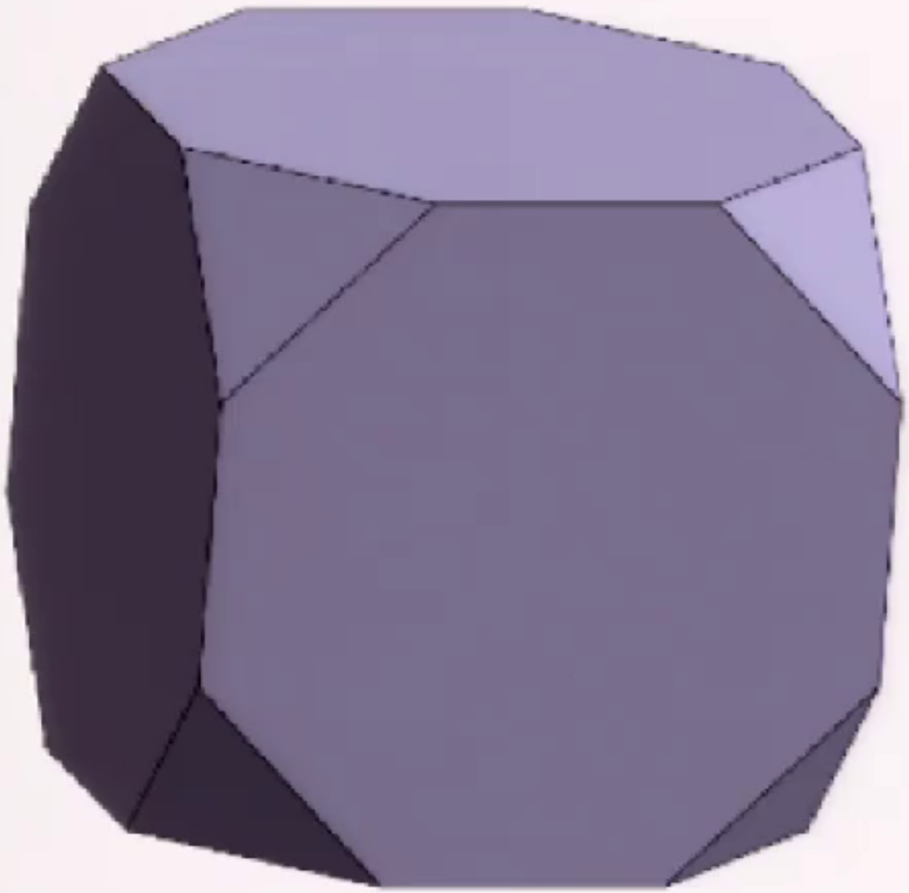
$$P(X = [i,j]) = X_{i,j}$$

The approximation

is that we form a rank-1 approximation of that stationary distribution object.

$$P(X = [i,j]) = x_i x_j$$

Why?



We want to analyze higher-order relationships and multi-way data and …

Things like

- Enron emails
- Regular hypergraphs

And there's three+ indices! So it's a higher-order Markov chain

Approximate stationary distributions of higher-order Markov chains

Due to Michael Ng and collaborators

The new problem

of computing an approx. stationary dist. is a tensor eigenvector

$$x_i = \sum_{jk} P_{ijk} x_j x_k \quad \text{or} \quad \mathbf{x} = \underline{\mathbf{P}}\mathbf{x}^2$$

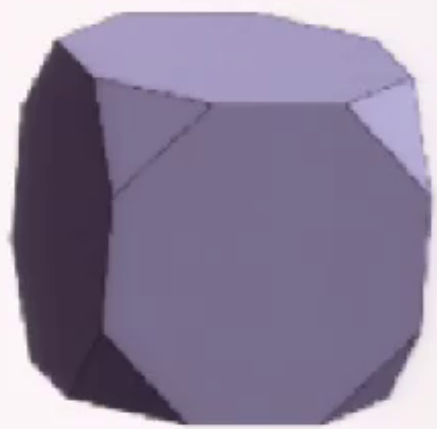
The new problem'

- *existence* is guaranteed under mild conditions
- uniqueness … **require heroic algebra**
- convergence … **(and are hard to check)**

Some small quick notes

A stochastic matrix **M** is a Markov chain

A stochastic hypermatrix / tensor / probability $\underline{\mathbf{P}}$ table is a higher-order Markov chain



One nagging question …

Is there a stochastic process that underlies this approximation?

We tried many

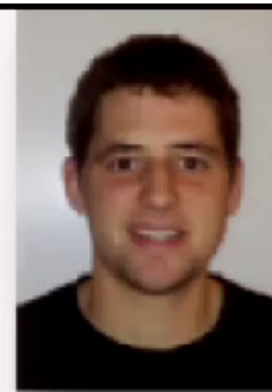
- *apriori* good and
- *retrospectively* bad

ideas for the second eigenvector

Austin and I were talking one day …

EUREKA!

The spacey random walk




Consider a higher-order Markov chain.

$$P(X_{t+1} = i \mid \text{history}) = P(X_{t+1} = i \mid X_t = j, X_{t-1} = k)$$

If we were perfect, we'd figure out the stationary distribution of that. *But we are spacey!*

- On arriving at state j, we promptly "space out" and forget we came from k.
- But we still believe we are "higher-order"
- So we invent a state k by drawing a random state from our history.

The spacey random walk

Let $C_t(k) = (1 + \sum_{s=1}^{t} \text{Ind}\{X_s = k\})$

$P(X_{t+1} = i \mid X_t = j \text{ and the right filtration on history})$

$$= \sum_k P_{i,j,k} C_k(t)/(t+n)$$

How often we've visited state k in the past

This is a vertex-reinforced random walk!
e.g. Polya's urn.

Pemantle, 1992; Benaïm, 1997; Pemantle 2007

Stationary distributions of vertex reinforced random walks

A vertex-reinforced random walk at time t transitions according to a Markov matrix $\mathbf{M}$ given the observed frequencies.

$$P(X_{t+1} = i \mid X_t = j \text{ and the right filtration on history})$$
$$= [\mathbf{M}(t)]_{i,j}$$
$$= [\mathbf{M}(\mathbf{c}(t))]_{i,j}$$

This has a stationary distribution, iff the dynamical system

$$\frac{d\mathbf{x}}{dt} = \pi[\mathbf{M}(\mathbf{x})] - \mathbf{x} \qquad \pi[\mathbf{M}] \text{ is a map to the stat. dist.}$$

converges.

A vertex-reinforced random walk at time t transitions according to a Markov matrix $\mathbf{M}$ given the observed frequencies.

$$P(X_{t+1} = i \mid X_t = j \text{ and the right filtration on history})$$
$$= [\mathbf{M}(t)]_{i,j}$$
$$= [\mathbf{M}(\mathbf{c}(t))]_{i,j}$$

This has a stationary distribution, iff the dynamical system

$$\frac{d\mathbf{x}}{dt} = \pi[\mathbf{M}(\mathbf{x})] - \mathbf{x}$$

$\pi[\mathbf{M}]$ is a map to the stat. dist.

converges.

The Markov matrix for Spacey Random Walks

We have all sorts of cool results on spacey random walks… e.g.

Suppose you have a Polya Urn with memory… Then it always has a stationary distribution!

Back to Multilinear PageRank

The Multilinear PageRank problem is what we call a spacey random surfer model.

- This is a spacey random walk
- We add random jumps with probability (1-alpha)

It's also a vertex-reinforced random walk.

Thus, it has a stationary probability if

$$\frac{d\mathbf{x}}{dt} = \pi[\mathbf{M}(\mathbf{x})] - \mathbf{x} \qquad \mathbf{M}(\mathbf{x}) = \alpha \sum_k \underline{\mathbf{P}}(:, :, k) x_k + (1 - \alpha)\mathbf{v}$$

converges. `Which occurs when alpha < 1/order !`

The Multilinear PageRank problem is what we call a spacey random surfer model.

- This is a spacey random walk
- We add random jumps with probability (1-alpha)

It's also a vertex-reinforced random walk.

Thus, it has a stationary probability if

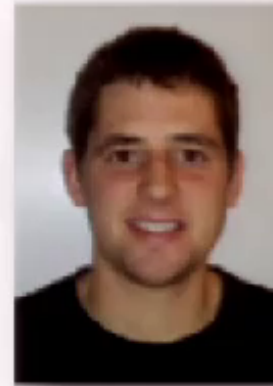
$$\frac{d\mathbf{x}}{dt} = \pi[\mathbf{M}(\mathbf{x})] - \mathbf{x} \qquad \mathbf{M}(\mathbf{x}) = \alpha \sum_k \underline{\mathbf{P}}(:,:,k) x_k + (1-\alpha)\mathbf{v}$$

converges. `Which occurs when alpha < 1/order !`

Some interesting notes about vertex reinforced random walks

Meanwhile …
Spectral clustering of tensors

SIAM Data Mining 2015, arXiv:1502.05058




Austin Benson (a colleague) asked

if there were any interesting method to "cluster" tensors.

"Conjecture" spectral clustering on tensors

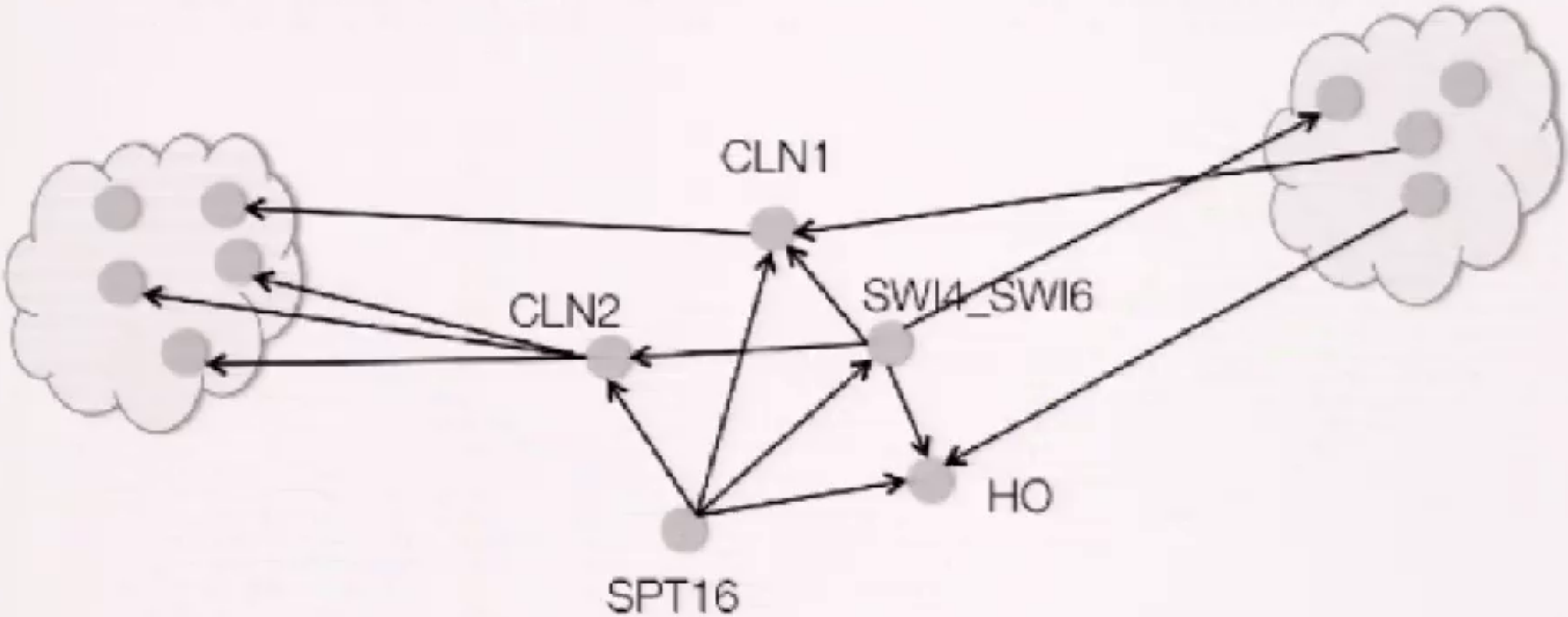
graph/tensor $\rightarrow$ higher-order random walk

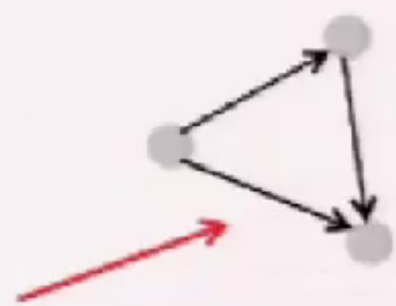
$\rightarrow$ second eigenvector $\quad \mathbf{M}(\mathbf{x})^T\mathbf{y} = \lambda_2\mathbf{y}$

$\rightarrow$ sweep cut partition

Use the asymptotic
Markov matrix!

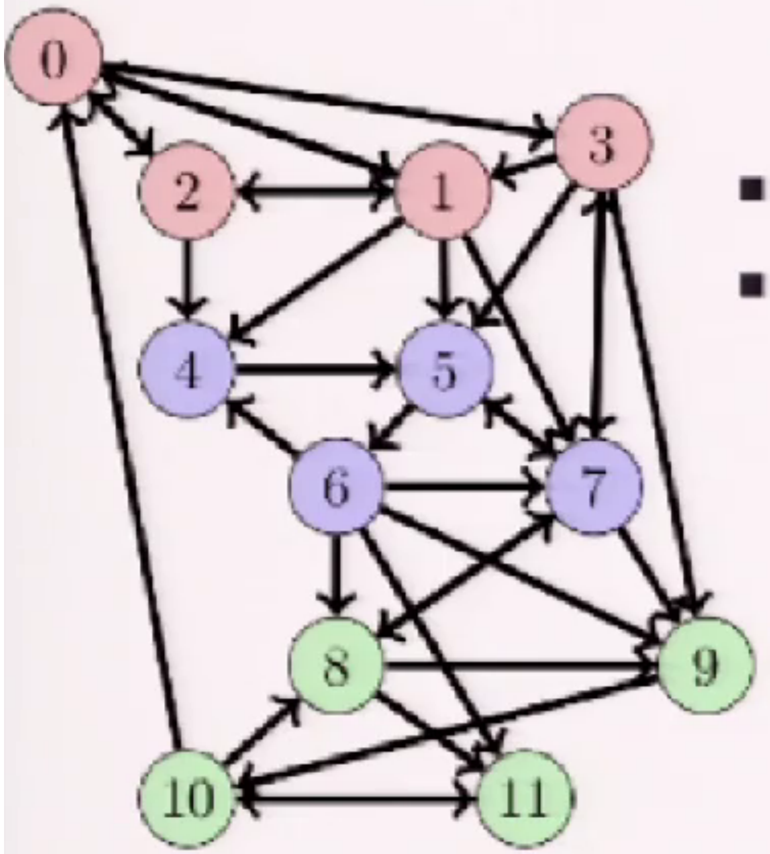
Problem current methods only consider edges



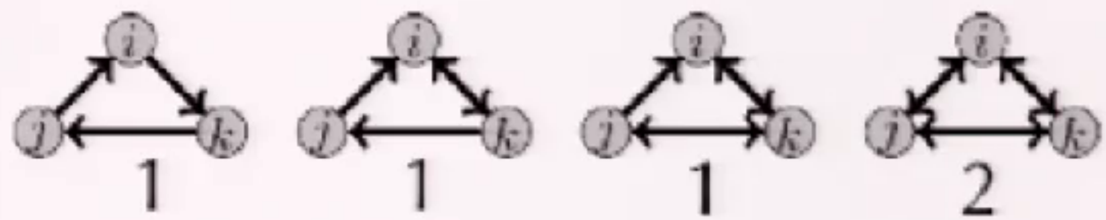


In transcription networks, the "feedforward loop" motif represents biological function. Thus, we want to look for clusters of this structure.

An example with a layered flow network



- The network "flows" downward
- Use directed 3-cycles to model flow



- Tensor spectral clustering: {0,1,2,3}, {4,5,6,7}, {8,9,10,11}
- Standard spectral: {0,1,2,3,4,5,6,7}, {8,10,11}, {9}

WAW2015

EURANDOM – Eindhoven – Netherlands

Workshop on Algorithms and Models for the Web Graph
(but it's grown to be all types of network analysis)
December 10-11

Winter School on Complex Network and Graph Models
December 7-8

Submissions Due July 25th!

Winter School on Complex Network and Graph Models
December 7-8

Submissions Due July 25th!

Time for Lots of Questions!

Manuscripts

Li, Ng. On the limiting probability distribution of a transition probability tensor. Linear & Multilinear Algebra 2013.

Gleich. PageRank beyond the Web. (accepted at SIAM Review)

Gleich, Lim, Yu. Multilinear PageRank. (under review…)