# Learning Random Walks to Rank Nodes in Graphs


**Alekh Agarwal    Soumen Chakrabarti**

IIT Bombay


## Motivation

- Typically studied in vector spaces
- Graphs model relationships naturally
- Should exploit/respect graph structure and links
- Web, XML, database search ...

## Interpretations of Edges

### Associative Networks

- Edges encode *similarity*
- Preference for *smooth* scoring functions
- Edge weights encode extent of similarity

### Random Walk Approach

- Edges indicate *endorsement*
- Motivated by Pagerank, widely used
- Each edge $(u,v)$ has associated transition probability $Q(v,u) = \Pr(v|u)$

## Training and Testing

- Given fixed graph $G$
- Training set of node-pairs "$u \prec v$", means we want score$(u) <$ score$(v)$
- Associative Networks: Learn a *good* scoring function on the nodes
- Random Walk: Discover underlying transition probabilities
- Test node pairs from same distribution
- Number of incorrect predictions measured

## Associative Scoring

- Let $\pi = Q\pi, \Pi = \mathrm{diag}(\pi)$
- Directed Graph Laplacian

$$L = \mathbb{I} - \left(\frac{\Pi^{1/2}Q\Pi^{-1/2} + \Pi^{-1/2}Q\Pi^{1/2}}{2}\right)$$

- Impose roughness penalty

$$f^\top L f = \sum_{(u,v)\in E} \pi(u)Q_{uv}\left(\frac{f(u)}{\sqrt{\pi(u)}} - \frac{f(v)}{\sqrt{\pi(v)}}\right)^2$$

- Scoring algorithm:

$$\min_{\substack{f:V\to\mathbb{R}\\ s=\{s_{uv}\geq 0: u\prec v\}}} \tfrac{1}{2} f^\top L f + B\sum_{u\prec v} s_{uv} \quad \text{subject to} \qquad \text{(Lap)}$$

$$f_v - f_u \geq 1 - s_{uv} \quad \forall u \prec v$$

## Ranking using Random Walks

- Pagerank $\pi$ induces flow $q_{uv} = \pi(u)Q_{uv}$ along graph edges
- Goal: learn flow $p$ close to $q$ that satisfies training preferences

$$\min_{\substack{\{0\leq p_{uv}\}\\ \{0\leq s_{uv}\}}} \sum_{(u,v)\in E} p_{uv}\log\frac{p_{uv}}{q_{uv}} + B\sum_{u\prec v} s_{uv} \qquad \text{(KL)}$$

$$\text{s.t.:} \quad \sum_{(u,v)\in E} p_{uv} = 1$$

$$\forall v\in V: \quad \sum_{(u,v)\in E} p_{uv} - \sum_{(v,w)\in E} p_{vw} = 0 \qquad \text{(Bal)}$$

$$\forall u \prec v: \quad \sum_{(w,u)\in E} p_{wu} - \sum_{(w,v)\in E} p_{wv} - s_{uv} \leq 0 \qquad \text{(Pref)}$$

## Laplacian-KL Correspondence

- $f(v) \propto \sqrt{\pi(v)}$ minimizes $f^\top L f$
- Both (Lap) and (KL) prefer same ordering in absence of training data

Let $p$ be a valid flow distribution on $G$, and let $q$ be the reference flow distribution induced by $\pi$. Let $f_p(u) = \sqrt{\sum_{\{u:(u,v)\in E\}} p_{uv}}$. Then

$$\mathrm{KL}(p\|q) \leq \epsilon \ \Rightarrow\ f_p^\top L f_p \leq 4\sqrt{2\epsilon\ln 2}.$$

- Proof follows from standard inequalities and $\mathrm{KL}(p\|q) \geq \frac{1}{2\ln 2}\|p-q\|_1^2$
- Small KL-distance $\Rightarrow$ low Laplacian roughness
- Justifies $\mathrm{KL}(\cdot\|q)$ as a regularizer

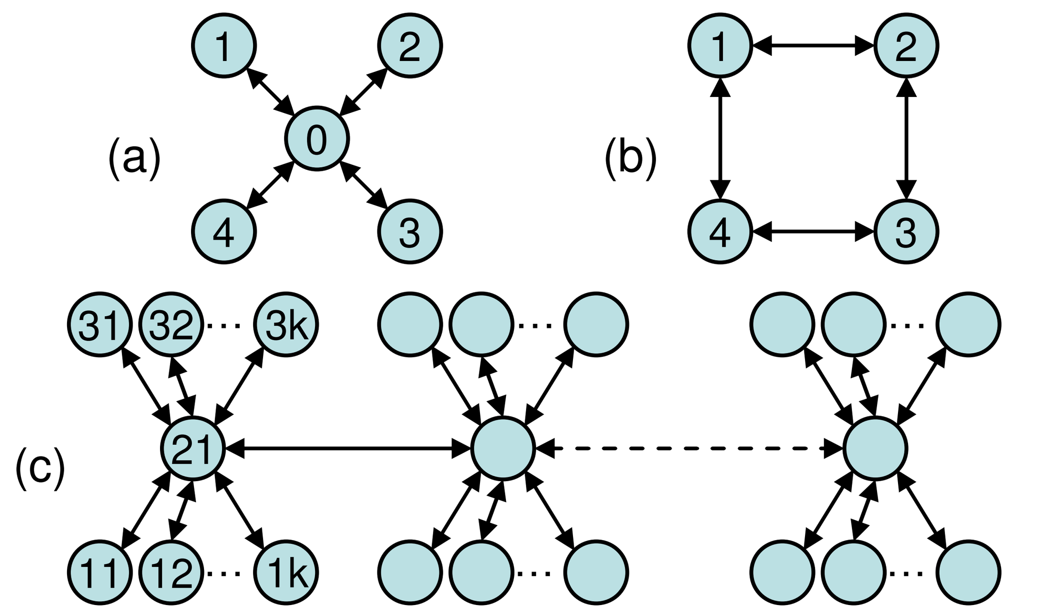
## Generalization Bounds

For any $m \geq 1$, any $\delta \in (0,1)$, w.p. $\geq 1-\delta$ over random draws of sample $\prec$ of size $m$,

$$R \leq R_{\mathrm{emp}} + 2\beta + (4m\beta + 1)\sqrt{\frac{\ln(1/\delta)}{2m}}$$

- Graph agnostic bound of $\beta = \frac{2\ln 2}{\lambda m}$ trivial to show

## Using Graph Properties




- $G$ may not always reduce function class, $(n-1)!$ possible orderings in (a)
- Bound on outdegree $D$ not enough: in (c) middle layer acts as a hub, by biasing outflows among neighbors in arbitrary ratios
- $G$ *may* help: $1 \prec 3 \prec 2 \prec 4$ not learnable in (b)

Eccentricity $\quad \rho = \max_{u\in V} \frac{\max_{v:(u,v)\in E} p_{uv}}{\min_{v:(u,v)\in E} p_{uv}}$

- Controls influence of single node on rankings, along with $D$
- Can find modified $\beta$, worsens with increasing $D, \rho$. Proof involves upper and lower bounding $p_{uv}$ using induced pagerank $\phi(u)$, $D$ and $\rho$
- Note: Relative loss bounds, not upper bound on 0-1 risk.

## Margin and 0/1 Loss Bound

- Hinge loss $\max(0, 1 + f(u) - f(v))$ upper bound on 0-1 loss. Training loss bounds empirical 0-1 risk
- (Pref) uses shifted hinge $\max(0, f(u) - f(v))$, not upper bound on 0-1 loss
- $\sum_{(u,v)} p_{uv} = 1$ keeps flows in [0,1]
- Power-law assumptions $\Rightarrow$ most nodes with small pageranks
- Arbitrary additive margin (e.g., "1" for hinge loss) infeasible
- Solution: Make $F = \sum_{(u,v)} p_{uv}$ variable
- Possible as $\mathrm{KL}(p\|q)$ is still well-defined. I.e., if $F \geq 1$, $\mathrm{KL}(p\|q) \geq 0$ and minimized at $p = Fq$ for a fixed $F$

$$\min_{\substack{\{p_{uv}\},\{s_{uv}\}\\ F\geq 1}} \sum_{(u,v)\in E} p_{uv}\log\frac{p_{uv}}{q_{uv}} + B\sum_{u\prec v} s_{uv} + B_1 F^2$$

$$\text{subject to} \quad 1 + \sum_{(w,u)\in E} p_{wu} - \sum_{(w,v)\in E} p_{wv} - s_{uv} \leq 0 \quad \forall u \prec v$$

- Small $F \Rightarrow$ large margin (effectively $\frac{1}{F}$)
- Generalization bound polynomial in $\delta$
- Worsens as upper bound on $F$ increases

## Cost-sensitive Ranking

- Tester wants no mistakes near top of ranked list
- Can penalize on true scores/ranks in tune with importance-weighted classification
- **But**, true scores/ranks not available, excessive cognitive burden on trainer (may be end-user)

Use algorithm's score estimates as surrogate

- Demands high confidence in ranks of nodes with high scores
- Cost-sensitive loss $g(f_u, f_v, \ell(f_u, f_v))$
- For example, $g(f_u, f_v, \ell(f_u, f_v)) = (\max(f_u, f_v) + \ell(f_u, f_v))^2$
- Can show $\Pr(g(f_u, f_v, \ell(f_u, f_v)) \geq \epsilon) \leq \delta(\epsilon) \Rightarrow \Pr(\ell(f_u, f_v) \geq \epsilon \wedge h(f_u, f_v) \geq \theta) \leq \delta(\epsilon\gamma\theta)$
- Generalization proved using stability wrt $g$
- Can easily extend (Lap) to implement this as a convex quadratic optimization

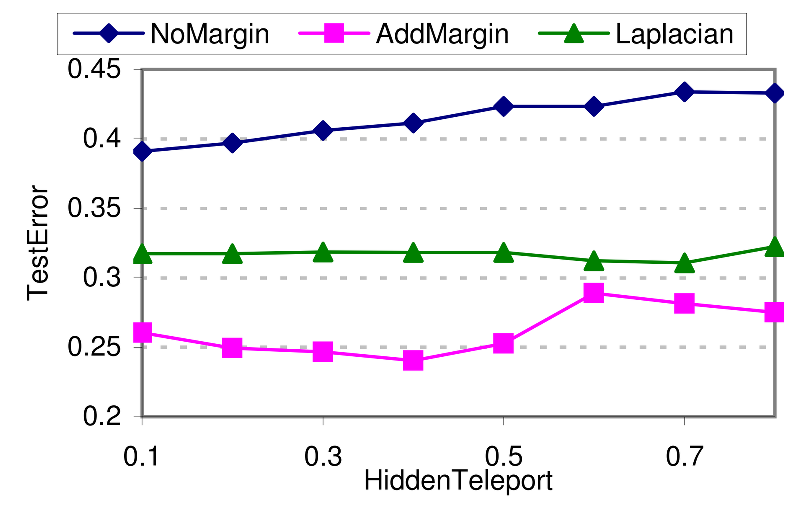
## Experimental Results

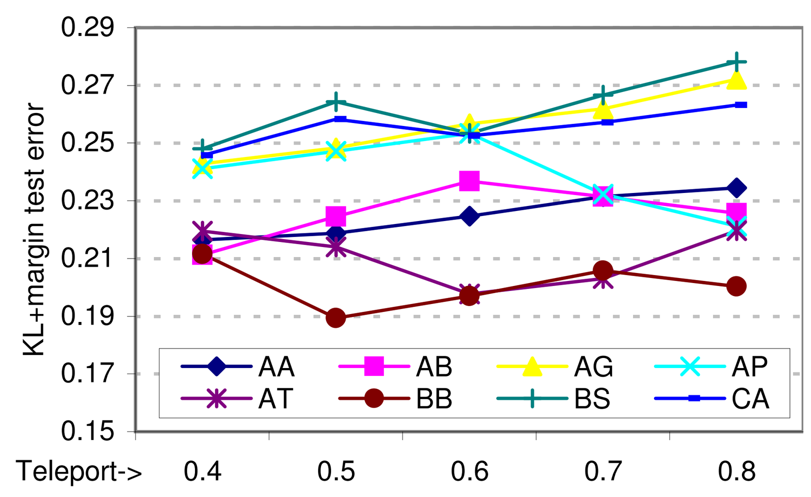
**Synthetic RMAT Graphs:** 1000-4000 nodes, 4000-16000 edges resembling real social networks

**Real Graphs:** Biological networks, directed graphs with social network like degree distributions

**Preferences:** Aim to discover a hidden favored "personalized" community

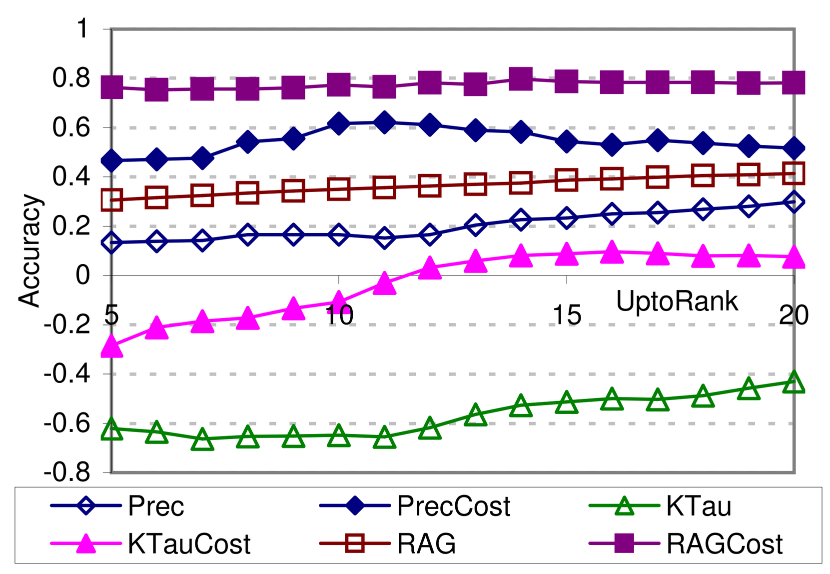
- Compute reference unweighted $\pi$
- Directed large teleport(0.1-0.8) into hidden random seed nodes, get "true" scores $\phi^*$
- Sample train and test sets using these, keep node-disjoint to remove transitivity




### Evaluating Cost-Sensitive Ranking

- Precision at $k$ = $|T_k^u \cap \hat{T}_k^u|/k$
- Relative Average Goodness (RAG) at $k$ = $\sum_{v\in T_k} \phi^*(v) / \sum_{v\in T_k^*} \phi^*(v)$
- Kendall's $\tau$ between true and computed ranks




## Summary

- Correspondences between Laplacian and random-walk ranking
- New generalization bounds for random-walk ranking
- Ranking with margin using network flow
- Cost-sensitive ranking framework

## References


[1] A. Agarwal, S. Chakrabarti, and S. Aggarwal. Learning to rank networked entities. In *SIGKDD Conference*, pages 14–23, 2006.

[2] S. Agarwal. Ranking on graph data. In *ICML*, pages 25–32, 2006.





Author contact: alekh@cse.iitb.ac.in, soumen@cse.iitb.ac.in