DivRank: Interplay of Prestige and Diversity in Information Networks

Qiaozhu Mei[1,2],   Jian Guo[3],   Dragomir Radev[1,2]

[1.]*School of Information*
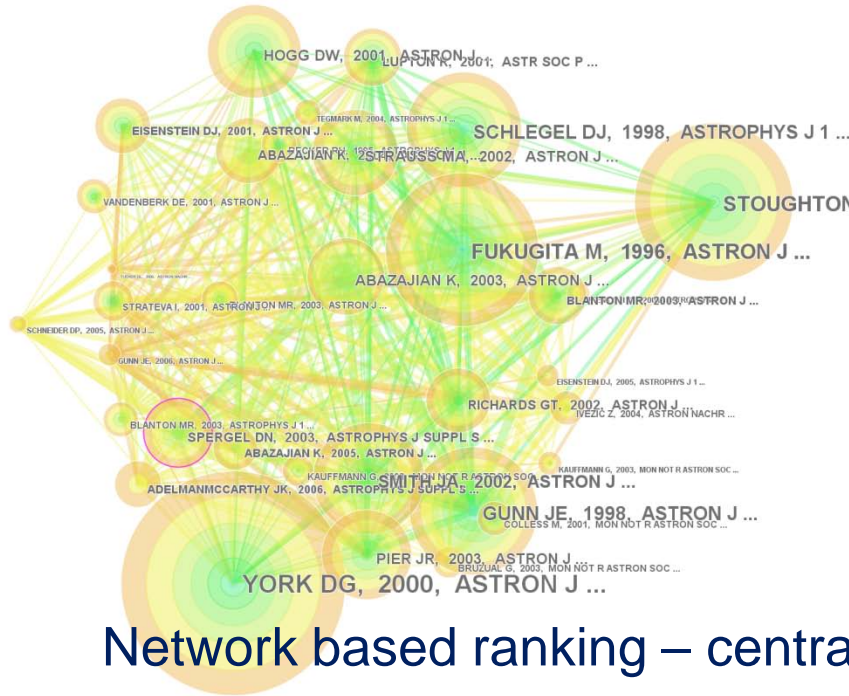
[2.]*Computer Science and Engineering*

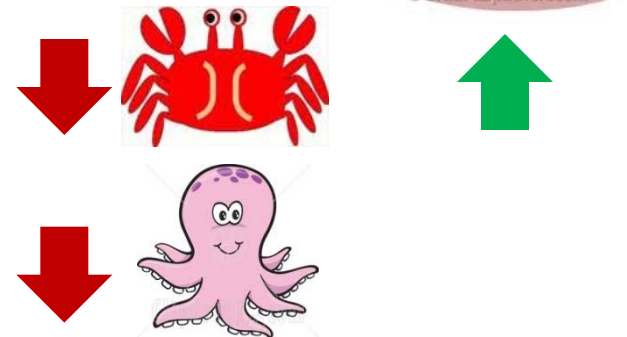[3.] *Department of Statistics*

University of Michigan

Diversity in Ranking

Network based ranking – centrality/prestige

Ranking papers, people, web pages, movies, restaurants…

Web search; ads; recommender systems …
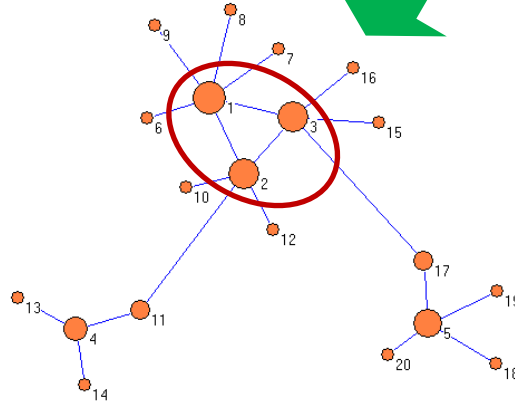
Ranking by Random Walks
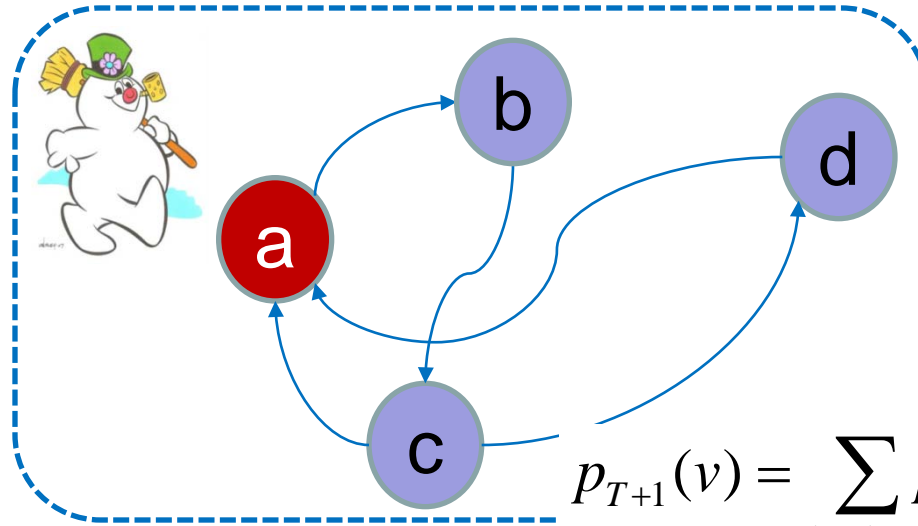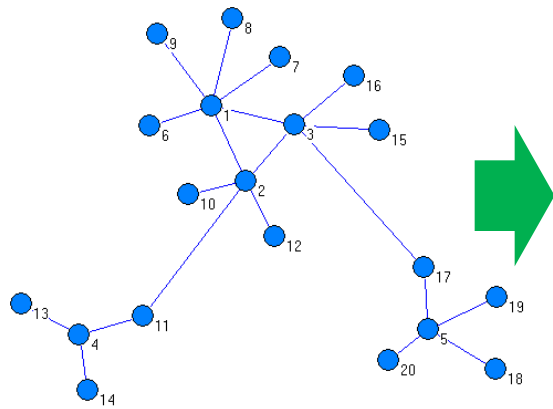
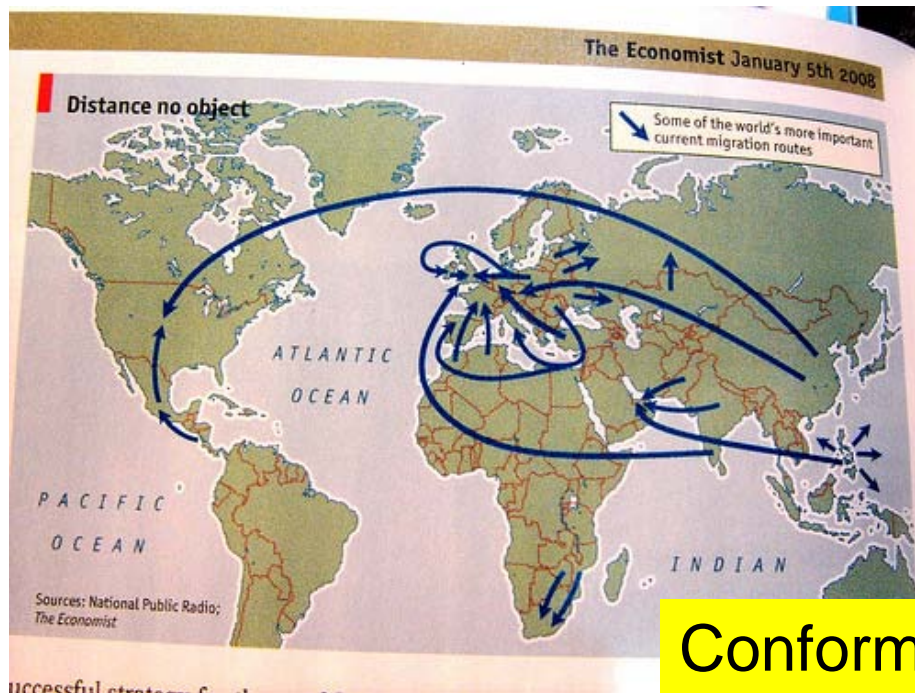$$p_{T+1}(v) = \sum_{(u,v) \in E} p(u,v) \, p_T(u)$$
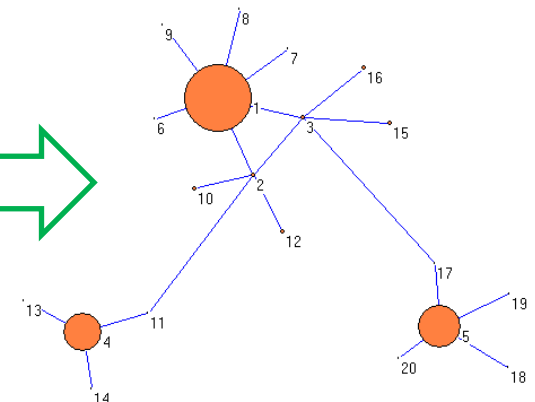
Ranking using
stationary distribution
E.g., PageRank

Reinforcements in Random Walks

- Random walks are not random - rich gets richer;
  - e.g., civilization/immigration – big cities attract larger population;
  - Tourism – busy restaurants attract more visitors;

Conformity!

Source - http://www.resettlementagency.co.uk/modern-world-migration/
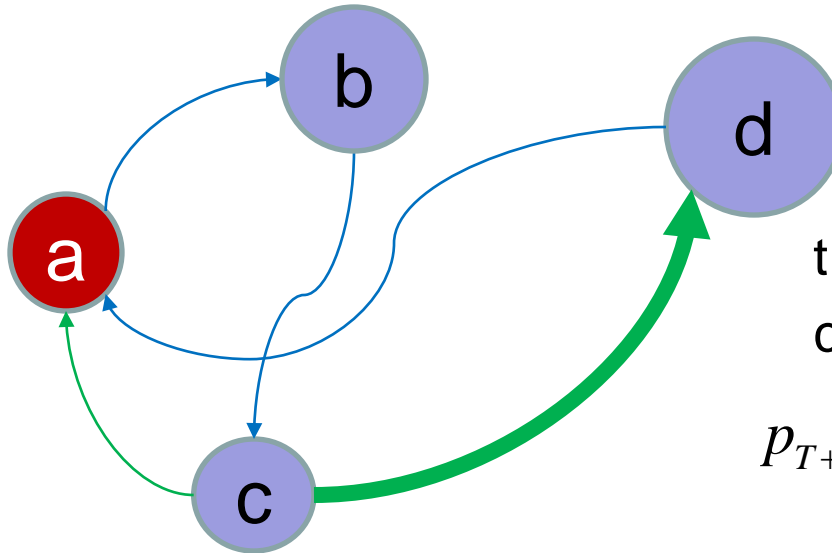
Vertex-Reinforced Random Walk
(Pemantle 92)

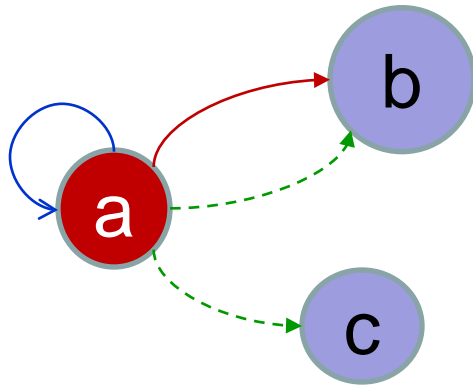transition probabilities
change over time

$$p_{T+1}(v) = \sum_{(u,v) \in E} p_T(u,v) \, p_T(u)$$

<u>Reinforced random walk</u>: transition probability is reinforced by the weight (number of visits) of the target state

$$p_T(u,v) \propto N_T(v)$$

DivRank

- A smoothed version of Vertex-reinforced Random Walk

$$p_T(u, v) = (1 - \lambda)\,p^*(v) + \lambda \cdot \frac{p_0(u, v)\,N_T(v)}{D_T(u)}$$

Random jump, could be personalized

"organic" transition probability

- Adding self-links;
- Efficient approximations: use $E[N_T(v)]$ to approximate $N_T(v)$

Cumulative DivRank:

$$E[N_T(v)] \propto \sum_{t=0}^{T} p_t(v)$$

Pointwise DivRank:

$$E[N_T(v)] \propto p_T(v)$$

Experiments

- Three applications
  - Ranking movie actors (in co-star network)
  - Ranking authors/papers (in author/paper-citation network)
  - Text summarization (ranking sentences)
- Evaluation metrics:
  - diversity: *density of subgraph*; *country coverage (actors)*
  - *quality: h-index (authors); # citation (papers);*
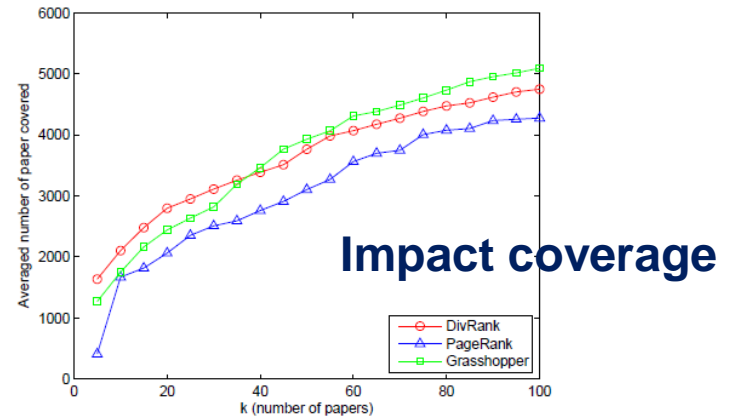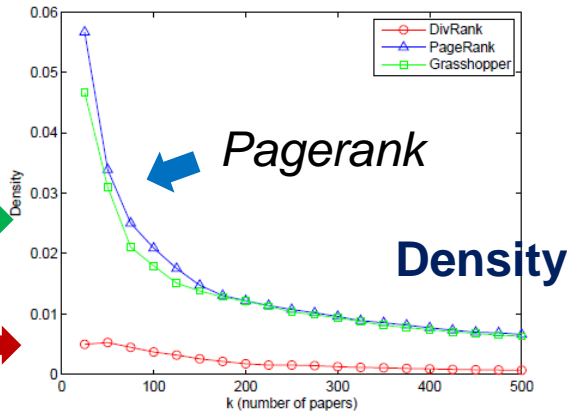 - *quality + diversity: movie coverage (actors); impact coverage (papers); ROUGE (text summarization)*

Results

- *Divrank >> Grasshopper/MMR >> Pagerank*

Paper citation:

Grasshopper →

Divrank →

Pagerank

Density

**Impact coverage**

Text Summarization:

Table 1: Results on DUC04 Task-2.

Method	Training		Testing	
	R-1	95% C.I.	R-1	95% C.I.
LR	0.359	[0.337, 0.381]	0.343	[0.318, 0.366]
PPR	0.378	[0.356, 0.398]	0.368	[0.350, 0.385]
MMR	0.363	[0.347, 0.379]	0.343	[0.318, 0.366]
GH	0.380	[0.360, 0.397]	0.356	[0.333, 0.378]
DR	**0.387**	[0.367, 0.404]	**0.379**	[0.366, 0.394]
CDR	0.384	[0.365, 0.401]	0.362	[0.342, 0.378]

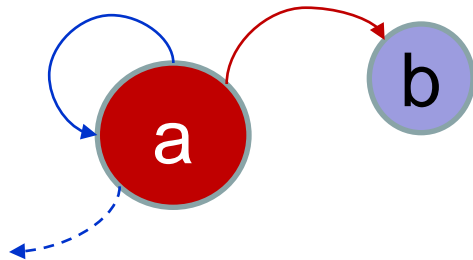University of Michigan School of
iNFORMATION

Why Does it Work?

- Rich gets richer
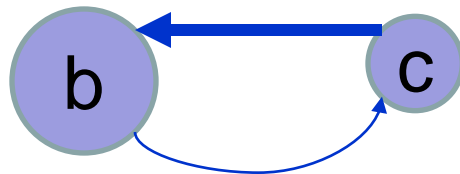
  – Related to *Polya's urn* and *preferential attachment*

- Compete for resource in neighborhood

Stay here or go to neighbors?

  – Prestigious node absorbs weights of its neighbors

- An optimization explanation

Summary

- DivRank – Prestige/Centrality + Diversity
- Mathematical foundation: vertex-reinforced random walk
- Connections:
  - Polya's Urn
  - Preferential Attachments
  - Word burstiness
- Why it works?
  - Rich-gets-richer
  - Local resource competition
- Future work: Query dependent DivRank;

Thanks!

2010 © *University of Michigan*