Improving Diversity in Ranking using Absorbing Random Walks

GRASSHOPPER

Andrew B. Goldberg

with Xiaojin Zhu, Jurgen Van Gael*, and David Andrzejewski

Department of Computer Sciences, University of Wisconsin, Madison

* Department of Engineering, University of Cambridge

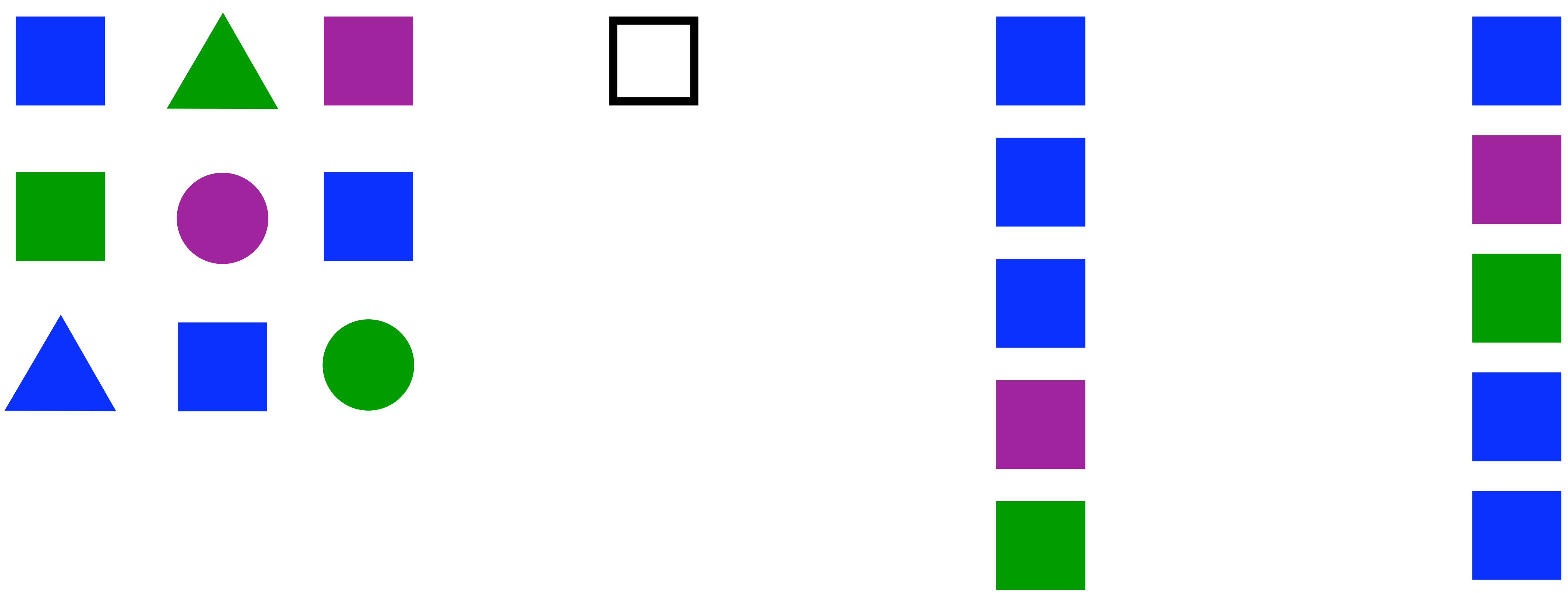
SIGIR IDR Workshop, July 23, 2009

Originally appeared at NAACL-HLT 2007

Problem Description

- Given set of items (e.g., sentences, news articles, people)
- Rank items such that highly ranked items are important but not too similar to each other

Documents: Query: Bad ranking: Good ranking:



Our Contribution

Previous work [MMR (Carbonell and Goldstein, 1998) & CSIS (Radev, 2000)]

- Treat centrality and diversity separately using heuristics
- Post-processing step to eliminate redundancy

Our approach (GRASSHOPPER)

- Use *absorbing* Markov chain random walks to simultaneously rank based on centrality *and* diversity

Diversity in Ranking can be Important

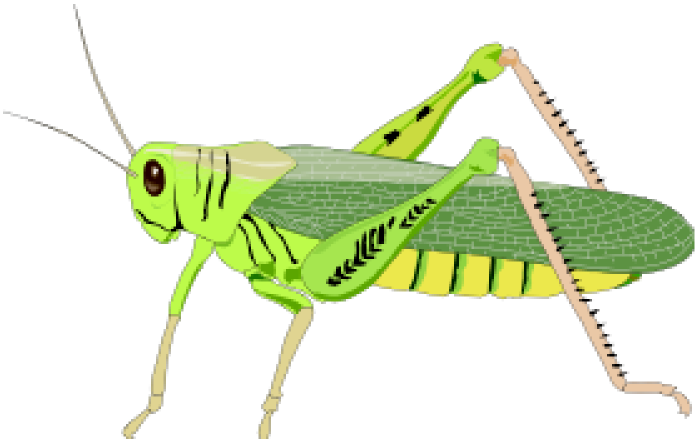
Examples

- Information retrieval
  - Do not want near identical articles
- Extractive text summarization
  - Do not want near identical sentences
- Social network analysis
  - Do not want people all from the same group

Our Algorithm

GRASSHOPPER = Graph Random-walk with Absorbing StateS that HOPs among PEaks for Ranking

- Centrality
- Diversity
- Prior ranking



Inputs to GRASSHOPPER

Three inputs

- Graph W
  - Relationships between the items to rank
- Prior ranking r (as a probability vector)
  - Prior knowledge of item importance
  - Uniform if no information is available
- Weight $\lambda \in [0, 1]$ to balance the two

The Graph W

To rank n items, we use an $n \times n$ weight matrix W

- w_{ij} the weight on edge from item i to item j
- Directed or undirected
- Non-negative weights ($w_{ij} = 0$ if i, j not connected)
- Self edges are ok
- Examples: cosine similarity between documents i, j; number of phone calls i made to j, etc.

The Prior Ranking r

Optional. Used for re-ranking.

- Probability vector $r = (r_1, \ldots, r_n)^\top$, $r_i \geq 0$, $\sum_i r_i = 1$
- r_i = P(picking item i as the most important item)
- More control than ranking: $(0.1, 0.7, 0.2)^\top$ vs. $(0.3, 0.37, 0.33)^\top$
- Uniform if no prior ranking
- Examples: (IR) document relevance scores to query; (summarization) position of sentence, etc.

Finding the First Item

Same as PageRank (Page et al., 1998)

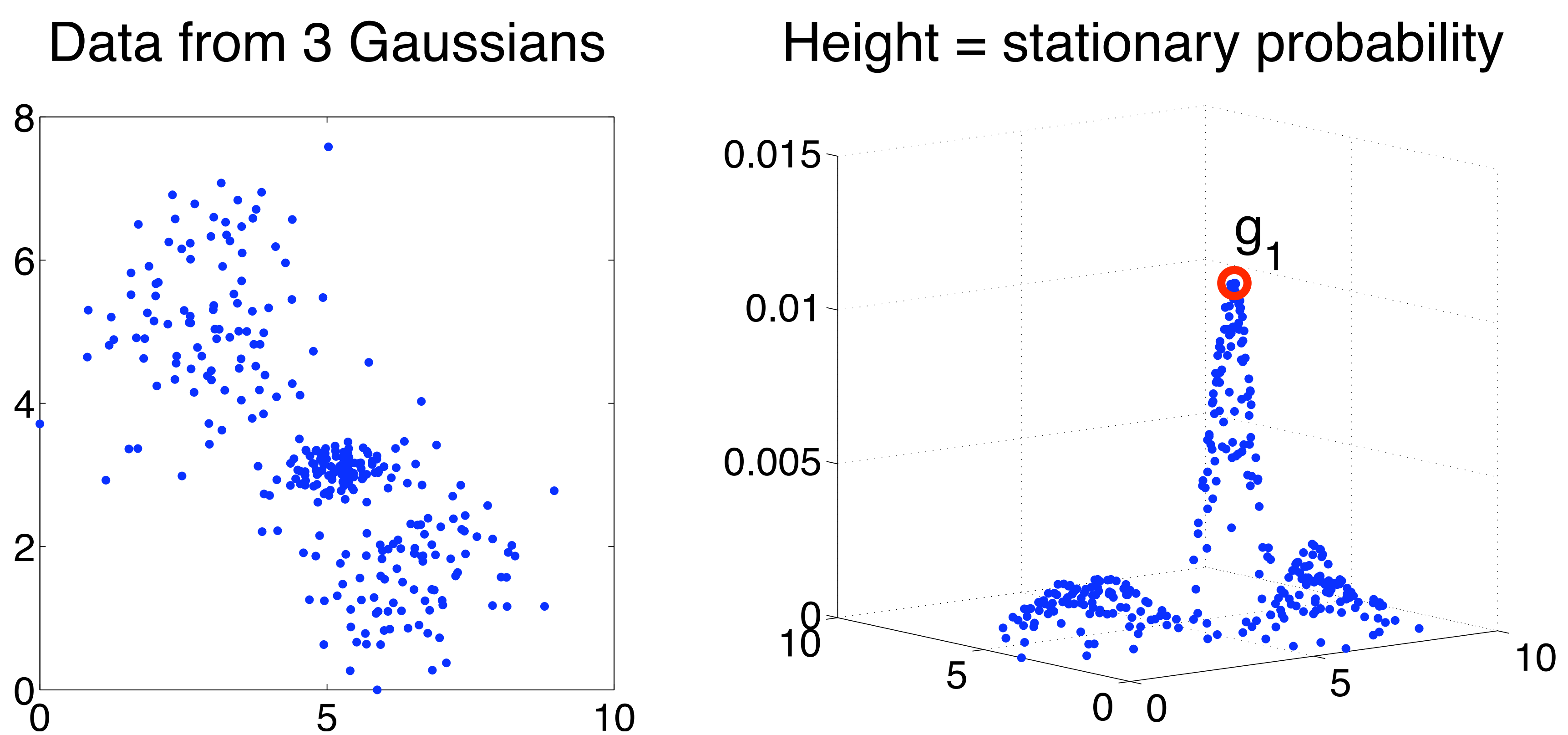
- Teleporting random walk
- Raw transition matrix $\tilde{P}$: $\tilde{P}_{ij} = w_{ij} / \sum_k w_{ik}$
- Teleporting transition matrix $P = \lambda \tilde{P} + (1 - \lambda) 1 r^\top$
- Stationary distribution $\pi = P^\top \pi$
- $g_1 = \arg\max_i \pi_i$

But, PageRank does not address diversity at all

(Note: First item can instead be specified by user.)

Finding the First Item

Toy example



- Problem: Stationary distribution lacks diversity

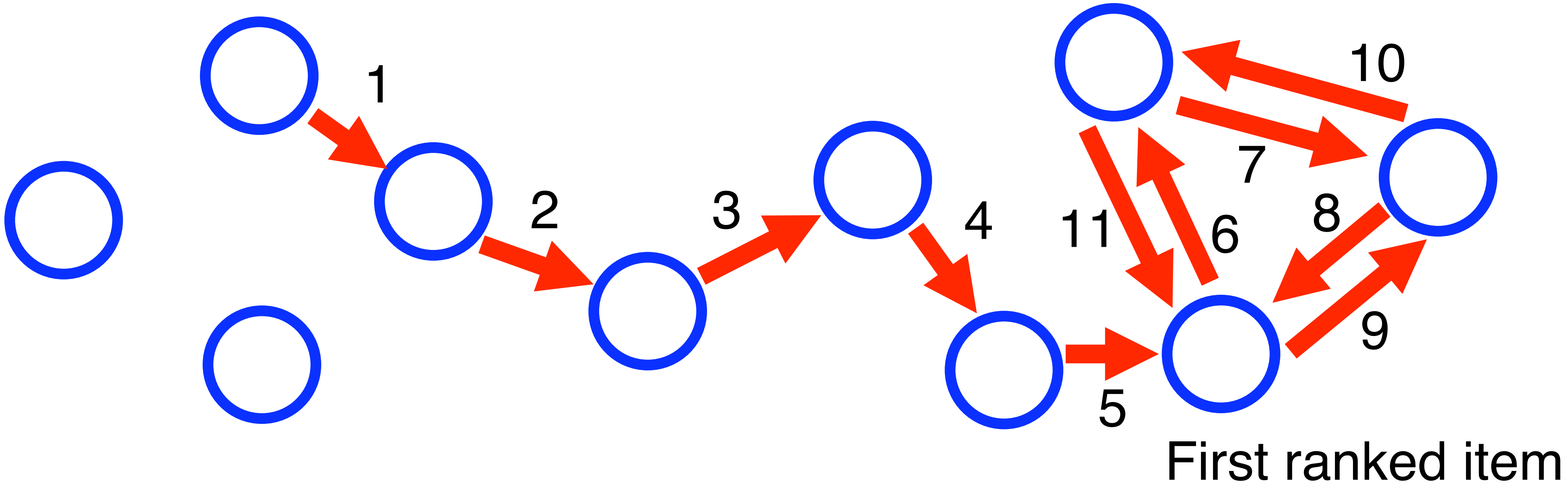
Ranking the Remaining Items

Idea: Turn ranked items into absorbing states

- Random walk ends when reaches absorbing state
- No more stationary distribution
- Rank by *expected number of visits* per walk

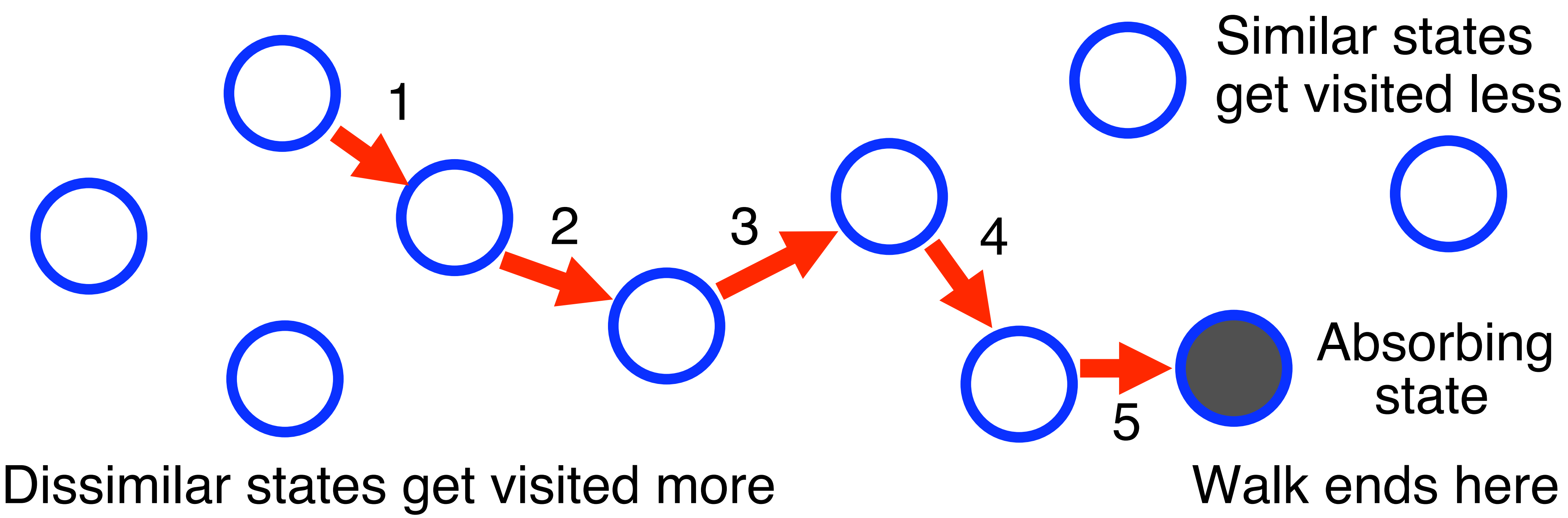
Absorbing States Illustrated

Initial random walk with no absorbing states



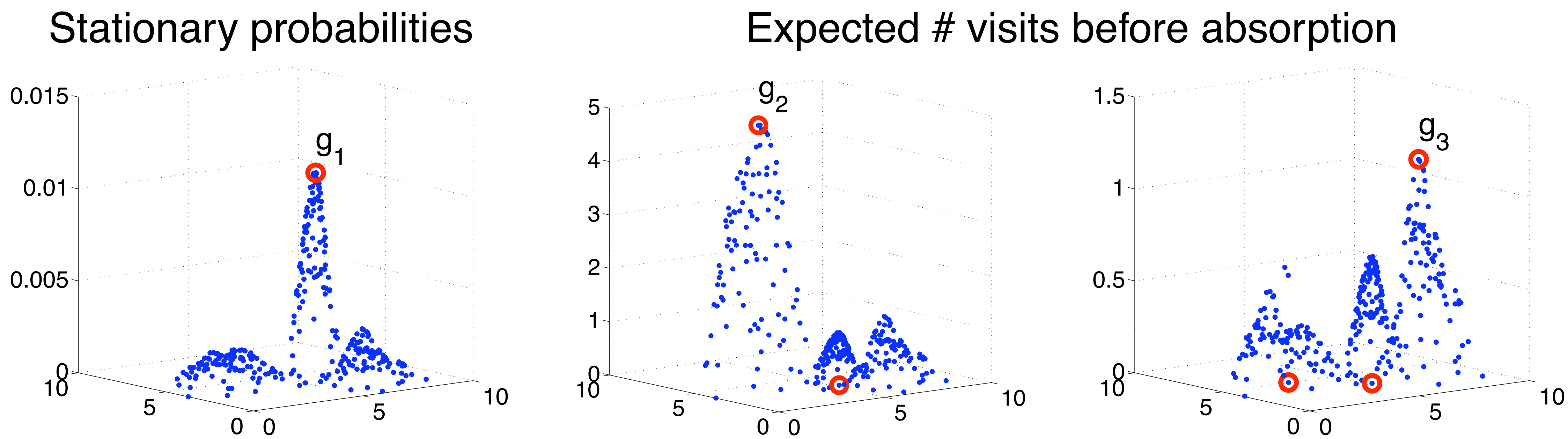
Absorbing States Illustrated

Absorbing random walk after ranking first item



Ranking the Remaining Items

GRASSHOPPER hops!



- States similar to absorbing states get fewer visits
- Next item ranked is diverse w.r.t. already ranked items

Expected Number of Visits

- For $g \in G$ (set of absorbing states), $P_{gg} = 1$, other entries 0
- Rearrange $P = \begin{array}{c} \left[\begin{array}{cc} I & 0 \\ R & Q \end{array} \right] \end{array} \begin{array}{c} \text{absorbing} \\ \text{non-absorbing} \end{array}$
- Fundamental matrix $N = (I - Q)^{-1}$
  - N_{ij}: if random walk starts from i, the expected number of visits to j, before absorption
- Average over starting states: $v = N^\top 1/(n - |G|)$
  - v_j: expected number of visits to j regardless of start
- Next ranked item: $g_{|G|+1} = \arg\max_i v_i$

Discussions

- Similar to the heuristic “cluster, take center in turn”
  - But no actual clustering is performed
  - No need to know *a priori* how many clusters exist
- Fast computation: Matrix inversion lemma
  - Initial fundamental matrix computed with real inverse
  - Follow-up inverses derived using simple updates

NLP Experiments

Multi-document extractive text summarization

- 2004 Document Understanding Conference data sets
- Items: all sentences in all documents on a topic
- Graph W with sparse (thresholded) cosine edges
- Prior ranking based on sentence position: $r_i \propto {p_i}^{-\alpha}$
- $\lambda = 0.5, \alpha = 0.25$ tuned on DUC 2003 data

Results on DUC 2004 Tasks

Evaluation based on ROUGE-1 metric (Lin and Hovy, 2003)

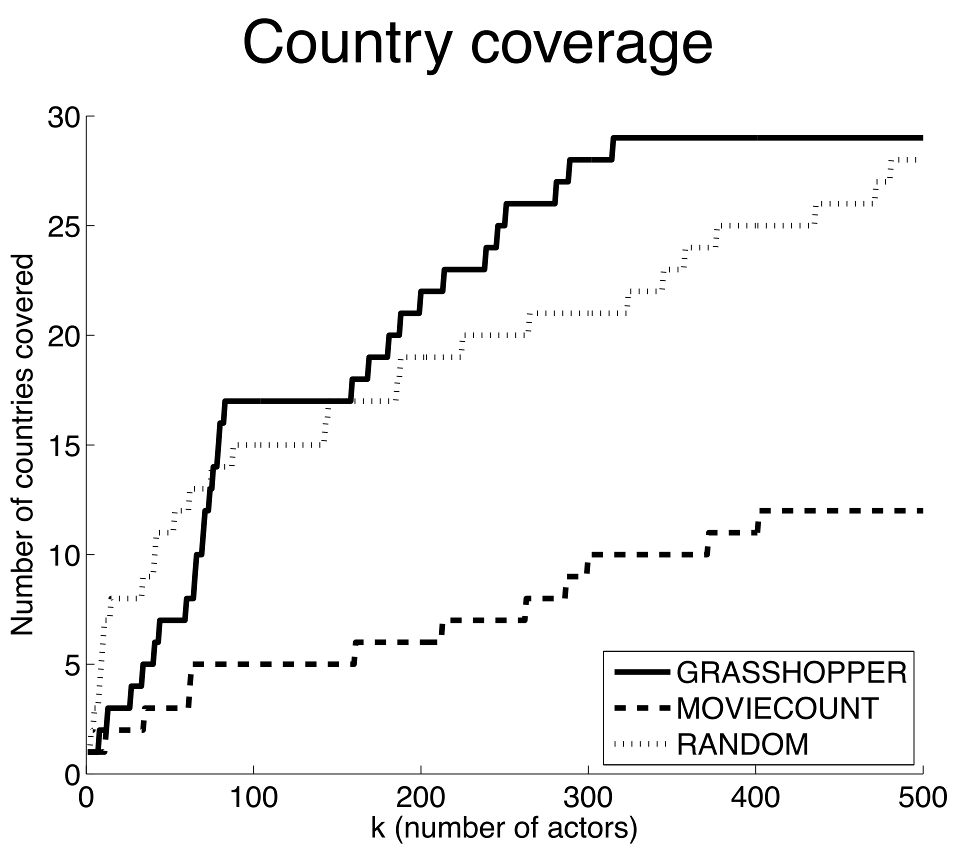
- Task 2: Between 1st & 2nd of 34 DUC competitors
- Task 4a: Between 5th & 6th of 11
- Task 4b: Between 2nd & 3rd of 11

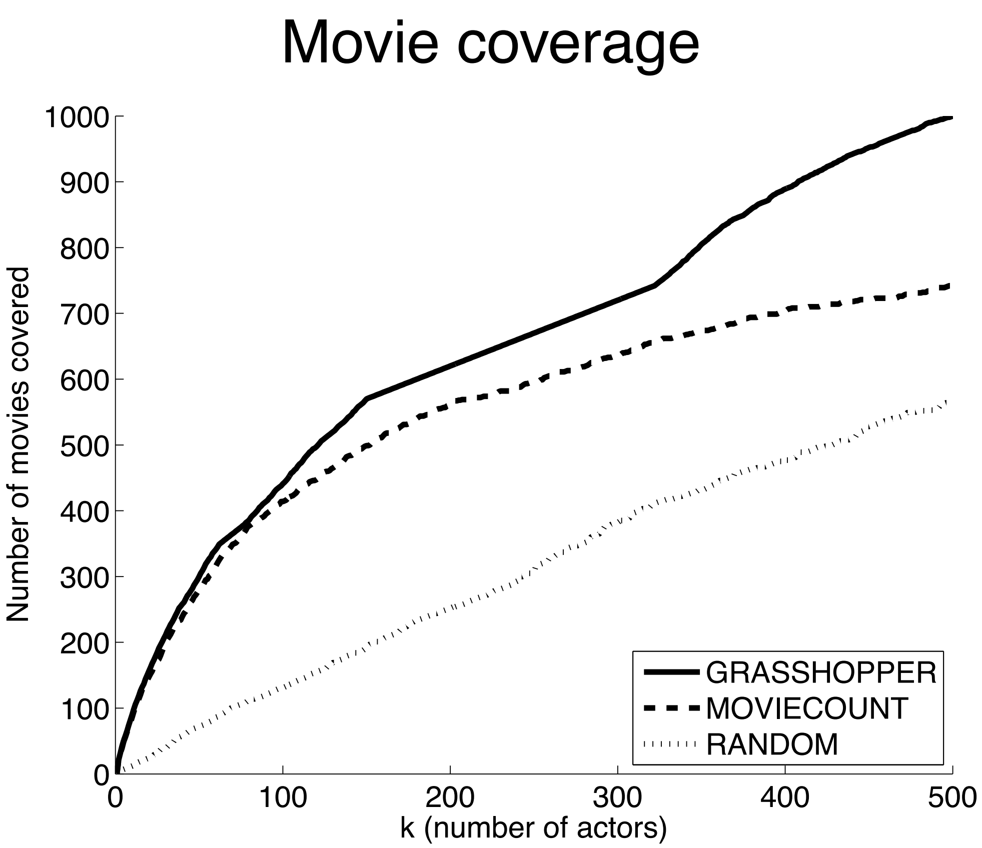
Social Network Experiments

- Goal: Rank prominent comedy stars from diverse countries
- Used 3000+ actors from films in 2000–2006

Movie Count Ranking			GRASSHOPPER Ranking		
1.	Ben Stiller	USA	1.	Ben Stiller	USA
2.	Anthony Anderson	USA	2.	Anthony Anderson	USA
3.	Eddie Murphy	USA	3.	Johnny Knoxville	USA
	...			...	
12.	Gerard Depardieu	France	8.	Gerard Depardieu	France
35.	Mads Mikkelsen	Denmark	13.	Mads Mikkelsen	Denmark
62.	Til Schweiger	Germany	27.	Til Schweiger	Germany
161.	Tadanobu Asano	Japan	34.	Tadanobu Asano	Japan
287.	Kjell Bergqvist	Sweden	44.	Kjell Bergqvist	Sweden

Social Network Analysis Results





- GRASSHOPPER highly ranks actors representing many diverse countries and who starred in many different movies

Conclusion

GRASSHOPPER ranking

- Centrality + diversity + prior in a unified framework of absorbing random walks

Download code:
http://www.cs.wisc.edu/~jerryzhu/pub/grasshopper.m