DivRank: Interplay of Prestige and Diversity in Information Networks

Qiaozhu Mei\textsuperscript{1,2}, Jian Guo\textsuperscript{3}, Dragomir Radev\textsuperscript{1,2}

\textsuperscript{1}School of Information
\textsuperscript{2}Computer Science and Engineering
\textsuperscript{3}Department of Statistics
University of Michigan
Diversity in Ranking

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

Web search; ads; recommender systems ...

Network based ranking – centrality/prestige
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;

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

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

\[ p_T(u, v) \propto N_T(v) \]

Transition probabilities change over time

\[ p_{T+1}(v) = \sum_{(u,v) \in E} p_T(u,v)p_T(u) \]
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:**

**Density Impact coverage**

**Text Summarization:**

<table>
<thead>
<tr>
<th>Method</th>
<th>Training R-1</th>
<th>95% C.I.</th>
<th>Testing R-1</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>0.359</td>
<td>[0.337, 0.381]</td>
<td>0.343</td>
<td>[0.318, 0.366]</td>
</tr>
<tr>
<td>PPR</td>
<td>0.378</td>
<td>[0.356, 0.398]</td>
<td>0.368</td>
<td>[0.350, 0.385]</td>
</tr>
<tr>
<td>MMR</td>
<td>0.363</td>
<td>[0.347, 0.379]</td>
<td>0.343</td>
<td>[0.318, 0.366]</td>
</tr>
<tr>
<td>GH</td>
<td>0.380</td>
<td>[0.360, 0.397]</td>
<td>0.356</td>
<td>[0.333, 0.378]</td>
</tr>
<tr>
<td>DR</td>
<td><strong>0.387</strong></td>
<td><strong>[0.367, 0.404]</strong></td>
<td><strong>0.379</strong></td>
<td><strong>[0.366, 0.394]</strong></td>
</tr>
<tr>
<td>CDR</td>
<td>0.384</td>
<td>[0.365, 0.401]</td>
<td>0.362</td>
<td>[0.342, 0.378]</td>
</tr>
</tbody>
</table>
Why Does it Work?

- Rich gets richer
  - Related to *Polya’s urn* and *preferential attachment*
- Compete for resource in neighborhood
  - 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!