DivRank: the Interplay of Prestige and Diversity in Information Networks
Qiaozhu Mei, Jian Guo, Dragomir Radev

Maggie Zhou
COMP 790 Data Mining Seminar, Spring 2011
Outline

• **Background**
  – What problem is DivRank solving?
  – What solution exists already? Why is suboptimal?
  – Example

• DivRank

• Other Models

• Experimental Comparisons

• Summary
Background: Problem Statement

• Many models primarily consider *prestige* in ranking results
  – *Prestige*: data items that are referred to by many items / connected to many items are more prestigious

• Diversity in results is also useful to the user
  – i.e. restaurant recommendations
Background: Example

Figure 1: An illustration of diverse ranking in a toy network.
Outline

• Background
• DivRank
  – Intuition & Principles
  – Form & Optimization Argument
• Other Models
• Experimental Comparisons
• Summary
DivRank: Intuition & Principles

• Mathematically: DivRank is a vertex-reinforced random walk
  – *Random walk*: Markov chain in the given network with each vertex represents a state and a walk moves from state to state based on a transition probability distribution
  – *Vertex-reinforced random walk*: Transition probability to one state is reinforced by the number of previous visits to that state
    • Ex. Actor accumulates prestige when acting in more movies, which gives the actor more opportunities
DivRank: Intuition & Principles

• In contrast: PageRank enforces regularization (transition probabilities do not change over time)

  – Ex. So, a movie actor has equally high prestige at the beginning of the career as at the end (PageRank assumes a true theoretical value that it’s aiming to find)
DivRank: Form & Optimization

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

\[ D_T(u) = \sum_{v \in V} p_0(u, v) N_T(v). \]

Idea: As random walk starts, nodes with a higher degree will get a higher weight, which results in a higher accumulative number of times visited weight \((N)\)

Reinforcement of probability of staying at current state based on number of times visited
Outline

• Background
• DivRank

• Other Models
  – PageRank (2001)
  – Grasshopper (2007)
  – MMR (Maximum Marginal Relevance) (1998)

• Experimental Comparisons
• Summary
Other Models: PageRank Revisited

• PageRank vs. DivRank: transition probabilities
• PageRank: smoothed stationary distribution to rank web pages

• *Smoothing*: This distribution is going to assign higher weights to vertices that are more prestigious (so a prestigious node’s neighbors are also likely to be visited in the random walk)

• DivRank differs because in addition to smoothing it has a *competing* element between the vertices
Other Models: Grasshopper

• Greedy approach that penalizes nodes for being visited recently.
  • Ex. **Green**: next selected node into “absorption set”.
  **Red**: absorption set, whose vertices aren’t used in running the random walk.
Other Models: MMR

• MMR: Maximum Marginal Relevance (1998)
  – Greedy vertex selection with diversity as aim
  – Selects most prestigious vertex & penalizes vertices already covered

• MMR vs. Grasshopper:
  – MMR compares previously selected vertices to remaining vertices using ‘similarity index’,
  – Grasshopper penalizes vertices around previously selected ones in the random walk
Outline

• Background
• DivRank
• Other Models
• Experimental Comparisons
  – Methodology
  – Results
• Summary
Experimental Methodology

• “there is no evaluation metric that seems to be universally accepted as the best for measuring the performance of algorithms that aim to obtain diverse rankings.” –SIGIR 2009

• Their ranking: They assume the density of the subgraph of top-ranked vertices is an inverse measure of diversity.
  – Density: number of edges in a network divided by the maximum number of edges in the network
Experimental Results

Comparison of Diversity Rankings:

![Graphs showing comparison of diversity rankings.](image)

Figure 2: Comparison of network-based ranking methods in ranking IMDb stars. Parameters: $\lambda$ (or $d$) = 0.9 in PageRank, Personalized PageRank, DivRank, Cumulative DivRank, and Grasshopper; $\alpha = 0.25$ in DivRank and Cumulative DivRank.

No good metric for both prestige and diversity.
Outline

• Background
• DivRank
• Other Models
• Experimental Comparisons
• Summary
Summary

• DivRank includes more diversity than other methods, without sacrificing prestige

• Questions:
  – How much diversity is important?
  – How much prestige is important?
    • i.e. if the best result is 3\textsuperscript{rd} or 4\textsuperscript{th} down, instead of first, does it matter?