Explaining & Reformulating Authority Flow Queries

Ramakrishna Varadarajan
Vagelis Hristidis
School of Computing and Information Sciences,
FLORIDA INTERNATIONAL UNIVERSITY
Miami
Louïqa Raschid
UNIVERSITY OF MARYLAND, COLLEGE PARK
Roadmap

• Motivation
• Explaining Query Results
• Query Reformulation
• Experimental Results
• Related Work
• Conclusions
Roadmap

• Motivation
• Explaining Query Results
• Query Reformulation
• Experimental Results
• Related Work
• Conclusions
Motivation – Authority Flow Queries

• Authority Flow – Effective Ranking Mechanism

• Authority originates from the authority sources and flows according to the semantic connections.

• Follows the Random Surfer Model.

• At any time step, the random surfer either:
  – Moves to an adjacent node
  – Randomly jumps to some node (different in Personalized PageRank and ObjectRank)

• Applications:
  – Web [unstructured] (PageRank, Personalized-PageRank)
  – Databases [structured] (ObjectRank)
• **Data Graph of Entities**

• **ObjectRank** Ranks Objects According to Probability of Reaching Result Starting from Base Set
Motivation - ObjectRank

Authority Transfer Data Graph (Keyword Query: [OLAP])

Paper H. Gupta et al.  
Index Selection for OLAP  
ICDE 1997

Conference ICDE  
Year 1997

Paper C. Ho et al.  
Range Queries in OLAP Data Cubes  
SIGMOD 1997

Paper J. Gray et al.  
Data Cube: A Relational…  
ICDE 1996

Paper R. Agrawal et al.  
Modeling Multidimensional Databases  
ICDE 1997

Author R. Agrawal

Base Set

cites 0.7  
contains 0.3  
contained 0.1  
has instance 0.3  
0.3

Schema Graph

Ramakrishna Varadarajan, Florida International University (FIU)
Motivation

**Limitations of ObjectRank** :

- No way to *explain* to the user why a particular result received its current score.
- Authority transfer rates have to be set manually by a domain expert.
- No *query reformulation* methodology to refine results.

**ObjectRank2** (*Slight modification of ObjectRank*)

- Random Surfer jumps to different nodes of base set with different probabilities.
- Probability for a node $v$ is proportional to $IRScore(v, Q)$
Roadmap

• Motivation
• Explaining Query Results
• Query Reformulation
• Experimental Results
• Related Work
• Conclusions
Explaining Query Results

- **Problem** – Given a target object $T$, explain to user why it received a high score.

- **Our Solution** – Display an explaining subgraph of Authority transfer data graph, for $T$.

- Explaining subgraph contains:
  - All Edges that transfer authority to $T$.
  - Edges are annotated with amount of authority flow.

- Done in two stages:
  - **Subgraph Construction Stage**
    - Bidirectional Breadth-First Search
  - **Authority Flow Adjustment Stage**
    - Adjust original authority flows – more challenging
Explaining Query Results – Explaining Subgraph

• Target Object – “Modeling Multidimensional databases” paper.

**Explaining Subgraph Creation**

1. Perform a BFS search in reverse direction from the target object.
2. Perform a BFS search in forward direction from base set objects (authority sources).
3. Subgraph will contain all nodes/edges traversed in the forward direction.
4. Compute the explaining authority flow along each edge by eliminating the authority leaving the subgraph (iterative procedure).
Roadmap

- Problem Statement & Motivation
- Explaining Query Results
- Query Reformulation
- Experimental Results
- Related Work
- Conclusions
Query Reformulation

Motivation

- Content-based Reformulation - Well studied in Traditional IR (Salton, Buckley 1990)
- Query Expansion is Dominant strategy
- No Method to Reformulate based on Link-Structure and Authority Flow Bounds.

STEPS:

1) System computes Top-\(k\) objects with high ObjectRank2 scores.
2) User marks relevant objects.
3) Compute explaining subgraph of feedback objects.
4) Reformulate based on (a) Content (b) Structure.
   - Content Reformulation based on traditional IR techniques on explaining subgraph
   - Structure Reformulation Achieved by Adjusting Authority Flow Bounds
5) Practically diameter is limited to a constant (L=3).
Roadmap

• Problem Statement & Motivation
• Explaining Query Results
• Query Reformulation
• Experimental Results
• Related Work
• Conclusions
Experimental Results – Internal Survey

- Dataset: DBLP (Nodes - 876,110 & Edges - 4,166,626)
- Query Reformulation types tested:
  - Content-based Reformulations ($C_f=0.0$ & $C_e=0.2$).
  - Structure-based Reformulations ($C_f=0.5$ & $C_e=0.0$).
  - Content & Structure-based Reformulations ($C_f=0.5$ & $C_e=0.2$).
- 2 stages of experiments:
  - Evaluate Reformulation types (User Surveys using residual collection method).
  - Evaluate how close the trained authority transfer bounds are to the ones set by domain experts in ObjectRank [VLDB04].

(a) **Average Precision**

(b) **Training transfer rates**
Experimental Results – External Survey

- External Survey – using only structure-based reformulation (as it performs the best).

- 5 iterations; 20 queries; 10 users.

(a) **Average Precision**  
(b) **Training transfer rates**

![Graph showing average precision and training transfer rates](image)

Ramakrishna Varadarajan, Florida International University (FIU)
Roadmap

• Problem Statement & Motivation
• Explaining Query Results
• Query Reformulation
• Experimental Results
• Related Work
• Conclusions
Related Work

1) Link-Based Semantics
- HITS [ACM Journal 99].
- ObjectRank for the database [VLDB02].
- XRANK [SIGMOD03] for XML databases.

2) Relevance Feedback & Query Reformulation
- Salton, Buckley introduced Relevance feedback [InformationSciences 90].
- Term selection, re-weighting, query expansion [SIGIR94, TREC95].
- Ruthven, Lalmas - Complete Relevance feedback Survey [know. Engg 2003]
- RF based on web-graph distance metrics [SIGIR06]
- Query-independent techniques to assign propagation factors -Nie et al. [WWW2005], Agarwal et al. [SIGKDD2006]
Roadmap

• Problem Statement & Motivation
• Explaining Query Results
• Query Reformulation
• Experimental Results
• Related Work
• Conclusions
Conclusions

• Efficient techniques to explain & reformulate authority flow query results were presented.

• Reformulation was based on (a) Content (b) Structure of the explaining subgraph.

• Techniques to automatically train authority transfer rates were presented.

• User Surveys were conducted to evaluate the effectiveness of the proposed techniques.
Thank You !!!

Questions ???