Large Graph Mining: Power Tools and a Practitioner's guide

Task 4: Center-piece Subgraphs

Faloutsos, Miller and Tsourakakis

CMU

Outline

- Introduction – Motivation
- Task 1: Node importance
- Task 2: Community detection
- Task 3: Recommendations
- Task 4: Connection sub-graphs
- Task 5: Mining graphs over time
- …
- Conclusions

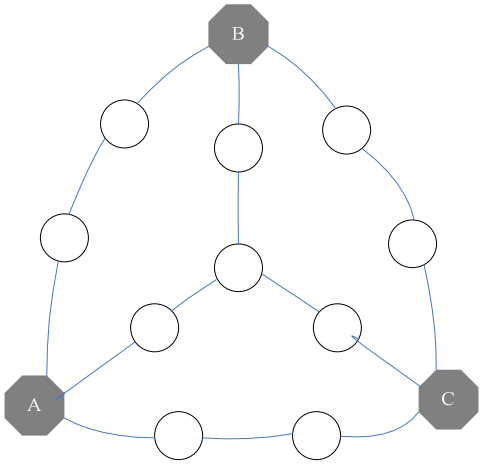
Detailed outline

- Problem definition
- Solution
- Results

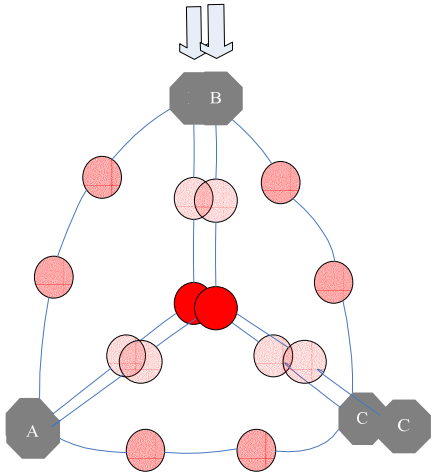
H. Tong & C. Faloutsos *Center-piece subgraphs: problem definition and fast solutions*. In KDD, 404-413, 2006.

Center-Piece Subgraph(Ceps)

- Given Q query nodes
- Find Center-piece ($\leq b$)



- Input of Ceps
  - Q Query nodes
  - Budget b
  - k softAnd number
- App.
  - Social Network
  - Law Inforcement
  - Gene Network
  - …



Challenges in Ceps

- **Q1: How to measure importance?**

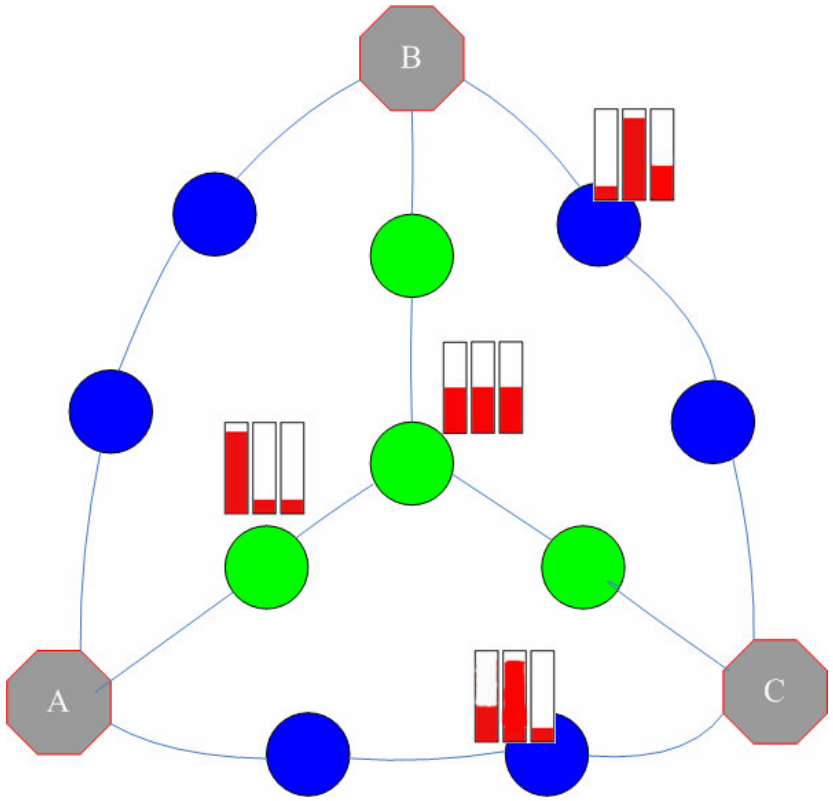
- (Q2: How to extract connection subgraph?
- Q3: How to do it efficiently?)

Challenges in Ceps

- **Q1: How to measure importance?**
- A: "proximity" – but how to combine scores?
- (Q2: How to extract connection subgraph?
- Q3: How to do it efficiently?)

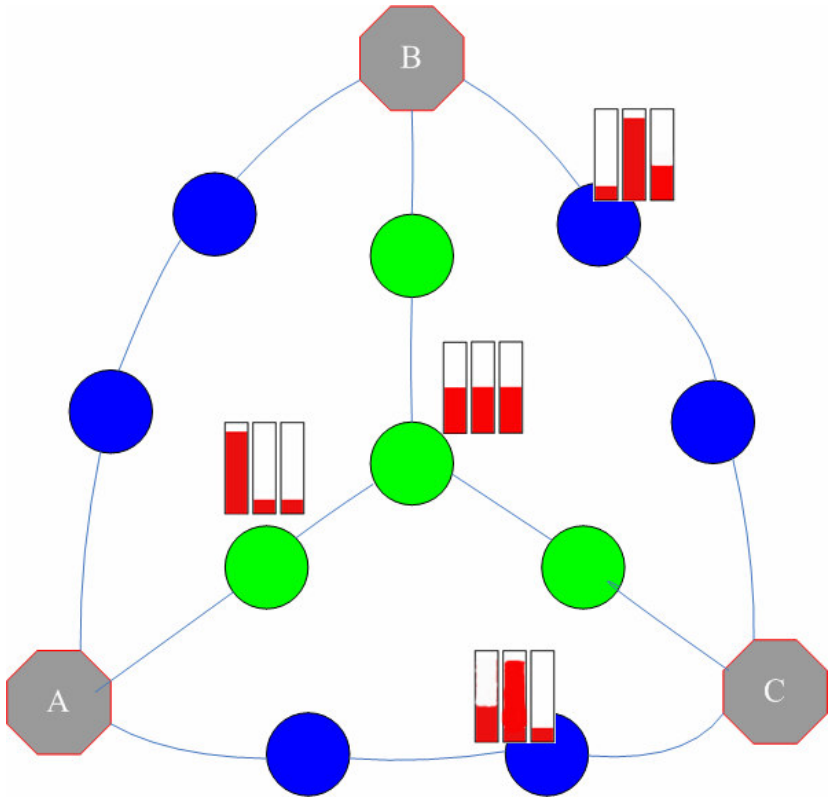
AND: Combine Scores

- Q: How to combine scores?

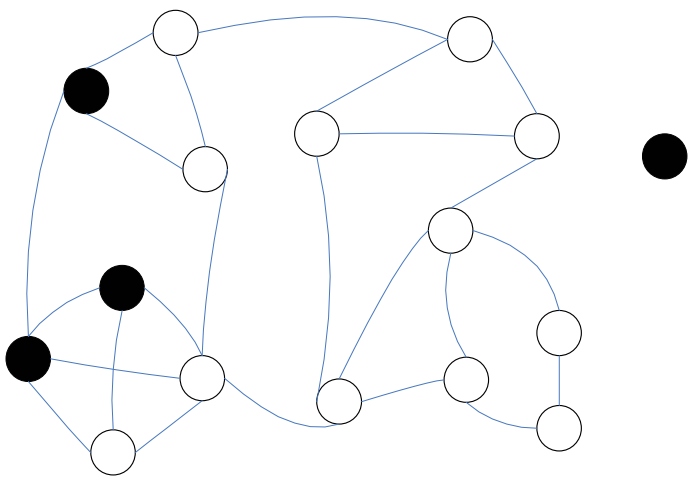


AND: Combine Scores

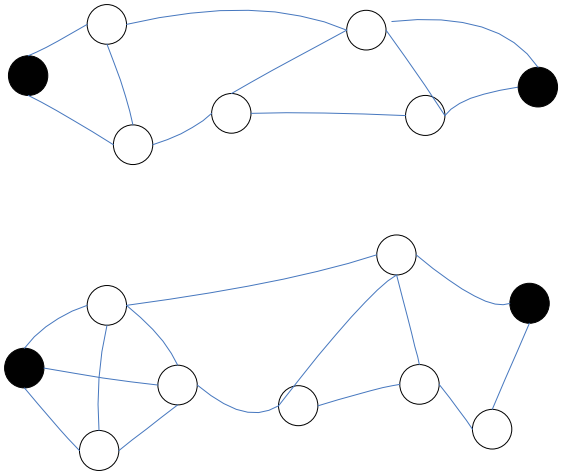
- Q: How to combine scores?
- A: Multiply
- …= prob. 3 random particles coincide on node *j*



K_SoftAnd: Relaxation of AND



Noise



Disconnected Communities

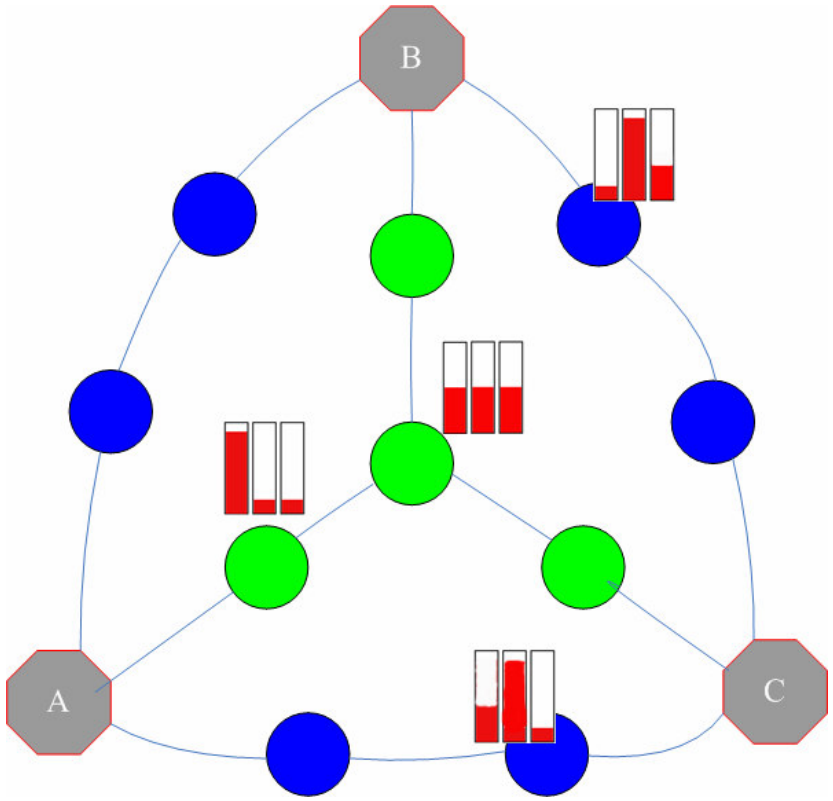
What if AND query ➔ No Answer?

K_SoftAnd: Combine Scores

Generalization – SoftAND:

We want nodes close to k of Q ($k<Q$) query nodes.

Q: How to do that?



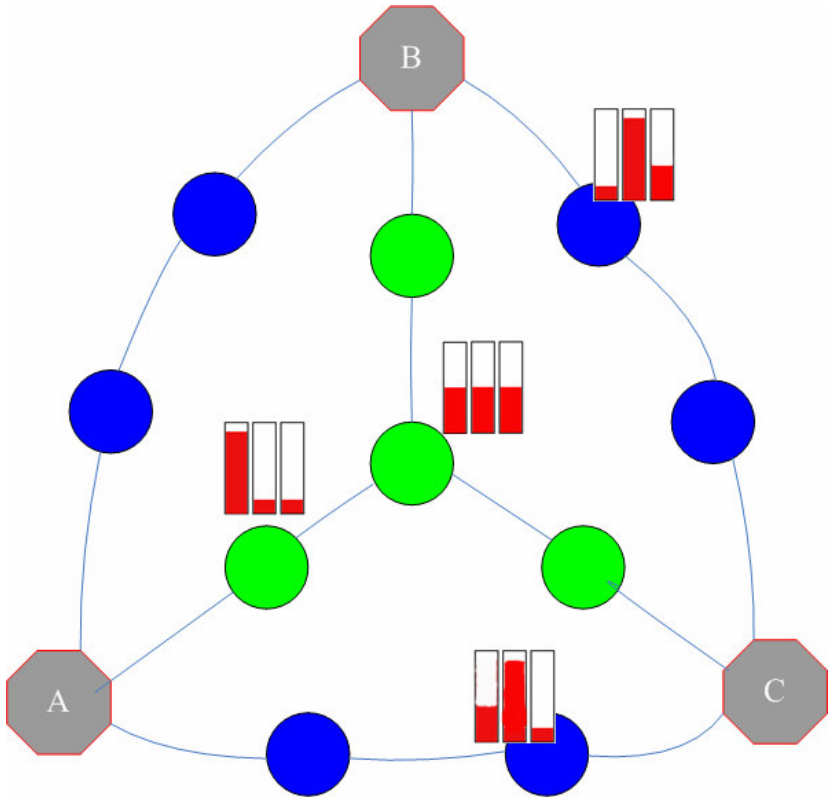
K_SoftAnd: Combine Scores

Generalization – softAND:

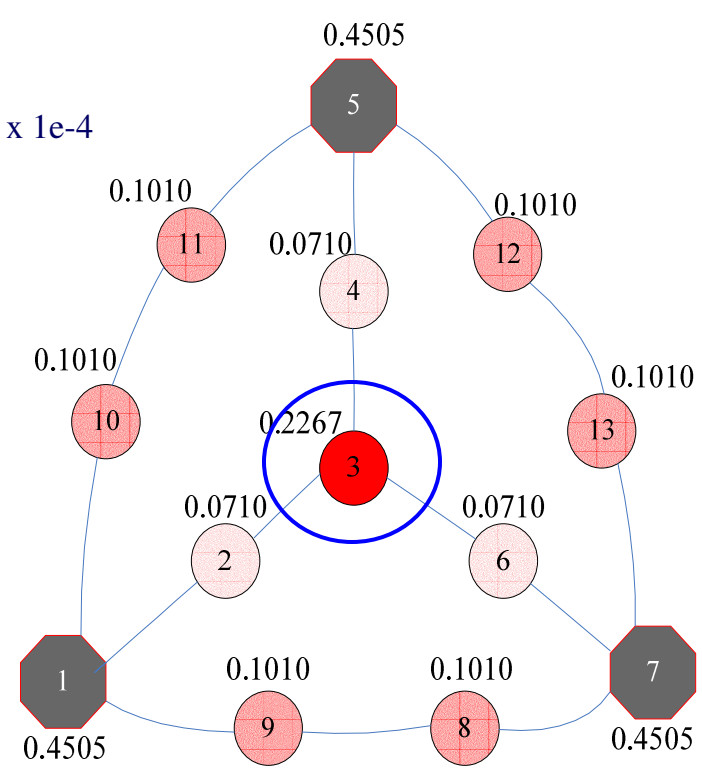
We want nodes close to k of Q ($k<Q$) query nodes.

Q: How to do that?

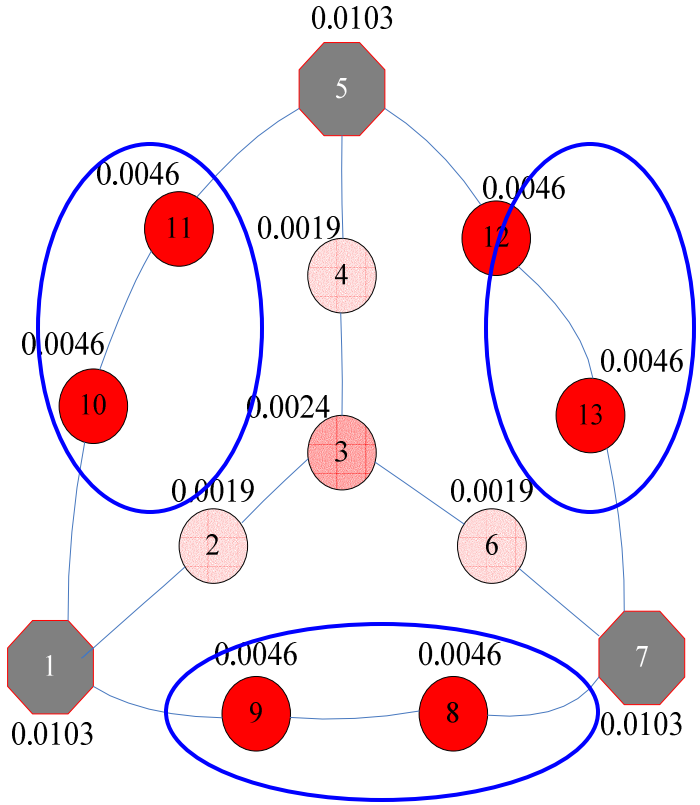
A: Prob(at least k-out-of-Q will meet each other at j)



AND query vs. K_SoftAnd query

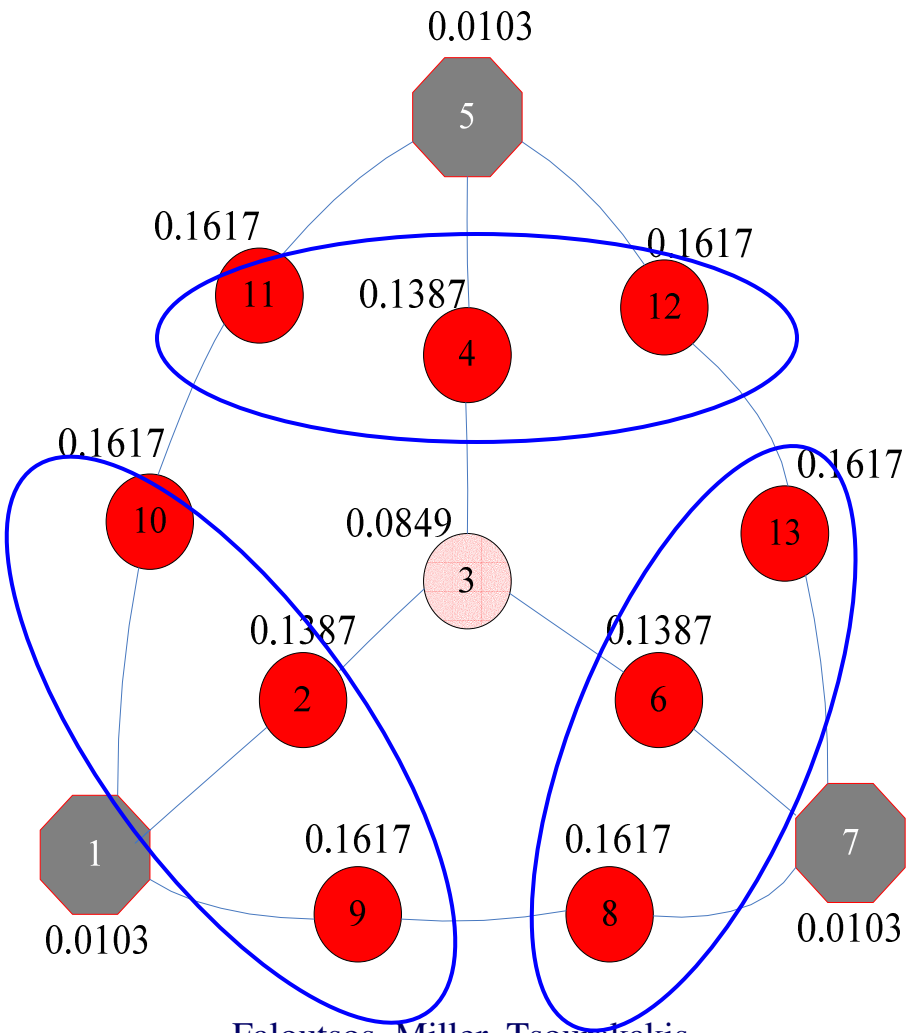


And Query



2_SoftAnd Query

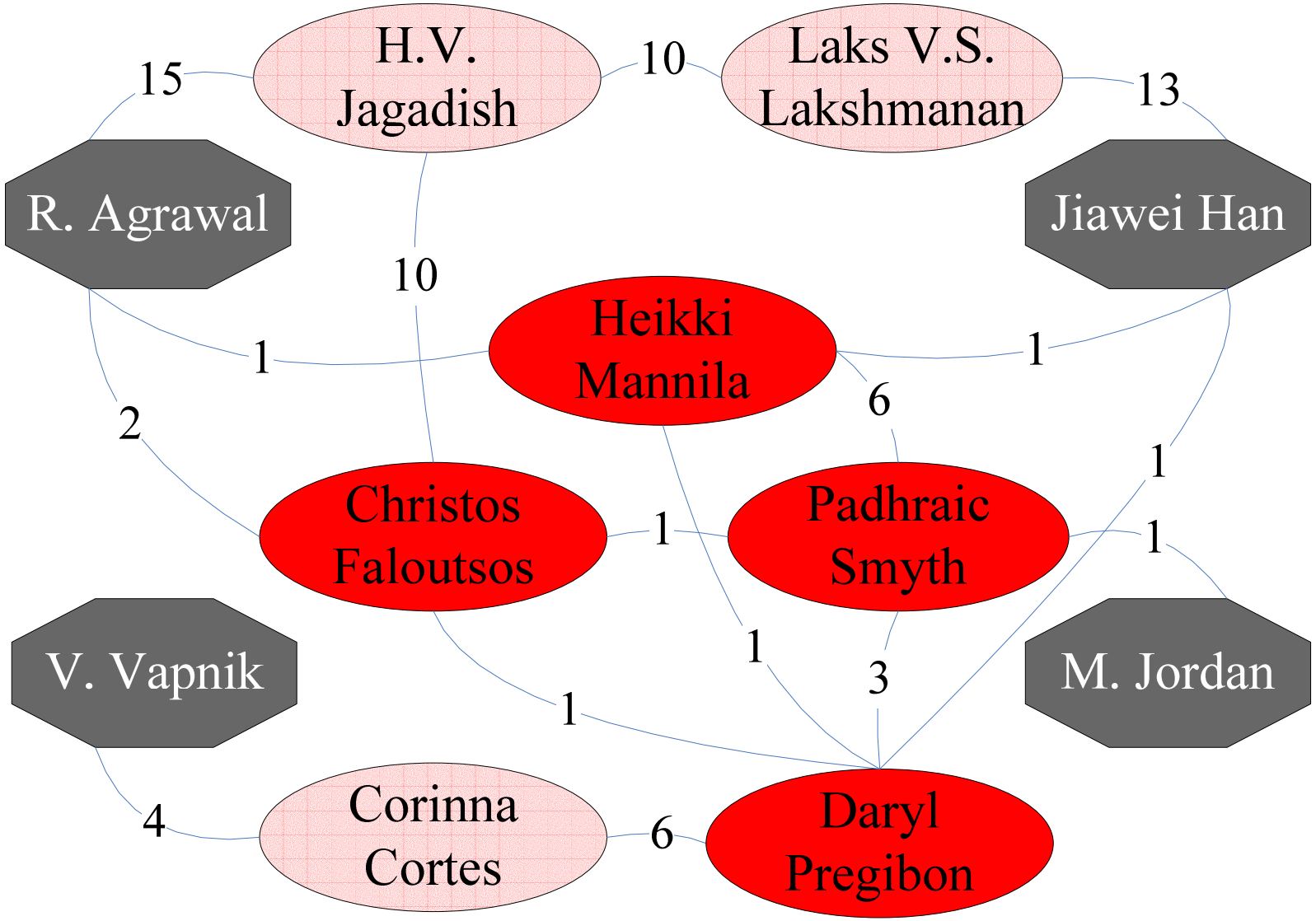
1_SoftAnd query = OR query



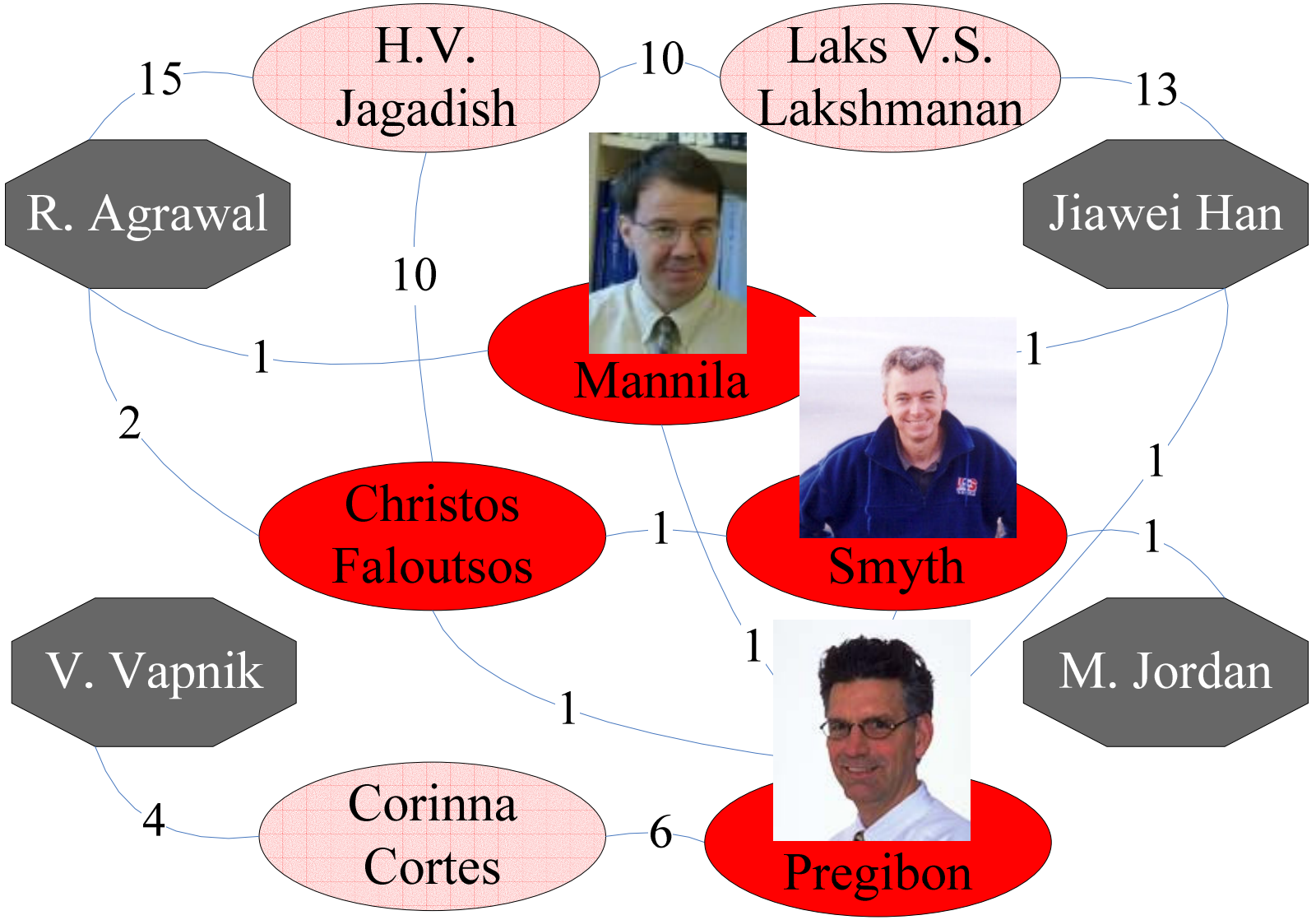
Detailed outline

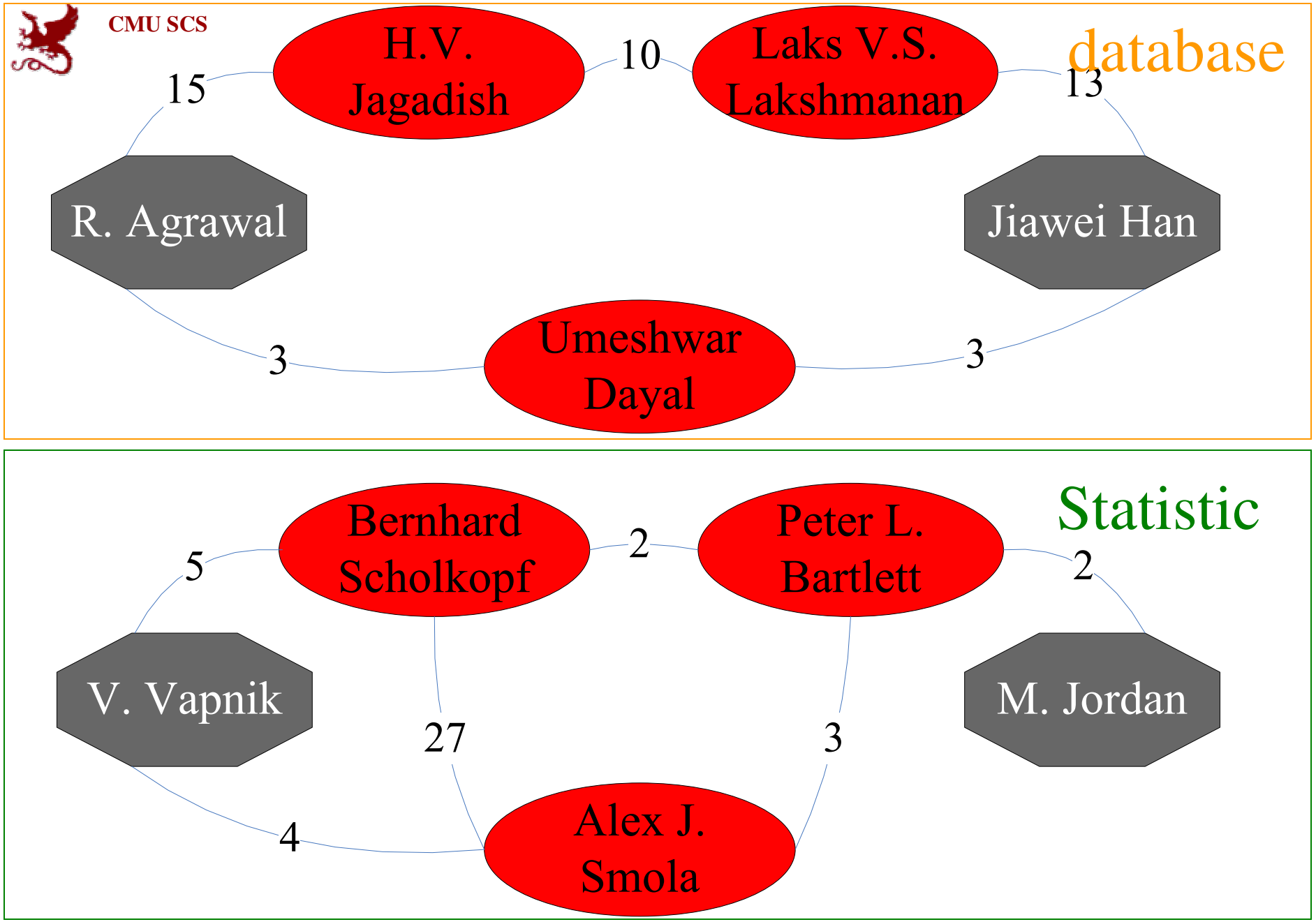
- Problem definition
- Solution
- Results

Case Study: AND query



Case Study: AND query





2_SoftAnd query

Conclusions

Proximity (e.g., w/ RWR) helps answer ‘AND’ and ‘k_softAnd’ queries