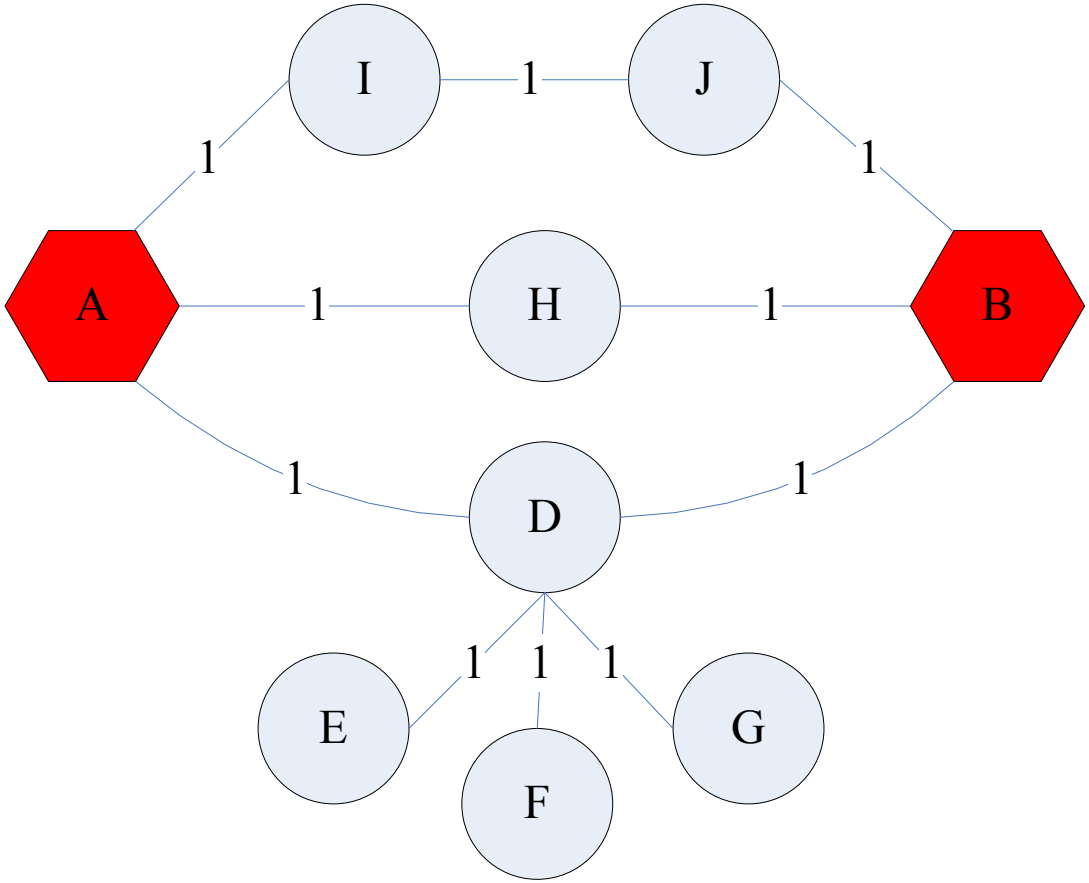
Application to Measuring Proximity in Graphs

Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University

[Tong-Faloutsos, '06]

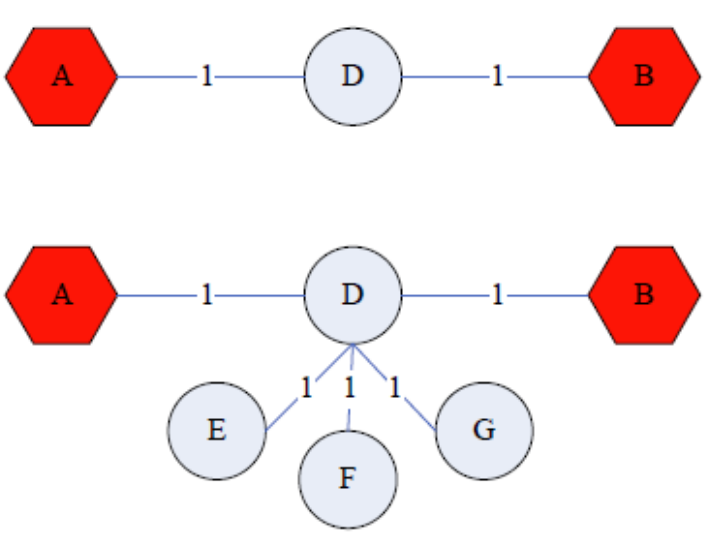
Proximity on Graphs

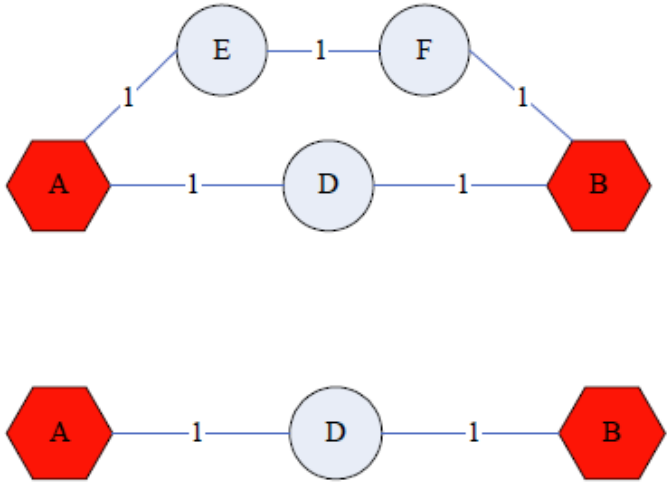


a.k.a.: Relevance, Closeness, 'Similarity'…

Good proximity measure?

- **Shortest path is not good:**

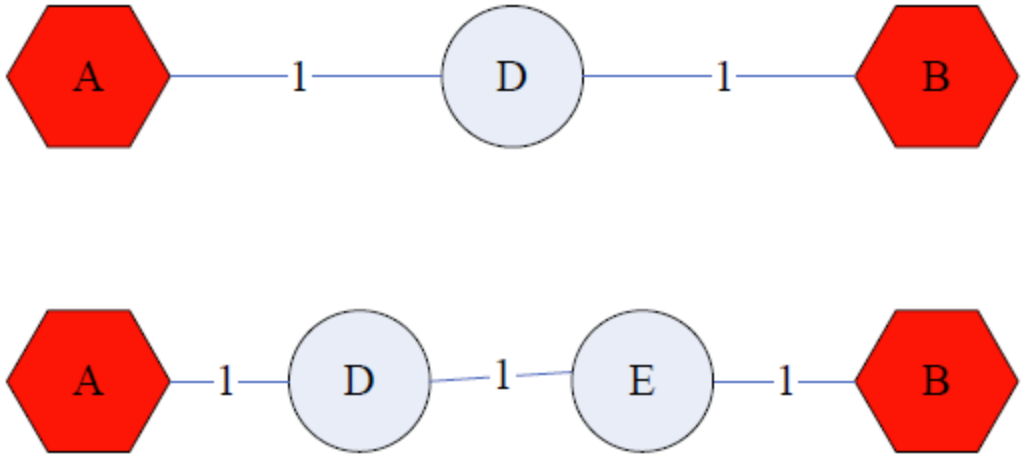




- **No effect of degree-1 nodes (E, F, G)!**
- Multi-faceted relationships

Good proximity measure?

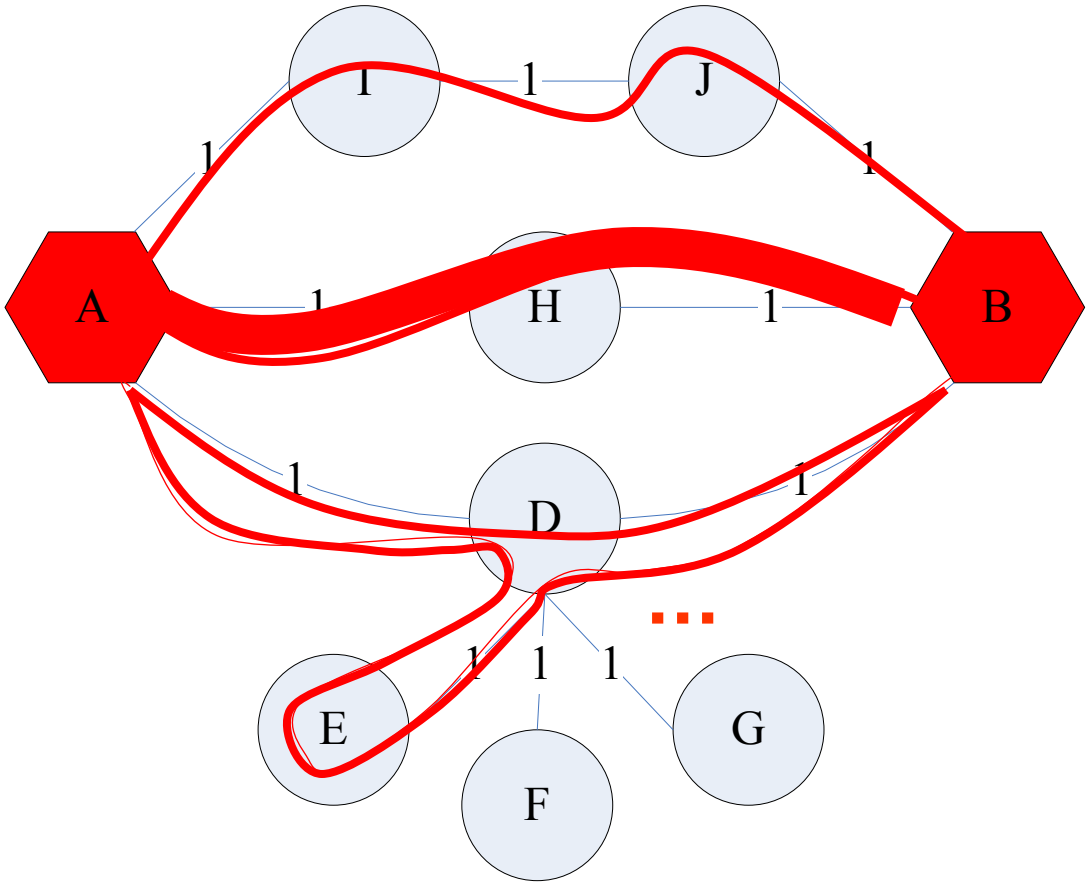
- **Network flow is not good:**



- **Does not punish long paths**

[Tong-Faloutsos, '06]

What is good notion of proximity?

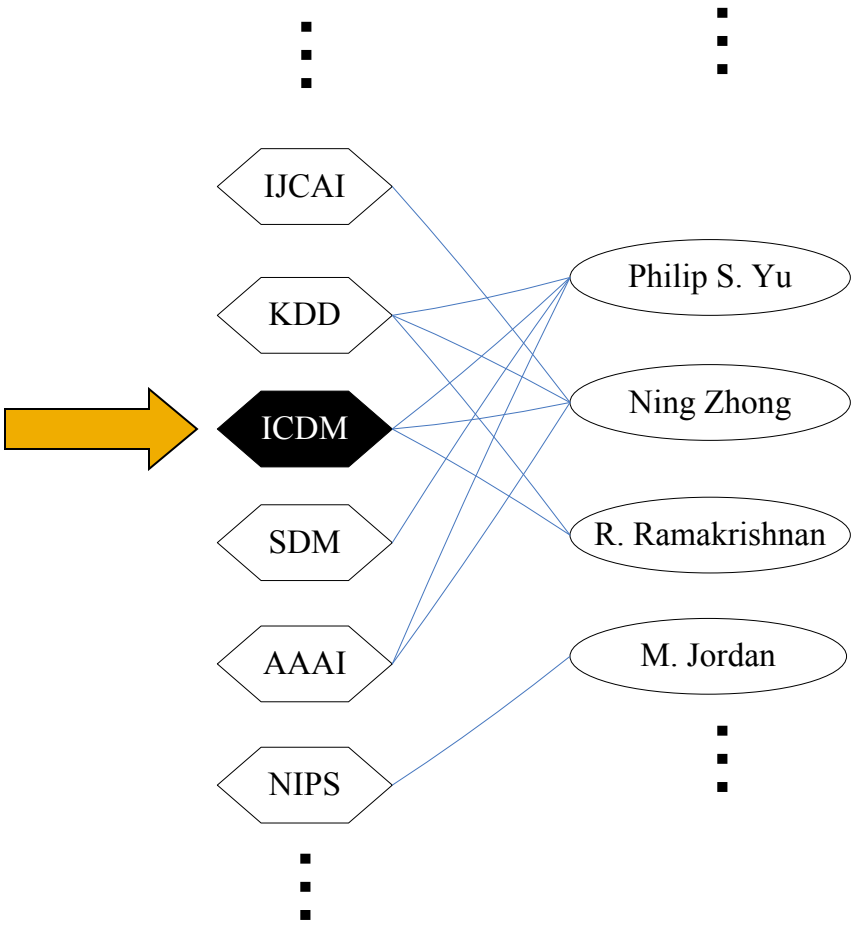


- Multiple Connections
- Quality of connection
  - Direct & In-direct connections
  - Length, Degree, Weight…

SimRank: Idea

- **SimRank:** Random walks from a **fixed node** on ***k***-partite graphs
- **Setting:** ***k***-partite graph with ***k*** types of nodes
  - Example: picture nodes and tag nodes
- Do a **Random Walk with Restarts** from node ***u***
  - i.e., **teleport set $S = \{u\}$**
- Resulting scores measures similarity to node ***u***
- **Problem:**
  - Must be done once for each node ***u***
  - Suitable for sub-Web-scale applications

SimRank: Example



Q: What is most related conference to **ICDM**?

A: Personalized PageRank with teleport set S={ICDM}

SimRank: Example