

# Community Detection

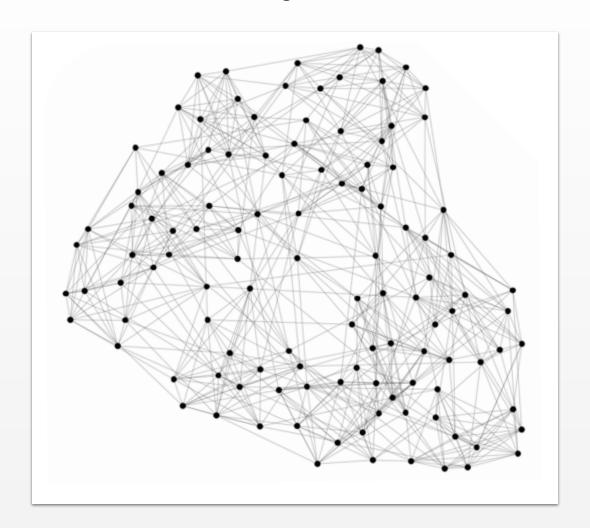
Shantanu Jain



# Centrality Measures

Betweenness, Closeness, Degree

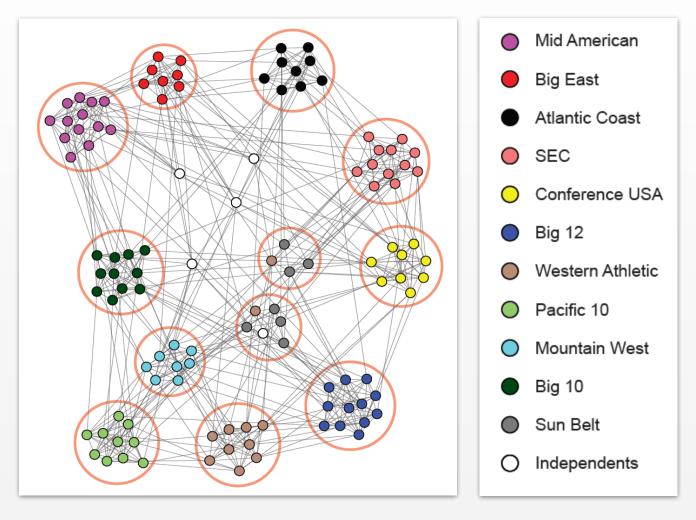
### Community Detection



**Problem:** Can we identify groups of densely connected nodes?

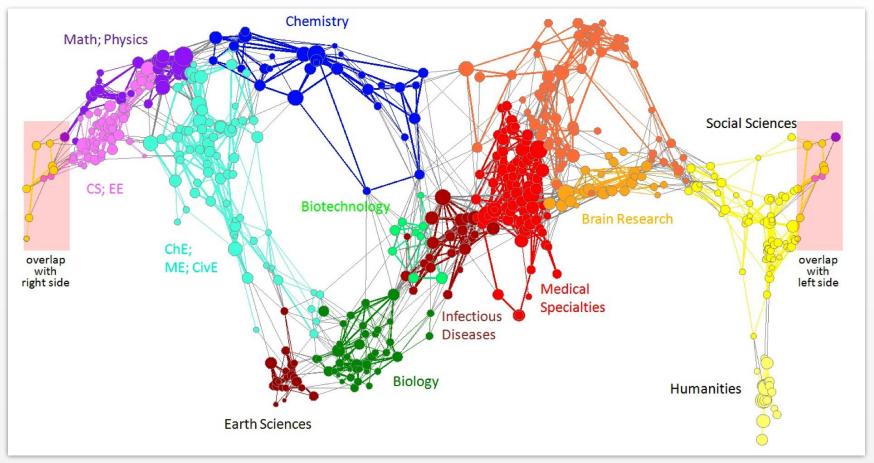
(Adapted from: Mining of Massive Datasets, <a href="http://www.mmds.org">http://www.mmds.org</a>)

#### Communities: Football Conferences



Nodes: Football Teams, Edges: Matches, Communities: Conferences

#### Communities: Academic Citations

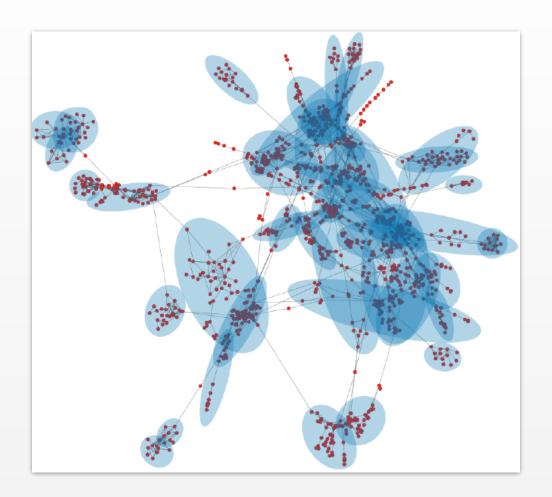


Source: Citation networks and Maps of science [Börner et al., 2012]

Nodes: Journals, Edges: Citations, Communities: Academic Disciplines

(Adapted from: Mining of Massive Datasets, <a href="http://www.mmds.org">http://www.mmds.org</a>)

#### Communities: Protein-Protein Interactions

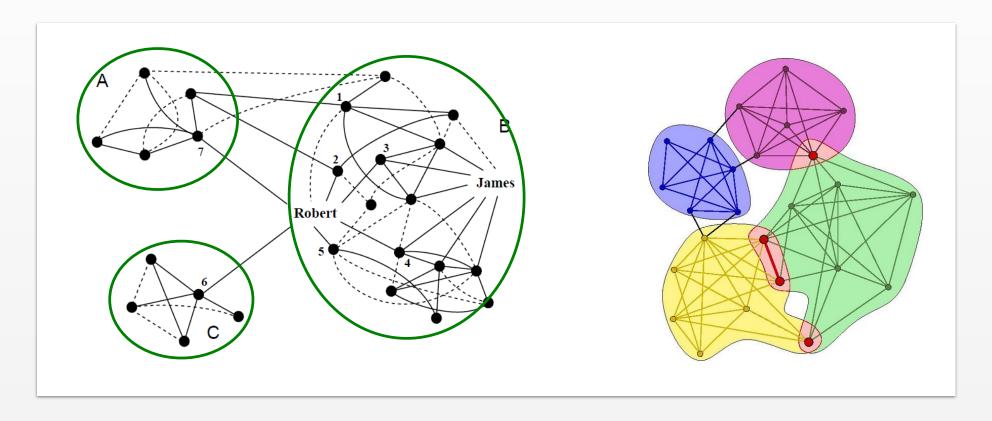


Nodes: Proteins, Edges: Physical interactions, Communities: Functional Modules

## Community Detection

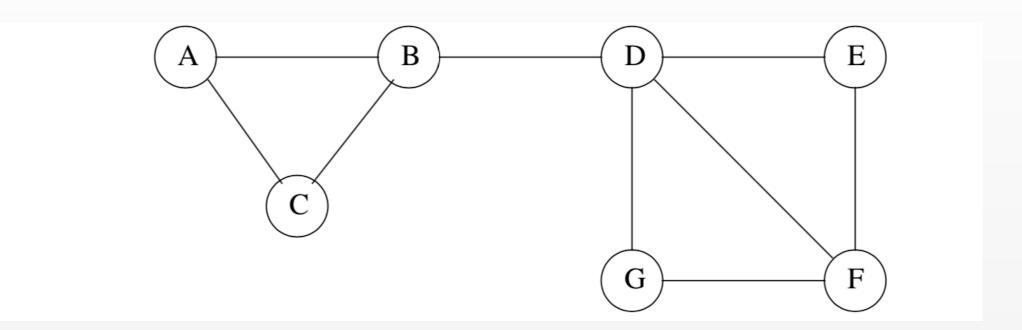
**Graph Partitioning** 

Overlapping Communities



We will work with undirected (unweighted) networks

#### Can we use clustering technique?

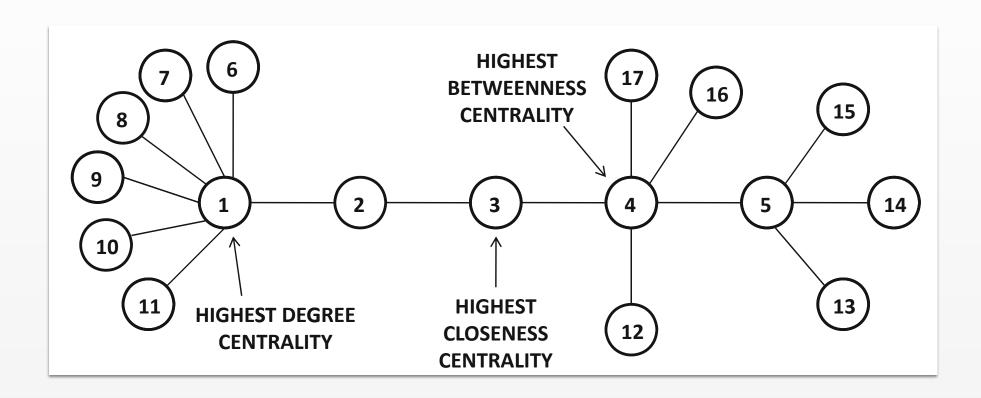


$$d(x,y) = \begin{cases} 0 & (x,y) \text{ is an edge} \\ 1 & (x,y) \text{ is not an edge} \end{cases}$$

Doesn't satisfy triangle inequality.

Since all edges are representing equal distances there is 1/9 chance that B-D will get connected in the first iteration of hierarchical agglomerative clustering, when they clearly belong to different clusters.

### Centrality Measures for nodes

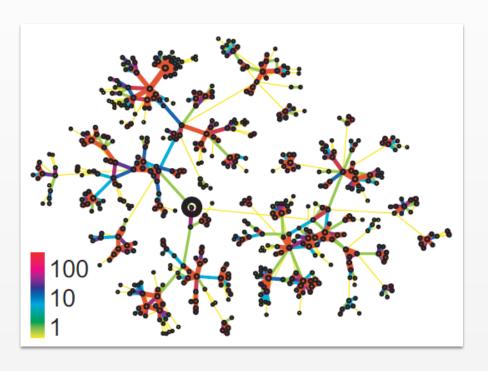


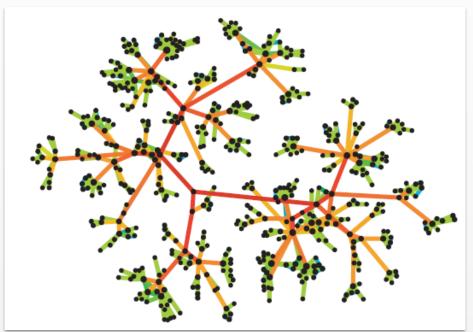
- Betweenness: Number of shortest paths passing through the node
- Closeness: Average distance to other nodes
- *Degree*: Number of connections to other nodes

### Edge betweenness

Edge Strength (call volume)

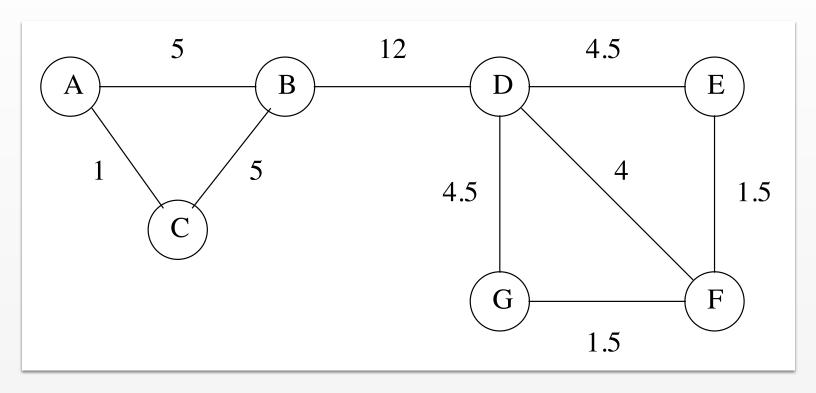
Edge Betweenness





• *Betweenness*: Number of shortest paths passing through a node or edge

#### Edge Betweenness



- Count number of shortest paths passing through each edge (can be done with weighted edges)
- If there are multiple paths of equal length, then split counts