Intro to Network Analysis

with Game of Thrones

Kacper Wieczorek

Outline of the tutorial

- Intro to networks and graphs
- The Network of Thrones project
- Building and analysing our own Games of Thrones network with jupyter notebook and networkx

Why networks?

Interactions between elements in a system

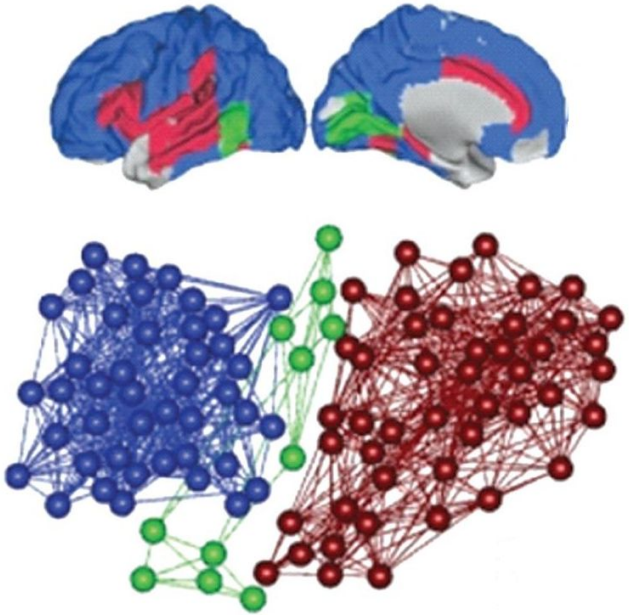
>

Each element in isolation

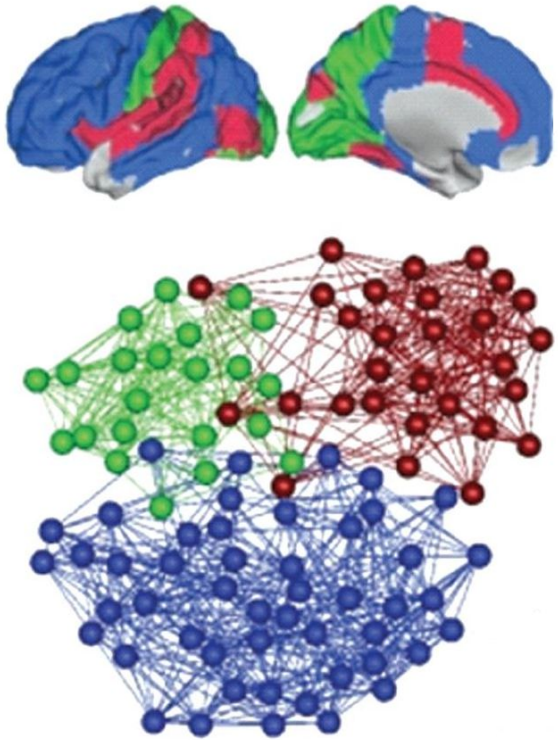
System as a whole

Brain networks

Controls

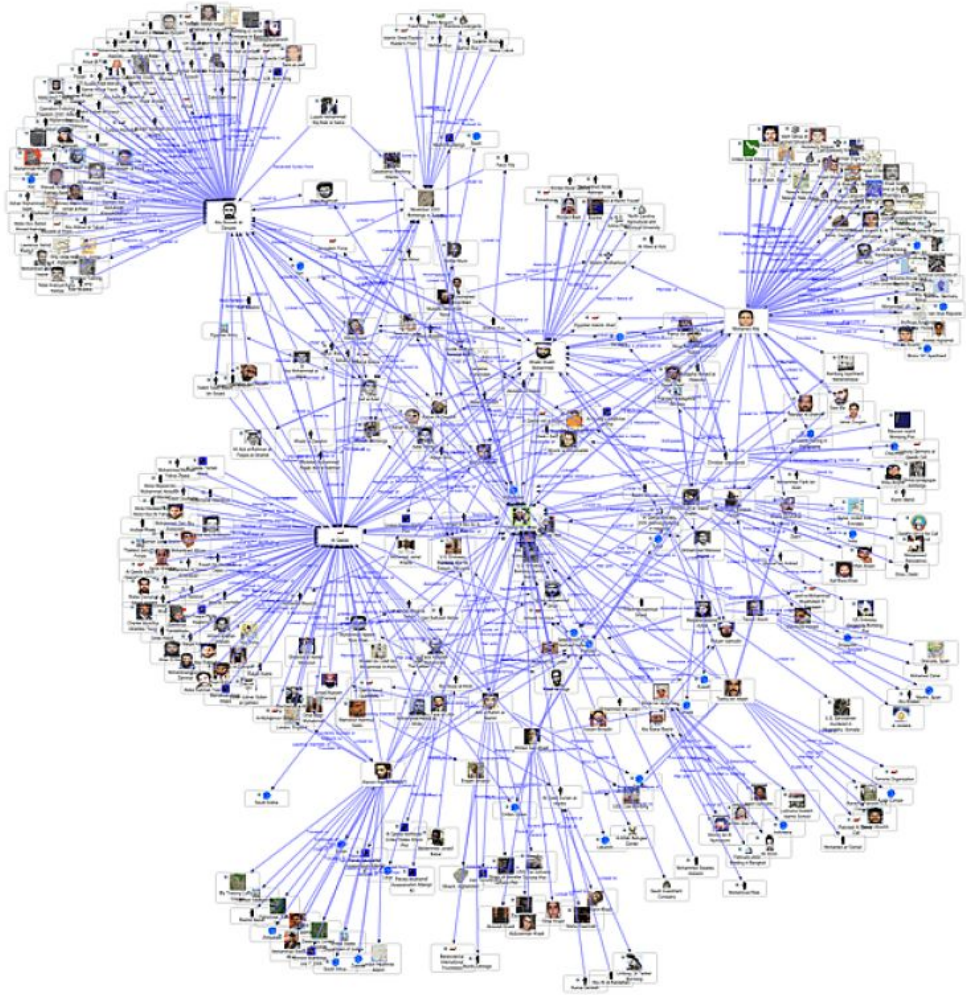


Patients



Micheloyannis, 2012

Terrorist networks



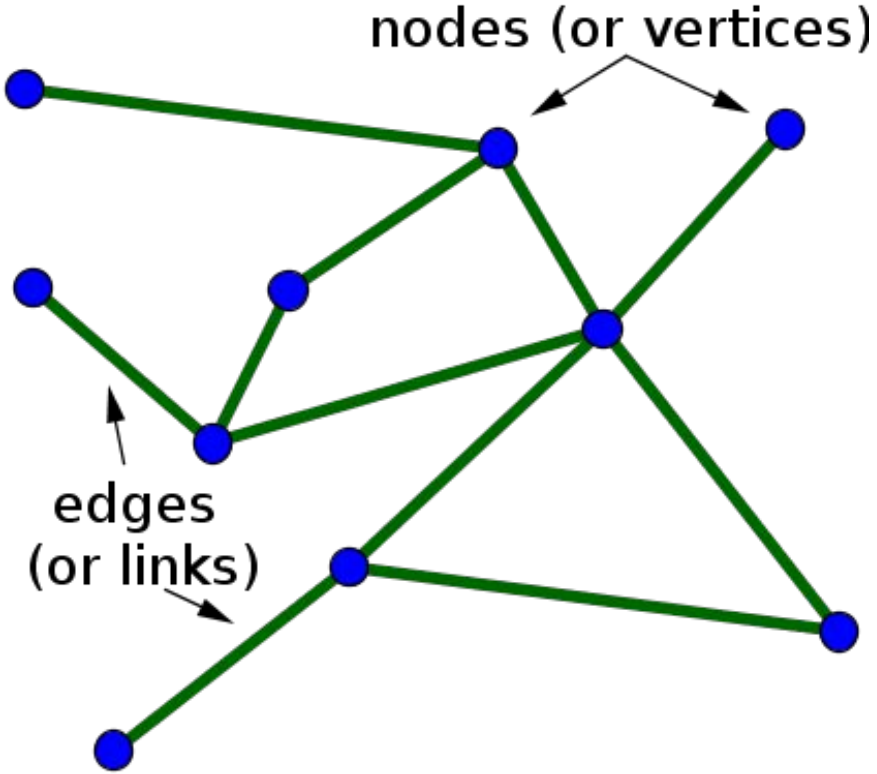
Network of 9/11 contacts (Krebs)

NETWORKS

are modelled using

GRAPHS

Graphs are made of nodes and edges



Graphs: undirected vs directed

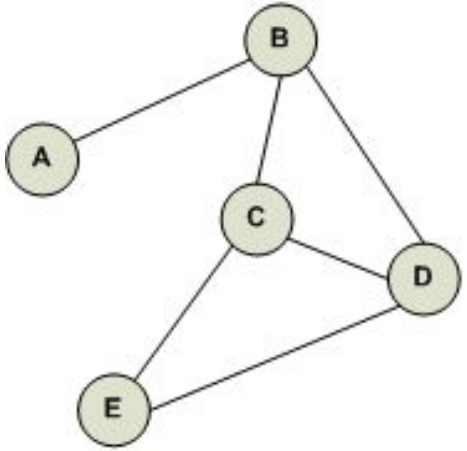


Fig 1. Undirected Graph

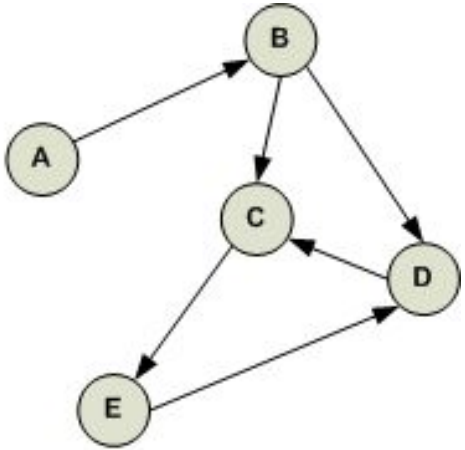
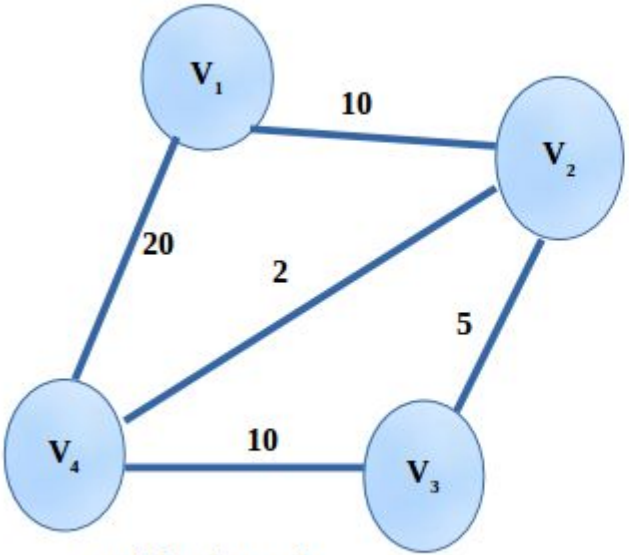
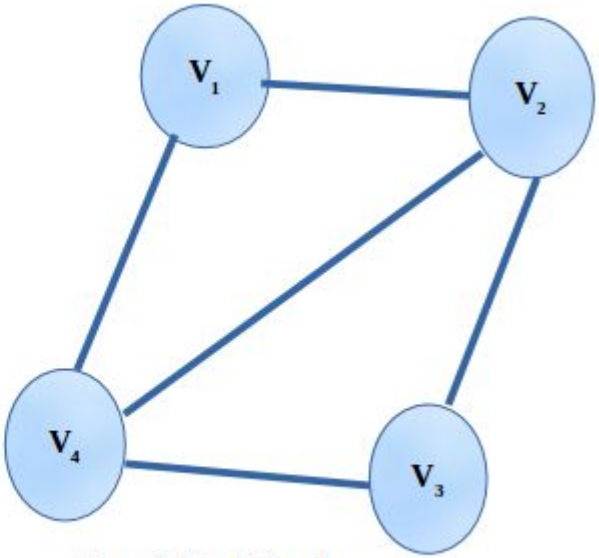


Fig 2. Directed Graph

Graphs: weighted vs unweighted

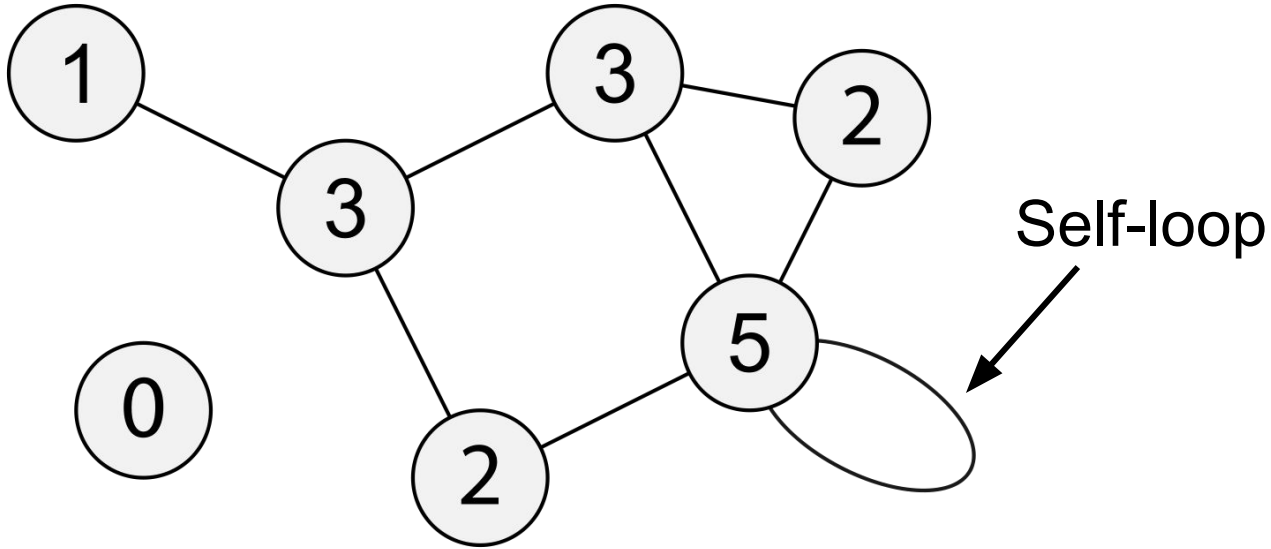


Weighted Graph

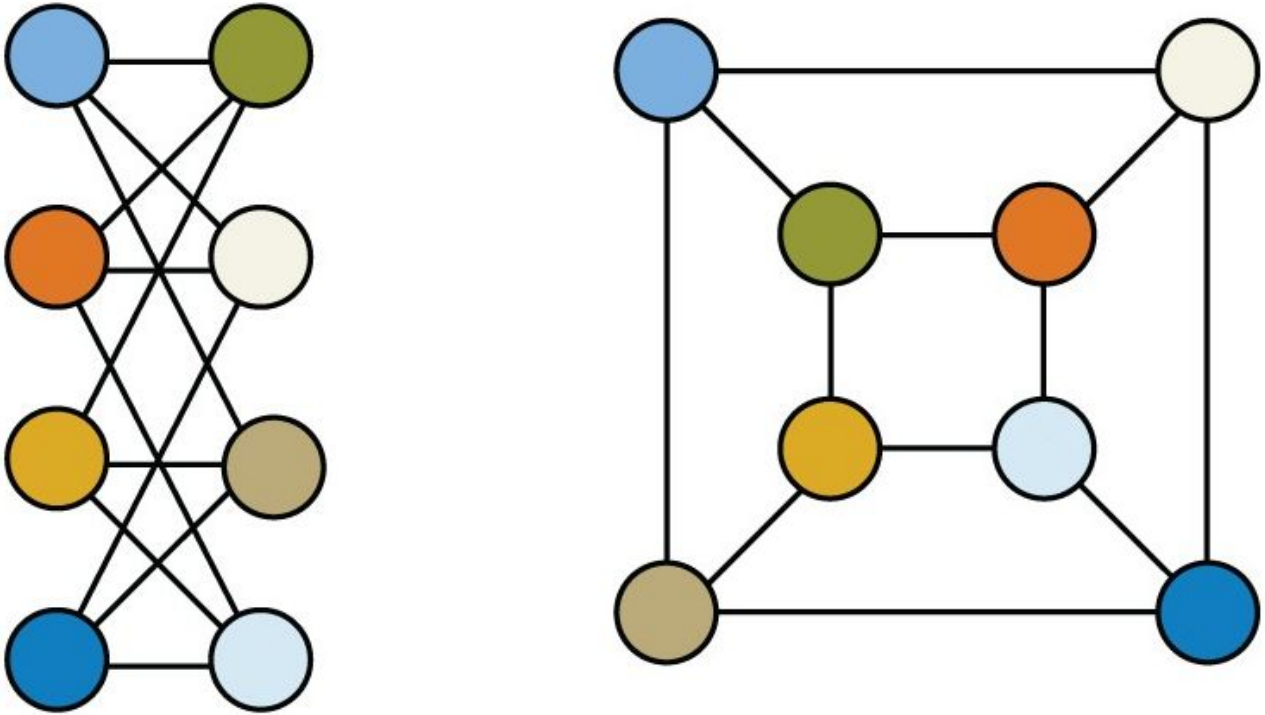


Unweighted Graph

Graphs: self-loops



These two graphs are exactly the same



By the way...

1. **Graph Databases** have become common computational tools and alternatives to SQL and NoSQL databases
2. Graphs are used to model analytics workflows in the form of DAGs (**Directed acyclic graphs**)
3. Some **Artificial Neural Network Frameworks** also use DAGs to model the various operations in different layers
4. It is used in **Clustering algorithms** – Specifically K-Means
5. **System Dynamics** also uses some Graph Theory concepts – Specifically loops
6. **Path Optimization** is a subset of the Optimization problem that also uses Graph concepts
7. Graphs offer **computational efficiency**. The Big O complexity for some algorithms is better for data arranged in the form of Graphs (compared to tabular data)

(from https://www.analyticsvidhya.com)

Network of thrones https://networkofthrones.wordpress.com/

- Characters (nodes) recognition and disambiguation
  - NLP, A Wiki of Ice and Fire
  - Which Jon, Walder, Brandon? Which king, queen, maester?
- Identifying edges (links) and weights between characters (nodes)
  - Link two characters each time their names (or nicknames) appear within 15 words of one another. Only edges with weights > 2 included
  - Undirected, weighted network
- Analysing and visualising the network
  - Gephi

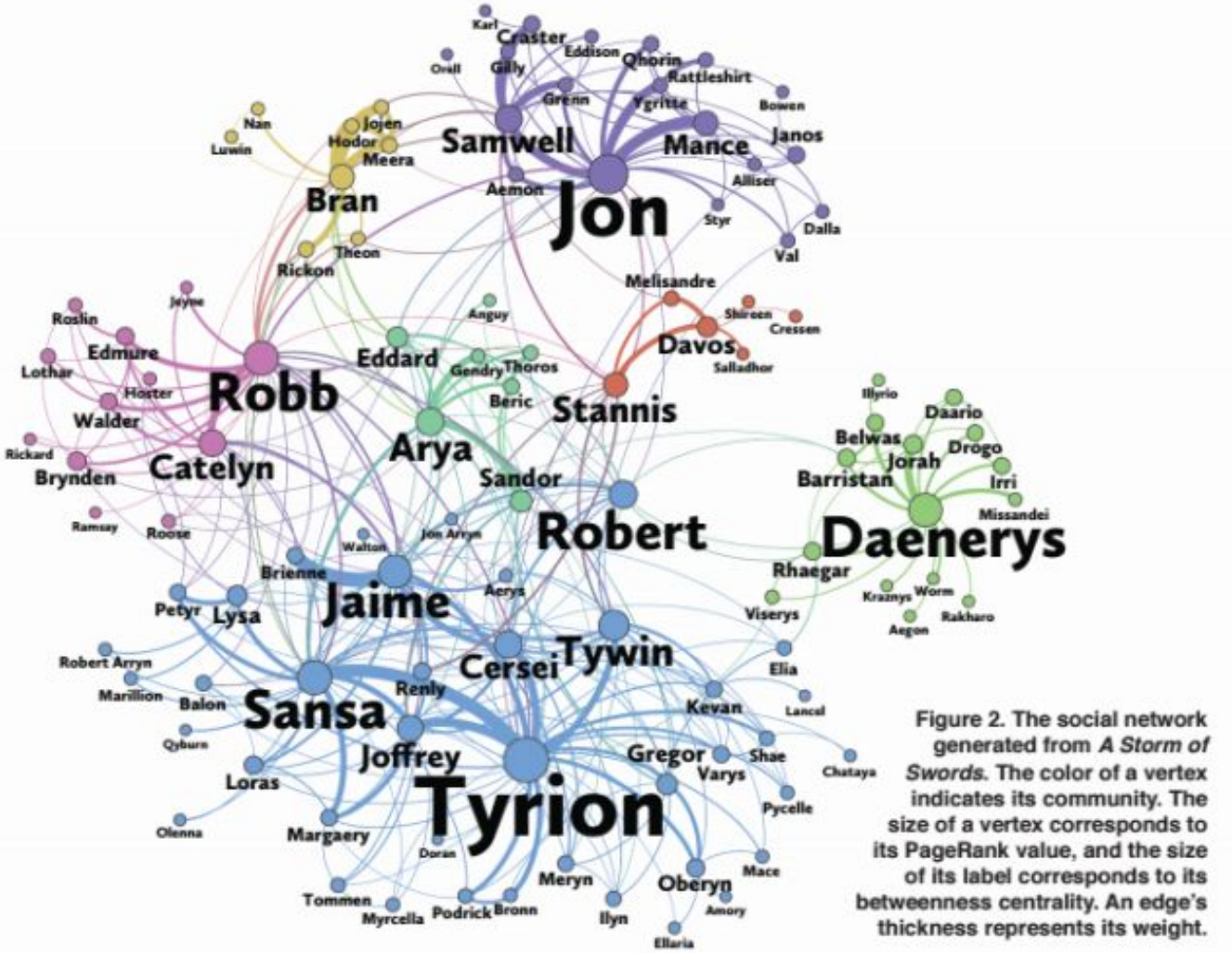


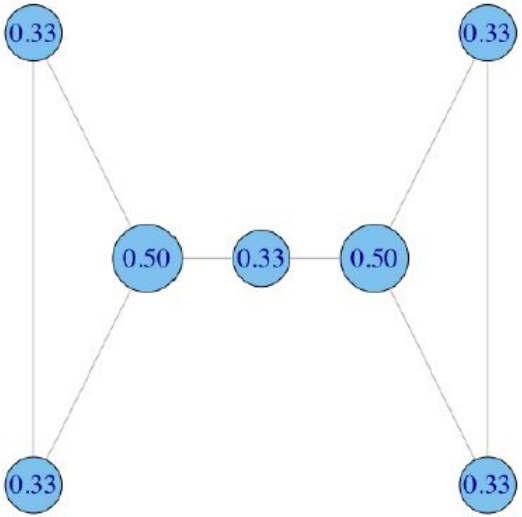
Figure 2. The social network generated from *A Storm of Swords*. The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.

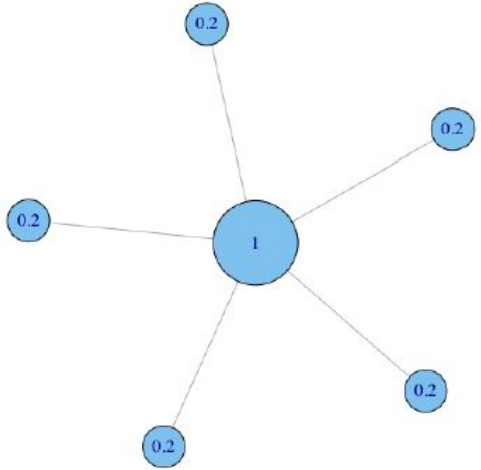
Mridul Seth

- Data Camp project
  - https://www.datacamp.com/projects/76
- GitHub repo
  - https://github.com/MridulS/pydata-networkx
- Tutorial at PyData London 2018
  - https://www.youtube.com/watch?v=oIlxXPebL-I

Degree: Normalized Degree Centrality

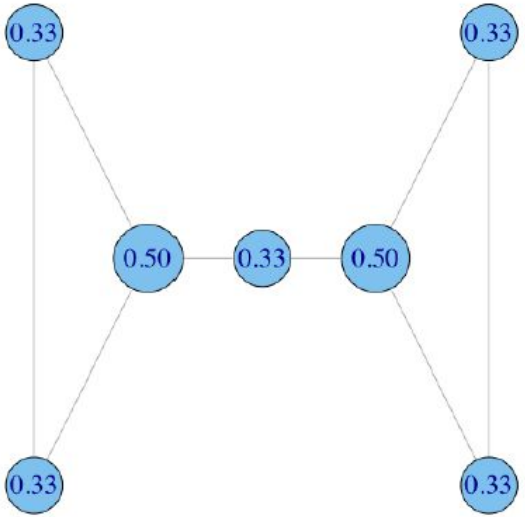
divide by the max. possible, i.e. (N-1)





When Degree Isn't Everything

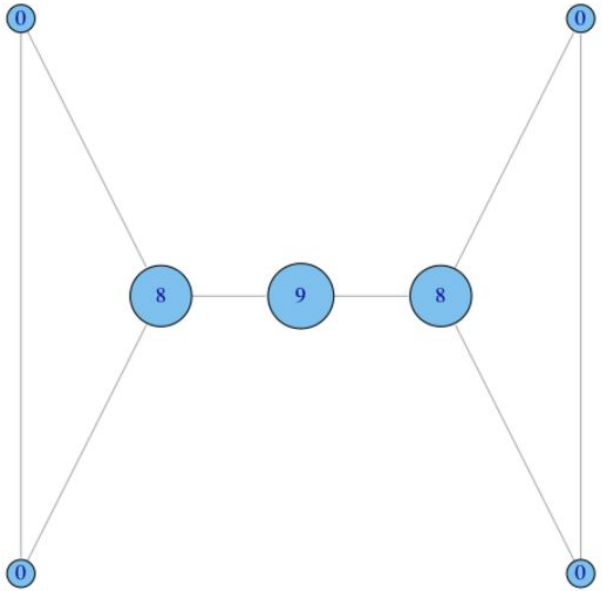
In what ways does degree fail to capture centrality in the following graphs?

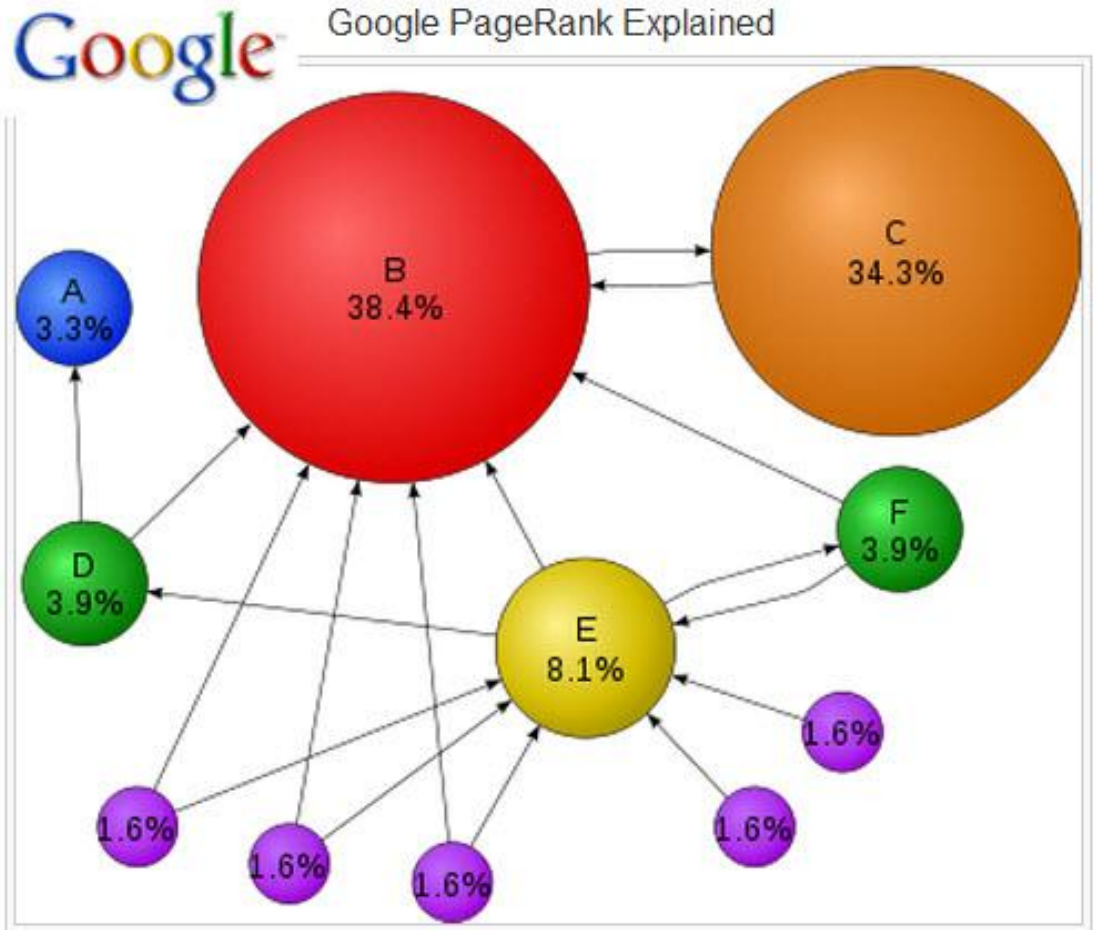


- ability to broker between groups
- likelihood that information originating anywhere in the network reaches you…

Betweenness: Another Centrality Measure

- Intuition: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?




Google
Google PageRank Explained
A
3.3%
B
38.4%
C
34.3%
D
3.9%
E
8.1%
F
3.9%
1.6%
1.6%
1.6%
1.6%
1.6%