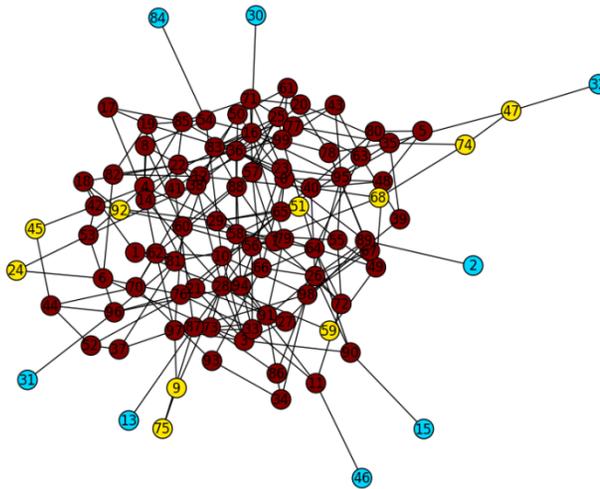


Topological Analysis (1)



Hiroki Sayama
sayama@binghamton.edu

Network data import & export

- `read_gml`
- `read_adjlist`
- `read_edgelist`
 - Creates undirected graphs by default; use `"create_using=Nx.DiGraph()"` option to generate directed graphs

Exercise

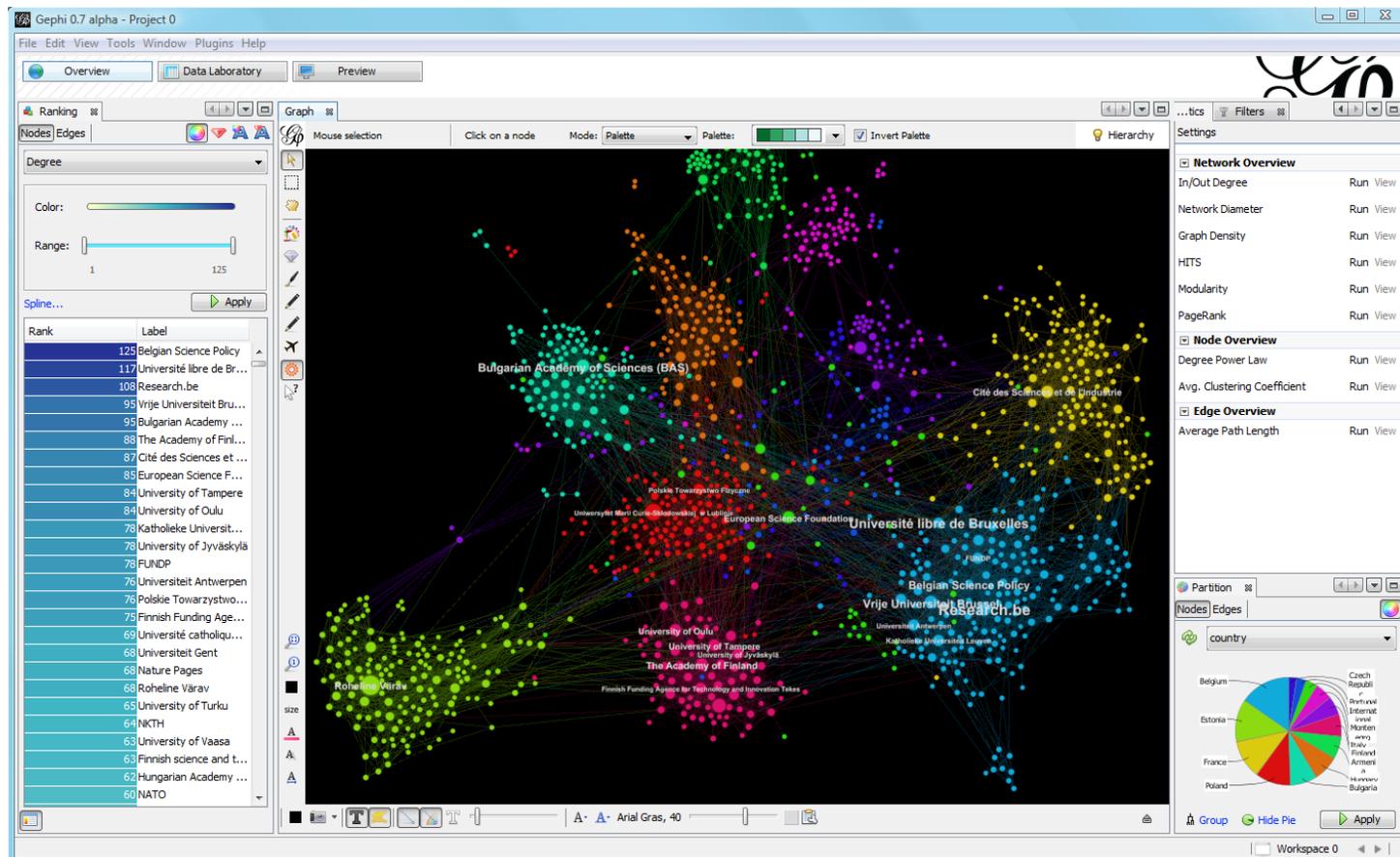
- **Import Supreme Court Citation Network Data into NetworkX**
(<http://jhfowler.ucsd.edu/judicial.htm>)
 - Import as an undirected graph
 - Import as a directed graph

Network visualization

- “nx.draw”
- Various layout functions
 - Spring, circular, random, spectral, etc.
- For visualization of large-scale networks, use “Gephi”

Gephi

- Network visualization & analysis tool



Basic Properties of Networks

Basic properties of networks

- Number of nodes
- Number of links
- Network density
- Connected components

Network density

- The ratio of # of actual links and # of possible links

- For an undirected graph:

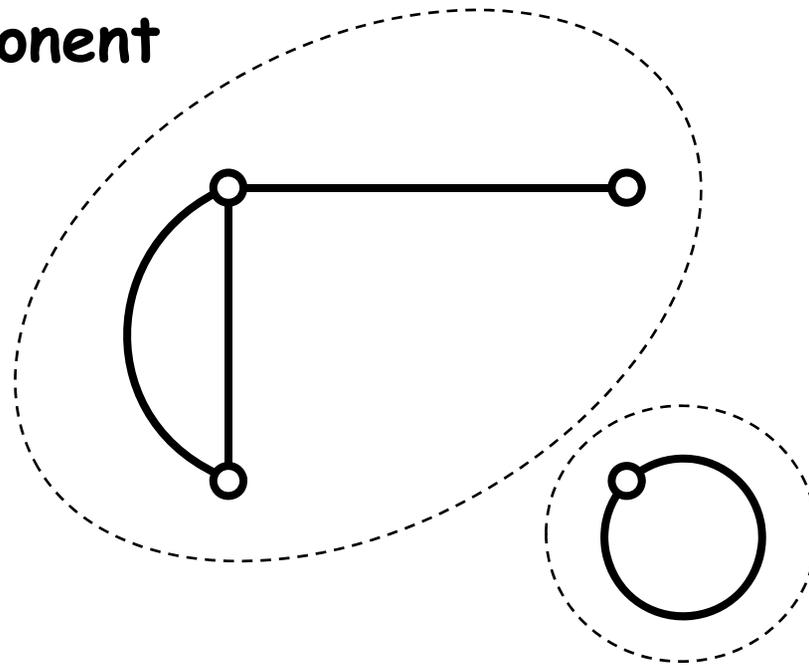
$$d = |E| / (|V| (|V| - 1) / 2)$$

- For a directed graph:

$$d = |E| / (|V| (|V| - 1))$$

Connected components

Connected
component



Connected
component

Number of
connected
components
= 2

Exercise

- Measure the following for the (undirected) Supreme Court Citation Network
 - Number of nodes, links
 - Network density
 - Number of connected components
 - Size of the largest connected component
 - Distribution of the sizes of connected components

Shortest path lengths, etc.

- **shortest_path**
- **shortest_path_length**
- **eccentricity**
 - Max shortest path length from each node
- **diameter**
 - Max eccentricity in the network
- **radius**
 - Min eccentricity in the network

Exercise

- Draw the Karate Club network with its nodes painted with different colors according to their eccentricity

Characteristic path length

- Average shortest path length over all pairs of nodes
- Characterizes how large the world represented by the network is
 - A small length implies that the network is well connected globally

Clustering coefficient

- For each node:
 - Let n be the number of its neighbor nodes
 - Let m be the number of links among the n neighbors
 - Calculate $c = m / \binom{n}{2}$

Then $C = \langle c \rangle$ (the average of c)

- C indicates the average probability for two of one's friends to be friends too
 - A large C implies that the network is well connected locally to form a cluster

Exercise

- Measure the average clustering coefficients of the following network:
 - Karate Club graph
 - Krackhardt Kite graph
 - Supreme Court Citation network
 - Any other network of your choice
- Compare them and discuss
 - Can you tell anything meaningful?

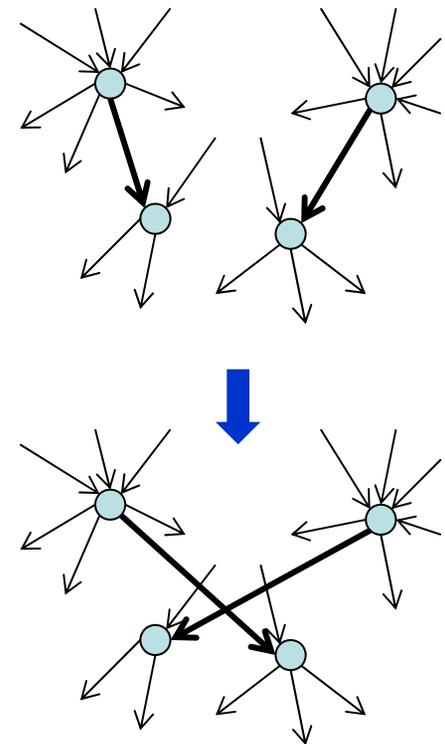
Randomizing networks

- Construct a “null model” network samples to test statistical significance of experimentally observed properties
 - Randomized while some network properties are preserved (e.g., degrees)
 - If the observed properties still remain after randomization, they were simply caused by the preserved properties
 - If not, something else was causing them

Randomization method (1)

- **Double edge swap method**

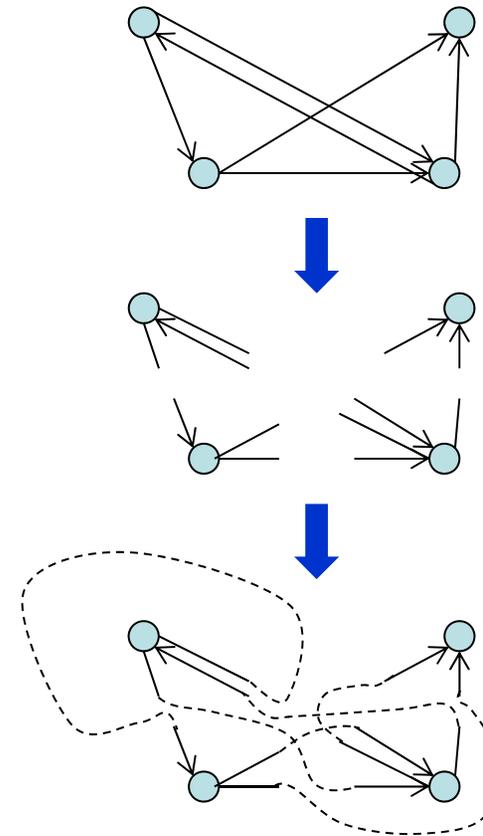
1. Randomly select two links
2. Swap its end nodes
 - (If this swap destroys some network property that should be conserved, cancel it)
3. Repeat above many times



Randomization method (2)

- Configuration model (Newman 2003)

1. Cut every link into halves (heads and tails)
2. Randomly connect head to tail
 - This conserves degree sequences
 - (Could result in multiple links and self-loops)



Other randomization methods

- Keeping only #'s of nodes and edges
- Degree sequence method
- Expected degree sequence method

Exercise

- Randomize connections in the Karate Club graph
- Measure the average clustering coefficient of the randomized network many times
- Test whether the average clustering coefficient of the original network is significantly non-random or not