Lecture 08
Network Analysis 3

Andrew Sonta

CIVIL 534: Computational systems thinking for sustainable engineering

16 April 2025

Outline

- Network measures
  - ~~Individual nodes: centrality~~
  - Groups of nodes
- Group definitions
  - Cliques, cores, components
- Clustering nodes
  - Global and local clustering coefficient
    - Random graph models
  - Similarity
  - Homophily
  - Community detection
- Motifs

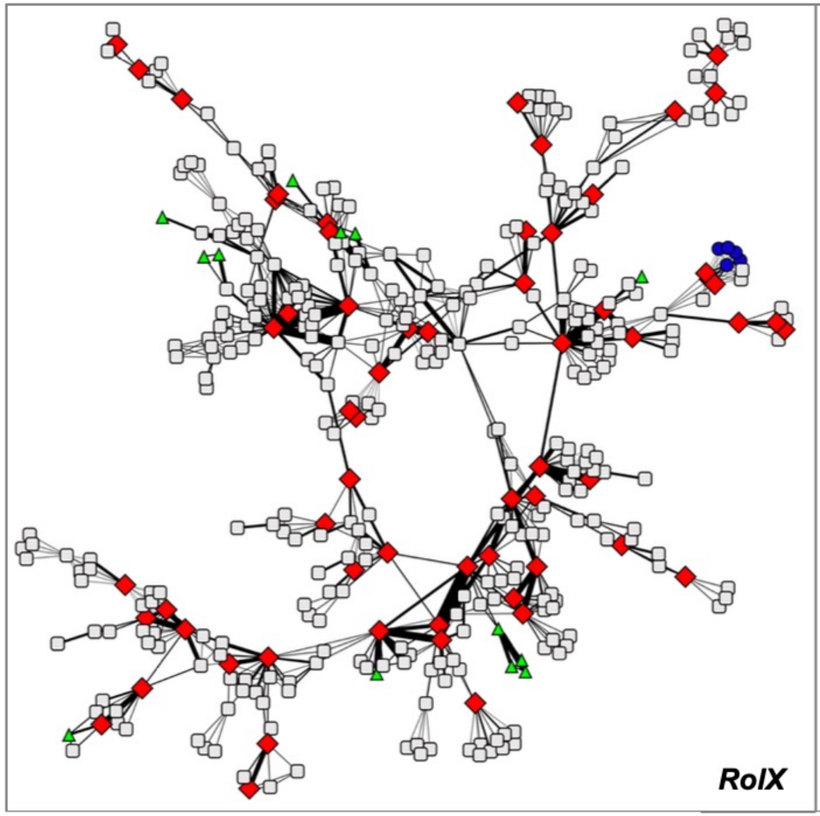
- Corresponding parts of Newman: 7.2-7.4, 7.6, 7.7
- Project milestone 2 out today
- My office hours today: shifted to 13:00-14:00

Motivation: Roles and Communities

How can we describe the structure of groups of nodes in a network?

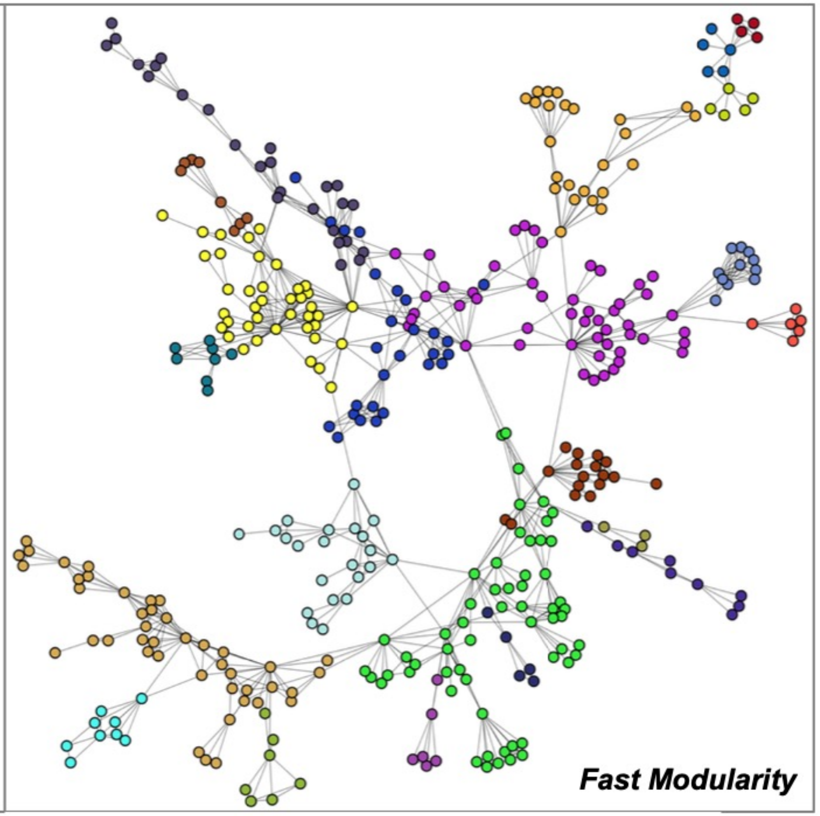
Is there an underlying structure to the network?

Roles



Henderson, *et al.*, KDD 2012

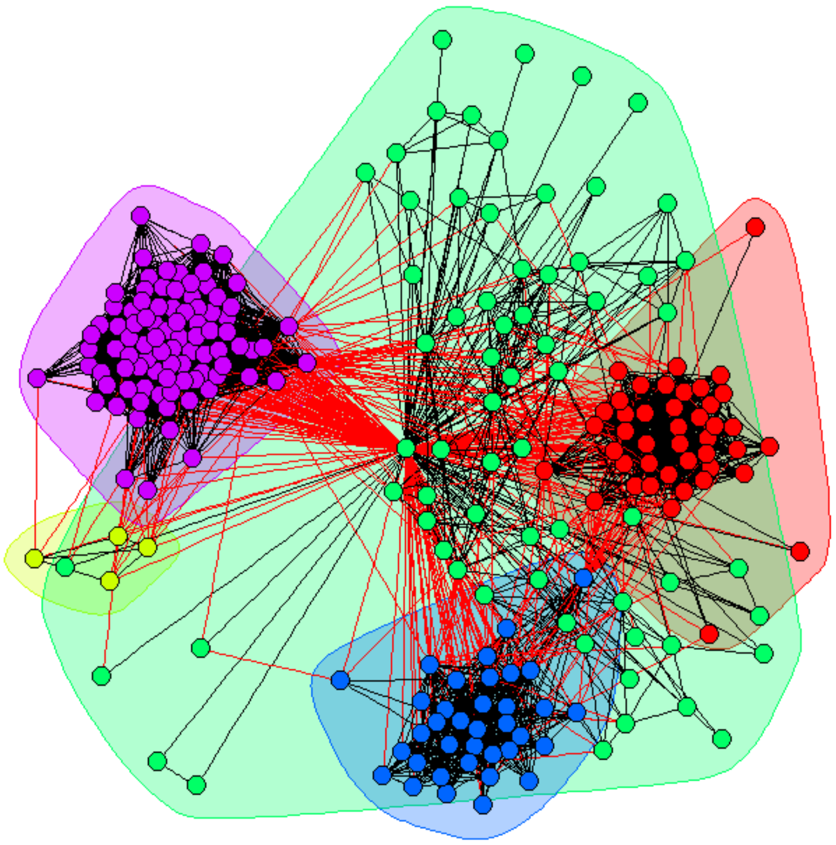
Communities



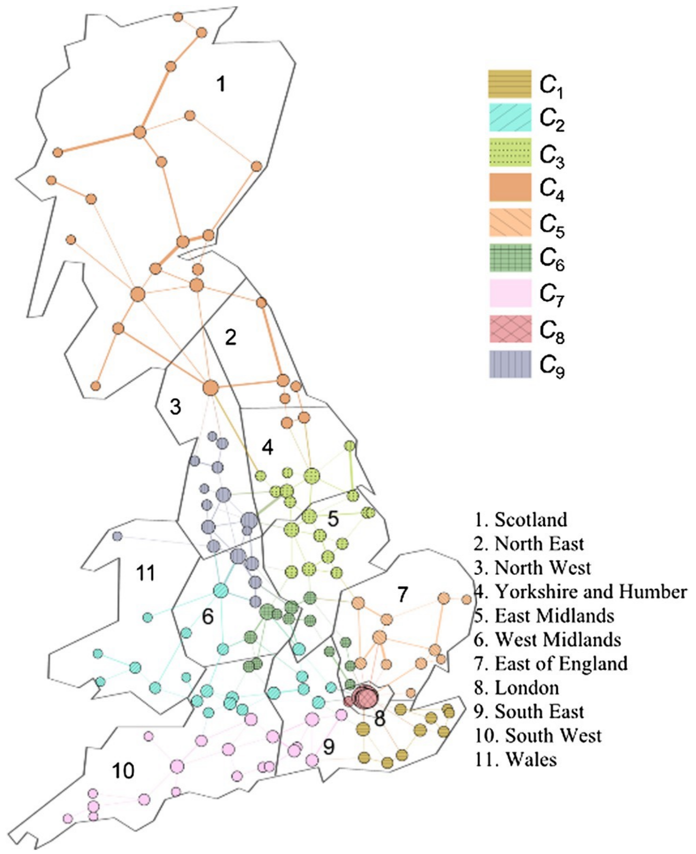
Clauset, *et al.*, Phys. Rev. E 2004

Communities - Applications

Online social network "ego-network"

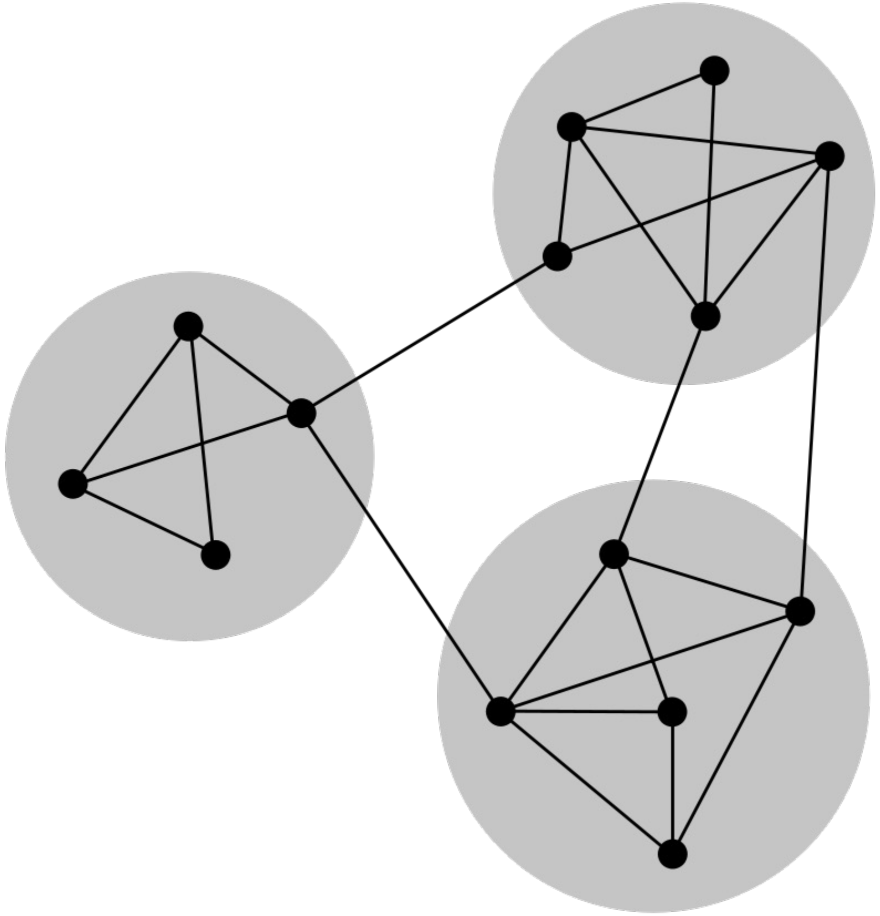


Urban transit network and locally-dependent segments



What led to our perception of the existence of communities?

- We think of many networks as looking like this →
- Mark Granovetter (PhD thesis in the 1960s)
  - Strong and weak ties
  - Strong ties are structurally embedded but redundant in terms of flow of information
  - Weak ties are longer range but greatly enhance access to information across the network

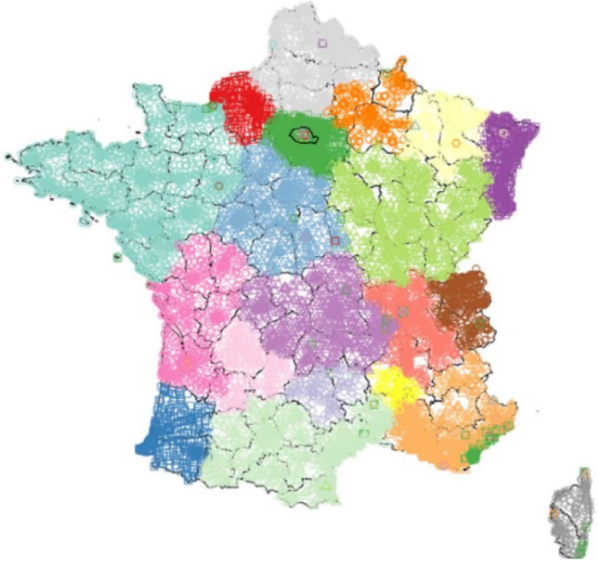


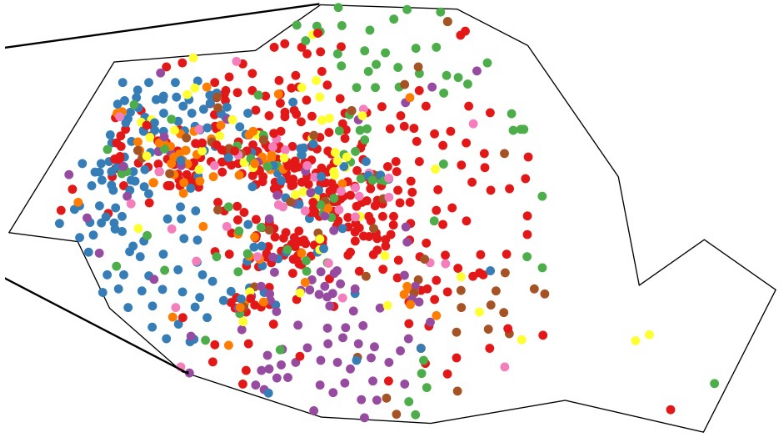
How do we find communities?

- Active area of research
- Many computational methods and algorithms
- Goal for today: introduce the underlying concepts so that you can understand and apply the different methods effectively

Community detection example

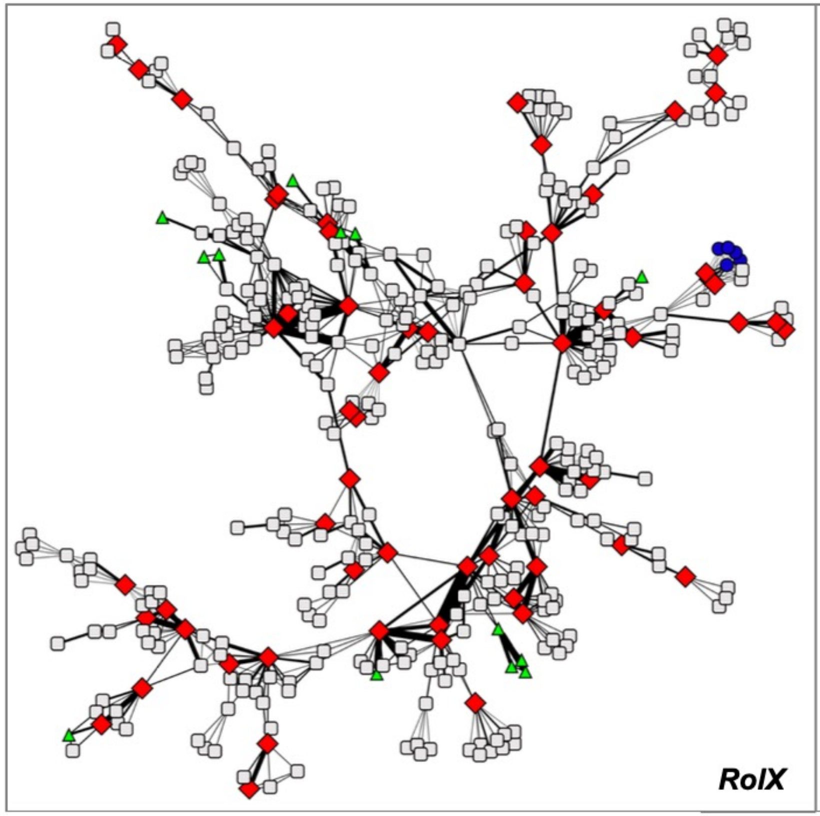
- Gonzalez (2015) created a network of mobile phone users (edges = # of calls)
- Ran automated detection of communities
  - Louvain algorithm – will discuss later
- Found that:
  1. Country-scale communities are largely geographically consistent (essentially redrew administrative boundaries)
  2. City-scale communities are more interspersed geographically → are there strong neighborhoods?





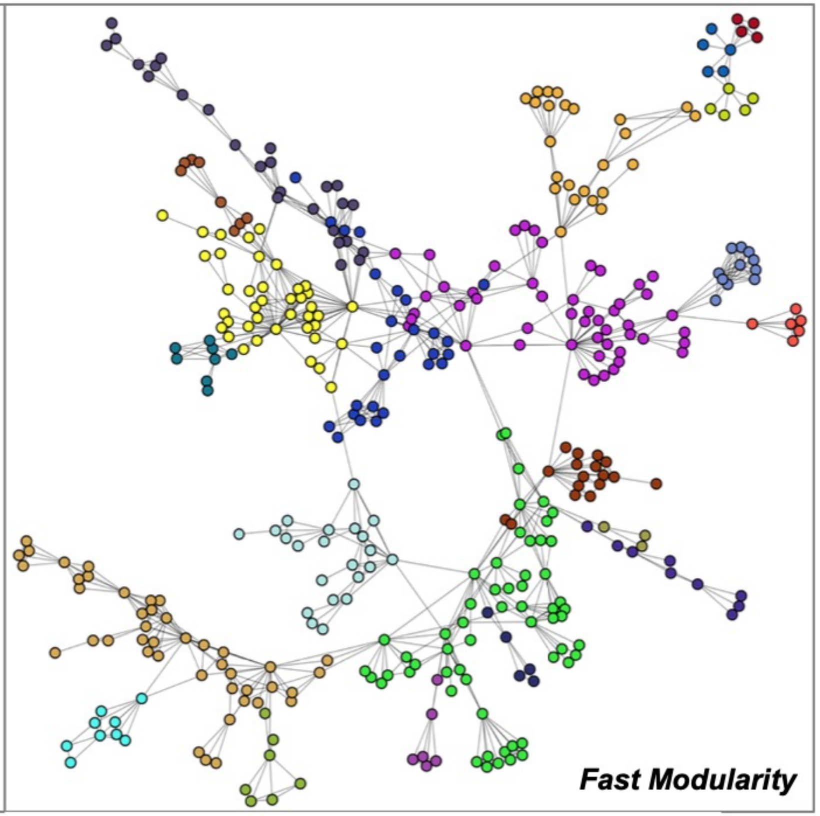
Herrera-Yagüe, C., Schneider, C., Couronné, T. *et al.* The anatomy of urban social networks and its implications in the searchability problem. *Sci Rep* **5**, 10265 (2015).

Roles



Henderson, *et al.,* KDD 2012

Communities



Clauset, *et al.,* Phys. Rev. E 2004

Roles

- Examples
  - Individuals in a company
  - Species in an ecosystem
  - Buildings in a city
- Intuition: how ***similar*** are these two nodes?

Similarity

- Two ways to describe similarity, both involving the notion of "equivalence"
- Structural equivalence: when nodes share the same neighbors
- Regular equivalence: when the neighbors of the nodes are themselves similar (similarity here open to interpretation)
  - Less well-developed than structural equivalence
  - Similar concept to eigenvector centrality and Katz centrality but adapted for similarity instead of centrality
  - We won't go in depth but see section 7.6.2 in Newman if you are curious

Structural equivalence

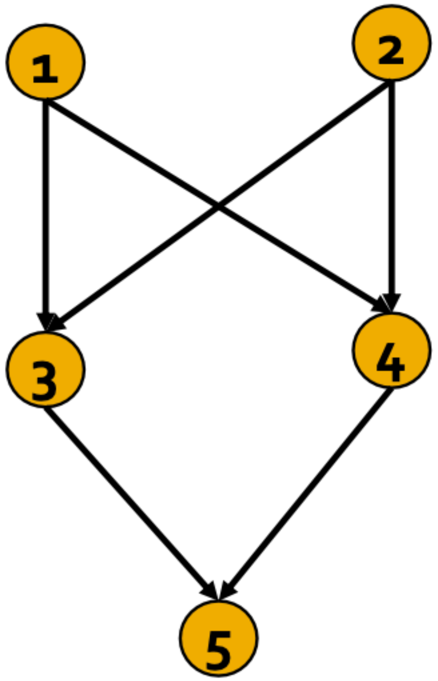
- Step 1: count number of common neighbors

$$n_{ij} = \sum_{k=1}^{n} A_{ik}A_{kj}$$

- Step 2: compute similarity (cosine similarity)

$$\sigma_{ij} = \frac{\sum_{k=1}^{n} A_{ik}A_{kj}}{\sqrt{\sum_{k=1}^{n} A_{ik}^2}\sqrt{\sum_{k=1}^{n} A_{jk}^2}} = \frac{\sum_{k=1}^{n} A_{ik}A_{kj}}{\sqrt{k_i k_j}}$$

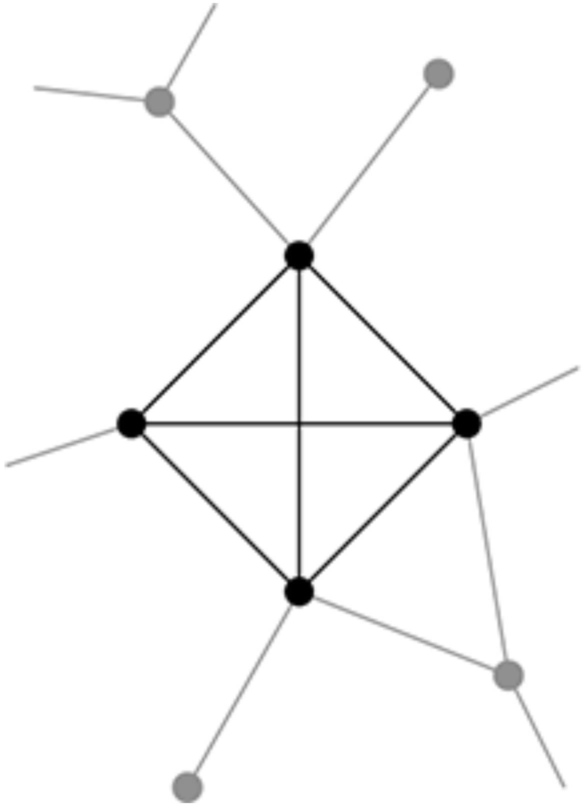
For undirected network



Groups of nodes

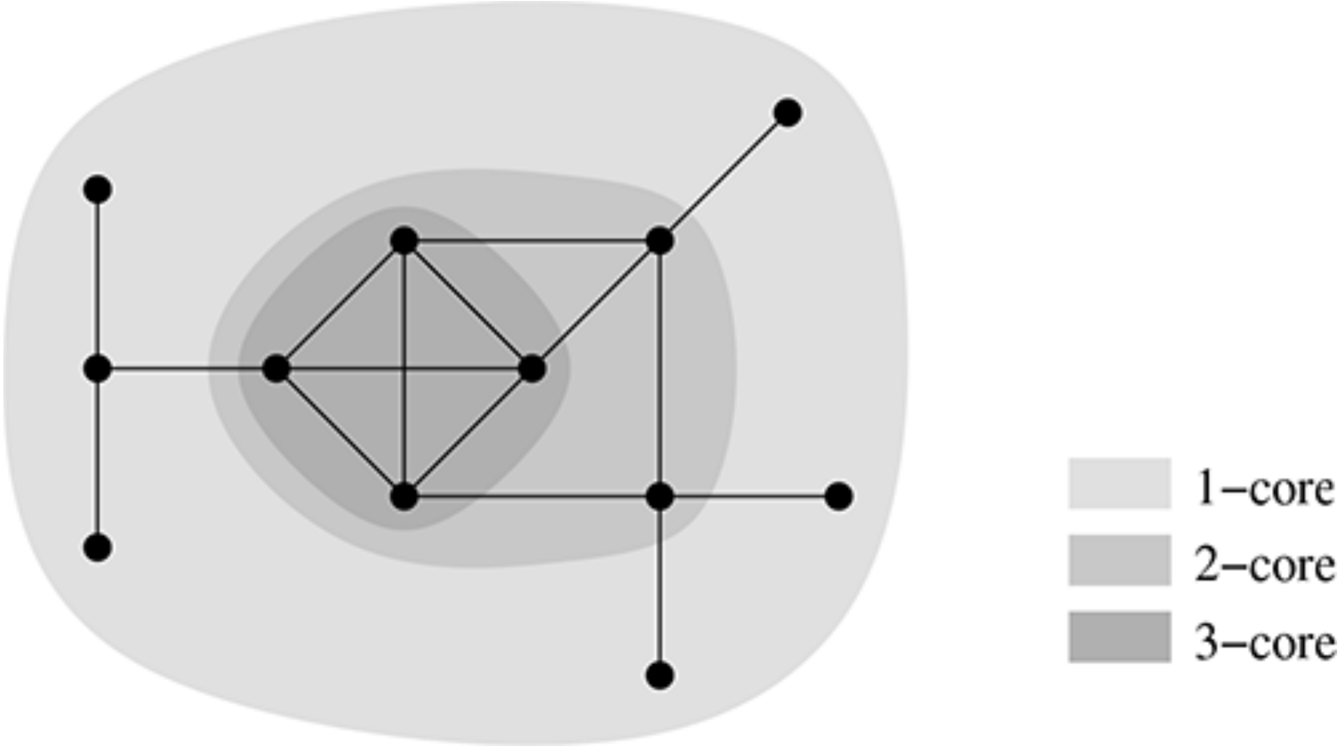
Cliques

- Set of nodes such that every member of the set is connected by an edge to every other
- Nodes can be part of multiple cliques
- Limitation: very stringent requirement



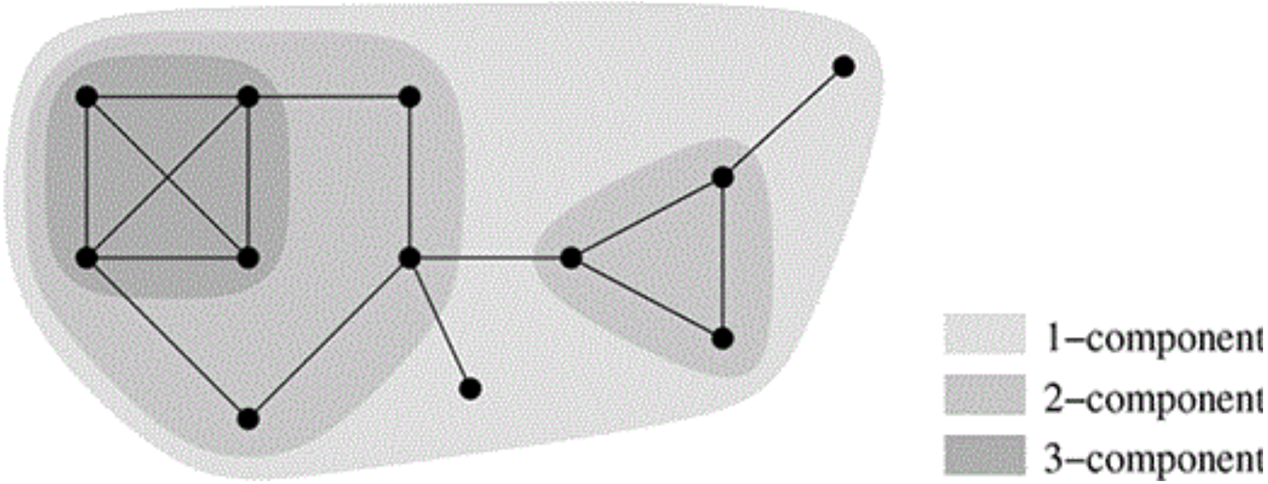
Cores (k-cores)

- Relaxation of the requirements of a clique
- *k*-core: each set of nodes in the group is connected to at least *k* of the others

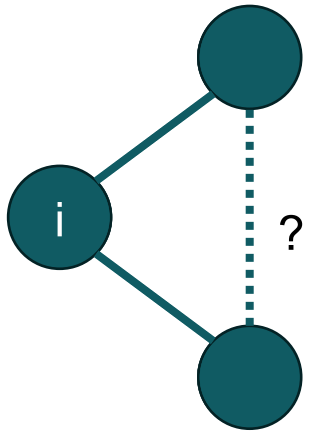


Components (k-components)

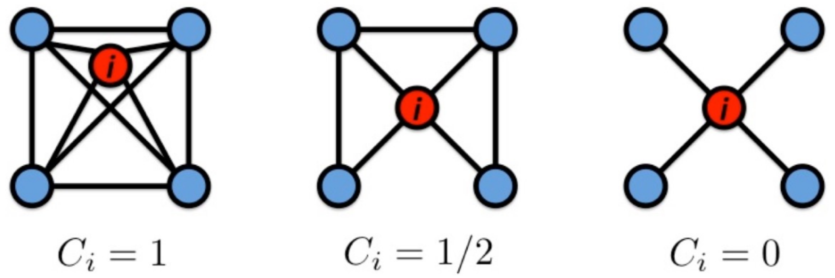
- From lecture 7: components are sets of nodes in which each node is reachable by some path from each of the others
- *k*-component: each node is reachable by at least *k* node-independent paths from each of the others



Clustering

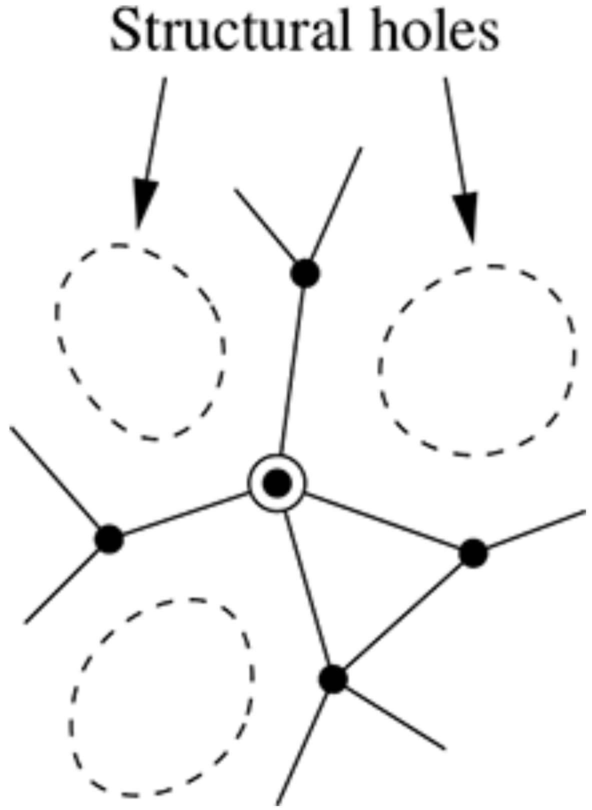


- Measure of extent to which any node's neighbors are themselves connected
  - How many triangles are closed
- For an individual node:
  - $c_i = \dfrac{\text{# of pairs of neighbors } i \text{ that are connected}}{\text{# of pairs of neighbors } i}$
  - $c_i = \dfrac{2e_i}{k_i(k_i-1)}$ where e_i is the number of edges between the neighbors of i
  - Also known as **local clustering**



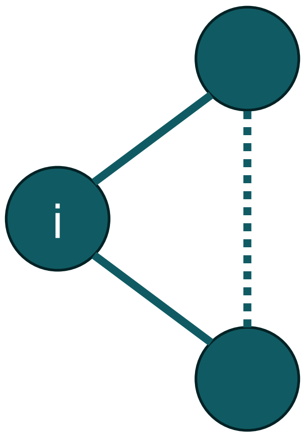
Local clustering coefficient

- Intuition: “the friend of my friend is also my friend”
- Offers empirical insights on network structure
- Indicates “structural holes” in a network
    - Expected connections that are missing
    - Indicates less redundancy/resilience (fewer alternative routes in the network)
    - Indicates when individual nodes hold higher control over flows
        - Local version of betweenness centrality



“Global clustering”

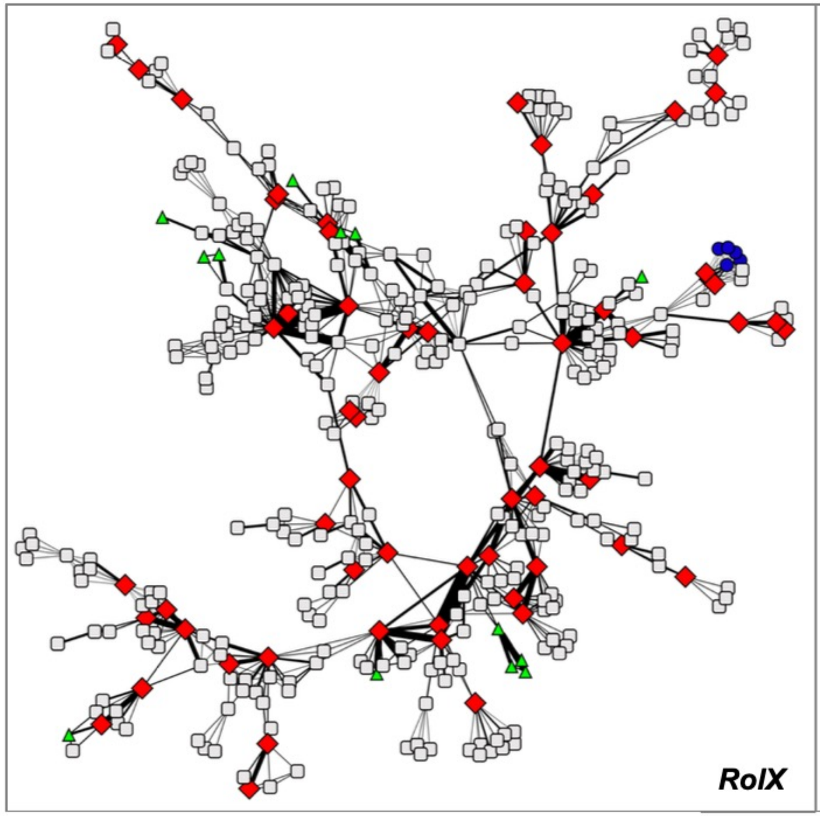
- Confusingly, there are two distinct metrics that are used to define clustering of the entire graph:



1. c_{avg} = Average of all the local clustering coefficients

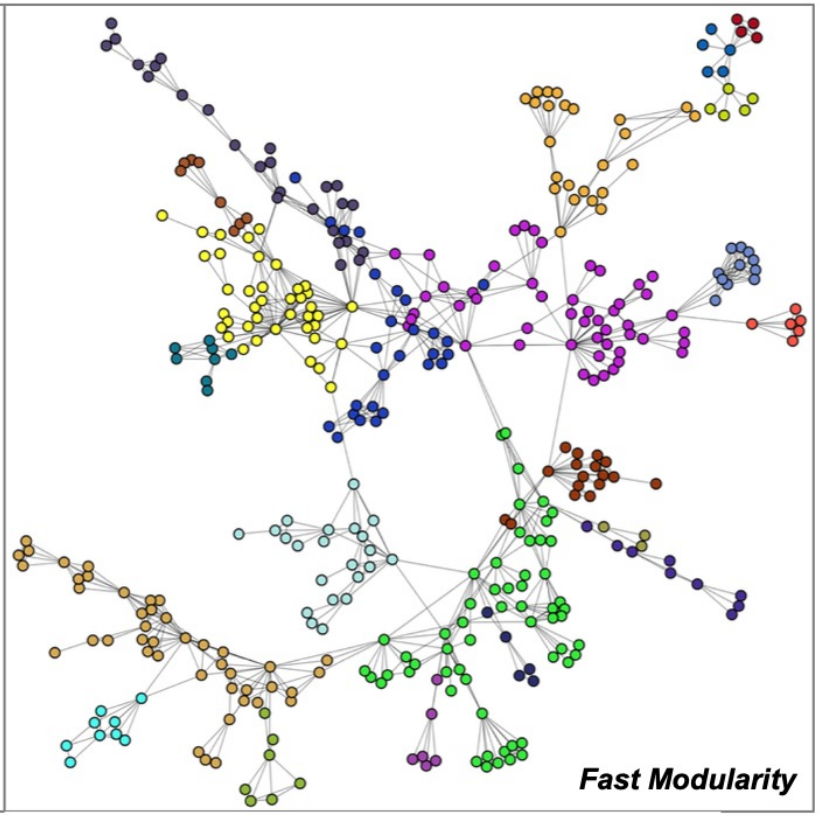
2. $C = \dfrac{\text{number of closed paths of length 2}}{\text{number of paths of length 2}}$

Roles



Henderson, *et al.*, KDD 2012

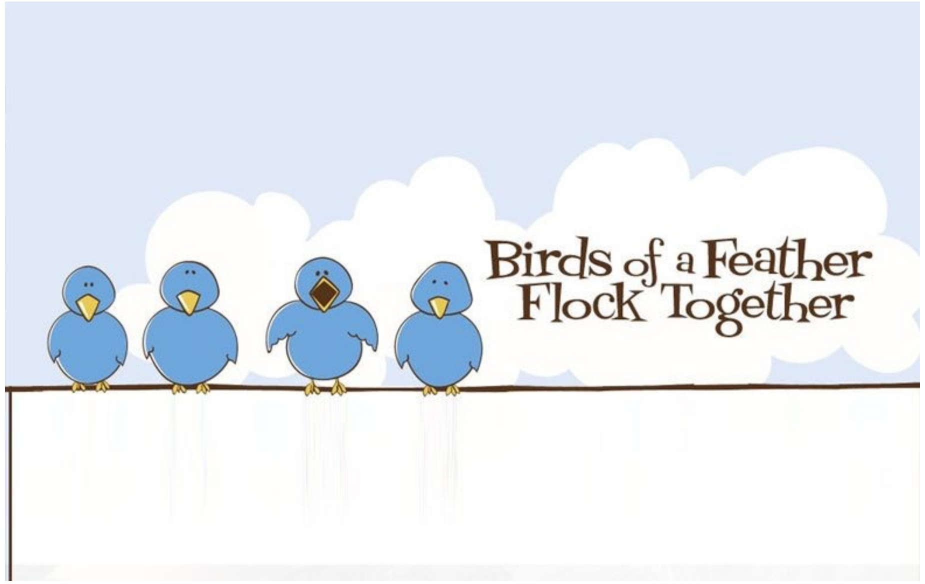
Communities



Clauset, *et al.*, Phys. Rev. E 2004

Homophily / Assortative mixing

- Tendency for nodes to have links to others that are similar in some way or are perceived to be similar
- Very common in social networks (classic example)
- Other examples:
    - Webpages
    - Urban systems with a social component (socio-technical)
        - Occupants in buildings
        - EV adoption and use



Our social networks can drive our behavior in the built environment

Network Effect	Definition	Energy Consumption Example
Homophily	A user tends to create relationships with other users who share similar characteristics	A user creates a relationship with a user who also enjoys computer gaming causing them to use their computer the same amount and have similar energy consumption
Confounding Factors	A user is exposed to similar external factors or stimuli as others in their peer network	Two users in the same peer network have the same work schedule causing them to adopt similar patterns of energy use and, as a result, to use similar amounts of energy
Influence	A user’s actions are triggered by the actions of another user in their peer network	A user uses less energy because they observe his/her friend to be using less energy

Modularity – measuring homophily

- Modularity measures how “well” the network is portioned into groups of different types

$$Q = \boxed{\text{Number of edges between nodes within a group}} - \boxed{\text{Number of edges expected at random within a group}}$$

Need a null model!

Example:
Nodes: locations
Edges: amount of travel between locations
Group: does the location have an EV charging station or not?

Modularity

1 if the groups are the same

Number of edges between nodes within a group

$$\sum_{\text{edges}(i,j)} \delta_{g_i g_j} = \frac{1}{2} \sum_{ij} A_{ij} \delta_{g_i g_j}$$

Modularity

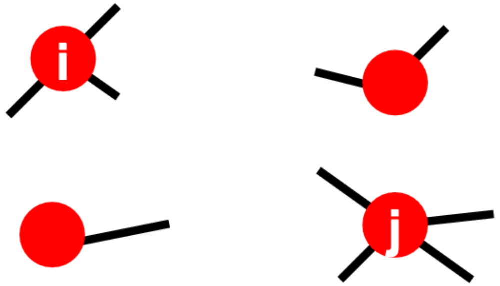
$$\tfrac{1}{2}\sum_{ij}\frac{k_i k_j}{2m}\delta_{g_i g_j}$$

Number of edges expected at random within a group

Need a null model!

Should be consistent with the characteristics of the network we are analyzing

number of edges coming from i = k_i



chance that a single edge from i attaches to j at random = k_j /2m

total number of ends of edges = 2m

Modularity

- Modularity measures how “well” the network is portioned into groups of different types

$$Q = \tfrac{1}{2}\sum_{ij} A_{ij}\delta_{g_i g_j} \quad - \quad \tfrac{1}{2}\sum_{ij} \frac{k_i k_j}{2m}\delta_{g_i g_j}$$

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta_{g_i g_j}$$

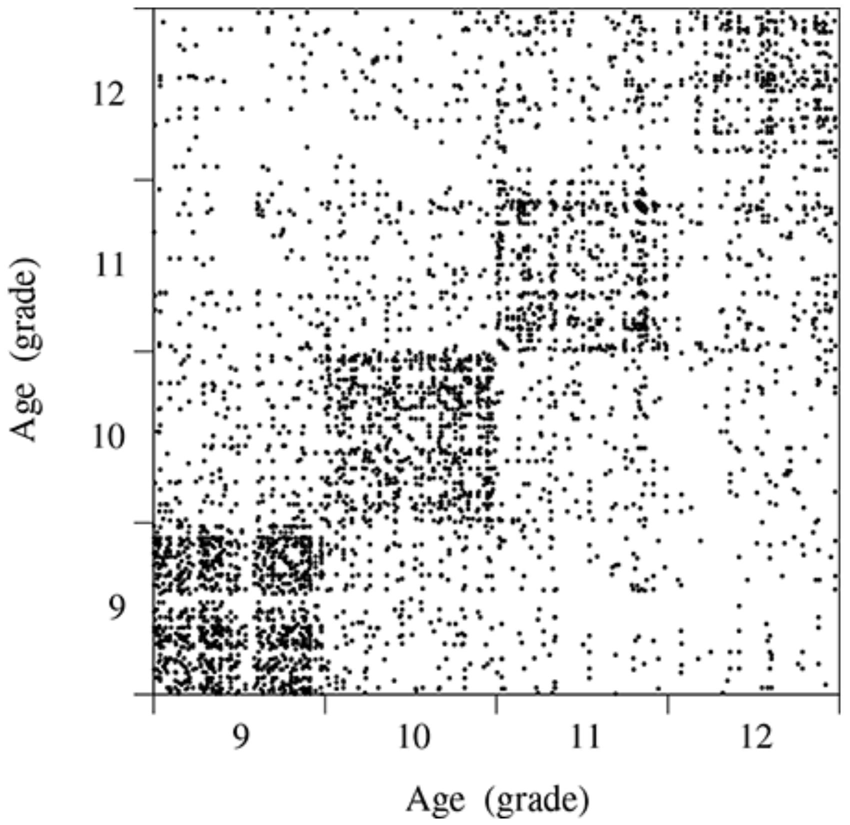
normalization so that -1 ≤ Q ≤ 1

Q between ~ 0.3 and 0.7 indicates high degree of assortative mixing

Modularity: extension to scalar “types”

- What if types are not categorical (e.g. age)?
  - Example: friendships in a school based on age
- Instead of Kronecker delta, we can use scalars indicating the values

$$\frac{1}{2m}\sum_{ij}\left(A_{ij}-\frac{k_i k_j}{2m}\right)x_i x_j$$



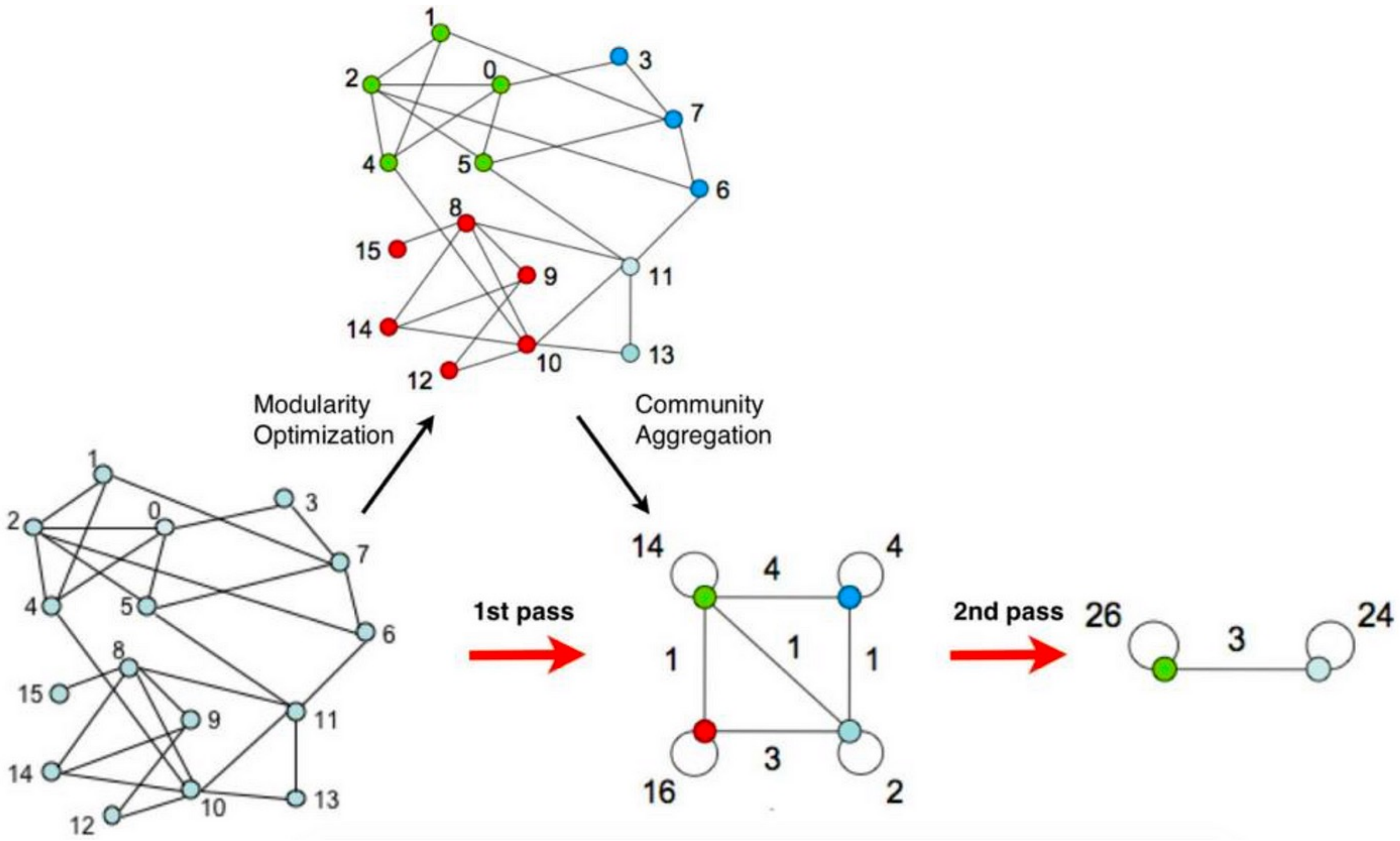
Modularity

- Indicates the extent to which we see assortative mixing / homophily in an existing network with assigned groups
- Useful if groups are naturally defined to characterize degree of homophily
- What if groups are not defined?
- Idea: try to assign groups in a way that maximizes modularity → discover the communities in the network!

Modularity maximization – community detection

- Different algorithms – now you have the tools to understand them!
- One common one is the Louvain algorithm (greedy algorithm):
  - Round 1: nodes:
    - Assign each node to its own group (n groups)
    - Pick one node and simulate the impact on modularity when that node is assigned to each of the other groups
    - Choose step that has the largest impact on modularity
    - Repeat
  - Round 2:
    - Same procedure but now for groups instead of nodes
  - Round X:
    - Repeat as many times as necessary until there are no moves that increase modularity

Louvain algorithm

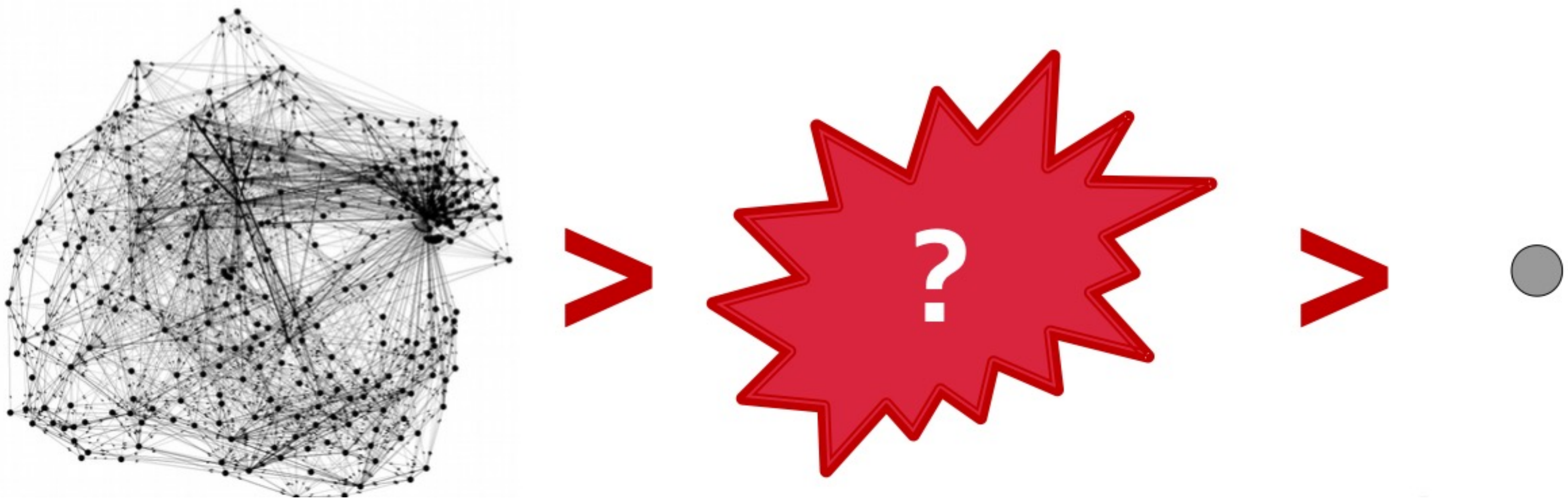


Network metrics

Network level
- Global clustering
- Modularity
- Communities
- Component properties

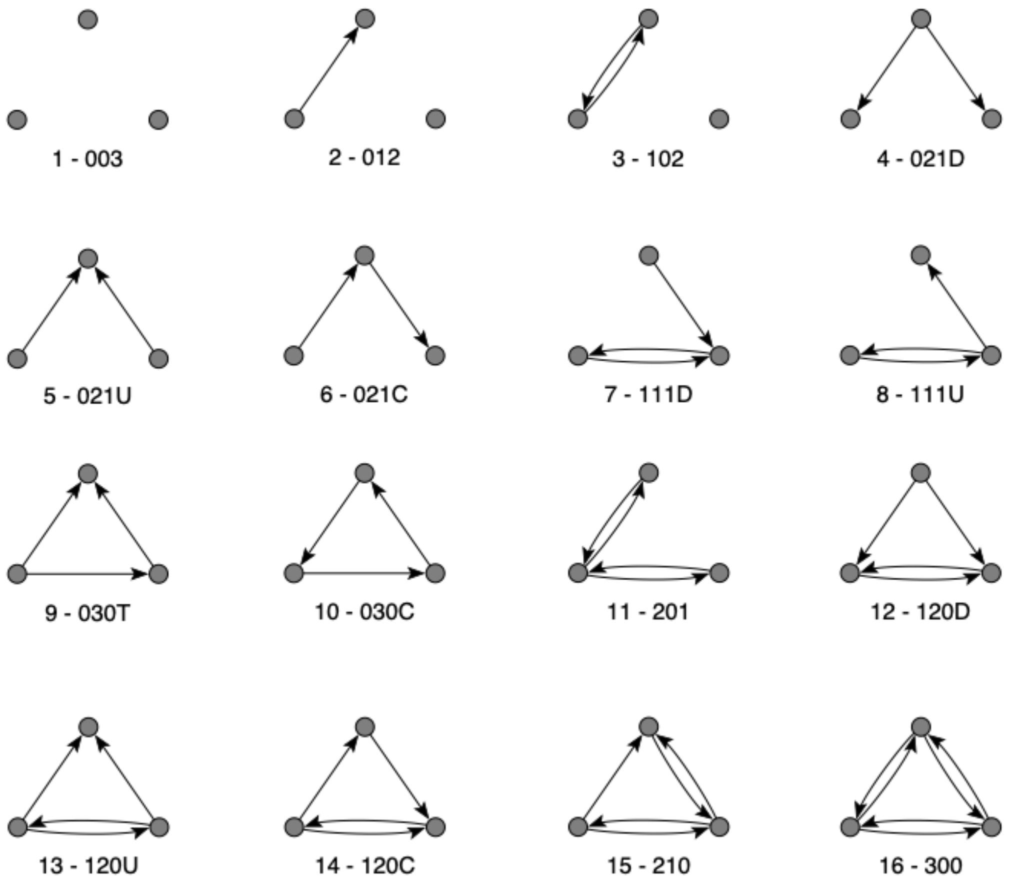
“Meso-scale”???

Node level
- Degree
- Centrality
- Node clustering
- …



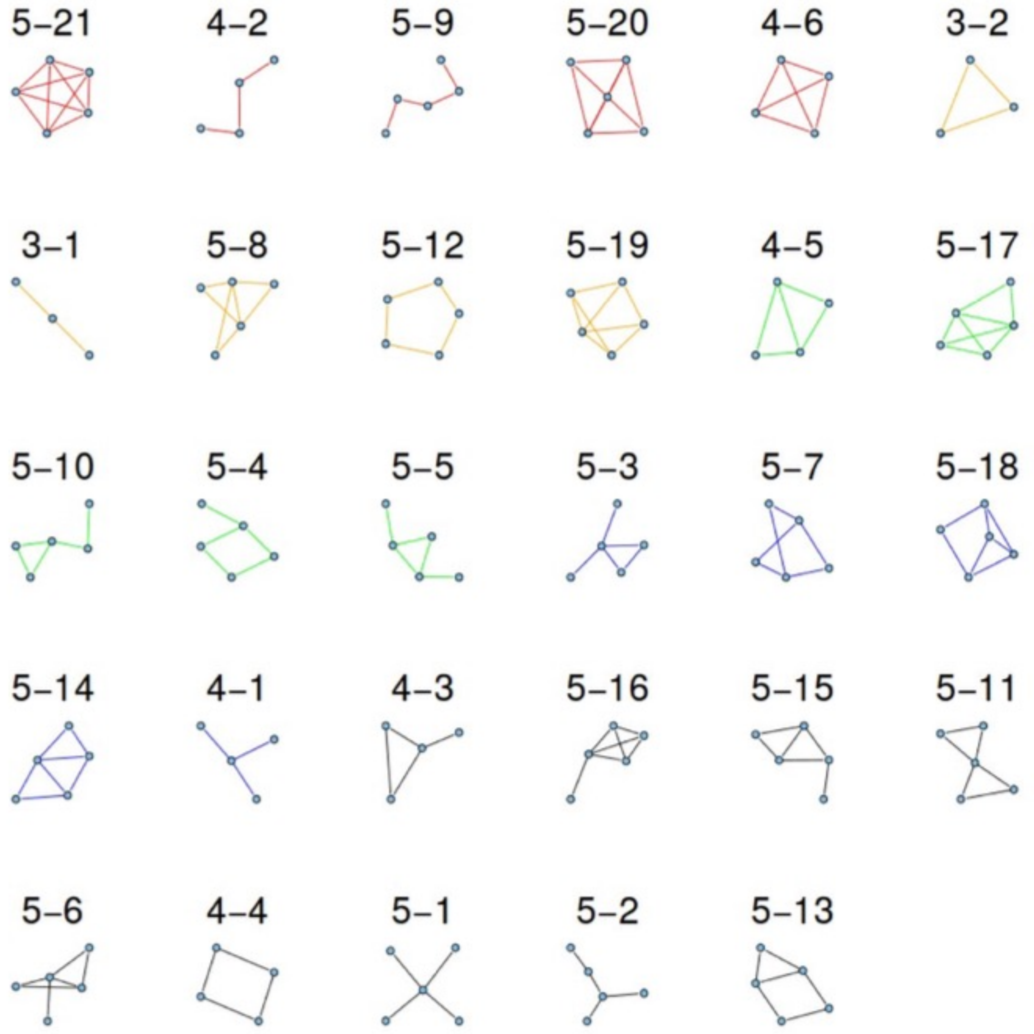
Network motifs

- Example: What are all the possible configurations of directed subgraphs of size 3?
- Are some of these over-represented or under-represented in a particular network?
- Idea: If certain structures occur more often than you would expect at random, they have some functional significance



Network motifs

- Undirected example of motifs of size 3-5



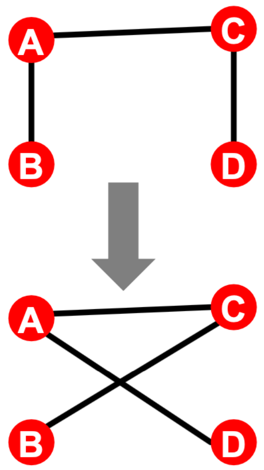
Network motif significance

- How does representation of a motif compare to a randomized network?
- Calculate Z-score for motif *i*:

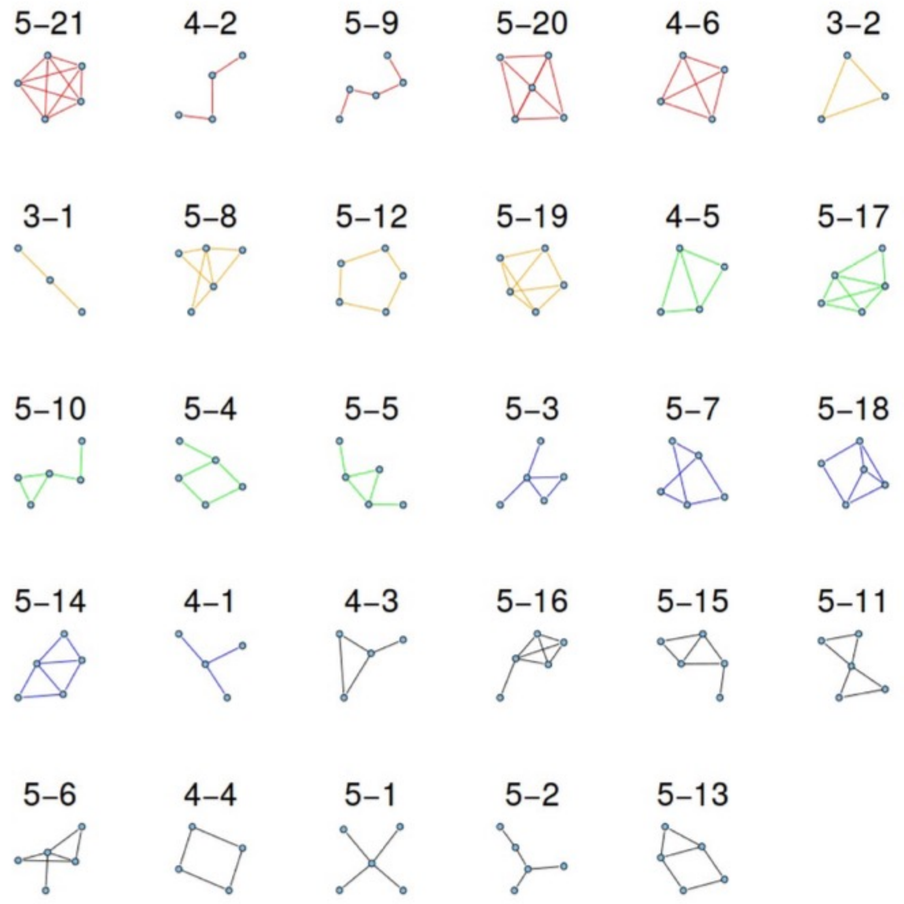
$$Z_i = \frac{\left(N_i^{\text{real}} - \overline{N_i^{\text{rand}}}\right)}{std\left(N_i^{\text{rand}}\right)}$$

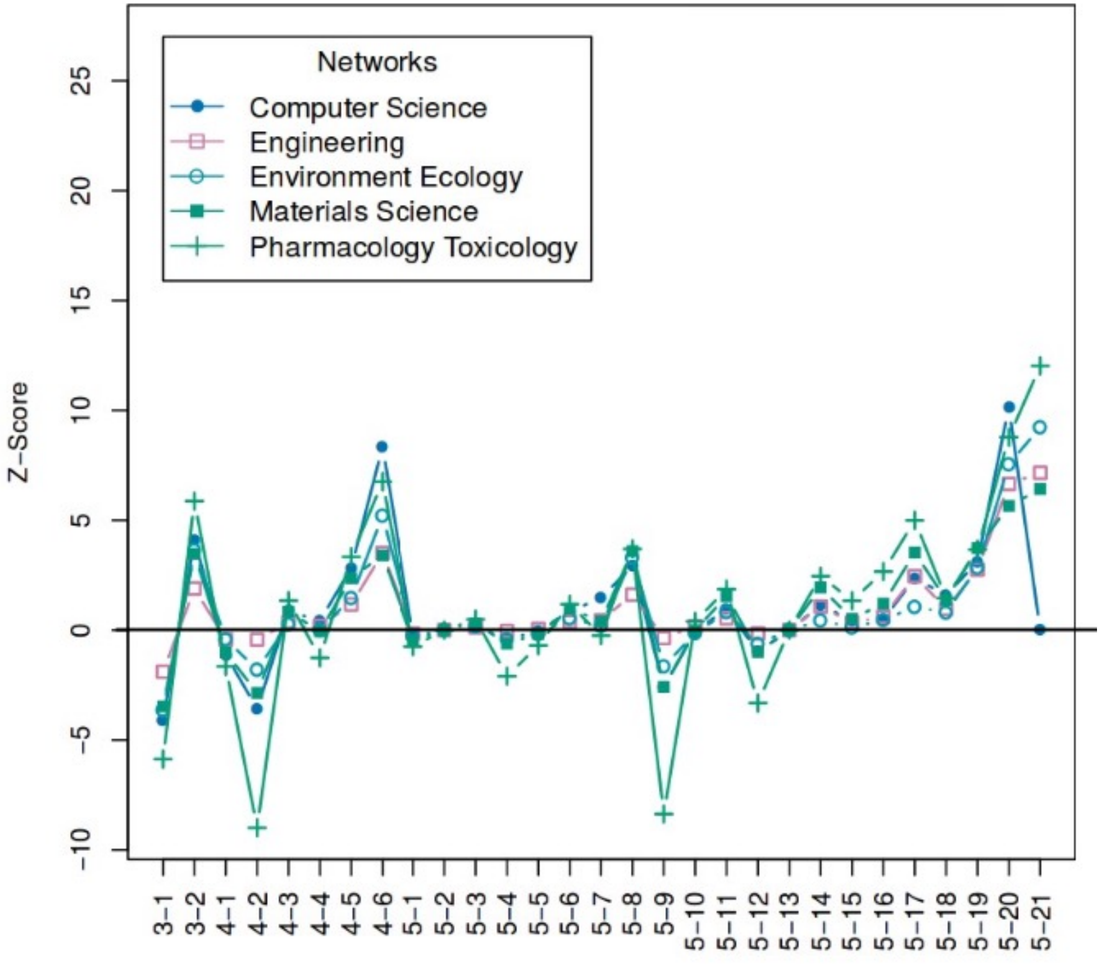
How to generate randomized network?

- Configuration model: enforce degrees of nodes are the same as the original network, but randomly assign edges
- Network “rewiring”:
  - Select 2 edges at random (A→B and C→D; assume directed in this case)
  - Exchange the endpoints to give A→D and C→B
    - Only do the exchange if it does not create a multi-edge or a self-loop
  - Result: randomly rewired graph



“Significance profile”





Network significance profile