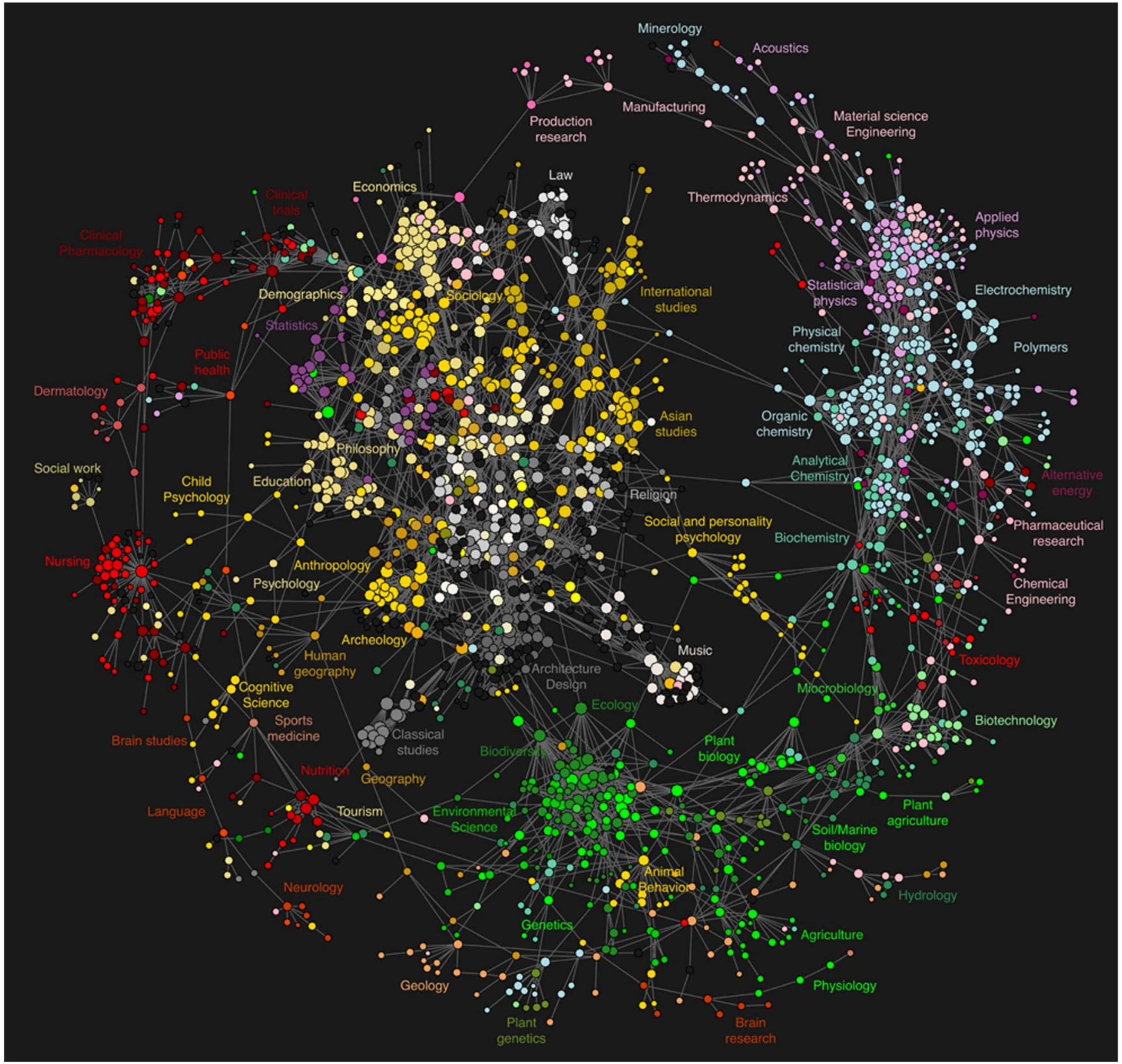
Subgroups

Goals

- Discuss why subgroup detection is important, useful
- Understand difference between groups that are internally cohesive, vs. groups defined by their structural equivalence
- See how to identify subgroups in igraph

Does this ‘map of science’ have subgroup structure?



https://doi.org/10.1371/journal.pone.0004803.g005

Subgroups

How to detect subgroups and communities in small and large networks

Subgroups- Introduction

- Fundamental idea
  - Identification of cohesive subgroups of actors in a larger network
  - Cohesive subgroups are subsets of actors which have relatively strong, direct, intense, frequent, or positive ties (W&F, p. 249)
- Theoretical roots
  - Cohesive subgroups are an attempt to define or model social groups
  - Important theoretical approaches
    - Sociology – emergence of consensus
    - Psychology – homophily, peer pressure

Classes of Subgroups

Class	Operationalization	Examples
Cohesive Subgroups	Group based on *internal* ties	Cliques, k-plex, cores
Roles	Group based on *external* ties	Structural equivalence
Communities	Group based on both *internal and external* ties	Louvain, Girvan-Newman

Operationalization

- Types of measures
  - Mutuality of ties
  - Closeness
  - Frequency of ties
  - Relative frequency of ties within a group compared to outside the group
- Cohesion methods are primarily concerned with internal ties; stands in contrast to concepts of social structure and roles

Approaches

- Some common approaches to subgroup identification and analysis
  - Components
  - Cliques
  - Cores
  - Plexes
  - Community detection

Cohesive subgroups

Detecting subgroups based on strong, frequent internal ties

Components

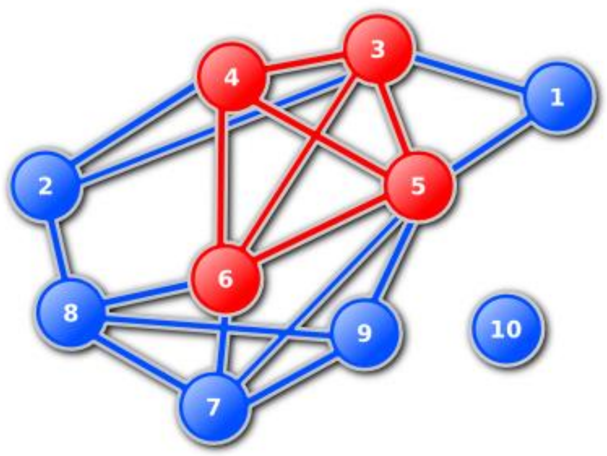
- Simple approach to subgroup detection. Components identify the disconnected ‘lumps’ in a network.
- Remember, a component is a maximally connected subgraph.
- May be more useful for directed graphs, where strong or weak components can be identified
- Component analysis can be useful to identify network islands, but the internal connections within those islands may not be particularly dense, reciprocated, etc.

Cliques

- Definition of clique based on complete mutuality. A clique is a subnetwork that has all possible ties present
- Formal definition – a clique is a maximally complete subgraph of three or more nodes.
  - Compare to definition of component
  - Maximally connected vs. maximally complete

Cliques

- Simplest, oldest subgroup definition
- Cliques may not be particularly useful in real applications
  - Conservative definition – loss of only one tie removes a clique
  - Not found in sparse networks
  - May not reflect real social groups



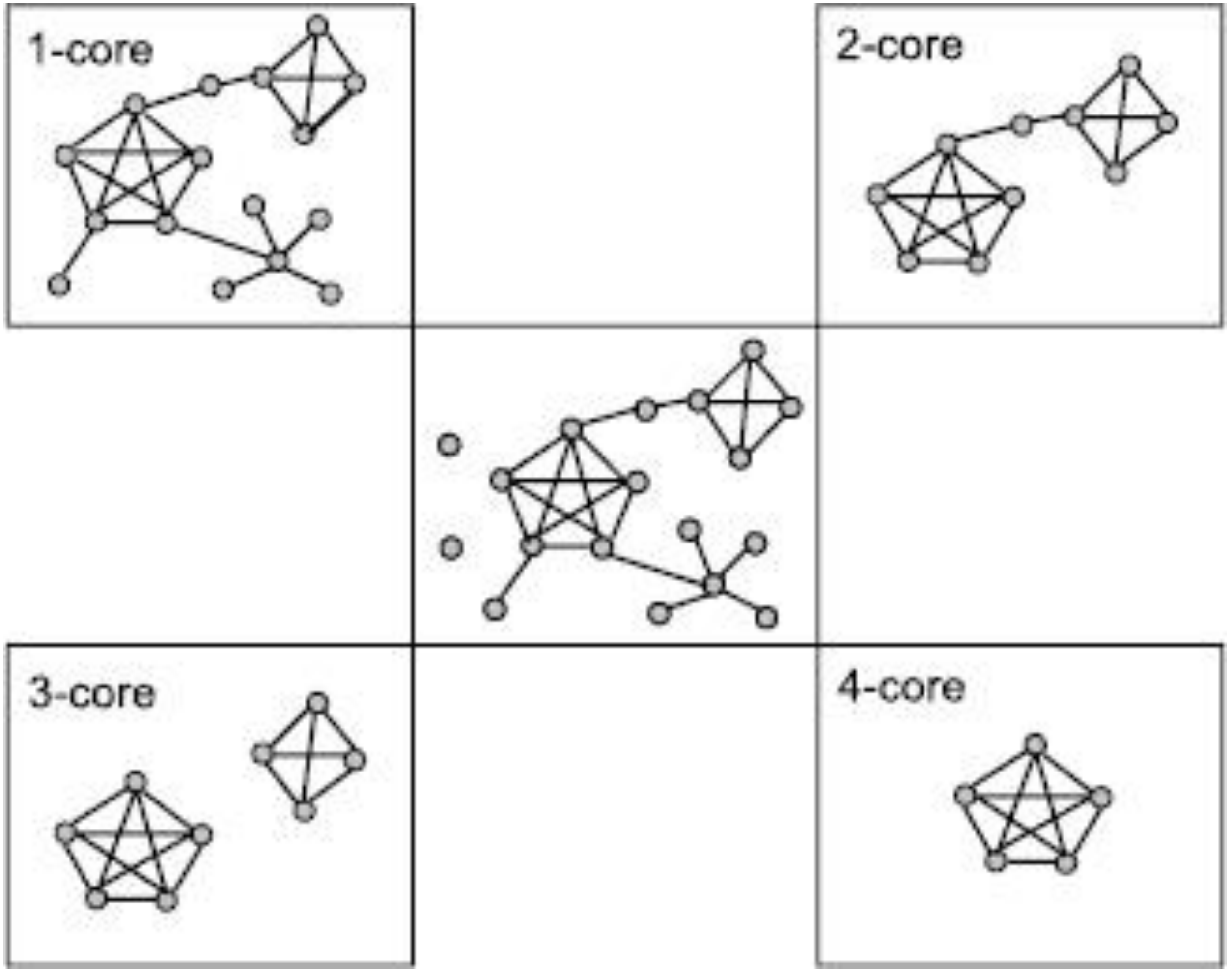
n-Cliques

- Based on geodesics
- An *n*-clique is a maximal subgraph where all geodesics are less than or equal to *n*.
  - A 1-clique, is simply a clique
  - A 2-clique is a subgraph where all pairs of actors are connected with a path no longer than 2 steps.

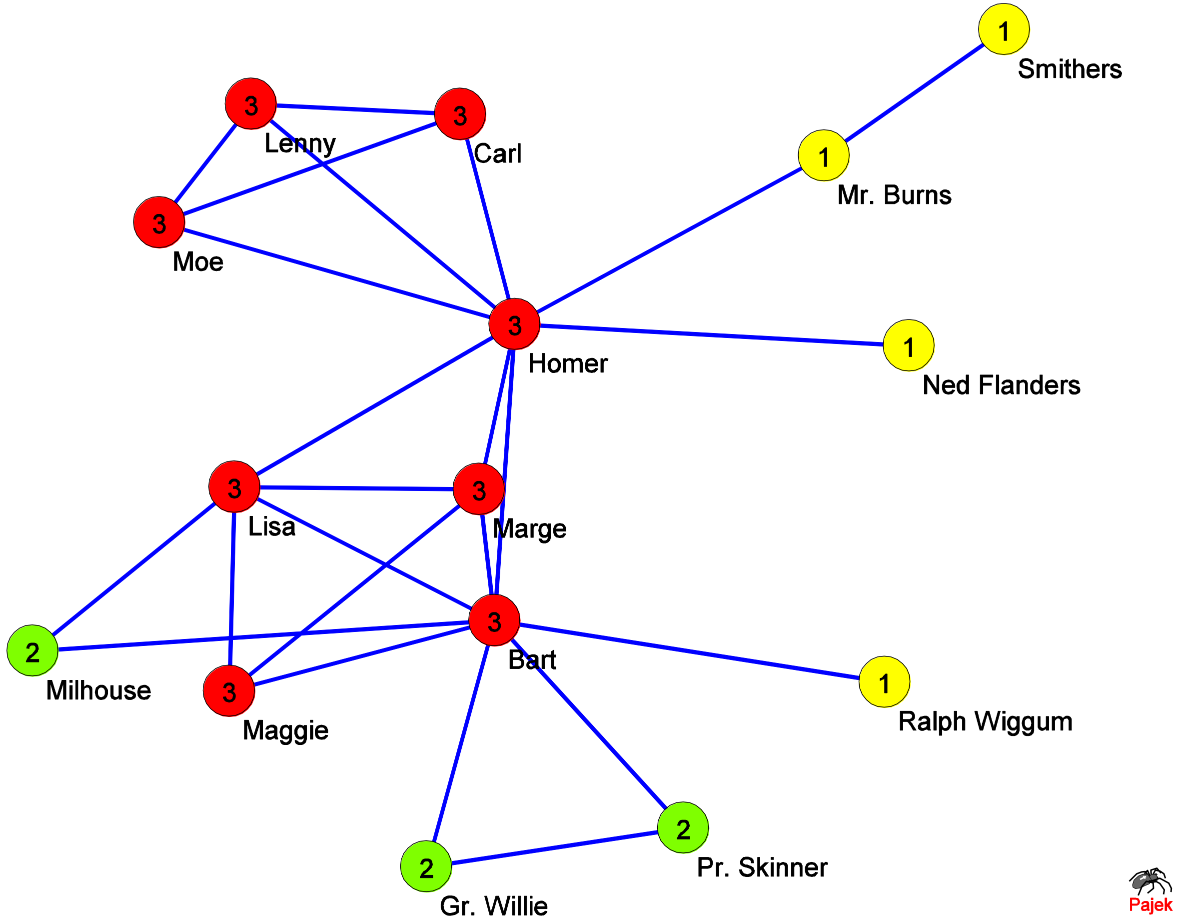
Cores

- Based on degree
- A *k*-core is a maximal subnetwork in which each node has at least degree *k* in the subnetwork.
- Hierarchical nature of cores (See de Nooy, pp. 71-72)
- Nice approach to working with cores: Eliminate lower-order cores until relatively dense subgroups are identified.

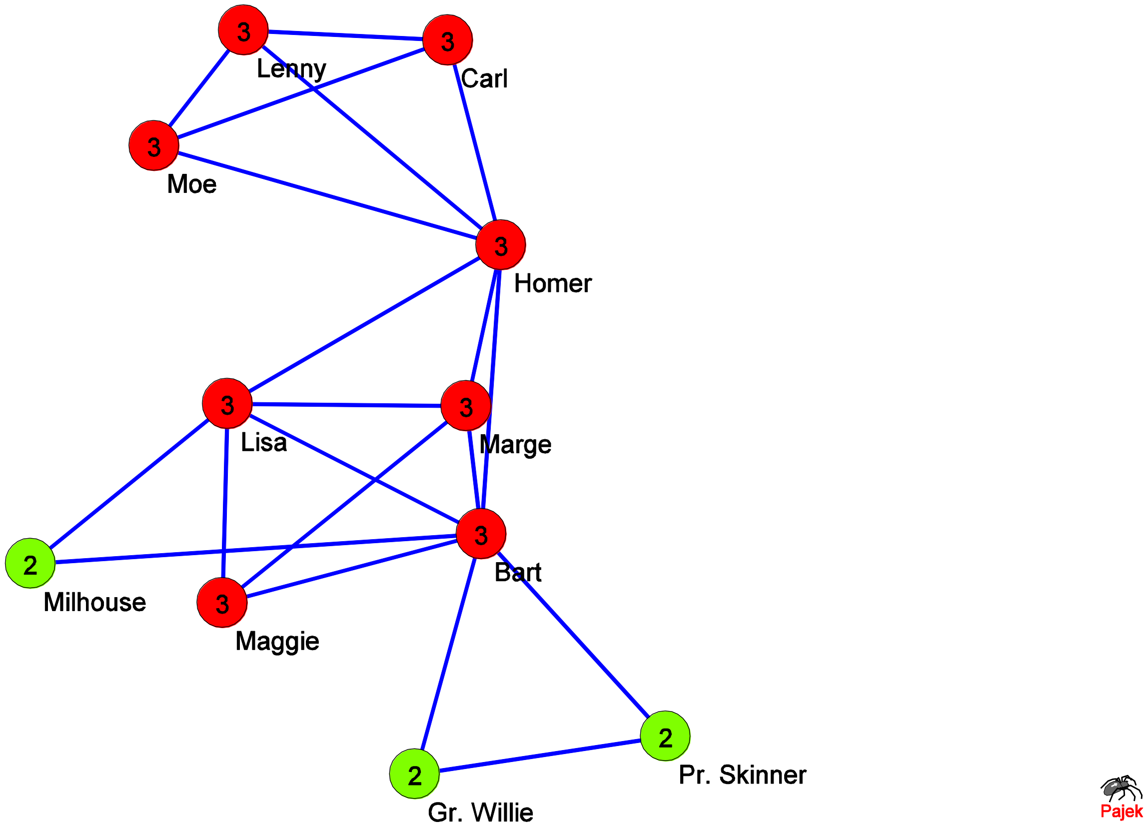
Cores - Example



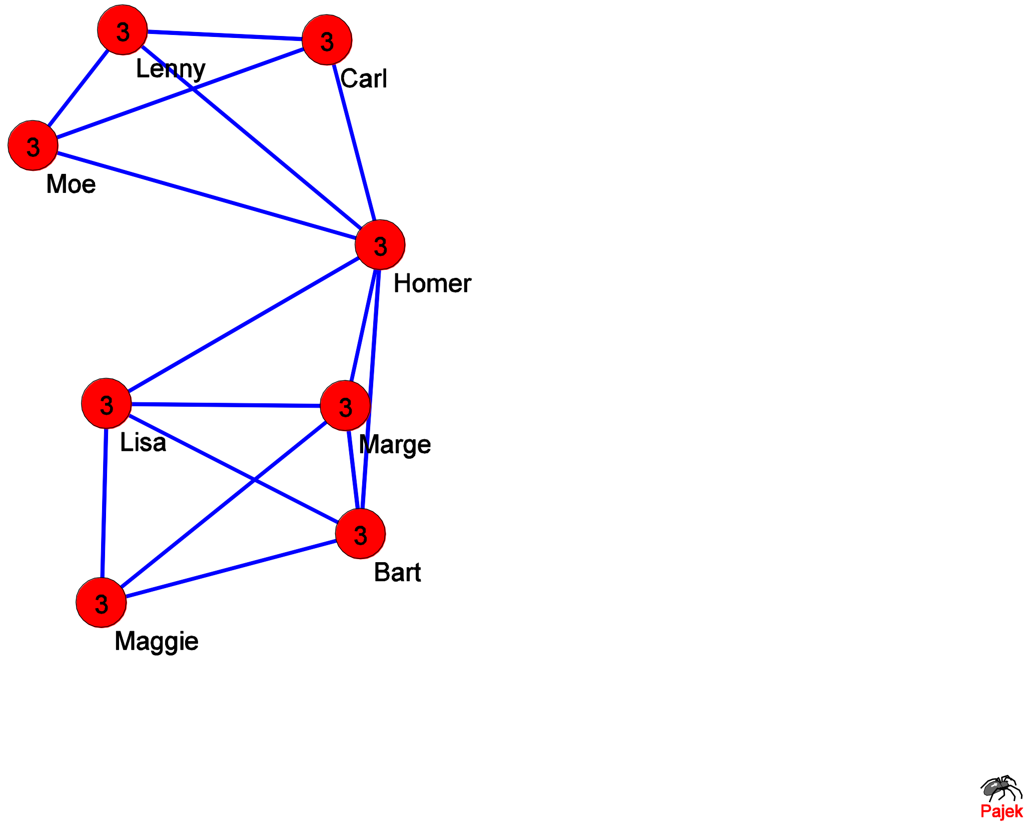
Simpsons graph – Cores 1, 2, 3



Simpsons graph – Cores 2, 3



Simpsons graph – Core 3



Plexes

- A *k*-plex is a maximal subgraph where no more than *k* direct connections are allowed to be missing between all pairs of actors
- Can be thought of as a complement to the concept of cores.
- Plexes can be useful to identify subgroups with small diameter, if *k* is small relative to the number of nodes.

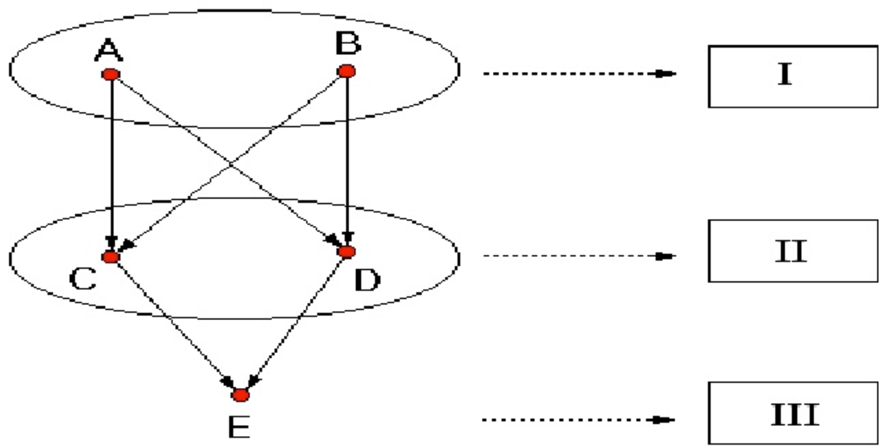
Roles

Detecting subgroups based on structural equivalence

Definition of structural equivalence

- Two actors are structurally equivalent if they have identical ties to and from all other actors in the network
- Tightly tied to notions of roles, positions, social class
- Methodologically tied to cluster analysis, block modeling

cluster, structural equivalence, block modeling



Communities

Detecting subgroups based on overall pattern of both internal (frequent) and external (not frequent) ties

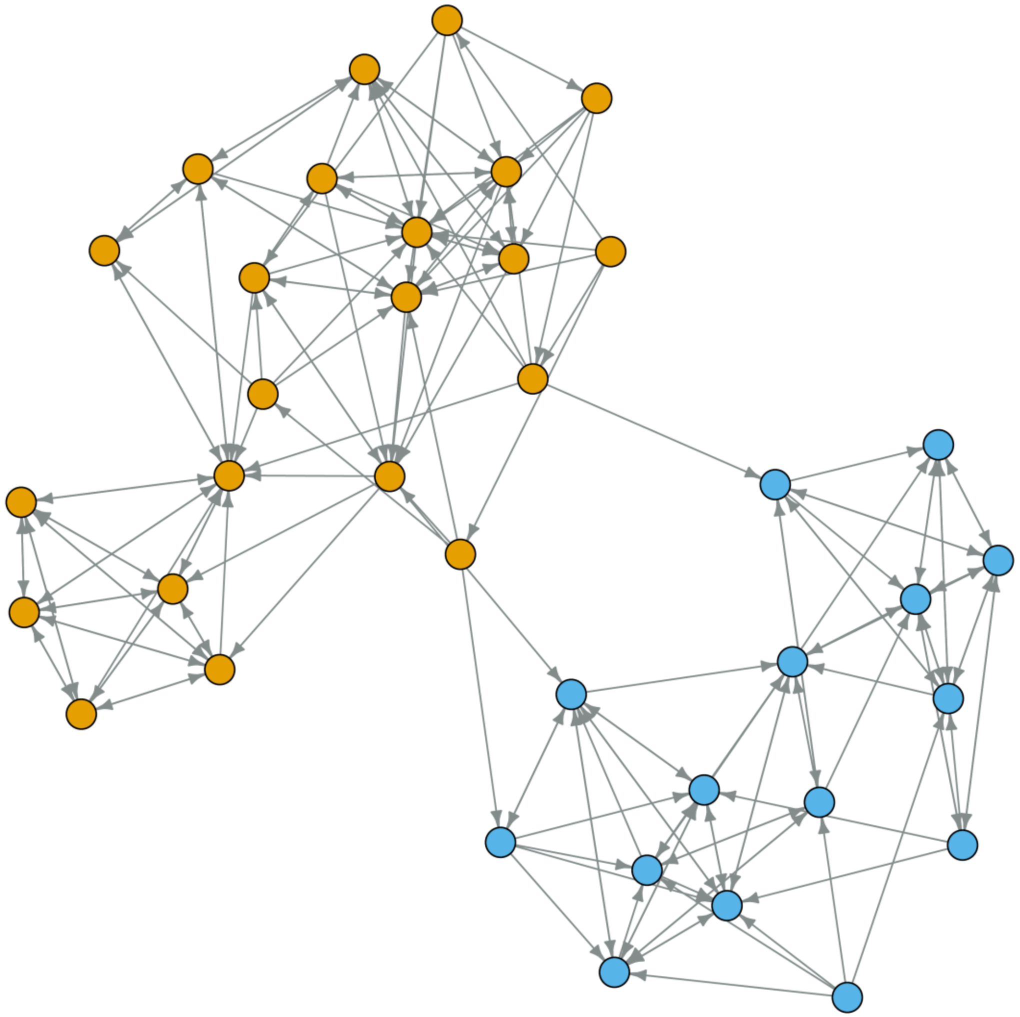
Detecting communities in networks

- Limitations of previous methods
- Look at both internal and external ties to determine groups (high internal, low external)
- Important ability, but challenging for large networks
- See http://en.wikipedia.org/wiki/Community_structure

Girvan-Newman Method

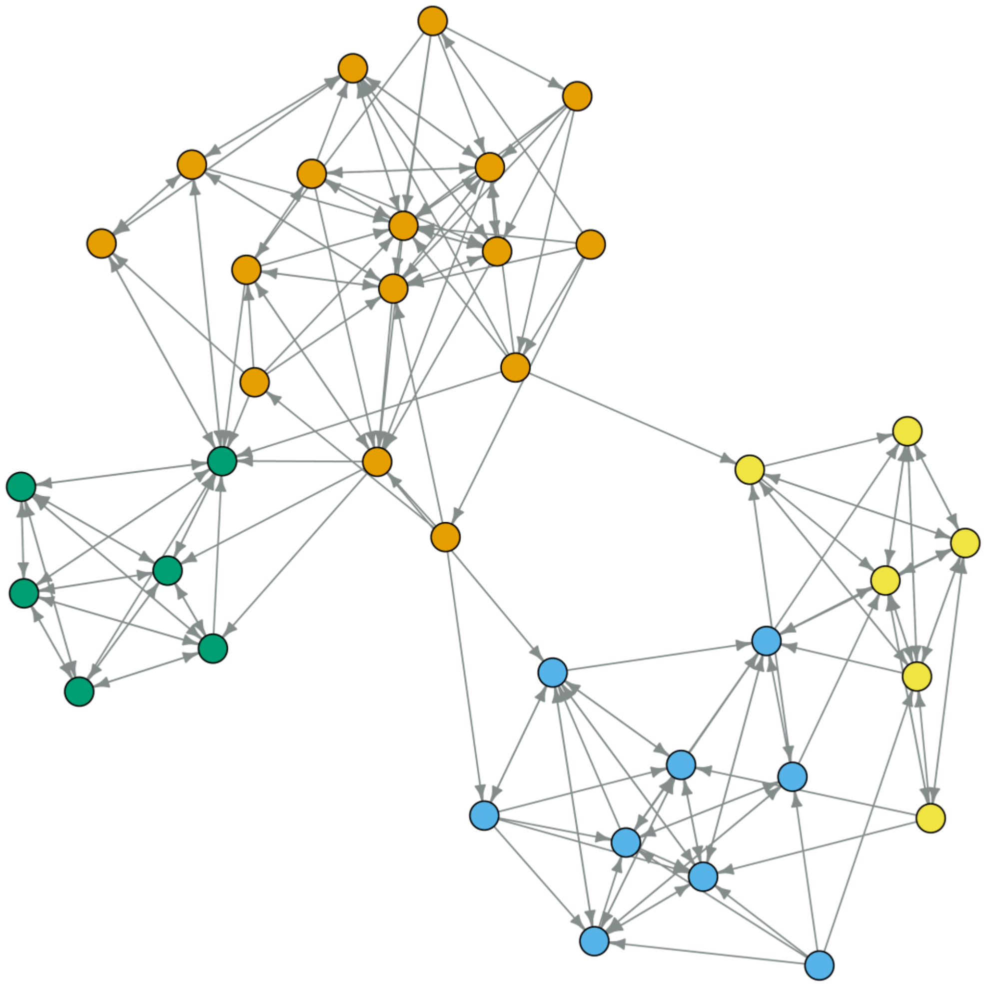
- Girvan-Newman
  - Calculate *edge* betweenness
  - Remove edge with highest betweenness
  - Repeat, while looking for components
  - See http://en.wikipedia.org/wiki/Girvan%E2%80%93Newman_algorithm
  - Called 'cluster_edge_betweenness' in igraph

Example – Friendship network

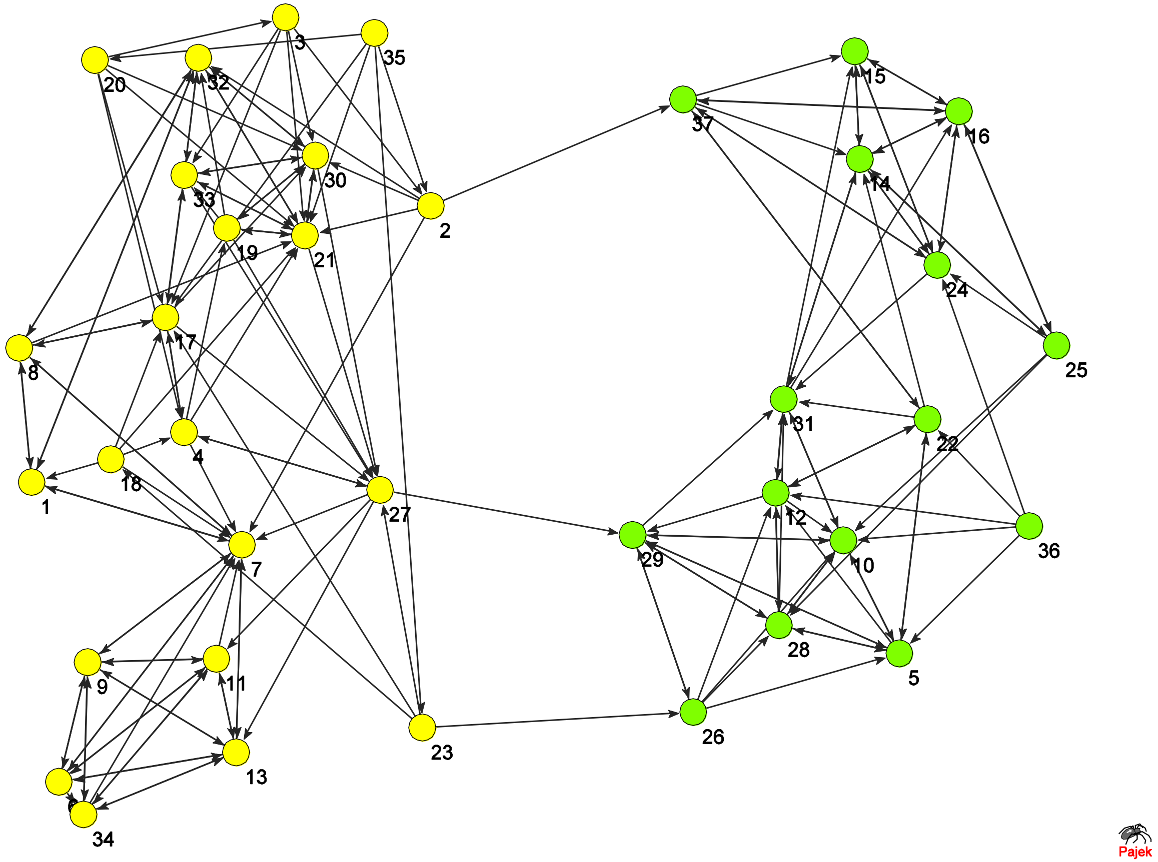


Data from Valente, 2010

Four communities detected

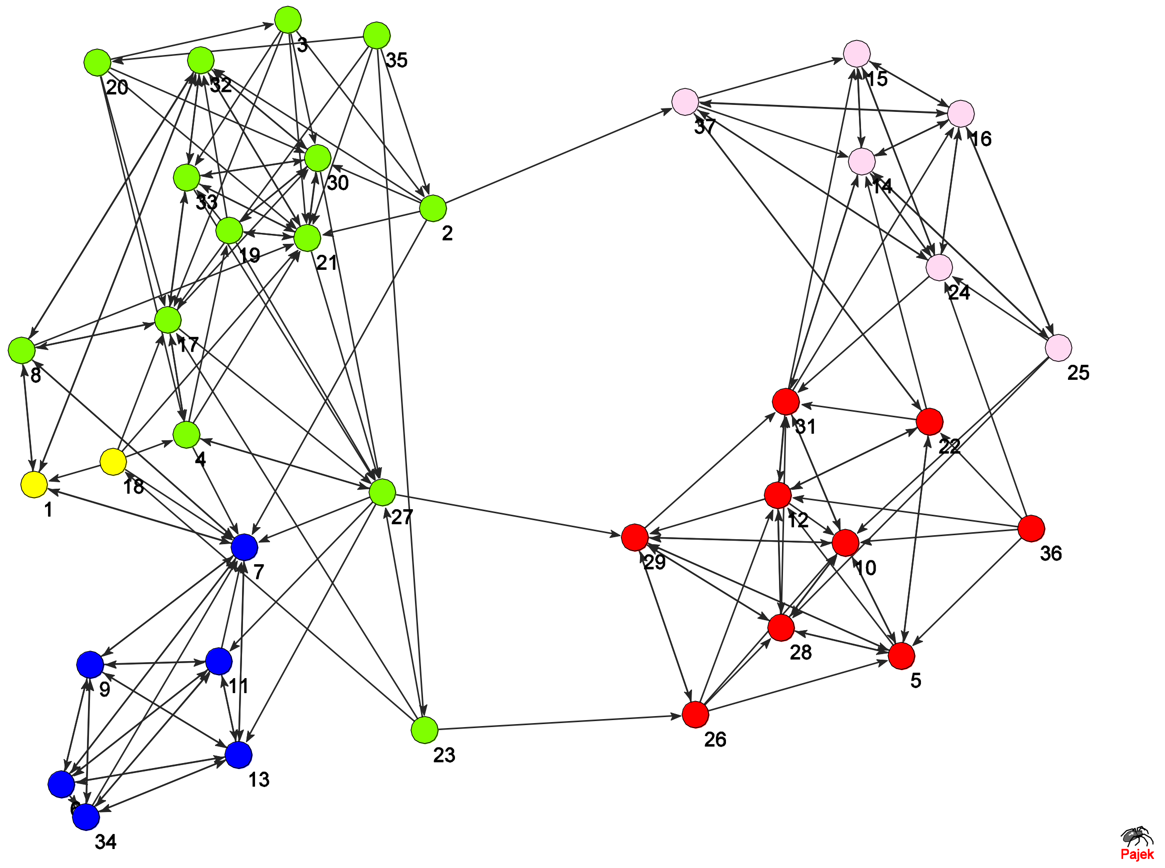


Example – Friendship network



Data from Valente, 2010

Five communities detected



Louvain Communities method

- Heuristic approach (greedy optimization)
- Two stages
  - Identifies small communities by optimizing *modularity* in a local way.
  - Aggregates nodes of the same community and builds a new network whose nodes are the communities
- Called ‘cluster_louvain’ in igraph
- Faster than Girman-Newman, thus useful for large networks

Modularity

- Basic measure of the structure of a network
  - Measures the strength of division of a network into clusters or communities.
  - Networks with high modularity have dense connections between the nodes within a cluster, but sparse connections between nodes in different clusters.
- Definition
  - Given a partition of the network, the modularity is the fraction of the edges that fall within the given clusters minus the expected such fraction if edges were distributed at random.
  - The value of the modularity lies in the range $[-1/2, 1]$

(From Blondel, 2008)

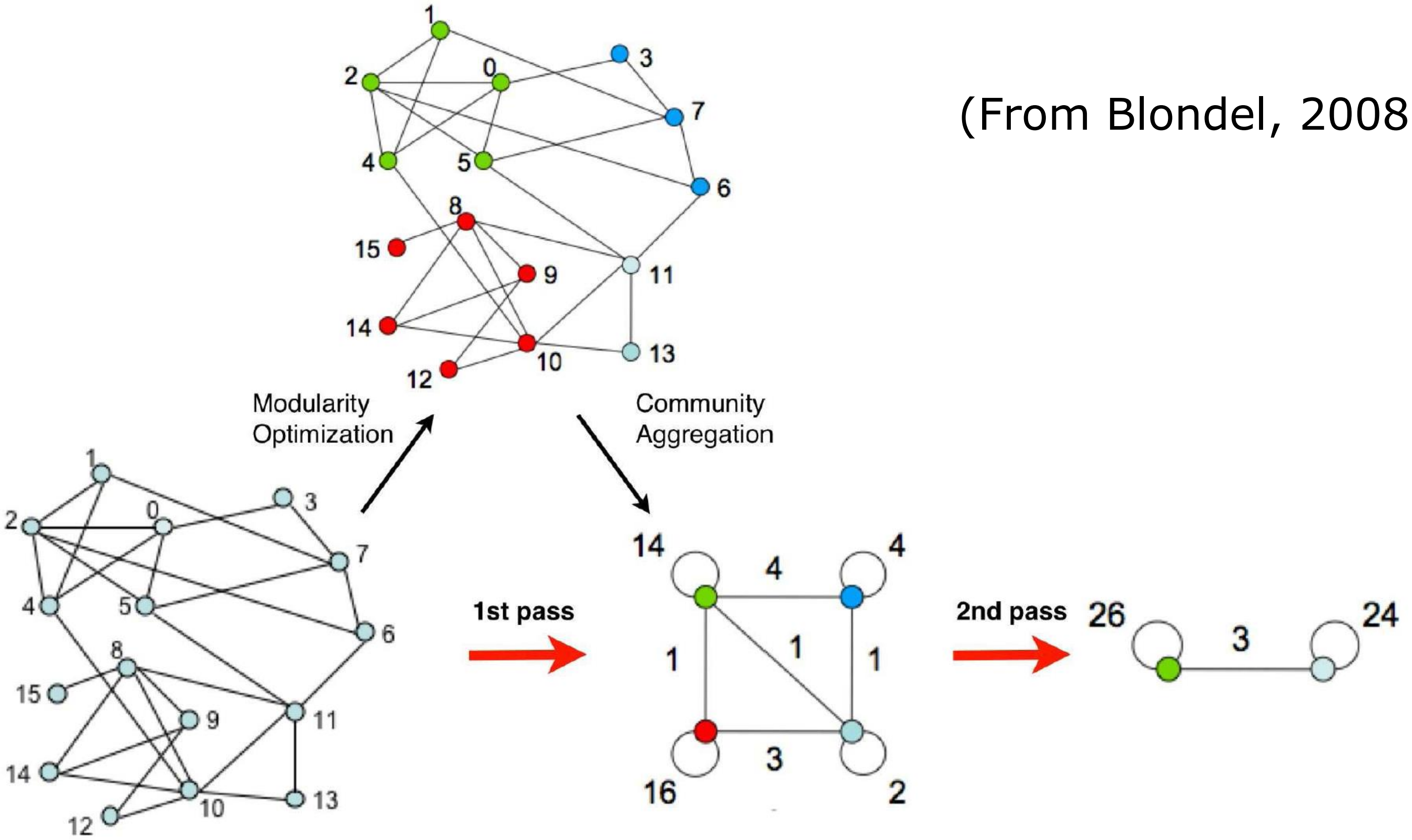
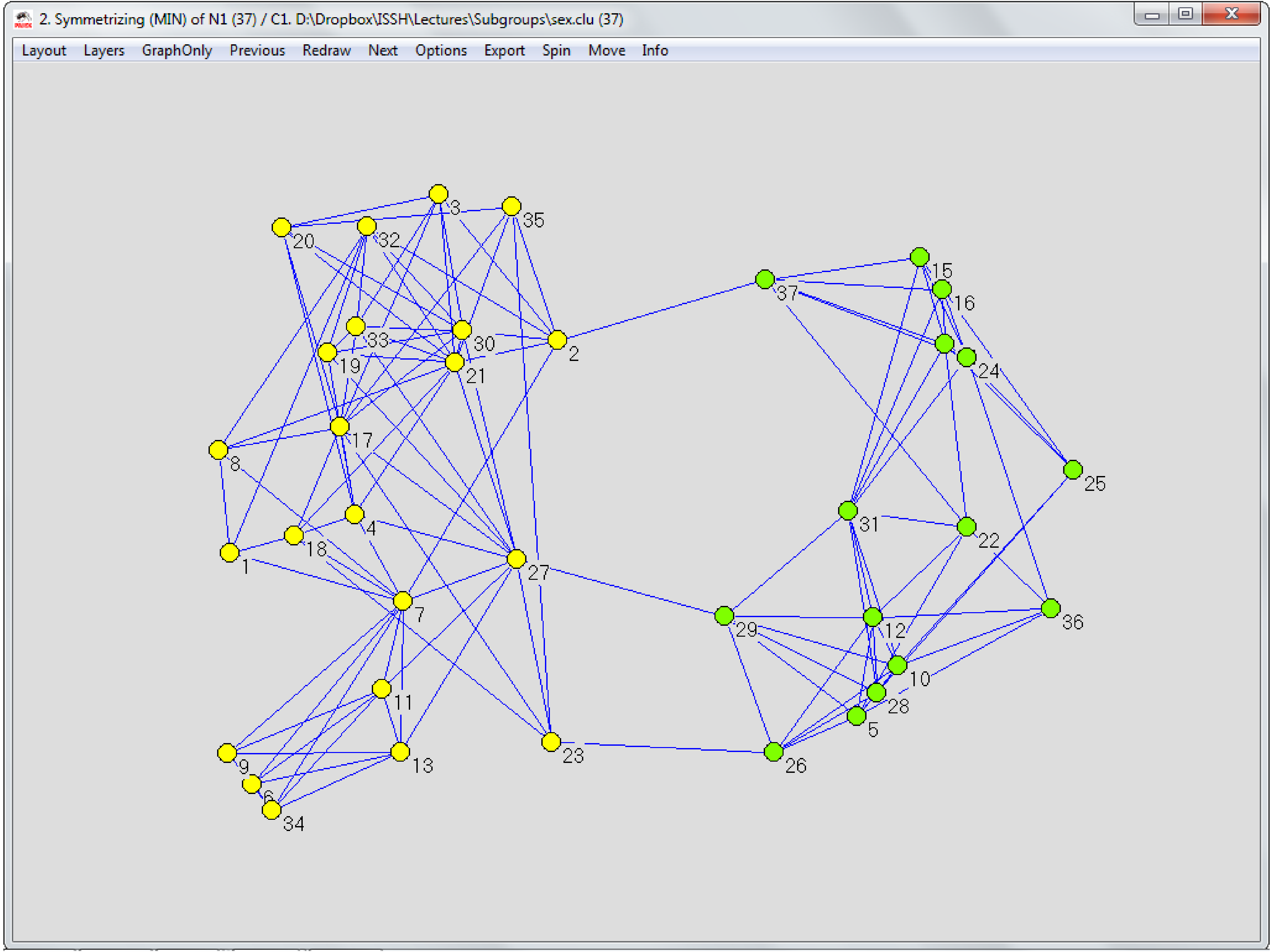
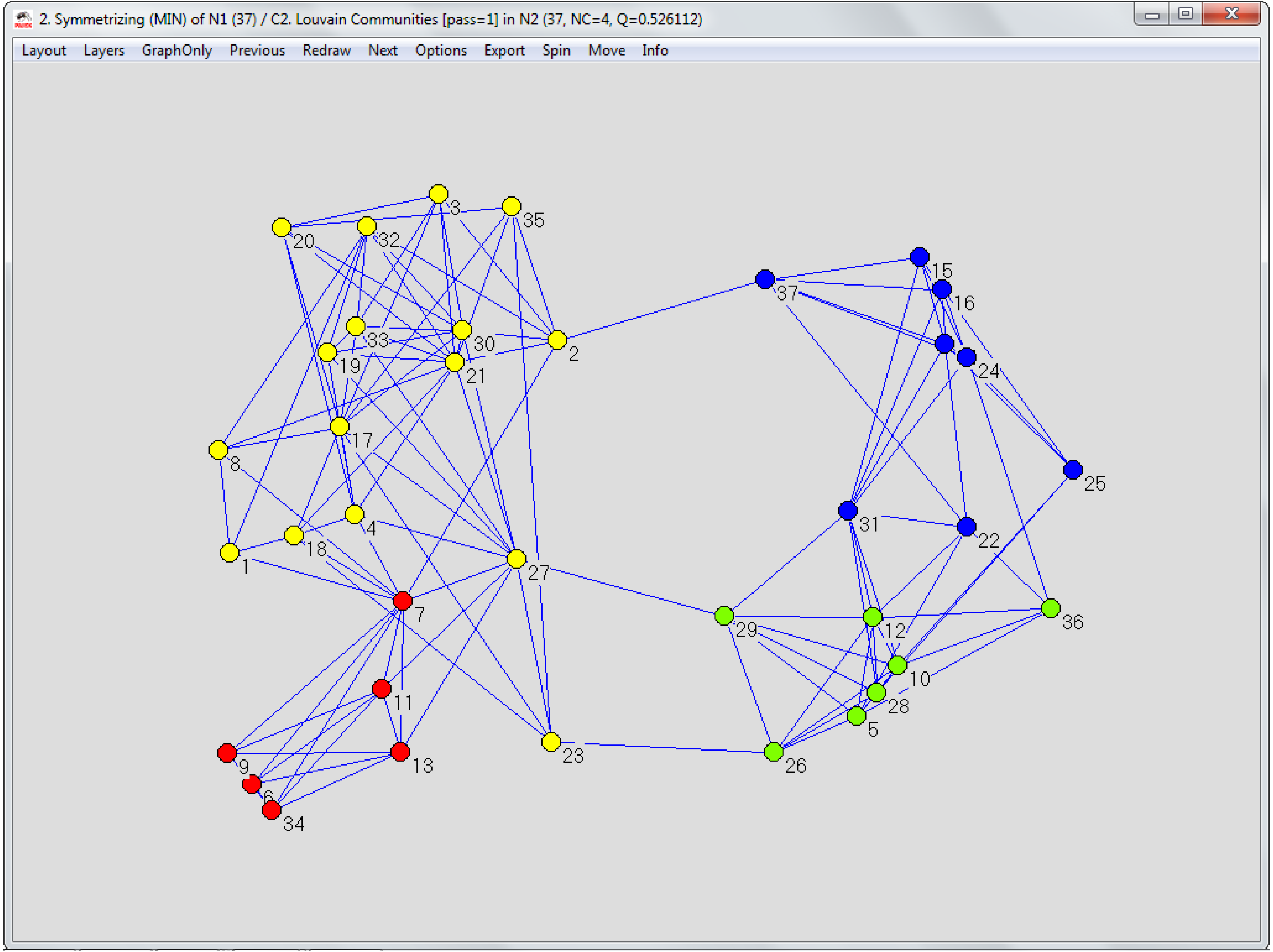
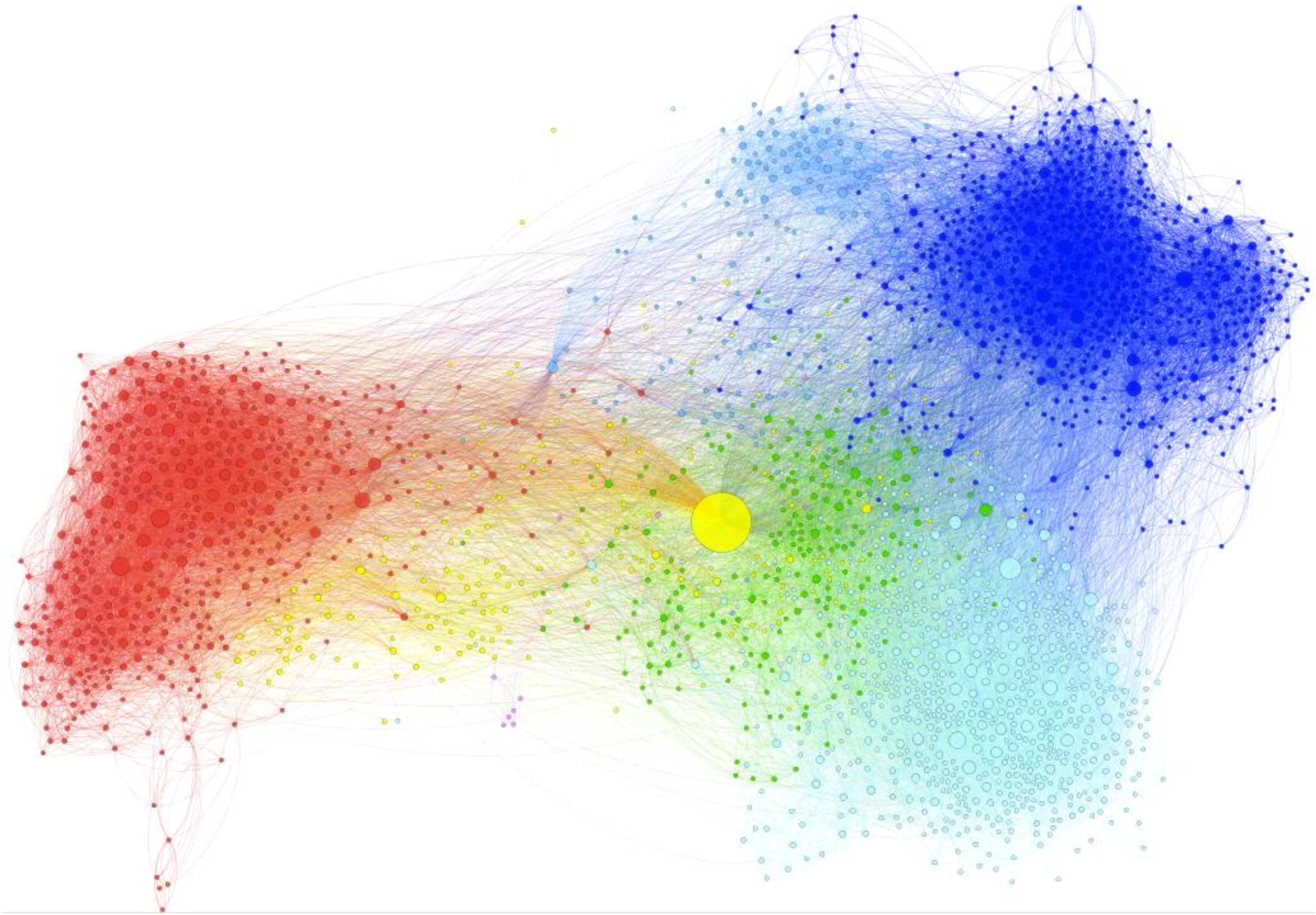


Figure 1. Visualization of the steps of our algorithm. Each pass is made of two phases: one where modularity is optimized by allowing only local changes of communities; one where the found communities are aggregated in order to build a new network of communities. The passes are repeated iteratively until no increase of modularity is possible.


2. Symmetrizing (MIN) of N1 (37) / C1. D:\Dropbox\ISSH\Lectures\Subgroups\sex.clu (37)
Layout
Layers
GraphOnly
Previous
Redraw
Next
Options
Export
Spin
Move
Info
3
35
20
32
33
30
2
19
21
17
8
4
18
1
27
7
11
9
13
6
34
23
15
37
16
24
25
31
22
29
12
36
10
28
5
26


2. Symmetrizing (MIN) of N1 (37) / C2. Louvain Communities [pass=1] in N2 (37, NC=4, Q=0.526112)
Layout Layers GraphOnly Previous Redraw Next Options Export Spin Move Info

Modularity & community detection useful for large networks



http://www.martingrandjean.ch/introduction-to-network-visualization-gephi/