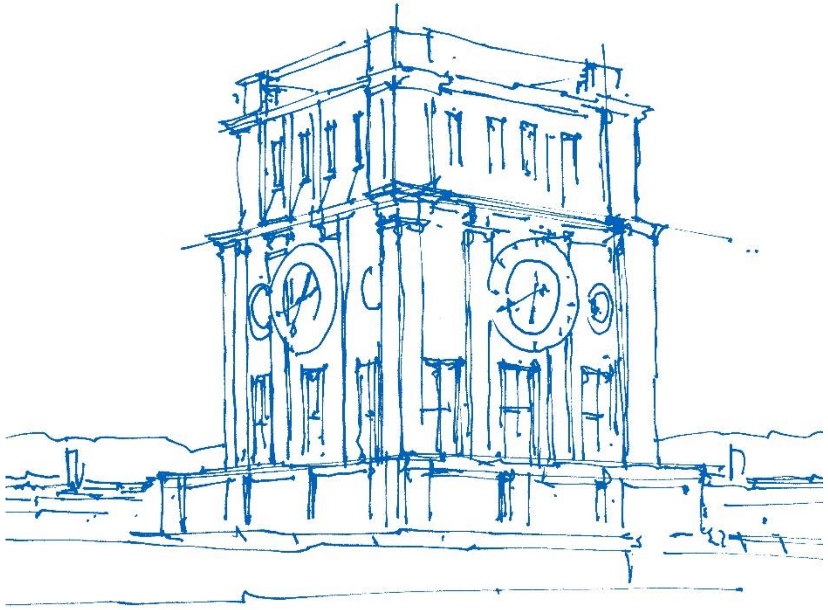
Globale Strukturen und Gruppen

Jürgen Pfeffer

Technical University of Munich

Bavarian School of Public Policy

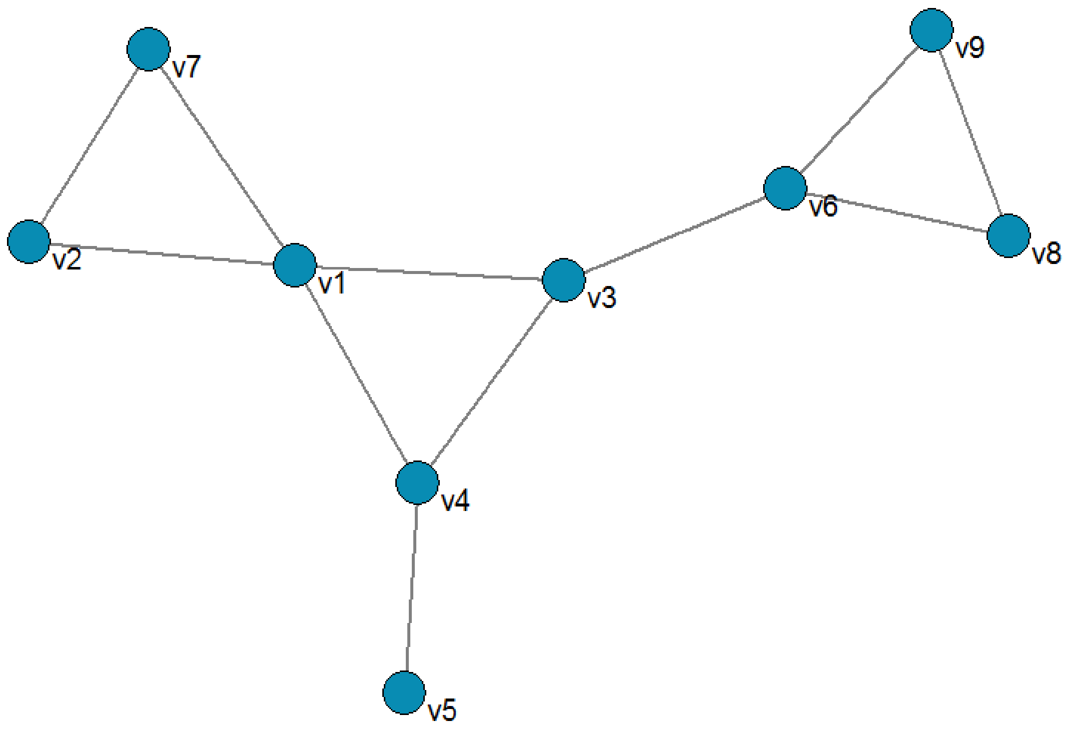
juergen.pfeffer@tum.de | @JurgenPfeffer



Uhrenturm der TUM

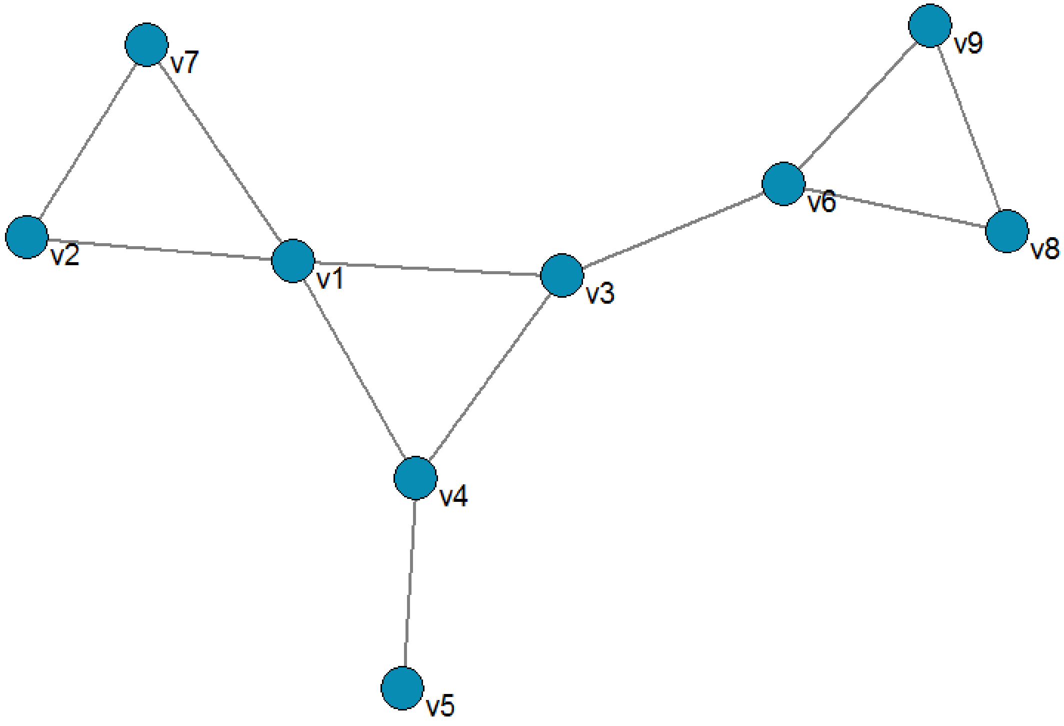
Cutpoints

Nodes which, if deleted, would disconnect net



Bridge

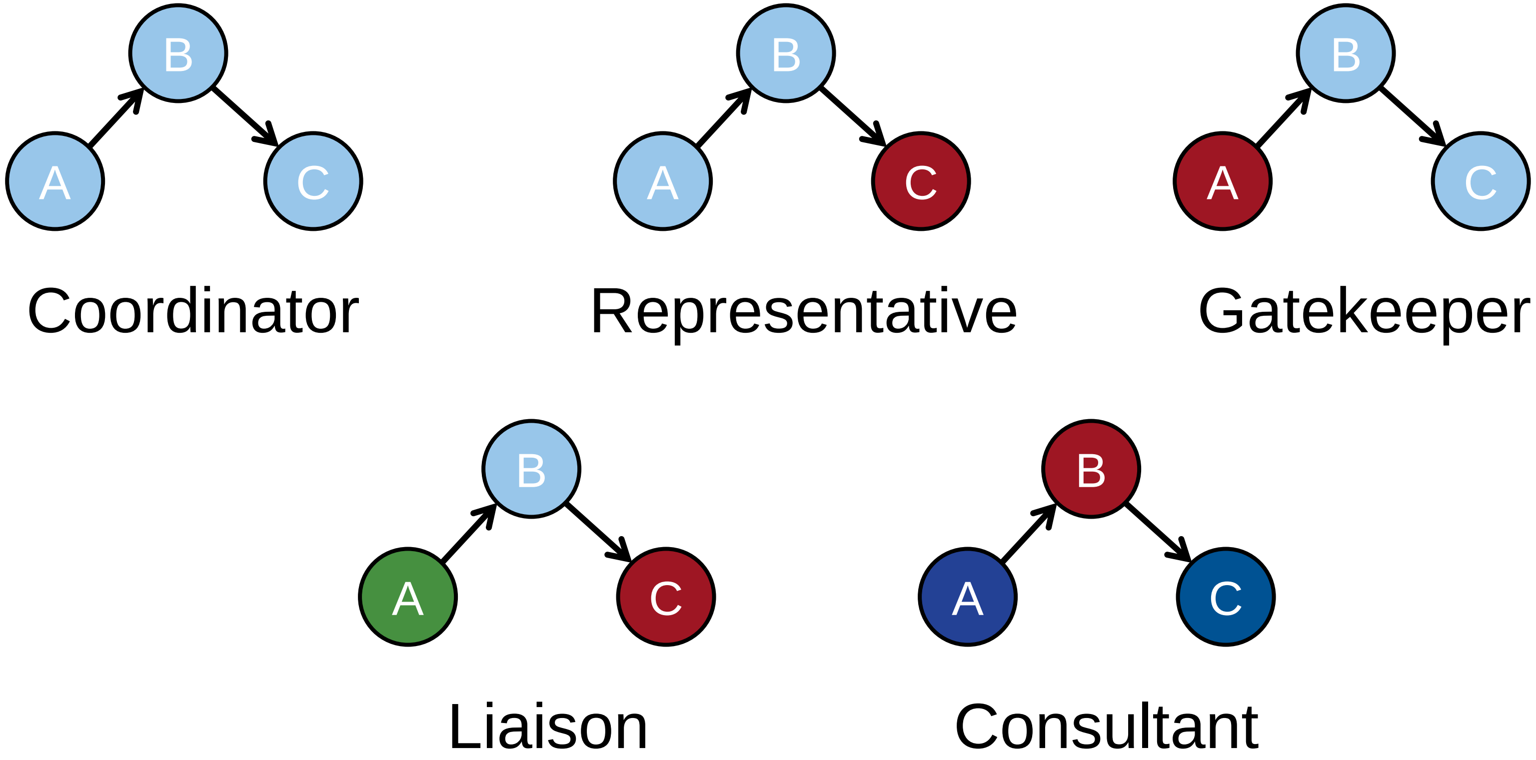
A tie that, if removed, would disconnect the network



Brokerage

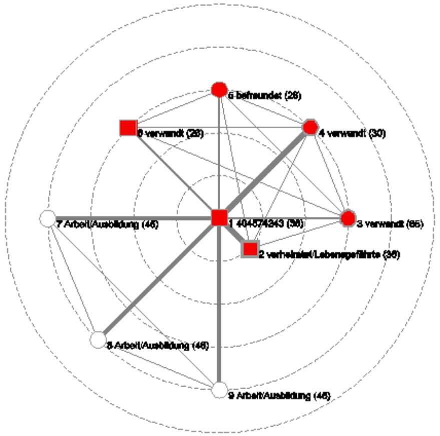
Burt, Ronald S. 1990. Detecting Role Equivalence, Social Networks, Nr.12.

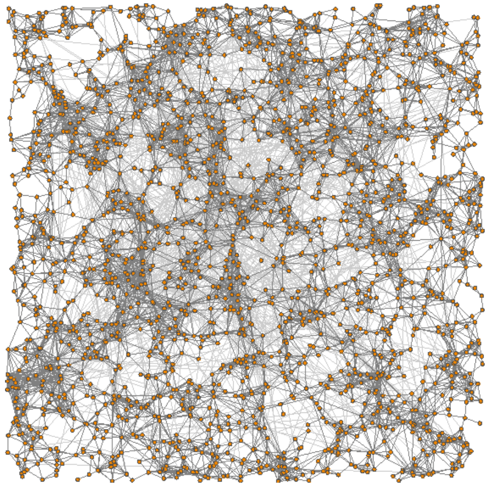
Gould, Roger V, & Fernandez R. M. 1989. Structures of Mediation, in: Sociological Methodology, San Francisco: Jossey-Bass.



Micro/Macro Connection

Local change can change network structure





Network change can change node level metrics

Micro-Macro Connection

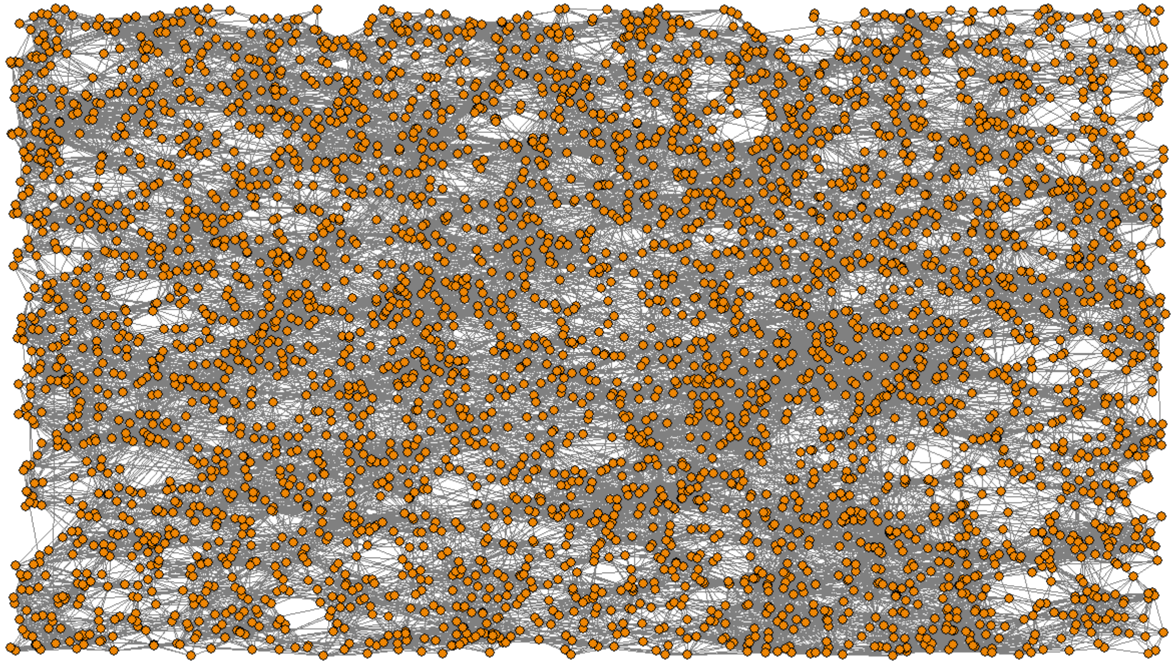
Micro (Ego Networks)	**Macro (Entire Network)**
Maximum number of alters	Maximum degree of network
Average degree of egos $\bar{e}$	1. Average degree of all nodes 2. Number of edges $E = \frac{\bar{e}N}{2}$ 3. Network density = $\frac{\bar{e}}{N-1}$
Degree distribution of egos	Degree distribution of entire network
Degree skewness, degree variance of egos	Degree skewness, degree variance of the entire network
Clustering coefficient C_v	Clustering coefficient $C = \frac{1}{N}\sum C_v$

Random Links in Structured Networks

Stylized interpersonal communication network

4k nodes, 47.7k edges

Avg. degree 23.8, max. degree 65



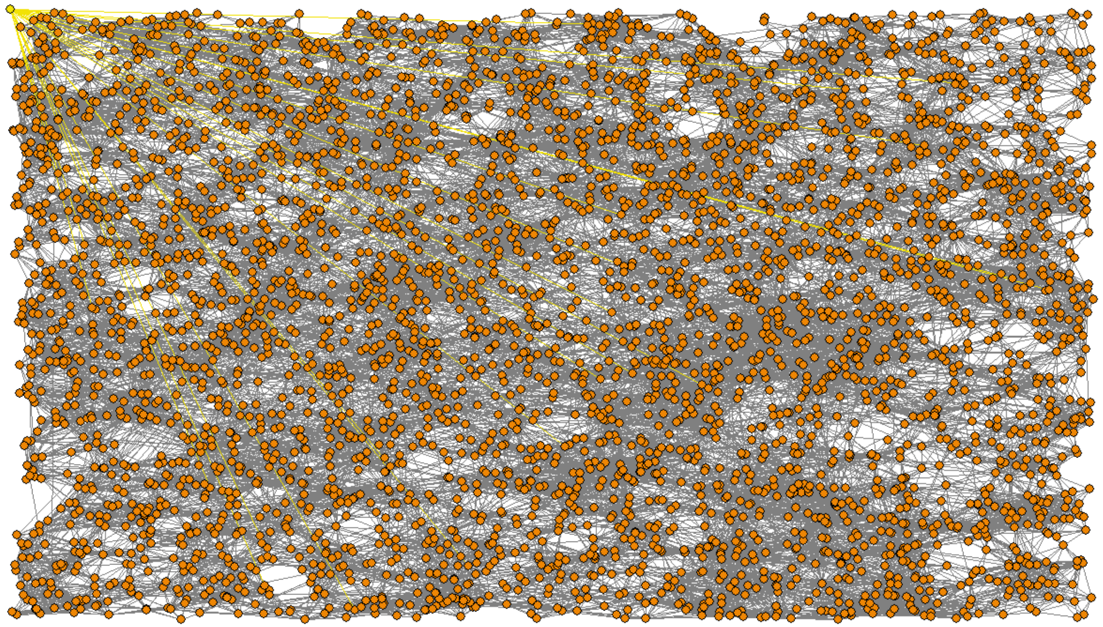
Random Links in Structured Networks

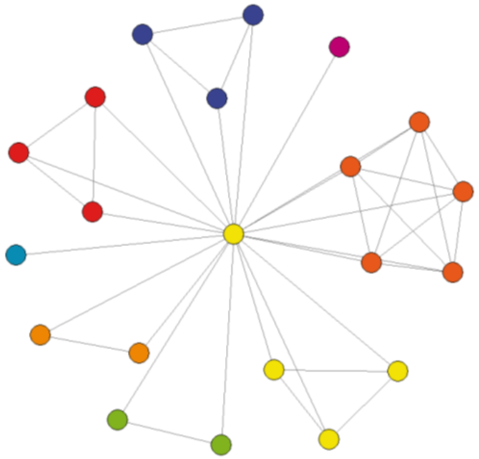
Stylized interpersonal communication network

4k nodes, 47.7k edges

Avg. degree 23.8, max. degree 65

Intervention: 1 new node + 23 new edges (0.05%)





We know a lot about this network

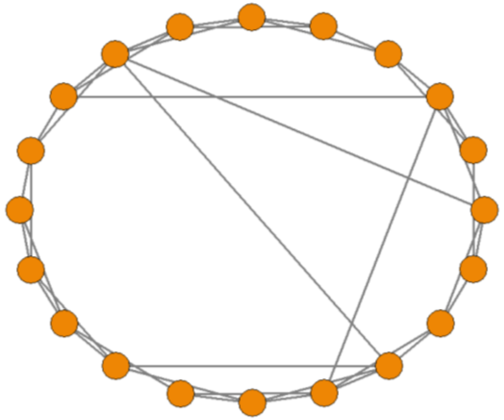
Stylized interpersonal communication network

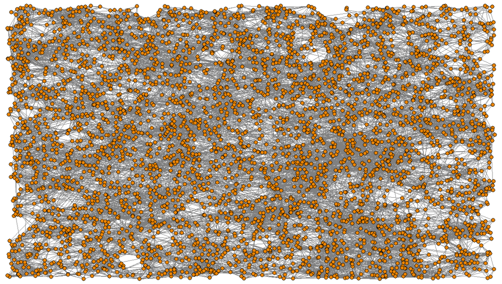
Limited degree, Low density, Differences in degree

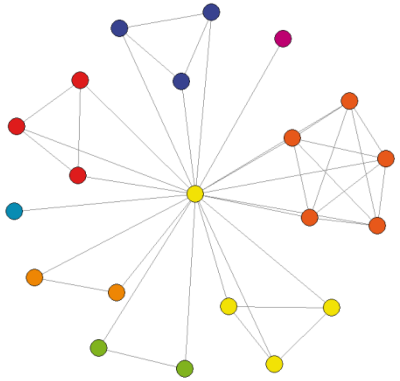
Right tailed degree distribution, Degree correlation

High clustering, Community structure

Short average path length

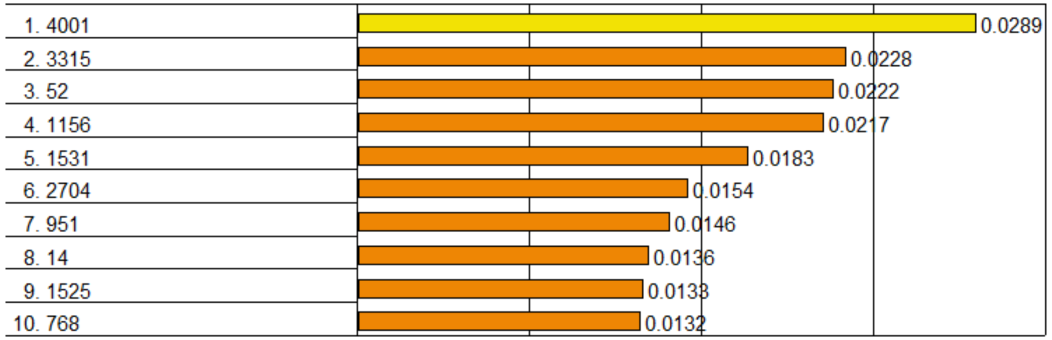






Random Links in Structured Networks

Betweenness Centrality in altered network:



Change in Betweenness Centrality of existing nodes:

- 17/20 relative winners are connected with the new node

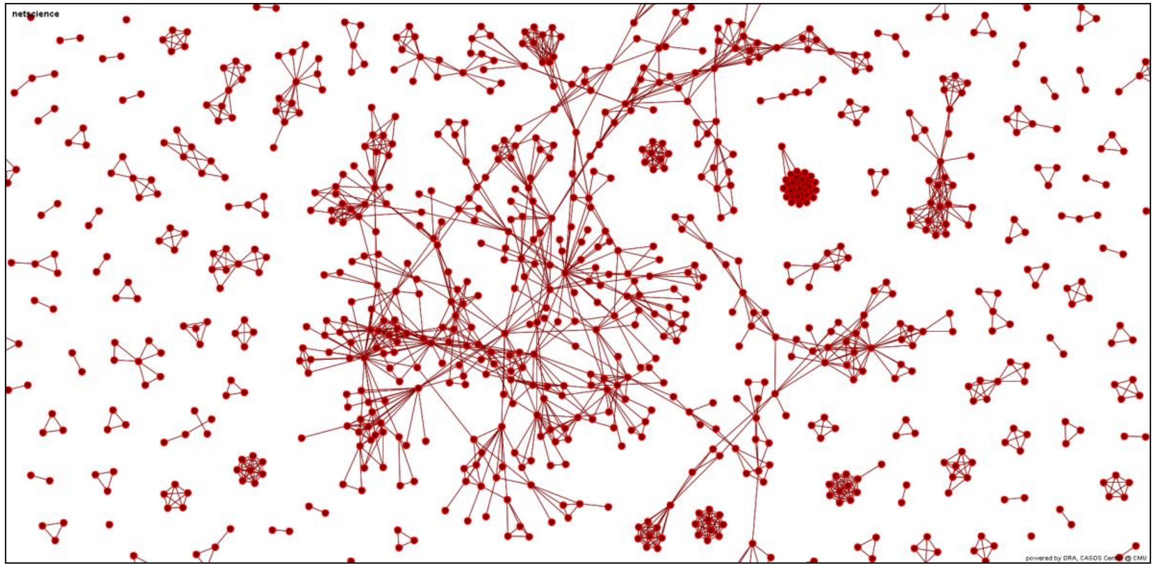
Groups & Communities

Different definitions of groups

How to detect communities?

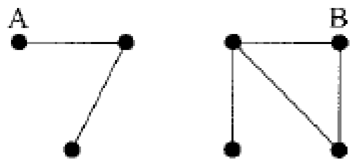
Different algorithms for different community definitions

To summarize / predict the high level structure of the graph



Components

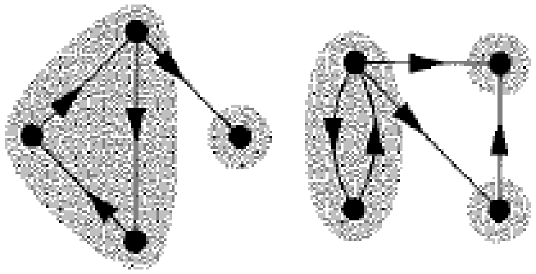
Disconnected network



$$\mathbf{A} = \begin{pmatrix} \boxed{} & 0 & \cdots \\ 0 & \boxed{} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

What about directed networks (e.g. www)?

- 2 undirected (weak) components
- 5 directed (strong) components
  - Contain cycles for (every) node

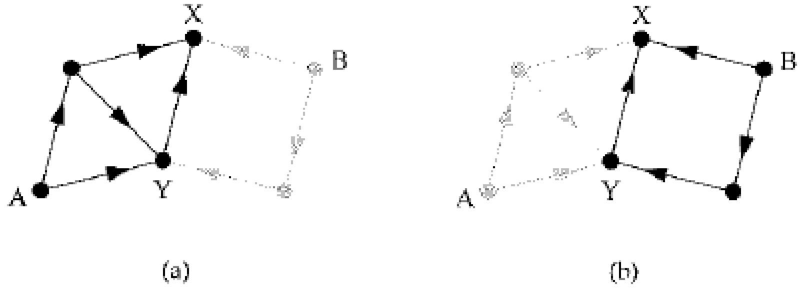


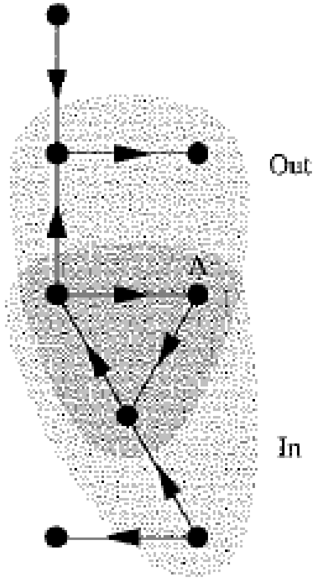
→ Acyclic directed networks have no strong components

Out-Component

How far can you get from one point (no walk back)?

Reachability of nodes

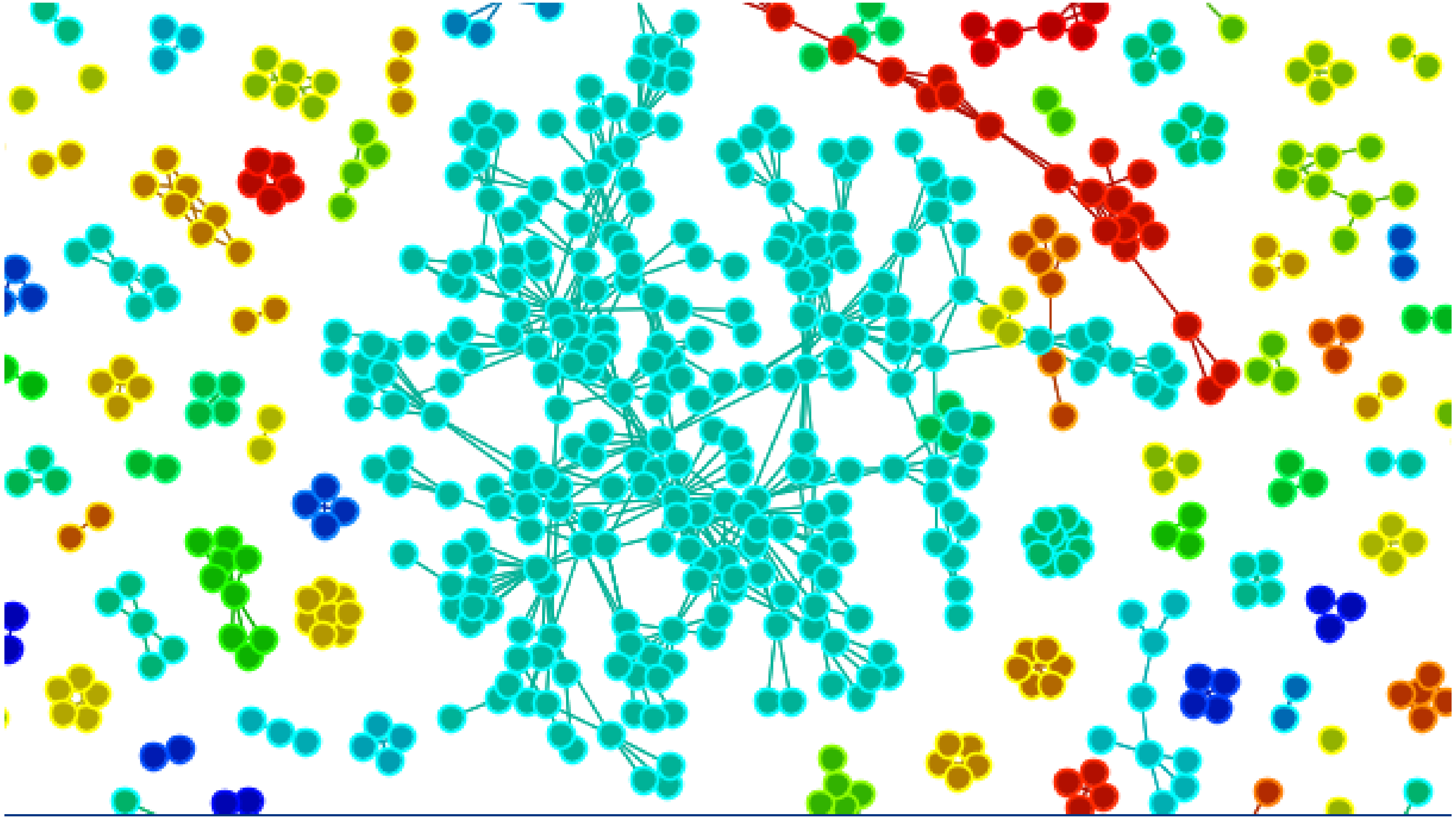




Strongly connected components have identical out-components

The intersection of in- and out-components are strong components

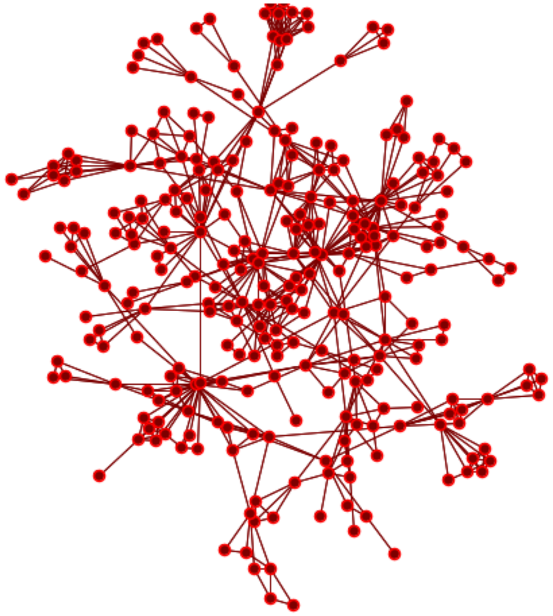
Visualize Components



Largest Component

In real-world networks, most of the time, 90+% of nodes are in largest component

In most studies, we focus on analysis of larges component

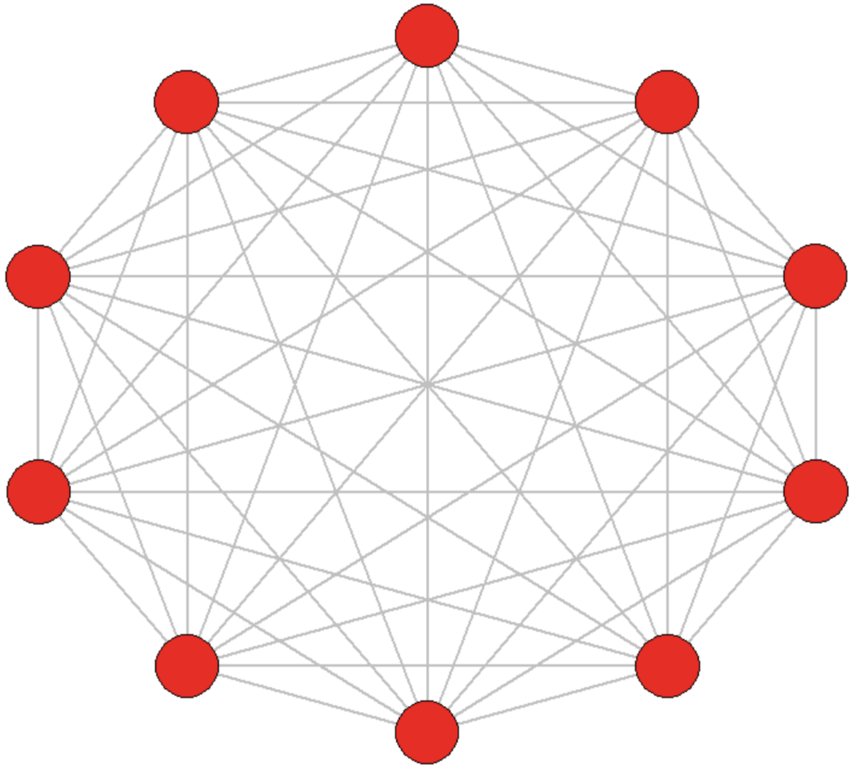


Cliques

Clique = Fully connected group of nodes

CON = Rarely found in real data

CON = Hard to calculate



Clique Relaxations

N-Clique = Connected with path distance N, normally 2

- CON: Nodes could be connected by nodes outside the clique

C-Clan = N-Clique + all links within group

K-Cores = Connected to at least k other members of a “clique”

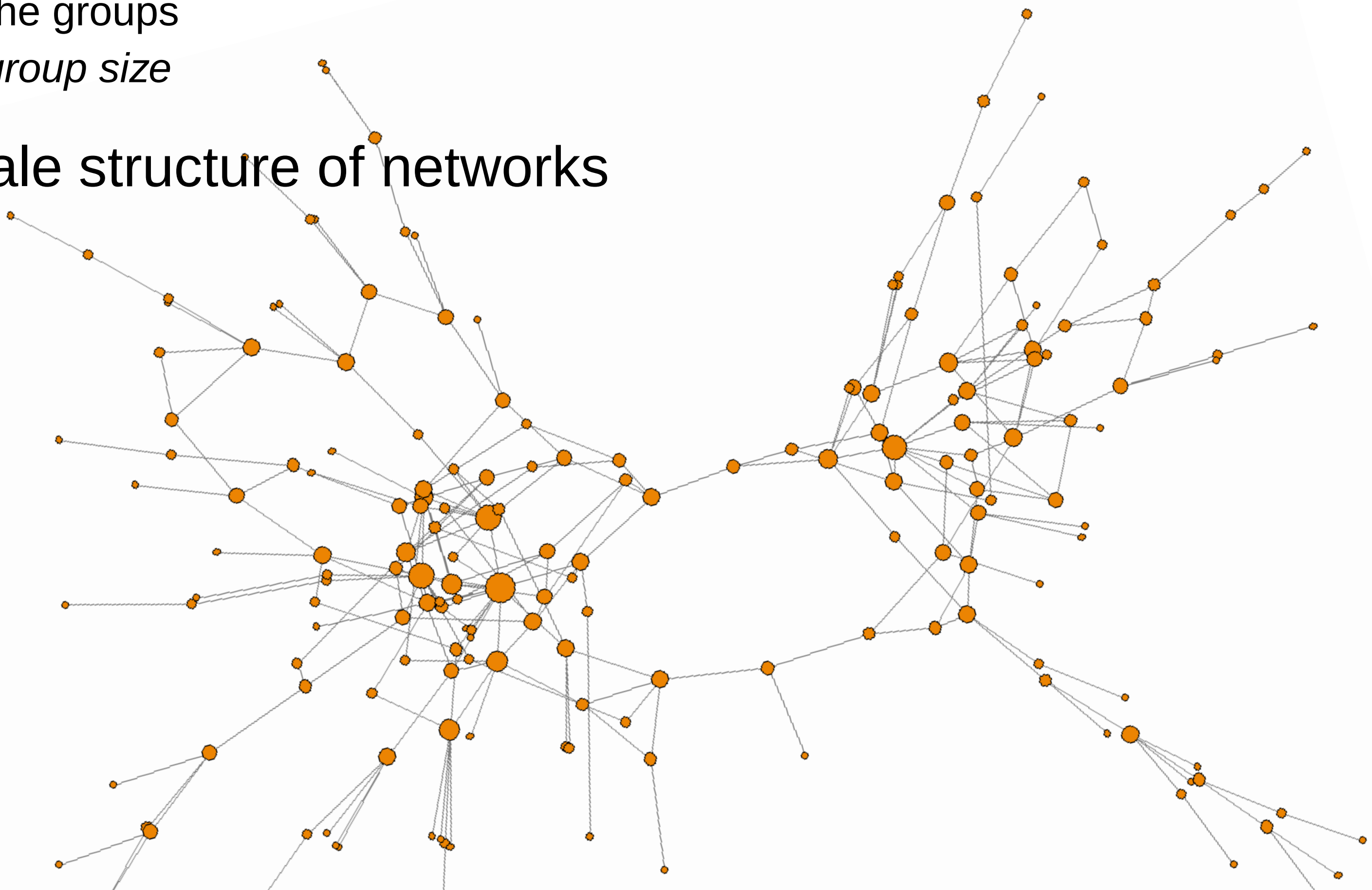
K-Plex = Node is a member of a “clique” of size n if it has direct ties to n-k members of that clique

Community Detection

Goals:

- Separate graph into groups of nodes
- Minimum connections among the groups
- *No fixed number of groups or group size*

Understanding large-scale structure of networks



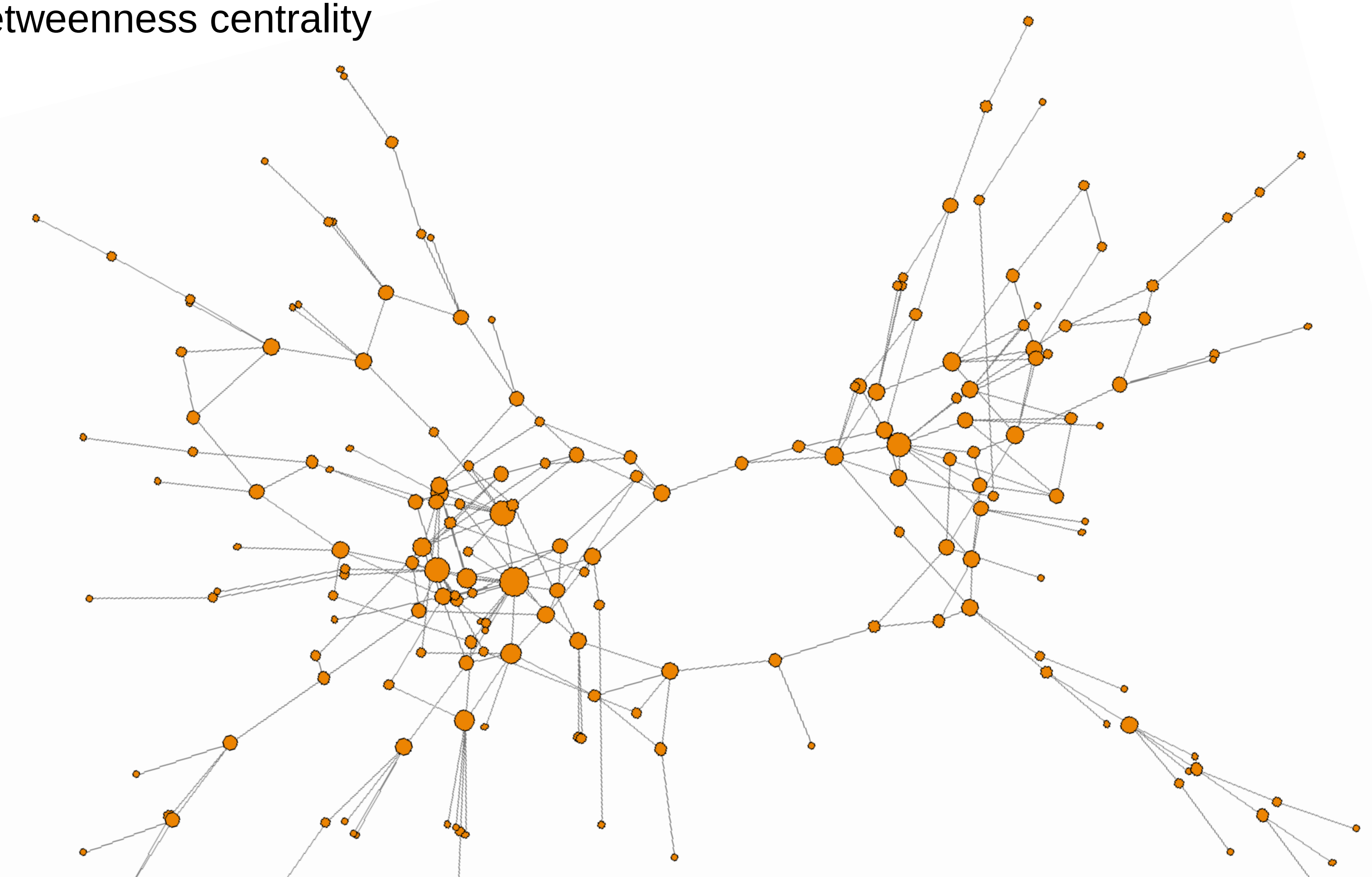
Newman/Girvan Grouping

Newman & Girvan [2004]

- Calculate edge betweenness centrality
- Remove edge with highest betweenness centrality
- Repeat process

When to stop?

- K-groups or modularity



How to Evaluate Grouping?

Fewer links between groups “than expected”

Count links within and between groups

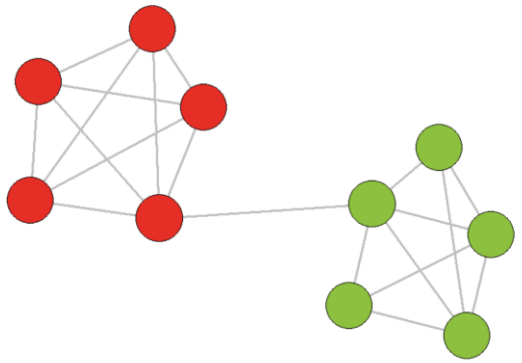
All = within + between

→ Goal: Optimize links within groups compared to what is expected

Modularity maximization: most commonly used

Perfect solution = exponential time complexity

Efficient heuristic optimization algorithms



Modularity

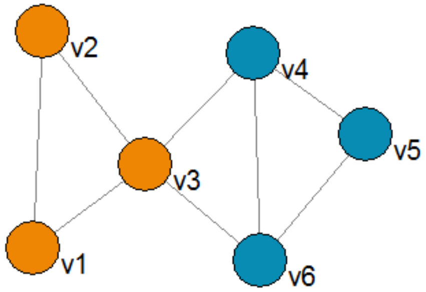
Fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random

Degrees k_i and k_j

S_i = 1 for group 1, s_j = -1 for group 2

2m = number of ends of edges

$$\frac{1}{4m}\sum_{ij}\left(A_{ij}-\frac{k_ik_j}{2m}\right)s_is_j,$$



Simple Modularity Maximization

Two random communities of equal size

Algorithm:

- For every node:
  - How much would modularity change if node would move
  - Move best node
- Repeat until no improvement

No constraint on group size

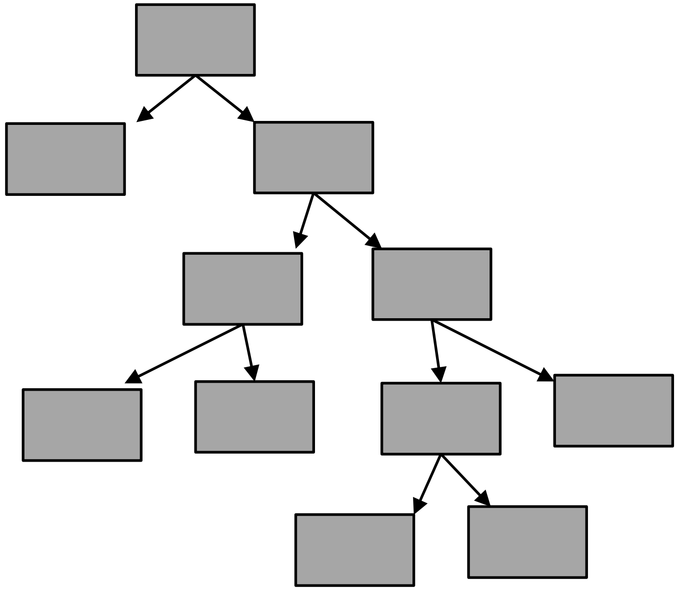
Quite fast $O(nm)$

More Than Two Groups

Modularity maximization works

Repeatedly bisecting the network

Stop when modularity does not increase anymore

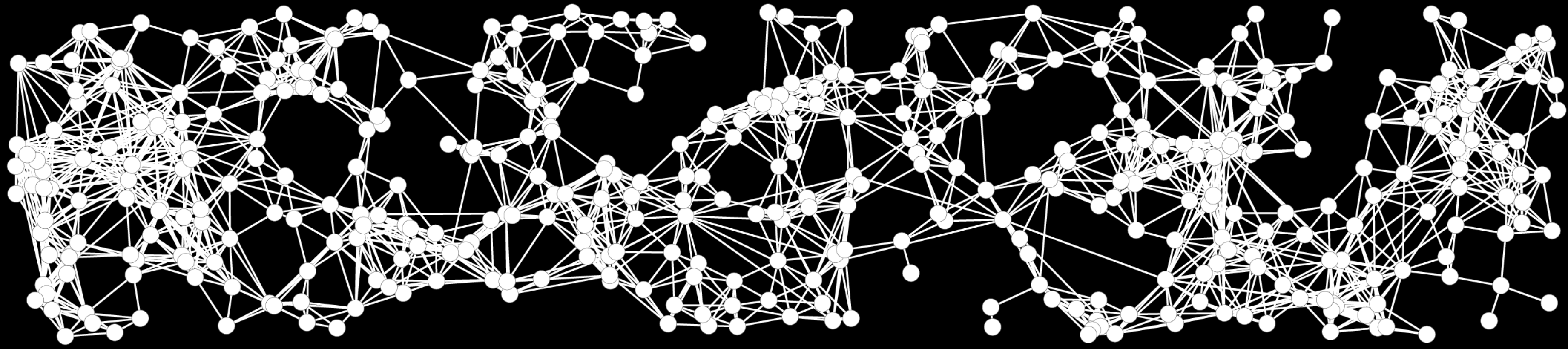


Important to Know!

Community/Partitioning algorithm find a best solution regardless whether a good solution exists!

Modularity value serves as a kind of a significance level for clustering

“Our mission is to go forward, and it has only just begun.
There's still much to do, still so much to learn. Engage!”
Jean-Luc Picard, Star Trek TNG, Season 1 Episode 26



Juergen.Pfeffer@tum.de @JurgenPfeffer
Mirco.Schoenfeld@tum.de