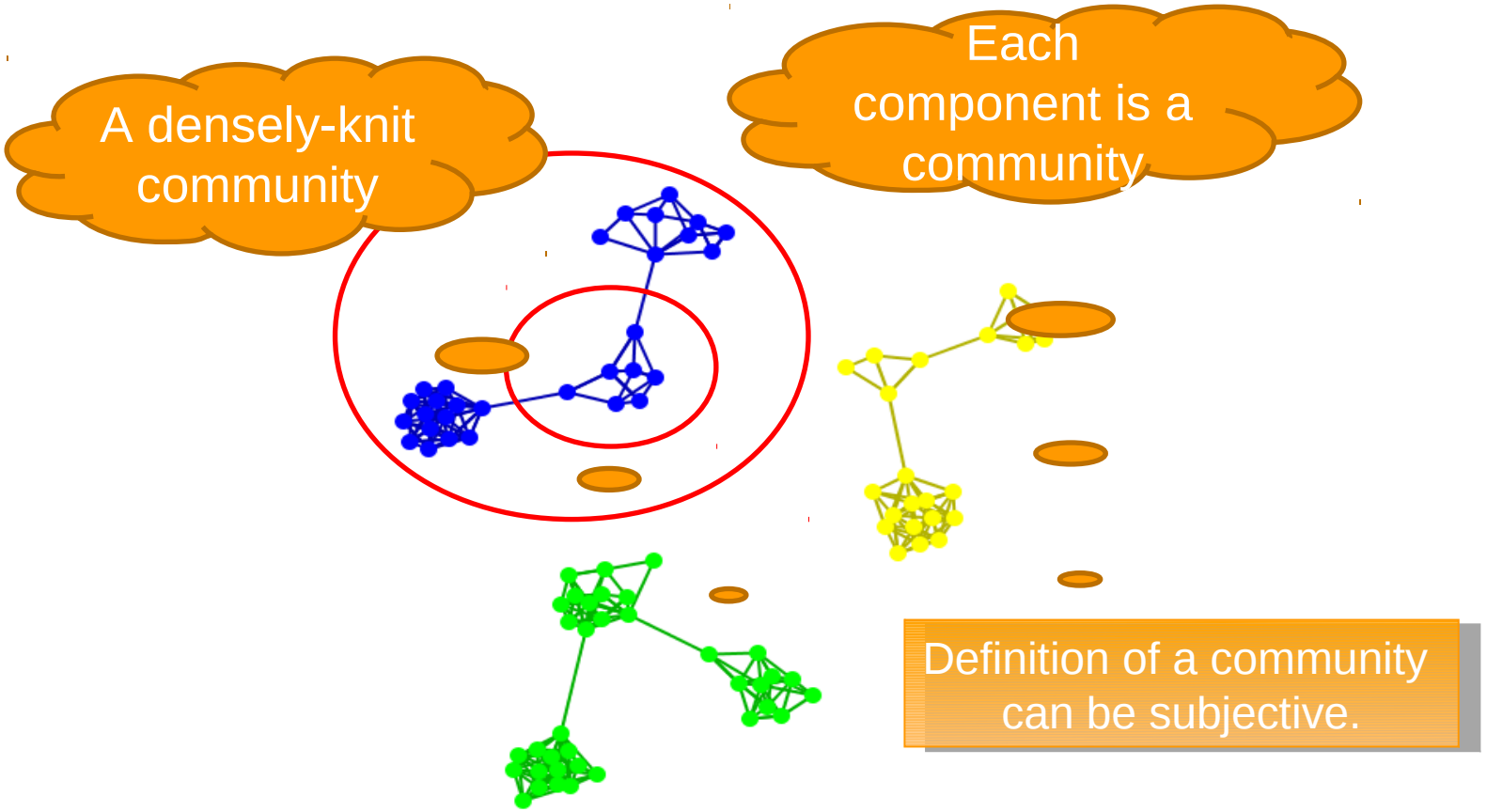
Community Identification/Clustering

- Groups of nodes that are densely connected amongst themselves while being sparsely connected to the rest of the network



Examples

Communities ← Facebook

Name: **Social Computing**
Type: Organizations
Members: 14 members

Name: **Social Computing**
Type: Internet & Technology
Members: 12 members

Name: **Social Computing Magazine**
Type: Internet & Technology
Members: 34 members

Name: **Trustworthy Social Computing**
Type: Internet & Technology
Members: 28 members

Name: **Social Computing for Business**
Type: Internet & Technology
Members: 421 members

Name: **UCLA Social Sciences Computing**
Type: Internet & Technology
Members: 22 members

Name: **Social Media and Computing**
Type: Organizations
Members: 6 members

Name: **Transition to Social Computing Era**
Type: Common Interest
Members: 1 member

Name: **Lab for Social Computing at the RIT Libraries**
Type: Organizations

Communities ← Flickr

I * Urban LIFE in Metropolis ////
4,286 members | 31 discussions | 89,645 items | Created 46 months ago | Join?
UrbanLIFE, People, Parties, Dance, Musik, Life, Love, Culture, Food and Everything what we could imagine by hearing that word URBANLIFE! Have some FUN! Please add... (more)

Islam Is The Way Of Life (Muslim World)
619 members | 13 discussions | 2,685 items | Created 23 months ago | Join?
The word islām is derived from the Arabic verb aslama, which means to accept, surrender or submit. Thus, Islam means submission to and acceptance of God, and believers must... (more)

*** THE CELEBRATION OF ~LIFE~ (Post1~Award1) [only living things]**
4,871 members | 22 discussions | 40,519 items | Created 21 months ago | Join?
WELCOME to THE CELEBRATION OF ~LIFE~ (Post1~Award1) PLEASE INVITE & COMMENT USING only THE CODES FOUND BELOW! ☆☆ This group is for sharing BEAUTIFUL, TOP QUALITY images... (more)

"Enjoy Life!"
2,027 members | 10 discussions | 39,916 items | Created 23 months ago | Join?
There are lovely moments and adorable scenes in our lives. Some are in front of you, and some are just waiting to be discovered. A gaze from someone we love, might touch the... (more)

Baby's life
2,047 members | 185 discussions | 30,302 items | Created 32 months ago | Join?
This group is designed to highlight milestones and important events in your baby's life (ie 1st time smiling/crawling/sitting in a high chair/reading/playing etc). It can also be... (more)

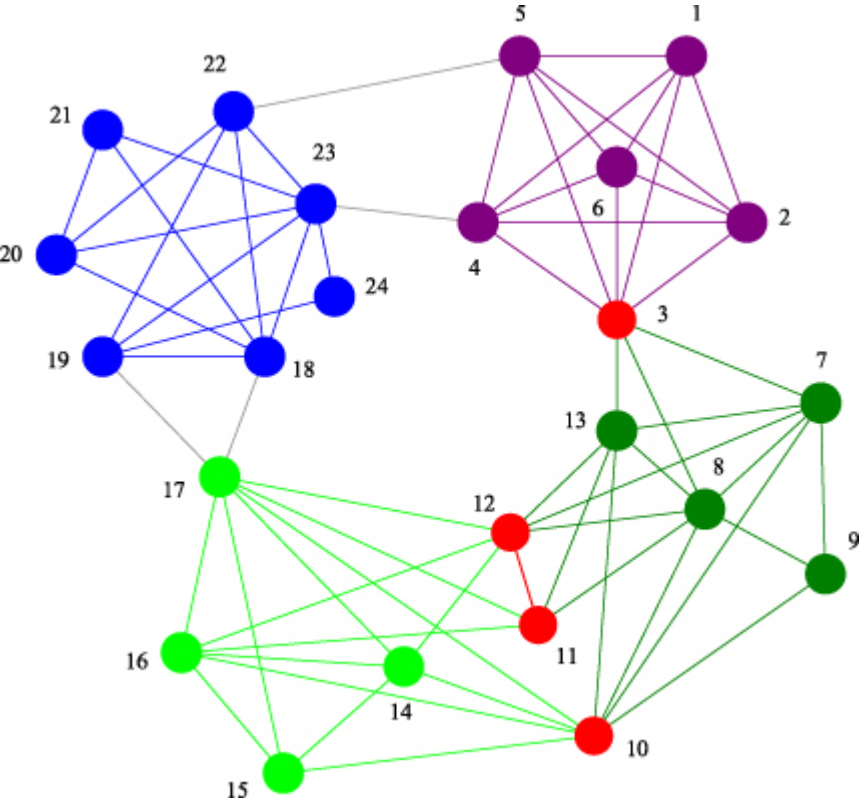
Only group members s
pool

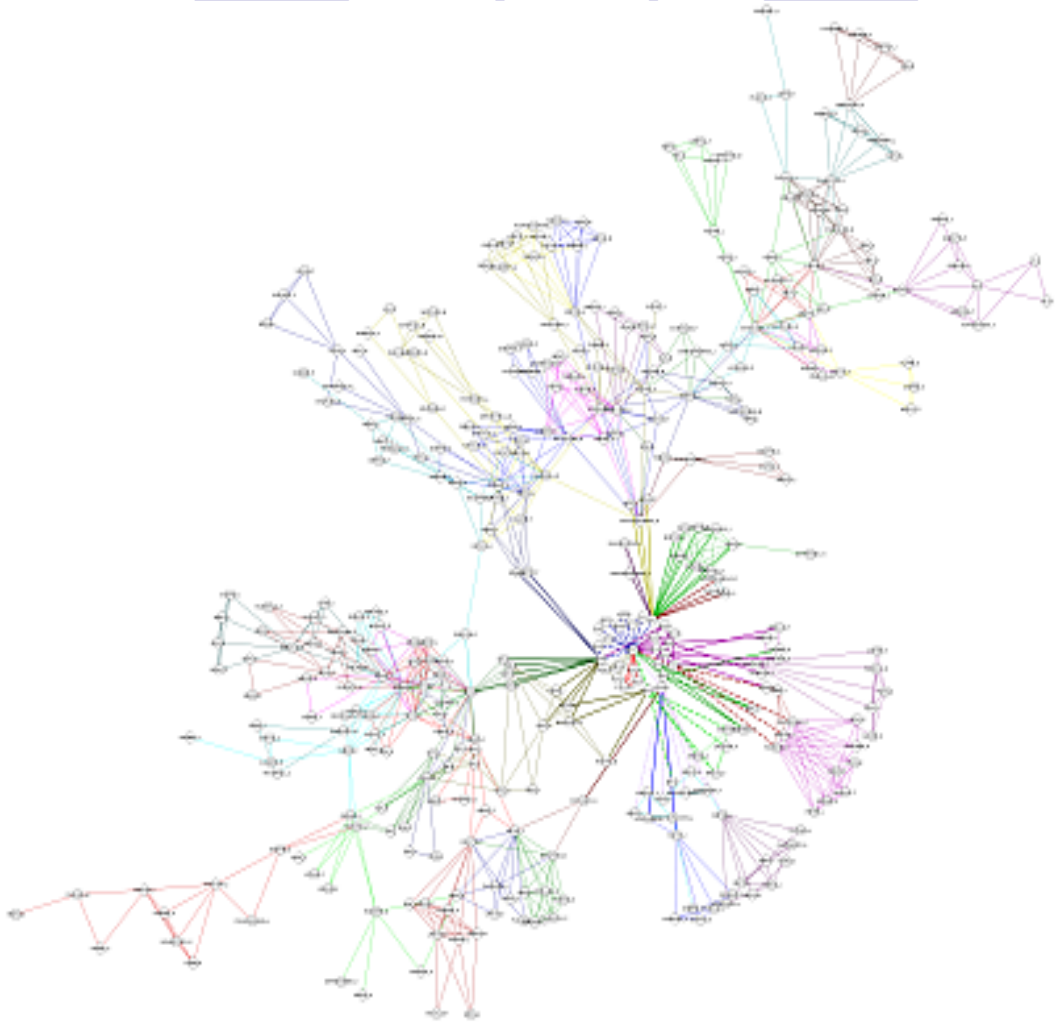
Pond Life
903 members | 20 discussions | 6,877 items | Created 32 months ago | Join?
Pic of the week: chosen from the pool by the group admins. Nuphar by guus timpers Pond Life is a group for all aquatic flora and fauna. Koi ponds, wildlife ponds, garden ponds,... (more)

Second Life
10,288 members | 773 discussions | 257,870 items | Created 61 months ago | Join?
Welcome to the Second Life pool, the biggest group on Flickr for residents/players of Second Life, the 3D virtual world from Linden Lab. This group is not endorsed or run in any... (more)

Life in Kuwait - Post 1 Award 1
637 members | 28 discussions | 3,233 items | Created 18 months ago | Join?
About kuwait: LOCATION: Kuwait lies on the northern tip of the Arabian Gulf. It is bordered by the Kingdom of Saudi Arabia to the south and south west, and the Republic of Iraq to... (more)

Might not be easy to see through ...





Applications

- Drug interactions
- Disease spreading
- CPU optimization
- Recommendation
- Layout optimization
- Lexical analysis
- Many others ....

Computational Metods

- Agglomerative
  - make an empty graph (N nodes, 0 edges)
  - add edges into empty graph maximizing something in original network
- Divisive
  - cut edges in prescribed order until communities separate
- Spectral
  - split graph based on eigenvalues/eigenvectors of Graph Laplacian

Similarity Measures

- Choosing (dis)similarity measures – a critical step in community finding/clustering
- Recall that the goal is to group together "similar" data – but what does this mean?

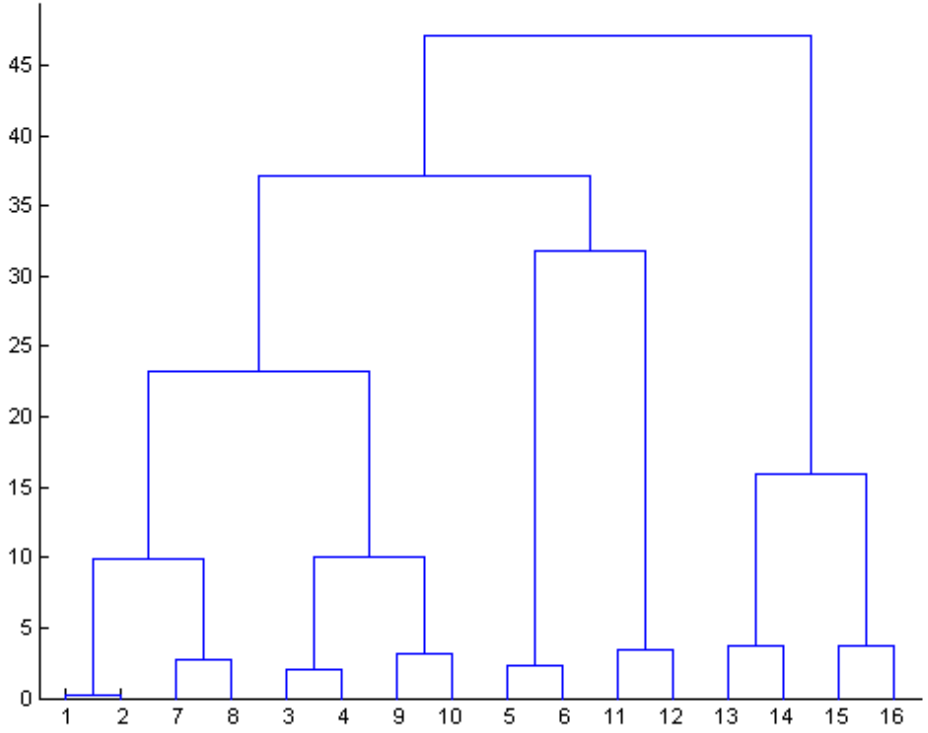
Similarity Measures

- Choosing (dis)similarity measures – a critical step in community finding/clustering
- Recall that the goal is to group together "similar" data – but what does this mean?
- No single answer – it depends on what we want to find or emphasize in the data; this is one reason why clustering is an "art"
- The similarity measure is often more important than the clustering algorithm used – don't overlook this choice!

Agglomerative Method

- Start with every data point in a separate cluster
- Keep merging the most similar pairs of data points/clusters until we have one big cluster left
- This is called a bottom-up or agglomerative method

Agglomerative Method



This produces a binary tree or ***dendrogram***

The final cluster is the root and each data item is a leaf

The height of the bars indicate how close the items are

Linkages

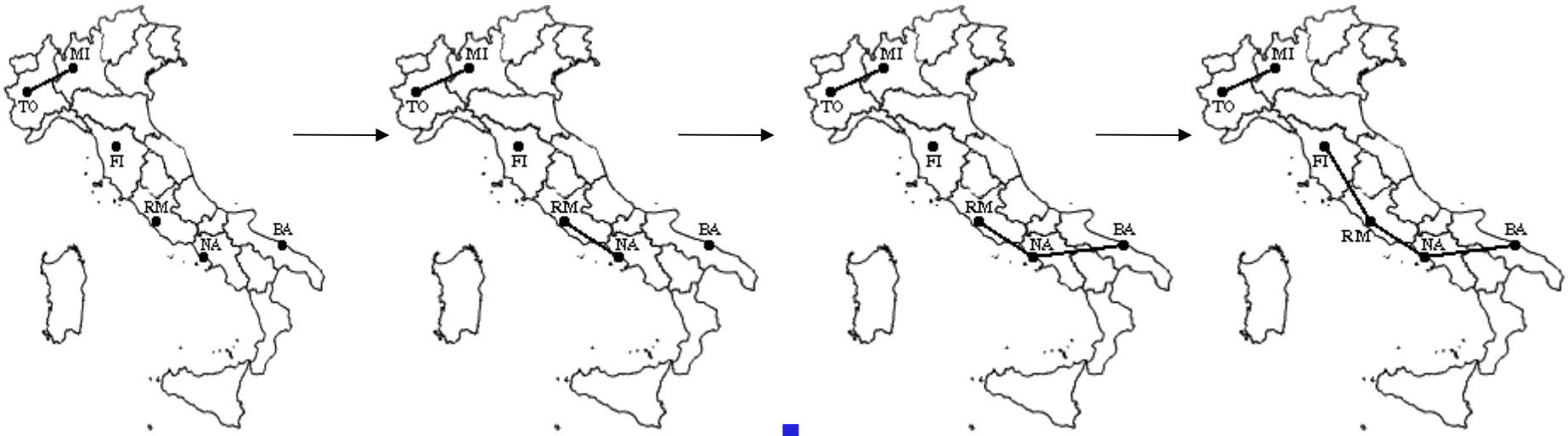
- We already know about distance measures between data items, but what about between a data item and a cluster or between two clusters?
- We just treat a data point as a cluster with a single item, so our only problem is to define a ***linkage*** method between clusters
- As usual, there are lots of choices…

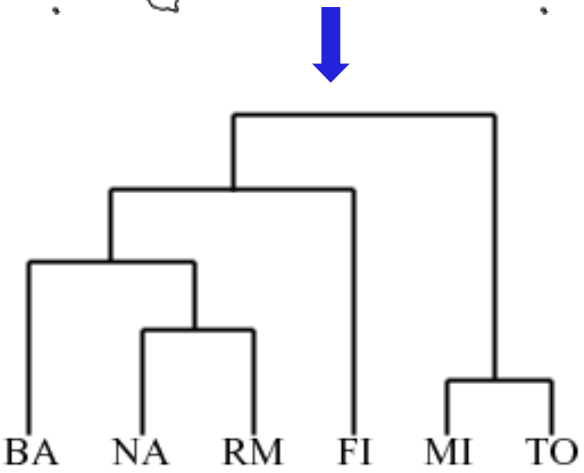
Types of Linkages

- Single Linkage: The minimum of all pairwise distances between points in the two clusters
- Complete Linkage: The maximum of all pairwise distances between points in the two clusters
- Average Linkage: Compute average of all pairwise distances between two clusters

Running Example

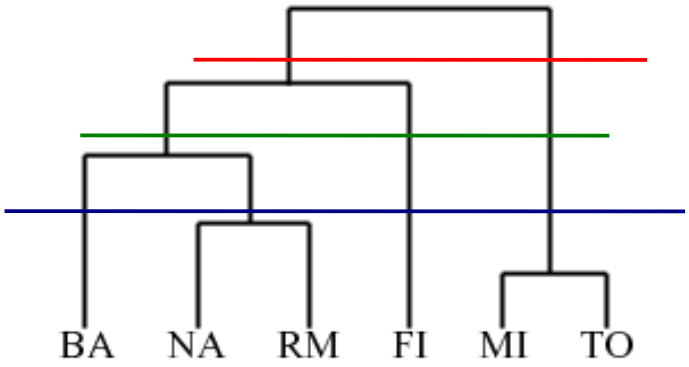
- Clustering of distances in kilometers between some Italian cities





Problems

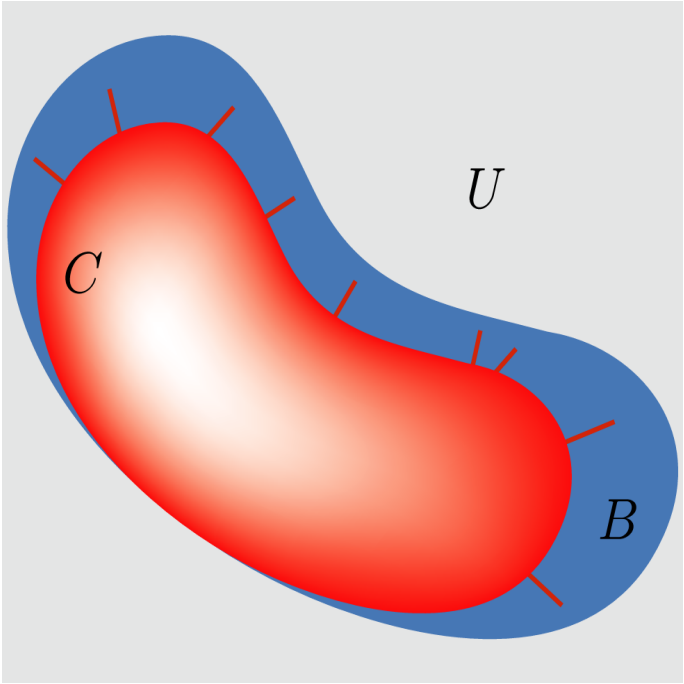
- They do not scale: $O(n^3)$
- Steps once done cannot be undone
- Arbitrary cut-offs need to set to arrive at a community structure



Local algorithm based on agglomeration

Bagrow, J. Stat. Mech., 2008, also: Bagrow, Bollt, Phys. Rev. E, 2005

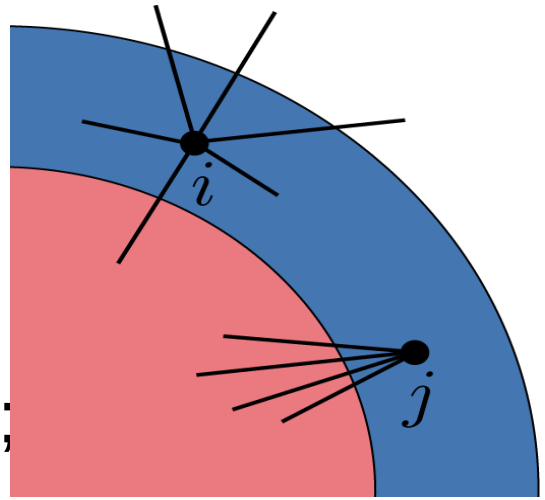
- Agglomerate nodes one at a time
- Maintain two groups, the community C and the border B
- nodes in B have been explored but are not yet in C
- Move nodes from B *to* C in specfied order, Outwardness (Ω)



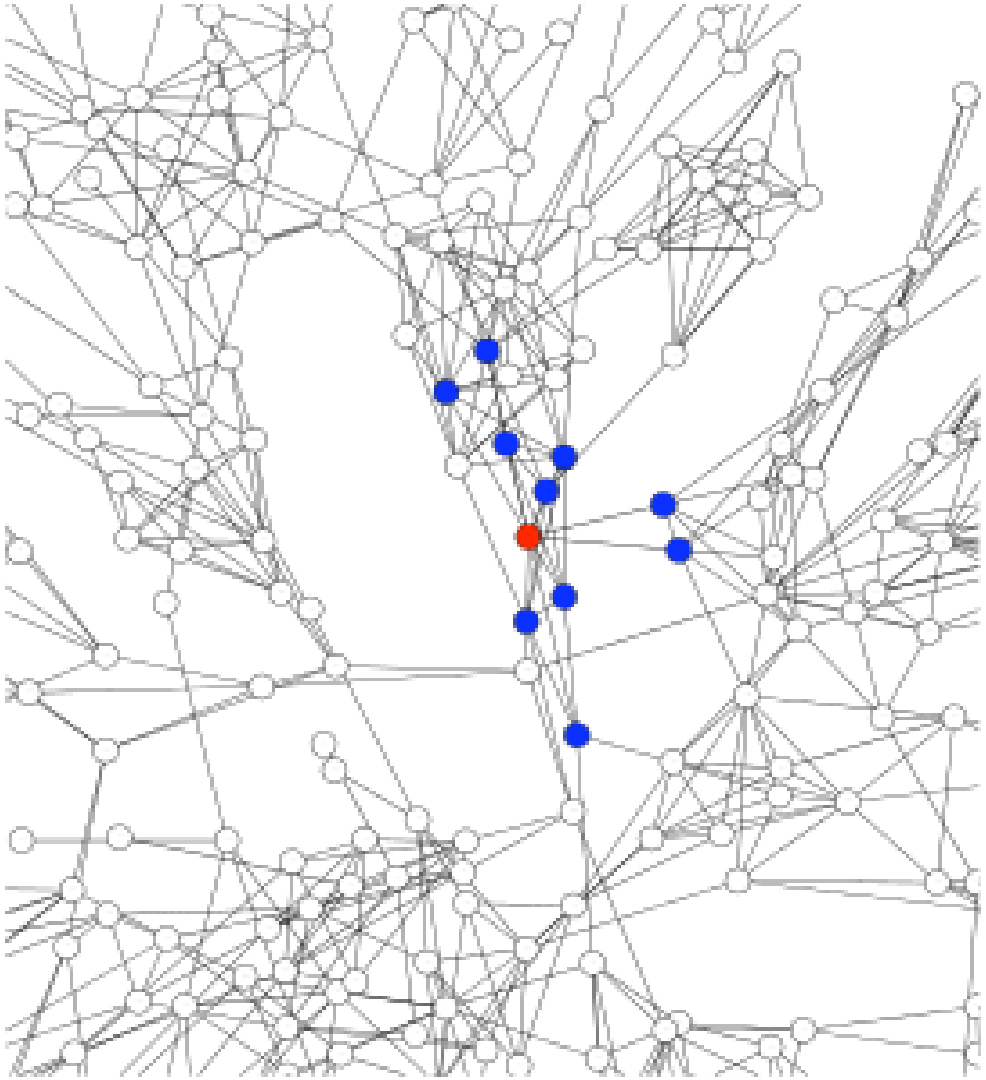
Local algorithm based on agglomeration

Bagrow, J. Stat. Mech., 2008, also: Bagrow, Bollt, Phys. Rev. E, 2005

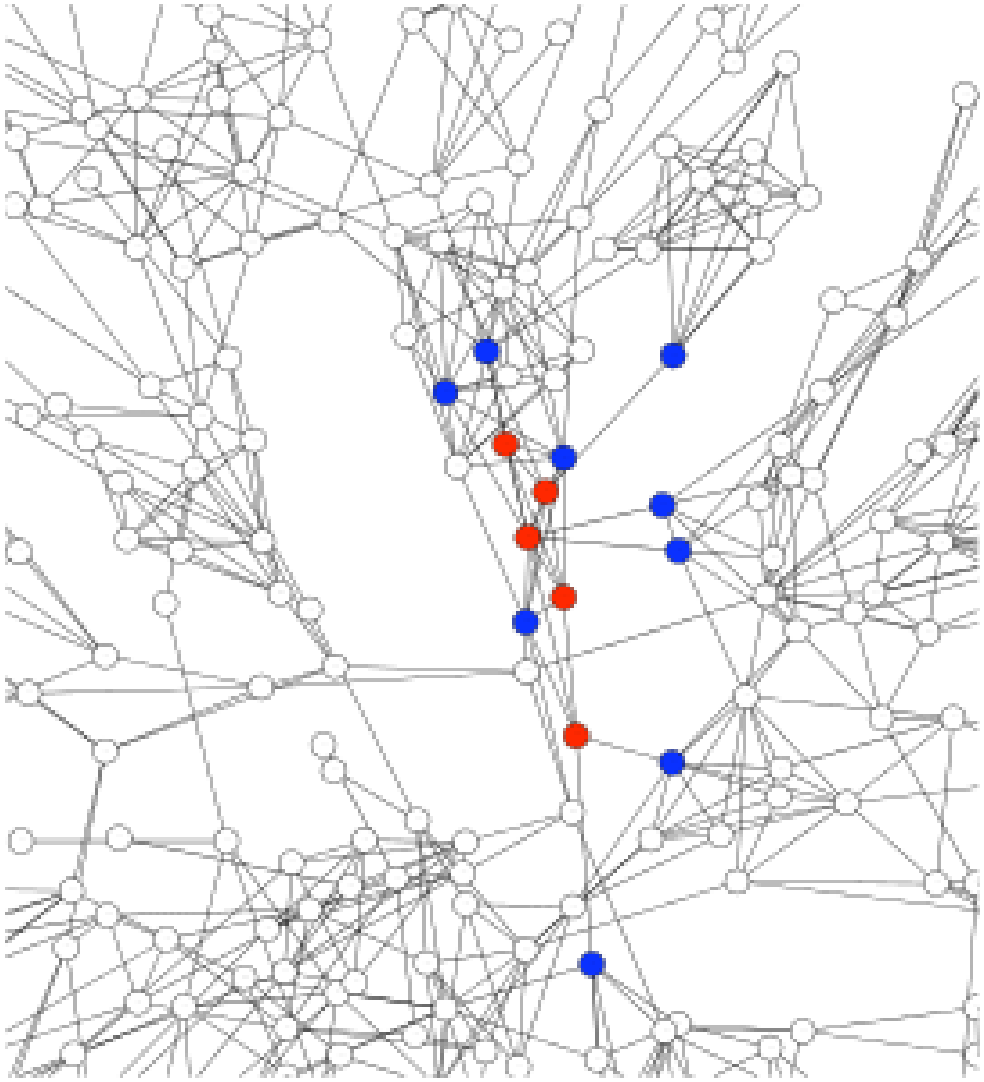
- Ω_v = (# of neighbors of v outside C - # of neighbors of v inside C)/k_v
- $\Omega_i = 2/3$, $\Omega_j = -1$
- Algorithm
  1. Choose starting node s: $C = \{s\}$; B = {neighbors of s};
  2. Add $v \in B$ to C, where $\Omega_v = \min\{\Omega\}$;
  3. Update B, Ω's, repeat from 2;



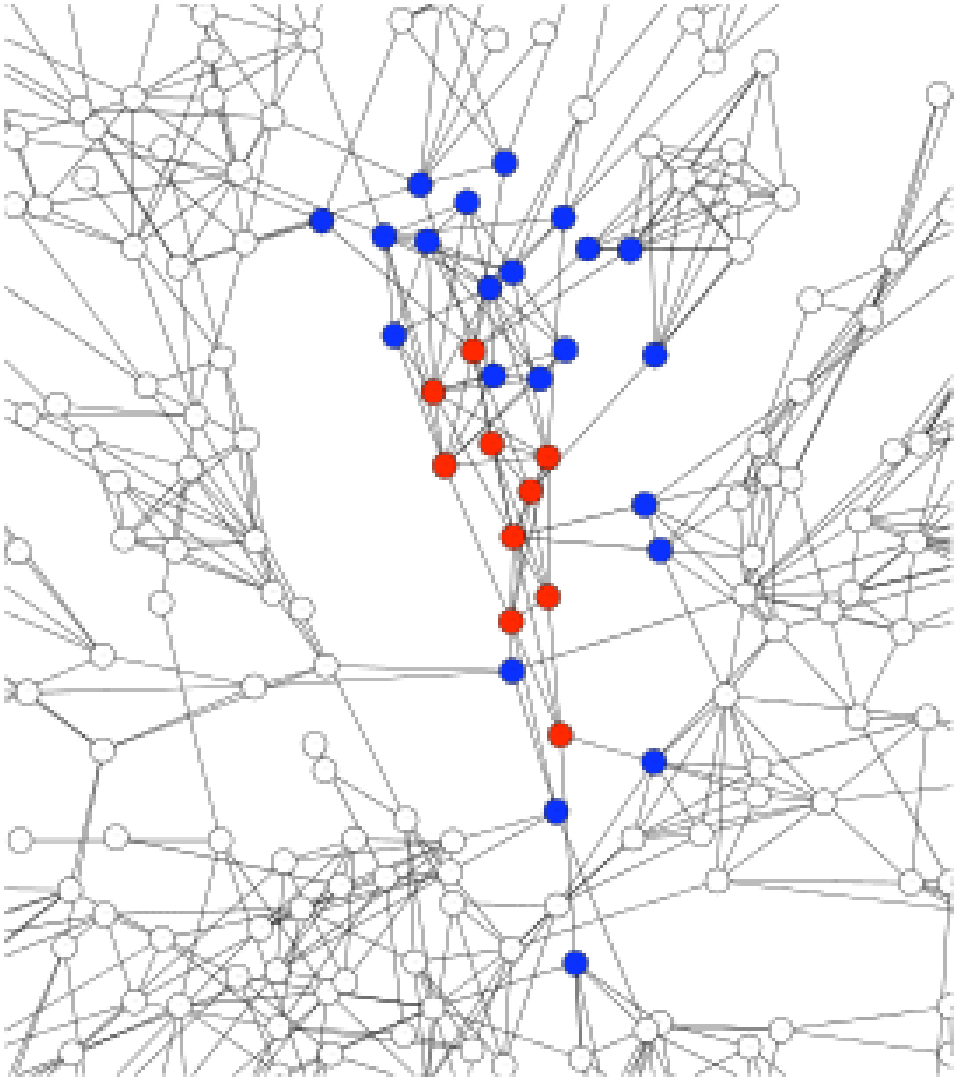
Local algo in execution



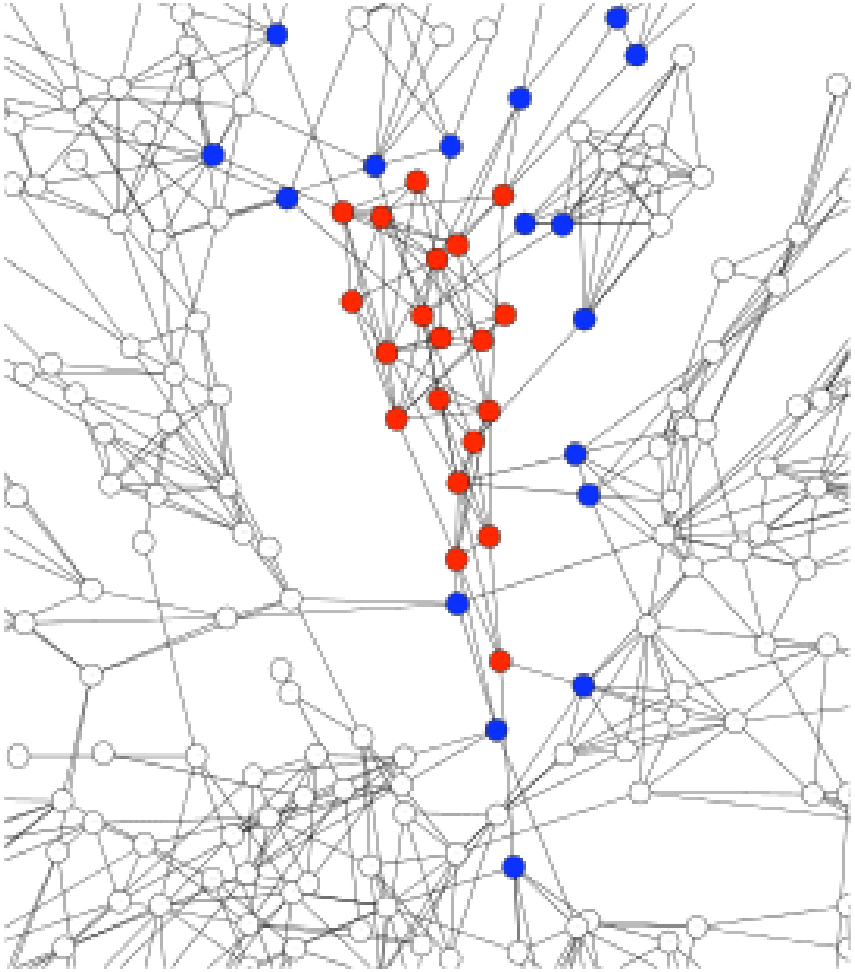
Local algo in execution



Local algo in execution

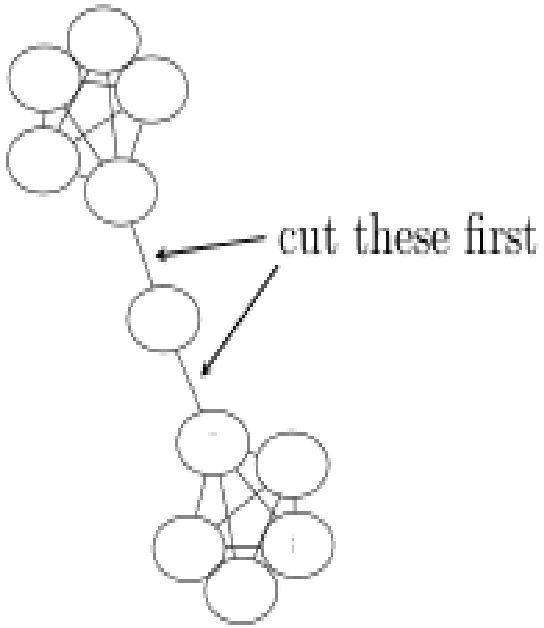


Local algo in execution

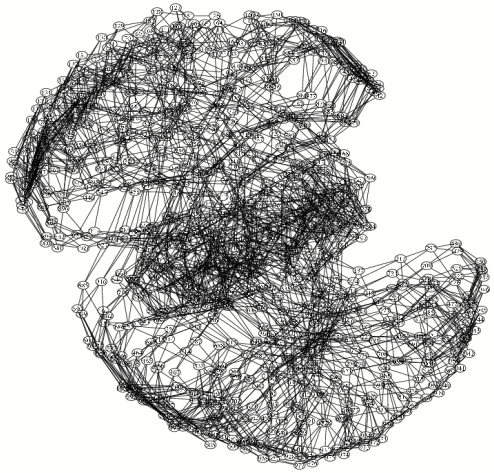


Divisive Method

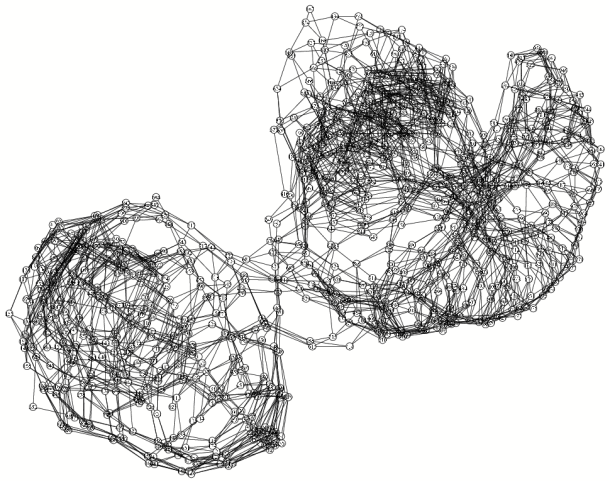
- Prototypical example: ***Edge betweenness*** based
- $betweenness(e_{ij})$ = number of times e_{ij} appears in all shortest paths
- High betweenness edges are more "central"



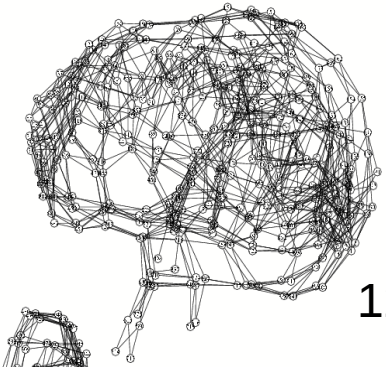
In Execution



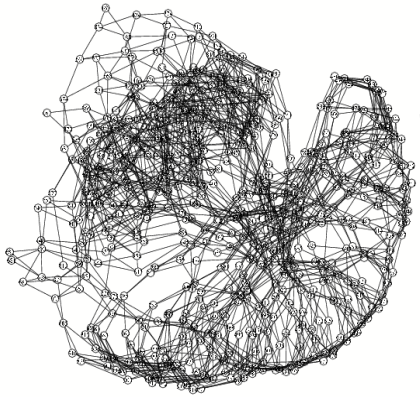
0 cuts

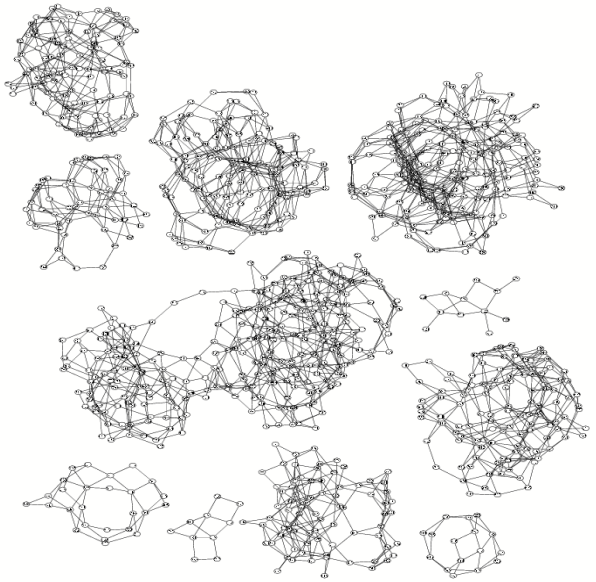


100 cuts



120 cuts





500 cuts

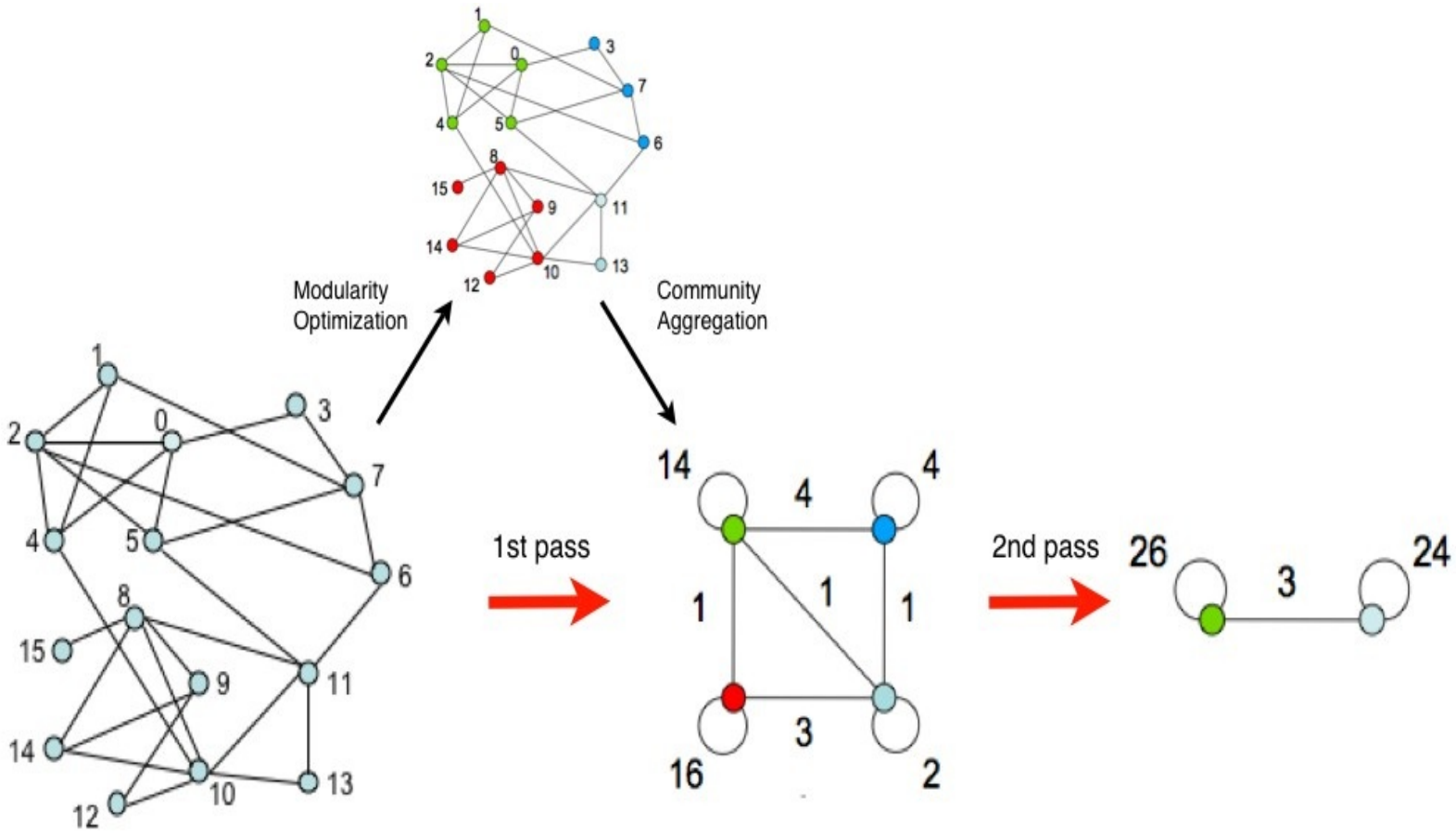
Calculating Betweenness

- Use BFS for shortest path
- $O(m)$ for each source node
- $O(mn)$ for all nodes
- For m cuts $O(m^2n)$
- Sparse graph $O(n^3)$

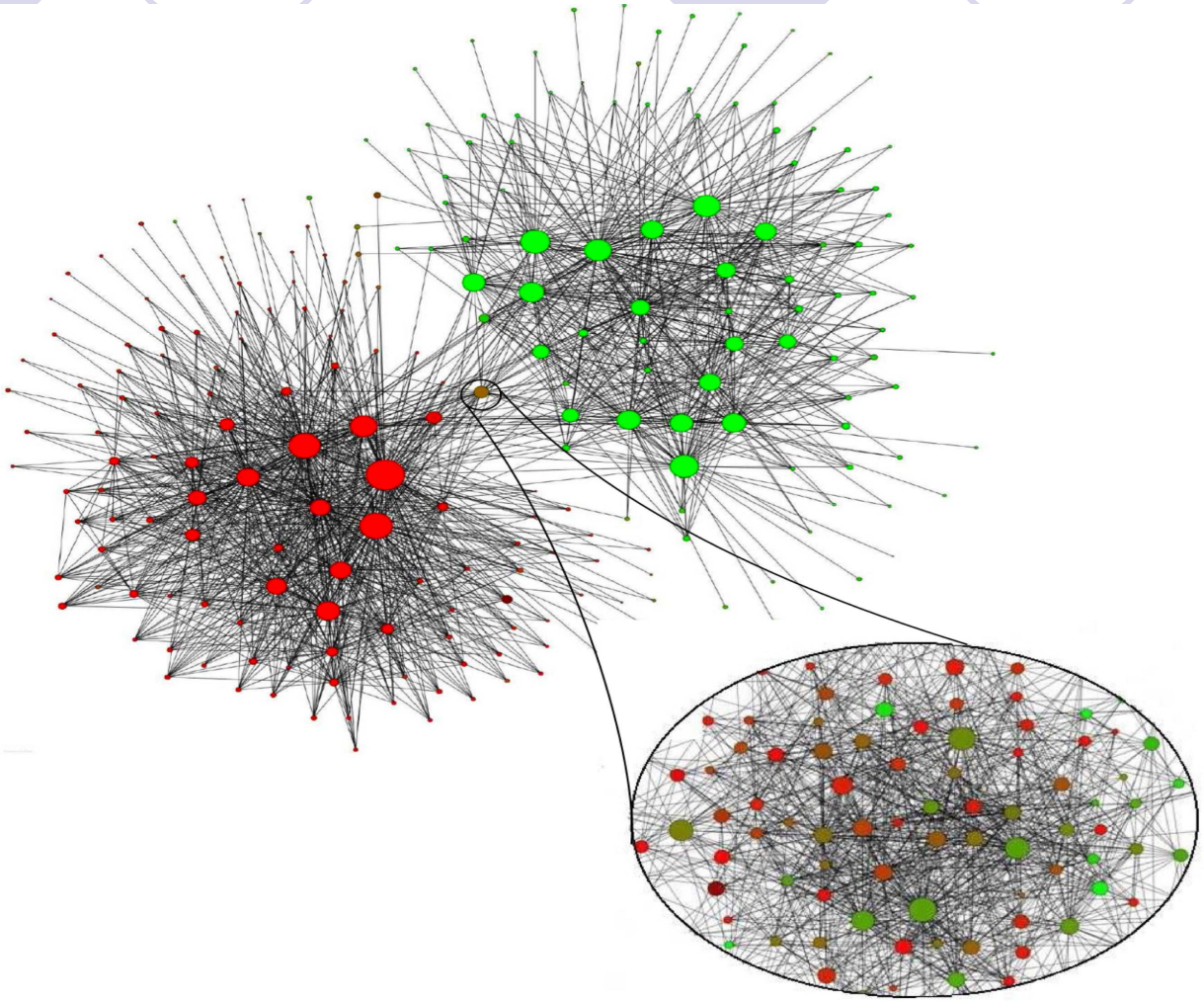
How good is a community

- Communities are dense compared to random case
- Measured in terms of modularity
- Total number of in-community edges – expected number of edges if there is no community structure

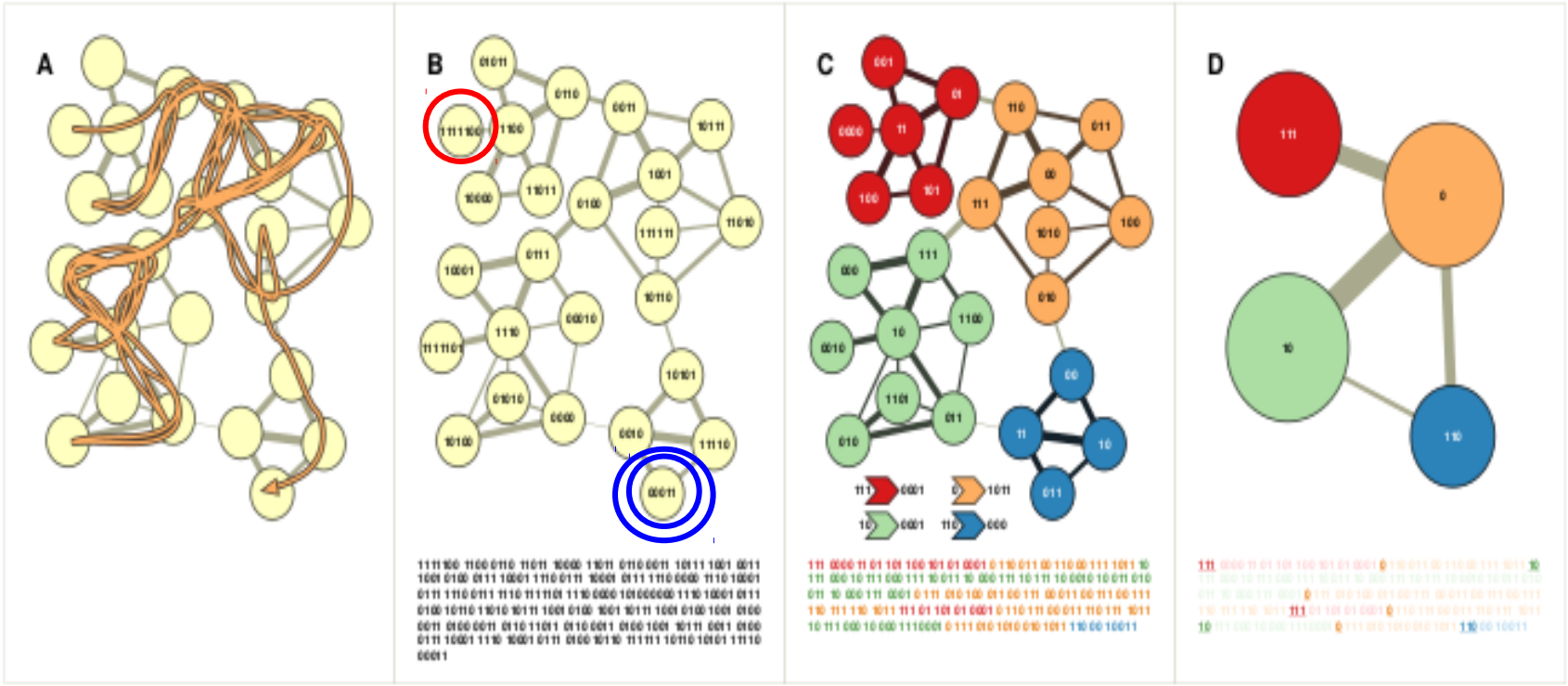
Modularity Optimization (Blondel et al.)



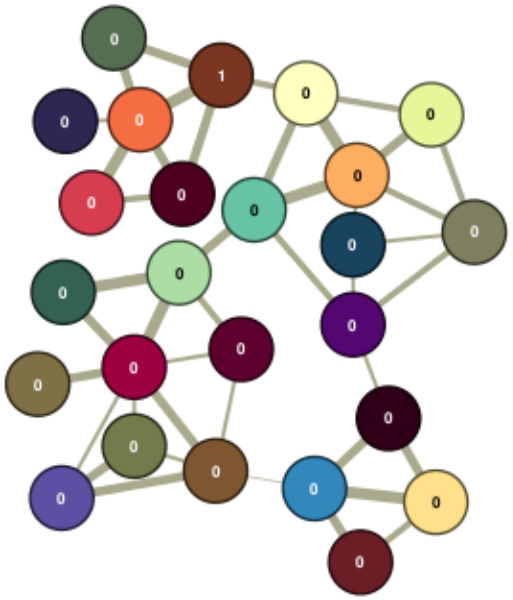
On Execution

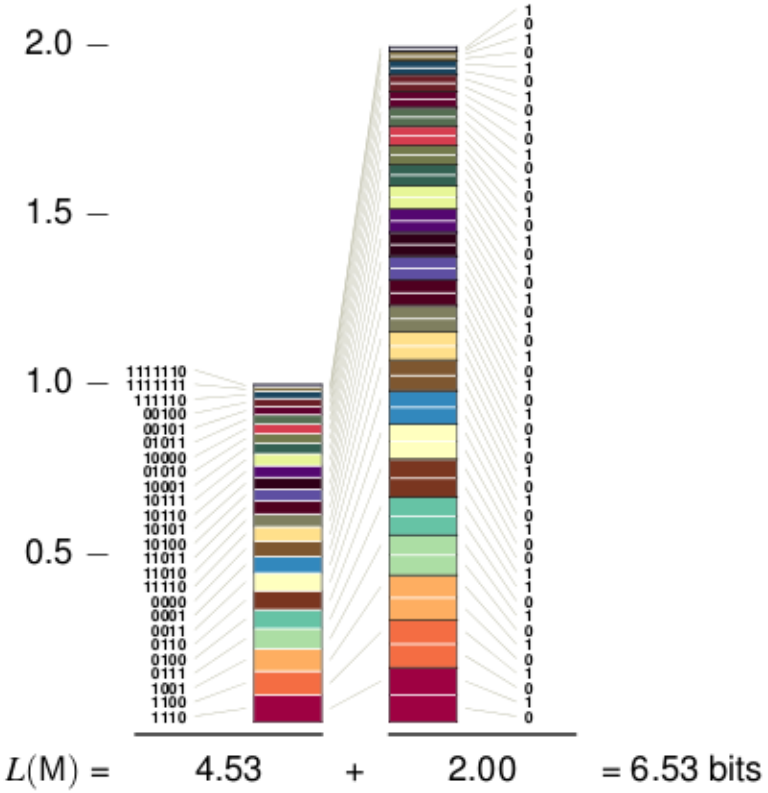


Infomap

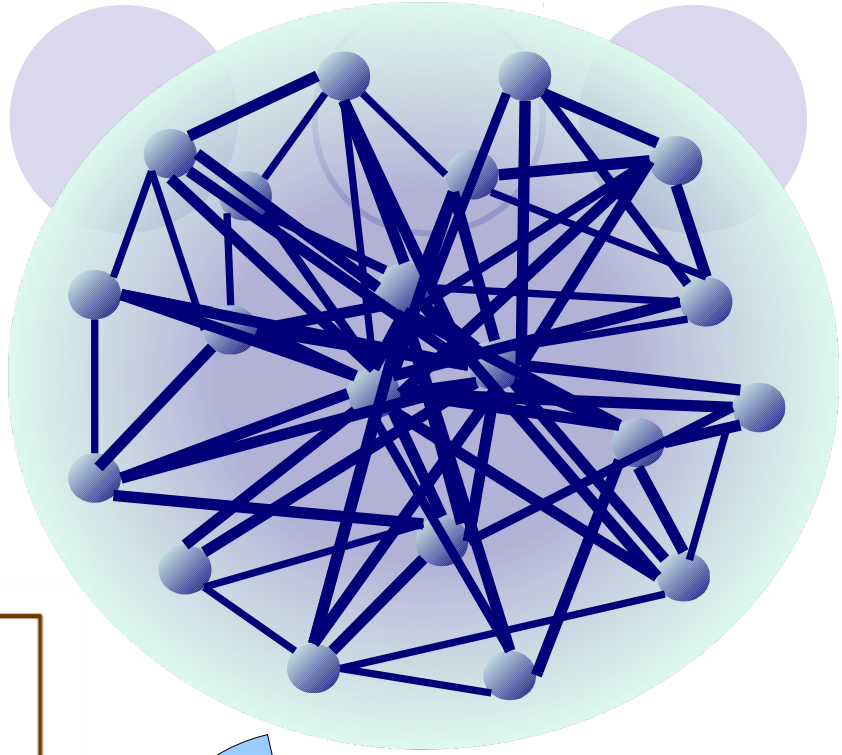


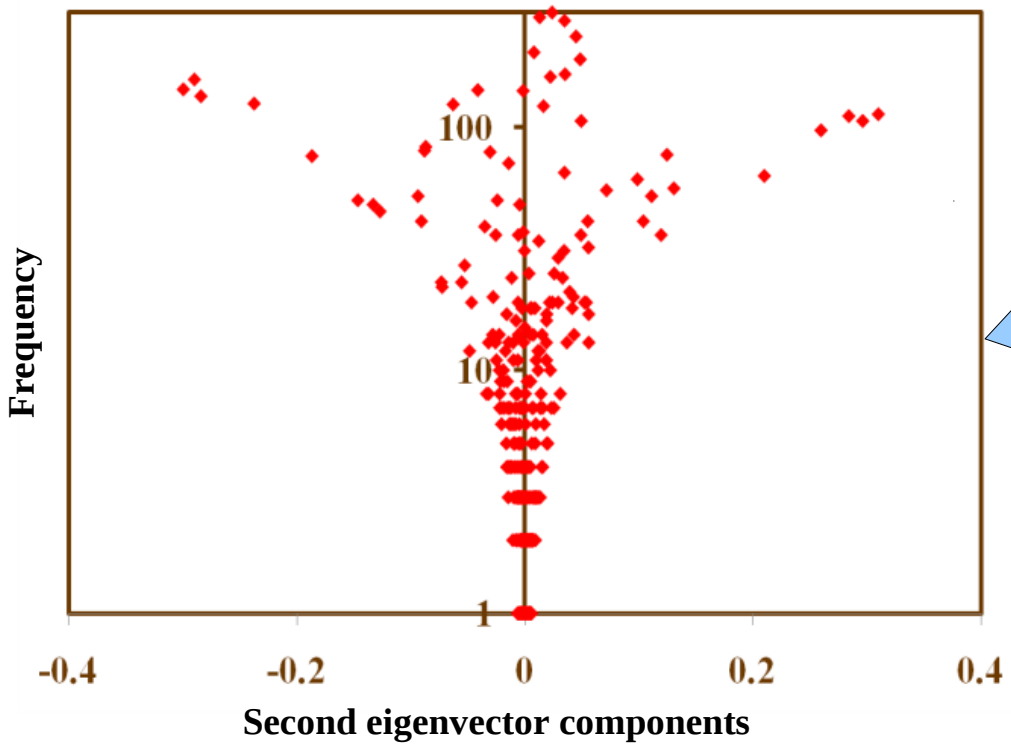
In Execution



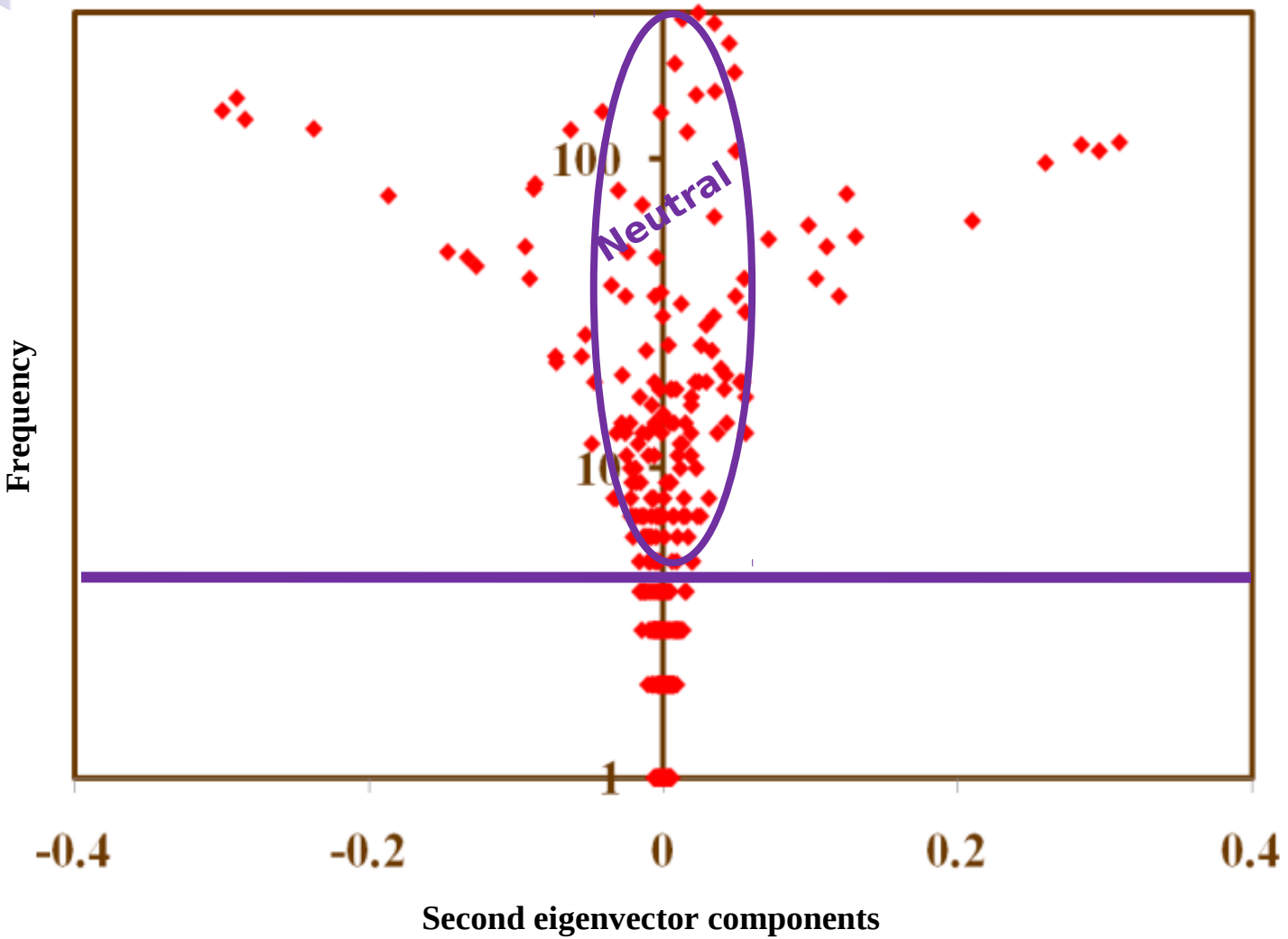


Spectral Bisection





In Execution



In Execution