COMMUNITY DISCOVERY

PART 1: A (BRIEF) INTRODUCTION

Giulio Rossetti
WMA - 4 May 2015

Community Discovery

The aim of Community Discovery algorithms is to identify communities hidden into complex network structures

Why Community Discovery?

- “Cluster” homogeneous nodes relying on topological information
  - (Clustering networked entities)

Major Problems:

- Community Discovery is an ill posed problem
  - Each algorithm models *different properties* of communities
- Different approaches comparison
- Context Dependency

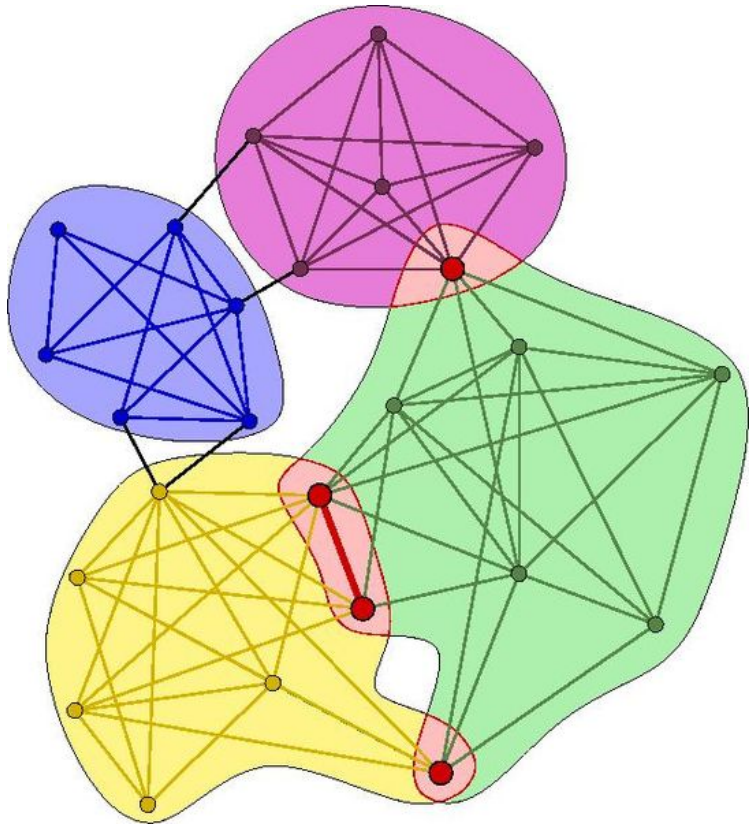
Community Charachteristics

Given the complexity of the problem a number of different typologies of approaches where proposed in order to:

Analyze:
- Directed\Undirected graphs
- Weighted\Unweighted graphs
- Multidimensional graphs
- …

Following:
- Top-Down\Bottom-Up partitioning
- …

Producing:
- Overlapping Communities
- Hierarchical Communities
- …



But…what is exactly a community?

Unfortunately does not exist a completely shared definition of what a community is…

A *general* idea is that a community represent:

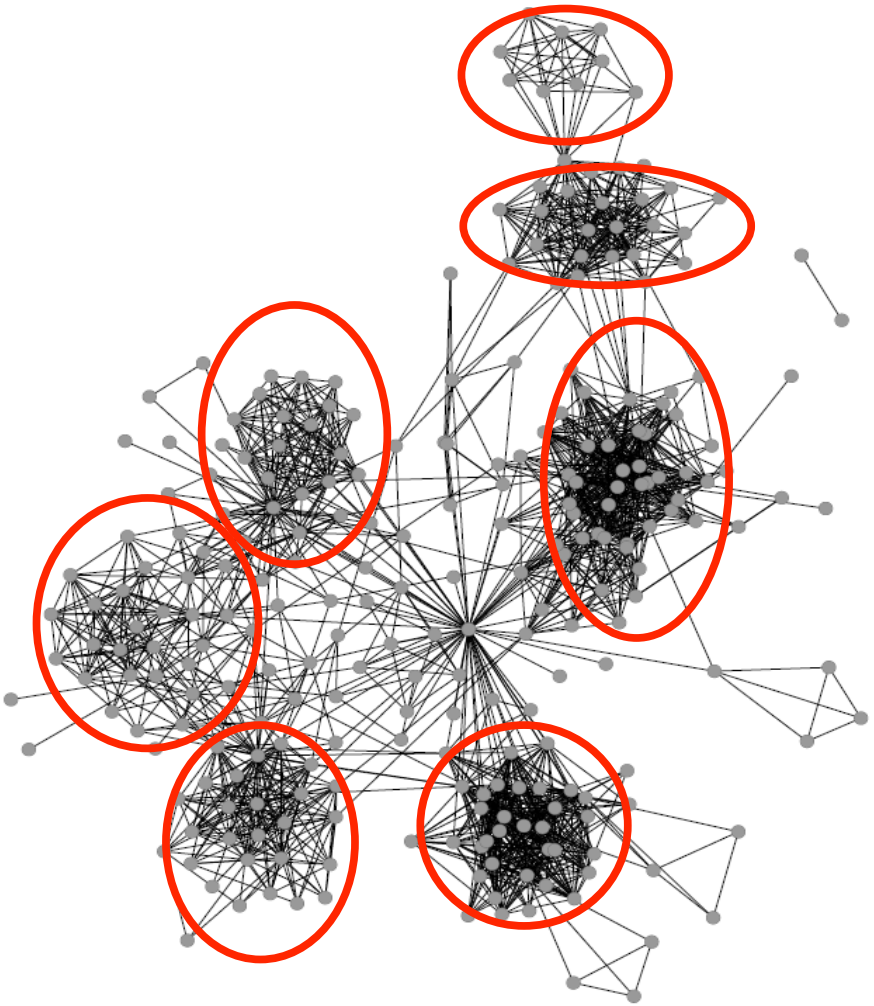
"A set of entities where each entity is closer, in the network sense, to the other entities within the community than to the entities outside it."

or

"A set of nodes tightly connected within each other than with nodes belonging to other sets."

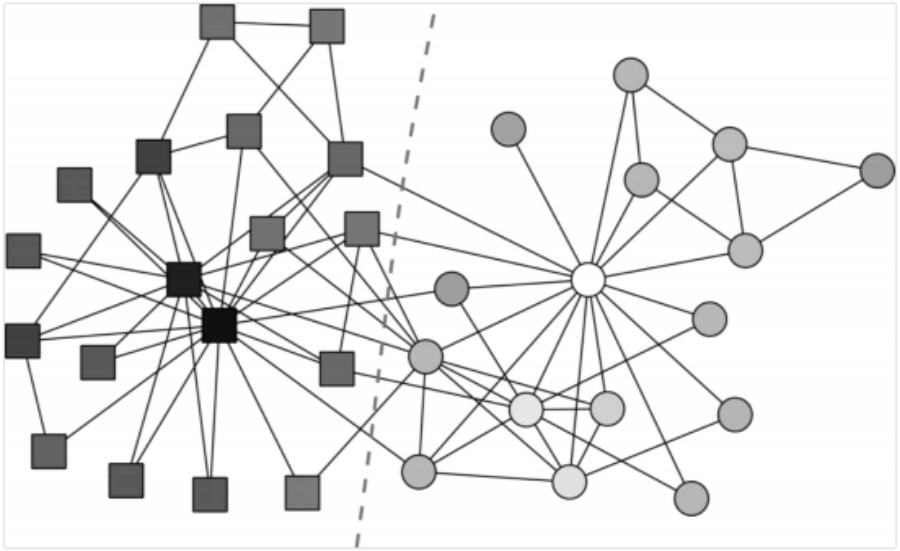
Communities in Complex Networks

- Communities can be seen as the basic bricks of a network
- In simple, small, networks it is easy identify them by looking at the structure…



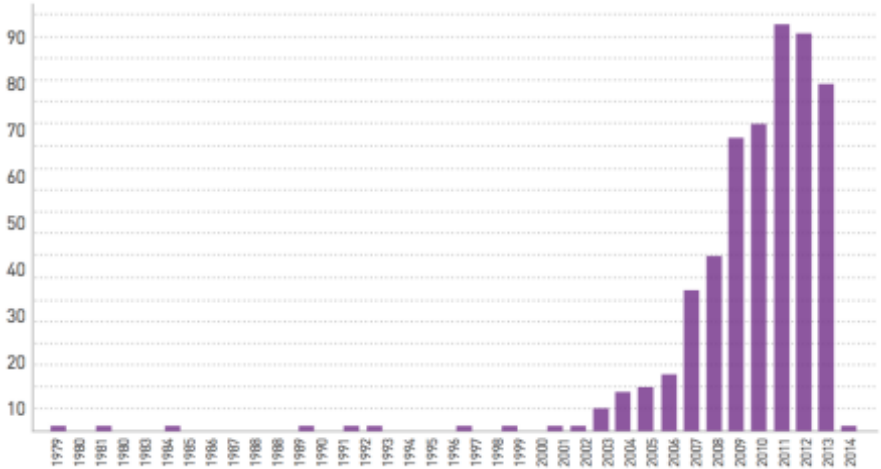
A first example…

Zachary's Karate Club

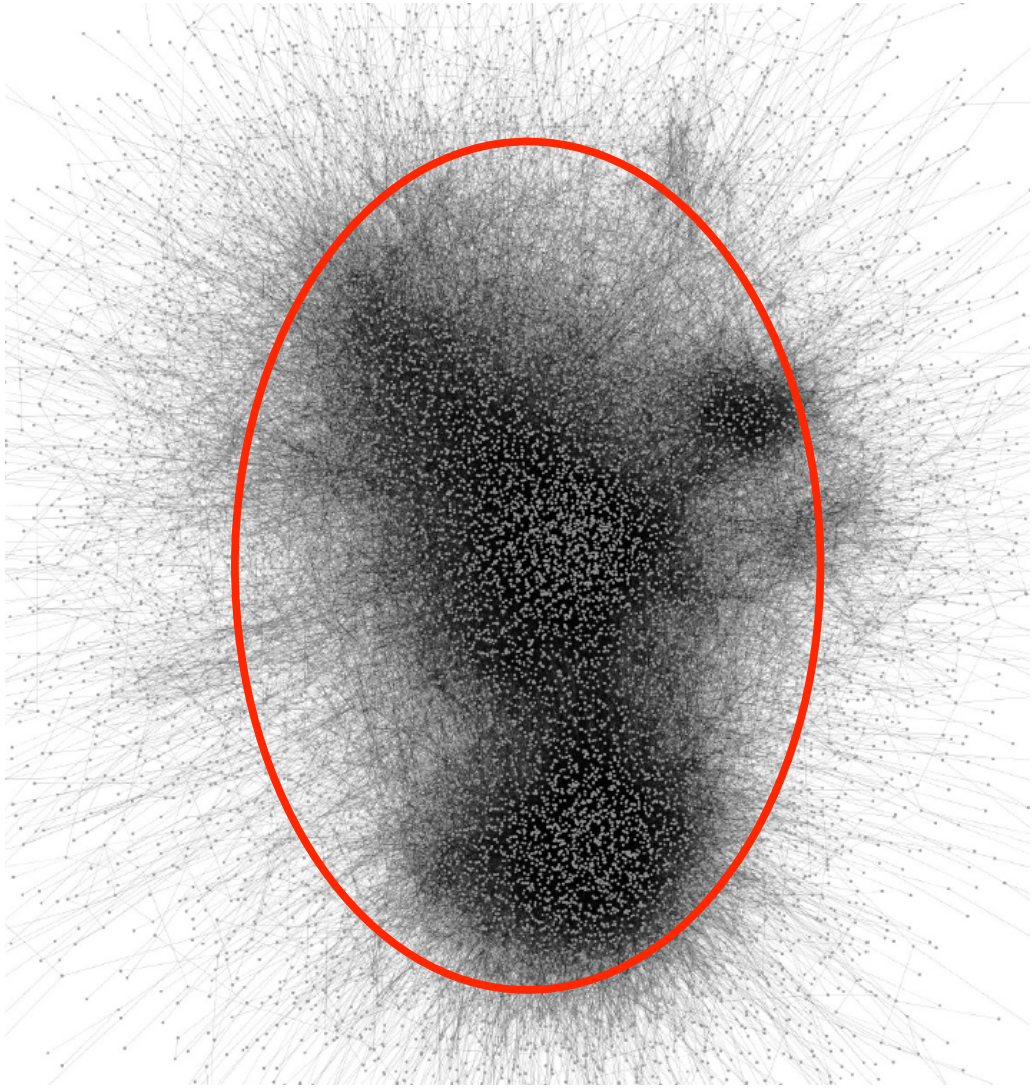


Communities emerge from the breakup of the Club

Citation history
of the Zachary's Karate club paper



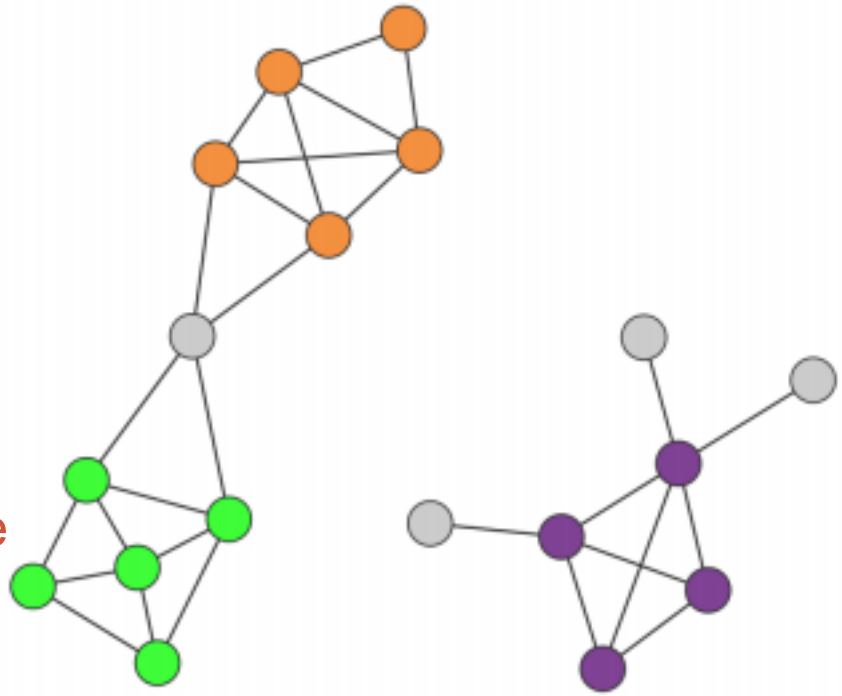
…however, real world networks are not often that “simple”…



- We can’t identify easily different communities
- Too many nodes and edges

Communities: some Hypotesis

- **H1**: The community structure is uniquely encoded in the wiring diagram of the overall network
- **H2**: A community corresponds to a connected subgraph
- **H3**: Communities are locally dense neighborhoods of a network



COMMUNITY DISCOVERY

PART 2: THE NIGHTMARE OF AN ILL POSED PROBLEM

Giulio Rossetti
WMA - 4 May 2015

Which kind of community do you like?

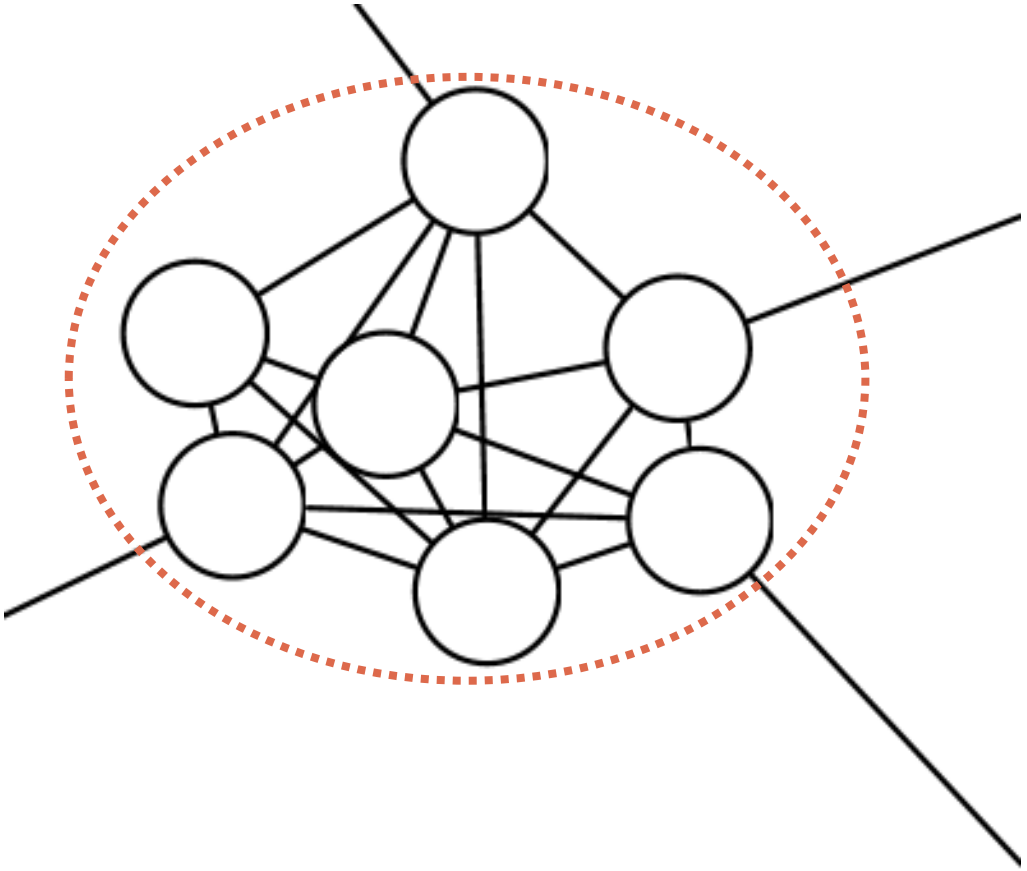
Can we classify CD algorithms according to a taxonomy of community definitions?

We can divide CD algorithms in:

- Internal density
- Bridge detection
- Feature distance
- Percolation
- Entity closeness
- Structure definition
- Link communities
- No a priori definition

Internal Density

"A community in a complex network is a set of entities that are densely connected"



Each community must have a number of edges significantly higher than the expected number of edges in a random graph

Internal density (cont'd)

How to assure high density?

General Idea:

- define a quality function measuring the density of a community and then try to maximize it

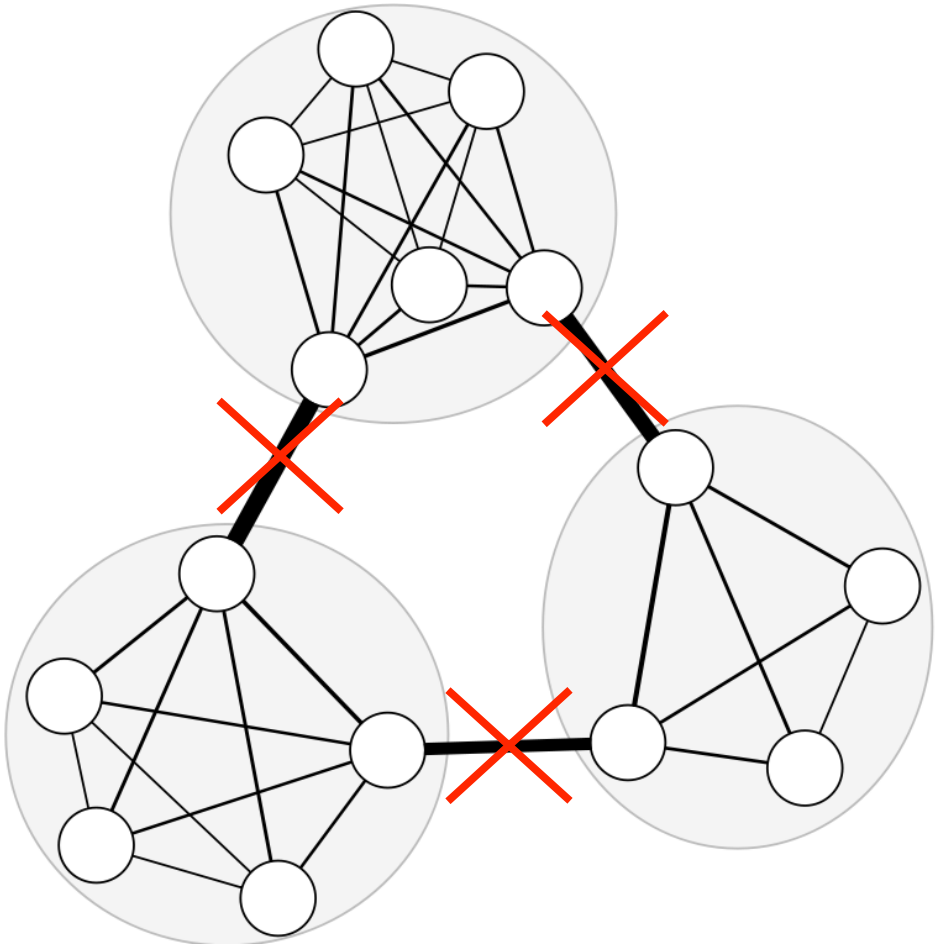
Popular concept:
Modularity optimization

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Example:
Louvain

Bridge detection

"A community is a component of the network obtained by removing from the structure all the sparse bridges that connects the dense parts of the network"



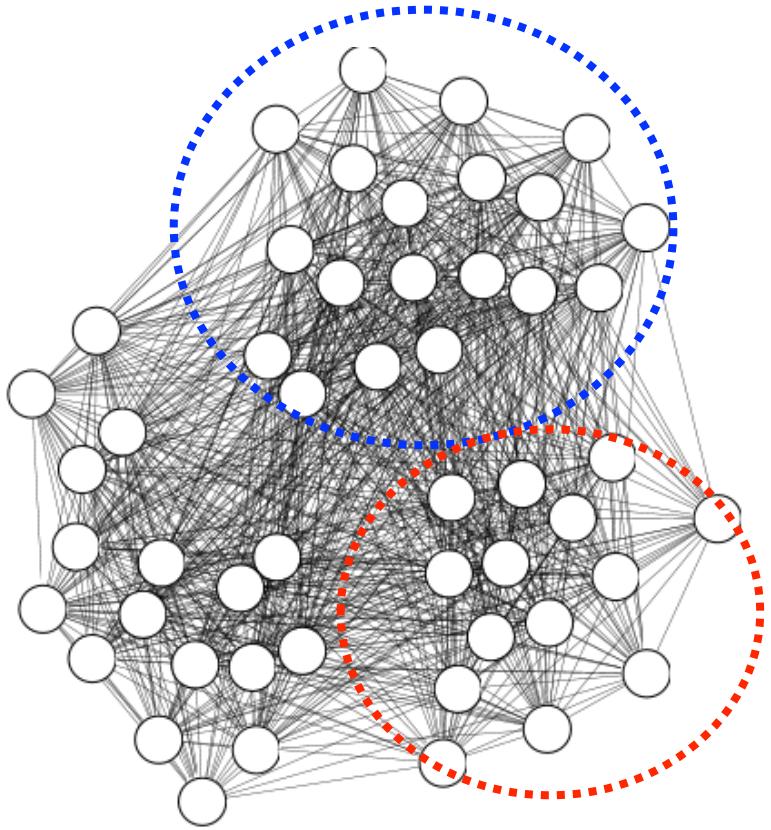
Partitioning approaches

Example: Girvan-Newman (*edge-betweenness*)

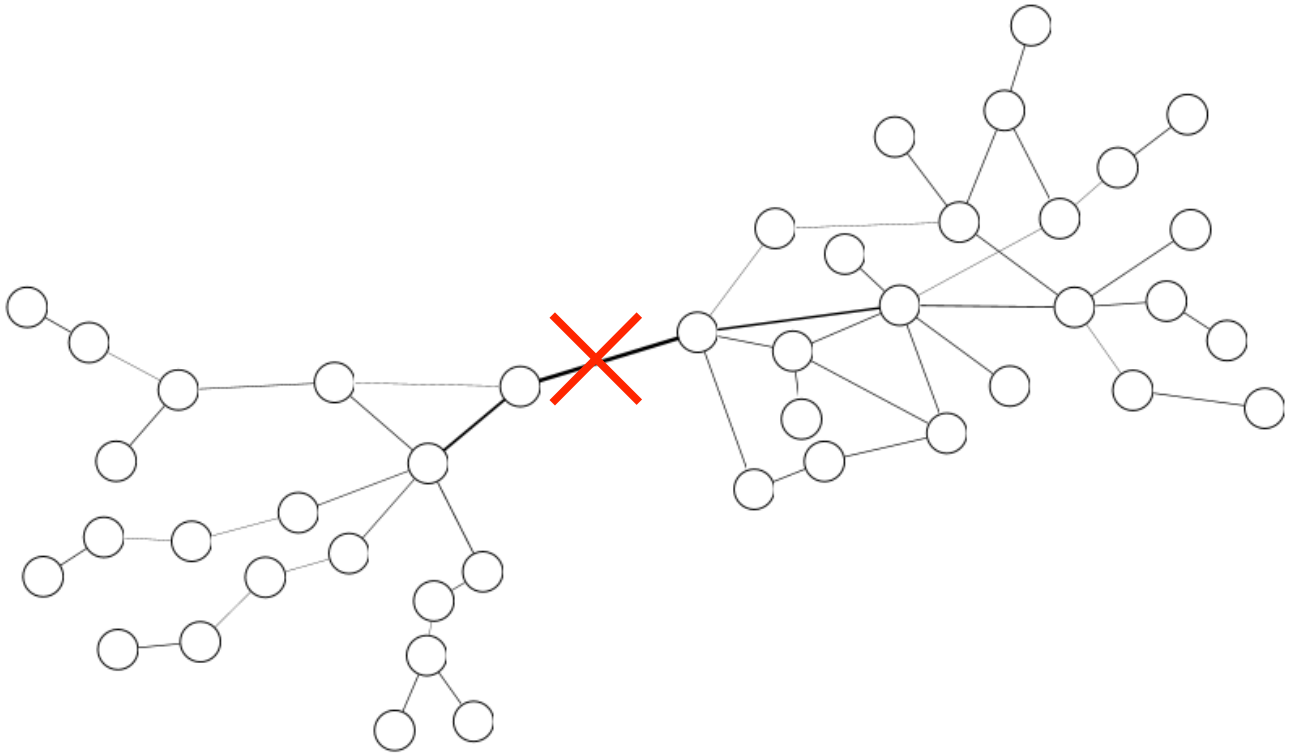
Density vs Bridges

These two definitions seems very similar... Are they equivalent?

- for some networks yes;
- for very dense network there are no clear bridges.



Density vs Bridges (cont'd)

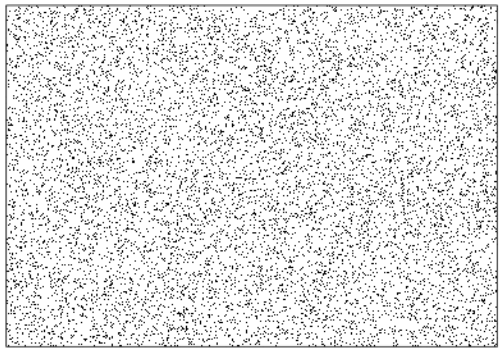


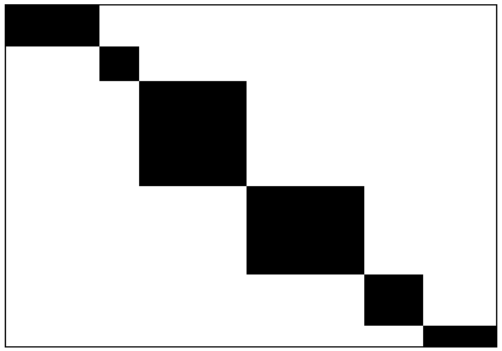
Moreover, for very sparse networks a density definition will fail even if we can detect some bridges

Feature distance

"A community is a set of entities that share a precise set of features"

Once defined a distance measure based on the values of the selected features, the entities within a community are very close to each other, more than entities outside the community.



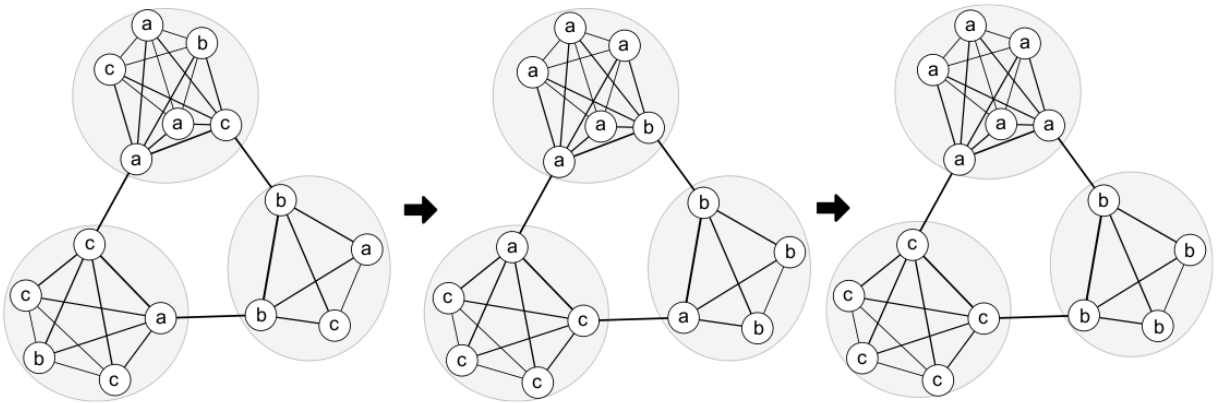


Clustering approach

- It considers any kind of vertex features, not only their adjacencies (in the latter case we can map this definition in the density one).

Percolation

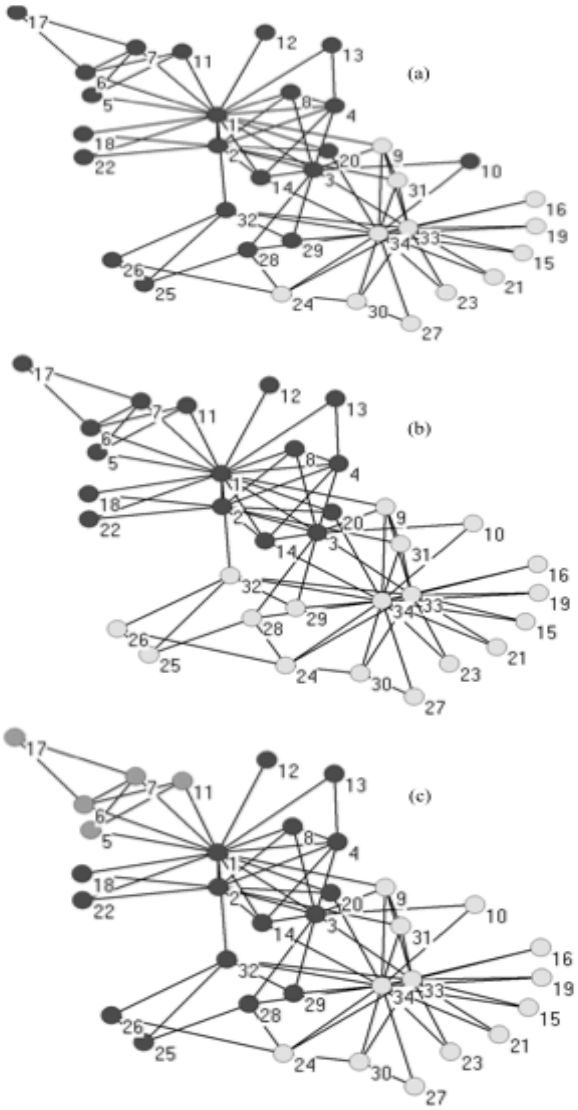
"A community is a set of nodes who are grouped together by the propagation of the same property, action or information in the network"



Example:
Label Propagation, DEMON

Label Propagation

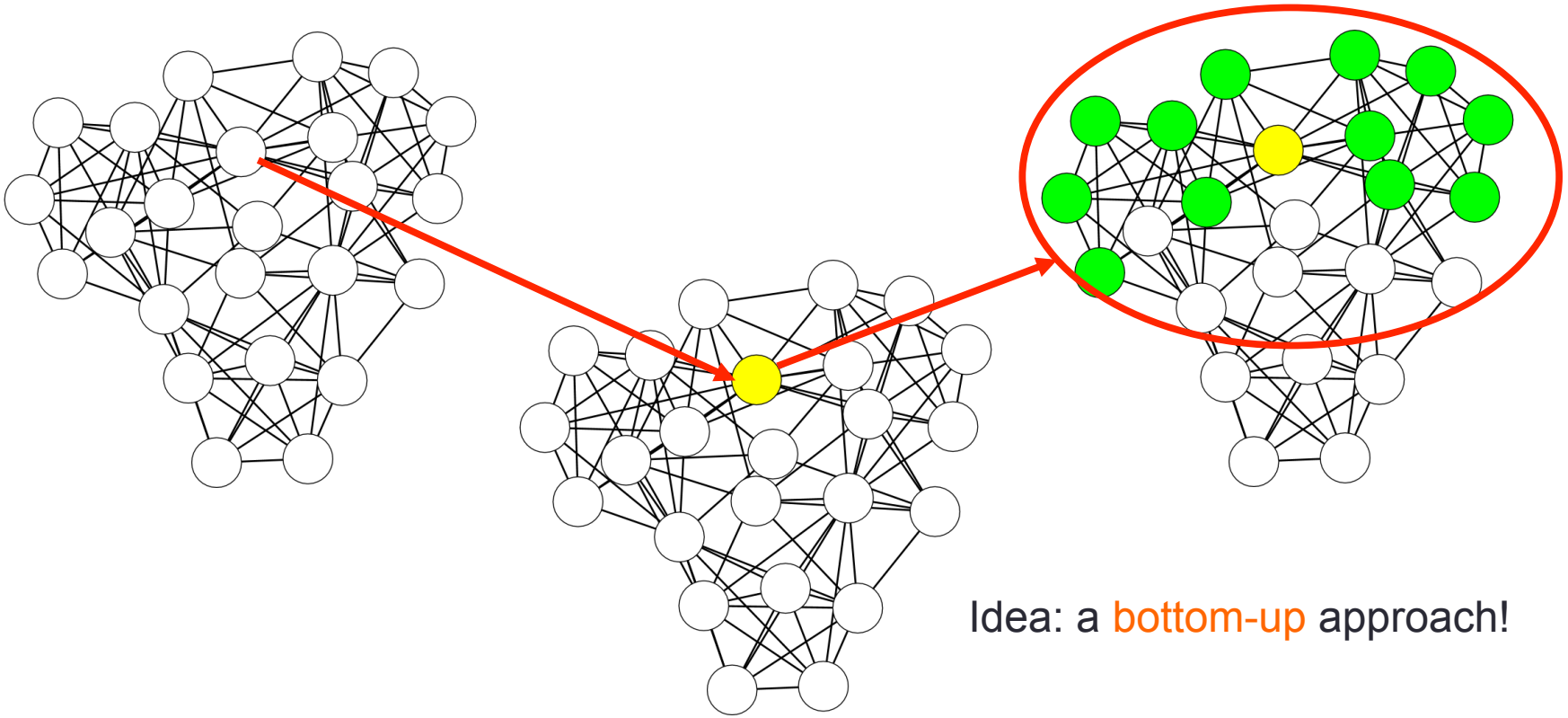
- Each node has an unique label (i.e. its id)
- In the first (setup) iteration each node, with probability α, change its label to one of the labels of its neighbors;
- At each subsequent iteration each node adopt as label the one shared *(at the end of the previous iteration)* by the majority of its neighbors;
- We iterate untill consensus is reached.



DEMON: A Matter of Perspective…

Locally, each node is able to identify its communities

Globally, we are tangled in complex overlaps



Idea: a bottom-up approach!

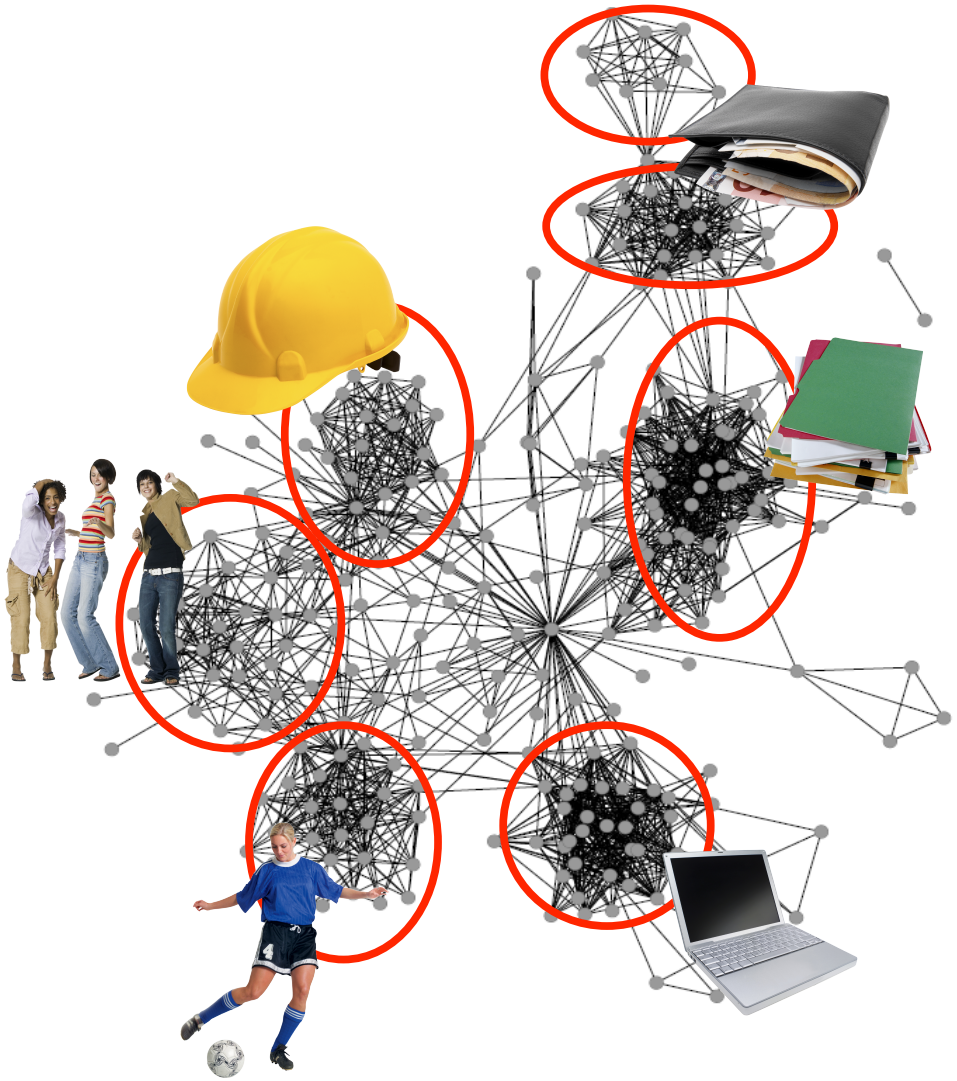
Reducing the complexity

Real Networks are Complex Objects

Can we make them “simpler”?

Ego-Networks

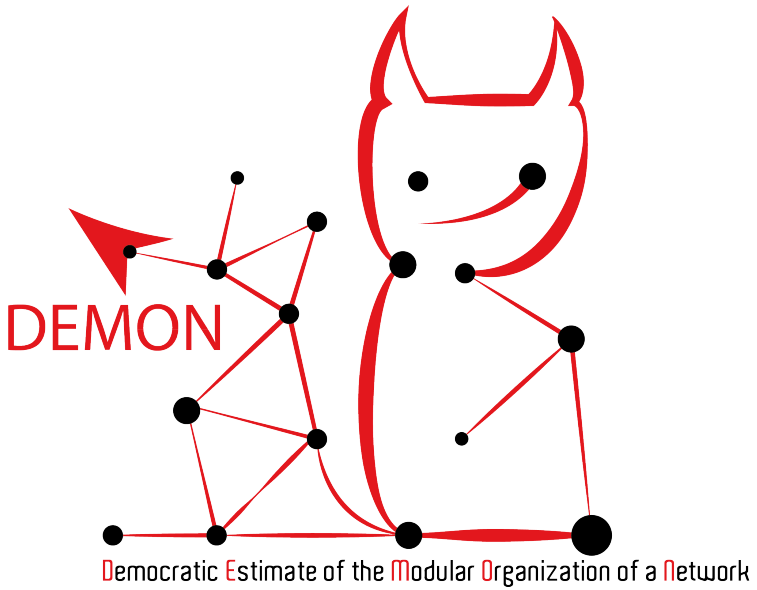
(networks builded upon a focal node , the "*ego*”, and the nodes to whom *ego* is directly connected to plus the ties, if any, among the alters)



DEMON Algorithm

For each node n:

1. Extract the Ego Network of n
2. Remove n from the Ego Network
3. Perform a Label Propagation
4. Insert n in each community found
5. Update the raw community set C



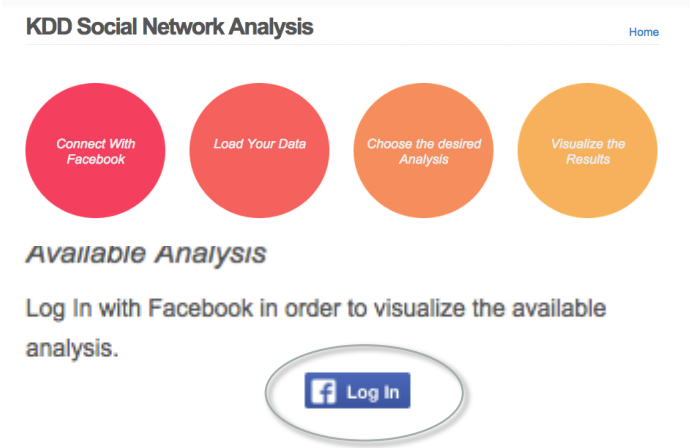
For each raw community c in C

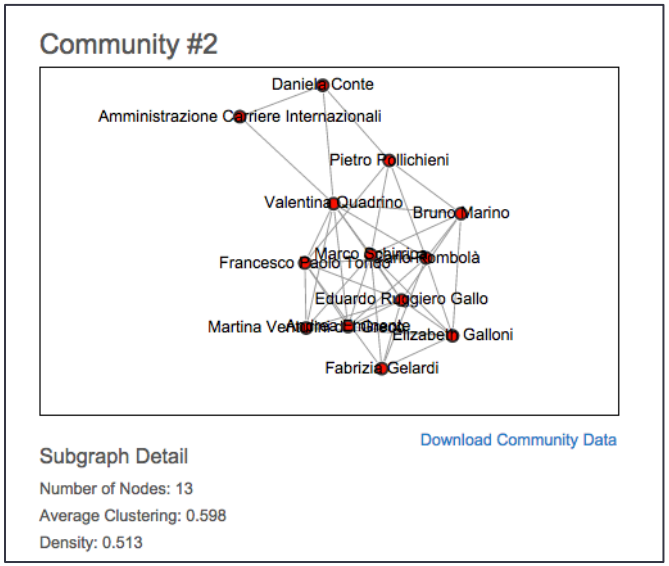
1. Merge with “similar” ones in the set (given a threshold)
   (i.e. merge iff at most the ε% of the smaller one is not included in the bigger one)

DEMON@Work

Personal Facebook Communities

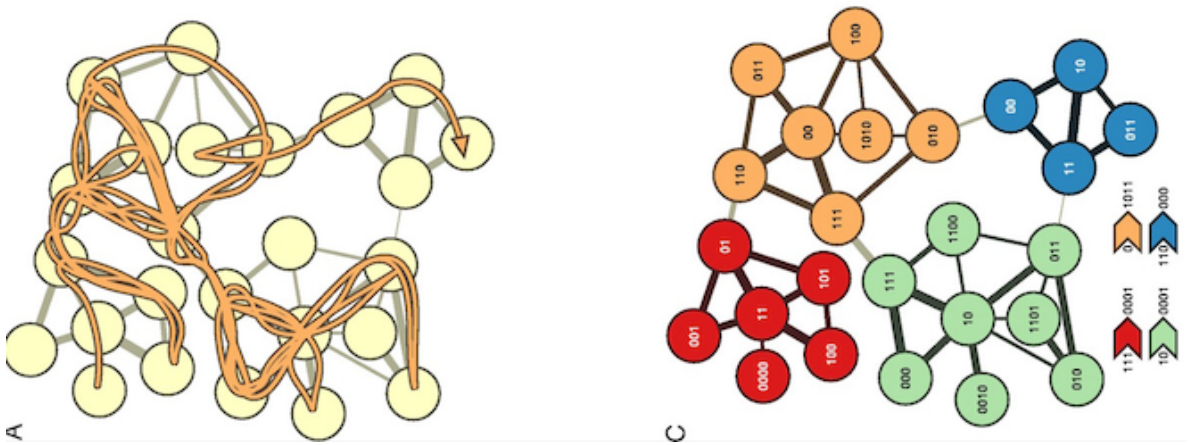
1. Log out from Facebook and clean your browser cookies
2. Visit: **kddsna.isti.cnr.it:8080**
3. Log In with Facebook
4. Select one of the two options:
   1. “Visualize your network”
   2. “Demon Communities”
5. **Wait** for the data to be collected and displayed
6. Zoom-in/out and drag communities with your mouse





Entity closeness

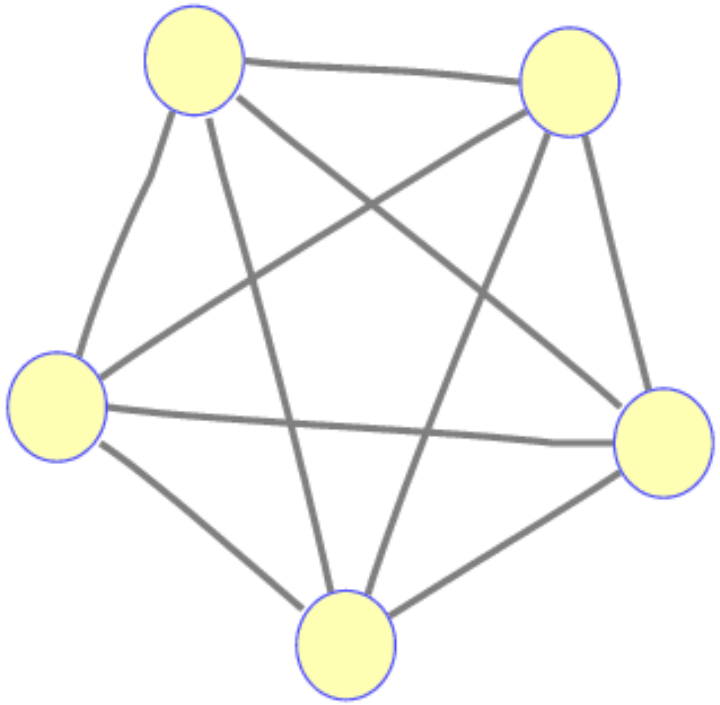
"A community is a set of nodes who can reach any member of its group crossing a very low number of edges, significantly lower than the average shortest path in the network"



Example:
Random Walks
(conductance analysis)

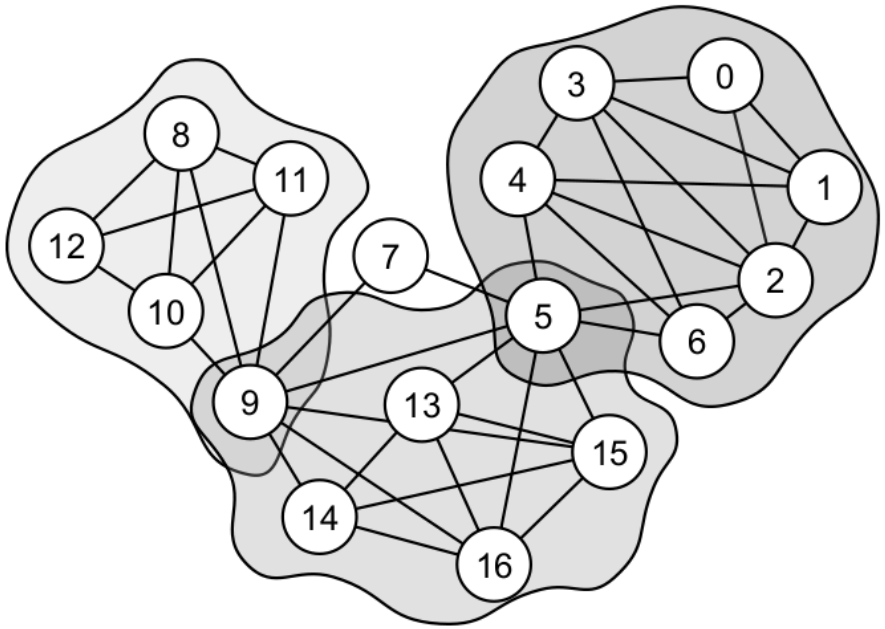
Structure definition

"A community is a set of nodes with a precise number of edges among them, distributed in a very precise topology defined by a number of rules"



Example:
k-cliques

Example: Clique percolation



- A very popular structure definition algorithm: **k-cliques**
- Also this case is different from the density definition: node 7 is in some sense “dense” (is in a triangle), but outside of any community

Link communities

"A community is a set of nodes which share a number of relations clustered together since they belong to a particular relational environment"

It is the relation that belongs to a community and then the nodes belong to the communities of their connections

No a Priori Definitions

"Communities are sets which present a number of particular features defined by an analyst"

COMMUNITY DISCOVERY

PART 3: EVOLUTIONARY COMMUNITY DISCOVERY

Giulio Rossetti
WMA - 4 May 2015

Are we missing something?

Real world networks evolve quickly:

- Social interactions
- Buyer-seller
- Stock-exchanges
- …

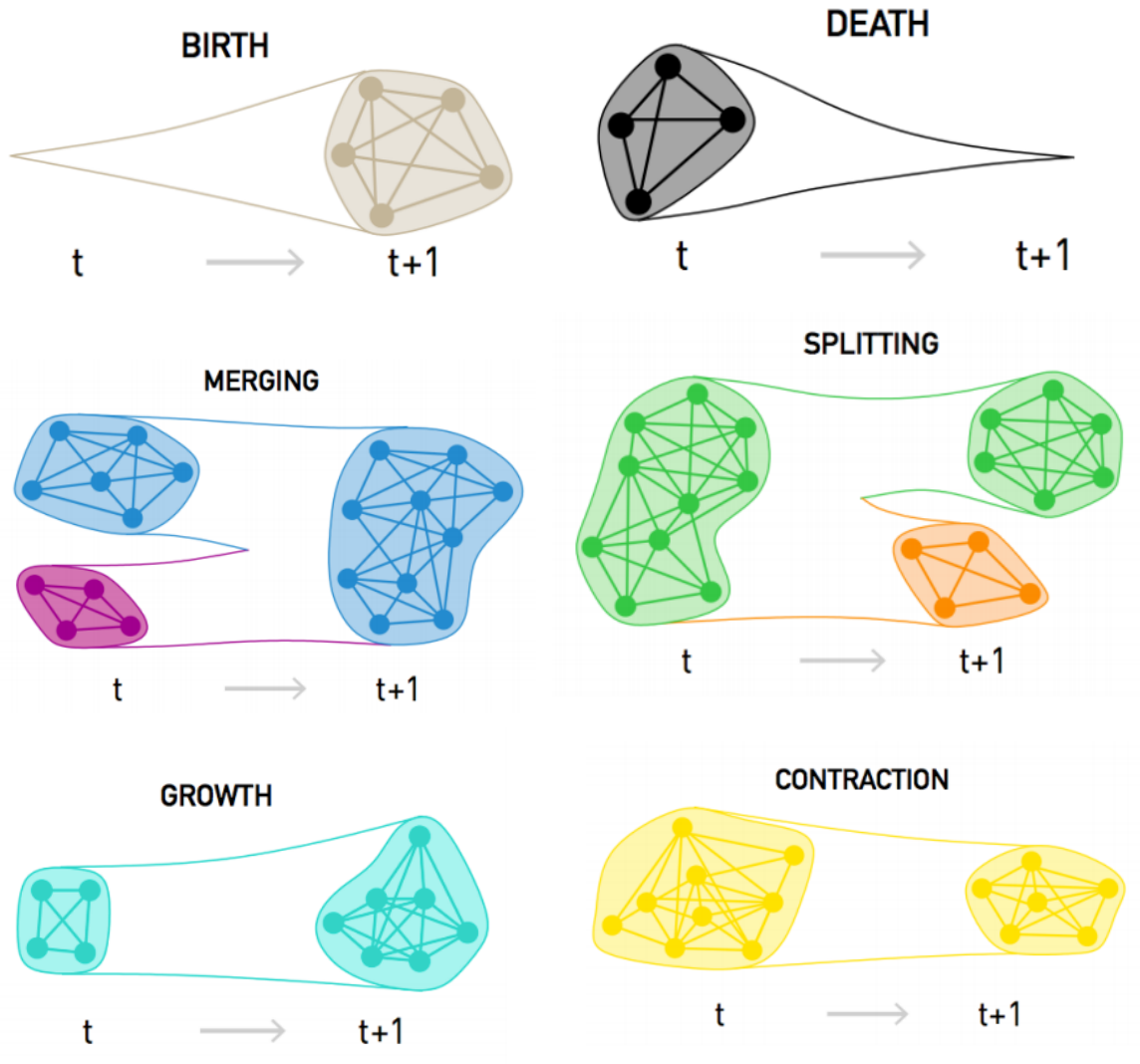
In these scenarios a QSSA (Quasi Steady State Assumption) rarely holds:

- Network cannot be *"frozen in time"*
  - Nodes and edges rise and fall producing perturbation on the whole topology
- The reduction to static scenarios trough temporal discretization is not always a good idea
  - How can we chose the temporal threshold?
  - To what extent can we trust the obtained results?

Community life-cycle

As time goes by the **rising** of novel nodes and edges (as well as the **vanishing** of old ones) led to network perturbations

Communities can be deeply affected by such changes



Conclusions

- Nowadays Community Discovery is, perhaps, the hottest topic in complex network analysis

- Major issues:
  - Problem definition
  - Community evaluation

- Problem specializations:
  - Multidimensional Community Discovery
  - Evolutionary Community Discovery
    - How communities evolve in dynamic networks?
  - …

Bibliography

- S. Fortunato. 2010. *Community detection in graphs*. Physics Reports 486 (Feb. 2010).

- M. Coscia, F. Giannotti, and D. Pedreschi. *A classification for community discovery methods in complex networks*. Statistical Analysis and Data Mining 4, 5 (2011), 512–546.