Local approaches for community detection in complex networks:

Preference networks, boundary nodes and gossip propagation

Mürsel Taşgın

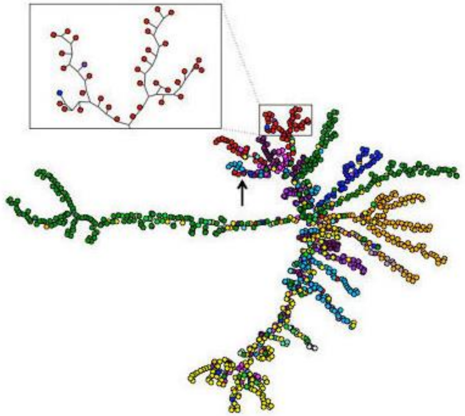
Department of Computer Engineering

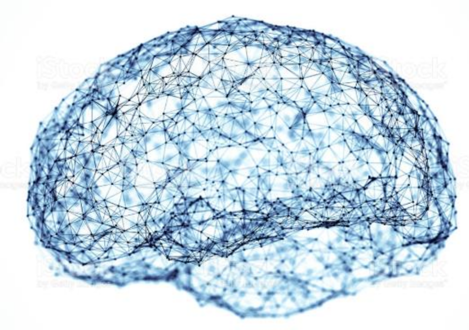
Boğaziçi University

What is a network?

A collection of entities (nodes) that are interconnected with links (edges).

- people that are friends
- group of interconnected computers
- brain cells that have similar functions
- web pages pointing to each other
- proteins interacting together
- phone calls between people
- authors collaborating on same paper
- airline traffic connections
- underground map of cities





What makes a network a “complex network”?

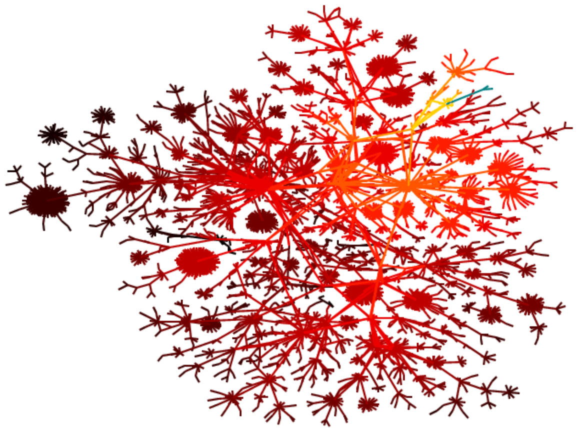
- Many agents or individuals interacting with each other

- Each component has its own internal structure and each one perform an specific behavior or function

- The behavior of a small part of the system affects in a non linear way the whole system.

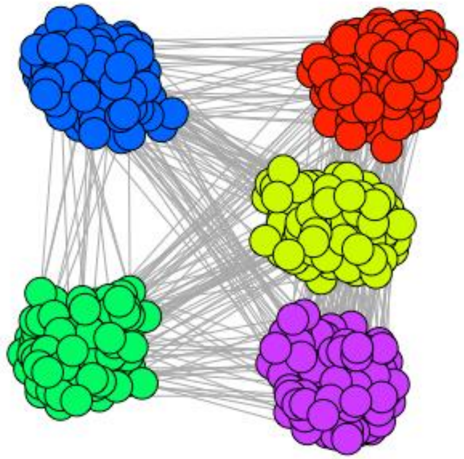
Complex networks present emergent behavior in such a way that any property of the system is not the **simple sum of its parts**.

Community and community detection



- Community

  Group of nodes more connected within community than rest or the network

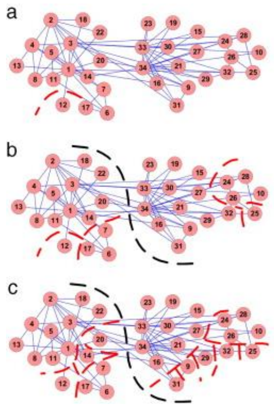


- Community detection

  Task of identifying communities in networks

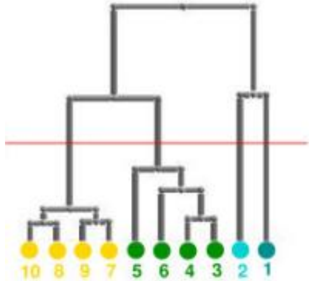
Community detection algorithms

Divisive algorithms



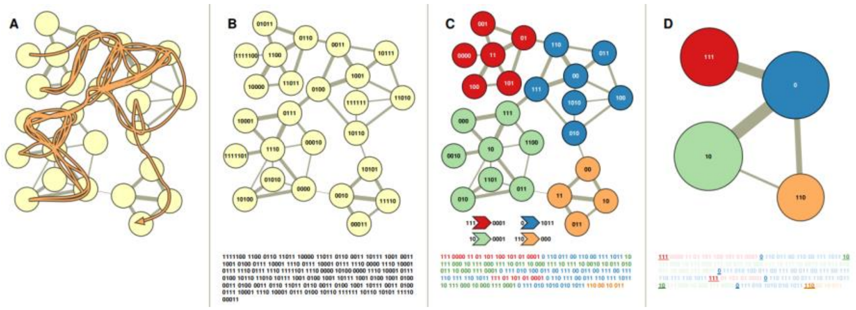
GN, centrality-based algorithms

Agglomerative algorithms



Newman's greedy algorithm

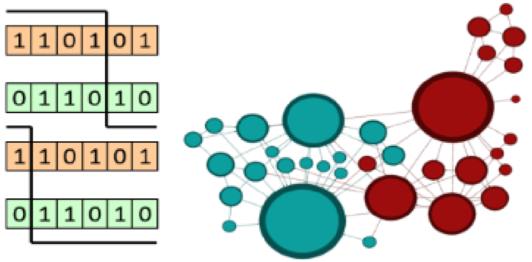
Information theoretic approaches



Infomap

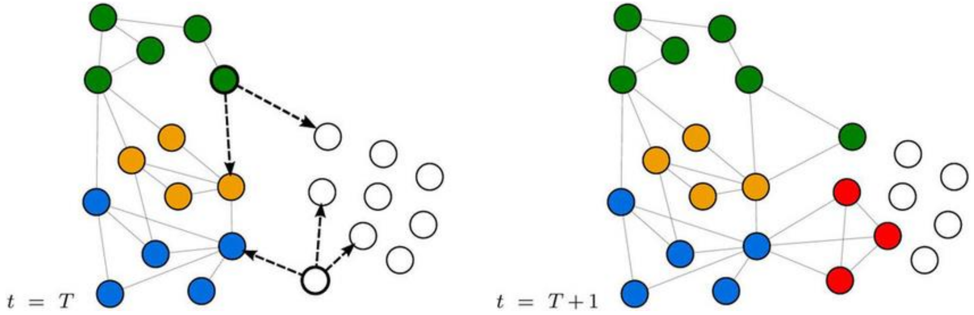
Some examples of community detection algorithms

Using AI methods



Community detection using genetic algorithm

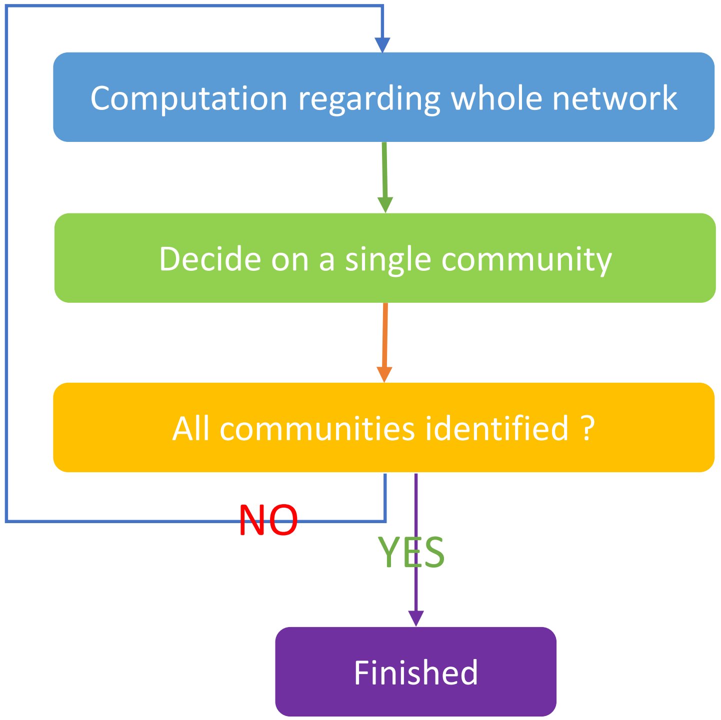
Label propagation



LPA

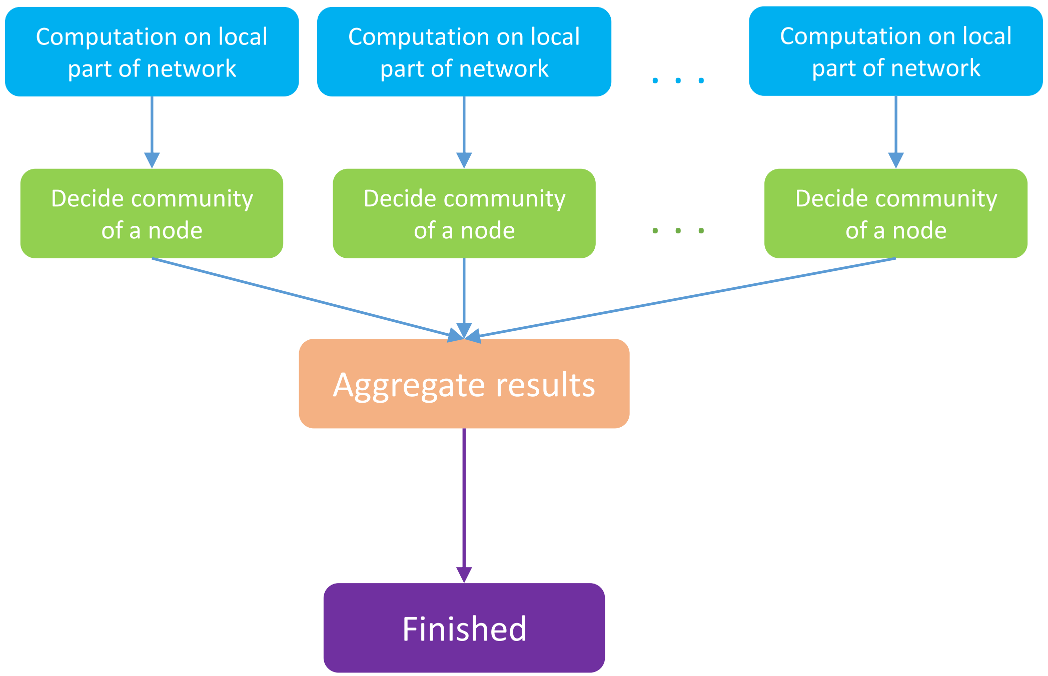
Global approach vs local approach

Global approach



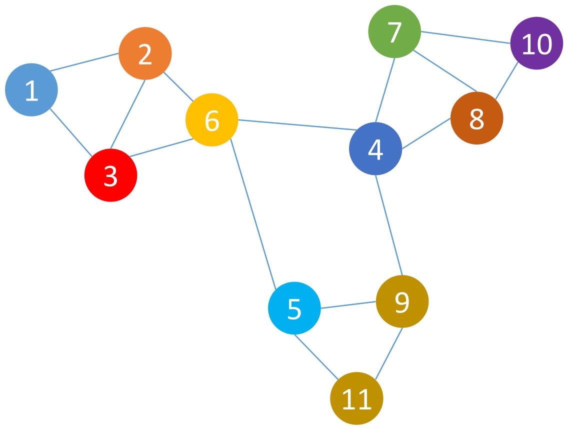
+ Decisions with more information
- High time-complexity
- No parallelism, not distributed

Local approach



- Decisions with less information
+ Low time-complexity
+ Parallel and distributed execution

Community detection using preference networks



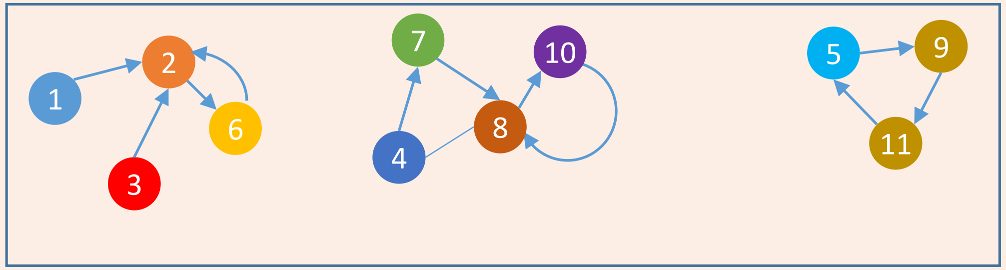
Original network

1 Ask each node

With whom do you prefer to be in same community?

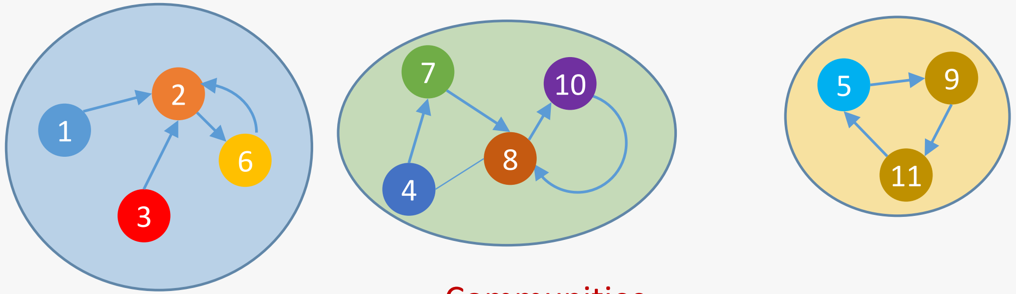
1	2	3	4	5	6	7	8	9	10	11	→ *Node ID*
2	6	2	7	9	2	8	10	11	8	5	→ *Preferred node ID*

2 Build a directed «preference network» using answers



Preference network

3 Identify «connected components»; *i.e. communities*



Communities

Tasgin, Mursel, and Haluk O. Bingol. "Community detection using preference networks." *Physica A: Statistical Mechanics and its Applications* 495 (2018): 126-136.

Community detection using preference networks

Deciding the preferred node

With whom do you prefer to be in same community?

Ask each node

How to decide the preferred node?

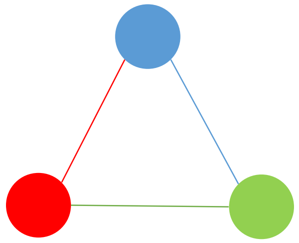
Given a node, assign a score to each neighbor and select the neighbor with highest score

$$p(i) = \underset{j \in A(i)}{\arg\max}\ s_i(j)$$

Score methods:

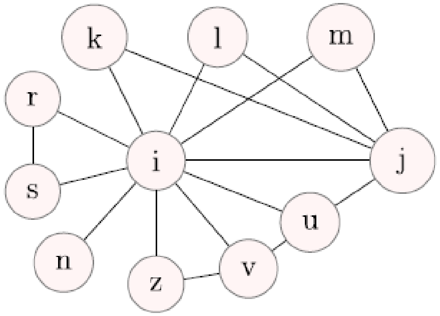
- Random number
- Degree of neighbor
- Clustering coefficient of neighbor
- Jaccard similarity
- Number of common neighbors with neighbor
- Gossip spreading capability of neighbor

Triangles and communities



"Strong social tie forms when being part of a triangle in a relation", Simmel 1950

"There exists many triangles within communities while few or no triangle exists between nodes of different communities", Radicchi et al. 2004



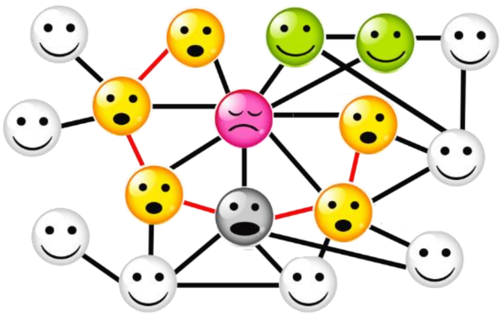
Triangle related metrics

- clustering coefficient
- # common neighbors
- Jaccard similarity
- Gossip spread capability

Community detection using preference networks

# triplets where *i* is the center	$\wedge_i$
# triangles around *i*	$\triangle_i$
1-neighborhood of *i*	$\Gamma(i)$
clustering coefficient of *i*	$CC(i) = \frac{\triangle_i}{\wedge_i}$
# common neighbors of *i* & *j*	$CN(i,j) = \lvert\Gamma(i) \cap \Gamma(j)\rvert$
Jaccard similarity of *i* & *j*	$J(i,j) = \frac{\lvert\Gamma(i) \cap \Gamma(j)\rvert}{\lvert\Gamma(i) \cup \Gamma(j)\rvert}$
# neighbors of *i* who hears gossip initiated by *j*	$\Gamma_j(i)$
Gossip spread capability of *j* around *i*	$\sigma_{i,j} = \frac{\lvert\Gamma_j(i)\rvert}{\lvert\Gamma(i)\rvert}$

Gossip spread capability



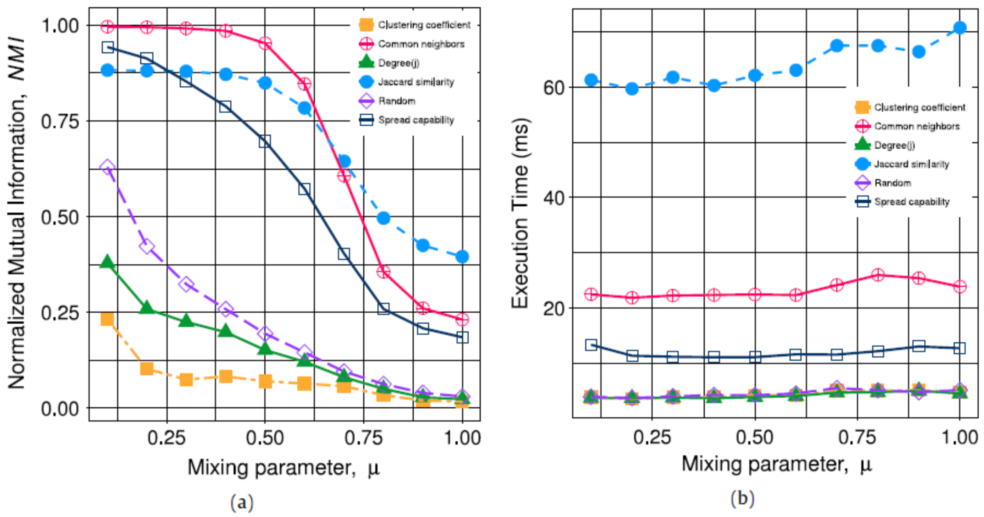
* Gossip algorithm of Lind et al.

```
Spread capability of gray
node for pink node: 6/8
```

* P. G. Lind, L. R. Da Silva, J. S. Andrade jr., H. J. Herrmann
A Letters Journal Exploring the Frontiers of Physics
EPL, 78(June 2007) 68005

Community detection using preference networks

Results on LFR generated networks of 1000 nodes



Different preference metrics

Findings

- Degree and clustering coefficient score metrics are not good
- Random score is surprisingly betten than clustering coefficient
- # common neighbors produces best result
- Jaccard similarity and spread capability also produce good results (Jaccard similarity have high execution time)

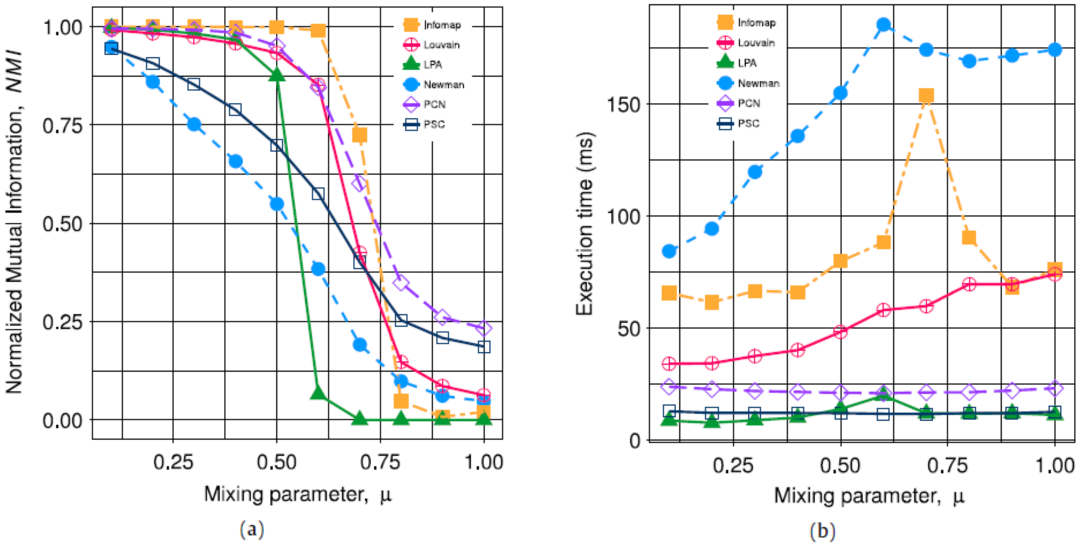
Conclusion

We select the following 2 score methods for our algorithm

- # common neighbors (PCN)
- Spread capability (PSC)

Community detection using preference networks

Results on LFR generated networks of 1000 nodes



Comparison with some known algorithms

Results

- Successfully identify dense communities (i.e. μ <0.5)
- Can identify subtle communities (i.e. μ >0.5)
- Have less execution times compared to most of the algorithms

Community detection using boundary nodes

A new algorithm based on label propagation algorithm (LPA)

LPA: Each node gets most popular label (*i.e. community identifier*) among the neighbors

Unnecessary operations of LPA is eliminated

Label update is possible only when interacting nodes have different labels

After a few iterations most of the nodes are surrounded by same labelled nodes

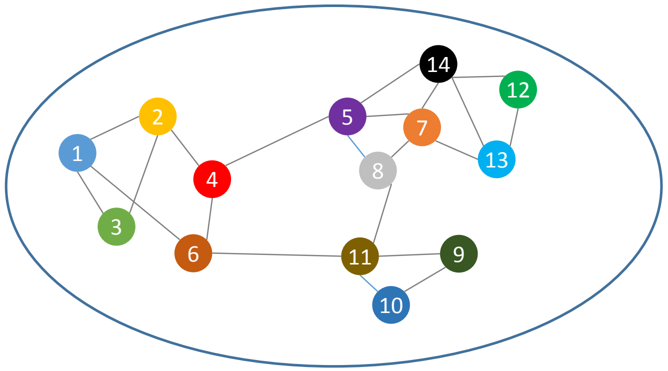
LPA should be applied to only ***boundary nodes*** (*i.e.* Node having at least 1 neighbor with different label*)*

Algorithm finds boundary nodes and their correct community label

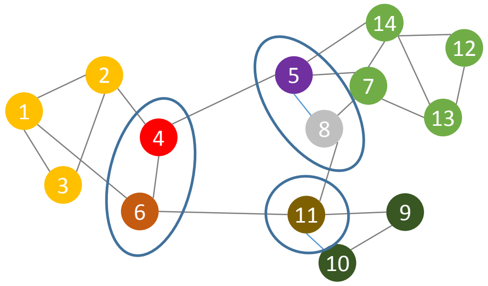
Boundary nodes will divide the network into **communities**

Label selection rule is enhanced with additional data

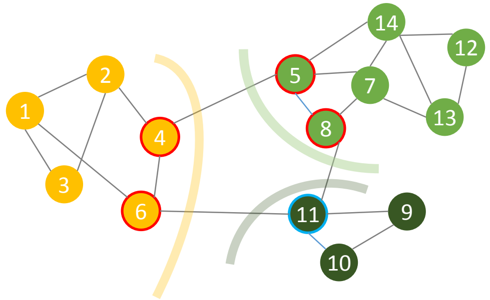
Majority rule + # common neighbors



Label propagation on all nodes



Label propagation only on boundary nodes

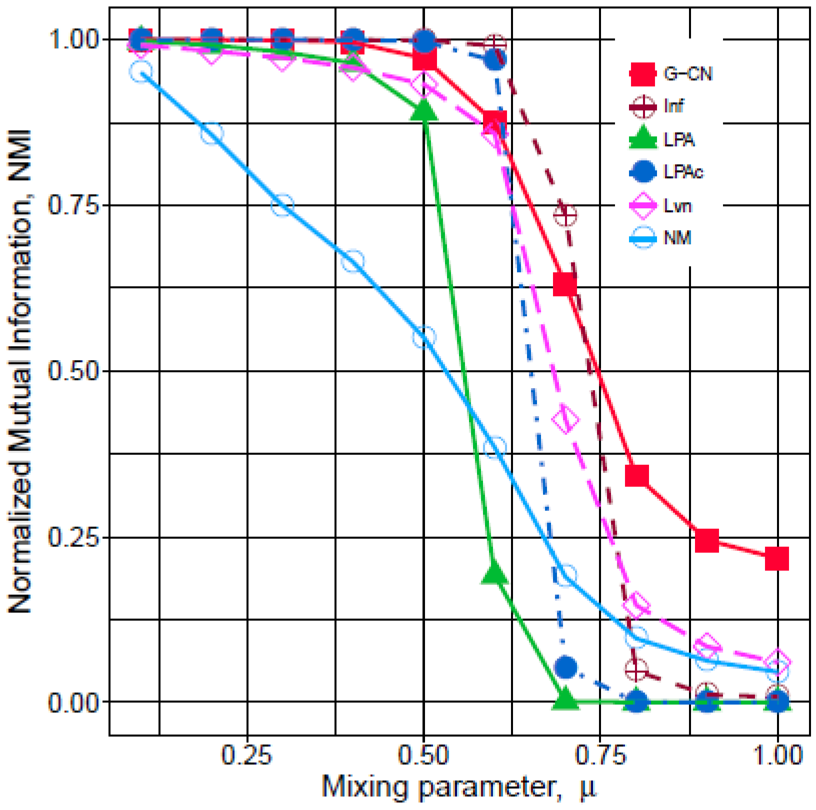


Communities identified

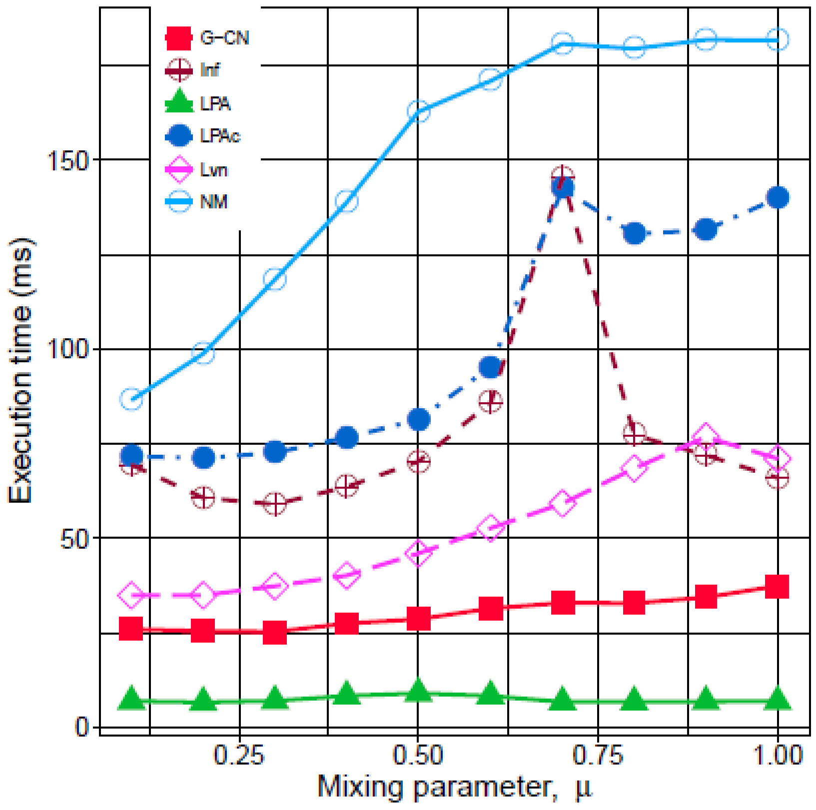
Tasgin, Mursel, and Haluk O. Bingol. "Community detection using boundary nodes in complex networks." *arXiv preprint arXiv:1802.09618* (2018).

Community detection using boundary nodes

Comparison with known algorithms on LFR generated networks of 1000 nodes



(a)



(b)