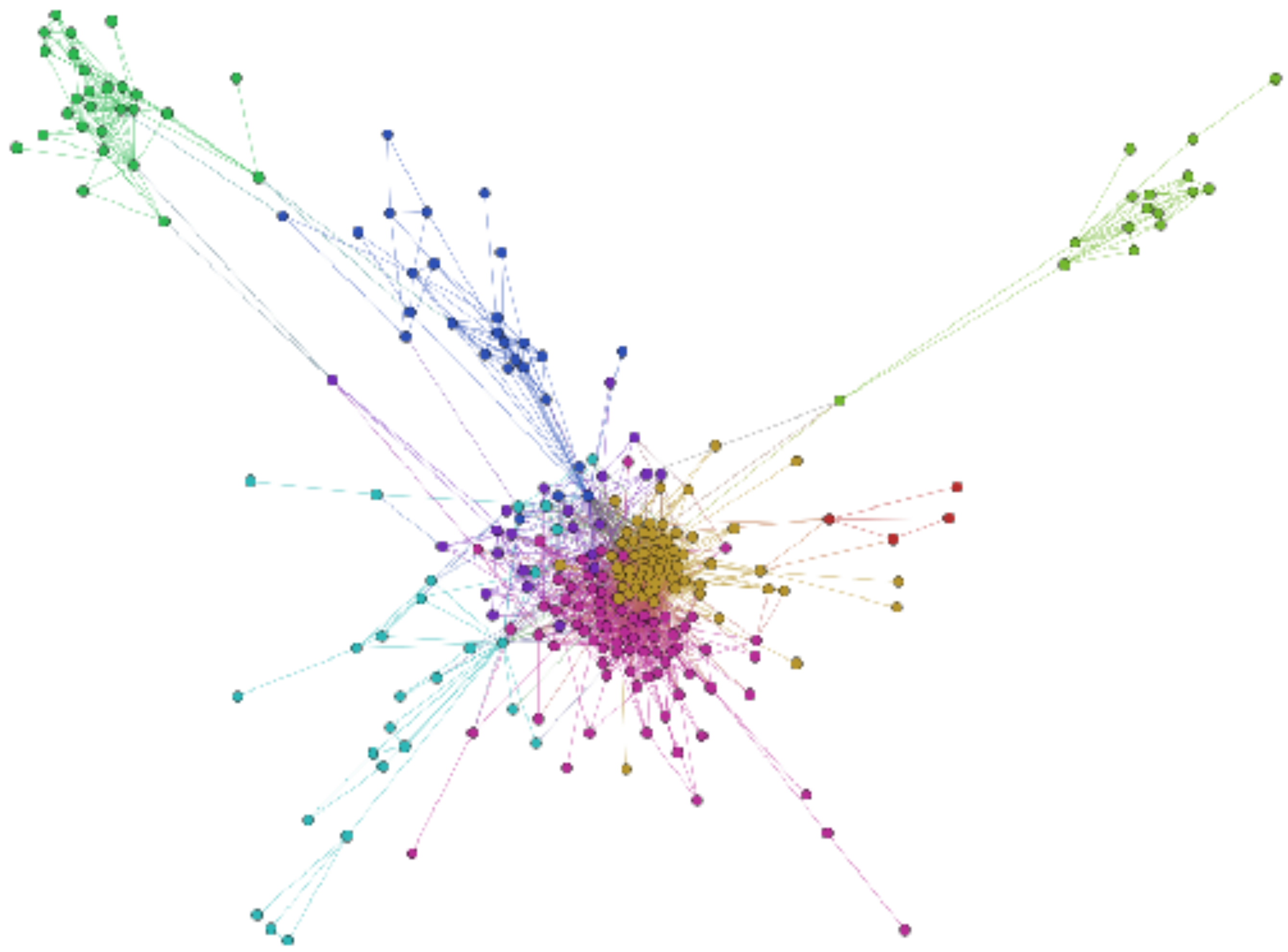
- Why Communities?

- Measuring and Finding Communities

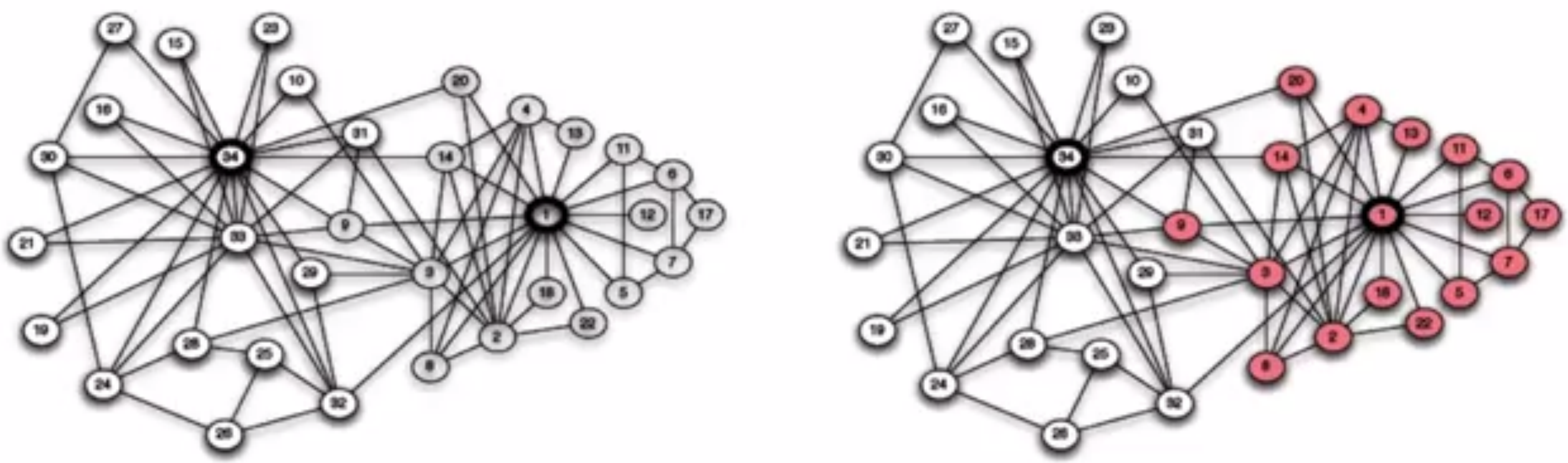
- Validating the detection Methods

Communities Usually Share Features



Facebook Network

Hidden Structures of the Network

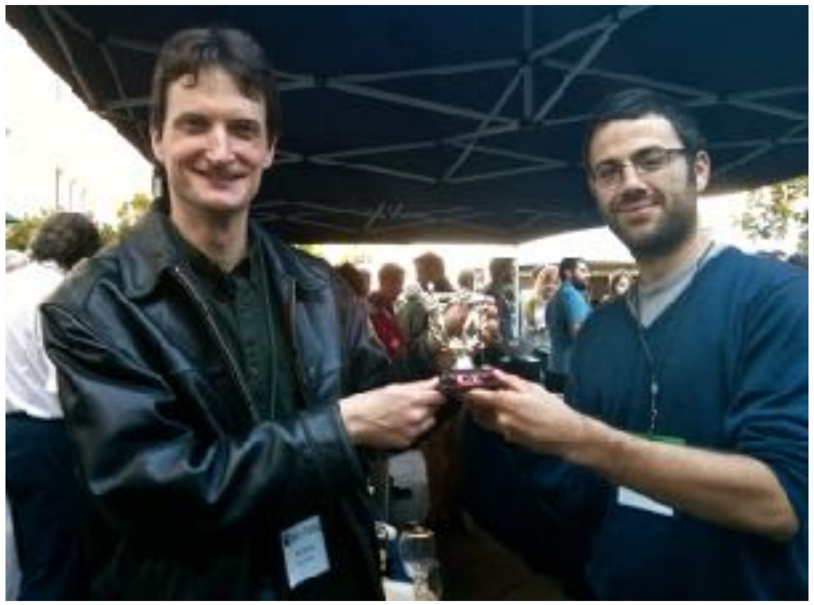


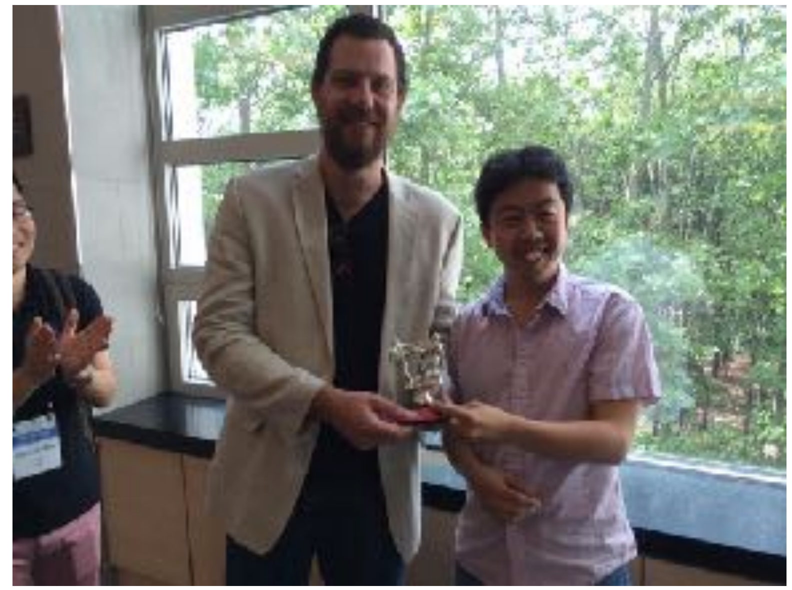
(a) *Karate club network*

(b) *After a split into two clubs*

Zachary Karate Club (W. W. Zachary 1977)

Zachary Karate Club Club

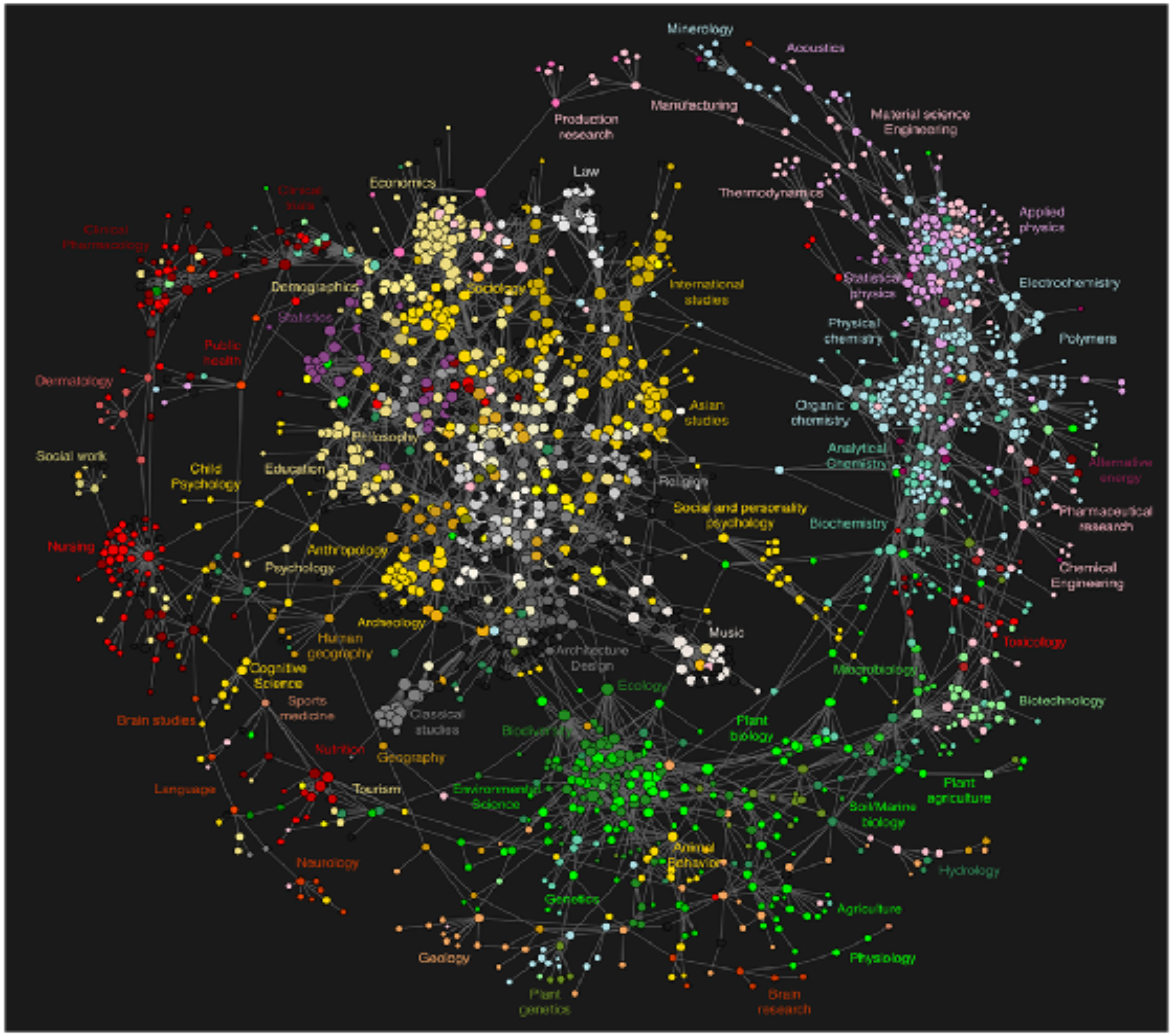




12th Amir Rubin (January 2017)
11th Federico Battiston (September 2016)
10th Giona Casiraghi (July 2016)
9th Filippo Radicchi (May 2016)
8th Qing Ke (September 2015)
7th Manlio De Domenico (July 2015)
6th Tiago Peixoto (June 2015)
5th Mark Newman (June 2014)
4th Marián Boguñá (September 2013)
3rd YY Ahn (July 2013)
2nd Mason Porter (June 2013)
1st Cristopher Moore (May 2013)

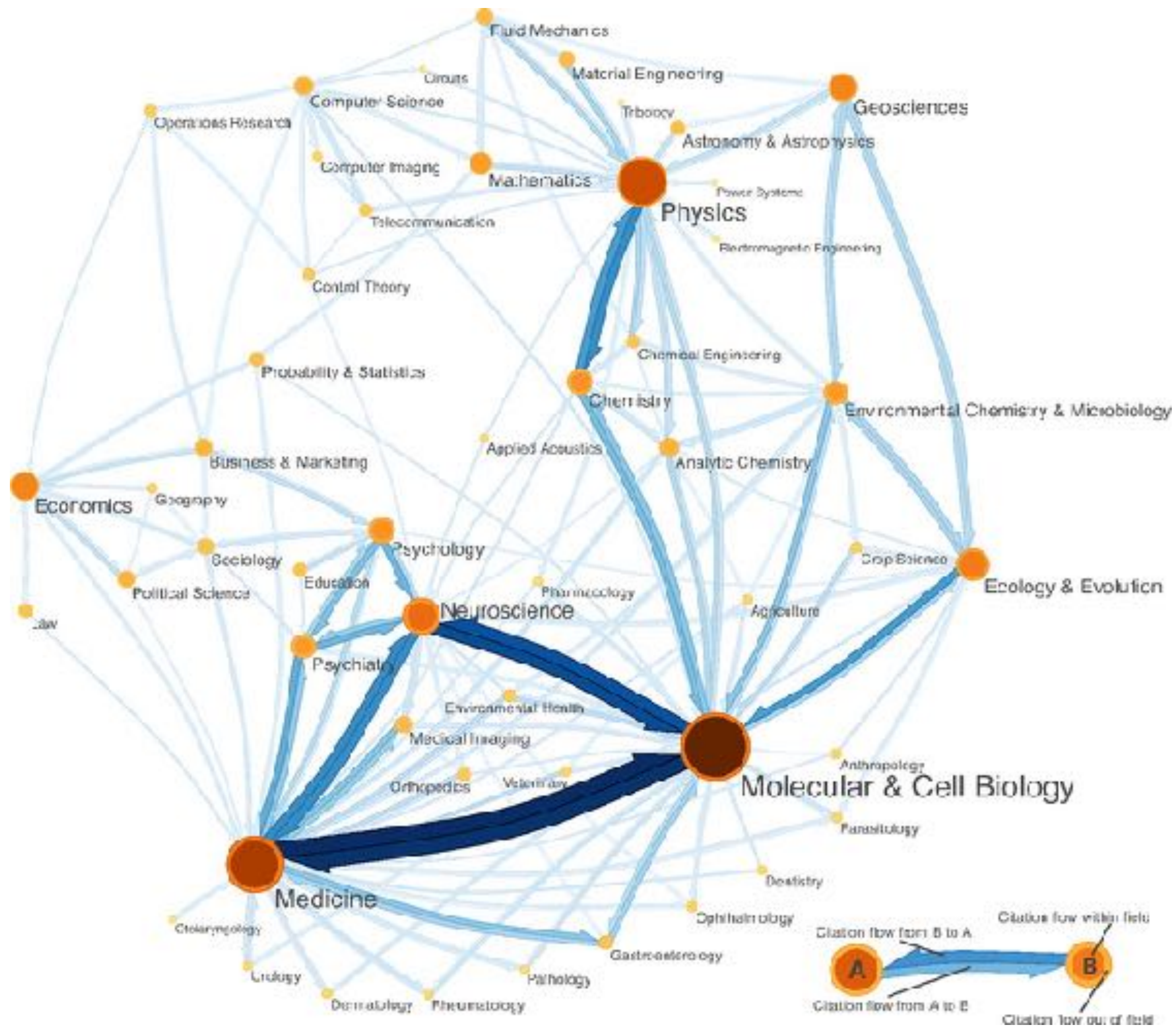
http://networkkarate.tumblr.com/

High Resolution Map of Science



Bollen et al PloS ONE (2009)

Coarse-Graining



Rosvall & Bergstrom, PNAS, 2008

How to find Communities

Author	Ref.	Label	Order
Eckmann & Moses	(Eckmann and Moses, 2002)	EM	$O(m\langle k^2\rangle)$
Zhou & Lipowsky	(Zhou and Lipowsky, 2004)	ZL	$O(n^3)$
Latapy & Pons	(Latapy and Pons, 2005)	LP	$O(n^3)$
Clauset et al.	(Clauset *et al.*, 2004)	NF	$O(n \log^2 n)$
Newman & Girvan	(Newman and Girvan, 2004)	NG	$O(nm^2)$
Girvan & Newman	(Girvan and Newman, 2002)	GN	$O(n^2 m)$
Guimerà et al.	(Guimerà and Amaral, 2005; Guimerà *et al.*, 2004)	SA	parameter dependent
Duch & Arenas	(Duch and Arenas, 2005)	DA	$O(n^2 \log n)$
Fortunato et al.	(Fortunato *et al.*, 2004)	FLM	$O(m^3 n)$
Radicchi et al.	(Radicchi *et al.*, 2004)	RCCLP	$O(m^4/n^2)$
Donetti & Muñoz	(Donetti and Muñoz, 2004, 2005)	DM/DMN	$O(n^3)$
Bagrow & Bollt	(Bagrow and Bollt, 2005)	BB	$O(n^3)$
Capocci et al.	(Capocci *et al.*, 2005)	CSCC	$O(n^2)$
Wu & Huberman	(Wu and Huberman, 2004)	WH	$O(n+m)$
Palla et al.	(Palla *et al.*, 2005)	PK	$O(\exp(n))$
Reichardt & Bornholdt	(Reichardt and Bornholdt, 2004)	RB	parameter dependent

TABLE I List of the algorithms used in the comparative analysis of Danon et al. (Danon *et al.*, 2005). The first column indicates the names of the algorithm designers, the second the original reference of the work, the third the symbol used to indicate the algorithm and the last the computational complexity of the technique. Adapted from Ref. (Danon *et al.*, 2005).

Author	Ref.	Label	Order
Girvan & Newman	(Girvan and Newman, 2002; Newman and Girvan, 2004)	GN	$O(nm^2)$
Clauset et al.	(Clauset *et al.*, 2004)	Clauset et al.	$O(n \log^2 n)$
Blondel et al.	(Blondel *et al.*, 2008)	Blondel et al.	$O(m)$
Guimerà et al.	(Guimerà and Amaral, 2005; Guimerà *et al.*, 2004)	Sim. Ann.	parameter dependent
Radicchi et al.	(Radicchi *et al.*, 2004)	Radicchi et al.	$O(m^4/n^2)$
Palla et al.	(Palla *et al.*, 2005)	Cfinder	$O(\exp(n))$
Van Dongen	(Dongen, 2000a)	MCL	$O(nk^2)$, $k < n$ parameter
Rosvall & Bergstrom	(Rosvall and Bergstrom, 2007)	Infomod	parameter dependent
Rosvall & Bergstrom	(Rosvall and Bergstrom, 2008)	Infomap	$O(m)$
Donetti & Muñoz	(Donetti and Muñoz, 2004, 2005)	DM	$O(n^3)$
Newman & Leicht	(Newman and Leicht, 2007)	EM	parameter dependent
Ronhovde & Nussinov	(Ronhovde and Nussinov, 2009)	RN	$O(m^\beta \log n)$, $\beta \sim 1.3$

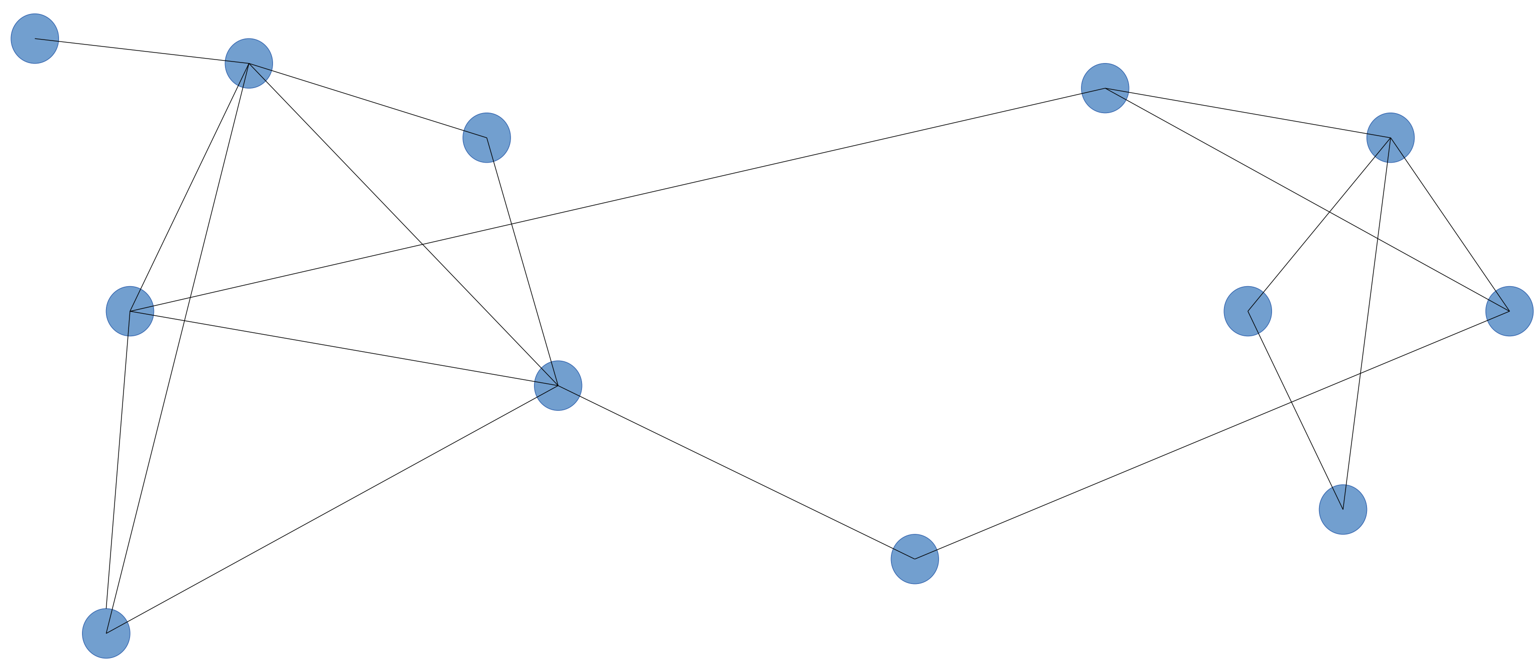
TABLE II List of the algorithms used in the comparative analysis of Lancichinetti and Fortunato (Lancichinetti and Fortunato, 2009). The first column indicates the names of the algorithm designers, the second the original reference of the work, the third the symbol used to indicate the algorithm and the last the computational complexity of the technique.

Santo Fortunato, Community detection in graphs, 2010

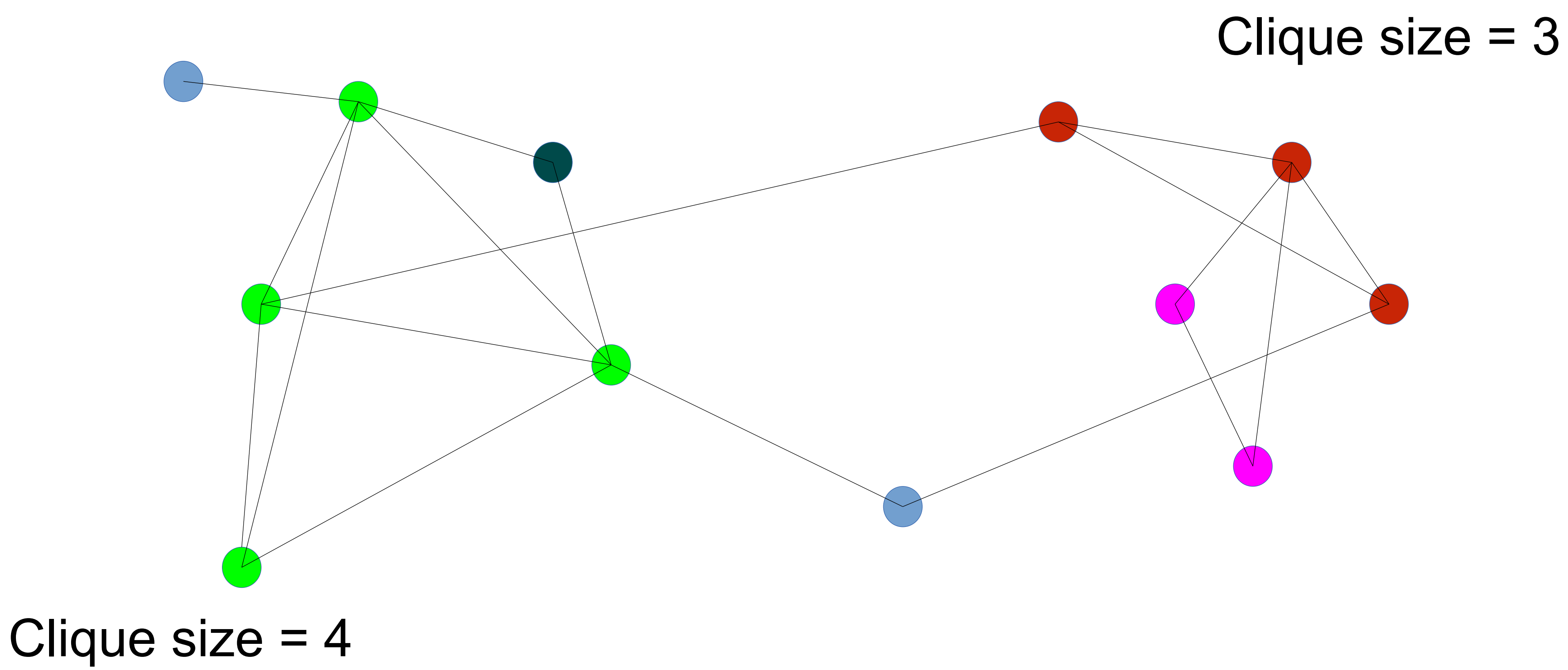
How to find Communities: Basic Concepts

- Full Connectivity
- Partial Connectivity
- Closeness
- High intra-group/inter-group connectivity

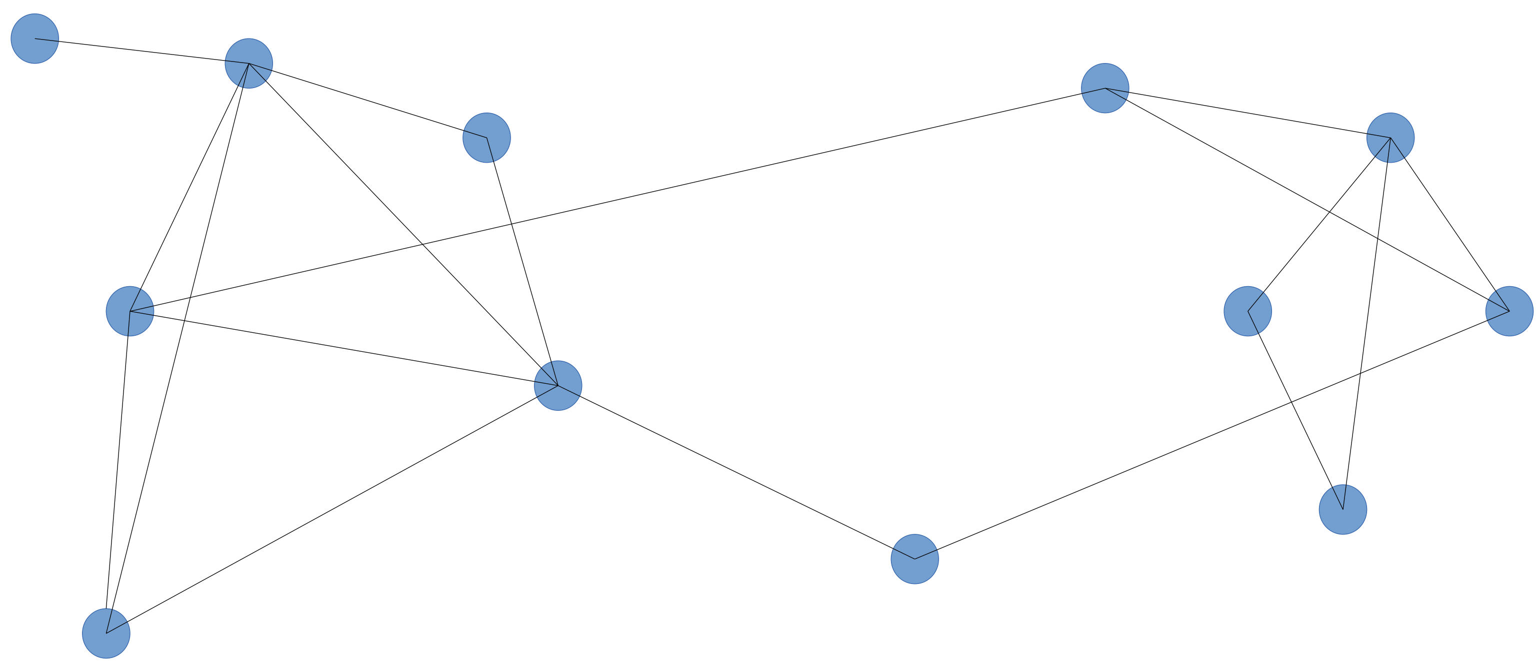
Cliques: fully connected sub-graphs



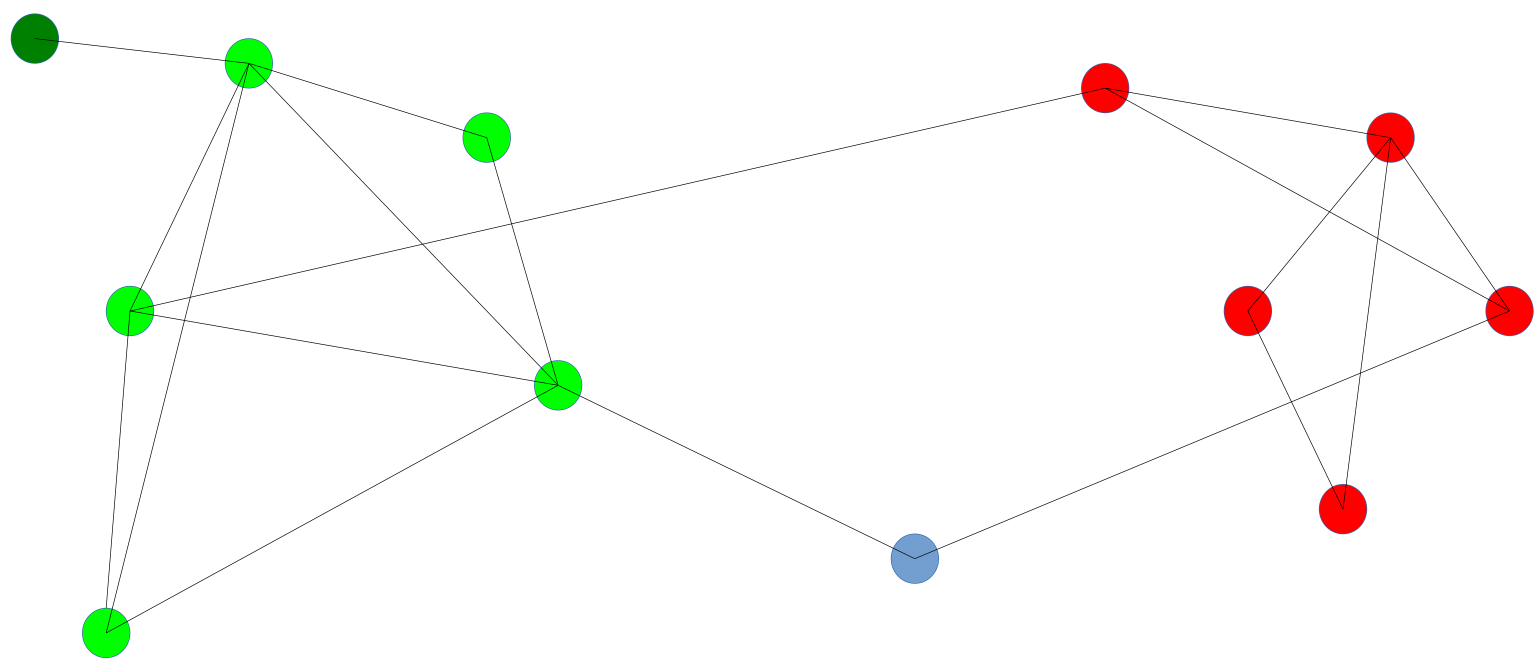
Cliques: fully connected sub-graphs



k-cores: sub-graphs in which each node is connected to k others



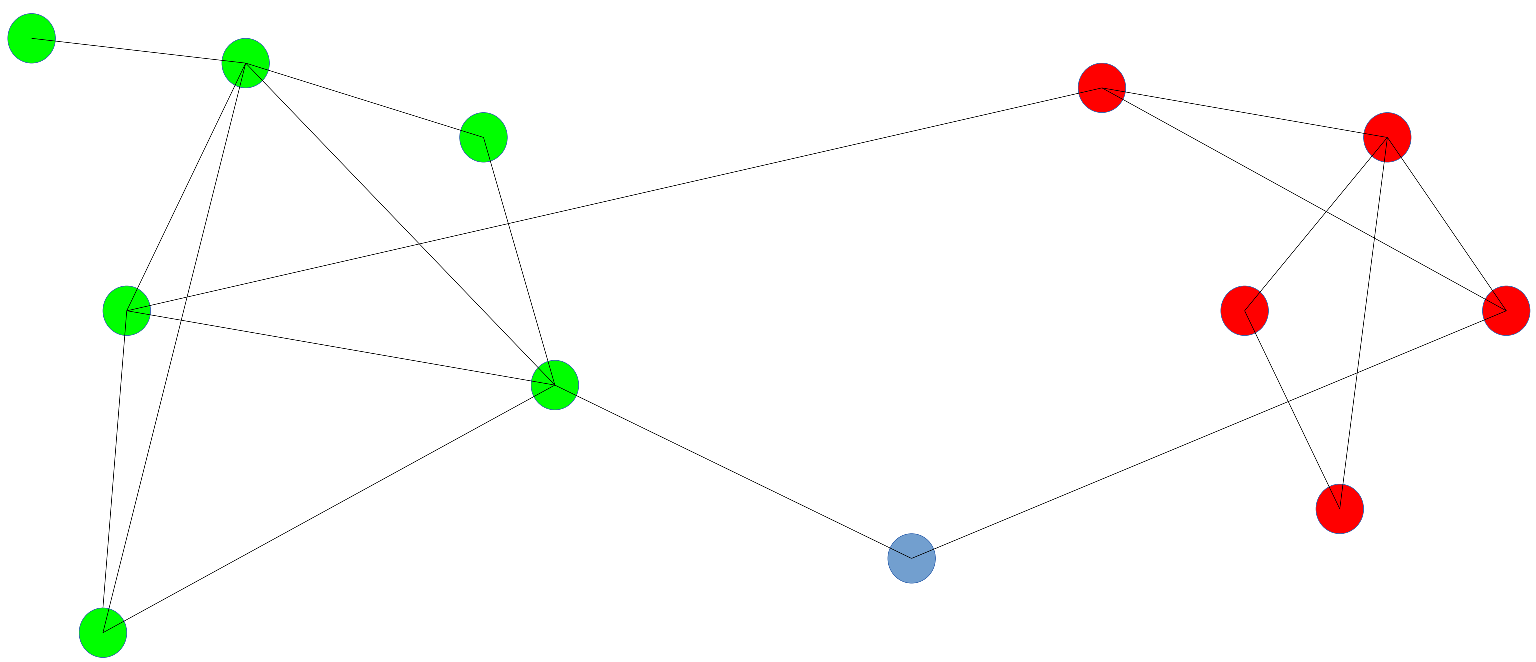
k-cores: sub-graphs in which each node is connected to k others



2-cores

n-cliques: sub-graphs with diameter of n

(each two nodes in sub-graph could be reached by n links)

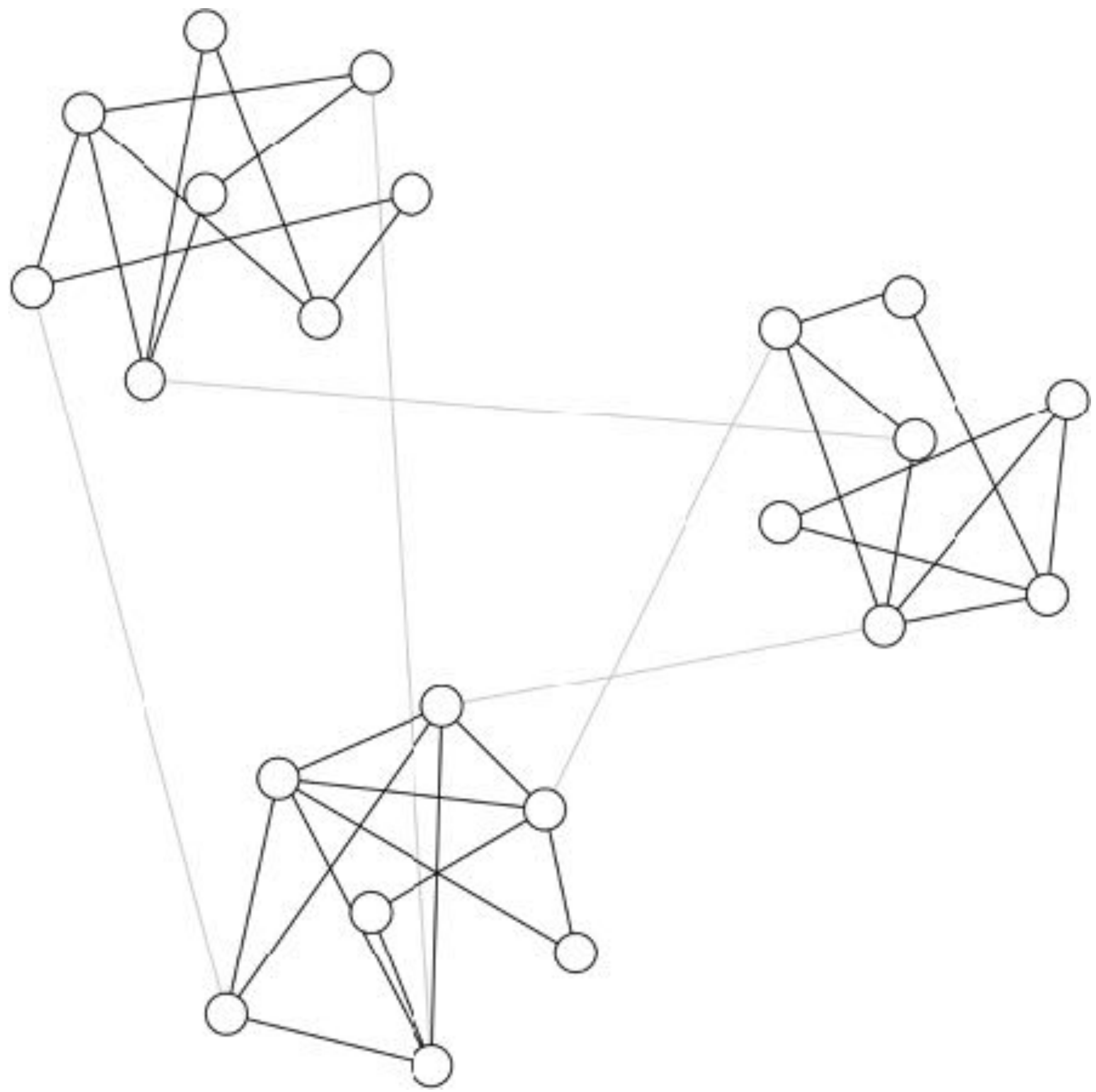


2-cliques

More Generalisations...

- p-cliques: sub-graphs in which each node is connected to a proportion p of all other nodes
- Using the weights of links
- Using the direction of links
- Using “structural similarity” of nodes

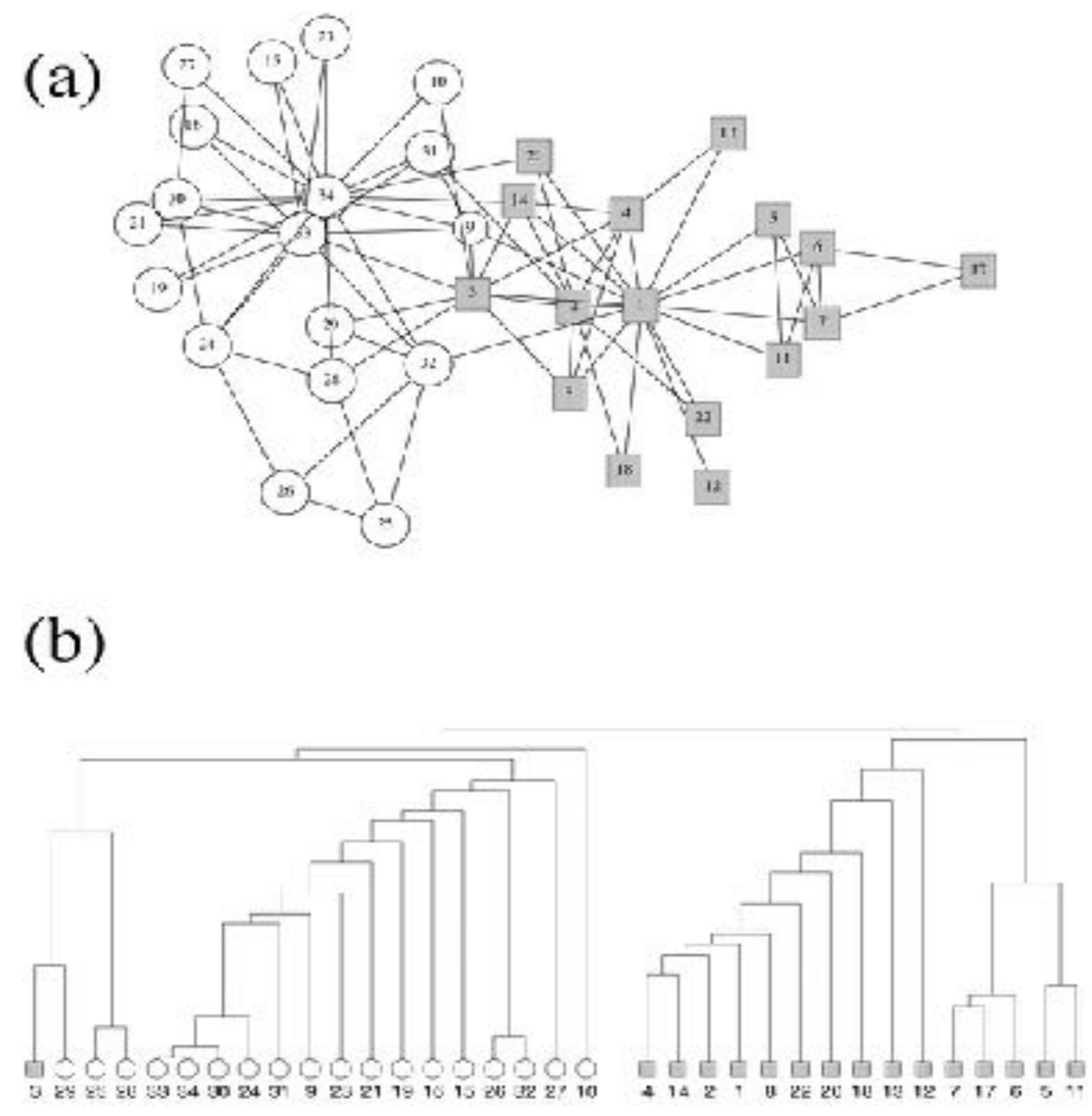
Betweenness Clustering



Girvan and Newman, PNAS, 2002.

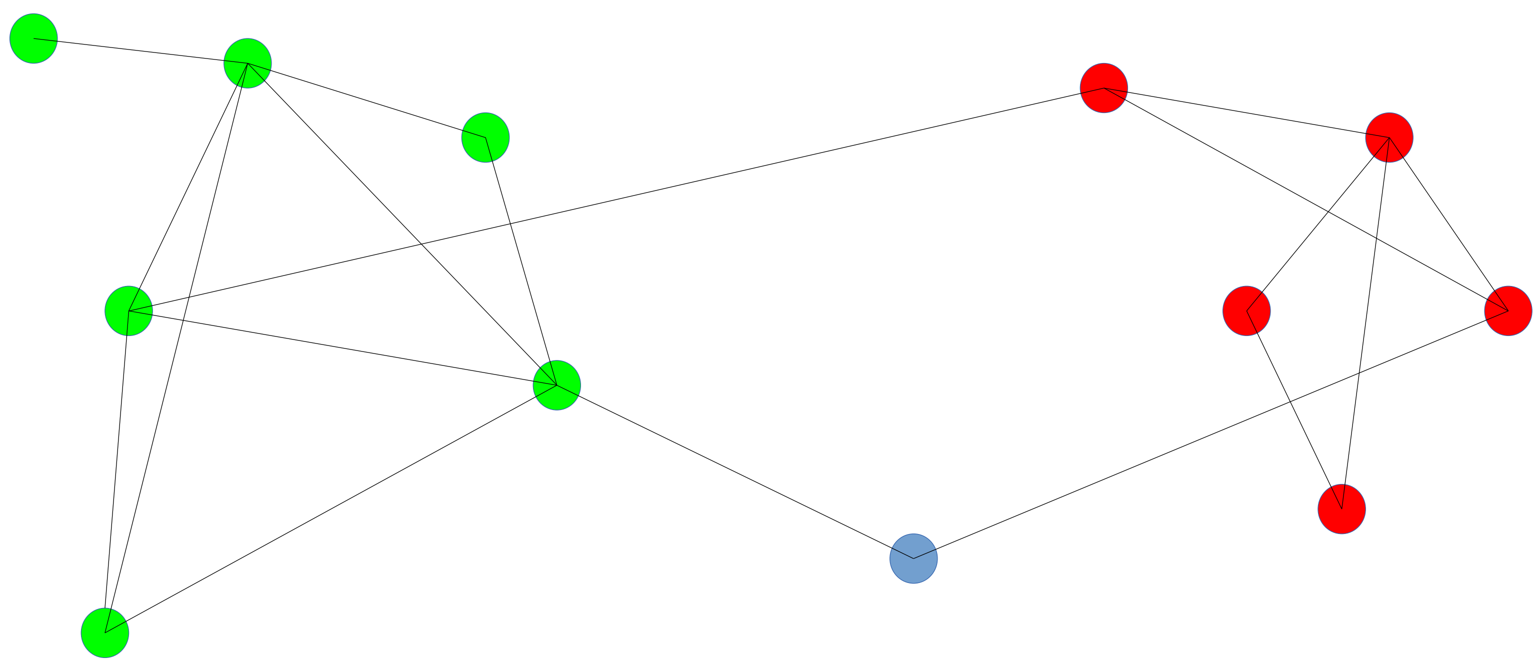
Betweenness Clustering

1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweennesses for all edges affected by the removal.
4. Repeat from step 2 until no edges remain.



Modularity

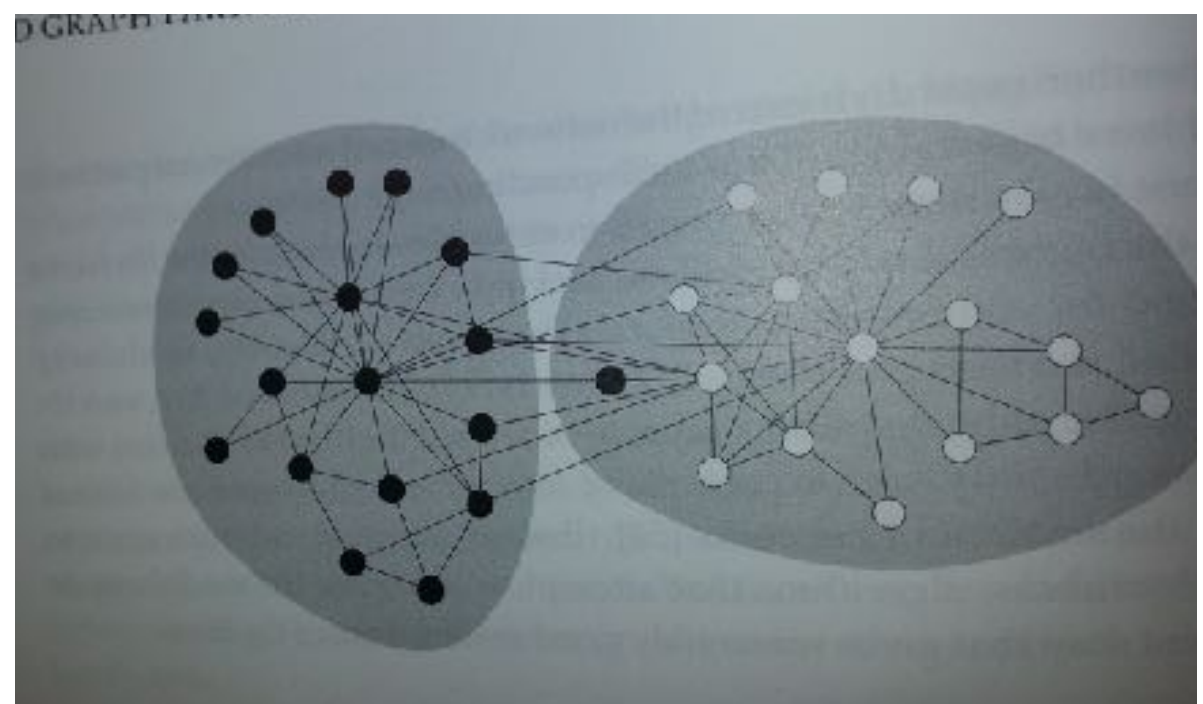
$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\delta(c_v, c_w).$$



Clauset, Newman, and Moore, Phys. Rev. E, 2004.

Modularity Maximisation Algorithm

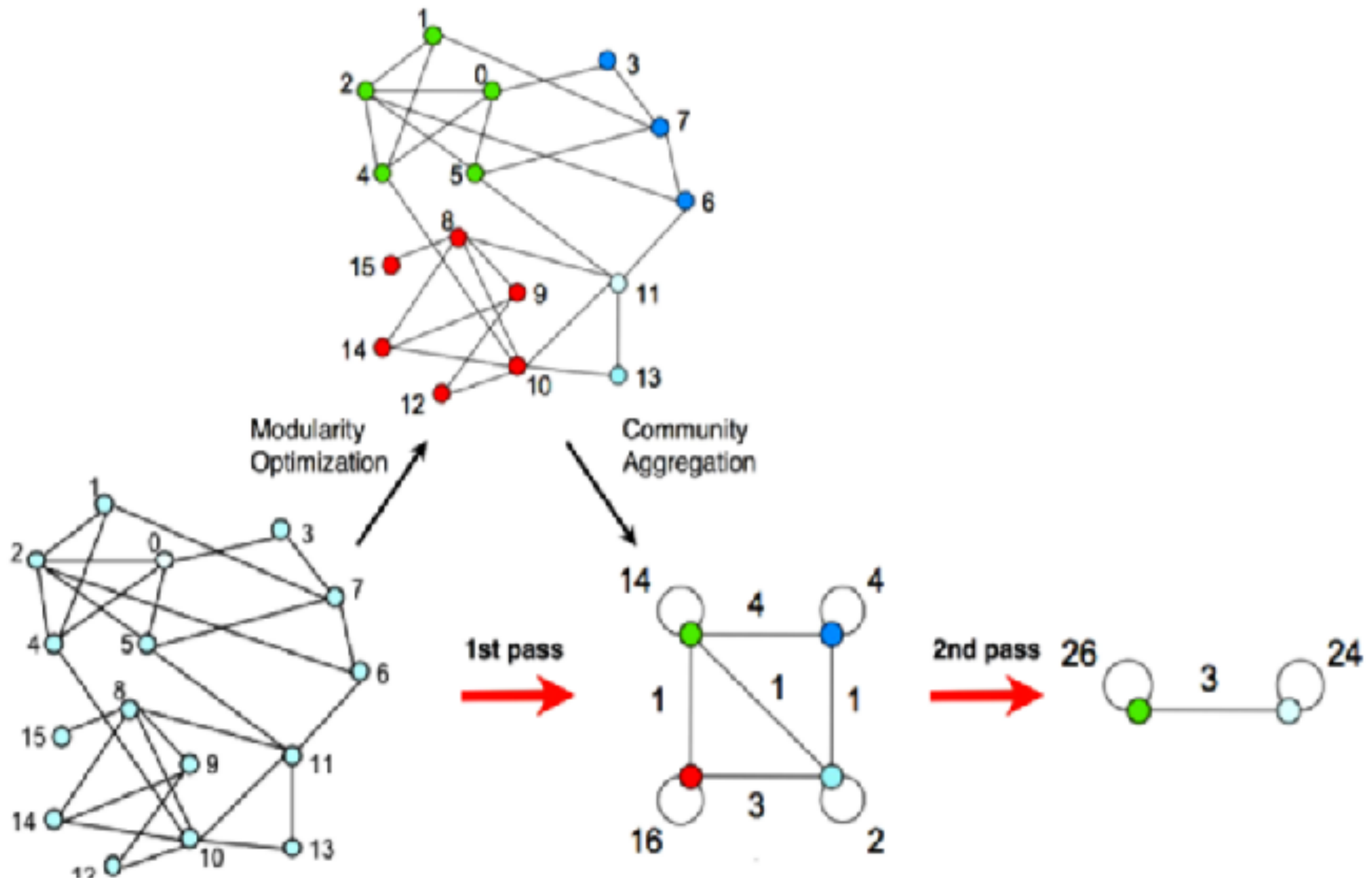
1) Divide the network in two
2) Check all the possible moves
3) Perform the most efficient one
4) Go to 2, until all the vertices are moved
5) Look at the history of Q and find the Max.



Zachary Karate Club (Newman 2010)

More than two communities

- Iterating the bisection algorithm
- Greedy Algo



Blondel et al, J Stat Mech, 2008, **Louvain Algorithm** (4849)

Connecting Dream Networks Across Cultures

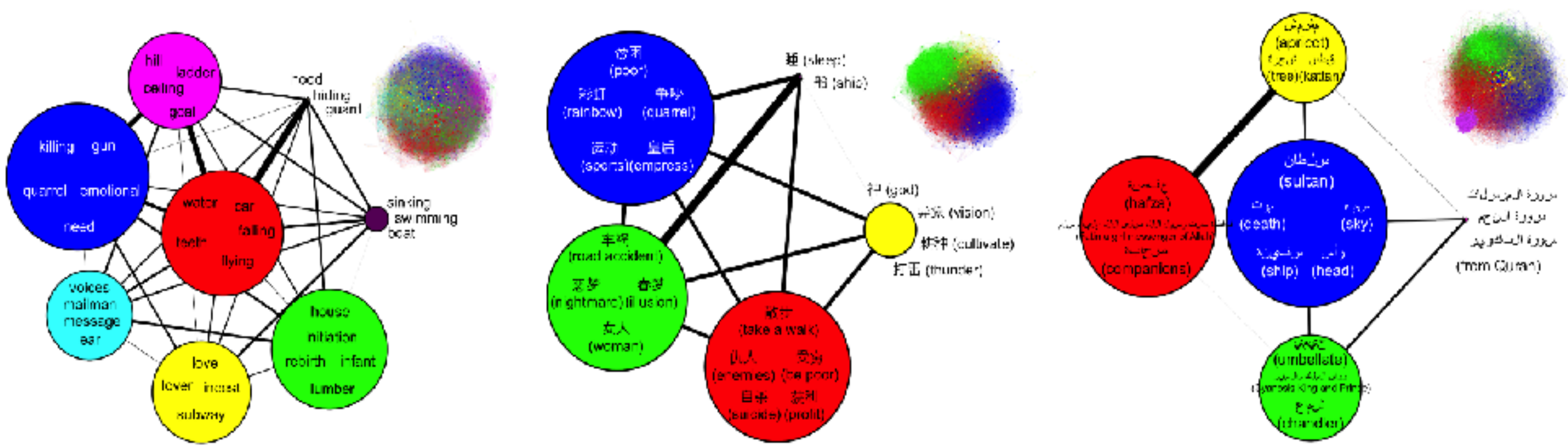
arXiv:1402.2297v1 [cs.SI] 10 Feb 2014

Onur Varol* Filippo Menczer

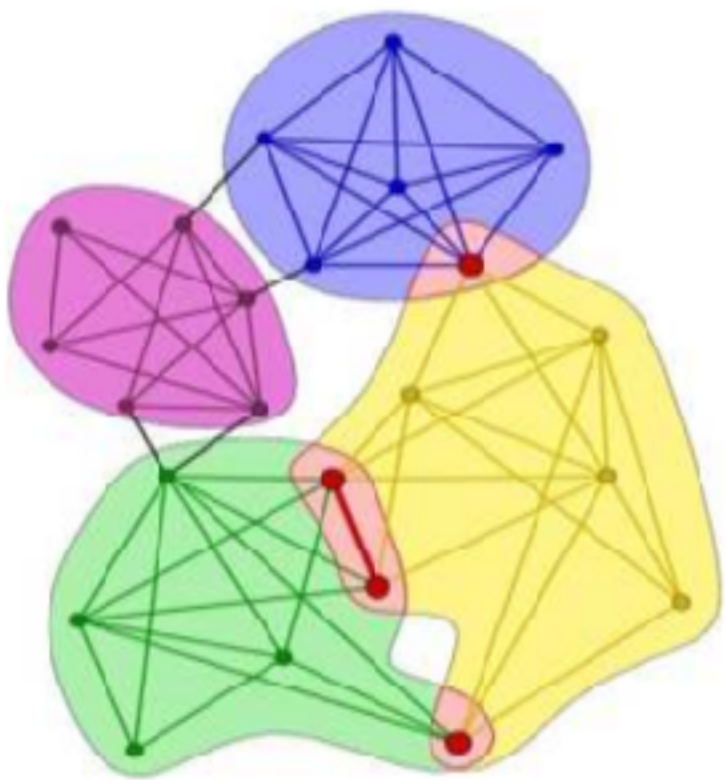
Center for Complex Networks and Systems Research
School of Informatics and Computing, Indiana University, Bloomington, USA

Rain: To see and hear rain falling symbolises forgiveness and grace.

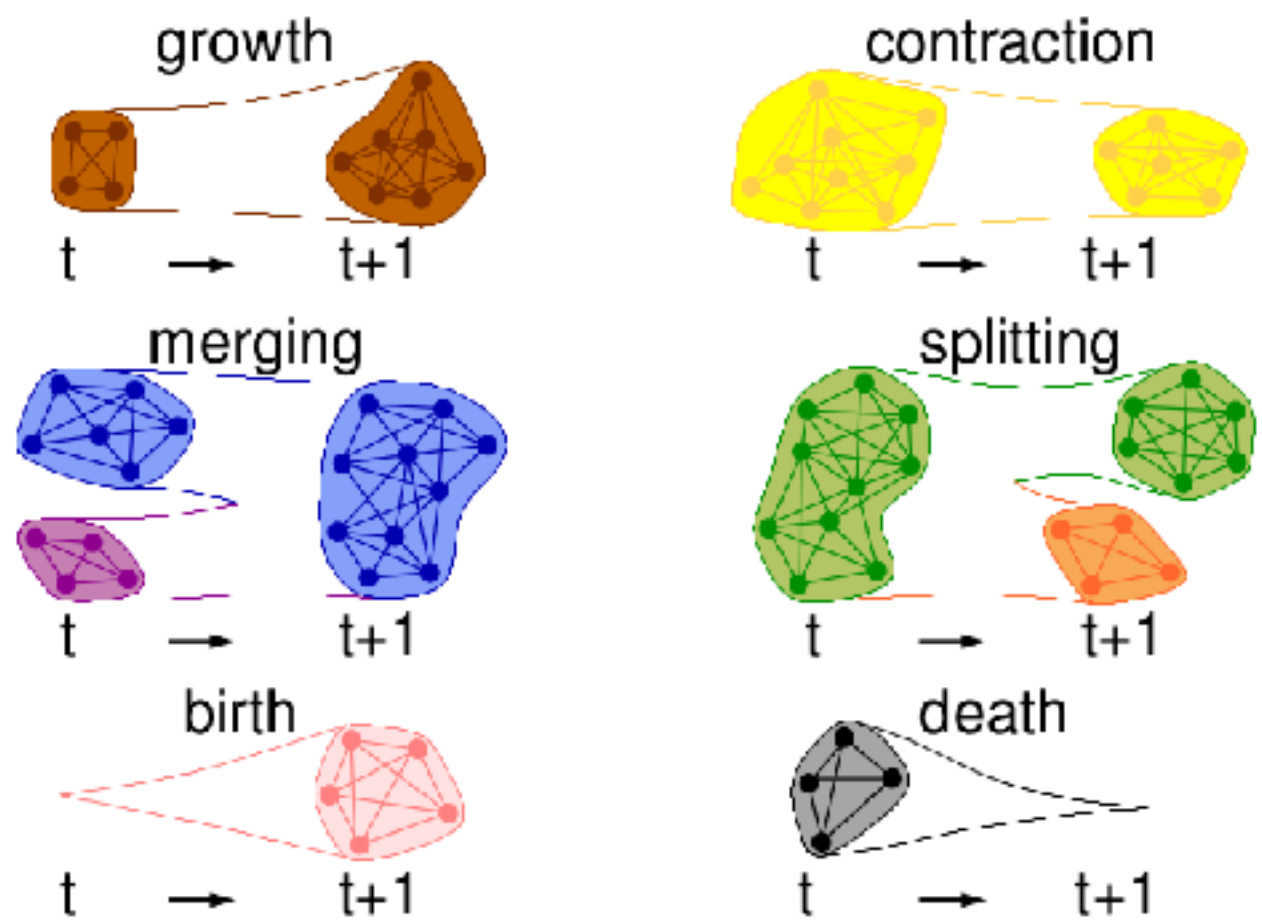
To Make Up: The dream may be a metaphor that you need to "make up" with someone. It is time to forgive and forget.



Overlapping and Evolving Communities



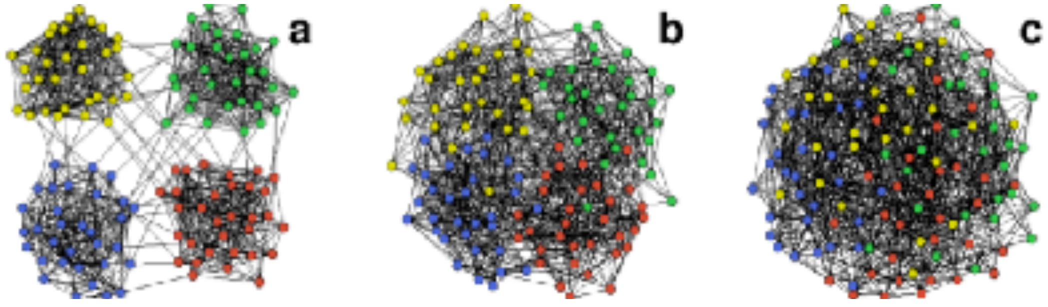
Clique Percolation
Palla et al, Nature, 2005.



Palla et al, Nature, 2007.

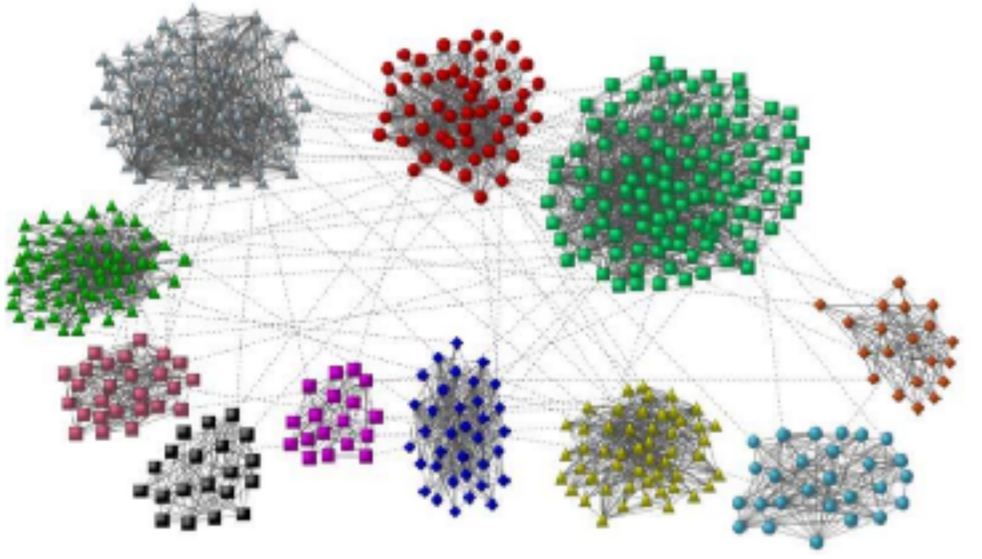
Algorithm Validation

- Test against external information
- Run on artificially synthesised networks
  - Erdos-Renyi-based Benchmark:
  - *planted l-partition model*
  - l groups, g vertices each
  - $<k> = p_{in}(g-1) + p_{out}\,g(l-1)$



Girvan and Newman, 2002

  - LFR Benchmark:
  - fat-tailed degree and cluster size
  - distributions



  - Lancichinetti et al., 2008