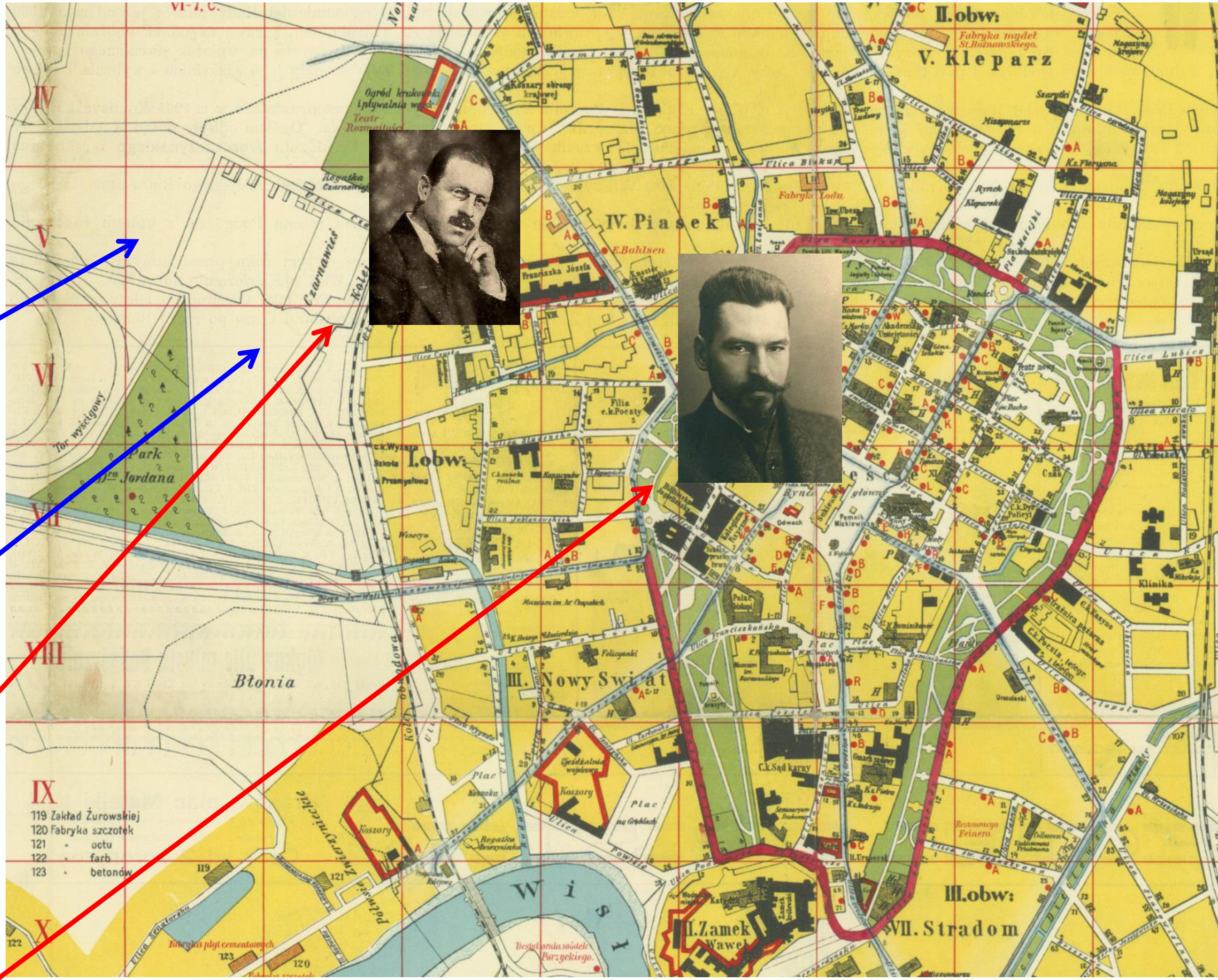
Emerging communities in networks – a flow of ties

Małgorzata J. Krawczyk, Przemysław Gawroński, Krzysztof Kułakowski

Complex Systems Group
Faculty of Physics and Applied Computer Science
AGH University of Science and Technology

27th Marian Smoluchowski Symposium
on Statistical Physics
Fundamentals, soft matter and biocomplexity
September 22-26, Zakopane, Poland



FIS AGH

FAIS UJ
till 2014

AG, 1919/35

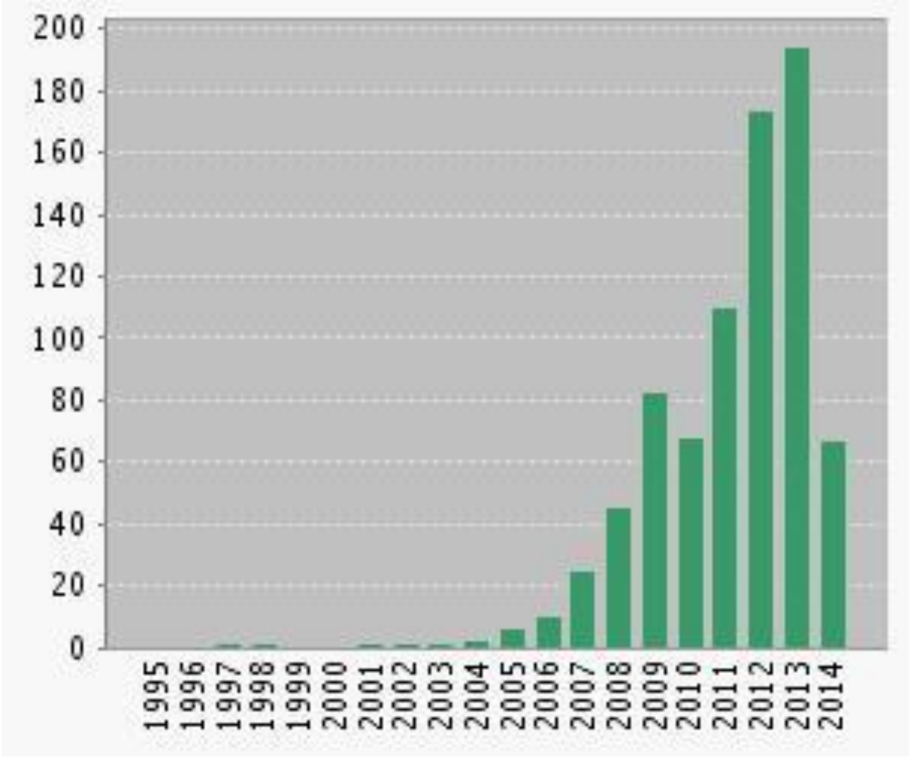
Collegium Physicum
1912-64

Kraków 1907

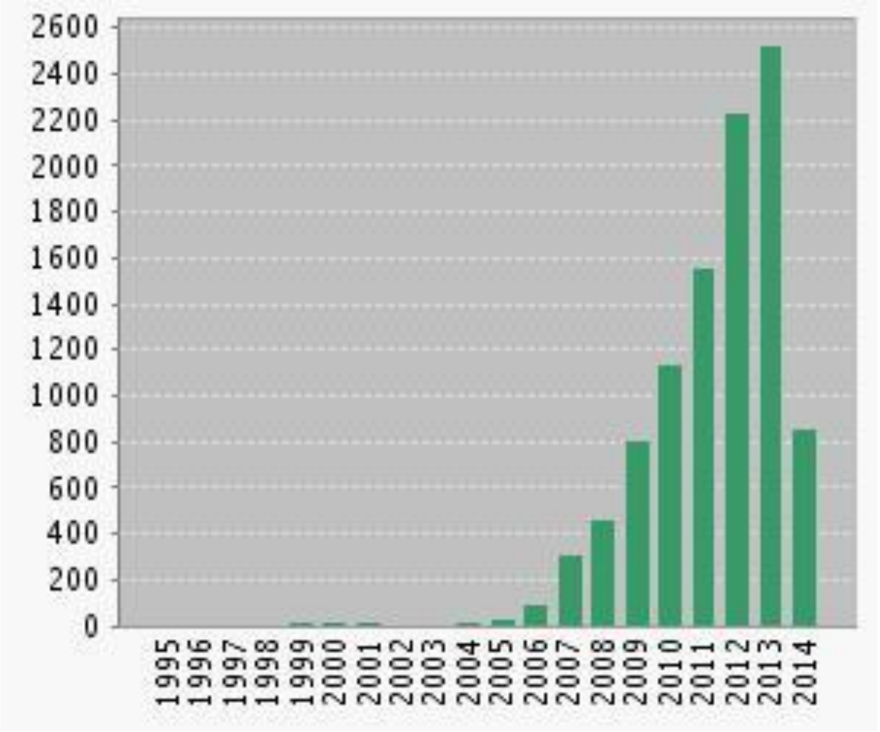
Communities (...) occur in many networked systems from biology, computer science, engineering, economics, politics, etc. In protein-protein interaction networks, communities are likely to group proteins having the same specific function within the cell (...), in metabolic networks they may be related to functional modules such as cycles and pathways in food webs they may identify compartments and so on.

[S. Fortunato, Community detection in graphs, Phys. Reports 486, 75-174 (2010)]

„community detection”



Published Items in Each Year



Citations in Each Year

[Web of Science, 16.06.2014]

Are there any communities ?

...there must be more edges "inside" the community than edges linking vertices of the community with the rest of the graph.

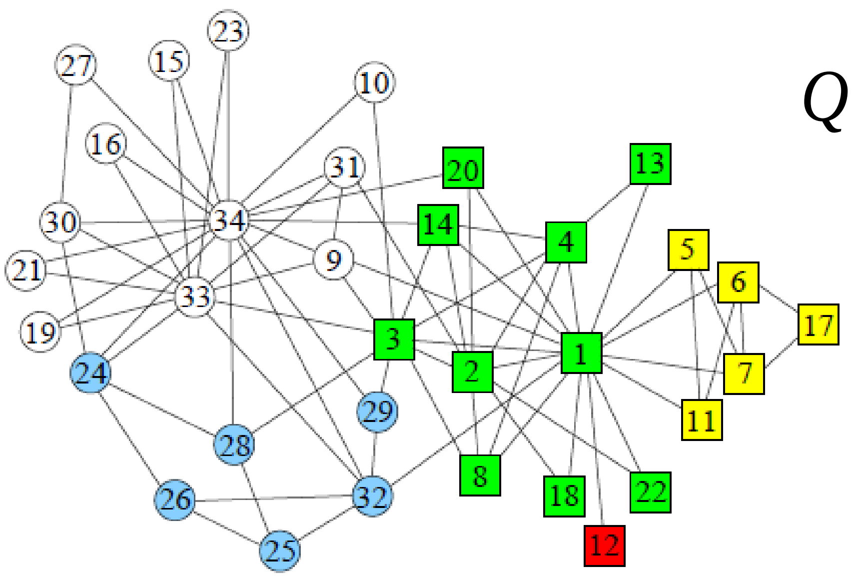


FIG. 4: Splitting of the Zachary club network. Squares and circles indicate the two communities observed by Zachary, colors denote the further subdivision found by our algorithm.

$$Q = \frac{1}{w}\sum_{ij}\left(A(i,j) - \frac{k_i k_j}{w}\right)\delta(c_i, c_j)$$

where

$$k_i = \sum_j A(i,j)$$

$$w = \sum_{ij} A(i,j)$$

[S. Fortunato, Community detection in graphs, Physics Reports 486 (2010) 75]
[L. Donetti, M. A. Muñoz, Detecting network communities: a new systematic and efficient algorithm, J. Stat. Mech. 2004, P10012.
[M. E. J. Newman, Analysis of weighted networks, Phys. Rev. E 70, 056131 (2004)]

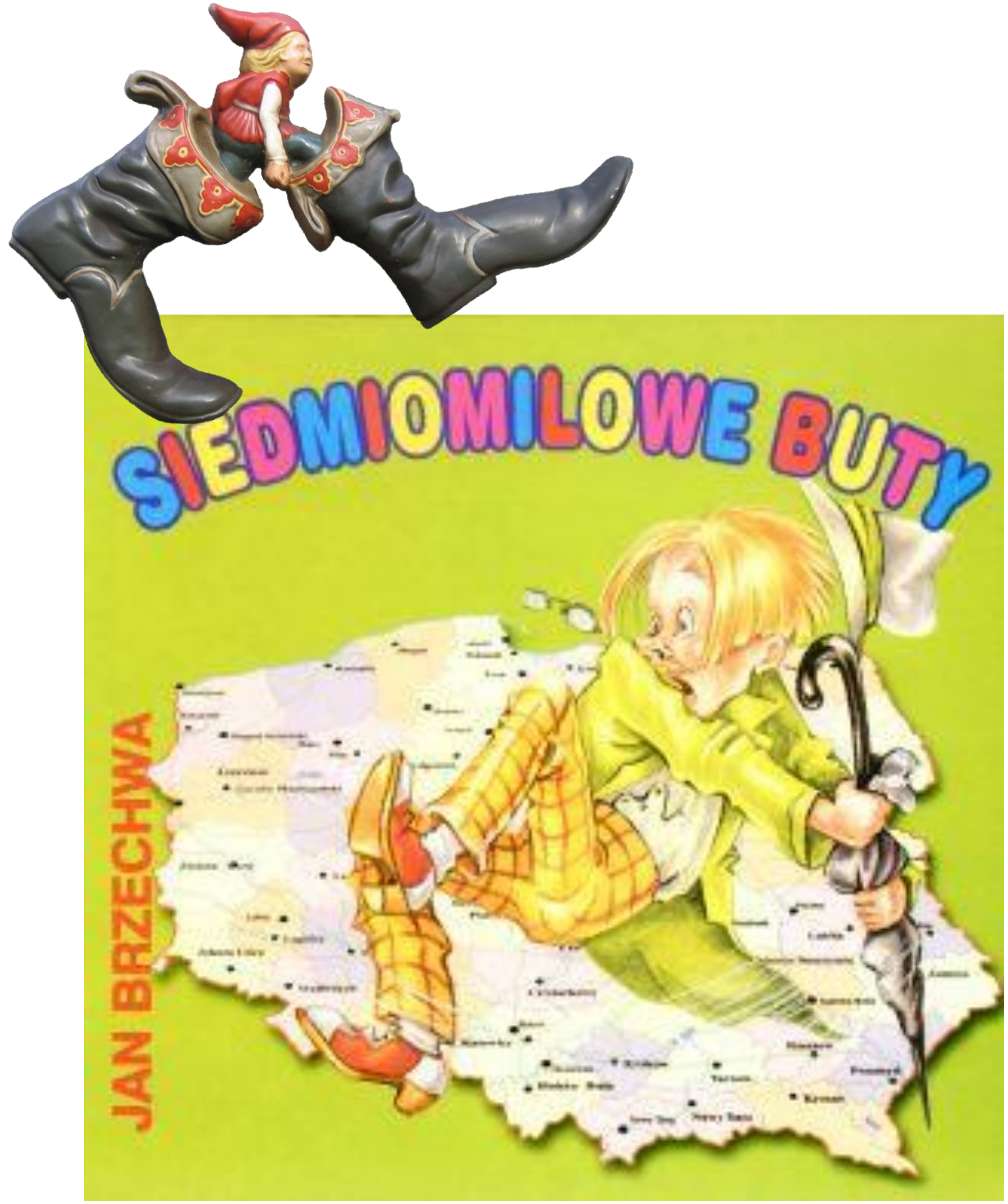
Outline

1. The goal :
   a spontaneous flow
   towards solution

2. Examples:

   - Heider balance

   - communities
   in the Sierpinski triangle
   with a bit of noise



[alejka.pl/siedmiomilowe_buty.html]
[www.giantsteps-project.eu/]

Cognitive dissonance
is a state of conflict in the mind, whereby you have two opposing views at the same time

Leon Festinger, 1919-1989

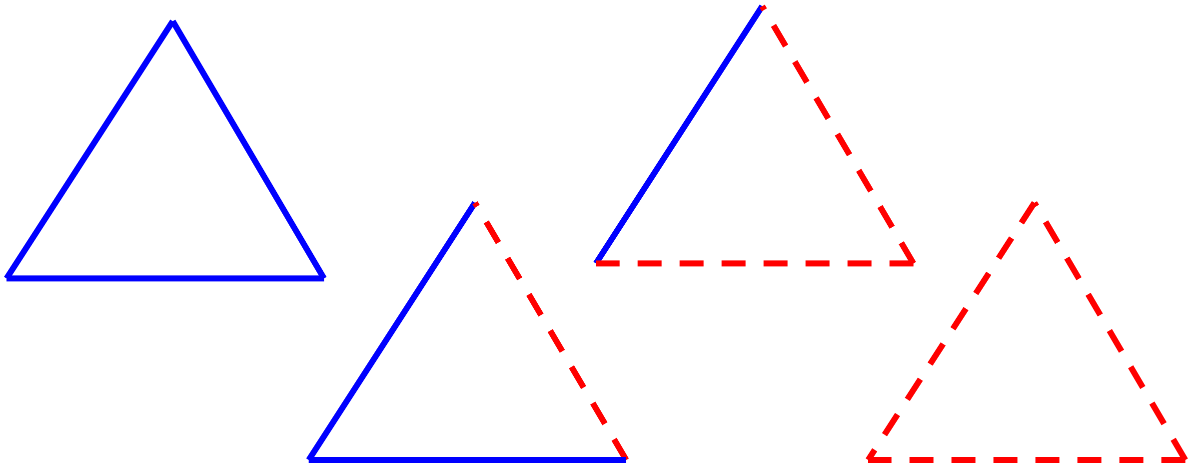
An example:

Sarah just bought a new car. Admittedly, she spent much more money than she should have and feels regretful and actually a little embarrassed as well (*buyer's remorse*).

Rather than continue feeling these undesirable emotions, she decides that the car is less likely to break down than her older one, and will actually save her loads of cash in the long run.

[L. Festinger, A Theory of Cognitive Dissonance, Stanford UP 1957]
[psychohawks.wordpress.com/]
[lasombradecharvaka.blogspot.com/]

Heider balance

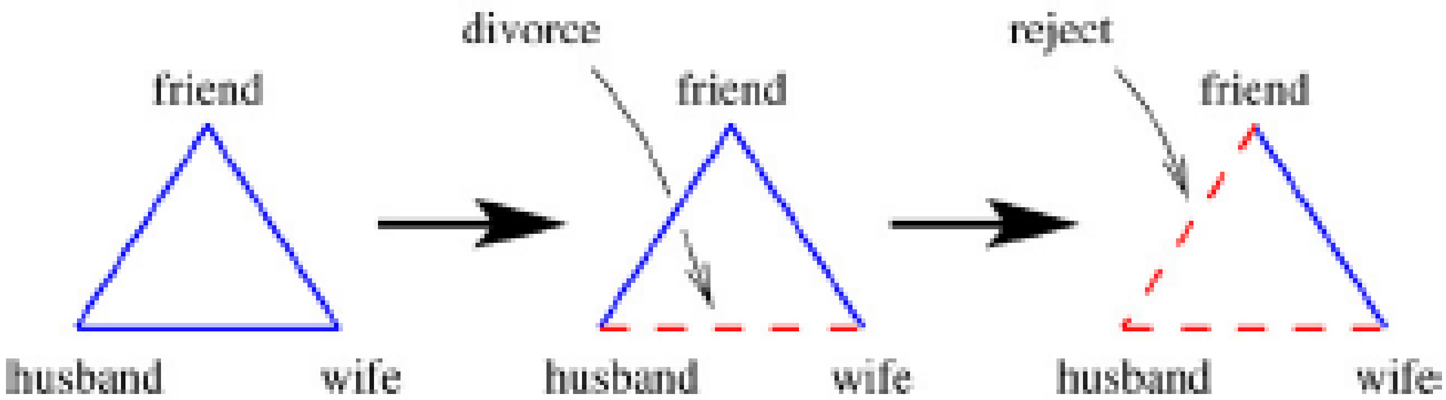


Fritz Heider, 1896-1988

A friend of my friend is my friend
A friend of my enemy is my enemy
An enemy of my friend is my enemy
An enemy of my enemy is my friend

*[F. Heider, Attitudes and Cognitive Organization, The Journal of Psychology **21** (1946) 107]*
[E. Aronson, V. Cope, My enemy's enemy is my friend, J. of Personality & Soc. Psych. 8 (1968) 8]
[compendium.open.ac.uk/]

How the balance can be restored



[T. Antal, P.L.Krapivsky, S. Redner, Social balance of networks: the dynamics of friendship and enmity, Physica D 224 (2006) 130]

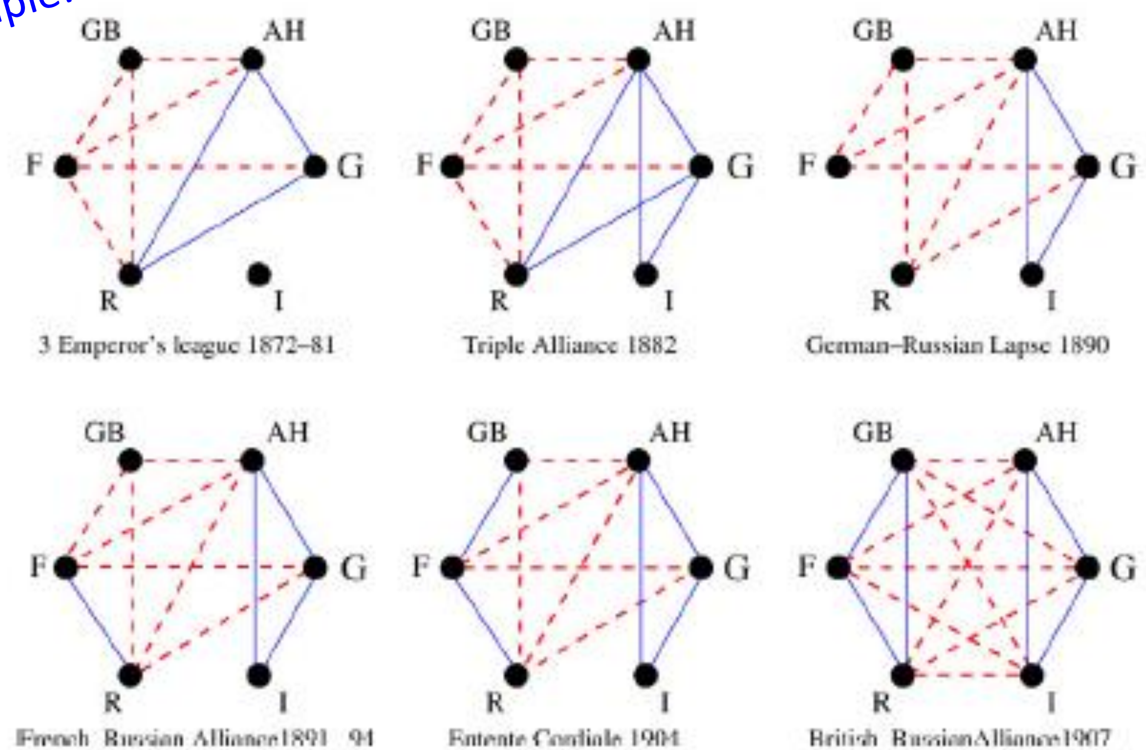
Cartwright - Harary theory

…on a complete graph, balanced societies are remarkably simple:

either all individuals are mutual friends (“*utopia*”),

or the network segregates into two mutually antagonistic but internally friendly cliques - a “bipolar” state.

An example:



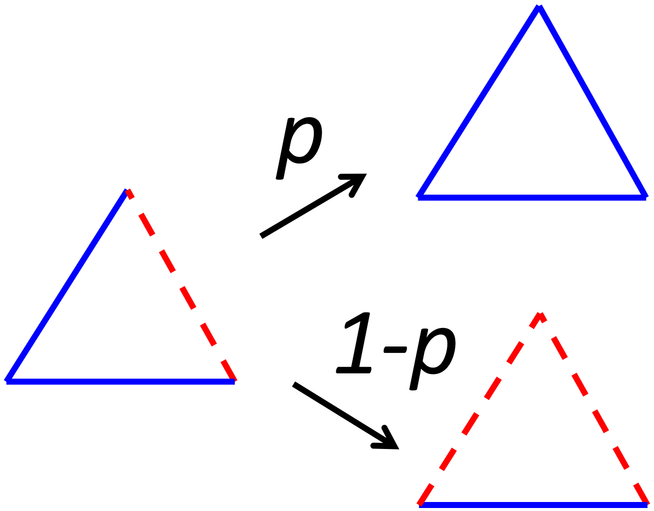
G= Germany
AH - Austria -Hungary
I = Italy
R = Russia
F = France
GB = GB

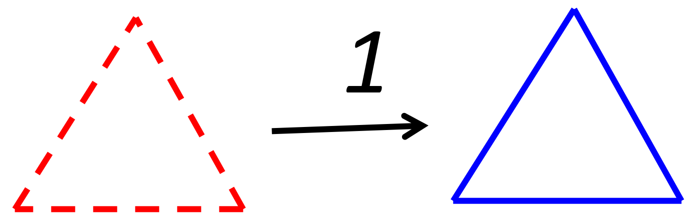
How the major relationship changed from 1872 to 1907 (Antal *et al*, 2006)

[T. Antal, P.L.Krapivsky, S. Redner, Physica D 224 (2006) 130]
[D. Cartwright, F. Harary, Psychol. Rev. 63 (1956) 277]

Discrete algorithms for complete graphs:

1. Local Triad Dynamics (LTD) : operates at randomly selected triads





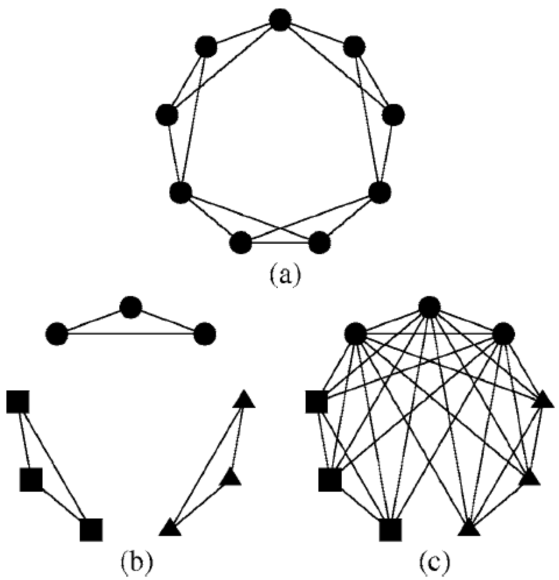
2. Constrained Triad Dynamics (CTD) : the whole number # of imbalanced triads is controlled
   - select a link randomly and check how the change of its sign does influence #
   - if # decreased, OK
   - if # increased, withdraw the change
   - if # is the same, change the link with probability 1/2

[T. Antal, P.L. Krapivsky, S. Redner, Dynamics of social balance on networks, PRE 72 (2005) 036121]

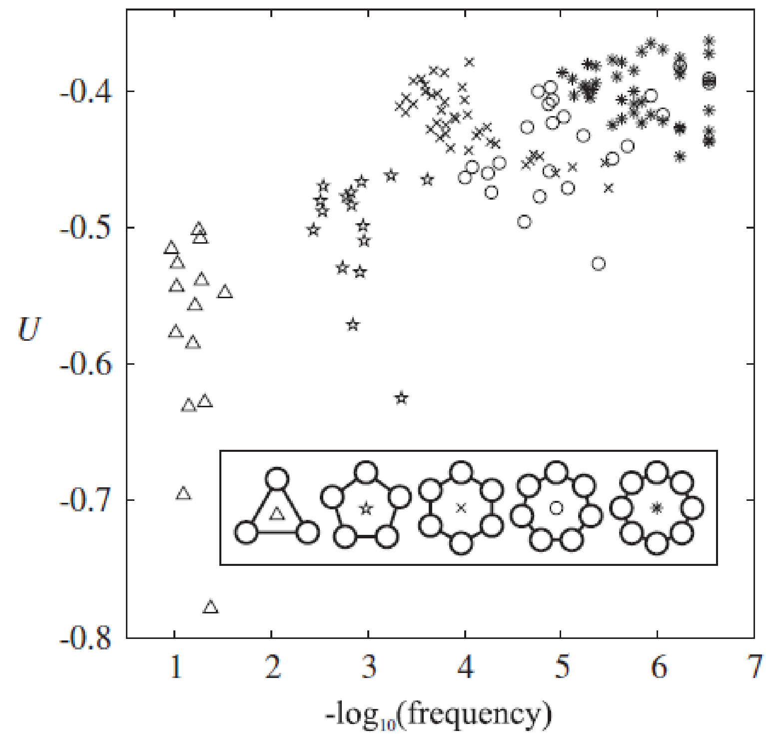
LTD: - always leads to a balanced state, but which one?
- psychologically not plausible: the dissonance can increase

CTD: - jammed states

$$U = -\binom{N}{3}^{-1} \sum_{ijk} s_{ij} s_{jk} s_{ki}$$



Examples of jammed states for 9 nodes
(only positive links are shown)



Jammed states for 26 nodes

[T. Antal, P.L. Krapivsky, S. Redner, Dynamics of social balance on networks , PRE 72 (2005) 036121]
[S. A. Marvel, S. H. Strogatz, J. M. Kleinberg, Energy landscape of social balance, PRL 103 (2009) 198701]

Proposition – a flow

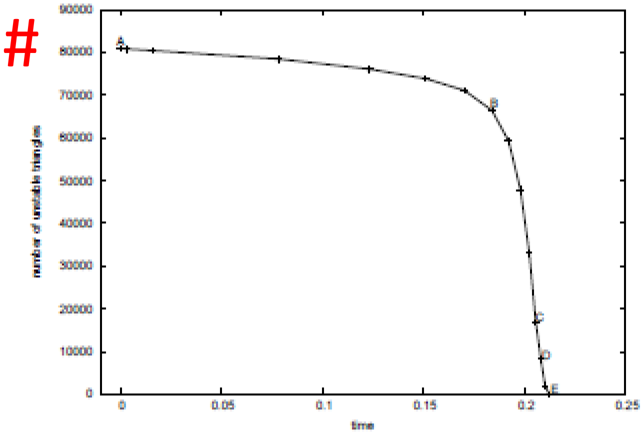
$$\frac{dx_{ij}}{dt} = G\left(\frac{x_{ij}}{R}\right)\sum_k x_{ik}x_{kj}$$

$$G(x) = 1 - x^2$$

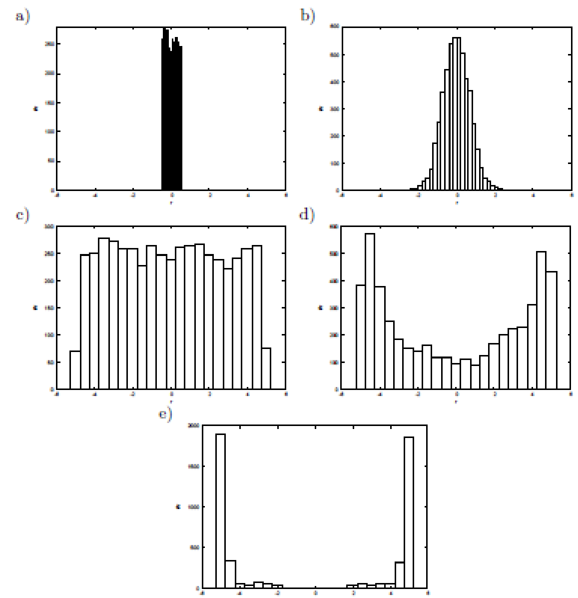
keeps the relation x_{ij} within the range (-R,R)

positive if x_{ik} and x_{kj} both friendly or both hostile, elsewhere negative

#



time



For $G(x)$=1, jammed states do not appear (Marvel *et al*).

[KK, P. Gawroński, P. Gronek, The Heider balance - a continuous approach, Int. J. Mod. Phys. C 16 (2005) 707]
[S.A. Marvel, J. Kleinberg, R.D. Kleinberg, S.H.Strogatz, Continuous-time model of structural balance, PNAS 108 (2011) 1771]

A test: Natchez women data

[L.C. Freeman, Finding Social Groups: A Meta-Analysis of the Southern Women Data, in Dynamic Social Network Modeling and Analysis, R. Breiger, K. Carley, and P. Pattison, (Eds.), US National Academies Press, 2003.]

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×	....	×	×	....	....	....	....	....
2. Miss Laura Mandeville	×	×	×	....	×	×	×	×	....	....	....	....	....	....
3. Miss Theresa Anderson	....	×	×	×	×	×	×	×	×	....	....	....	....	....
4. Miss Brenda Rogers	×	....	×	×	×	×	×	×	....	....	....	....	....	....
5. Miss Charlotte McDowd	....	....	×	×	×	....	×	....	....	....	....	....	....	....
6. Miss Frances Anderson	....	....	×	....	×	×	....	×	....	....	....	....	....	....
7. Miss Eleanor Nye	....	....	....	....	×	×	×	×	....	....	....	....	....	....
8. Miss Pearl Oglethorpe	....	....	....	....	....	×	....	×	×	....	....	....	....	....
9. Miss Ruth DeSand	....	....	....	....	×	....	×	×	×	....	....	....	....	....
10. Miss Verne Sanderson	....	....	....	....	....	....	×	×	×	....	....	×	....	....
11. Miss Myra Liddell	....	....	....	....	....	....	....	×	×	×	....	×	....	....
12. Miss Katherine Rogers	....	....	....	....	....	....	....	×	×	×	....	×	×	×
13. Mrs. Sylvia Avondale	....	....	....	....	....	....	×	×	×	×	....	×	×	×
14. Mrs. Nora Fayette	....	....	....	....	....	×	×	....	×	×	×	×	×	×
15. Mrs. Helen Lloyd	....	....	....	....	....	....	×	×	....	×	×	×	....	....
16. Mrs. Dorothy Murchison	....	....	....	....	....	....	....	×	×	....	....	....	....	....
17. Mrs. Olivia Carleton	....	....	....	....	....	....	....	....	×	....	×	....	....	....
18. Mrs. Flora Price	....	....	....	....	....	....	....	....	×	....	×	....	....	....

Meta-analysis – comparison of 21 methods

	Code	Analysis	Closeness to the Matching Criterion
1	DGG41	Davis, Gardner and Gardner, Ethnography	0.920
2	HOM50	Homans, Intuition	0.854
3	P&C72	Phillips and Conviser, Information Theory	0.968
4	BGR74	Breiger, Algebra	0.933
5	BBA75	Breiger, Boorman and Arabie, CONCOR	0.927
6	BCH78	Bonacich, Boolean Algebra	0.841
7	DOR79	Doreian, Algebraic Topology	0.923
8	BCH91	Bonacich, Correspondence Analysis	0.968
9	FRE92	Freeman, G-Transitivity	0.926
10	E&B93	Everett and Borgatti, Regular Coloring	0.916
11	FR193	Freeman, Genetic Algorithm 1	0.968
12	FR293	Freeman, Genetic Algorithm 2	0.842
13	FW193	Freeman and White, Galois Lattice	0.917
14	FW293	Freeman and White, Galois Sub-Lattice	0.954
15	BE197	Borgatti and Everett, Bi-Clique	0.916
16	BE297	Borgatti and Everett, Taboo Search	0.968
17	BE397	Borgatti and Everett, Genetic Algorithm	0.968
18	S&F99	Skvoretz and Faust, p* Model	0.957
19	ROB00	Roberts, SVD with Normalization	0.968
20	OSB00	Osbourn, VERI Algorithm	0.543
21	NEW01	Newman, Weighted Co-Attendance	0.932

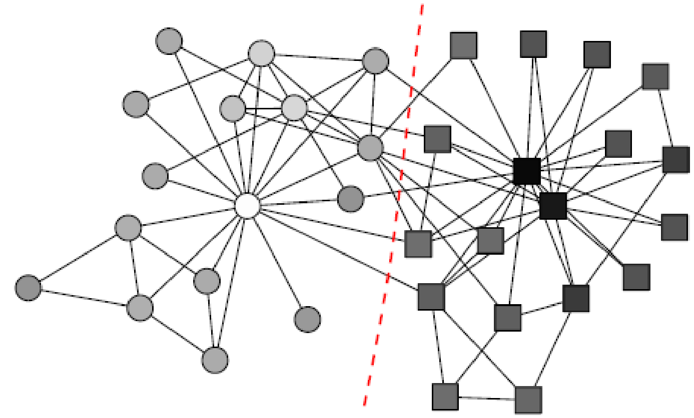
we get the same value

[L.C. Freeman, Finding Social Groups: A Meta-Analysis of the Southern Women Data, in Dynamic Social Network Modeling and Analysis, R. Breiger, K. Carley, and P. Pattison, (Eds.), US Nat. Acad. Press, 2003.]

A test: Zachary karate club

0	4	5	3	3	3	3	2	2	0	2	3	1	3	0	0	0	2	0	2	0	2	0	0	0	0	0	0	0	0	0	2	0	0
4	0	6	3	0	0	0	4	0	0	0	0	0	5	0	0	0	1	0	2	0	2	0	0	0	0	0	0	0	0	2	0	0	0
5	6	0	3	0	0	0	4	5	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	2	0
3	3	3	0	0	0	0	3	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	5	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	2	5	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3	4
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
2	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	5	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4
0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1
2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	4	0	3	0	0	5	4
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	0	2	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	7	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	2
0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	0	0	0	0	0	0	0	0	4
0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	4	0	0	0	0	0	4	2
0	2	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	0	0	2	0	0	0	4	4
0	0	2	0	0	0	0	0	3	0	0	0	0	0	3	3	0	0	1	0	3	0	2	5	0	0	0	0	0	4	3	4	0	5
0	0	0	0	0	0	0	0	4	2	0	0	0	3	2	4	0	0	2	1	1	0	3	4	0	0	2	4	2	2	3	4	5	0

Weighted (0,1,...,7)



$$x_{ij} \to x_{ij} - \varepsilon$$

$$\frac{dx_{ij}}{dt} = [1-(x_{ij}/R)^2]\sum_k x_{ik}x_{kj}$$

The obtained split almost exactly as real:

1-8,11-14,17,18,20,22

vs

9,10,15,16,19,21,23-34

for R=5 and any ε, if only $-R < x_{ij} < R$

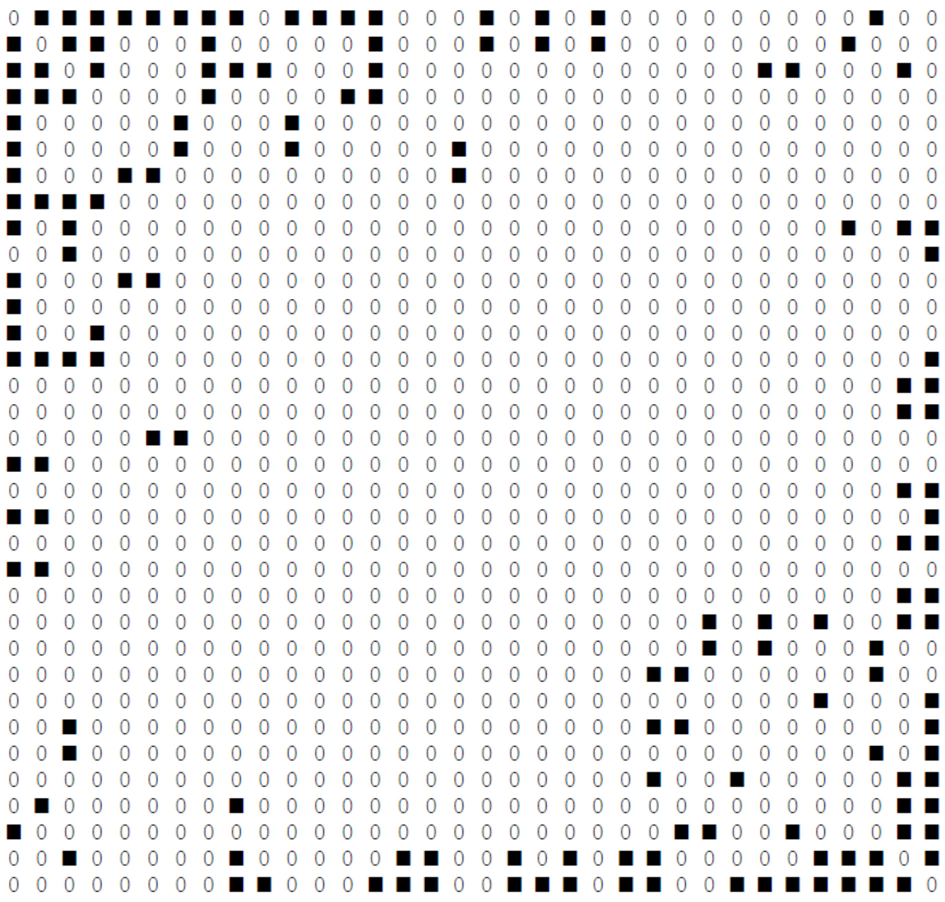
[W.W. Zachary, An information flow model for conflict and fission in small groups, J. Anthr.Res. 33 (1977) 452]

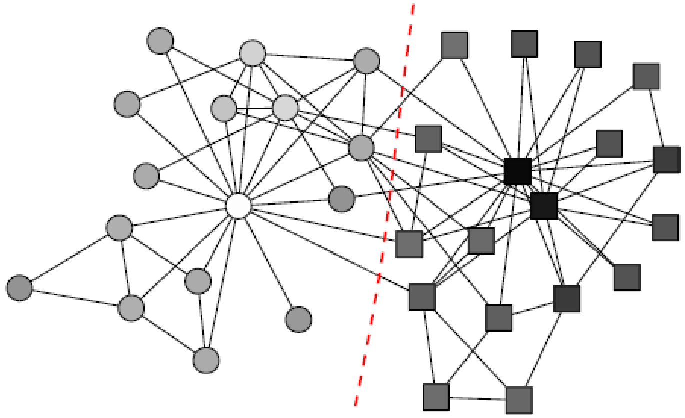
[http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm#zachary]

[P. Gawronski, K. K., Heider Balance in Human Networks, AIP Conf. Proc.,vol. 779, 2005, pp.93–95]

[M. E. J. Newman, Modularity and community structure in networks, PNAS 103 (2006) 8577]

A test: Zachary karate club





$$x_{ij} \rightarrow x_{ij} - \varepsilon$$

$$\frac{dx_{ij}}{dt} = [1-(x_{ij}/R)^2]\sum_k x_{ik}x_{kj}$$

The obtained split almost exactly as real:

1-8,11-14,17,18,20,22

vs

9,10,15,16,19,21,23-34

for R=5 and any ε, if only $-R < x_{ij} < R$

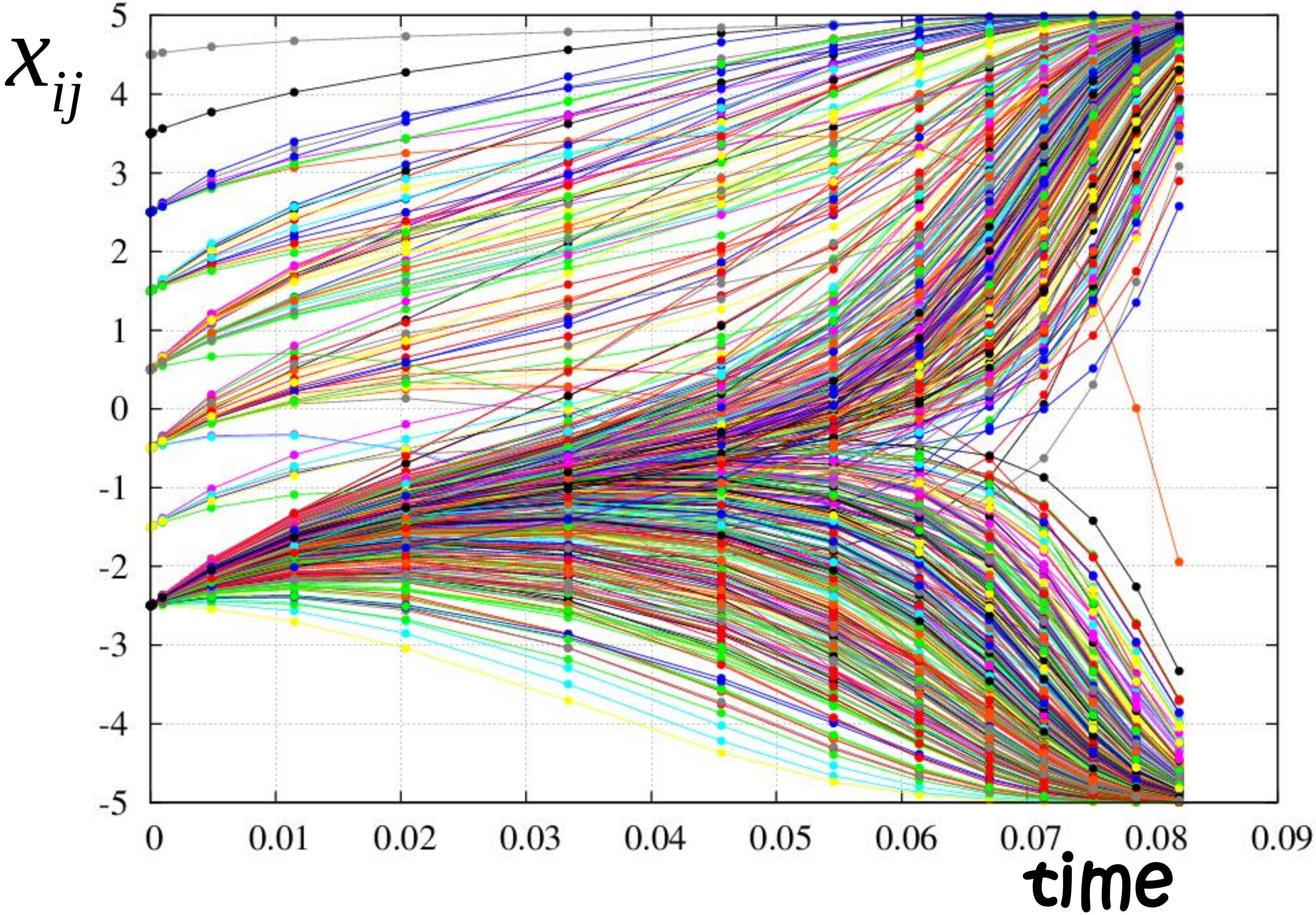
[W.W. Zachary, An information flow model for conflict and fission in small groups, J. Anthr.Res. 33 (1977) 452]

[http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm#zachary]

[P. Gawronski, K. K., Heider Balance in Human Networks, AIP Conf. Proc.,vol. 779, 2005, pp.93–95]

[M. E. J. Newman, Modularity and community structure in networks, PNAS 103 (2006) 8577]

Zachary karate club + calculated relations



More communities

$$\frac{dA_{ij}}{dt} = G(A_{ij}) \sum_{k \neq i,j} (A_{ik} A_{kj} - \beta)$$

where A - the connectivity matrix,

$$G(x) = \Theta(x)\Theta(1-x)$$

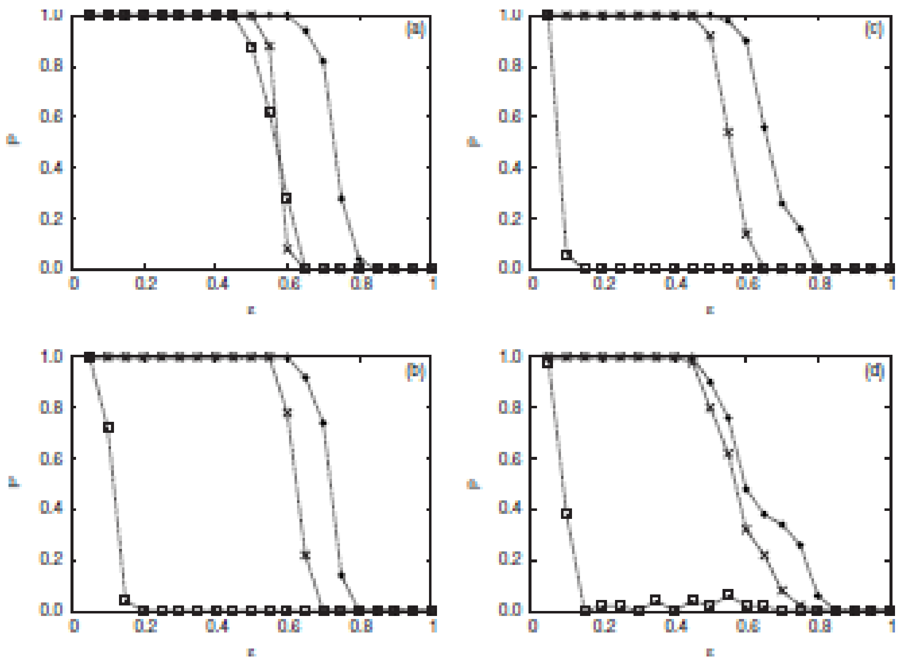


FIG. 4. Results for clusters of different sizes: (a) $N=110$, clusters of 50, 48, and 12 nodes, (b) $N=130$, clusters of 22, 34, 11, 10, and 53 nodes, (c) $N=130$, clusters of 19, 60, 45, and 6 nodes, (d) $N=110$, clusters of 18, 22, 30, 38, and 2 nodes [symbols denote: □, the Newman algorithm; ×, Eq. (2) with $\beta=0.25$; ♦, Eq. (2) with $\beta=0.4$].

[M. J. Krawczyk, Differential equations as a tool for community identification, PRE 77 (2008) 065701R]
[M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, PRE 69 (2004) 026113]

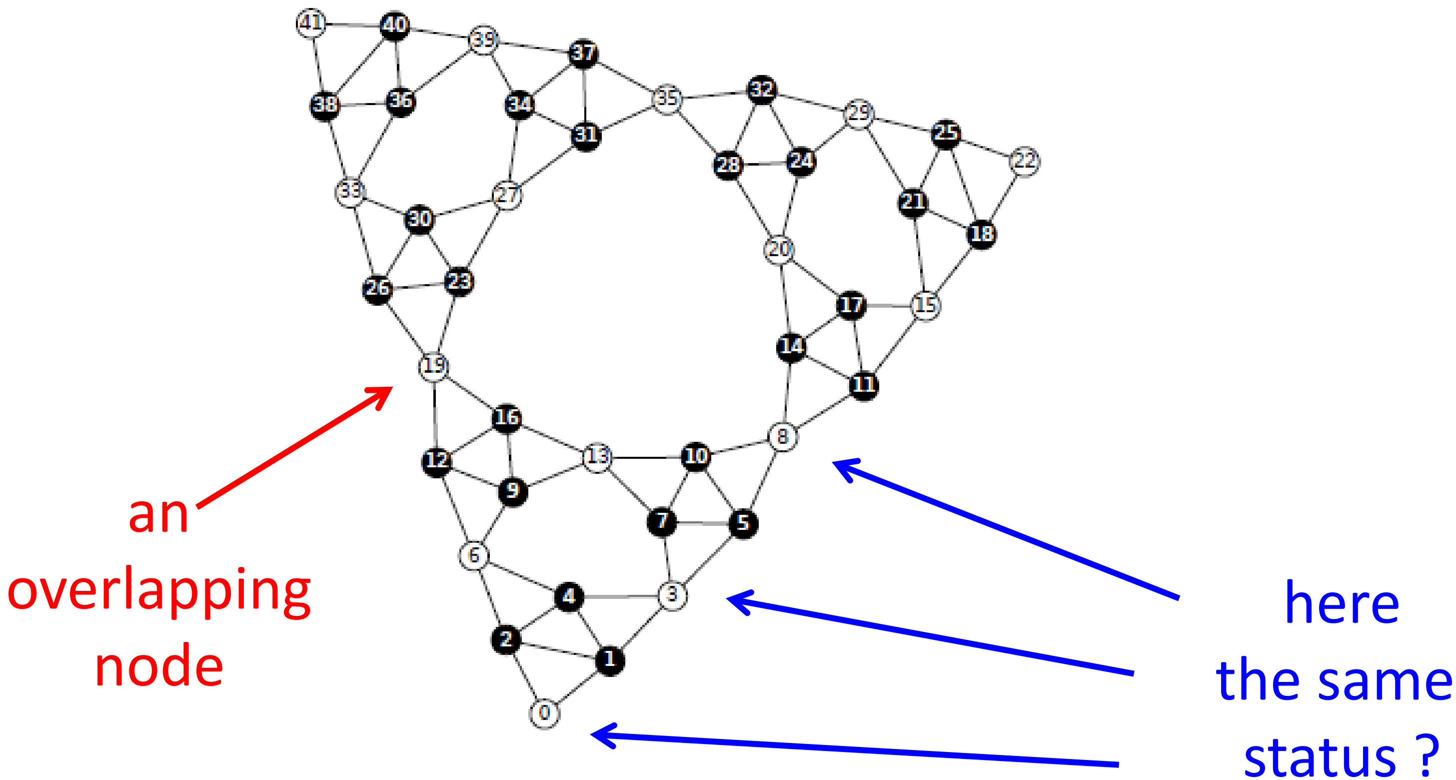
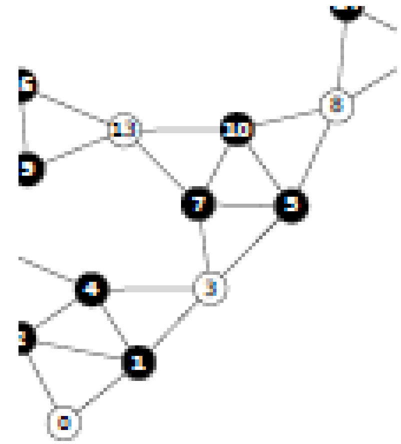


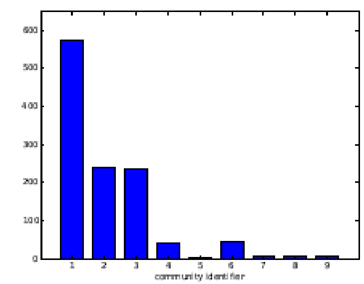
Figure 3: The Sierpinski triangle network with indication of communities for $N = 42$, here we have 9 communities of 3 nodes each (black nodes) and 15 communities of 1 node each (white nodes).

[M.J.Krawczyk, Communities and classes in symmetric fractals, IJMPC (2014), in print (arXiv:1404.7416)]

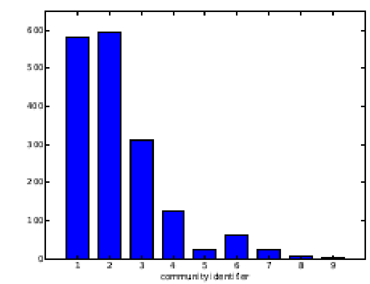
How to detect overlapping nodes?



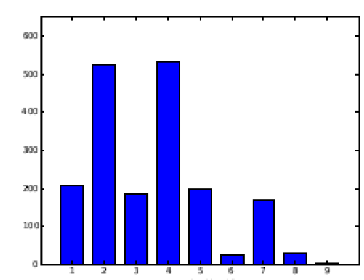
- add a small noise to the initial matrix of connectivity
- run the equations many times
- construct a histogram: how often a given node belongs to particular community?
- if more than one community appear with the largest peaks, the node is overlapping



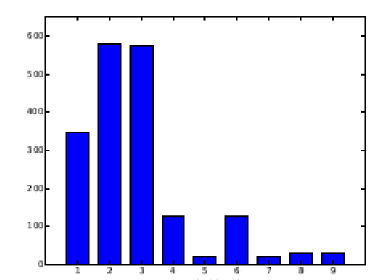
(a) Node number 0. Analogous histograms are obtained for nodes 22 and 41.



(b) Node number 3. Analogous histograms are obtained for nodes 6, 15, 29, 33 and 39.

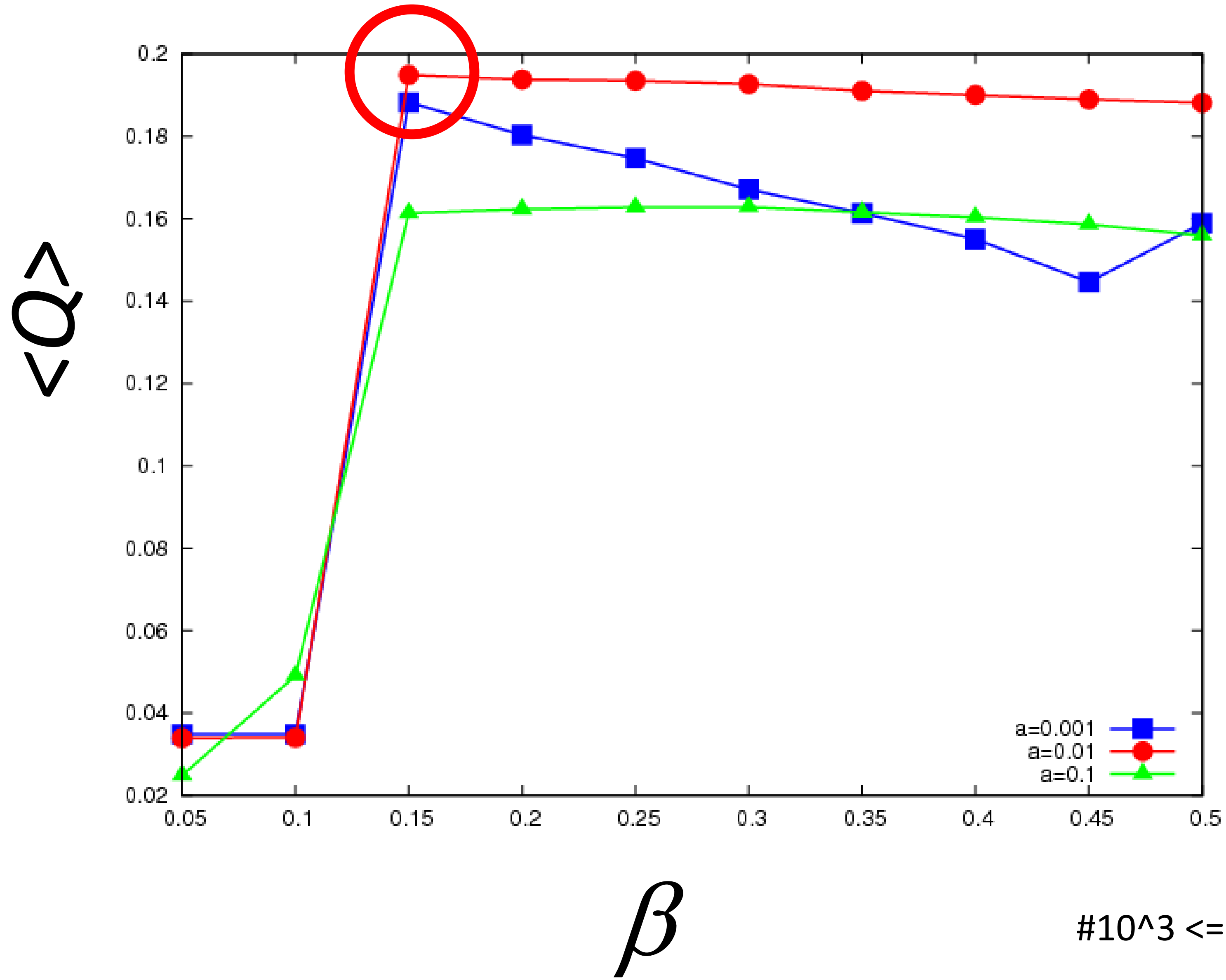


(c) Node number 8. Analogous histograms are obtained for nodes 19 and 35.



(d) Node number 13. Analogous histograms are obtained for nodes 20 and 27.

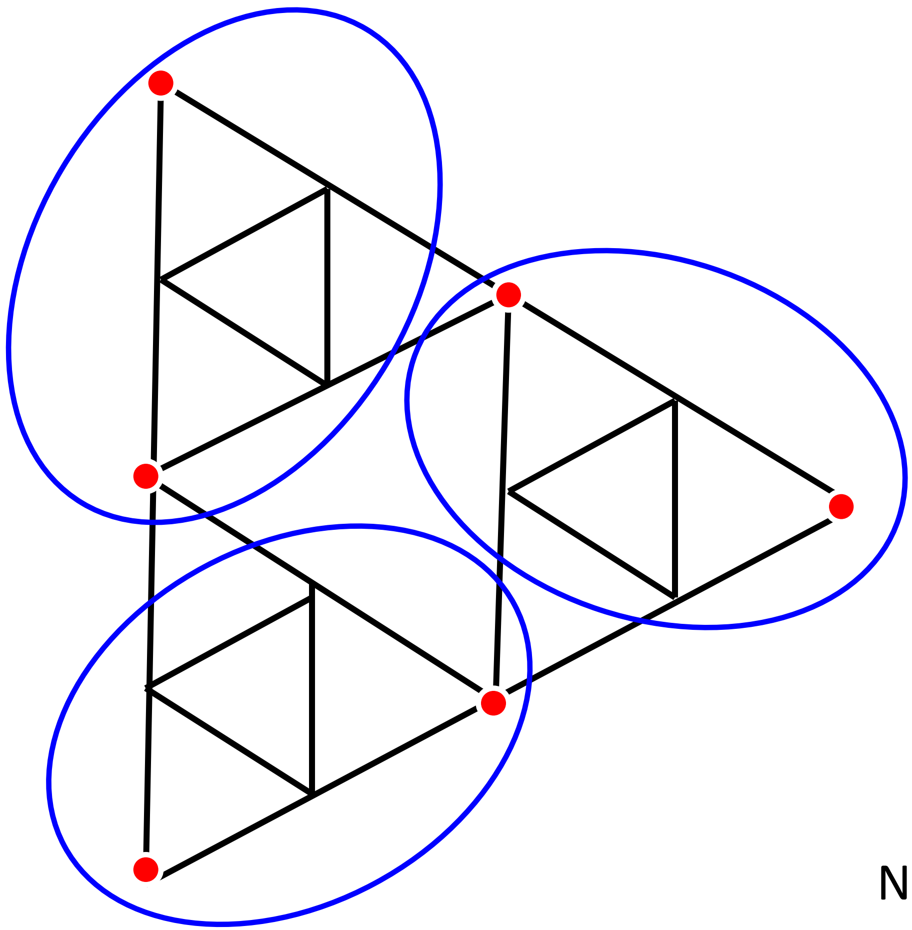
Sierpiński triangle with noise, N=15

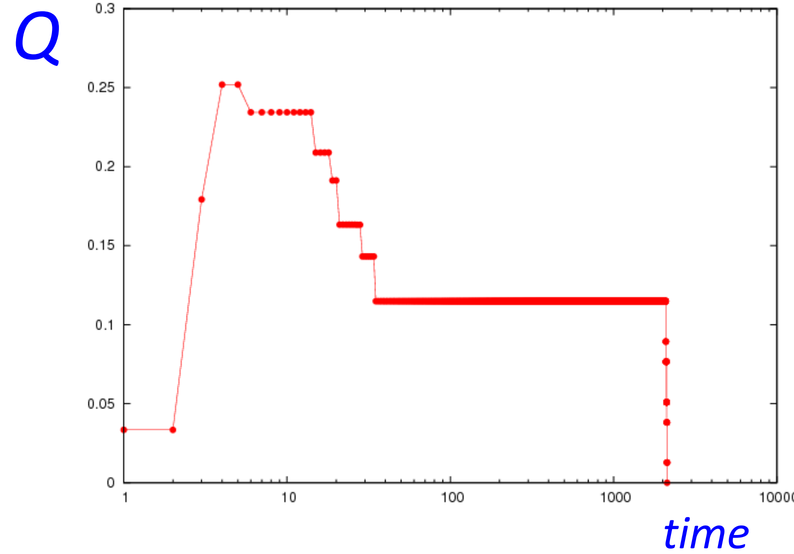


#10^3 <=> #10^4

N = 15

β=0.15





No noise:

- isolated communities

Q = 0.12

Noise a = 0.01:

<Q> = 0.19

- communities

Q_{max}= 0.25

No noise:

- communities

Q = 0.26

Conclusions:

- the designed flow of links drives the system to a set of separate clusters

- the 4-th axiom of Heider precludes more than two communities

- an application of noise breaks symmetry, allows to reveal overlapping nodes and leads to better partitions

Thank you