Report on the OTKA Project 60456:

STRUCTURE AND DYNAMICS OF COMPLEX NETWORKS

February 2006 – June 2010
PI: János Kertész

I. Introduction

According to Steven Hawking the 21-st century will be the century of complexity [1]. While such statements are always somewhat shaky, there is much to expect from complexity science in the near future. However, even the definition of complex systems seems to be an extremely difficult task, so usually a kind of characterization is made without the sake of completeness. Typically mentioned features are: Many interacting components, nonlinearity, feedback, cooperativity, emergent phenomena [2].

In statistical physics we like to concentrate on universal features, where details of interactions become unimportant. It is a focus question to what extent this can be done when analyzing complex systems. Clearly, a real breakthrough in this context was achieved by network theory [3,4], which is a truly holistic approach, where – in its simplest form – the interactions have only a binary role: Either they exist or they do not. It turned out that a considerable amount of complexity comes from the topology of the scaffold of the systems, which can be described by complex networks. Moreover, we have learned that networks describing very different systems like the cell, the internet or the air traffic show remarkable similarities – a real joy for a statistical physicist. A spectacular development started at the turn of the millenium, which is still going on as demonstrated by the unbroken impetus of research as measured, e.g., by the number of papers in this field.

The main challanges this project have been related to are the following:
- How to make a step toward reductionism from the entirely holistic approach of binary networks? In other words: How can we develop network theory in order to be able to answer *more system specific* questions?
- What are the relevant *mesoscopic structures* and how can we identify them?
- How to *apply* network theory to a broad range of systems where – due to the enormously developing computarization large amount of data are available?
- What are the most important *dynamic processes* and how are they related to the topology of the networks?
- In what respect are (developing) networks *optimal* and how can they be improved?

II. Characterization of networks

Weighted networks

The first step toward reductionism from the holistic approach of binary networks is to introduce a measure for the strength of the interactions as edge or link weights. Earlier we introduced new measures to characterize weighted networks like subgraph intensity and coherence. The intensity is the geometric mean of the weights in a

subgraph and the coherence is the ratio of the geometric and the arithmetic mean. Using these of measures weighted motifs could be defined and it turned out that the weighted motif statistics leads to surprising results as compared to the case when we do not take weights into account [5].

It was natural to apply these ideas to the clusters of weighted networks as obtained from correlation matrices, e.g., of stock market data. If groups of stocks belonging to a given business sector are considered as a fully connected subgraph of the final network, their intensity and coherence can be monitored as a function of time (Fig. 1). This approach indicates to what extent the business sector classifications are visible in market prices, which in turn enables us to gauge the extent of group-behavior exhibited by stocks belonging to a given business sector. It should be emphasized that this approach is widely applicable if large correlation matrices have to be analyzed.

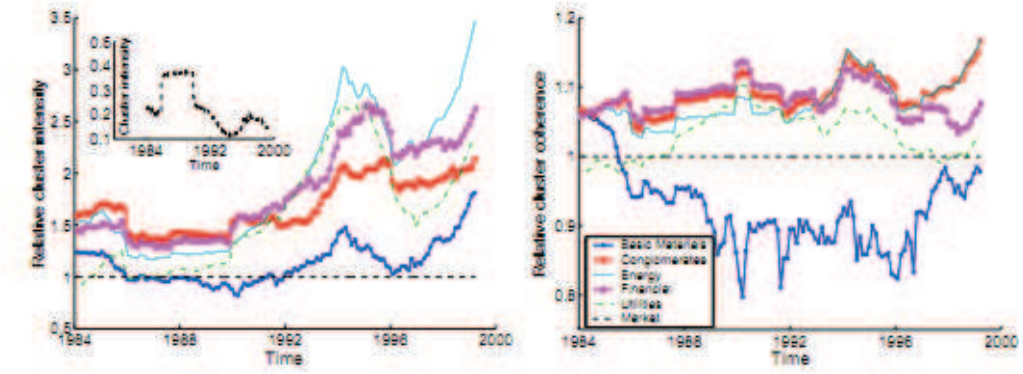

Figure 1. (a) Relative cluster intensity as a function of time for select clusters. In the inset: The absolute cluster intensity for the whole market used for normalisation. (b) Relative cluster coherence as a function of time.

In Fig. 1(a) we show the relative cluster intensity as a function of time for selected business sector clusters. Here the values above unity indicate that the intensity of the cluster is higher than that of the market. This implies that in most cases stocks belonging to a given business sector are tied together in the sense that intra-cluster interaction strengths are considerably stronger than the whole market interaction. In the inset of Fig. 1(a) we have depicted the absolute cluster intensity for the whole market, which shows high values roughly between 1986 and 1990. This is caused by stock market crash (Black Monday, 1987) when the market behaves in a unified manner. It should be noted here that although the crash is a localized event, in our analysis it covers an extended period due to the moving window length being four years. From Fig. 1(a) we also see that the crash compresses the relative cluster intensities, which means that the cluster-specific behaviour is temporarily suppressed by the crash. After the market recovers, the clusters regain their normal structural features.

The clustering coefficient is an important measure for binary graphs:

$$C_i = \frac{2t_i}{k_i\,(k_i-1)},$$

where k_i is the degree of node i and t_i is the number of triangles there. It is an interesting question whether one can generalize this quantity to weighted networks.

There have been a number of suggestions (including ours, [5]). In a comparative study we defined a number of plausible requirements such a generalization should obey and analyzed the different formulas from that point of view. The following table shows the results:

Coeff.	Motivation
$\tilde{C}_B$	Reflects how much of vertex strength is associated with adjacent triangle edges
$\tilde{C}_O$	Reflects how large triangle weights are compared to network maximum
$\tilde{C}_Z$	Purely weight-based; insensitive to additive noise which may result in appearance of "false positive" edges with small weights
$\tilde{C}_H$	Similar to $\tilde{C}_Z$, based only on edge weights

Feature	$\tilde{C}_B$	$\tilde{C}_O$	$\tilde{C}_Z$	$\tilde{C}_H$
1) $\tilde{C} = C$ when weights become binary	X	X	X	
2) $\tilde{C} \in [0,1]$	X	X	X	
3) Uses global $max(w)$ in normalization		X	X	X
4) Takes into account weights of all edges in triangles		X		X
5) Invariant to weight permutation for one triangle		X		
6) Takes into account weights of edges not participating in any triangle	X		X	X

Table 1. Comparison of different generalizations of the clustering coefficient to the case of weighted networks.

From this table is clear that while our suggestion, using triangle intensity [5]

$$\tilde{C}_{i,O} = \frac{1}{k_i\,(k_i-1)} \sum_{j,k} \left(\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk}\right)^{1/3}$$

gives the best performance, there is no ultimate solution to this problem, indicating that it is not possible to describe the property of clustering by a single measure.

Community detection

One of the focus problems in the theory of complex networks has been the detection of communities. These are meso-scale parts of the networks, which are denser then the average and which are supposed to play important functional role. There have been hundreds of papers devoted to this issue during the last decade; for a recent review we refer to [7]. We have contributed in several ways to the clarification of the related problems.

Since the recent interest concentrates on huge networks, computational demands are crucial in deciding about the efficiency of a method. That is the reason why the so-called „label propagation method" awaked interest: It was clearly a very fast method. It defines a community as a set of nodes such that each node has at least as many neighbors in its own community as in any other one. In the initial stage of the method,

all nodes form a distinct community (each node has an own \label"). Then, at each time step, the nodes join that community to which the largest fraction of their neighbors belong, by adopting the corresponding label. If there are multiple choices, a random decision is made with uniform distribution. We have shown [9] that this method is equivalent to finding the ground state of a simple Potts model.

The ground state is ferromagnetic, i.e., all spins are in the same state. However, the configuration may freeze into a metastable state, where more than one states are present. The latter depends on (besides the adjacency matrix) the random updating sequence and on the choices made in ambiguous cases. These metastable states are identified with the community structure. As it is clear by construction and demonstrated on real datasets the method leads to a proliferation of partitions and it is non-trivial to decide about their quality.

One of the most popular community detection methods is that of Newman and Girvan [10] based on modularity

$$Q = \sum_s e_{ss} - a_s^2,$$

where e_{ss} is the fraction of links that fall into community s and a_s is the number of links from community s to other communities. The method is to find the partitioning, which maximizes Q. Clearly, large Q means that the number of links within the community is large, i.e., it is a dense region. While this method received much attention because its conceptually appealing features, its relation to other physics and computer science problems, like NP completeness, there are severe problems with it. As pointed out by Fortunato and Barthélemy [11], the method has a resolution limit, i.e., for large networks it is unable to detect small modules.

The relationship between the modularity approach and other physics problems is best established by the mapping of modularity optimization to a Potts spin glass [12] with the following Hamiltonian:

$$\mathcal{H} = -\sum_{i \neq j} (A_{ij} - \gamma p_{ij})\, \delta(\sigma_i, \sigma_j),$$

where γ is a parameter of the model A_{ij} denotes the adjacency matrix p_{ij} denotes the link probability between nodes i and j according to a null model. Modularity optimization corresponds to taking the configuration model as null model with $p_{ij} = \frac{k_i k_j}{2L}$, where L is the total number of links. We have shown [13] that for very general conditions and with any null model the method suffers from the resolution limit problem: The minimum size of the communities detectable goes as $\sqrt{L}$, with the prefactors depending on the null models and the parameter γ. The effect is illustrated in Fig. 2.

The prefactor depends also on γ. This enables to introduce a multiresolution method by using γ as a running parameter. At different values of γ the resolution limit will also be different, enabling to monitor the whole range of community sizes [14]. Fig 3. shows the result of such a study.

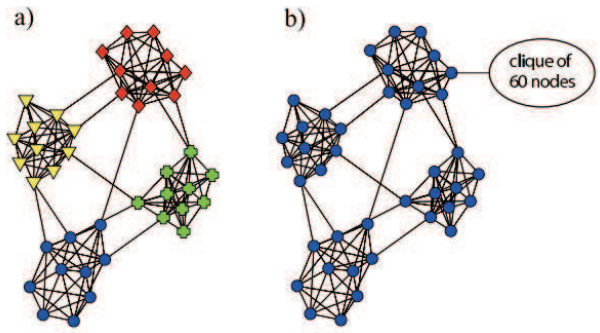

Figure 2. a) the Potts spin glass method is able to identify the ovious communities in a small network. b) Just by attaching a clique of 60 nodes to the network the previously separated communities get merged. $\gamma = 1$ in these calculations.

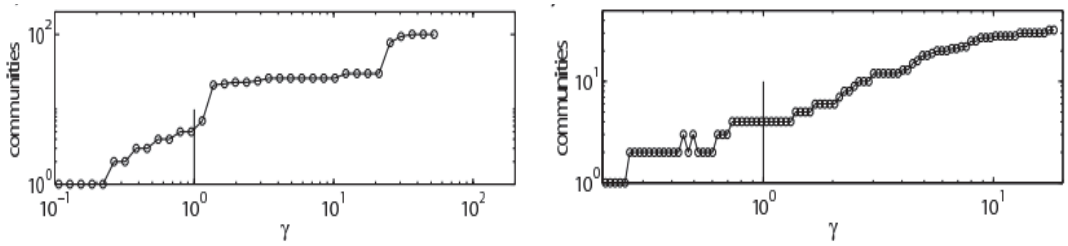

Figure 3. Multiresolution method. a) hierarchical scale free network with 125 nodes. b) Zachary karate club.

Whenever we see a plateau as a function of the parameter, there is a stable community structure. The fact that multiple plateaus are seen shows that the network has a nested, hierarchical structure.

The problem of community detection is already difficult enough if the network is very large but no additional complications occur. In real networks, however, there are two further problems to face with. The first is has already been mentioned, namely that very often the community structure is hierarchical. The second problem is that a node may belong to more than one communities, as easily seen at the example of social networks [16]. Both problems (see Fig. 4) indicate that partitioning is not necessarily the best solution. We set the goal to introduce a method, which could cope with these problems and, at the same time it has an acceptable computational time.

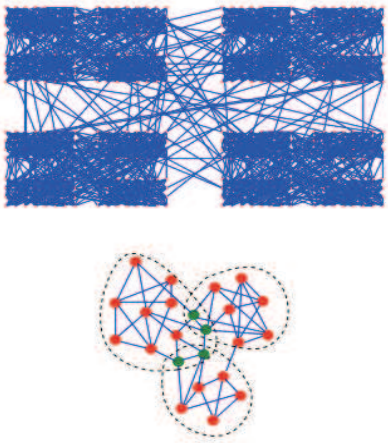

Figure 4. a) Hierarchically structured communities in a network; b) example for overlapping communities

Our method is local, i.e., there is no global optimization related. We define the fitness of a subgraph $\mathcal{G}$ as

$$f_{\mathcal{G}} = \frac{k_{in}^{\mathcal{G}}}{(k_{in}^{\mathcal{G}} + k_{out}^{\mathcal{G}})^{\alpha}},$$

where $k_{in}^{\mathcal{G}}$ ($k_{out}^{\mathcal{G}}$) is the sum of internal (external) degrees of the nodes of $\mathcal{G}$, and α is a parameter, which is for simplicity chosen to be 1 at first. The fitness of node A in $\mathcal{G}$ is

$$f_{\mathcal{G}}^{A} = f_{\mathcal{G}+\{A\}} - f_{\mathcal{G}-\{A\}}$$

where the two terms on the rhs mean the fitness of the community with and without the node A. The procedure is to find the community around a node, which maximizes the fitness of the community. This is done in a local, iterative way, which turns out to be rather economic. E.g., the community structure of the .gov domain of the web (consisting 774, 908 nodes and 4, 711, 340 links) could be identified within less than 40 hours of CPU time on a small PC (Figure 5).

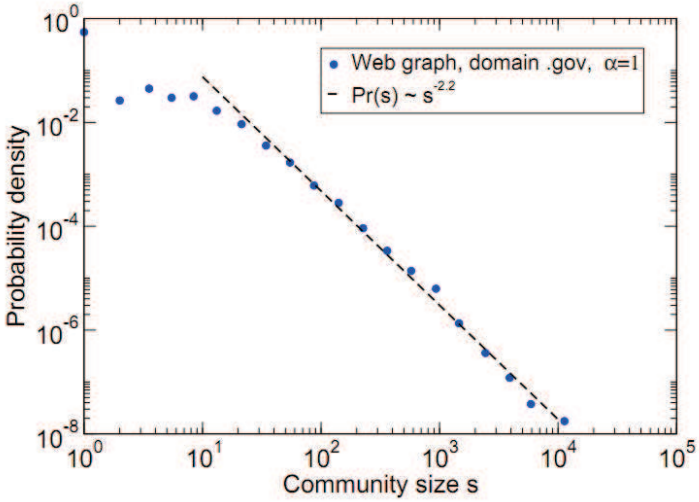

Figure 5. Community structure of the .gov domain of the WWW.

The method allows for overlaps by definition. The parameter α can be used as a resolution parameter. Again, if changing α there is a plateau in the total fitness, the cover (a tiling with overlaps) is stable. This allows for introducing „community spectroscopy" as illustrated in Figure 6.

Hierarchical structure of networks

The method described above is able to capture hierarchically organized community structures. Different peaks in Fig. 6 may refer to different levels of the hierarchy. The quality of the identification of hierarchical structures can be seen if the method is applied to a benchmark graph like in Fig. 4a, and the ratio inter-community links is gradually increased. Measuring the so-called normalized mutual information (which compares the results to the a priori known structure) the goodness of the method can be measured (Fig. 7).

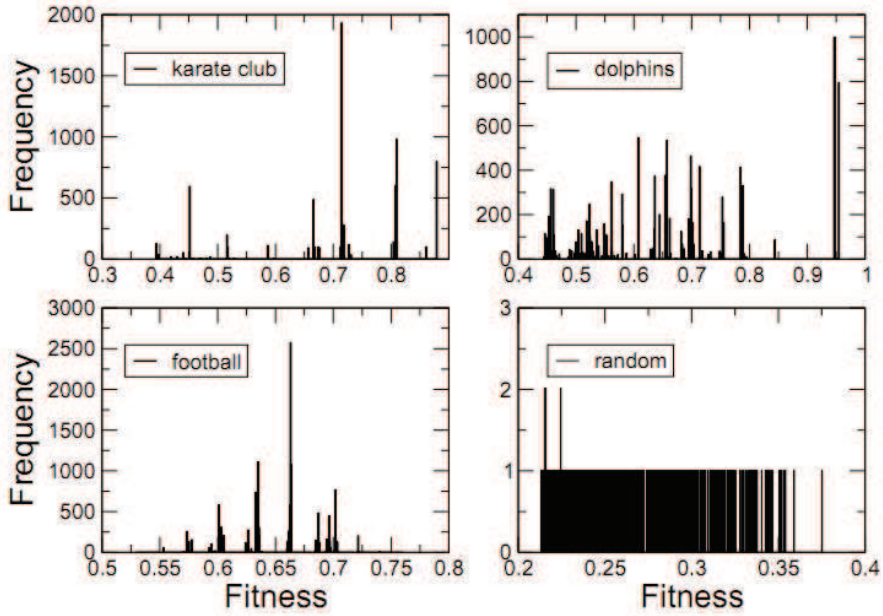

Figure 6. The peaks in this plots indicate the values of fitness parameters α, where there is a stable cover. Three frequently studied examples are shown (a,b,c). One of the advantages of the method is in grap d): no stable community structure in a random graph.

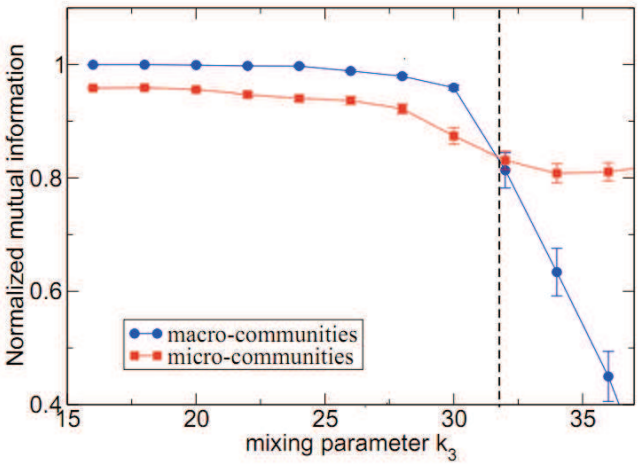

Figure 7. Measuring the ability of the method to identify hierarchical structures. k_3 is the parameterby which the number of the inter-community links is controlled. Around $k_3 = 32$ the hierarchical structure breaks down.

An interesting question is the relation between hierarchical organization and randomness. In a fully random system the different levels of the hierarchy can hardly be identified. We addressed this question by randomizing a fully deterministic (ordered) hierarchical network (Fig. 8, [17]) in a controlled way [18].

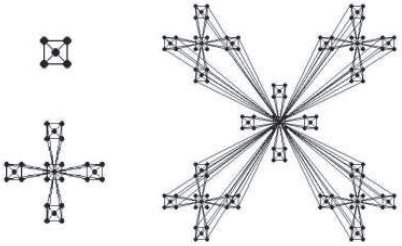

Figure 8. Deterministic hierarchical network

We studied the topological overlap matrix as a function of a randomization parameter *p*. The *i, j* element of this matrix, T_{ij} equals the number of mutual neighbours of the nodes *i* and *j* (plus 1 if *i* and *j* are connected), normalized by the minimum degree of *i* and *j*, so $T_{i\,j}$ is between 0 and 1. Fig. 9 shows the Fourier transform of T_{ij}-s averaged over many samples as a function of *p*.

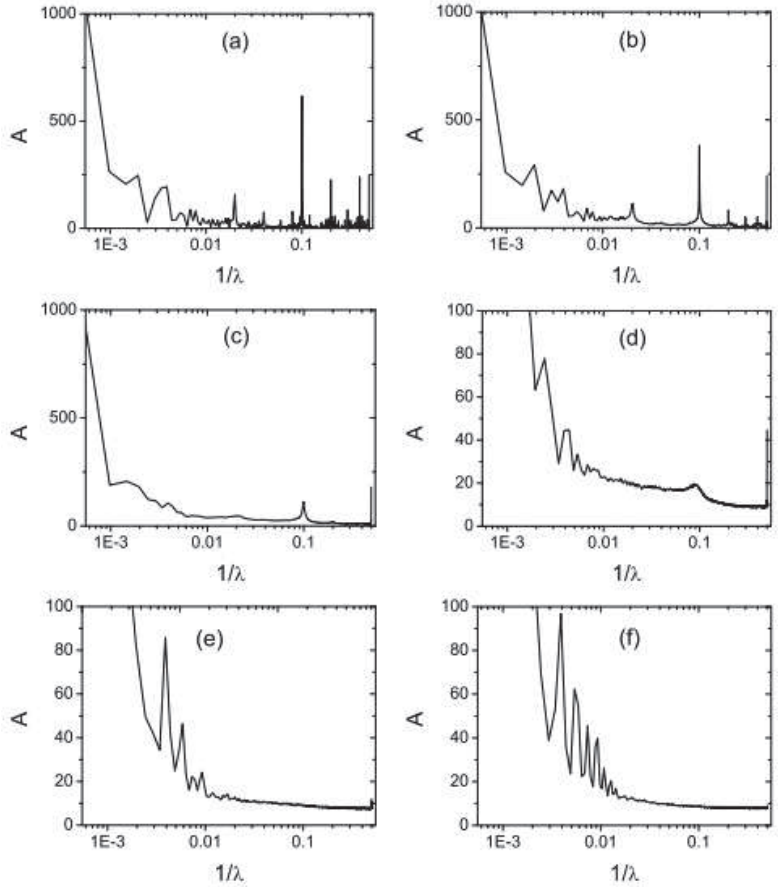

Figure 9. FFT spectra of topological overlap matrices. Each curve represents an averaged spectrum over 500 networks of 625 nodes for p = 0 (a), 0.01 (b), 0.05 (c), 0.2 (d), 0.4 (e), 1 (f).

The peaks in Fig. 9 correspond to well separated levels of hierarchical organization, which get smeared out when randomness is increased. The fact that natural and human made networks show clear hierarchical features shows that – while these networks are all random to some extent – they are far from total randomness.

III. Analysis and modeling large complex networks

Recent developments in Information Communication Technology (ICT) provide a flood of data about one of the most complex systems – the human society. Electronic databases, from phone to emails logs, produce detailed records of human communication patterns, offering novel avenues to map and explore the structure of social and communication networks. We examined [19] the communication patterns of millions of mobile phone users, allowing us to simultaneously study the local and the global structure of a society-wide communication network. We observed a coupling between interaction strengths and the network's local structure, with the consequence that social networks are robust to the removal of the strong ties, but fall

apart following a phase transition if the weak ties are removed. We showed that this coupling significantly slows the diffusion process, resulting in dynamic trapping of information in communities, and find that when it comes to information diffusion, weak and strong ties are both simultaneously ineffective.

Since the coverage by mobile phones is close to 100% in the adult population of modern societies, the call network as constructed from the records can be considered as a proxy of social relationships, the social links. Moreover, the call duration or frequency tells us about the intensity of these links. In his classic paper Granovetter formulated his hypothesis about the relationship of the topology and the tie strengths [20]: The closer a relationship between two persons, the larger is the overlap between their other friends. This hypothesis suggests a plausible structure of the society: There are communities, which are strongly "wired" and these are connected by weak links – hence the term "the strength of weak ties".

We were able to proof Granovetter's hypothesis for the first time on a network of societal size. We translated the verbal terms into mathematical concepts, measured them on the empirical network and showed that the overlap is indeed a monotonous function of the tie strength. Figure 10 shows a little part of the whole network to illustrate the main features.

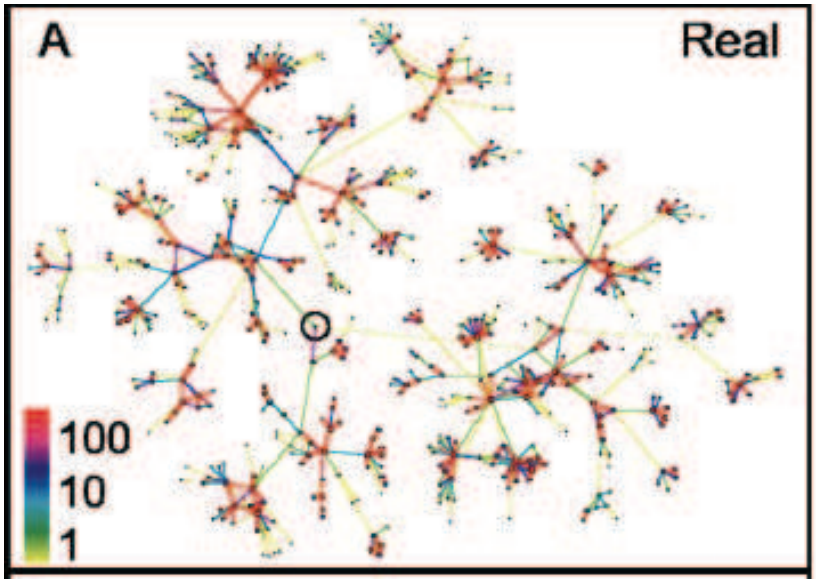

Figure 10. Part of the call network illustrates the validity of the picture of the "strength of weak ties"

This local relationship between the tie strength and the topology has consequences to the global properties. If we make the percolation experiment that we remove the links one by one in the increasing and decreasing order of their strength, we arrive at totally different results. Removing the weak links first results in fragmentation, i.e., in a percolation transition, since the communities become separated. In the opposite order there is no phase transition – the weak links hold the whole system together (Fig. 11).

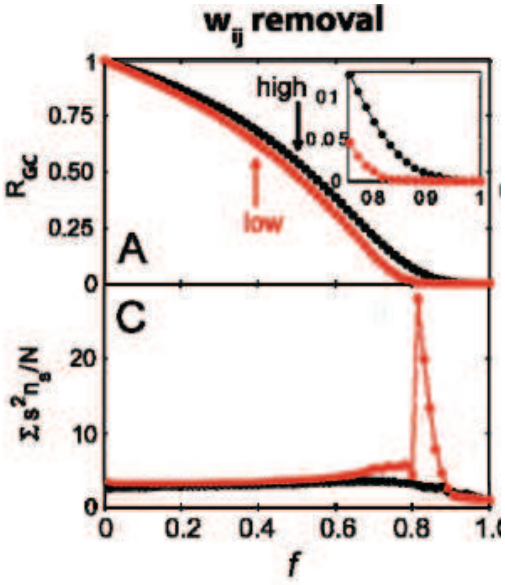

Figure 11. Order parameter and "susceptibility" of the percolation transition. For increasing order (red) we there is a transition, for decreasing (black) there is none.

This relationship between the weights of the links and the topology has severe consequences on the dynamic processes on the network. We examined the spreading properties by using an SI (susceptible-infected model) in two ways: a) we assumed spreading on a link with the frequency proportional to the link weight; b) we used uniform spreading frequencies as a reference system. The result is shown in Fig. 12.

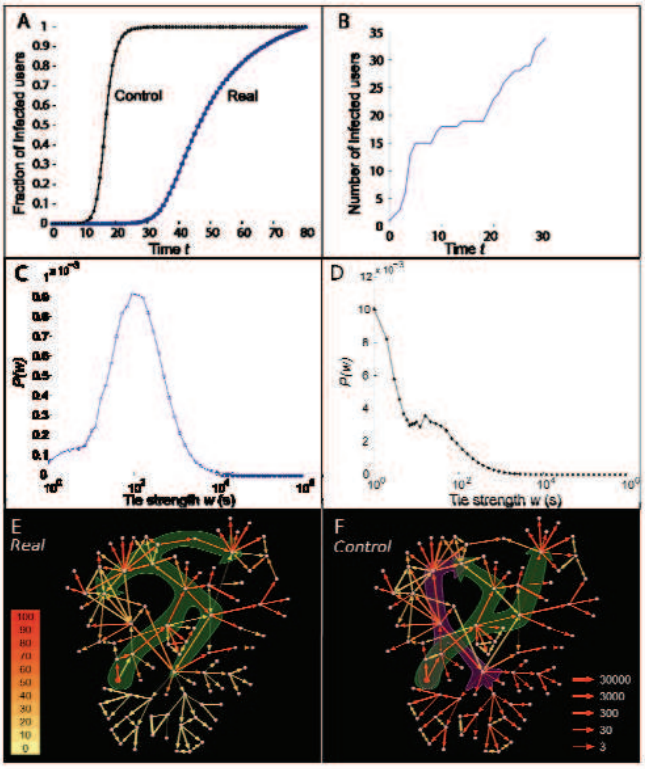

Figure 12. a) Comparison of spreading with frequencies as taken from the empirical weights and the reference system. b) The stepwise spreading shows the effect of trapping in communities. c) Weight distribution of first infection transmitting links in the inhomogeneous d) and in the reference systems. e) Spreading pattern for the "real" and f) the reference system.

Fig. 12 also shows what we called "the weakness of weak *and* strong ties". The statistics over the strength of the links transmitting the new information (or disease) shows that neither the strong ones (within the communities there is usually no news) nor the very weak ones (which have too low transmission frequency) play the main role but the medium strength links have the highest transmission power. This observation is relevant from the point of view of searching in a network. Our conclusion is that the structure of human society is not optimized toward spreading. Optimization would require high throughput (i.e., weights) on the links with high betweenness centrality, which are just the intercommunity bridges (see Fig. 13).

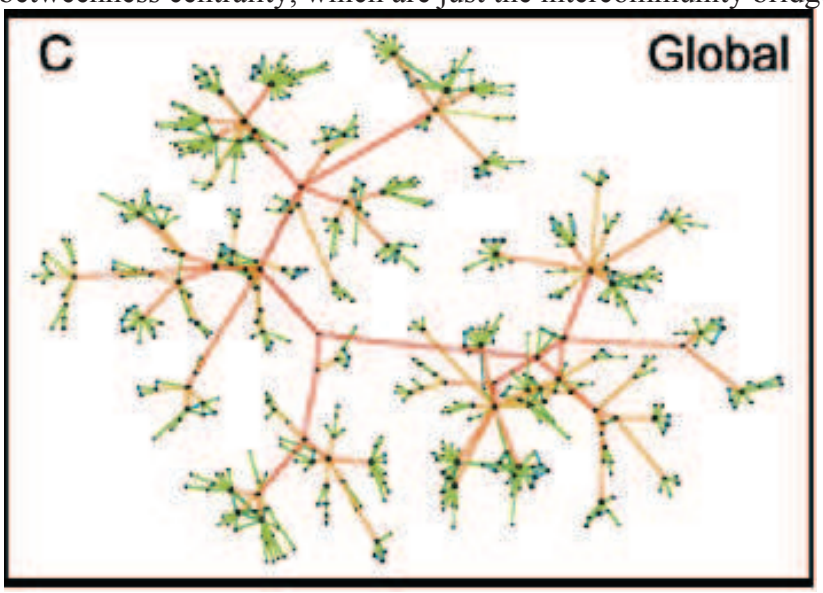

Figure 13. Network, having the same topology as the call network but with optimized throughput. The coloring is the same as in Fig. 10.

We also examined the self-organizing complex network of human communication by our method of the narrative network interview. We arrived at the conclusion that the canonization of the awareness of the community, the network is the result of a multidimensional process, where static and dynamic elements play equally important role [21].

Emergence of communities in a weighted network

The discovery of the intimate relationship between topology, link weigh and communities made it necessary to construct a simple model where all these elements are present. Our aim [22] was to introduce such a model on the base of sociological considerations [23]. Figure 16 shows the main linkage steps.

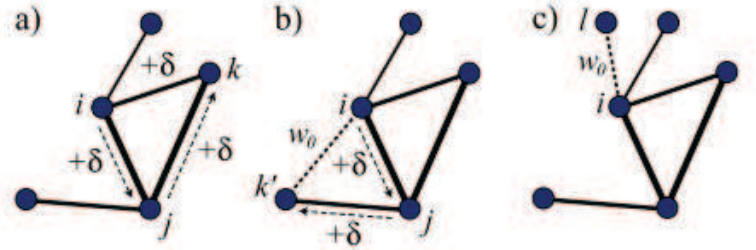

Figure 16: The model algorithm. (a): a weighted local search starts from *i* and proceeds first to *j* and then to *k*, which is a neighbor of *i*. (b): the local search from i

ends to *k'*, which is not a neighbor of i. In this case link $w_{ik'}$ is established with probability p_Δ. (c): node i creates a random link to random node *l* with probability p_r . In cases a) and b) the weights of involved links are increased by δ. The initial weight on a new link is w_o.
The model is defined for a fixed number of nodes. Nodes are eliminated and and put into the system without links with some small probability in order to maintain stationarity. The essential parameter of the model is δ, expresses the fact that relationships get reinforced when used. This reinforcement parameter controls the community formation in the network (Fig. 17).

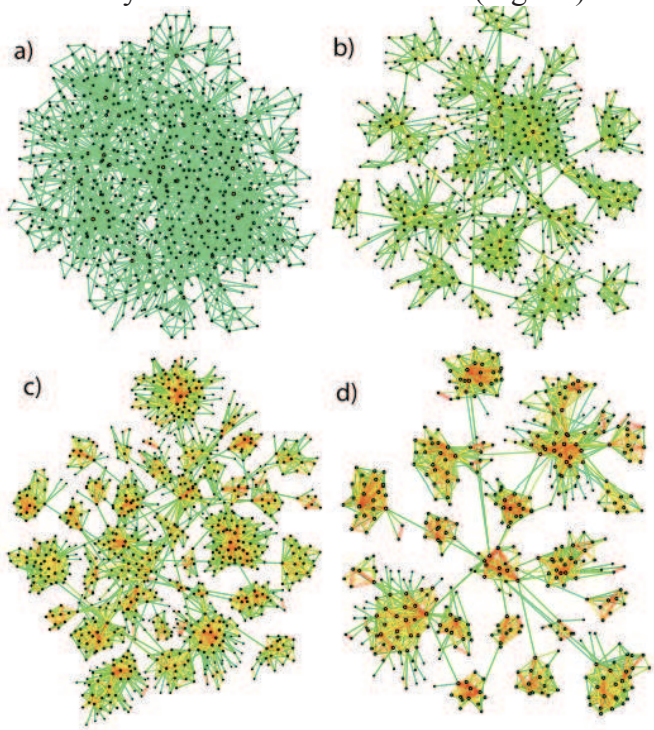

Figure 17. Snowball samples of networks with (a) δ = 0, (b) δ = 0.1, (c) δ = 0.5, and (d) δ = 1. Link colors change from green (weak links) to yellow and red (strong links).

In this model the weights establish a coupling between network structure and interaction strengths: addition of a new link depends on the existing weights, and once a new link is added the weights that led to its formation are strengthened. Communities will emerge only if this strengthening is large enough, i.e., if nodes favor sufficiently their strong connections in the process of establishing new ones. Our study supports the notion that communities result from initial structural fluctuations, which become amplified by repeated application of the microscopic processes. In addition to fulfilling the topological properties of social networks, the model networks exhibit a coupling between network topology and interaction strengths, which is compatible with the weak links hypothesis. We were able to verify that the model has properties, which are in good agreement with our previous empirical studies on degree and strength distributions and percolation properties.

Community structure in school friendship networks

Recently developed concepts and techniques of analyzing complex systems provide new insight into the structure of social networks. Uncovering recurrent preferences and organizational principles in such networks is a key issue to characterize them. We

investigated school friendship networks from the Add Health database [25]. Applying threshold analysis, we found [26] that the friendship networks did not form a single connected component through mutual strong nominations within a school, while under weaker conditions such interconnectedness was present. We extracted the networks of overlapping communities at the schools (c-networks, see Fig. 18) and found that they are scale free and disassortative in contrast to the direct friendship networks, which have an exponential degree distribution and are assortative. Based on the network analysis we studied the ethnic preferences in friendship selection. The clique percolation method we used revealed that when in minority, the students tend to build more densely interconnected groups of friends. We also found an asymmetry in the behavior of black minorities in a white majority as compared to that of white minorities in a black majority (Fig. 19).

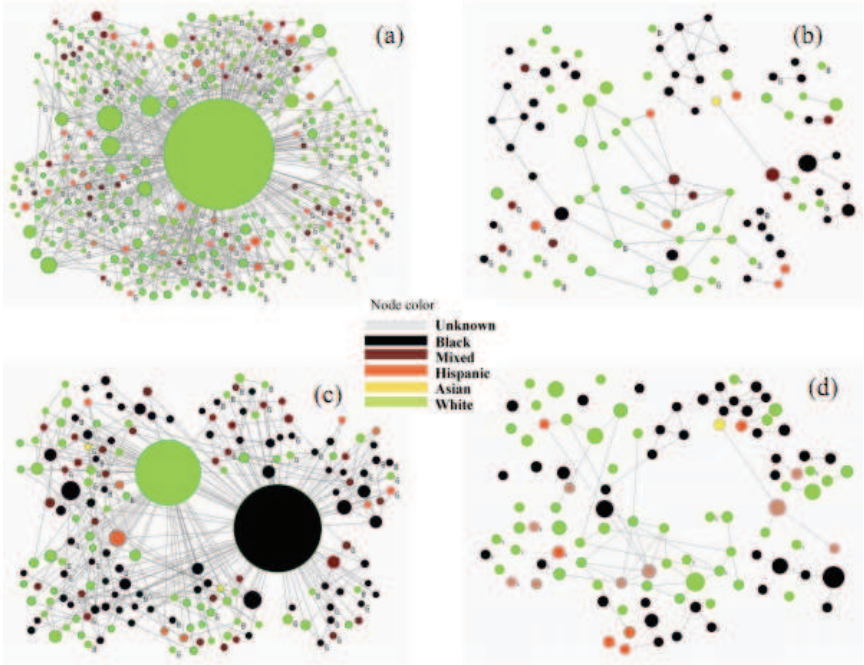

Figure 18. C-networks of 3-clique communities at School 1 (average school, white majority, 70%) (a) and School 2 (blacks overrepresented, 40%) (c). Compared to the corresponding c-networks of 4-clique communities ((b) and (d) respectively). The color is assigned according to the race of the majority of nodes in the community. The node size is proportional to the square root of the number of nodes in the community. Although, each community can have students from different races, we assign to it the color of the majority of the members of the community.

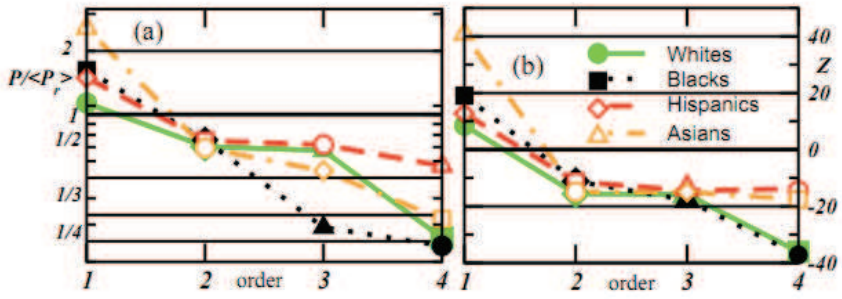

Figure 19. Measuring preferences of inter-racial connections $r - r'$. $P(r, r')$ is the relative frequency of directed links from Whites (full green line), Blacks (dotted black

line), Hispanics (dashed red line) and Asians (dashed-dotted yellow line) to each of the races r′ = W (circles), B (squares), H (diamonds), and A (triangles). Racial preferences manifest themselves as systematic deviations of the ratio $P(r, r')/P_r(r, r')$ from 1, P_r is the corresponding relative frequency in the randomized samples; (a) P/P_r in decreasing order from 1 to 4, for the nominations made from r to r'. (b) The corresponding Z-scores. The combination of (a) and (b) reveals relations $r - r'$ that are significantly absent. The results are the average over the 84 School networks.

Large events in complex systems

Large transportation networks like the road system, pipelines and the electrical power grid are sensitive to local failures. When failures occur the transport on these networks must be redistributed on the still intact part of the network, occasionally exceeding the local capacity and causing further failure. The resulting avalanches may finally end in major breakdowns: mega-jams in vehicular traffic or blackouts in the electrical power system.

We studied [27] the size distribution of power blackouts for the Norwegian and North American power grids. We found that for both systems the size distribution follows power laws with exponents −1.65 ± 0.05 and −2.0 ± 0.1 respectively. We presented a model with global redistribution of the load when a link in the system fails, which reproduced the power law from the Norwegian power grid if the simulations were carried out on the Norwegian high-voltage power grid (Fig. 20). The model was also applied to regular and irregular networks and gave power laws with exponents −2.0 ± 0.05 for the regular networks and −1.5 ± 0.05 for the irregular networks. A presented mean field theory is in good agreement with these numerical results.

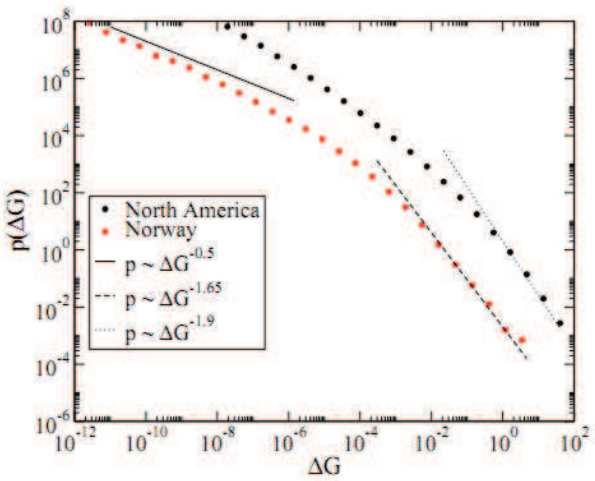

Figure 20. The distribution of the conductance drop ΔG in our model as implemented on the Norwegian (1220 nodes) and North American Power (4941 nodes) grids. The straight lines correspond to the power laws observed in the data.

Large events are crucial in the complex system of finances, where the underlying network is not plausible. We studied large events systematically by using data from the limit order book (LOB). This is the file, where all so called downstairs orders (with amount, price, possible expiration) are put in. Trades are performed when sell and buy orders match. We first studied [28] the data of 12 liquid stocks of the LSE for the period May 2000 to December 2002.

We were interested in the behavior of the book around large price changes. We filtered out such events by relative and absolute filters, averaged the normalized quantities after adjusting the events and examined a number of quantities. e.g., the volatility, the bid-ask spread, the bid-ask imbalance, the number of queuing limit orders, the activity (number and volume) of limit orders placed and canceled, etc. The relaxation of the quantities was generally found to be very slow and could be described by a power law of exponent ≈ 0.4 (Table 2). We introduced a numerical model in order to understand the empirical results better. We found that with a zero intelligence deposition model of the order flow the empirical results can be reproduced qualitatively. This suggests that the slow relaxations might not be results of agents' strategic behavior. Studying the difference between the exponents found empirically and numerically helps us to better identify the role of strategic behavior in the phenomena.

After we got access to the data of the Shanghai stock market, we have repeated the calculations [29]. We found a significant reversal of price for both intraday price decreases and increases with a permanent price impact. The volatility, the volume of different types of orders, the bid-ask spread, and the volume imbalance increased before the extreme events and decay slowly as a power law, which formed a well-established peak. The volume of buy market orders increased faster and the corresponding peak appeared earlier than for sell market orders around positive events, while the volume peak of sell market orders lead buy market orders in the magnitude and time around negative events. When orders were divided into four groups according to their aggressiveness, we found that the behaviors of order volume and order number were

variable	exponent
volatility	0.38 ± 0.01
bid-ask spread	0.38 ± 0.03
limit orders placed - bid	0.37 ± 0.01
limit orders placed - ask	0.40 ± 0.01
cancelations - bid	0.30 ± 0.02
cancelations - ask	0.42 ± 0.02

Table 2. Exponents obtained from fits of the relaxation curves of the corresponding variables.

similar, except for buy limit orders and canceled orders that the peak of order number postponed two minutes later after the peak of order volume, implying that investors placing large orders are more informed and play a central role in large price fluctuations. We also studied the relative rates of different types of orders and found differences in the dynamics of relative rates between buy orders and sell orders and

between individual investors and institutional investors. There was evidence showing that institutions behave very differently from individuals and that they have more aggressive strategies. Combing these findings, we conclude that institutional investors are more informed and play a more influential role in driving large price fluctuations.

Opinion formation and modular structure

One of the main questions in social psychology, which is closely related to network science is: How are opinions formed? This is not only important from a sociological point of view, but it is also highly relevant for politics, innovation spreading, decision making, and the general well feeling of people. This complex process depends on various factors or components like confidence, attitudes, communities or media effects. Recently, much effort has been invested in modeling different aspects of opinion dynamics and these models are in many ways related to those of physics [30]. Unfortunately, the empirical observations are rather sparse. Therefore, the usual strategy is to concentrate on some particular features by making plausible assumptions for a model, and comparing its results with expectations.

In human societies opinion formation is mediated by social interactions, consequently taking place on a network of relationships and at the same time influencing the structure of the network and its evolution. To investigate this coevolution of opinions and social interaction structure we developped a dynamic agent-based network model [31,32], by taking into account short range interactions like discussions between individuals, long range interactions like a sense for overall mood modulated by the attitudes of individuals, and external field corresponding to outside influence. Moreover, individual biases could be naturally taken into account. In addition the model included the opinion dependent link-rewiring scheme to describe network topology coevolution with a slower time scale than that of the opinion formation. With this model comprehensive numerical simulations and mean field calculations were carried out and they showed the importance of the separation between fast and slow time scales resulting in the network to organize as well-connected small communities of agents with the same opinion.

The basic equation of the model is as follows:

$$\frac{dx_i}{dt} = \frac{\partial x_i}{\partial t} + \sum_j \hat{O}(x_i, x_j, g) A_{ij},$$

where x_i is the opinion of individual i, $\hat{O}$ stands for an operator that changes the entries and/or the size of the adjacency matrix, g is the time separation parameter. The fast mode is governed by

$$\frac{\partial x_i}{\partial t} = \alpha_i f_0(\{x_j\}) + f_1(\{x_j\}) x_i + h_i,$$

where α_i is an attitude parameter describing the i-th individual's relation to the overall mood as described by the long range interaction f_0 (giving larger weight to closer neighbors) and f_1 stands for short range interaction. In the simulations we made simplifying assumptions like taking $h = 0$. While sort range interactions are always ferromagnetic ("homophily" in sociology), the attitude can be both ferromagnetic and

anti-ferromagnetic, introducing frustration into the system. Our main finding is that a) the time separation parameter g is relevant; b) there is a regime in g where many communities are formed (see Fig. 21). In a mean field type model we could account for the time evolution of the number of undecided individuals.

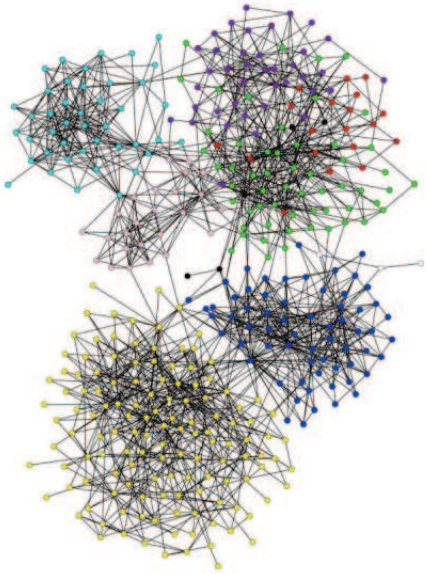

Figure 21. Communities found by algorithm [16]. Different colors mean different communities, irrespective of the opinion value.

Further examination of the model [32] revealed that within the communities the opinions are alike, however, they differ from the point of view of the attitude parameter α. We started the calculations with a homogeneous distribution of α-s. Large communities have positive α-s, while small ones have negative α-s. The message is that everyone likes to get along with his/her closest neighbor but those, who accept overall opinions build larger groups than those, who have aversion to the overall opinion and are more satisfied in smaller groups.

IV. Dynamic analysis

In some of the papers reported above the dynamic aspect was already emphasized [20,22,24,27,28,29]. Here we would like to summarize to somewhat technical papers, which provide tools to dynamic analysis.

Correlations of asynchronous data

The estimation of the correlation between time series is often hampered by the asynchronicity of the signals. Such data are produced in the mobile phone call network or on the stock market. Cumulating data within a time window suppresses this source of noise but weakens the statistics. We presented a method to estimate correlations without applying long time windows [33]. We decomposed the correlations of data cumulated over a long window using decay of lagged correlations as calculated from short window data. This increased the accuracy of the estimated correlation significantly and decreases the necessary efforts of calculations both in real and computer experiments. Here we present only the final formula without derivation:

$$\rho_{\Delta t}^{A/B} = \left(\sum_{x=-n+1}^{n-1} (n-|x|) \left\langle r_{\Delta t_0}^A(t) r_{\Delta t_0}^B(t+x\Delta t_0) \right\rangle - n^2 \left\langle r_{\Delta t_0}^A(t) \right\rangle \left\langle r_{\Delta t_0}^B(t) \right\rangle \right) \times$$
$$\left(\sum_{x=-n+1}^{n-1} (n-|x|) \left\langle r_{\Delta t_0}^A(t) r_{\Delta t_0}^A(t+x\Delta t_0) \right\rangle - n^2 \left\langle r_{\Delta t_0}^A(t) \right\rangle^2 \right)^{-1/2} \times$$
$$\left(\sum_{x=-n+1}^{n-1} (n-|x|) \left\langle r_{\Delta t_0}^B(t) r_{\Delta t_0}^B(t+x\Delta t_0) \right\rangle - n^2 \left\langle r_{\Delta t_0}^B(t) \right\rangle^2 \right)^{-1/2}$$

where $\rho_{\Delta t}^{A/B}$ is the correlation coefficient between time series r^A and r^B, Δt is the time window, n is the maximum number of time steps, Δt_0 is a shorter time scale than Δt. This formula decomposes the correlation coefficient for Δt into lagged correlation functions as obtained for some smaller Δt_0. As the latter usually decay fast, their accurate estimation is easier. In such cases the method is the following: We calculate the lagged correlations on a scale Δt_0 and derive the correlation coefficient from the above formula. Fig. 22 shows the power of this method on two examples.

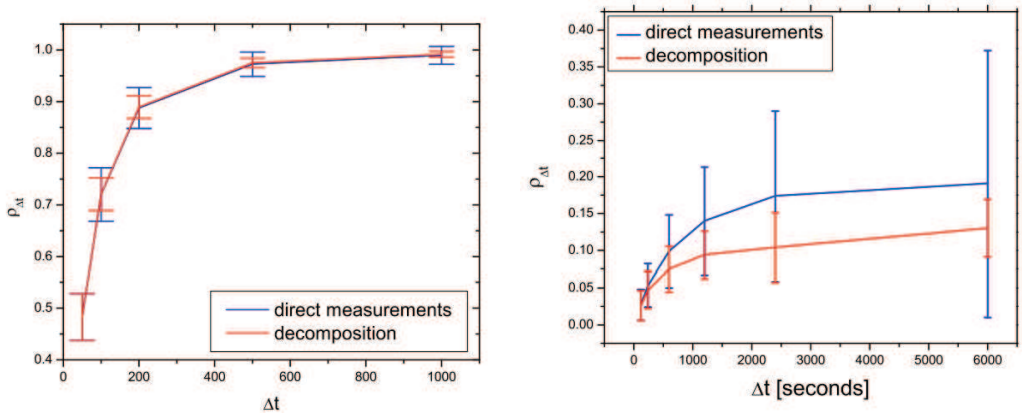

Figure 22. Comparison of the calculation of the correlation coefficient for asynchronous data. a) Generated correlated random walks. b) Stock return data (CO/PEP).

Taylors law and beyond

Complex systems consist of many interacting elements which participate in some dynamical process. The activity of various elements is often different and the fluctuation in the activity of an element grows monotonically with the average activity. This relationship is often of the form "fluctuations ~ average$^{\alpha}$ ", where the exponent α is predominantly in the range [1/2, 1]. This power law has been observed in a very wide range of disciplines, ranging from population dynamics through the Internet to the stock market and it is often treated under the names Taylor's law or fluctuation scaling.

We have investigated fluctuation scaling in several ways [34]. We have shown how general the above scaling relationship is by surveying the literature, as well as by reporting some new empirical data and model calculations. We also showed some basic principles that can underlie the generality of the phenomenon. We presented a mean-field framework based on sums of random variables. In this context the emergence of fluctuation scaling is equivalent to some corresponding limit theorems. In certain physical systems fluctuation scaling can be related to finite size scaling. In some cases the data can be described by multiscaling. The limiting cases (α=1/2 or 1)

can be related to the behavior of independent random variables or synchronized dynimcs. We also discussed fluctuation multiscaling and showed the conditions leading to it.

From technical point of view fluctuation scaling is a tool to see how much the dynamics of the variables in a complex system is coupled together. The limiting case $\alpha=1$ points toward strong coupling, while $\alpha=1/2$ can result from independent random variables. Fig. 23 shows, how fluctuation scaling analysis can be used to identify different mechanisms in a dynamic system.

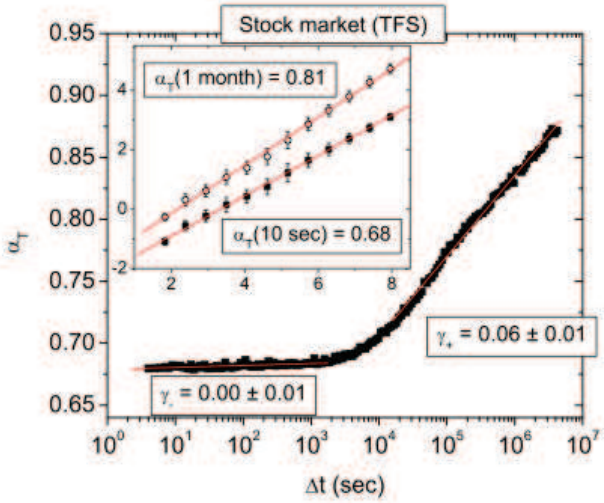

Figure 23. Fluctuation scaling in stock market traded volumes. The scaling exponent α depends on the time window Δt. After a characteristic time (roughly one day) the exponent starts to increase, indicating stronger mutual influence.

V. Summary and outlook

Within this project we have contributed to the theory and applications of network science. We clarified important aspects of weighted networks and community detection problems. We analysed large empirical networks, clarified the relationship between weights and topology in social networks and introduced a new model to describe our findings. We also pointed out the influence between the found effects and the dynamic properties of large social networks. We analysed spreading and concluded that it is far from optimal on such networks: Optimization would require large throughput on links with high betweennes centrality, which have low overlap. We have also developed an efficient technique for calculating correlations of asynchronous data and developed the theory of fluctuation scaling.

During the course of the project we got into contact with two industrial partners, the Morgan Stanley and Nokia-Siemens Hungary. Using our expertise in network theory we could enter into fruitful collaborations with these firms. On the other hand, together with four more scientific centers (Helsinki, Oxford, Turin, Warsaw) we launched an FP7 FET EU project on the interaction fo ICT and society. In the coming years we want to focus on the dynamic aspects of complex systems, including complex networks.

References

[1] http://clinton4.nara.gov/Initiatives/Millennium/shawking.html

[2] Melanie Mitchell: Complexity. A guided tour (Oxford UP, 2009)
[3] Albert-László Barabási: Linked (Pegasus Books, 2002)
[4] Mark E.J. Newman: Networks. An introduction (Oxford UP, 2010)
[5] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, Phys. Rev. E 71, 065103, 2005
[6] Gergely Tibély, Jukka-Pekka Onnela, Jari Saramaki, Kimmo Kaski and János Kertész, Physica A370, 145-50, 2006
[7] S. Fortunato, Phys. Rep. 486, 75-174, 2010
[8] U. N. Raghavan, R. Albert, and S. Kumara, Phys. Rev. E 76, 036106, 2007
[9] G. Tibély and J. Kertész, Physica A387 4982-4984, 2008
[10] M.E.J. Newman, M. Girvan, Phys. Rev. E. 69, 026113, 2004
[11] S. Fortunato, M. Barthélemy, PNAS 104, 36-41, 2007
[12] J. Reichardt, S. Bornholdt, Phys. Rev. E 74, 016110, 2006
[13] J.M. Kumpula, J.Saramaki, K.Kaski, J.Kertész, EPJB 56, 41-45, 2007
[14] J.M. Kumpula, J. Saramäki, K. Kaski, J. Kertész, Fluctuation and Noise Letters 7, L209-14, 2007
[15] Palla G, Derényi I, Farkas I and Vicsek T, Nature 435, 814, 2005
[16] A. Lancichinetti, S. Fortunato, J. Kertész, New J. Phys. 11, 033015 (2009)
[17] E. Ravasz, A-L. Barabási, Phys. Rev. E 67, 026112, 2003
[18] D. Nagy, G. Tibély and J. Kertész, FRACTALS 14, 101-110, 2006
[19] M. Granovetter, Am J Sociol 78: 1360-1380, 1973
[20] J.-P. Onnela, J. Saramäki, J. Hyvonen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, PNAS 104, 7332-7336, 2007
[21] Zs. Szvetelszky, The Self Organizing Complex Network of Human Communication. 24th EGOS Colloquium (European Group for Organizational Studies, 2008).
[22] J.M. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, J. Kertész , Phys. Rev. Lett. 99, 228701, 2007
[23] G. Kossinets and D. J. Watts, Science 311, 88, 2006.
[24] J.M. Kumpula, J.P. Onnela, J. Saramaki, J. Kertész, K. Kaski, Comp. Phy. Com. 180, 517-22, 2009
[25] http://www.cpc.unc.edu/addhealth
[26] M. C. González, H.J. Herrmann, J. Kertész, and T. Vicsek, Physica A 379, 307-316. 2007
[27] J. O. H. Bakke, A. Hansen, J. Kertész, Europhys. Lett. 76, 717-723, 2006
[28] Bence Tóth, J. Kertész and J. Doyne Farmer, Eur. J. Phys. B 71, 499-510 (2009)
[29] Guo-Hua Mu, Wei-Xing Zhou, Wei Chen, János Kertész, New J. Phys. 12 075037, 2010
[30] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. 81, 591, 2009
[31] Gerardo Iñiguez, János Kertész, Kimmo Kaski, Rafael A.Barrio, Phys. Rev. E 80, 066119, 2009
[32] G. Iñiguez, R. A. Barrio, J. Kertész, K. Kaski, Comput. Phys. Com. 180, 517-22, 2009
[33] B. Tóth, J. Kertész, Physica A 389, 1696-1705, 2009
[34] Z. Eisler , I Bartos and J. Kertész., Advances in Physics 57, 89-142, 2008