Statistical physics of the collective behaviour of organisms

OTKA: K49674 final report by Tamás Vicsek (PI)

In the original research plan proposed about 5 years ago 3 major topics were described including a number of individual sub-projects. The main goals have been achieved and published in high profile international journals. Some of the originally planned sub-projects have not been carried out because they have been substituted by a number of newer ones found, as the investigations progressed, to be more promising. In the following I shall summarize the results achieved point by point in accordance with the originally planned three major directions all involving the applications of statistical physics methods to phenomena common in life. At the end I also briefly overview our results which do not fit well any of the major directions (but are related and gave rise to perspective developments in our work).

1. Collective motion.

Remarkably, just after we had started our research on flocking (collective motion) in the framework of OTKA in various vaguely related branches of research (such as control theory, robotics, ethology and statistical physics), a great increase in the level of interest in this topic occurred.

Introduction (motivation, basic observations and questions)

Most of us must have been fascinated by the eye catching displays of collectively moving animals. Schools of fish can move in a rather orderly fashion or change direction amazingly abruptly. Under the pressure from a nearby predator the same fish can swirl like a vehemently stirred fluid. Flocks of hundreds of starlings can fly to the fields as a uniformly moving group, but then, after returning to their roosting site, produce turbulent, puzzling aerial displays. There are a huge number of further examples both from the living and the non-living world for the rich behaviour in systems consisting of interacting, permanently moving units. Although the advantages of collective motion have not been completely explored yet, it is likely that this behaviour is optimal from several points of view including defense from predators, foraging or collective decision making in further situations. Persistent motion is one of the conspicuous features of life, but recently several physical and chemical systems have also been shown to possess interacting, "self-propelled" units.

Are these observed motion patterns very much system specific? We are quite used to such a conclusion in biology. Or, alternatively, are there only a few typical classes into which all of the collective motion patterns can be classified? This would be a familiar thought for a statistical physicist dealing with systems of an enormous number of molecules in equilibrium. In fact, collective motion is one of the manifestations of a more general class of phenomena called collective behavior [1]. The studies of the latter have identified a few general laws related to how new, more complex qualitative features emerge as many simpler units are interacting [2].

There is an amazing variety of systems (including examples from physics, chemistry, but mainly from biology) of such units bridging over many orders of magnitude in size systems. Will they still exhibit the same motion patterns? If yes, are there any underlying universal principles predicting that this has to be so (e.g., non-conservation of moments during interactions).

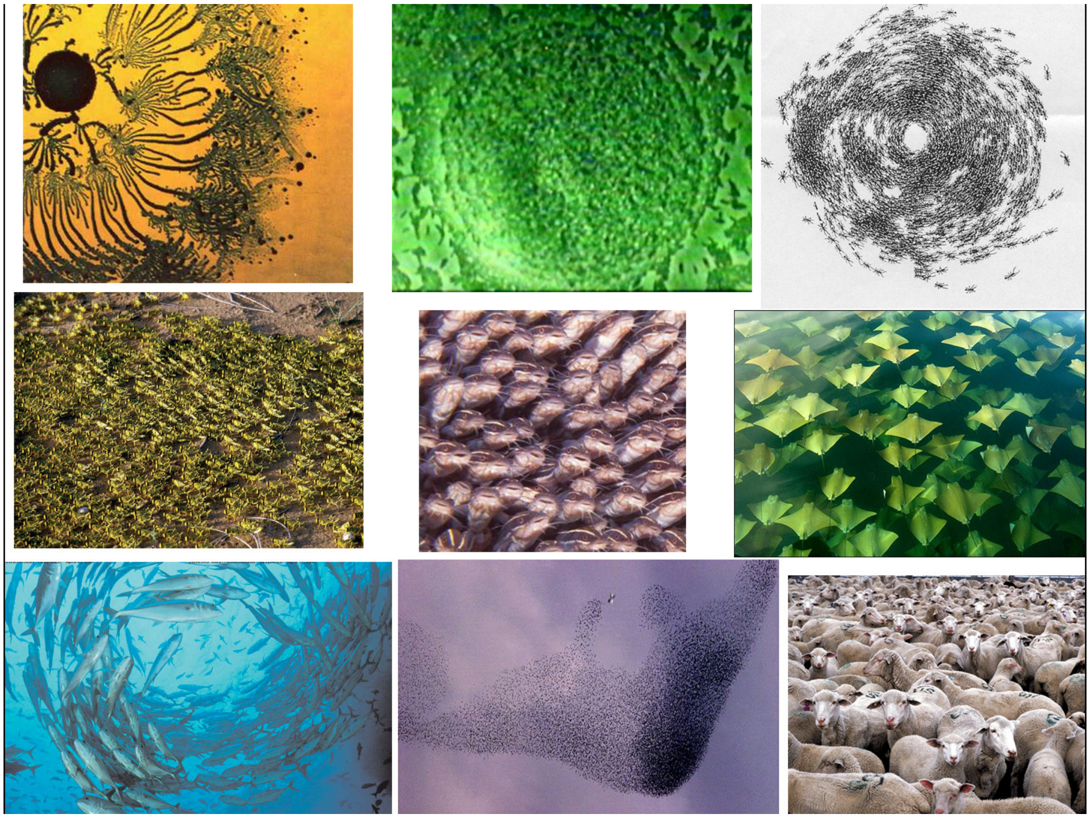


Fig. 1. A gallery of images related to collective behaviour. Among others, it illustrates the possible existence of general behavioral patterns. *First row from left to right***:** A bacterial colony growing as the black spots (containing thousands of bacteria rotating around the centre of the disk) “glide” outward and leave a trace of additional bacteria behind. Zooming on one spot on a) thousands of bacteria moving in an orderly fashion (along circles) can be seen. A rotating colony of South American fire ants. *Second row*: Wingless locusts are marching in the field**.** A particularly crowded school of fish bounded to move orderly. A nice three-dimensional array of Golden Rays. *Third row*: Fish are known to produce such vortices. Before roosting, thousands of starlings are taking part in a fascinating aerial display. They are also trying to avoid a predator bird close to the central finger-like structure. A counter-example: Although sheep are known to move very coherently, just as our theory predicts, when simply hanging around (no motion), well developed orientational patterns cannot emerge.

Results, experimental

Collective motion of cells

Our experimental setup allowed us to make long-term videomicroscopy of keratocytes together with a quantitative analysis of their motion [3]. *Cell Culture:* 2–4 fish scales were collected from living 5– 15 cm long goldfish *(Carassius auratus)* with tweezers, and placed external side up in a 35-mm Petri dish. After a few seconds scales were flood with 1 ml culture medium: RPMI 1640 supplemented with 10% fetal calf serum (Gibco), 40 ^{I}g/ml

Gentamycin (Gibco), 0.5 ^{1}g/ml Amphotericin B (Sigma). Scales were kept at room temperature in 5% CO2 atmosphere overnight to let epidermal keratocyte cells migrate out from the scales. Before timelapse microscopy, scales were removed, and cultures were flushed with fresh medium. To gain low density cultures we applied phosphate saline buffer (PBS) for 5–10 min and/or 0.25% Trypsin (Gibco) for 1–2 min. *Time lapse microscopy:* We used a home developed fully computer controlled time-lapse microscope, which enabled us to monitor cell motility up to four cultures in several fields of view with a frequency of 1 shot per min. During the typically 24–h long time-lapse microscopy cells were kept in a room temperature custom made mini-incubator to maintain 5% CO2 atmosphere and avoid steaming up. *Data processing and analysis*: In each time-lapse frame 20–30 representative cells were randomly selected from the population. Positions of the cells were determined manually along 100–200 frames by using a custom written software. This process resulted in the r_{ij} positon vectors of cells, where i denotes the index of the cell, $j\,\Delta t$ is the elapsed time from the beginning of the cell's trajectory. Δt stands for the time difference between the frames.

Our *experimental results* are based on monitoring the cell migration from very low cell densities to confluent (nearly full coverage) cultures. A relatively sharp transition could be observed from the random motility in sparse cultures to the ordered collective migration of dense islands of cells. Fig. 1 shows the typical behavior of cells in case of three different densities.

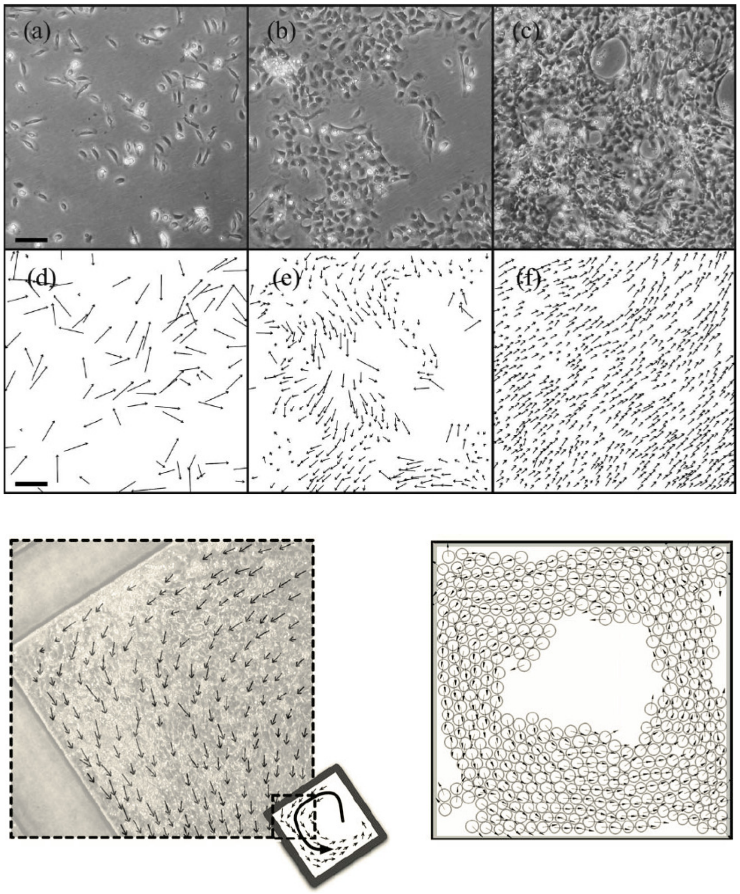


FIG. 2. Top: Phase contrast images showing the typical behavior of cells for three different densities. (a) 1.8, (b) 5.3, (c) 14.7 cells/100x100 μm^2. We observed that as cell density increases cell motility

undergoes collective ordering. The speed of single cells is higher than that of cells moving in coherent groups. Scale bar 200 μm. (d–f) Velocity of cells. Scale bar 50 μm/min. Bottom: Experimental snapshot of a circulating cell group with the velocities of cells in the corner of a 2 x2 mm^2 square shaped microfabricated arena. Right: Simulations of model cells confined to a square arena show the emergence of circular motion over a wide range of model parameters.

Thus, as the overall density of the migrating cells was increased, we observed a kinetic phase transition from a disordered (low density) into an ordered (at larger densities) state. Close to the transition a complex picture emerges with interacting clusters of cells moving in groups. We also introduced a novel flocking model in which no assumption about the knowledge of the direction of motion of the neighbors was incorporated. In order to interpret our experimental results, our model assumes repulsive and adhesive forces only and leads to results being in good agreement with the observations (more details are given in the section on modelling).

Flocking of pigeons

We have started another novel series of experiments on the group flight of homing pigeons [4].

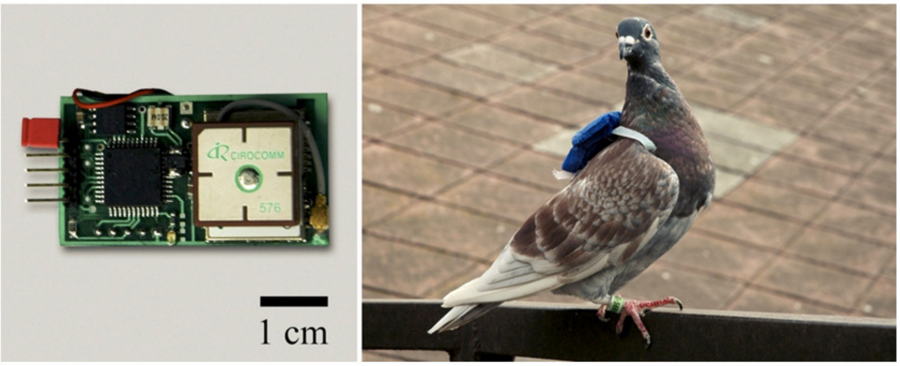


Fig. 2. Left: Our new generation, ultra light (16gram) GPS device. Right: the device is in the small blue backpack attached to the pigeon with the help of a special ribbon.

Animals moving around in groups display various fascinating motion patterns thought to be a result of intricate communication processes among the group members. Phenomena like splitting of a group, a collective sudden change in the direction of motion, or, in the case of birds, a synchronized landing are all signs of a quick collective decision making by the members of the group typically on a very short time scale. Although the most plausible way of investigating and interpreting these phenomena would be the collection of high resolution spatiotemporal data for the moving individuals, such data are extremely scarce and are virtually non-existent for three dimensional group motion because of the technological difficulties involved. Below we present results of experiments in which track logs a dozen of homing pigeons flying in a flock have been obtained by lightweight GPS devices and analyzed using a variety of correlation functions inspired by approaches common in statistical physics. We find a well defined hierarchy among the pigeons from data concerning leading roles in pair wise interactions. The average spatial position of a pigeon within flock strongly correlates with its place in the hierarchy. However, overall hierarchical relations can also be violated on a daily basis.

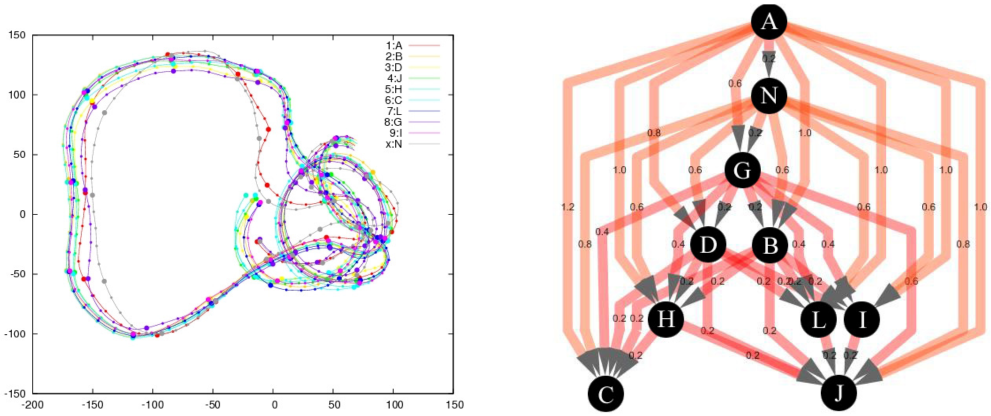


Figure 3: Left: Two minutes long part from the flight of birds flying around their loft. The smaller and the larger dots indicate every 1s and 5s, respectively. Each path begins near the center of the plot. Right: Hierarchical network of the flock for the same day as on left panel. A directed edge goes from the leader to the follower; the numbers on the edges show the time difference of their motion. Individual birds are denoted by capital letters.

Results, modelling

During the last 5 years a sudden increase of interest in our original 1995 SPP (Self-Propelled Particles) model [5] has occurred (resulting in several hundred new citations). This has also motivated us to become part of the renewed investigations of some of the basic aspects of the model.

The individual based behavioral rules, determining collective motion, are of particular interest. A frequent assumption in models is that the information, "perceived" by the particles, is restricted to the velocity of their neighbors. The assumption of reflecting only on the momentary velocity may not be enough for adequately describing a number of biologically relevant situations. We expect that the behavior of the SPP model is significantly extended if we also incorporate a term corresponding to memory on short time scales. This can be achieved by introducing an acceleration term into the equations (rules), corresponding to separating time scales by assuming that the particles differentiate between two kind of information, the first being their actual velocity, the second one corresponding to recent change in their direction of motion. For example, in the case of birds, reacting to acceleration may mean that birds can give signals to their neighbors about their intended changing of their flight direction by quickly modifying their velocity.

Thus, we have considered a three dimensional, generalized version of the original SPP model for collective motion [6]. By extending the factors influencing the ordering, we have investigated the case when the movement of the self-propelled particles (SPP-s) depends on both the velocity and the acceleration of the neighboring particles, instead of being determined solely by the former one. By changing the value of a weight parameter s determining the relative influence of the velocity and the acceleration terms, the system undergoes a kinetic phase transition as a function of a behavioral pattern. Below a critical value of *s* the system exhibits disordered motion, while above it the dynamics resembles that of the SPP model. We argue that in nature evolutionary processes can drive the strategy variable **s** towards the critical point, where information exchange between the units of a system is maximal.

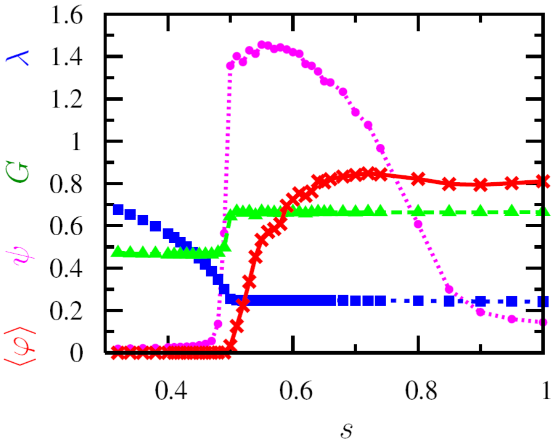


Fig. 4. The order parameter (φ, x marks), the Binder cumulant (G, triangles), the information exchange between particles (ψ , circles) and the average curvature of trajectories (λ, squares) as a function of the strategy variable *s*.

We found that the information exchange between particles was maximal at the critical point. Due to the important role of information exchange in animal societies, this might indicate that the critical point corresponds to an optimal behavioral strategy. In a more general context, this result implies that biological evolution may drive individual traits corresponding to critical values.

Intrigued by the possibility of finding density waves in the SPP model, we re-investigated the small velocity regime at larger system sizes and significantly longer simulation times [7]. The flocks have reached their steady sizes. The nature of the disorder–order phase transition was characterized by the probability distribution function (PDF) of the order parameter (average particle velocity). As shown in Fig. 5, the PDF was one humped, signaling a second order phase transition in accord with the earlier results. Furthermore, we also determined the corresponding Binder-cumulant G. We found that G did not exhibit a significant minimum, corroborating the second order nature of the phase transition. On the other hand, the density waves, described recently in the large velocity regime but claimed to exist for all velocities) did not occur in the small velocity regime.

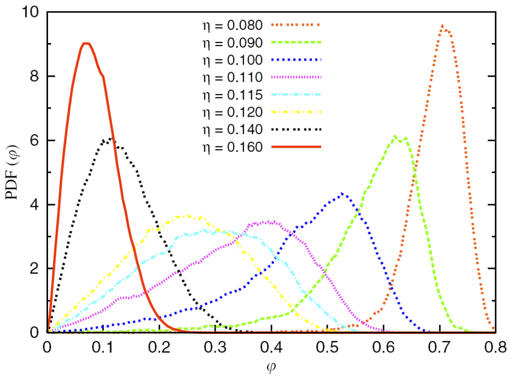


Fig. 5. The PDF of the order parameter in the small velocity regime for noise values around the critical point. The one humped character of the curves demonstrates a second order phase transition. The curves were obtained for systems with linear system size L=512, and we used more than 10 times longer simulational times than the previously recorded relaxation time.

We have also considered disorder-order phase transitions in the three-dimensional version of the original SPP model [8]. Our results show similarities with the findings in two dimensions.

For small velocity ($v < 0.1$) a continuous, second order phase transition is observable, while the diffusion of nearby particles is isotropic. By increasing the particle velocities the phase transition changes to first order, and the diffusion becomes anisotropic. The first order transition in the latter case is probably caused by the interplay between anisotropic diffusion and periodic boundary conditions, leading to an artificial symmetry breaking of the solutions.

2. Synchronization

Here we consider the adjustment of the phase of a behavioural pattern (displaying activity for a short time) as a special kind of synchronization.

Similarly to collective motion, the number of participants in collective opinion formation and decisions is often very large; a key factor is interaction (influence and imitation) between the participants, which strongly reduces the number of possible global patterns suggesting that the number of relevant parameters is small. Statistical physics based models have been successfully applied in the analysis of collective opinion formation and decisions as well. Unanimous and undecided election results and cooperation phenomena have been described by models containing particles with a small number of allowed states and simple rules of interaction plus external fields. Surprisingly, within the same modelling framework one can explain the shape of the transition measured for the cumulated binary decisions of millions of humans in several further collective decision processes of high public interest, such as birth rate and cell phone purchases. Here we consider the Mexican wave (also called La Ola) as a simple example for spatiotemporal behavioural pattern exhibited by tens of thousands of people.

The Mexican wave is launched by a small group of people, each of them standing up within a short time interval, raising their hands high above their heads and then sitting down. As this motion is repeated consecutively by groups of close neighbors, within a few seconds a stable, linear wave with constant amplitude, width and speed develops. The mechanism of this rapid self-organizing process can serve as a paradigm for situations involving limited interaction and the selection of one option out of a small number of possible choices, e.g., route choice behaviour in vehicle or the selection of exits during pedestrian escape panic. While the Mexican wave rolls, spectators try to predict when the (nearest) wave will arrive at their seats and leap to their feet at that moment. During stationary propagation around the stadium, both the wave and its velocity are well-defined and spectators can easily synchronize themselves to the wave's arrival time.

During the short time interval of the initiation, however, the wave's direction is not yet known. In principle, after initiation, the wave should travel in both directions. However, all video recordings available to us show and an overwhelming majority of our online visitors report that already short after the triggering event only one wave is present. Therefore, an intriguing question is how one of the two waves is suppressed and the other is selected so rapidly. According to the excitable media-type model we developed to address this problem [9] the key effect in selecting the wave's direction so quickly is a long range interaction: if the active region (perturbation, wave) is moving towards (away from) an excitable unit (person), then this will make the activation of that unit more (less) likely.

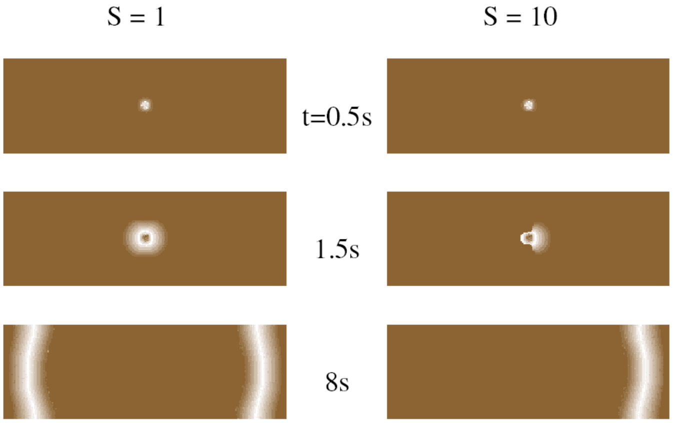


Fig. 6. Spontaneous symmetry breaking in the Mexican wave simulations: each spectator is represented as one particle in a rectangular lattice. Shown are parts of the simulation area at 0:5s, 1:5s and 8s after the triggering event. Excitable particles are colored dark. The increasing color brightness of active particles represents the different stages while standing up, the decreasing color brightness of the first 10 refractory states represents the stages of sitting down, while the dark color of the remaining 10 refractory states indicates that the person is already sitting, though not yet activable again. Left column: If the control parameter, *S* (which is the relative weight of global interactions), is low, the stable solution contains two waves moving in opposite directions. Right column: At higher values of the control parameter the asymmetric solution (one wave moving either left or right) becomes stable.

3. Complex networks

A widespread approach to the analysis of complex natural, social and technological phenomena is to assemble the participating molecules, individuals or electronic devices and their interactions into a network (nodes and links) and to infer functional characteristics of the entire system from this static web of connections. This approach is rooted in, among others, statistical physics, where often the thermodynamic limit ($N \rightarrow \infty$, where N is the number of nodes) is considered, and the overall (large-scale) structure of connections is studied rather than the details at the level of nodes and links. Accordingly, over the past few years, several broadly studied large- scale properties of real-world webs have been uncovered, e.g., a low average distance combined with a high average clustering coefficient, the broad (scale-free) distribution of node degree (number of connections of a node) and various signatures of hierarchical/modular organization [10]. In addition, detailed analyses of the small- scale behaviour of the same complex webs have revealed overrepresented local structures: graph motifs, i.e. , small groups of nodes (typically of size 3–5) with specifically arranged connections among them. The identified small- and large-scale properties are both closely related to the dynamical behaviour of the corresponding complex system. Nodes with many connections (hubs) often have a central role in traffic, while motifs act as building blocks performing distinct basic information processing tasks.

Clusters/modules

In recent major contributions to the exploding literature on the subject of community finding, it has been demonstrated that most real networks, ranging from protein interaction and gene expression networks to groups of people, typically contain parts in which the nodes (units) are significantly more highly connected to each other than to the rest of the network. The sets of such nodes are usually called clusters, communities, cohesive groups, or modules, with no widely accepted, unique definition. In spite of this ambiguity the presence of communities in networks is a relevant and informative signature of the hierarchical nature of complex systems. The existing efficient methods for finding communities in *large* networks are useful if the community structure is such that it can be interpreted in terms of separated sets of communities. However, most real networks have a complex community structure characterized by well defined statistics of *overlapping and nested communities*. In fact, such a statement can be best demonstrated by the numerous communities each of us belongs to, including those related to our scientific activities or personal life (school, hobby, family) and so on, as illustrated in Fig.7. Furthermore, members of our communities have their own communities, resulting in an extremely complicated web of the communities themselves. This has long been appreciated by sociologists, but has never been studied systematically for large networks. Another, biological example is that a large fraction of proteins belong to several protein complexes simultaneously, indicating that any kind of grouping of proteins has a natural tendency to correspond to an overlapping network structure.

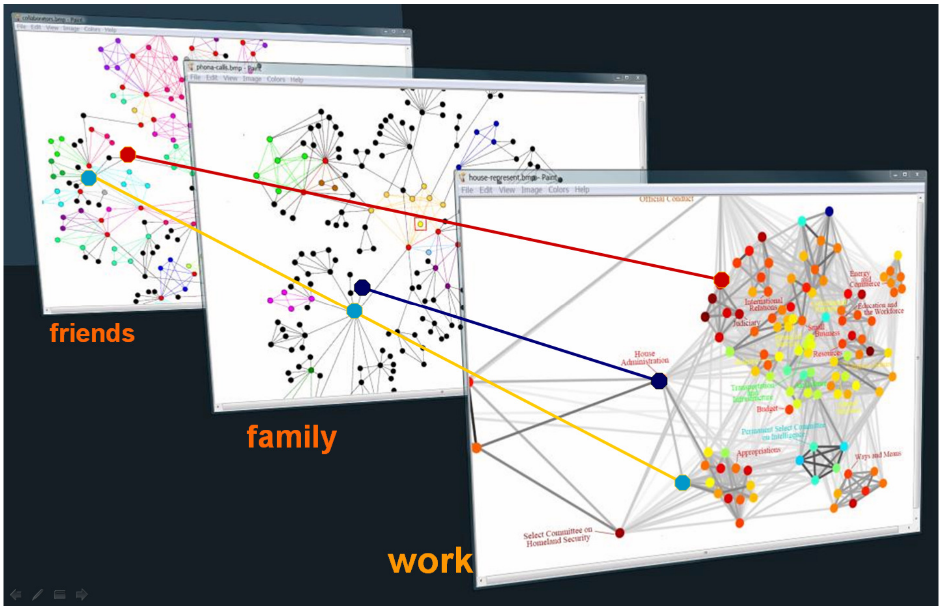


Fig. 7. Illustration of the possible complexity of networks. Each of us has several sub-networks made of overlapping communities (more densely connected sub-units) and these sub-networks of our full network of social ties also overlap (a colleague can be a friend or a relative, etc.)

We have proposed a link-density based approach to finding overlapping communities efficiently and named it as the Clique Percolation Method (CPM) [11]. In this approach, the definition of the modules is based on k-cliques (complete subgraphs of size k in which each node is connected to every other node). A k-clique is a sub-graph with maximal possible link density, therefore it is a good starting point for defining modules. However, a method accepting only complete sub-graphs as modules would be too restrictive. Therefore, k-cliques are “loosen up” in the following way. Two k-cliques are said to be adjacent if they share k−1 nodes (or in other words, if they differ only in a single node), and a module is defined as the union of k-cliques that can be reached from each other through a series of adjacent k-cliques. Such modules can be best visualized with the help of a k-clique template (an object isomorphic to a complete graph of k vertices). Such a template can be placed onto any k-clique in the graph, and rolled to an adjacent k-clique by relocating one of its vertices and keeping its other k – 1 vertices fixed. Thus, the k-clique modules (k-clique communities) of a graph are all those subgraphs that can be fully explored by rolling a k-clique template in them, but cannot be left by this template, as illustrated in Fig.8. The algorithm used for the implementation of this technique is very efficient for most real networks, and provides the full list of overlapping modules in a short amount of time. We have developed a stand-alone, user-friendly, freely downloadable software for finding overlapping communities using CPM [12].

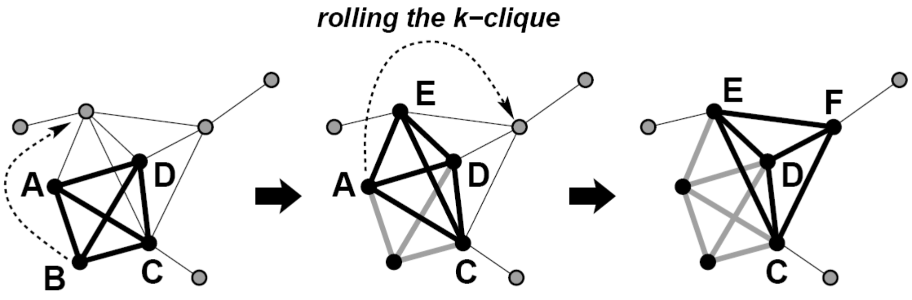
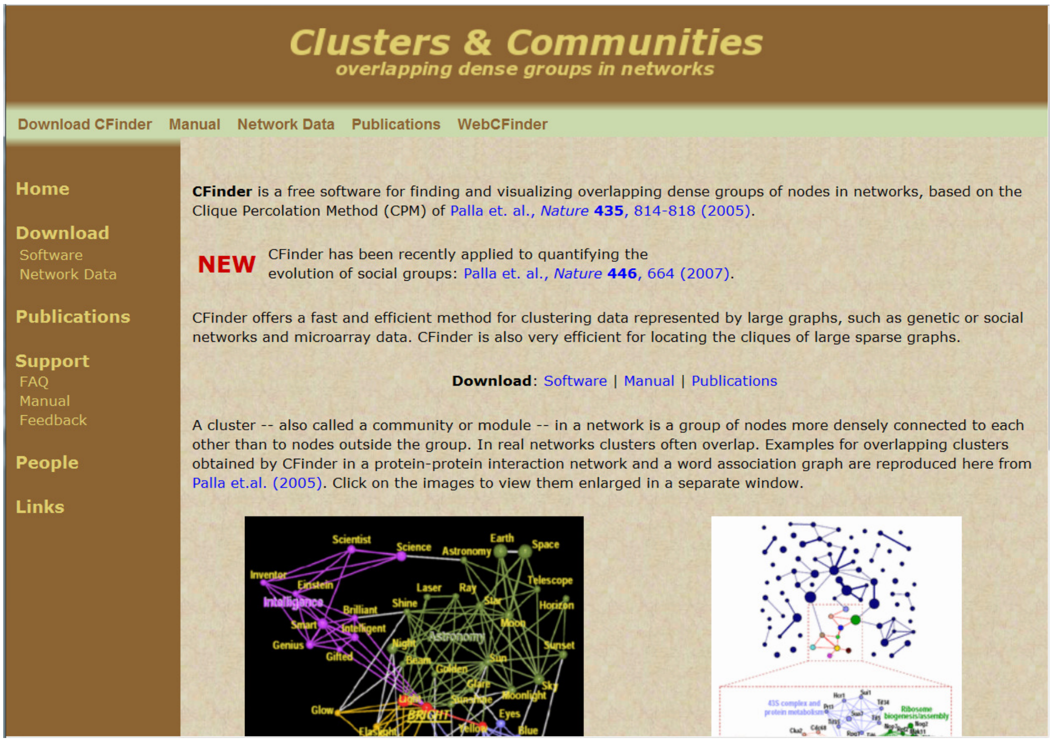


Fig. 8. Illustration of the Clique Percolation Method (CPM) with k-clique template rolling in a small undirected graph for k = 4. Initially the template is placed on A-B-C-D (left panel) and it is "rolled" onto the subgraph A-C-D-E (middle panel). The position of the k-clique template is marked with thick black lines and black nodes, whereas the already visited links are represented by thick gray lines and dark gray nodes. Observe that in each step only one of the nodes is moved and the two 4-cliques (before and after rolling) share k – 1 = 3 nodes. At the final step (right panel) the template reaches the subgraph C-D-E-F, and the set of nodes visited during the process (A-B-C-D-E-F) are considered as a module identified by the CPM at k = 4.



Clusters & Communities

overlapping dense groups in networks

CFinder is a free software for finding and visualizing overlapping dense groups of nodes in networks, based on the Clique Percolation Method (CPM) of Palla et. al., *Nature* **435**, 814-818 (2005).

NEW CFinder has been recently applied to quantifying the evolution of social groups: Palla et. al., *Nature* **446**, 664 (2007).

CFinder offers a fast and efficient method for clustering data represented by large graphs, such as genetic or social networks and microarray data. CFinder is also very efficient for locating the cliques of large sparse graphs.

Download: Software | Manual | Publications

A cluster -- also called a community or module -- in a network is a group of nodes more densely connected to each other than to nodes outside the group. In real networks clusters often overlap. Examples for overlapping clusters obtained by CFinder in a protein-protein interaction network and a word association graph are reproduced here from Palla et.al. (2005). Click on the images to view them enlarged in a separate window.

Fig. 9. Since our clustering method turned out to be one of the (if not the) most efficient procedures of its kind giving a more complete information about the modular structure of huge real-life (i.e., sparse) graphs, we decided to make it available for both the potential scientific and the profit oriented users.

The software (CFinder) has been downloaded since 2005 several thousand times and used in dozens of publications to uncover the underlying community structure in a wide variety of networks.

Using CPM allows us the determination a set of new characteristic quantities for the statistics of communities. We have found that overlaps are indeed very significant, and the distributions we introduce reveal novel universal features of networks. Our studies of collaboration, word association, and protein interaction graphs demonstrated that the *web of communities* has highly non-trivial correlations and specific scaling properties.

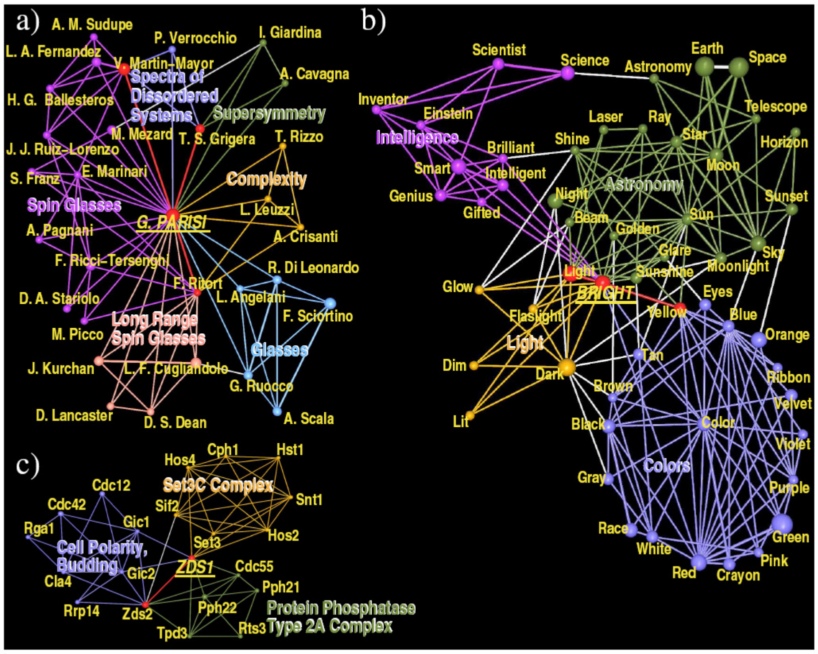


Fig.10. Demonstration of the outcomes of our module finding algorithms. Local network communities around selected nodes are shown. Since the co-authors of G. Parisi, the meanings of the word BRIGHT and the functions of many of the proteins related to the one called ZDS1 are approximately known, these sub-networks can serve as direct tests for the effectiveness of the method.

The scaling of the distributions we find is a novel signature of the hierarchical nature of the systems we study. It means that if we consider a network of communities instead of the nodes themselves, we still observe a degree distribution with a fat tail, but a characteristic scale is introduced, below which the distribution is exponential. This is in agreement with our understanding of a complex system having different levels of organisation with units specific to each level. In addition, in the present case the principle of organisation (scaling) is preserved (with some specific modifications) when going to the next level in good agreement with the recent finding of the self-similarity of many complex networks.

To uncover the subtleties of many social networks we have also investigated the school friendship networks from the Add Health database (data from 83 USA high schools). Applying threshold analysis, we found that the friendship networks do not form a single connected component through mutual strong nominations within a school, while under weaker conditions such interconnectedness is present [13]. We extracted the networks of overlapping communities at the schools 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.

We have also introduced a clustering algorithm (CPMw, Clique Percolation Method with weights) for weighted networks based on the concept of percolating k-cliques with high enough intensity [14]. The algorithm allows overlaps between the modules. First, we gave detailed analytical and numerical results about the critical point of weighted k-clique percolation on (weighted) graphs. Then for a scientist collaboration web and a stock correlation graph we computed three-link weight correlations and, with the CPMw the weighted modules. After reshuffing link weights in both networks and computing the same quantities for the randomised control graphs as well, we showed that groups of 3 or more strong links prefer to cluster together in both original graphs.

A search technique locating network modules, i.e. , internally densely connected groups of nodes in directed networks has also been introduced by us extending the Clique Percolation Method originally proposed for undirected networks [15]. After giving a suitable definition for directed modules we investigated their percolation transition in the graph both an Erdős-Rényi analytically and numerically. We also analysed four real-world directed networks, including Google's own web-pages, an email network, a word association graph and the transcriptional regulatory network of the yeast *Saccharomyces cerevisiae*. The obtained directed modules were validated by additional information available for the nodes. We found that directed modules of real-world graphs inherently overlap and the investigated networks can be classified into two major groups in terms of the overlaps between the modules. Accordingly, the word association network and Google's web pages the overlaps are likely to contain in-hubs, whereas the modules in the email- and transcriptional regulatory network tend to overlap via out-hubs.

We also studied the centrality properties of directed module members in social networks [16].

Percolation of complete subgraphs

We have addressed the general question of subgraph percolation in the E-R model. We obtained analytic and simulation results related to the appearance of a giant component made of complete subgraphs of k vertices (k-cliques) [17]. In particular, we provided an analytic expression for the threshold probability at which the percolation transition of k-cliques takes place. The transition is continuous, characterized by non-universal critical exponents, which depend on both k and the way the size of the giant component is measured. Our analytic calculations are in full agreement with the corresponding numerical simulations.

When the probability p of two nodes being connected is above a certain threshold $p_c(k)$, the complete subgraphs of size k (the k-cliques) are organized into a giant cluster. By making some assumptions that are expected to be valid below the threshold, we determined the average size of the k-clique percolation clusters, using a generating function formalism [18].

Tagged networks

We investigated the fundamental statistical features of tagged networks (also called as annotated networks) having a rich variety of attributes associated with their nodes [19]. Tags (attributes, annotations, properties, features, etc.) provide essential information about the entity represented by a given node and taking them into account is an important step towards a more complete description of the structure of large complex systems. Our main goal here was to uncover the relations between the statistical properties of the node tags and those of the graph topology. In order to better characterize the networks with featured nodes, we introduced a number of new quantities including tag-assortativity (being large if two very similar nodes are linked with an above the average probability) and node uniqueness (which is large if a node has only a few very rarely occurring tags). We appled our approach to three large networks representing very different domains of complex systems. A number of the tag-related quantities display analogous behaviour (e.g., the networks we studied are tag-assortative, indicating possible universal aspects of tags versus topology), while some other features, like the distribution of the node uniqueness shows variability from network to network allowing for pin-pointing large scale specific features of real-world complex networks.

Evolving networks

The graph of communities is a network emerging above the level of individual nodes in the hierarchical organisation of a complex system. In this graph the nodes correspond to communities (highly interconnected subgraphs, also called modules or clusters), and the links refer to members shared by two communities. Our analysis indicated that the development of this modular structure is driven by preferential attachment, in complete analogy with the growth of the underlying network of nodes [20]. We studied how the links between communities are born in a growing co-authorship network, and introduced a simple model for the dynamics of overlapping communities.

Thanks to frequent changes in the activity and communication patterns of individuals, the associated social and communication network is also subject to constant evolution. Our knowledge of the mechanisms governing the underlying community dynamics is limited, but is essential for a deeper understanding of the development and self-optimization of society as a whole. We have developed an algorithm based on clique percolation that allows us to investigate the time dependence of overlapping communities on a large scale, and thus uncover basic relationships characterizing community evolution [21]. Our focus was on networks capturing the collaboration between scientists and the calls between mobile phone users. We found that large groups persist for longer if they are capable of dynamically altering their membership, suggesting that an ability to change the group composition results in better adaptability. The behaviour of small groups displays the opposite tendency—the condition for stability is that their composition remains unchanged. We also showed that knowledge of the time commitment of members to a given community can be used for estimating the community's lifetime. These findings offer insight into the fundamental differences between the dynamics of small groups and large institutions.

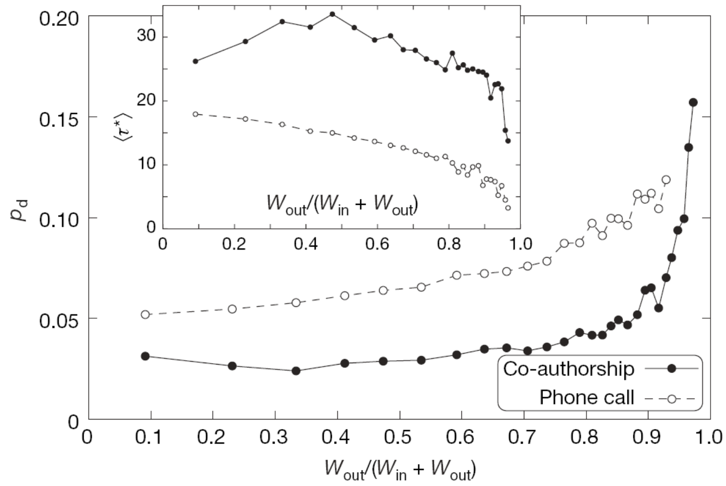


Fig. 11. The probability p_d of a community disintegrating in the next step as a function of the ratio of the aggregated weights of links from the community to other parts of the network (W_{out}) and the aggregated weights of all links starting from the community $(W_{in}+W_{out})$. The inset shows the average life-time of communities as a function of $W_{out}/(W_{in}+W_{out})$.

In summary, our results indicated the significant difference between smaller collaborative or friendship circles and institutions. At the heart of small cliques are a few strong relationships, and as long as these persist, the community around them is stable. It appears to be almost impossible to maintain this strategy for large communities. Our calculations showed that the condition for stability of large communities is continuous change, so that after some time practically all members are exchanged. Such loose, rapidly changing communities are reminiscent of institutions, which can continue to exist even after all members have been replaced by new members. For example, in a few years most members of a school or a company could change, yet the school and the company will be detectable as a distinct community at any time during its existence.

Other related investigations

Thermalling

The subject of thermalling is rather different from our originally planned topics, however, the quantitative study of this phenomenon in the case of birds involved the building of ultra light weight GPS devices which, in turn, can be used for studying the group flights (collective motion) of avians.

Gliding saves much energy and to make large distances using this form of flight only represents a great challenge for both birds and people. The solution is to make use of the so called thermals which are localized, warmer regions in the atmosphere moving upwards with a speed exceeding the descent rate of bird and plane. While birds use this technique mainly for foraging, humans do it as a sporting activity. Thermalling involves efficient optimization including the skilful localization of thermals, trying to guess the most favorable route, estimating the best descending rate, etc. In our related study we addressed the question whether there are any analogies between the solutions birds and humans find to handle the above task [22]. High resolution track logs were taken from thermalling falcons and paraglider pilots to determine the essential parameters of the flight patterns. We find that there are relevant common features in the ways birds and humans utilize thermals. In particular, falcons seem to reproduce the MacCready formula widely used by gliders to calculate the best slope to take before an upcoming thermal.

As for our original question, the result seems to be a draw. All of the parameters we determined were nearly the same for both humans and birds. Thus, as it happens, evolving flight strategies of birds and human calculations lead to virtually the same outcome.

References

1. T. Vicsek: A question of scale, *Nature*, Vol. **411** (2001) 421
2. T. Vicsek, "The bigger picture" (Concepts essay), *Nature,* vol. **418,** 131 (2002)
3. B. Szabó, G. Szőlősi, B. Gönci, Zs. Jurányi, D. Selmeczi and T. Vicsek. Phase transition in the collective migration of tissue cells: experiments and models, *Physical Review* **E 74**, (2006) 061908
4. M. Nagy, Zs Akos, D. Biro and T. Vicsek: Quantitative analysis of the group dynamics in homing pigeon flocks (submitted for publication)
5. T Vicsek, A Czirók, E Ben-Jacob, and I Cohen ``Novel type of phase transition in a system of self-driven particles" *Phys. Rev. Lett.* **75** 1226 (1995)
6. P. Szabó, M. Nagy and T. Vicsek: Transitions in a self-propelled-particles model with coupling of accelerations**.** *Phys. Review*. **E 79** (2009) 021908.
7. M. Nagy, I. Daruka and T. Vicsek, New aspects of the continuous phase transition in the scalar noise model (SNM) of collective motion, *Physica A* **373** (2007) 445-454
8. B. Gönci, M. Nagy and T. Vicsek: Phase transition in the scalar noise model of collective motion in three dimensions, *European Physical Journal,* **157** (2008) 53
9. Initiating a Mexican wave: An instantaneous collective decision with both short- and long-range interactions, *Physica A,* **369** (2006) 830-840
10. T. Vicsek. Phase transitions and overlapping modules in complex networks *Physica A* **378** (2007) 20-32.
11. G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, (2005) 81
12. B. Adamcsek, G Palla, I. J. Farkas, I Derenyi and T Vicsek, CFinder: locating cliques and overlapping modules in biological networks, *BIOINFORMATICS*, **22** (2006) 1021-1023
13. M. C. Gonzalez, H. J. Herrmann, J. Kertesz and T. Vicsek, Community structure and ethnic preferences in school frienship networks, *Physica A*, **379**, (2007) 307-316
14. Farkas, D. Ábel, G. Palla and T. Vicsek, Weighted network modules, *New J. Phys.* **9** (2007) 180
15. G. Palla, I. J Farkas, P. Pollner, I. Derényi and T. Vicsek, Directed network modules, *New J. Phys.* **9** (2007) 186
16. P. Pollner, G. Palla, D. Ábel, A. Vicsek, I. J. Farkas, I. Derényi and T. Vicsek, Centrality properties of directed module members in social networks *Physica* A 3**87** (2008) 4959–4966
17. Derényi, G. Palla and T. Vicsek: Clique Percolation in Random Networks *Phys Rev. Lett* **94**, (2005) 160202
18. G. Palla, I. Derényi, and T. Vicsek: The critical point of k-clique percolation in the Erdős-Rényi graph, *J. Stat. Phys*. **128,** (2007) 219-227_

19. G. Palla, I. Farkas, P. Pollner, I. Derényi,and T. Vicsek , Fundamental statistical features and self-similar properties of tagged networks, *New J. Phys.* **10** (2008) 123026
20. P. Pollner, G. Palla and T. Vicsek: Preferential attachment of communities: the same principle, but a higher level *Europhys. Lett* **73** (3), (2006) 478-484
21. G. Palla, A-L. Barabasi and T. Vicsek: "Quantifying social group evolution", *Nature* , **446** (2007) 664-667,
22. Zs. Ákos, M. Nagy, and T. Vicsek: Comparing bird and human soaring strategies, *Proc. Nat. Acad. Sci. USA.,* **105,** No. 11 (2008) 4139-4143