COSIN

IST–2001-33555

Coevolution and Self-Organization in Dynamical Networks

http://www.cosin.org

STATE OF THE ART

Deliverable Number:	D10
Delivery Date:	March 2004
Classification:	Public
Partner owning:	**CR3** (UB University of Barcelona)
WorkPackage	WP4
WorkPackage Coordinator	Albert Diaz-Guilera
Contact authors:	Andrea Capocci INFM, Albert Diaz Guilera, UB
Project Co-ordinator:	Guido Caldarelli (INFM) *guido.caldarelli@roma1.infn.it*, Istituto Nazionale Fisica per la Materia
Partners:	**CR2** (UDRLS) Università "La Sapienza" *Italy* **CR3** (UB) Universitat de Barcelona, *Spain* **CR4** (UNIL), Université de Lausanne *Switzerland* **CR5** (ENS) Ecole Normale Superieure, *France* **CR7** (UNIKARL), University of Karlsruhe, German*y* **CR8** (UPSUD) Université de Paris-Sud, *France*

Project funded by the European Community under the "Information Society Technologies" Programme (1998-2002)

Abstract

The study of the state of the art for a new field of research such the one of the growing networks is a crucial task in order to keep updated the scientific challenges of a research project. Here we present some analysis of the field as perceived by the project members as well as the opinion of the major experts worldwide as collected during our midterm conference at month 18.

Contents

Introduction	**2**
New Topics	**3**
Round Table of Experts	**7**
Bibliography	**11**

- **INTRODUCTION**

After an initial boost, the scientific production about complex networks has apparently reached a steady-state. The observation of the cond-mat archive (http://xxx.lanl.gov) gives a rough measure of this: the number of submitted preprints whose abstract includes the words "network" or "graph" were 46 in the last 12 months (March 04), whereas 60 have been submitted in 2004, and 65 in 2003. For a comparison, such papers were as many as 214 in 2000.

Network research, in recent times, has been focussing on deeper analysis of networks structure, with new approaches in real data surveys and in mathematical models able to reproduce them. Recent empirical studies concerned mainly social networks, that is, networks built by human interactions. Moreover, new models for complex networks have been introduced, in order to reproduce the features of real networks where growth and preferential attachment (believed to be responsible of the power law degree distribution) are absent. Other features of real networks such as the clustering have also been studied, having relevant practical applications. Finally, the heterogeneity of the interactions has been considered, by the phenomenological study of weighted networks as well as the construction of a new model for complex weighted networks.

In the following we will present the main results obtained in literature in the last two years (2002-2003) during the first period of the life of COSIN. Some of these results have been obtained within the project by the people of the consortium. For the others we tried to import within the consortium the most valuable ideas and approaches by looking for collaboration and support with the rest of the community. After the presentation of these results, we also report the results of a virtual round table held in Rome for our midterm meeting.

- **THE MOST RECENT TOPICS**

  - **Social Networks**

Social networks are built by human interactions. Such networks arise because of many mechanisms. Collaboration produce complex networks, a phenomenon well known since decades[1]. The study of scientific collaboration networks, where scientists are nodes connected by co-authorships of scientific literature, has been one of the most studied instance of social network, since data are well standardized and widely available. Similarly, other examples of interactions have been studied, including sexual contacts, opinion sharing and economic exchanges: all these cases display complex features, that is, a short network diameter (the average distance between two nodes) and a scale-free distribution of the degree $P(k)$, that is, the frequency $P(k)$ of nodes with k links follows a power-law behavior

$$P(k) \sim k^{-\gamma}.$$

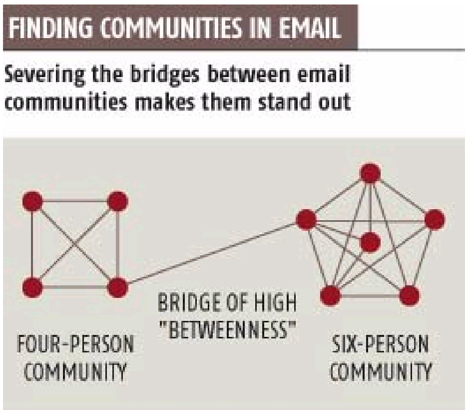

Figure 1: A particular situation where knowledge on social networks and communities detection are needed in order to solve the problem.

Recently, surveys carried on collaboration networks[2,3,4] (authors linked by co-authorships), communication networks (Internet users exchanging e-mail messages)[5,6,7] and markets (economic interactions connecting market agents)[8,9,10], have confirmed that also social networks display complexity in the degree distribution $P(k)$.

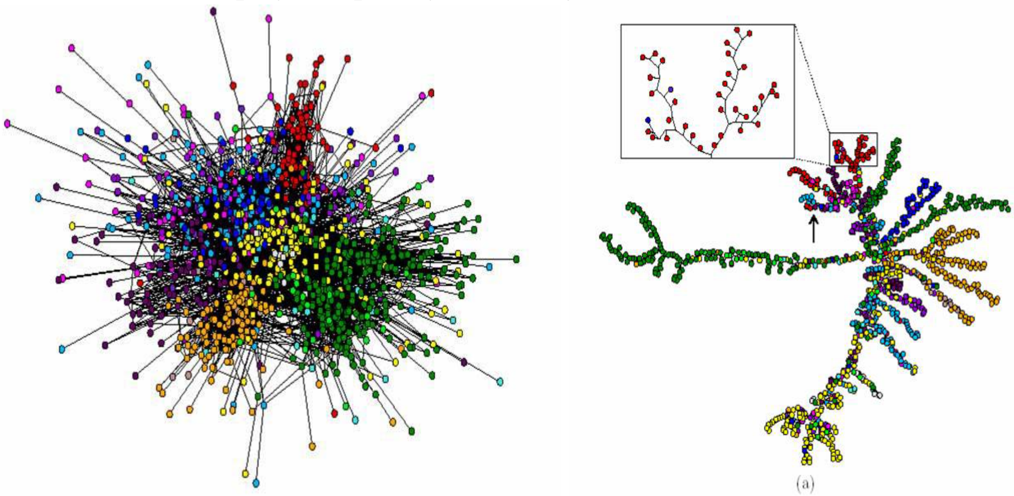

Figure 2: A collection of e-mails exchanged in the University of Tarragona, different colours refer to different Departments. On the right the result of a taxonomical representation of the different communities as obtained with the Newman-Girvan Algorithm (From A. Arenas et al.)

Moreover, as it has been verified by simulation, opinion formation models such as the Sznajd model[11], when set on a scale-free network instead of a regular lattice, produce results compatible with observations of real situations[12].
This suggests that the modelization of human interactions by scale-free graphs[13] is reasonable.

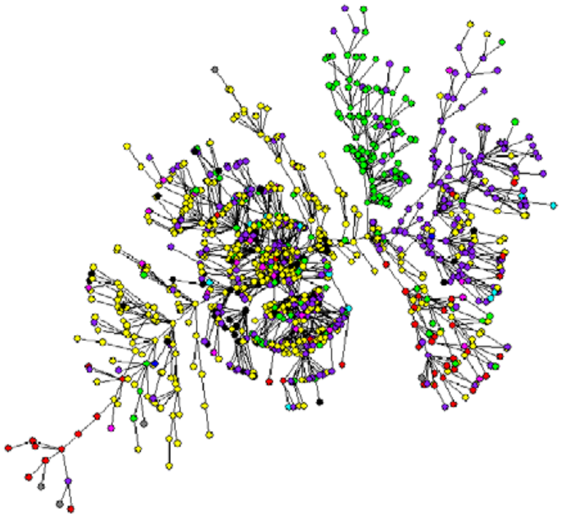

Figure 3 A network of correlation in stock prices at NYSE (From G. Bonanno et al.)

- **Static Models of Networks**

Mechanisms generating complex features in networks (above all, the power-law degree distribution) have been deeply studied in the previous years, focussing mainly on evolving networks, whose size grows with time. This caused a very rapid development of the scientific field at its start. A seminal paper by A.-L. Barabási et al.[14] introduced the idea that growth and preferential attachments were necessary to the occurrence of fat tails in the degree distribution.

Nowadays, such approach is no more a priority: growing networks with various additional mechanisms have been introduced and reproduce almost any the parameters measured in real networks. Nevertheless, the hypothesis of a dynamically evolving network is not always verified in reality.
Therefore, a number of scientists are now looking for network models displaying a scale-free distribution of the degree without network growth, since this could be more appropriate to represent a number of real network instances.
Many papers have shown that this is possible indeed. A class of static models, based on random intrinsic fitness, have been introduced, simulated and analytically studied[15-20]. Not only is the degree distribution correctly reproduced by such model: in addition, other typical relations found in social networks are consistent with such model, such as a positive correlation between connected nodes' degrees[20]. This characteristics is called "assortative mixing" in the sociologists' literature, meaning that individuals with similar properties tend to be connected with high probability.

- **Communities and Clustering**

Measurements and exact results concerning the clustering patterns of networks mainly concern the occurrence of regular motifs[21-24] and their correlations[15,25,26].
However, many social and information networks, such as the World Wide Web, turn out to be approximately partitioned into communities of irregular shape: for example, web pages focusing on similar topics are strongly mutually connected and have a weaker

linkage to the rest of the Web. The design of methods to partition a graph into several meaningful highly inter-connected components have then become a compelling application of graph theory to biological, social and information networks[27-29].

Detecting the community structure in information networks allows one to mine information in a more efficient way, narrowing the exploration of a network as large as the World Wide Web (about 10^9 nodes) to a limited portion of it. When used in the analysis of large collaboration networks, such as company or universities, communities reveal the informal organization and the nature of information flows through the whole system[30,31].

For this reason, algorithms able to find communities in large networks could have a strong scientific and technological impact. Methods have been introduced to solve this problem[32-37], with different approaches.

Two main classes of algorithms exist: communities in networks can be identified by recursively splitting the whole original graph. In this case, links are progressively removed until disconnected components corresponding to different clusters appear. Links to be removed are usually chosen according to their centrality, that is, the fraction of shortest paths passing through each of them[32,36].

Alternatively, other authors study the spectral properties of the adjacency matrix or related ones. These methods are often inspired by the spectral analysis of stochastic processes such as Markov chains. As shown in references[33,34], the structure of the eigenvectors associated to the largest eigenvalues reveal how nodes can be optimally clustered, according to different criteria.

- **Weighted networks**

While complex networks are usually characterized by their topological complexity, they also often display a large heterogeneity in the capacity and intensity of the connections. In the Internet or in the Web, in ecosystems, or in the world-wide airport network, the strength of interactions varies greatly. This diversity in the weights of the interaction adds a complexity which cannot be overlooked in the study and description of these networks. New models of complex networks explaining this heterogeneity are therefore necessary.

In this context, a first phenomenological study[37] has analyzed two typical examples of complex weighted networks. The first example is a prototype of large infrastructure network: the airline connection network. The second one is the scientific collaboration network which can be considered as a good proxy for social networks. In both cases, it is possible to assign a weight to each link and in the airline network example, the weight is the number of available seats (per week) for a given direct connection between two airports. New appropriate metrics have also been defined, combining weighted and topological observables, in order to characterize the complex statistical properties and heterogeneity of the actual strength of edges and vertices.

Moreover, a model for the growth of weighted networks has been proposed[38], that couples the establishment of new edges and vertices and the weights' dynamical evolution. The model is based on a simple weight-driven dynamics and generates networks exhibiting the statistical properties observed in several real-world systems. In particular, the model yields a non-trivial time evolution of vertices' properties and scale-free behavior with exponents varying in the interval [2,3] for the weight, strength and degree distributions.

The simplicity of the model and the diversity of the results obtained suggest that it could be a very important ingredient in the formation of large networks. This model and its possible variations will be a very important tool in studies of complex networks, since it incorporates in a natural way the diversity of interactions, and allows for the study of their effect on diverse phenomena such as virus propagation or congestion failures.

Virtual Round Table on Ten Leading Questions

in Network Research

*The following discussion is an edited summary of the public debate started during the conference "*Growing Networks and Graphs in Statistical Physics, Finance, Biology and Social Systems*" held in Rome in September 2003. Drafts documents were circulated electronically among experts in the field and additions and follow-up to the original discussion have been included. Among the scientists participating to the discussion, L.A.N. Amaral, A. Barrat, A.L. Barabasi, G. Caldarelli, P. De Los Rios, A. Erzan, B. Kahng, R. Mantegna, J.F.F. Mendes, R. Pastor-Satorras, A. Vespignani are acknowledged for their contributions and editing.*

The last few years have witnessed a tremendous activity devoted to the characterization and understanding of networked systems. Indeed large complex networks arise in a vast number of natural and artificial systems. Ecosystems consist of species whose interdependency can be mapped into intricate food webs. Social systems are best represented by graphs describing various interactions among individuals. The Internet and the World-Wide-Web (WWW) are prototypical examples of self-organized networks emerging in the technological world. Large infrastructures such as power grids and the air transportation network are critical networked systems of our modern society. Finally, the living cell is not an exception either, its organization and function being the outcome of a complex web of interactions among genes, proteins and other molecules.

For a long time all these systems have been considered as haphazard set of points and connections, mathematically framed in the random graph paradigm. This situation has radically changed in the last five years, during which the study of complex networks has received a boost from the ever-increasing availability of large data sets and the increasing computer power for storage and manipulation. In particular, mapping projects of the WWW and the physical Internet offered the first chance to study the topology of large complex networks. Gradually, other maps followed describing many networks of practical interest in social science, critical infrastructures and biology. Researchers thus have started to have a systematic look at these large data sets, searching for hidden regularities and patterns that can be considered as manifestations of underlying laws governing the dynamics and evolution of these complex systems. Indeed, when studying the structure of complex networks, one finds out that in spite of the apparent complexity and randomness of the underlying systems, clear patterns and regularities emerge, which can be expressed in mathematical and statistical fashion.

Specifically, many of these systems show the small-world property, which implies that in the network the average topological distance between the various nodes increases very slowly with the number of nodes (logarithmically or even slower). A particularly important finding is the realization that many networks are characterized by the statistical abundance of "hubs"; i.e. nodes with a large number of connections to other elements. This feature has its mathematical roots in the observation that the number of elements with k links follows a power-law distribution, indicating the lack of any characteristic scale. This has allowed the identification of the class of scale-free networks, whose topological features turn out to be extremely relevant in assessing the physical properties of the system as a whole, such as its robustness to damages or vulnerability to malicious attack.

The attempt to model and understand the origin of the observed topological properties of real networks has led to a radical change of perspective, shifting the focus from static graphs, aiming to reproduce the structure of the network in a certain moment, to modeling network evolution. This new approach is the outcome of the realization that most complex networks are the result of a growth process. As a result, we currently view networks as dynamical systems that evolve through the subsequent addition and deletion of vertices and edges. The set of dynamical rules defining these processes thus outlines the dynamical theory required for the description of the macroscopic properties of networks. This methodology that is akin to the statistical physics approach to complex phenomena appears as a revolutionary path in our understanding of networked systems and provides new techniques to approach conceptual and practical problems in this field.

While the advances that we have witnessed in the past few years were truly amazing, both in their impact on basic science and practical implications, they have highlighted the incompleteness of our knowledge as well. We are therefore in a position to ask a series of important questions that require the community's attention. Our goal here is to formulate some of these questions, offering a loose guide to the community and ourselves. We should emphasize that we are aware of the fact that advances in all sciences are often induced by the ability of its practitioners to ask novel questions. Thus these questions should by no means be seen as a way to limit new ideas, or to channel our thinking into narrowly defined directions. We feel, however, that formulating these questions would offer a valuable guide to both practitioners and those interested in network theory, thus it is worthwhile elaborating on them.

- Are there formal ways of classifying the structure of different growing models?
Many networks models, or classes of models, have been recently formulated and empirically studied by numerical simulations or approximate analytical methods. While this corresponds to a great advancement in the modeling and representation of networks, a rigorous understanding of the topology of these models is far less developed. Questions concerning the universality of some topological properties, the correlations introduced by the dynamical process and the interplay between clustering, hierarchies and centrality in networks are still only partially answered. A full and general understanding of this issue amounts to the development of rigorous methods to uncover the mathematical structure of growing networks.

-Are there further statistical distributions that can provide insights on the structure and classification of complex networks?
In the study of complex networks a definite set of statistical distributions and observable quantities are customarily used in the real data analysis and in model characterization and validation. These distributions usually rely on the analysis of the network's degree spectrum. The degree distribution $P(k)$ of vertices, the clustering coefficient and the degree-degree correlations expressed as the joint probability $P(k, k')$ of having an edge between two vertices of degree k and k' provide a general classification of the connectivity pattern properties. The ever-increasing evidence in networks for the presence of communities, motifs and modular ordering, however, calls for the developing of new ways to quantitatively characterize these features in precise mathematical terms.

-Why are most networks modular?
The hierarchical nature of networks goes hand in hand with the existence of a modular architecture. In this perspective, networks can be fragmented in different groups of

interconnected elements, or modules, each one being responsible for different functions, quantitatively identified by highly interlinked communities of elements. Modules can be repeated at different hierarchical levels and interconnected via the “hubs” of the system. How modularity emerges across many different networks and how it can be reconciled with the other properties of networks are basic questions of network theory.

- Are there universal features of network dynamics?

Networks are not only specified by their topology but also by the dynamics of information or traffic flow taking place along the links. For instance, the heterogeneity in the intensity of connections may be very important in understanding social systems. Analogously, the amount of traffic characterizing the connections of communication systems or large transport infrastructure is fundamental for a full description of these networks. The final aim is a mathematical characterization that might uncover very general principles describing the networks’ dynamics .

- How do the dynamical processes taking place on a network shape the network topology?

The network provides the substrate on top of which the dynamical behavior of the system must unfold At the same time, however, the various dynamical processes are expected to affect the network’s evolution. Dynamics, traffic and the underlying topology are therefore mutually correlated and it is very important to define appropriate quantities and measures capable of capturing how all these ingredients participate in the formation of complex networks. To carry out this task we need to develop large empirical datasets that simultaneously capture the topology of the network and the time-resolved dynamics taking place on it.

- What are the evolutionary mechanisms that shape the topology of biological networks?

While in technological and large infrastructure networks it is possible to uncover the fundamental dynamical rules governing the network evolution, in biological systems this task is much more difficult. In particular, the role of evolution and selection in shaping biological networks (with emphasis on cell biology) is still unclear, especially if we want a quantitative dynamical implementation of evolutionary principles in network modeling.

- How to quantify the interaction between networks of different character (networks of networks)?

Most networks are interconnected among them forming networks of networks. This is the case of the Internet, where diverse networks interconnect with each other and communicate through a common protocol. In more complicated situations, however, the interconnected networks are very different in nature and their interactions and reciprocal influence has been completely ignored. For instance, the energy and power distribution networks represent a physical layer over which many critical infrastructures, such as information systems, transportation networks and other public services are lying. In the biological world, the gene network is interconnected with the protein-protein interaction network and the metabolic network. The understanding and characterization of the complicate set of regulatory and feedback mechanisms connecting various networks is probably one of the most ambitious tasks in network research. In this case we are also in need of novel quantities and mathematical tools tailored to describe and model the structure of networks of networks.

- Is it possible to develop tools to address in a systematic fashion the robustness and vulnerability of large technological and infrastructural networks?
Complex networks react in different ways to different perturbations. In general they are robust to random damages but weak to attacks targeting some key elements of the system. Furthermore, networks have a dynamics as well as structure. For instance, in power grids, which transport energy, the failure of one component increases the burden on other elements, potentially overloading them, and disrupting their functions. In this way failures may cascade through the network, causing far more disruption that one would expect from the initial cause. A systematic theory of network resilience and robustness needs to address both local (individual failures) and global vulnerabilities (cascading failures).

- How to characterize small networks?
Many real world networks are far from being large scale objects, well described by statistical measures. In this case we have to stop relying on a statistical characterization and look for a detailed description of the particular structure. Concepts such as scaling behavior or average properties are not well defined and new unifying concepts and mathematical tools need to be devised.

Why are social networks all assortative, while all biological and technological networks disassortative?
A series of recent measurements indicate the social networks are fundamentally different from all other networks: while in social networks there is a tendency for the hubs to be linked together (assortativity), in biological and technological network the hubs show the opposite tendency, being primarily connected to less connected nodes. On one end it is not clear if this is a universal property (the measurements do indicate so, however). On the other end, the origin of this difference is not understood. Is there a generic explanation for the observed patterns, or it represents a feature that needs to be addressed in each network individually?

The above list represents an attempt to limit ourselves to ten questions that we consider as potentially playing an important role in the field at this point. Their order does not reflect their relative importance, but a gradual shift from the more theoretical the predominantly practical and applied problems. The list cannot be exhaustive and naturally it should be seen just as a starting point, highlighting some of the challenges in front of us. They have been collected through discussions and debates with several colleagues, thus they may be considered as expressing a (albeit inherently imperfect) consensus of the network community.

- **BIBLIOGRAPHY**

1. **D. De Solla Price,** *Networks of Scientific Papers*. Science **149** 510-515 (1965)
2. **F. Putsch,** *Analysis and modeling of science collaboration networks,* ArXiv:cond-mat/0305112 (2003)
3. **M. Catanzaro, G. Caldarelli, L. Pietronero**, *Assortative Model for Social Networks,* ArXiv:cond-mat/0308073 submitted to Physical Review E (2003).
4. **Y.Fan, M. Li, J. Chen, L. Gao, Z. Di, J, Wu,** *Network of Econophysicists: a weighted network to investigate the development of econophysics*, ArXiv:cond-mat/0401054
5. **G. Caldarelli F. Coccetti, P. De Los Rios**, *Preferential Exchange: Strengthening Connections in Complex Networks,* ArXiv:cond-mat/0312236 (2003)
6. **Alex Arenas, Leon Danon, Albert Diaz-Guilera, Pablo M. Gleiser, Roger Guimera,** *Community analysis in social networks* ArXiv:cond-mat/0312040 (2003)
7. **H. Ebel, L.-I. Mielsch, S. Bornholdt,** *Scale-free topology of e-mail networks*, Physical Review E **66** 056118 (2002).
8. **D. Garlaschelli, M. I. Loffredo,** *Wealth Dynamics on Complex Networks,* ArXiv:cond-mat/0402466 (2004).
9. **D. Garlaschelli, S. Battiston, M. Castri, V. D. P. Servedio, G. Caldarelli,** T*he scale-free topology of market investments*, Arxiv:cond-mat/0310503 (2003).
10. **G. Bonanno, G. Caldarelli, F. Lillo, S. Micciche`, N. Vandewalle, R. N. Mantegna,** *Networks of equities in financial markets* Arxiv:cond-mat/0401300 (2004).
11. **K. Sznajd-Weron, R. Weron,** *A simple model of price formation* Int. J. Modern Physics C, **13**, 115-123 (2002).
12. **M.C. Gonzalez, A.O. Sousa, H.J. Herrmann**, *Opinion Formation on a Deterministic Pseudo-fractal Network* Int.J.Mod.Phys. C **15**, (2004).
13. **Ph. Blanchard, A. Ruschhaupt, T. Krueger,** *Small World Graphs by the iterated "My Friends are Your Friends" Principle*ArXiv:cond-mat/0304563.
14. **A.-L. Barabási, R. Albert, H. Jeong,** ***Mean-field theory for scale-free random networks*** Physica A **272**, 173-187 (1999).
15. **G. Caldarelli, A. Capocci, P. De Los Rios, M. A Muñoz,** *Scale-free networks from varying vertex intrinsic fitness* Physical Review Letters **89**, 258702 (2002).
16. **M. Baiesi, S. S. Manna,** *Scale free networks from a Hamiltonian dynamics* cond-mat/0305054 (2003).
17. **G. Mukherjee, S.S. Manna,** *Quasistatic Scale-free Networks* Phys. Rev. E **67**, 012101 (2003).
18. **S. S. Manna, Parongama Sen,** *Modulated Scale-free Network in the Euclidean Space* Phys. Rev. E. **66**, 066114 (2002).
19. **D.-S. Lee, K.-I. Goh, B. Kahng, D. Kim,** On the evolution of scale-free graphs ArXiv:cond-mat/0312336 (2003).
20. **R. Albert, A.-L. Barabási,** *Statistical mechanics of complex networks,* Reviews of Modern Physics **74**, 47 (2002).
21. **S.N. Dorogotsev, J.F.F. Mendes,** *Evolution of networks*, Adv. Phys. 51, 1079-1187 (2002)**.**
22. **J.P. Eckmann, E. Moses,** *Curvature of co-links uncovers hidden thematic layers in the World Wide Web PNAS,* **99** (9), 5825 (2002).
23. **G. Bianconi and A. Capocci,** *Number of Loops of Size h in Growing Scale-Free Networks Phys. Rev. Lett.* **90**, 078701 (2003).
24. **A. Capocci, G. Caldarelli, P. De Los Rios,** *Quantitative description and modeling of real networks* Phys. Rev. E **68** 047101 (2003).
25. **G. Caldarelli, R. Pastor-Satorras, A. Vespignani,** *Cycles structure and local ordering in complex networks* cond-mat/0212026(2002).
26. **I. Simonsen, K. A. Eriksen, S. Maslov, K. Sneppen,** *Diffusion on Complex Networks : A way to probe their large scale topological structures* cond-mat/0312476 (2003), to appear in *Physica A*
27. **S.R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins**, *Trawling the web for emerging cyber-communities*. Computer Networks, **31** 1481-93. (1999).

28. **M. E. J. Newman**, *Finding communities in Social Networks* Eur. Phys. J. B, in press.
29. **B. Huberman, J. Tyler, D. Wilkinson**, in *Communities and technologies*, M. Huysman, E. Wegner and V. Wulf, eds. Kluwer Academic (2003).
30. **R. Guimerà, L. Danon, A. Diaz-Guilera, F. Giralt, A. Arenas**, *Self-similar community structure in organizations* Phys. Rev. E **68** 065103 (2003)
31. **M. Girvan, M. E. J. Newman**, *Community structure in social and biological networks* PNAS **99**, 7821-7826 (2002).
32. **A. Capocci, V.D.P. Servedio, G. Caldarelli, F. Colaiori**, *Detecting communities in large networks*, submitted for publication, ArXiv:cond-mat/0402499 (2004)
33. **J. Kleinberg,** *Authoritative Sources in a Hyperlinked Environment* Journal of the ACM **46** (5) 604 (1999).
34. **F. Wu, B. A. Huberman**, *Finding Communities in Linear Times* accepted by Eur. Phys. J. B (2003)
35. **F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi**, *Defining and identifying communities in networks* PNAS, **101** 2658-2663 (2004).
36. **J. Tyler, D. Wilkinson, B. Huberman,** *E-Mail as a spectroscopy automatic discovery of community structures within organization*. in *Communities and technologies*, M. Huysman, E. Wegner and V. Wulf, eds. Kluwer Academic (2003).
37. A**. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani,** *The architecture of complex weighted networks,* PNAS **101** 3747 (2004).
38. A**. Barrat, M. Barthélemy, A. Vespignani**, *Weighted evolving networks: coupling topology and weights dynamics*, ArXiv:cond-mat/0401057, to appear in Phys. Rev. Lett. (2004).