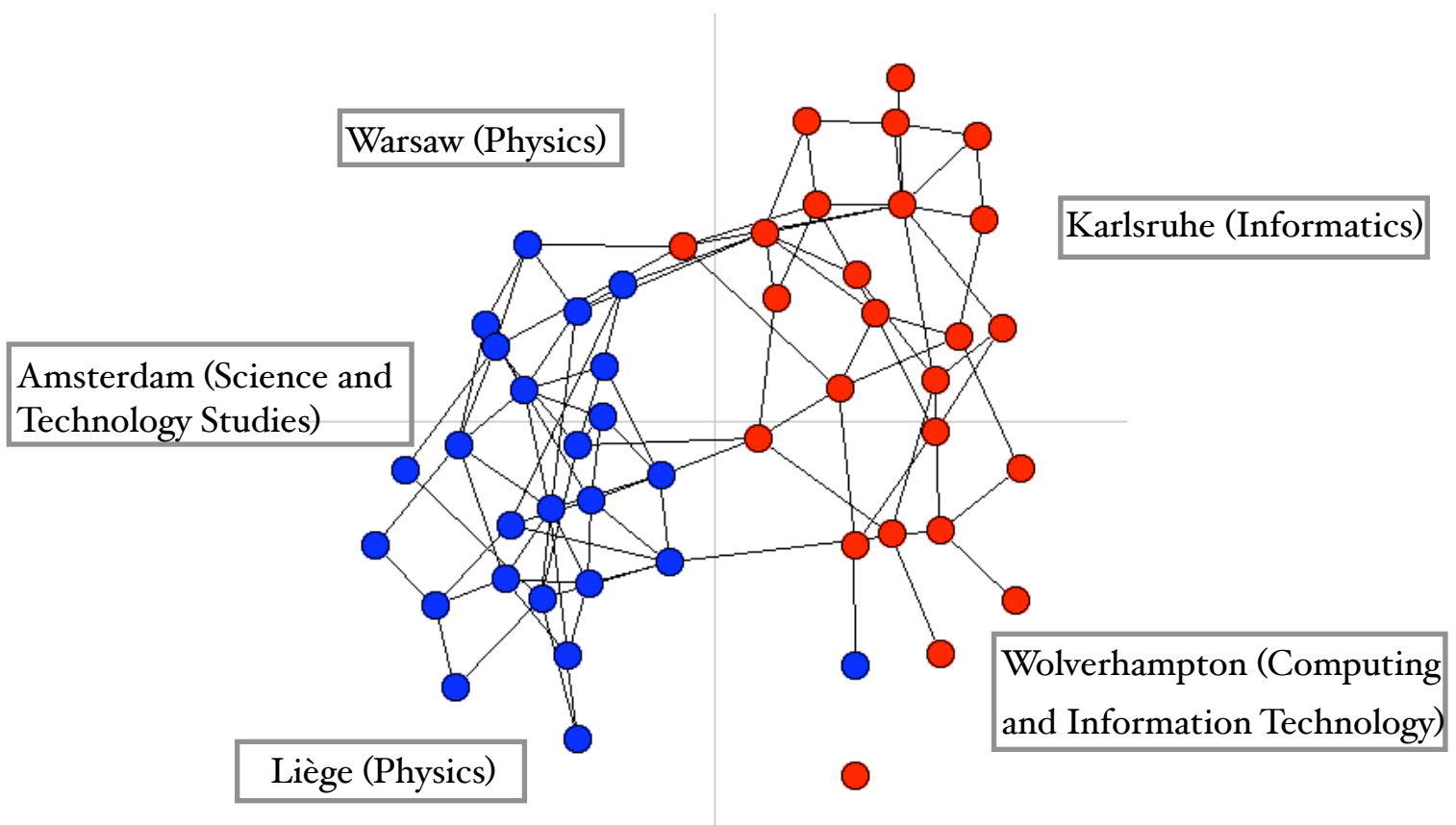# CREEN

Critical Events in Evolving Networks
www.creen.org

Project of NEST Action in 6th EU Framework Programme

SELECTED PROJECT RESULTS 2

1 Jan. 2005 - 31 Dec. 2006

Warsaw (Physics)

Karlsruhe (Informatics)

Amsterdam (Science and Technology Studies)

Wolverhampton (Computing and Information Technology)

Liège (Physics)

NEST = New and Emerging Science and Technology, www.cordis.lu/nest/home.html

CSR Complex Systems Research

WYDZIAŁ FIZYKI POLITECHNIKA WARSZAWSKA

Royal Netherlands Academy of Arts and Science

UNIVERSITY OF WOLVERHAMPTON

SupratecS

University Karlsruhe
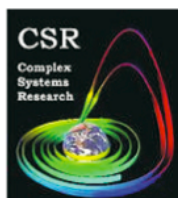
Table of content

I. Project description

What is CREEN?

The project, that is part of the sixth EU Framework Programme, unites research teams from different countries and different scientific backgrounds in order to share their respective areas of expertise and to promote scientific collaborations through Europe.

The participating groups are:
- Faculty of Physics and the Center of Excellence for Complex Systems Research at Warsaw University of Technology (WUT), Poland
- SUPRATECS at University of Liège (ULg), Belgium
- Statistical Cybermetrics Research Group at University of Wolverhampton (UK), England
- Netherlands Institute for Scientific Information Services (NIWI) of the Royal Netherlands Academy of Arts and Sciences (KNAW), Netherlands
- Faculty of Informatics at University of Karlsruhe (SCIT), Germany

The Project is supported by the NEST action of the 6th EU Framework Programme. NEST aims to support unconventional and visionary research with the potential to open new fields for European science and technology, as well as research on potential problems uncovered by science.

The Project CREEN is coordinated by Action GIACS (General Integration of the Applications of Complexity in Science). http://www.giacs.org

Project Abstract

CREEN focuses on social networks and more particularly on the spreading of information in scientific and public communication networks. We see as critical events the emergence of information avalanches linked to the emergence of a collective behaviour in large groups of social actors. The challenge of this interdisciplinary project is to combine models of information avalanches in mediated networks developed in the social sciences – in communication theory, media theory, and science and technology studies - with probabilistic models of data mining in complex networks and mathematical models about the evolution of complex networks developed in physics.

Empirically, we concentrate on the issue of how in science different topics appear, spread out through the scientific community and lead to epidemic-like behaviour (scientific avalanches) and how such scientific avalanches trigger and resonate with avalanches of information about science in the wider public. Data gathering in the project are based on both bibliometric and webometric techniques.

One of the goals of the project is to develop policy recommendations based on a new and innovative understanding of critical events in mediated social networks with regard to scientific avalanches in science and the public understanding of science.

Project objectives

Social networks are characterized by the ambiguity of the actors involved. Human beings are able to participate in many networks, acting in different roles, at the same time. In a real situation isolated networks almost do not exist. One problem in the application of abstract graph theoretical models to social systems is therefore the appropriate definition of the nodes (actors or agents) and of the processes creating links between them.

If one considers information flows in communication networks than it is evident that different networks interrelate with each other. In system theory scientists communicate in certain circles (journals or invisible colleges) using a certain code (scientific articles), whilst in the public media journalists follow other rules in their communication behavior. However, scientific innovation often appears at the boundaries of certain specialties in science, combining different scientific discourses. The public perception of science emerges at the boundary between science communication and public communication, both often interwoven.

Therefore, the issue of systems consisting of many coupled networks is very important for the analysis of social phenomena. Theoretically, on the level of abstract complex networks the question is not yet solved. This project aims to overcome this gap.

We create a theoretical and numerical description of multi-network systems. We localize internal variables at networks nodes to describe dynamics of corresponding agents. We consider cases of a symbiosis and of a competition between different networks. We analyze static and dynamical properties of such systems with the aid of analytic methods as the mean field approach and compare to results of numerical simulations.

As a special case one can consider the influence of external forces on a network. With this, the influence of the second network on the first one is expressed in terms of an overall acting driving force.

Empirically, we identify different actors and the networks they are involved in, in a given scientific debate and its public perception. One can find very different representations of communication processes and accordingly social scientists have produced many complementary descriptions of communication processes. Examples are media theory, communication and information theories, and sociology. In the project our empirical description is in social science terms but guided by concepts of the physics of complex networks. This way we can create a projection of social structures which are complex but not complicated.

Beside network structures another approach in complexity science seems to be very relevant in modeling social phenomena. Recently, the notion of active agents has been introduced to describe a special kind of collective phenomena. In this case, the structures occur because of non-linear interactions between the agents which are able to additionally influence the interactions by themselves. This possibility of self-determination is in particular suitable for social processes where the individuals have to balance between their own motivations driving their actions and the social constraints creating the boundary conditions for their actions.

Project members

Janusz Holyst (*1955) is CREEN Coordinator. He is Professor at the Physics Faculty of Warsaw University of Technology (since 1998) where he leads the Center of Excellence for Complex Systems Research. His current research field includes the simulation of evolving networks, collective opinion formation, econophysics, self-organized criticality, analysis of chaotic economical data, and stochastic resonance.
jholyst@if.pw.edu.pl, http://www.if.pw.edu.pl/~jholyst

Marcel Ausloos (*1943) is Scientific Director of the SUPRATECS group. His current research includes theoretical and applied statistical physics (transport properties, phase transitions, fractals, evolution, growth, econophysics), theoretical condensed matter physics (magnetism, superconductivity, optics) and nonlinear dynamical phenomena (self-organized criticality, extinctions and mutations in models of evolution).
marcel.ausloos@ulg.ac.be, http://www.ulg.ac.be/supras/groupe/Staff/ausloos.html

Andrea Scharnhorst (*1961) is Senior Researcher Fellow in the group „Networked Research and Digital Information" at the Netherlands Institute for Scientific Information Services, at the Royal Netherlands Academy of Arts and Sciences and member of ASCoR, UvA. Her areas of expertise entail scientometrics, evolutionary economics and science and technology studies.
andrea.scharnhorst@vks.knaw.nl, www.virtualknowledgestudio.nl

Loet Leydesdorff (*1948) reads science & technology dynamics at the Amsterdam School of Communications Research (ASCoR) of the University of Amsterdam and associated to the research group at NIWI-KNAW. He has published extensively in scientometrics, systems dynamics, and the sociology of innovation.
loet@leydesdorff.net, http://www.leydesdorff.net

Mike Thelwall (*1965) is Professor in the School of Computing and InformationTechnology at the University of Wolverhampton, UK, where he leads the Statistical Cybermetrics Research Group. His current research field includes identifying and analysing web phenomena using quantitative-lead research methods.
m.thelwall@wlv.ac.uk, http://www.scit.wlv.ac.uk/~cm1993/

Dorothea Wagner is full Professor of computer science at the University of Karlsruhe, Faculty of Informatics; her research areas are network algorithms, combinatorial optimization, visualization, network analysis and transport optimization.
dwagner@ira.uka.de, http://i11www.informatik.uni-karlsruhe.de/algo/people/dwagner/

Piotr Fronczak (*1974) holds a PhD in physics at Faculty of Physics at Warsaw University of Technology. He is a member of the Research Group Nonlinear Dynamics of Complex Systems. His current research field includes the theoretical and applied physics of evolving networks, chaos control and synchronization in dynamical systems as well as the nonlinear dynamics of nonstationary and spatially extended systems.
fronczak@if.pw.edu.pl, http://www.if.pw.edu.pl/~agatka

Agata Fronczak (*1975) holds her PhD in physics at Faculty of Physics at Warsaw University of Technology. She is a member of the Research Group Nonlinear Dynamics of Complex Systems. Her current research field concentrates on nonequilibrium statistical physics including the domain of complex networks, socio- and econophysics.
agatka@if.pw.edu.pl, http://www.if.pw.edu.pl/~agatka

Renaud Lambiotte (*1977) holds a PhD in physics at Université Libre de Bruxelles. His background in non-equilibrium statistical mechanics and molecular dynamics has led him to study granular fluids, anomalous velocity distributions and complex networks modeling (dynamical and static features). He is also interested in data mining and the detection of trends in online communities.
renaud.lambiotte@ulg.ac.be, http://www.lambiotte.be

Iina Hellsten (*1969) is a researcher at the Royal Netherlands Academy of Arts and Sciences, in the Virtual Knowledge Studio research group. She holds a PhD in social sciences. Her background is in communication sciences and the social studies of science and technology. Her areas of expertise include science communication, public understanding of science, media theory and communication sciences.
iina.hellsten@vks.knaw.nl, http://www.virtualknowledgestudio.nl

Rudy Prabowo (*1972) holds a PhD in Computer Science at the University of Wolverhampton. He is a member of the Statistical Cybermetrics Research Group.
His research interest involves automatic text classification, ontology-based reasoning and RSS data analysis.
Rudy.Prabowo@wlv.ac.uk, http://cybermetrics.wlv.ac.uk/people.htm

Thomas Schank (*1970) is a research assistant in the algorithmics group of Dorothea Wagner at the University of Karlsruhe. He is interested in algorithmic aspects of network analysis and in graph visualisation.
Thomas Schank,
http://i11www.informatik.uni-karlsruhe.de/members/index.php?algouser=schank

Malgorzata Zajac (*1977) holds a position of CREEN secretary. Her education background is Management in European Union.
peggy@mars.if.pw.edu.pl

CREEN meetings

Three CREEN internal workshops have taken place in order to strenghen the links and collaborations between the national teams, as well as to discuss details of project research tasks. They have occurred at:

Warsaw, January 13-14, 2005

Karlsruhe, June 2-4, 2005

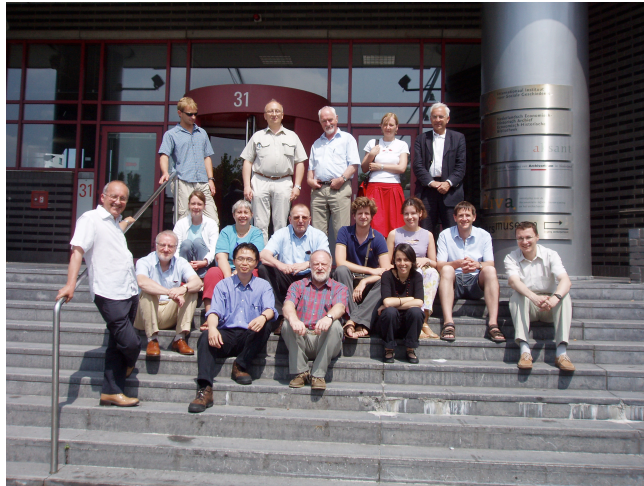**Wolverhampton, January 18-21, 2006**

Amsterdam, July 04-07, 2006

Andrea Scharnhorst co-organized a workshop "Innovation Networks - New Approaches in Modelling and Analyzing" Augsburg, Germany, October 10-14 2005 that was related to the CREEN project.
(http://www.wiwi.uni-augsburg.de/vwl/hanusch/forschung/Exystence_Home.htm)

COMPLEX SYSTEMS NETWORK OF EXCELLENCE

II. Selection of papers[1]

Fluctuation-dissipation relations for complex networks

In the paper we study fluctuations over several ensembles of maximum-entropy random networks. We derive several fluctuation-dissipation relations characterizing susceptibilities of different networks to changes in external fields. In the case of networks with a given degree sequence, we argue that the scale-free topologies of real-world networks may arise as a result of self-organization of real systems into sparse structures with low susceptibility to random external perturbations. We also show that the ensembles of networks with noninteracting links (both uncorrelated and with two-point correlations) are equivalent to random networks with hidden variables.
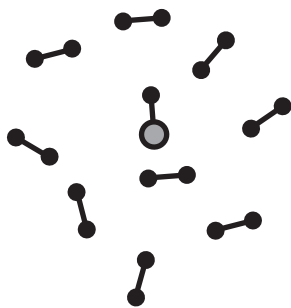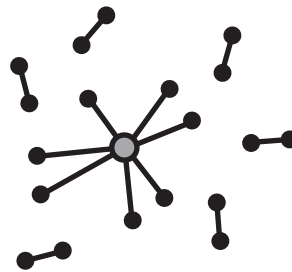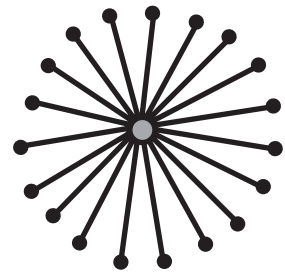
$<k^*>=1$ $<k^*>=7$ $<k^*>=N$

Schematic representation of networks possessing N=19 nodes with one nearest neighbor <k>=1 and the supernode (the gray one) for various connectivities <k*>.

Self-organized criticality and co-evolution of network structure and dynamics

P. Fronczak, A. Fronczak and J. Holyst,

http://www.creen.org/articles/CREEN/warsawSOC.pdf

We investigate, by numerical simulations, how the avalanche dynamics of the Bak-Tang Wiesenfeld (BTW) sandpile model can induce emergence of scale-free (SF) networks and how this emerging structure affects dynamics of the system. We also discuss how the observed phenomenon can be used to explain evolution of scientific collaboration.
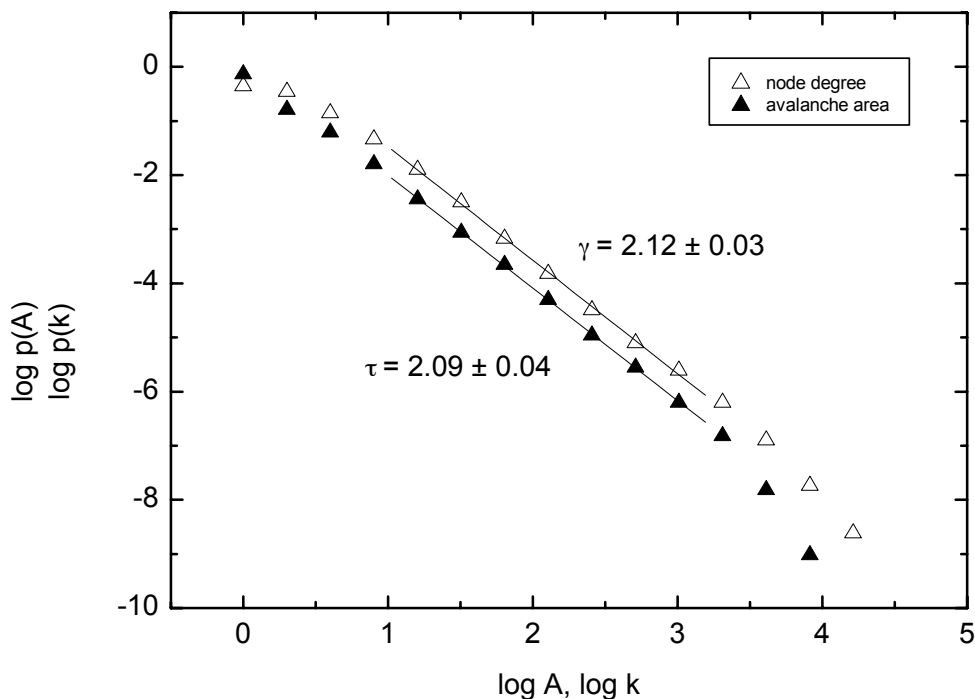
Distributions of avalanche area and node degree in time. Data are logarithmically binned. Lines are linearly fitted with the values indicated at the figure.

Structural phase transitions in a model of scientific projects

P. Fronczak, A. Fronczak and J. Holyst,
submitted (2006)

In this paper we present a simple model of social collaboration networks with clustering and preferential node degree which is a member of family of exponential random graphs. We show by numerical and analytical calculations how a very simple Hamiltonian can lead to surprisingly complicated and eventful phase diagram. We have shown that some configurations formed in the system remind existing structures of European projects, including Coordination Actions.

Diagram representing possible transitions between network configurations. Similarity of configuration *I* to the network of EU projects is shown. Two highly connected Coordination Actions (GIACS and ONCE-CS) are surrounded by plenty of small research projects.
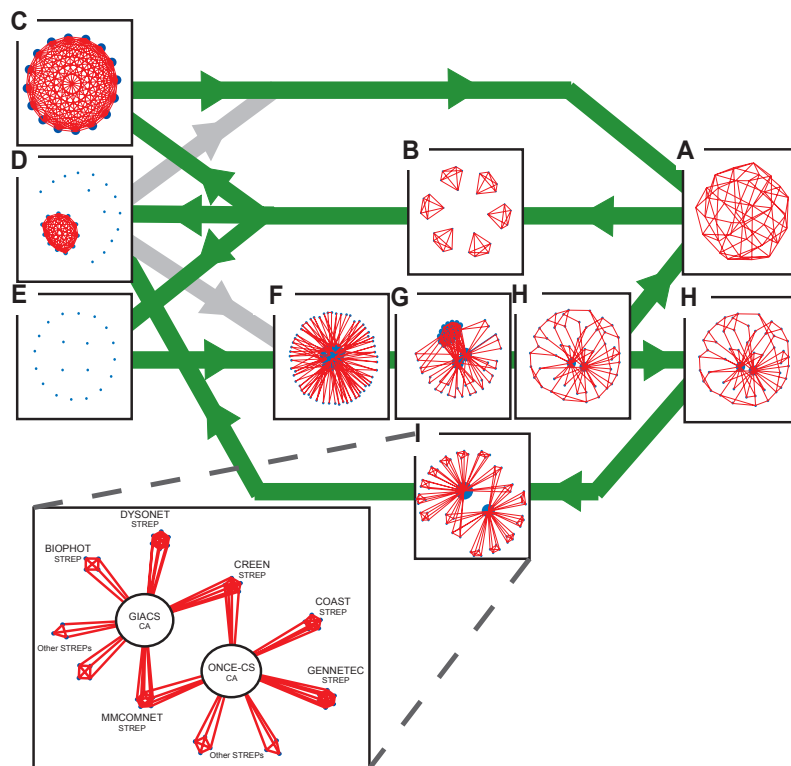
Publish or perish: analysis of scientific productivity using maximum entropy principle and fluctuation-dissipation theorem

P. Fronczak, A. Fronczak and J.A. Holyst,
Phys. Rev. E, in press

Using data retrieved from the INSPEC database we have quantitatively discussed a few syndromes of the publish-or-perish phenomenon, including continuous growth of rate of scientific productivity, and continuously decreasing percentage of those scientists who stay in science for a long time. Making use of the maximum entropy principle and fluctuation-dissipation theorem, we have shown that the observed fat-tailed distributions of the total number of papers x authored by scientists may result from the density of states function $g(x;\tau)$ underlying scientific community. Although different generations of scientists are characterized by different productivity patterns, the function $g(x;\tau)$ is inherent to researchers of a given seniority τ, whereas the publish-or-perish phenomenon is caused only by an external field θ influencing researchers.
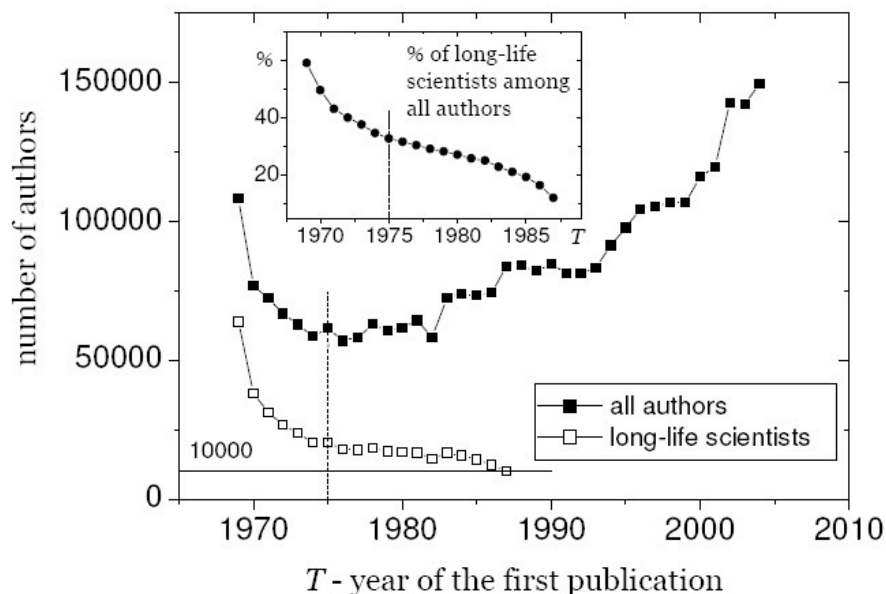
Number of all authors listed in the INSPEC database and the number of long-life scientists versus the year of the first publication T.

Majority Model on a network with communities

R. Lambiotte, M. Ausloos and J. Holyst,
submitted (2006)

We focus on the majority model in a topology consisting of two coupled fully-connected networks, thereby mimicking the existence of communities in social networks. We show that a transition takes place at a value of the inter-connectivity parameter. Above this value, only symmetric solutions prevail, where both communities agree with each other and reach consensus. Below this value, in contrast, the communities can reach opposite opinions and an asymmetric state is attained. The importance of the interface between the sub-networks is shown.
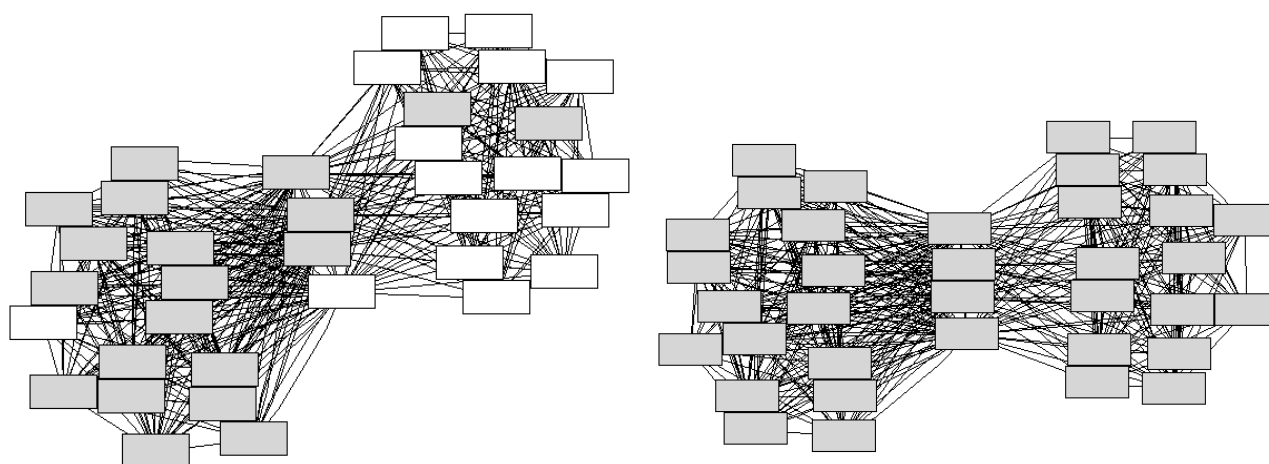
Depending on the initial conditions and on the model parameters, the system can reach either a mixed state, composed of two populations with different opinions or global consensus.

N-body decomposition of bipartite co-authorship networks

R. Lambiotte and M. Ausloos,

In this paper we present a method to project co-authorship networks, that accounts in detail for the geometrical structure of scientists collaborations. By restricting the scope to 3-body interactions, we focus on the number of triangles in the system, and show the importance of multi-scientists (more than 2) collaborations in the social network. This motivates the introduction of generalised networks, where basic connections are not binary, but involve arbitrary number of components. We focus on the 3-body case, and study numerically the percolation transition.
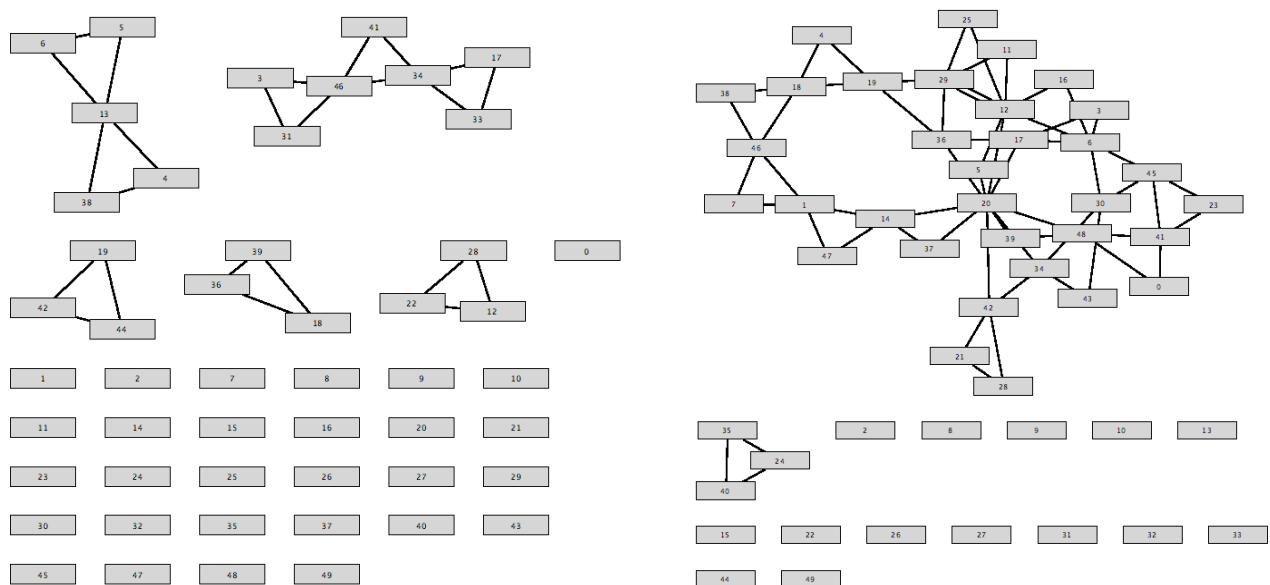
Percolation transition for a triangular network.

Uncovering collective listening habits and music genres in bipartite networks

We analyze web-downloaded data on people sharing their music library, that we use as their individual musical signatures (IMS). The system is represented by a bipartite network, nodes being the music groups and the listeners. Music groups audience size behaves like a power law, but the individual music library size is an exponential with deviations at small values. In order to extract structures from the network, we focus on correlation matrices, that we filter by removing the least correlated links. This percolation idea-based method reveals the emergence of social communities and music genres, that are visualised by a branching representation. Evidence of collective listening habits that do not fit the neat usual genres defined by the music industry indicates an alternative way of classifying listeners/music groups. Finally, a personal identification - community imitation model (PICI) for growing bipartite networks is outlined, following Potts ingredients.
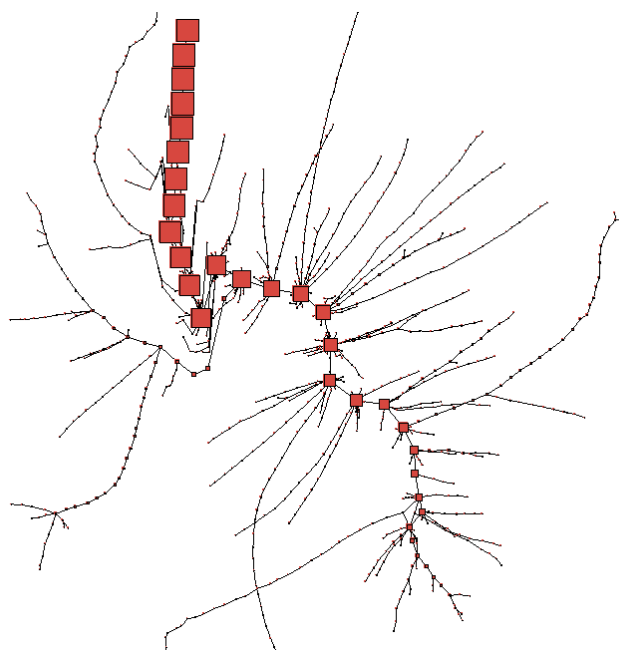
Branching representation of the music groups correlation matrix.

From sand to networks: a study of multi-disciplinarity

R. Lambiotte and M. Ausloos,
submitted (2006)

In this paper we study empirically co-authorship networks of neighbouring scientific disciplines, i.e. granular media and networks. We describe the system, data taken from arXives, by two coupled networks. By considering the 1998-2004 time interval and scanning the network with a mobile time window, we focus on the properties of the interface between both disciplines, as well as on the time evolution of the co-authorship network. We observe drastic jumps in the order parameter, i.e. the link proportion of a given type, that remind of features appearing during nucleation-growth processes and percolation. Therefore, we present an agent based ferro-electric-like model (CDFM), involving bond redistribution between nodes, that reproduces qualitatively the structuring of the system in homogeneous phases. Finally, we define a local parameter accounting for the specialisation of the scientists

Collision and merging of two collaboration networks.

Endo- vs. Exo-genous shocks and relaxation rates in book and music "sales"

R. Lambiotte and M. Ausloos,

In this paper we analyze the response of music and book sales to an external field and to buyer herding. We distinguish endogenous and exogenous shocks. We focus on some case studies, whose data have been collected from ranking on amazon.com. We show that an ensemble of equivalent systems quantitatively respond in a similar way to a similar "external shock", indicating roads to universality features. In contrast to Sornette et al. [Phys. Rev. Lett. 93, 228701 (2004)] who seemed to find power law behaviors, in particular at long times, - a law interpreted in terms of an epidemic activity, we observe that the relaxation process can be as well seen as an exponential one that saturates toward an asymptotic state, itself different from the pre-shock state. By studying an ensemble of 111 shocks, on books or records, we show that exogenous and endogenous shocks are discriminated by their short - time behaviour: the relaxation time seems to be twice shorter in endogenous shocks than in exogenous ones. We interpret the finding through a simple thermodynamic model with a dissipative force.
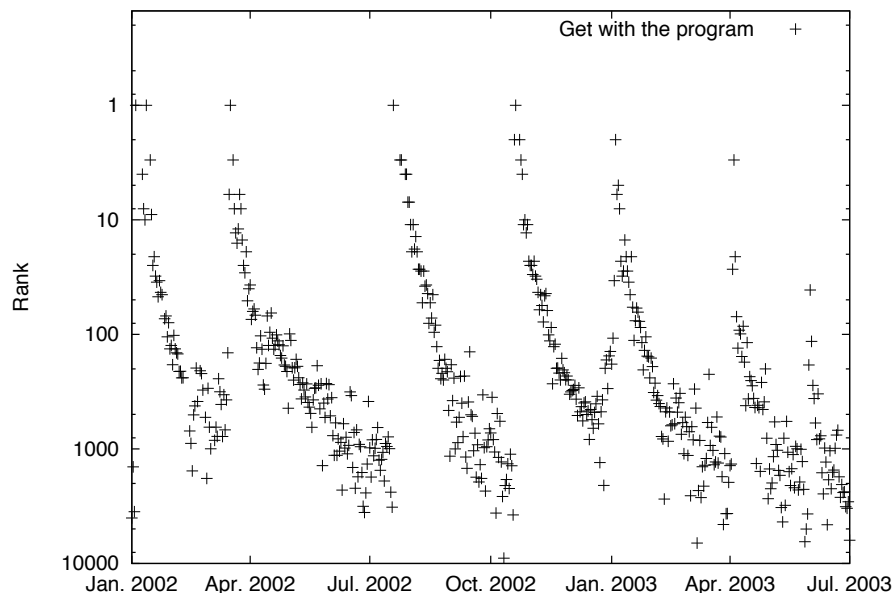
Time evolution of the rank of "Get with the program" in Amazon book sales.

On the genre-fication of Music: a percolation approach

R. Lambiotte and M. Ausloos,
E.P.J.B. **50**, 183 (2006)

We analyze web-downloaded data on people sharing their music library. By attributing to each music group usual music genres (Rock, Pop...), and analysing correlations between music groups of different genres with percolation-idea based methods, we probe the reality of these subdivisions and construct a music genre cartography, with a tree representation. We also discuss an alternative objective way to classify music, that is based on the complex structure of the groups audience. Finally, a link is drawn with the theory of hidden variables in complex networks.
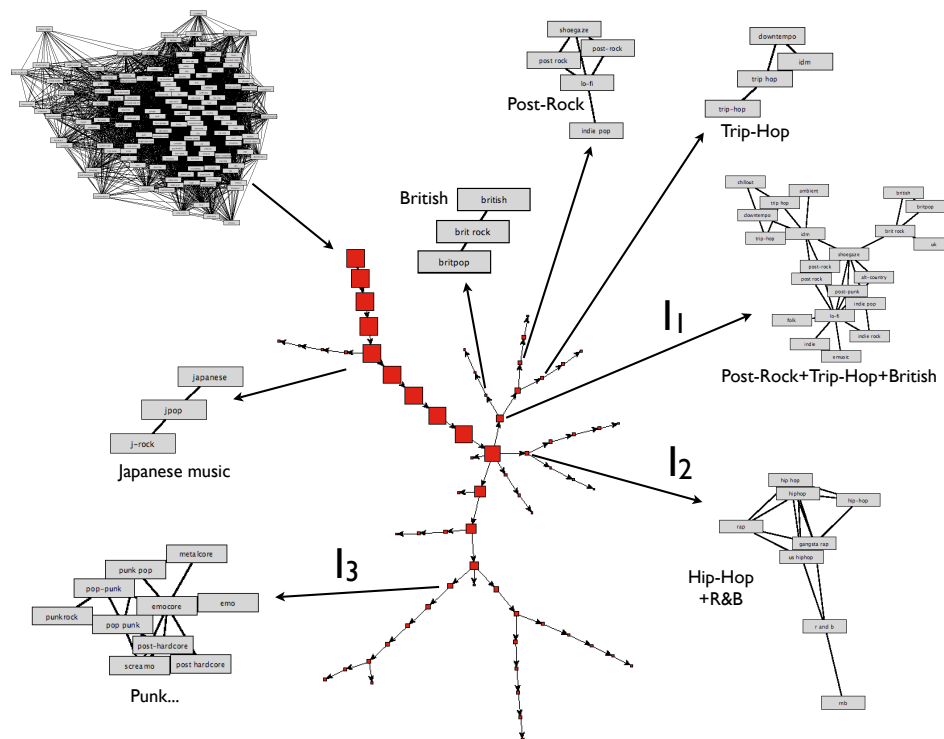
Branching representation of the genres correlation matrix. The tree obviously shows the emergence of homogeneous branches, that are composed of alike music-subdivisions, thereby showing evidence of genre families.

Unanimity Rule on networks

R. Lambiotte, S. Thurner and R. Hanel,
submitted (2006)

We present a model for innovation-, evolution- and opinion dynamics whose spreading is dictated by unanimity rules. The underlying structure is a directed network and the (binary) state of a node will change only if all of its incoming links come from nodes having the same corresponding state. It is shown that a transition takes place depending on the initial condition of the problem. In particular, a critical number of initially activated nodes is necessary for the whole system to get activated in the long-time limit. The influence of the degree distribution of the nodes is naturally taken into account. For simple network topologies we solve the model analytically; the cases of random, small-world and scale-free are studied in detail.
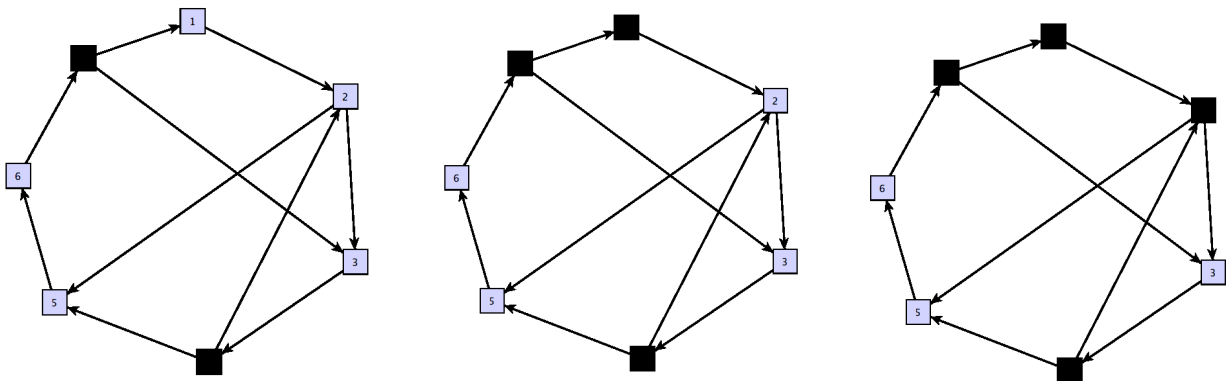
First time steps of the unanimity rule

Growing network with *j*-redirection

R. Lambiotte and M. Ausloos,

http://xxx.lanl.gov/abs/physics/0612148

A model for growing information networks is introduced where nodes receive new links through *j*-redirection, i.e. the probability for a node to receive a link depends on the number of paths of length *j* arriving at this node. In detail, when a new node enters the network, it either connects to a randomly selected node, or to the *j*-ancestor of this selected node. The *j*-ancestor is found by following *j* links from the randomly selected node. The system is shown to undergo a transition to a phase where condensates develop. We also find analytical predictions for the height statistics and show numerically the non-trivial behaviour of the degree distribution.
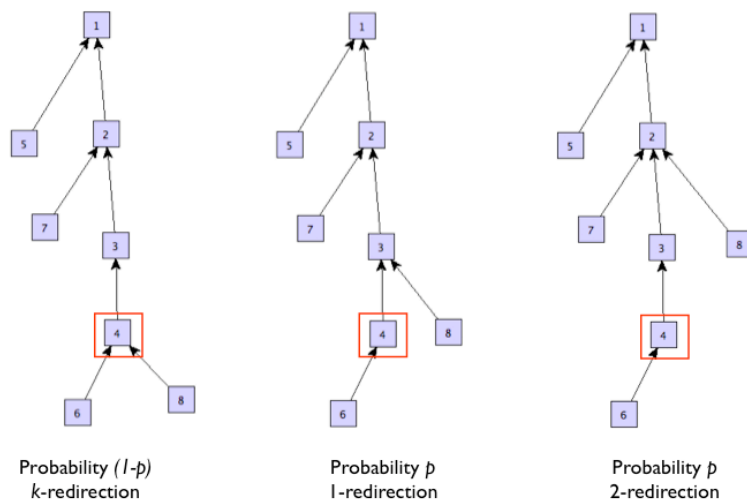
Sketch of a time step of the model with 1-redirection or 2-redirection.

Activity ageing in growing networks

R. Lambiotte,
submitted (2007)

We present a model for growing information networks where the ageing of a node depends on the time at which it entered the network and on the last time it was cited. The model is shown to undergo a transition from a small-world to large-world network. The degree distribution may exhibit very different shapes depending on the model parameters, e.g. delta-peaked, exponential or power-law tailed distributions.
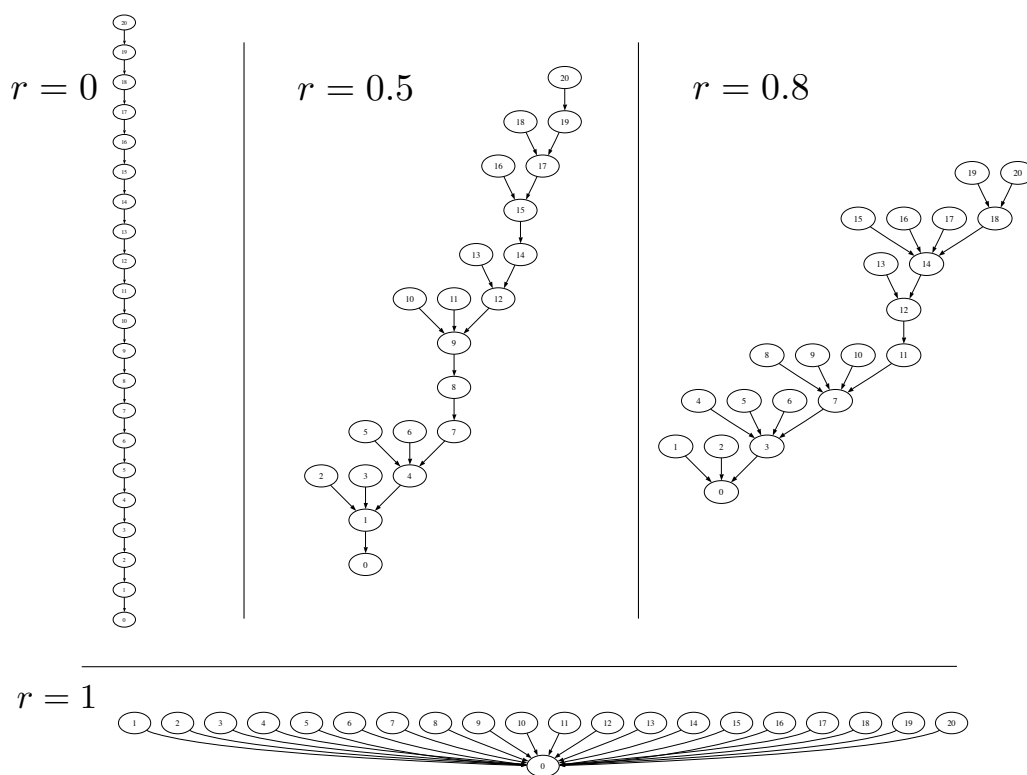
Typical realisations of the model for which the average height evolves linearly with time. One observes a large range of behaviours, from a aligned network (r=0) to a star network (r=1).

Self-citations, co-authorships and keywords: A new approach to scientists' field mobility?

I. Hellsten, R. Lambiotte, A. Scharnhorst and M. Ausloos,
Scientometrics (2007, in press)

This paper introduces a new approach to detecting scientists' field mobility by focusing on an author's self-citation network, and the co-authorships and keywords in self-citing articles. Contrary to much previous literature on self-citations, we will show that author's self-citation patterns reveal important information on the development and emergence of new research topics over time. More specifically, we will discuss self-citations as a means to detect scientists' field mobility. We introduce a network based definition of field mobility, using the Optimal Percolation Method (Lambiotte & Ausloos, 2005). The results of the study can be extended to self-citation networks of groups of authors and, generally also for other types of networks.
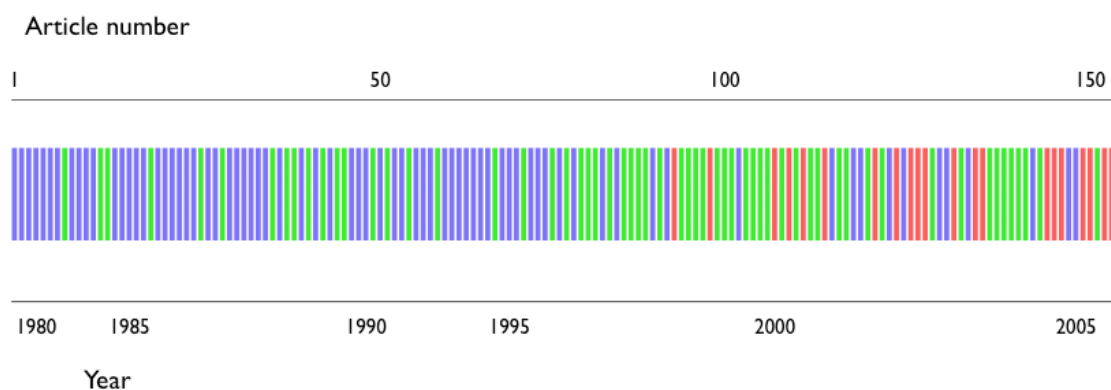
Colour-based graph that represents the evolution of W. Ebeling's centers of interest.

Metaphors and Diaphors in Science Communication: Mapping the Case of 'Stem-Cell Research'

"Stem-cell research" has become a subject of political discussion in recent years because of its social and ethical implications. The intellectual research program, however, has a history of several decades. Therapeutic applications and patents on the basis of stem-cell research became available during the 1990s. Currently, the main applications of stem-cell research are found in marrow transplantation (e.g., for the treatment of leukemia). In this study, the various meanings of the words "stem cell" are examined in these different contexts of research, applications, and policy debates. Translation mechanisms between contexts are specified and a quantitative indicator for the degree of codification is proposed.
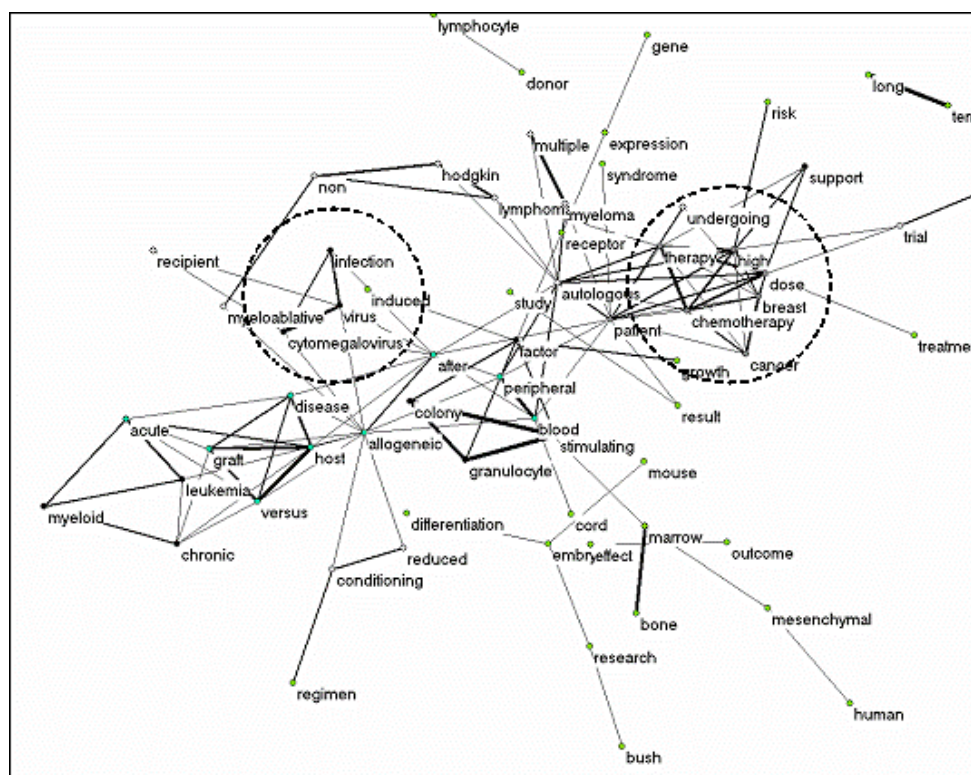
Representations using 67 title word distributions from MED-LINE 2001 with cosine≥0.2.

Measuring the Meaning of Words in Contexts: An automated analysis of controversies

L. Leydesdorff and I. Hellsten,
Scientometrics **67**, 231 (2006)

Co-words have been considered as carriers of meaning across different domains in studies of science, technology, and society. Words and co-words, however, obtain meaning in sentences, and sentences obtain meaning in their contexts of use. At the science/society interface, words can be expected to have different meanings: the codes of communication that provide meaning to words differ on the varying sides of the interface. Furthermore, meanings and interfaces may change over time. Given this structuring of meaning across interfaces and over time, we distinguish between metaphors and diaphors as reflexive mechanisms that facilitate the translation between contexts. Our empirical focus is on three recent scientific controversies: Monarch butterflies, Frankenfoods, and stem-cell therapies. This study explores new avenues that relate the study of co-word analysis in context with the sociological quest for the analysis and processing of meaning.
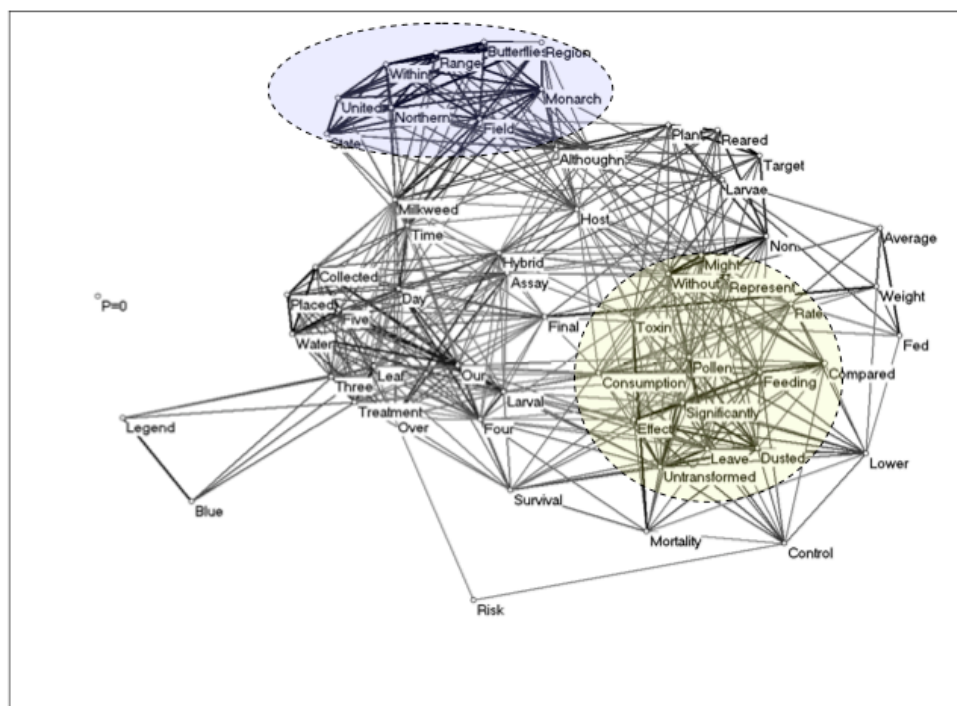
The cosine map of 59 words used more than once in the Scientific Correspondence published in Nature, 399: 214 on May 20, 1999.

Multiple Presents: How Search Engines Re-write the Past

I. Hellsten, L. Leydesdorff and P. Wouters,

Internet search engines function in a present which changes continuously. The search engines update their indices regularly, overwriting Web pages with newer ones, adding new pages to the index, and losing older ones. Some search engines can be used to search for information at the internet for specific periods of time. However, these 'date stamps' are not determined by the first occurrence of the pages in the Web, but by the last date at which a page was updated or a new page was added, and the search engine's crawler updated this change in the database. This has major implications for the use of search engines in scholarly research as well as theoretical implications for the conceptions of time and temporality. We examine the interplay between the different updating frequencies by using AltaVista and Google for searches at different moments of time. Both the retrieval of the results and the structure of the retrieved information erodes over time.

46 words related at the level of cosine≥0.2 and occurring more than 12 times in the 2106 records collected with the AltaVista Advance Search Engine in January 2004.

"Betweenness Centrality" as an Indicator of the "Interdisciplinarity" of Scientific Journals

L. Leydesdorff,
Journal of the American Society for Information Science and Technology
http://www.leydesdorff.net/list.htm

In addition to science citation indicators of journals like impact and immediacy, social network analysis provides a set of centrality measures like degree, betweenness, and closeness centrality. These measures are first analyzed for the entire set of 7,379 journals included in the Journal Citation Reports of the Science Citation Index and the Social Sciences Citation Index 2004, and then also in relation to local citation environments which can be considered as proxies of specialties and disciplines. Betweenness centrality is shown to be an indicator of the interdisciplinarity of journals, but only in local citation environments and after normalization because otherwise the influence of degree centrality (size) overshadows the betweenness-centrality measure. The indicator is applied to a variety of citation environments, including policy-relevant ones like biotechnology and nanotechnology. The values of the indicator remain sensitive to the delineations of the set because of the indicator's local character. Maps showing interdisciplinarity of journals in terms of betweenness centrality can be drawn using information about journal citation environments which is available online.
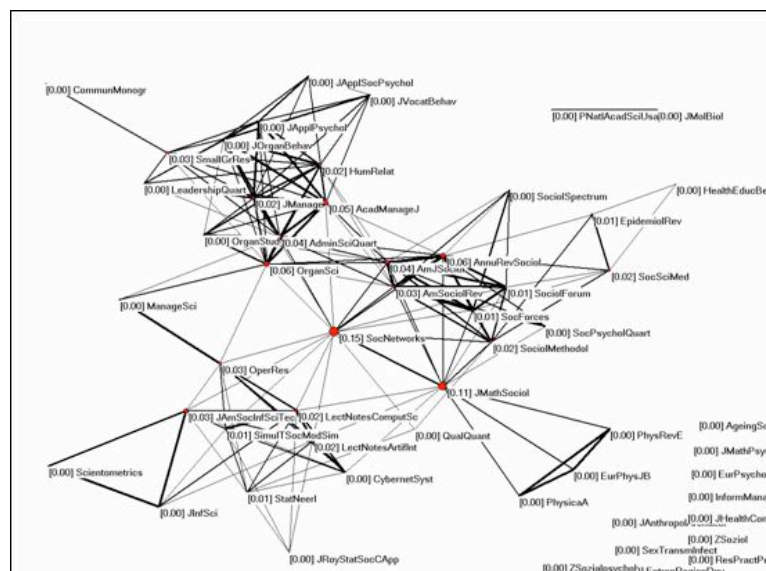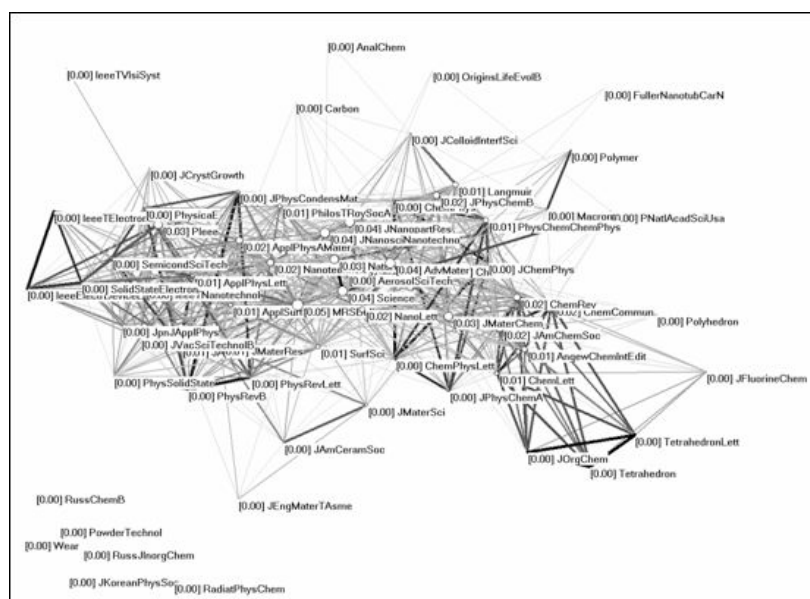
Betweenness centrality of 54 journals in the vector space of the citation impact environment of Social Networks.

Nanotechnology as a Field of Science: Its Delineation in terms of Journals and Patents

L. Leydesdorff and P. Zhou,

The Journal Citation Reports of the Science Citation Index 2004 were used to delineate a core set of nanotechnology journals and a nanotechnology-relevant set. In comparison with 2003, the core set has grown and the relevant set has decreased. This suggests a higher degree of codification in the field of nanotechnology: the field has become more focused in terms of citation practices. Using the citing patterns among journals at the aggregate level, a core group of ten nanotechnology journals in the vector space can be delineated on the criterion of betweenness centrality. National contributions to this core group of journals are evaluated for the years 2003, 2004, and 2005. Additionally, the specific class of nanotechnology patents in the database of the U.S. Patent and Trade Office (USPTO) is analyzed to determine if non-patent literature references can be used as a source for the delineation of the knowledge base in terms of scientific journals. The references are primarily to general science journals and letters, and therefore not specific enough for the purpose of delineating a journal set.

Betweenness centrality among the 67 nano-relevant journals (cosine ≥ 0.2).

The BBC, Daily Telegraph and Wikinews Timelines of the Terrorist Attacks of 7th July 2006 in London: A Comparison with Contemporary Discussions

M. Thelwall and I. Hellsten,

Introduction. We assess the extent to which published media timelines reflect contemporary electronic discussions of major media events, using the London Attacks of July 2005 as a case study. The main objective is to judge whether timelines could, in principle, be automatically generated from contemporary debates. A secondary objective is to analyse the reasons for differences between contemporary debates and retrospective media timelines.

Results. We found that the timelines tended to ignore the role of communication during the event itself, to use less emotionally charged language, and to gloss over to some extent the wider context of the event. Hence some aspects (frames) of the unfolding discussion were ignored by the timelines. Moreover, all sources included a range of relatively trivial details, the timelines apparently using them as concept markers (metonymy) for broader issues.

Conclusions. It seems that it would be difficult to automatically generate media timelines from contemporary discussions because of frame changes, and because of changes in terminology and the difficulty in separating highly discussed relatively insignificant details from the key facts. Nevertheless, the comparative analysis techniques reported in this paper may provide an interesting new window on contemporary discussions and media representations of major events.
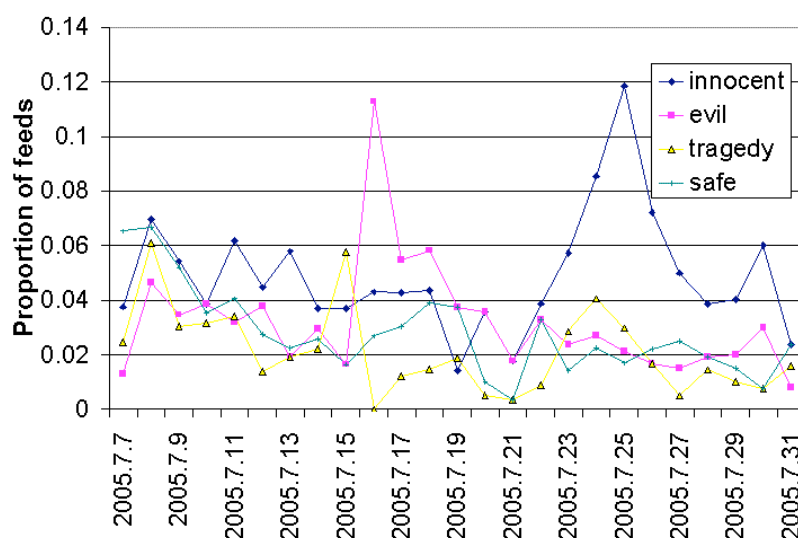
Time series for selected emotion-related terminology.

Triangulation of Network Metaphors: New Ways of Making Scientific and Public Avalanches Visible

I. Hellsten and A. Scharnhorst,
Innovation Networks – Lecture Notes Springer, Eds. A. Pyka, A. Scharnhorst (forthcoming)

In recent decades, the concept of "network" has gained popularity in various scientific and scholarly research fields. In the last years in particular the aspect of "complex networks" has gained attention. First used in the social psychology, life sciences and mathematics, it is now used in the various specialties in the social sciences and has become a scientific 'fashion' in physics. Moreover, physics has increasingly moved its research object from physical world to social phenomena largely due to access to large scale data sets on online databases.

The concept of "network", however, means different things for physicists, economists, social scientists and mathematicians. While "network" seems to provide a common interest for the different fields of research, it may also cause miscommunication between them. It functions in this sense as a metaphor that circulates between the various scientific domains of use. We aim at providing a meta-perspective on the concept of network by taking it as a metaphor, and wish to clarify the similarities and differences of the use of network. This may also facilitate the communication between the various fields of network research.

New species in evolving networks - stochastic theory of sensitive networks and applications on the metaphorical level

W. Ebeling, R. Feistel, I. Hartmann-Sonntag, L. Schimansky-Geier and A. Scharnhorst,

In this paper we develop a theory to describe stochastic influences on the fate of new species with nonlinear growth rates in evolutionary processes. We develop a theoretical framework based on a metaphoric use of notions as species, network, innovation, competition, survival and fitness. We introduce a stochastic picture describing the role of fluctuations for the survival of new species in non-linear systems. In particular we consider the fate of new species with nonlinear growth. As an application of the general model framework we consider the fate of 'rare species' in early biological evolution. Thus we move from the metaphoric level of the use of certain notions to a concrete context of interpretation. In general, we show that hyper-cycle systems do not represent the end of the evolutionary process as they may evolve further in small niches. This has implications for different types of applications ranging from biological and - on the metaphoric level - to socio-technological systems.

Citation and Hyperlink Networks

A. Scharnhorst and M. Thelwall,

From the very beginning the principle of citations indexing incorporated the idea of visualizing scientific information in the form of a graph. One prominent example is the concept of historiograph proposed by Garfield himself . In a historiograph, an evolutionary tree is constructed with a key (or parent) paper at the top followed by subsequent temporal layers of citing papers. Co-citation graphs set a mark in the visualization of scientific specialties and research fronts. Accompanied by co-author graphs, citation graphs and co-word graphs all these graphs have been used to visualize and analyze the growth of specialties, the structure inside scientific communities and the flow of information in science. Recently, a new branch in information science emerged which devotes itself to the visualization of knowledge domains. In this area tools have been developed to automatically create graphs, to explore different visualization approaches and to navigate through massive quantities of scientific information. Citation landscapes built in analogy to fitness landscapes make the occupation and evaluation of different scientific specialities visible and allow both for navigation through research fronts and identification of possible innovative areas. The visualization of emergent structures is not the only useful way to analyse collections of interconnected scientific documents, however. It is also possible to explore network structures by means of social network theory and measures developed in statistical physics. Complex network theory, a branch of statistical physics, mainly concentrates on analysing degree distributions (e.g. the numbers of citations documents in a collection receive), clustering coefficients and abstract theoretical mathematical models to explain the empirical findings. Social network analysis, in contrast, concentrates more on the interpretation of the social natures of the units and of the links between them. In this paper we will address both issues.

Circulation of Topics on the Web: The Bird Flu Hype

I. Hellsten and R. Prabowo,
in preparation

This paper focuses on the anatomy of the hype over bird flu in the e-news media, medical journals, and blogs in between January 1997 and June 2006. The argument is that public hypes feed themselves from sudden increase in the interactions between the three domains of medical sciences, news media and blogs, and this information avalanche between the domains may enhance the news avalanche. In particular, we are interested in finding new ways to detect such interactive momentums during public hypes. Web based media of communication, such e-news and blogs are expected to play potentially a crucial role in opening up an interactive momentum, first because they create context for rapid spread of discussions and topics, and second, because they facilitate copying some news accounts over and over again over the blogosphere, and third because they hence may accelerate the interaction and diffusion of certain pieces of information instead of others.

Emergent Forms of Crisis Communication: Detection and Comparison

M. Thelwall and D. Stuart,
submitted (2006)

Crises often have their own dynamics for information diffusion and past crises have revealed the particular value of emerging communication forms, such as blogs. In this paper we propose a method to identify and assess the usage patterns of emergent forms of crisis communication, using word usage statistics from Blogs and news feeds. Identification is achieved by manually scanning a list of words experiencing a sudden increase in usage during a crisis, selecting terms describing new forms of communication. Emergent forms of crisis communication are then assessed by comparing word usage statistics across crises and between different forms.
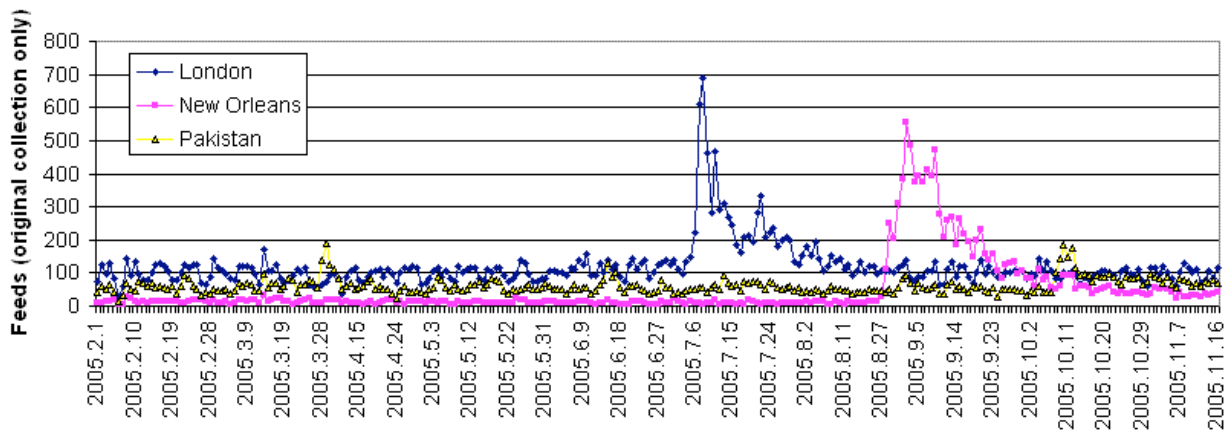
Number of Blog feeds encompassing specific words (London, New Orleans, Pakistan) as a function of time.

# Language Evolution and the Spread of Ideas on the Web

Word usage is of interest to linguists for its own sake as well as to social scientists and others seeking to track the spread of ideas, for example in public debates over political decisions. The historical evolution of language can be analysed with the tools of corpus linguistics through evolving corpora and the web. But word usage statistics can only be gathered for known words. In this article, techniques are described and tested for identifying new words from the web, focussing on the case when the words are related to a topic and have a hybrid form with a common sequence of letters. The results highlight the need to employ a combination of search techniques and show the wide potential of hybrid word family investigations in linguistics and social science.
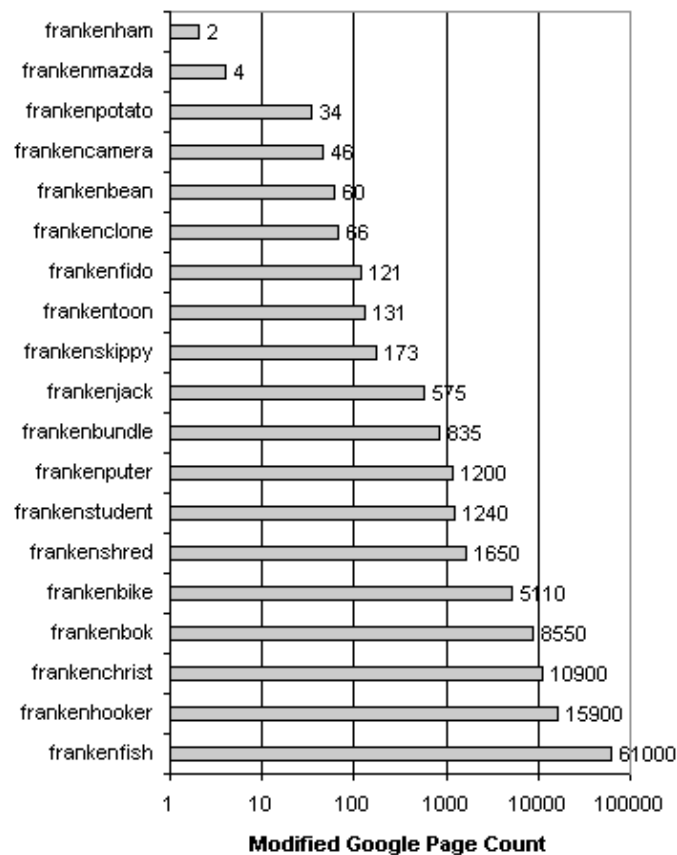
Modified Google Page Count of some word variations.

Are Raw RSS Feeds Suitable for Broad Issue Scanning? A Science Concern Case Study

M. Thelwall, R. Prabowo and R. Fairclough,
Journal of the American Society for Information
Science & Technology, **57(12)**, 1644-1654 (2006)

Broad issue scanning is the task of identifying important public debates arising within a given broad issue; Rich Site Syndication (RSS) feeds are a natural information source for investigating broad issues. RSS, as originally conceived, is a method for publishing timely and concise information on the Internet, for example about the main stories in a news site or the latest postings in a blog. RSS feeds are potentially a non-intrusive source of high quality data about public opinion: monitoring a large number may allow quantitative methods to extract information relevant to a given need. In this paper we describe an RSS feed-based co-word frequency method to identify bursts of discussion relevant to given broad issue. A case study of public science concerns is used to demonstrate the method and assess the suitability of raw RSS feeds for broad issue scanning (i.e. without data cleansing).
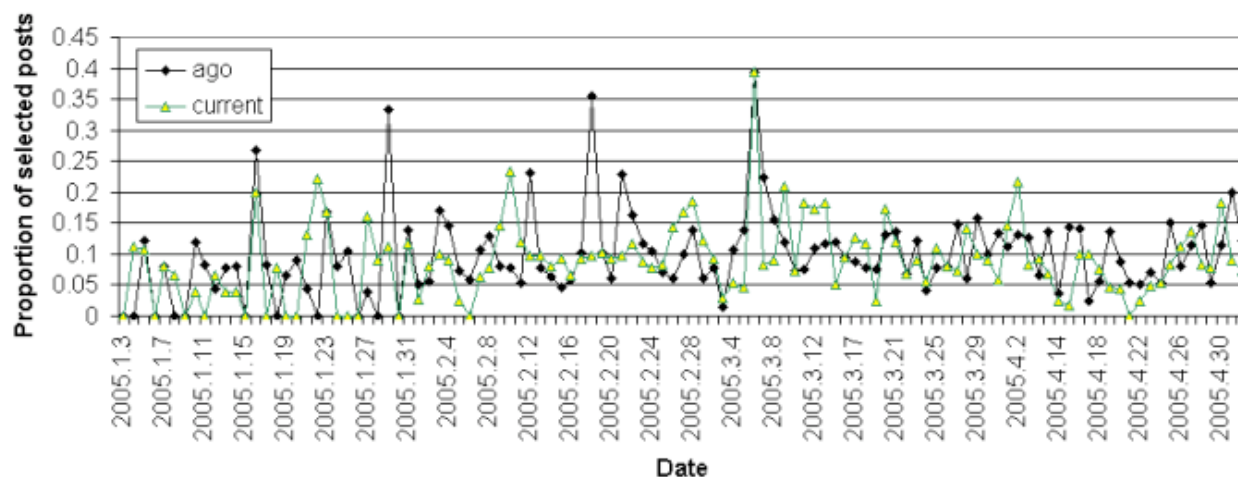
Science concern co-word time series for "ago" and "current", sharing a common main spike due to both occurring in the original post for a space shuttle story thread.

Identifying and characterising public science-related concerns from RSS feeds

M. Thelwall and R. Prabowo,
Journal of the American Society for Information Science & Technology (2007, to appear)

http://www.creen.org/articles/CREEN/wolverhamptonIdentifying.doc

A feature of modern democracies is public mistrust of scientists and the politicisation of science policy, for example concerning stem-cell research and genetically-modified food. Whilst the extent of this mistrust is debatable, its political influence is tangible. Hence science policy researchers and science policymakers need early warning of issues that resonate with a wide public. In this paper a semi-automatic method for identifying significant public science-related concerns from a corpus of internet-based RSS feeds is described and shown to be an improvement on a previous similar system because of the introduction of feed-based aggregation. In addition, both the RSS corpus and the concept of a public science-related concern are deconstructed, revealing hidden complexity.
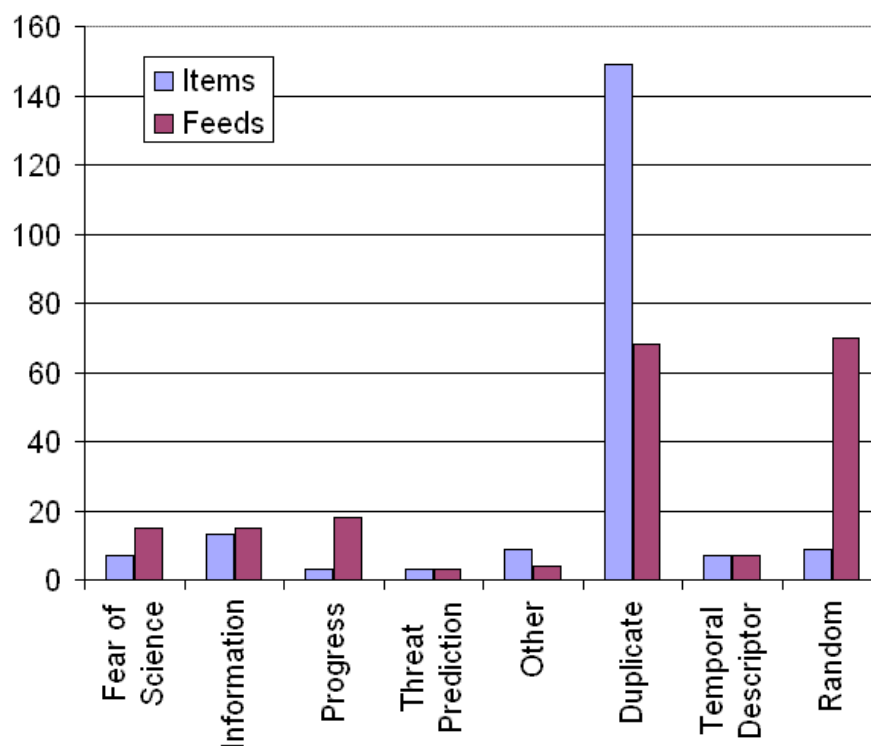
Classified story types associated with the top 200 science fear words.

A comparison of feature selection methods for an evolving RSS feed corpus

R. Prabowo and M. Thelwall,

http://www.creen.org/articles/CREEN/wolverhamptonComparison.pdf

Previous researchers have attempted to detect significant topics in news stories and blogs through the use of word frequency-based methods applied to RSS feeds. In this paper, the three statistical feature selection methods: $\chi 2$, Mutual Information (MI) and Information Gain (I) are proposed as alternative approaches for ranking term significance in an evolving RSS feed corpus. The extent to which the three methods agree with each other on determining the degree of the significance of a term on a certain date is investigated as well as the assumption that larger values tend to indicate more significant terms. An experimental evaluation was carried out with 39 different levels of data reduction to evaluate the three methods for differing degrees of significance.

$\chi 2$	I	MI	Date
4.06	3.60e-06	4.32	2004-10-12
46.20	1.20e-05	4.61	2004-12-26
287.60	3.71e-05	5.63	2004-12-27
34.70	1.07e-05	4.24	2004-12-28

An excerpt of the $\chi 2$, I, MI values, assigned to the term 'asian nations' which are sorted according to date. This table represents the functionality of feature selection methods.

Generating overview timelines for major events in an RSS corpus

R. Prabowo, M. Thelwall and M. Alexandrov,
Journal of Informetrics (2007, to appear)

Really Simple Syndication (RSS) is becoming a ubiquitous technology for notifying users of new content in frequently updated web sites, such as blogs and news portals. This paper describes a feature-based, local clustering approach for generating overview timelines for major events, such as the tsunami tragedy, from a general-purpose corpus of RSS feeds. In order to identify significant events, we automatically (1) selected a set of significant terms for each day; (2) built a set of (term – co-term) pairs, and (3) clustered the pairs in an attempt to group contextually related terms. The clusters were assessed by 10 people, finding that the average percentage apparently representing significant events was 68.6%. Using these clusters, we generated overview timelines for three major events: the tsunami tragedy, the US election and bird flu. The results indicate that our approach is effective in identifying predominantly genuine events, but can only produce partial timelines.

Bloggers during the London attacks: Top information sources and topics

M. Thelwall,

Blogs are probably most associated with the high profile postings of a few highly popular bloggers who debate or comment on major news stories, but for each 'A–lister' there are numerous faceless bloggers who write about their own daily lives and/or interests. Hence it is interesting to investigate the extent to which an event with extensive media coverage, such as the London attacks, is reflected in blogspace as a whole. This paper reports a descriptive analysis of blog postings around the London attacks of July 7, 2005. The core of this study is the development of methods to identify and report on bloggers' activities in a way that is not dominated by prolific bloggers or repetitive blog postings. We report daily trends for the top links and topics for three sets of data: all bloggers' postings; the postings of bloggers who mentioned London at least once; and the blog postings mentioning London. Although only 5% of active bloggers ever mentioned London by name, the attacks appeared to be the most significant event in blogspace during the two weeks after the initial bombings. Bloggers who posted about London were found to be atypical, linking and posting much more frequently than general bloggers. The results suggest a dichotomy between externally–focused, news–aware approximately daily bloggers and internally–focused diary–like approximately weekly bloggers.
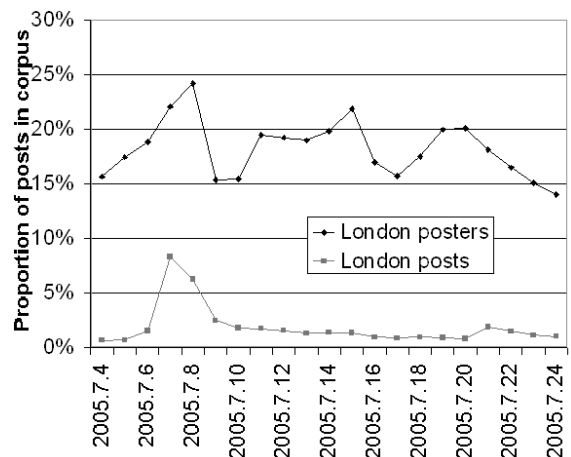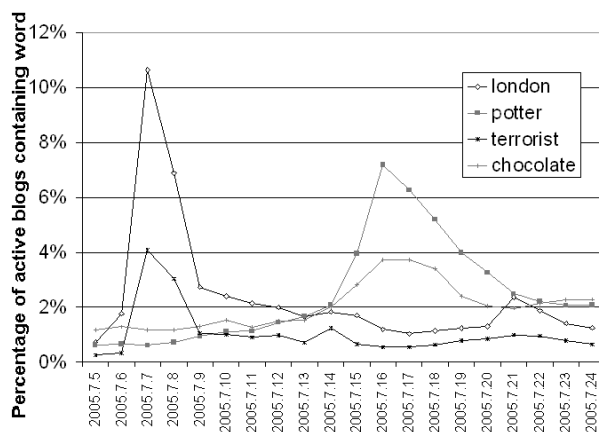
Proportion of active blogs in the full corpus

Blog searching: The first general–purpose source of retrospective public opinion in the social sciences?

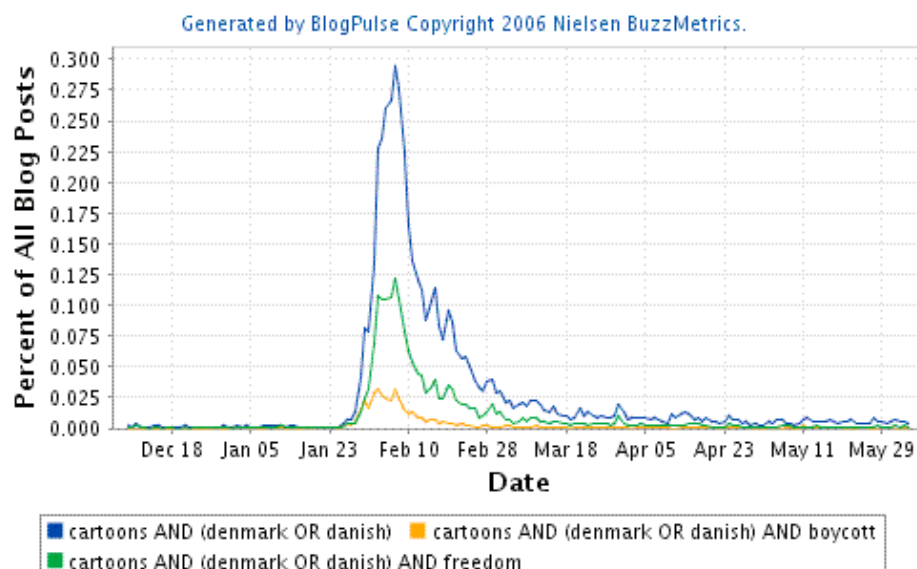M. Thelwall,
Online Information Review (2007, to appear)

Purpose – To demonstrate how blog searching can be used as a retrospective source of public opinion.

Design/methodology/approach -In this paper a variety of blog searching techniques are described and illustrated with a case study of the Danish cartoons affair.

Findings - A time series analysis of related blog postings suggests that the Danish cartoons issue attracted little attention in the English-speaking world for four months after the initial publication of the cartoons, exploding only after the simultaneous start of diplomatic sanctions and a commercial boycott.

Research limitations/implications – Blogs only reveal the opinions of bloggers, and blog analysis is language-specific. Sections of the world and the population of individual countries that do not have access to the internet will not be adequately represented in blogspace. Moreover, bloggers are self-selected and probably not representative of internet users.

Originality/value - The existence of blog search engines now allows researchers to search blogspace for posts relating to any given debate, seeking either the opinions of blogging pundits or casual mentions in personal journals. It is possible to use blogs to examine topics before they first attracted mass media attention, as well as to dissect ongoing discussions. This gives a retrospective source of public opinion that is unique to blog search engines.

The cartoons debate: volume of blog postings related to freedom and boycotts (reproduced with permission).

Which types of news story attract bloggers?

M. Thelwall, A. Byrne and M. Goody,
submitted (2007)

Blogs have been hailed as potential transformers of journalism and news values. Nevertheless, despite some major stories gestating in blogs, it is unclear how what types of news is discussed in blogs and hence the extent of potential blogspace influence. In response, we sampled 556 stories from four news web site home pages in June 2006. Each story was classified by topic, event typeand geography, and the number of relevant blog postings from the publication day was estimated. The results showed a surprisingly close average match between blogger interests and BBC, CNN, LA Times and Fox News coverage, probably because news sites tend to publish more stories of popular types. Further analysis suggested that blogs favour participatory events and right−wing stories, and hence may pull mainstream news in this direction. Additionally, we recommend simple guidelines for assessing whether individual news stories attract above average interest in blogspace.
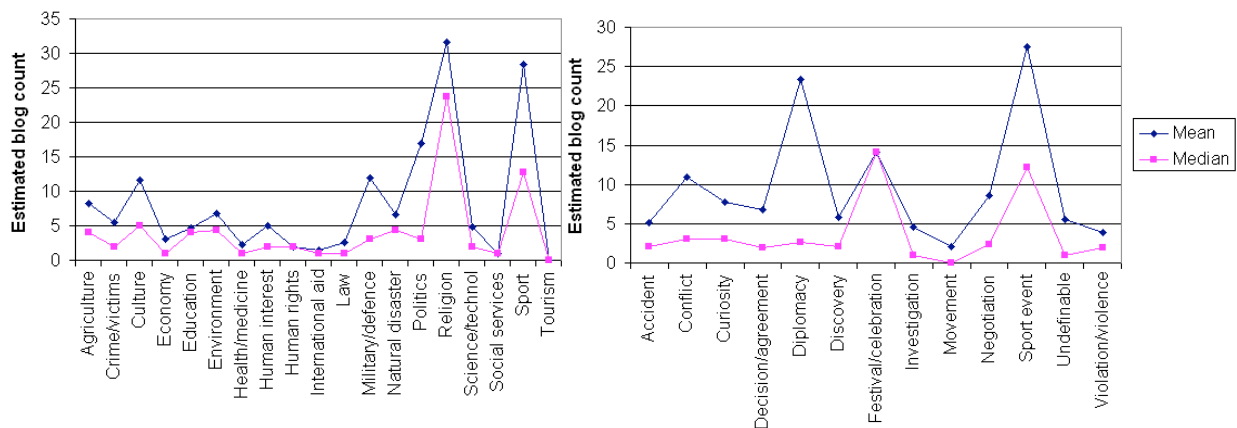
Event and topic category results (without the soccer world cup)

Approximating Clustering Coefficient and Transitivity

T. Schank and D. Wagner,
J. of Graph. Algor. and Appl. **9**, 265 (2005)

Since its introduction in the year 1998 by Watts and Strogatz, the clustering coefficient has become a frequently used tool for analyzing graphs. In 2002 the transitivity was proposed by Newman, Watts and Strogatz as an alternative to the clustering coefficient. As many networks considered in complex systems are huge, the efficient computation of such network parameters is crucial. Several algorithms with polynomial running time can be derived from results known in graph theory. The main contribution of this work is a new fast approximation algorithm for the weighted clustering coefficient which also gives very efficient approximation algorithms for the clustering coefficient and the transitivity. We namely present an algorithm with running time in $O(1)$ for the clustering coefficient, respectively with running time in $O(n)$ for the transitivity. By an experimental study we demonstrate the performance of the proposed algorithms on real-world data as well as on generated graphs. Moreover we give a simple graph generator algorithm that works according to the preferential attachment rule but also generates graphs with adjustable clustering coefficient.

Algorithm 1: *forward*

Input: ordered list (high degree first) of vertices $(1, \ldots, n)$; Adjacencies $Adj(v)$
Data: Node Data: $A(v)$;
for $v \in V$ **do**
   $A(v) \leftarrow \emptyset$
for $s \in (1, \ldots, n)$ **do**
   for $t \in Adj(s)$ **do**
     if $s < t$ **then**
       foreach $v \in A(s) \cap A(t)$ **do**
         output triangle $\{v, s, t\}$;
       $A(t) \longleftarrow A(t) \cup \{s\}$;

Pseudo-code of the Algorithm forward, that is an improvement of edge-iterator.

Finding, Counting and Listing all Triangles in Large Graphs

T. Schank and D. Wagner,
Lect. Notes in Comp. Sc. **3503** (2005)

http://www.creen.org/articles/CREEN/karlsruheTriangles.pdf

The two known standard Algorithms node-iterator and edge-iterator are asymptoticly equivalent. However, the Algorithm edge-iterator can be implemented with a much lower constant overhead. It works very well for graphs where the degrees do not differ much from the average degree. If the degree distribution is skewed, refined algorithms are required. The Algorithm forward shows to be the best compromise. It is asymptotically efficient and can be implemented to have a low constant factor with respect to execution time.
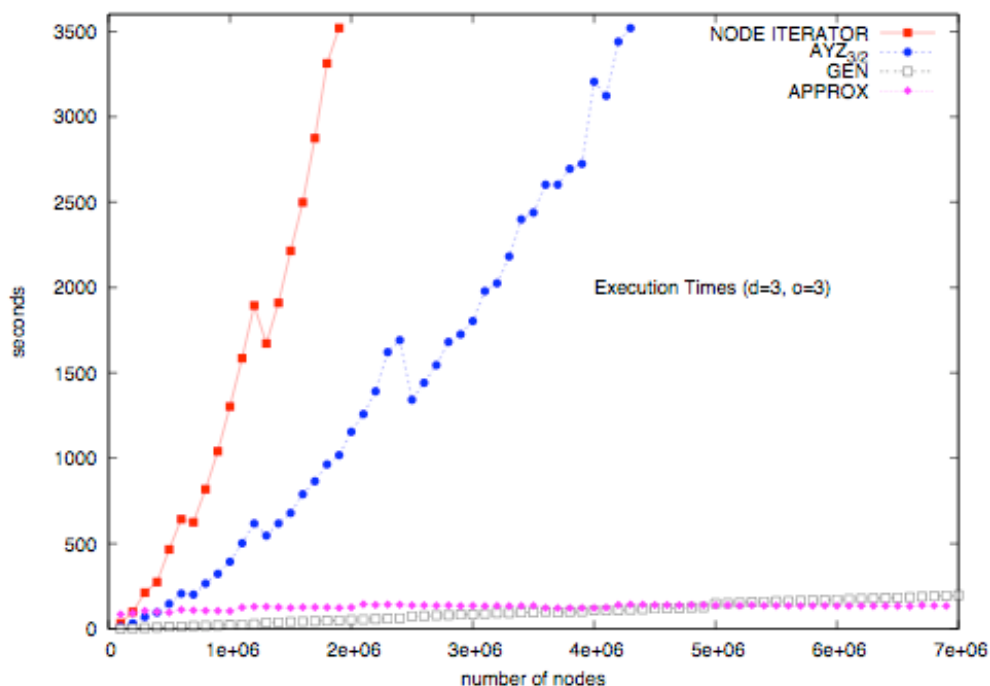
Execution times for a series of graphs.

A Hybrid Model for Drawing Dynamic and Evolving Graphs

M. Gaertler and D. Wagner,

http://www.creen.org/articles/CREEN/karlsruheHybrid.pdf

Dynamic processes frequently occur in many applications. Visualizations of dynamically evolving data, for example as part of the data analysis, are typically restricted to a cumulative static view or an animation/sequential view. Both methods have their benefits and are often complementary in their use. In this article, we present a hybrid model that combines the two techniques. This is accomplished by 2.5D drawings which are calculated in an incremental way. The method has been evaluated on collaboration networks.
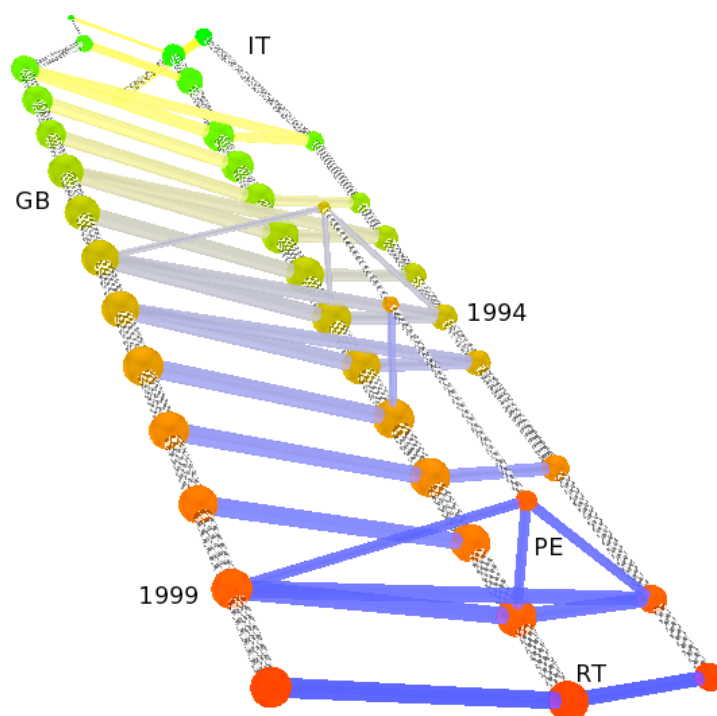
Collaboration between Giuseppe Di Battista (GB), Ioannis Tollis (IT), Peter Eades (PE) and Roberto Tamassia (RT) between 1986 and 2000.

Analysis of Overlay-Underlay Topology Correlation using Visualization

V. Aggarwal and A. Feldmann and M. Gaertler and R. Gorke and D. Wagner,

In the design and implementation of the overlay architecture most peer-to-peer (P2P) systems rely on the underlay network to provide them with basic connectivity. Therefore, the intrinsic features of the underlay network determine the efficiency of the overlay. Accordingly, studying the interdependency of the overlay and underlay networks leads to a better understanding of P2P behaviour. We present a visualization-driven analysis for evaluating the overlay architecture with respect to the underlay. Using Gnutella as a case study, our analysis confirms that Gnutella's topology differs from a randomly generated network and that there is an implicit correlation between the overlay and underlay topologies.

Comparison of occuring communication in the Gnutella network (left) and a randomly (right) generated network.

Generating Significant Graph Clusterings

D. Delling and M. Gaertler and D. Wagner,
Proceedings of the European Conference of Complex Systems ECCS '06

Many applications such as experimental evaluations of clustering algorithms require the existence of a significant reference clustering. This task is dual to finding significant clusterings of a given graph. We present several generators for pre–clustered graphs based on perturbation and geometry. In an experimental evaluation we confirm the applicability of our generators. Furthermore, the presented results lead to a better understanding of the correlation between the degree of perturbation and significance.
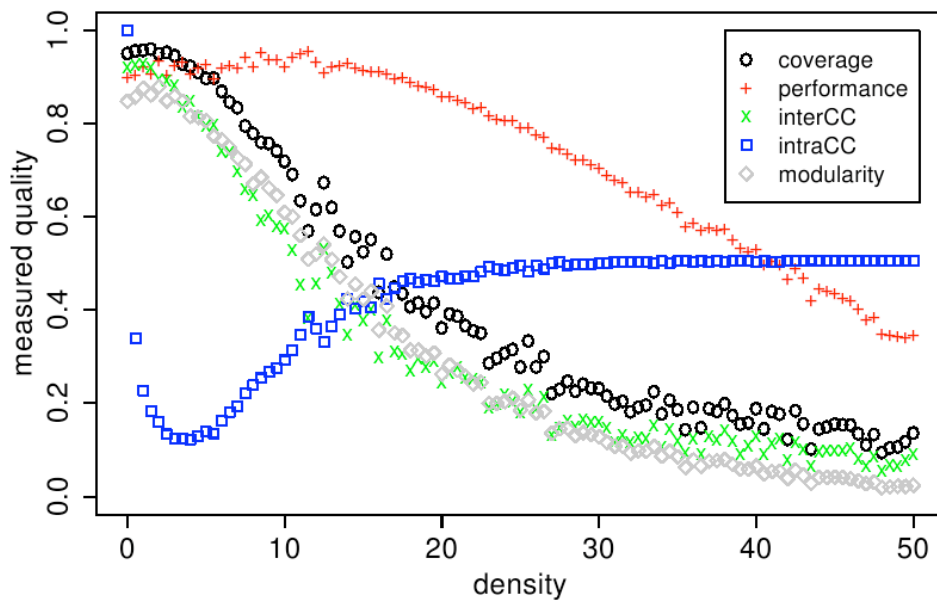
The measured quality for attractors

III. CREEN in press

I. Article "Naukowe Lawiny" (Scientific Avalanches) about CREEN project by Ewa Chybińska published in scientific magazine "Forum Akademickie" No 3/2005 (http://www.forumakad.pl/archiwum/2005/03/02-kronika.htm);

II. Article "Fizycy podgrzewają atmosferę" ("Physicists warm up the atmosphere") published in WUT's monthly magazine;

III. Article "Politechnika kieruje europejskim projektem" (University of Technology co-ordinates European Project) published in one of the most popular Polish daily newspaper "Gazeta Wyborcza" No 15/1/05 (http://www.creen.org/);

IV. Article "Des physiciens de l'Universite de Liege ont etudié le partage des musiques sur les sites internet" (University of Liege physicists have studied music sharing through internet websites) published in journal "Le soir" (11/01/2006, http://www.lesoir.be);

V. Article "La musique dans tous ses états. Nouvelles musiques, nouvelles sciences, nouveaux réseaux" (Music in all its states. New musics, new sciences, new networks) published in the Liège University journal (http://www.ulg.ac.be/le15jour/149/musique.shtml);

VI. Article "What music do you like" published in the volume of October of Physicsweb (http://physicsweb.org/articles/news/9/9/7/1);

VII. Article "De la physique statistique à la musique " (From statistical physics to music) published in journal "La Libre Belgique" (18/01/2006, http://www.lalibre.be/article.phtml?id=5&subid=104&art_id=263160);

VIII. Renaud Lambiotte has been invited to two national radios in order to present results obtained in the context of CREEN (13/01/2006 on Vivacité and 23/01/2006 on Pure FM).