Drastic events make evolving networks

Ausloos, M.; Lambiotte, R.

Published in:
European Physical Journal B

DOI:
10.1140/epjb/e2007-00159-6

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (HARVARD):
Ausloos, M & Lambiotte, R 2007, 'Drastic events make evolving networks' European Physical Journal B, vol. 57, no. 1, pp. 89-94. https://doi.org/10.1140/epjb/e2007-00159-6

Eur. Phys. J. B **57**, 89–94 (2007)
DOI: 10.1140/epjb/e2007-00159-6

**THE EUROPEAN
PHYSICAL JOURNAL B**

Drastic events make evolving networks

M. Ausloos[a] and R. Lambiotte[b]

GRAPES, Université de Liège, B5 Sart-Tilman, 4000 Liège, Belgium

Received 5 February 2007 / Received in final form 23 April 2007

Abstract. Co-authorship networks of neighbouring scientific disciplines, i.e. granular (G) media and networks (N) are studied in order to observe drastic structural changes in evolving networks. The data is taken from arXives. The system is described as coupled networks. By considering the 1995–2005 time interval and scanning the author-article network evolution with a *mobile* time window, we focus on the properties of the links, as well as on the time evolution of the nodes. They can be in three states, N, G or multi-disciplinary (M). This leads to drastic jumps in a so-called order parameter, i.e. the link proportion of a given type, forming the main island, that reminds of features appearing at percolation and during metastable (aggregation-desaggregation) processes. The data analysis also focuses on the way different kinds (N, G or M) of authors collaborate, and on the kind of the resulting collaboration.

PACS. 89.75.Fb Structures and organization in complex systems – 89.75.Hc Networks and genealogical trees – 87.23.Ge Dynamics of social systems

1 Introduction

Since the pioneering works of Barabási and Albert [1,2], *"complex networks"* have become a more and more active field, attracting physicists from the many sub-fields pertaining to non-equilibrium statistical physics. Those complex structures are usually composed of a large number of internal components (the nodes), and describe a wide variety of systems of high intellectual and technological importance, examples including the Internet [3], business relations between companies [4], ecological networks [5], airplane route networks [6] ... As a paradigm for large-scale networks, people usually consider co-authorship networks [7], namely networks where nodes represent scientists, and where a link is drawn between them if they co-authored a common paper. Their study has been very active recently, due to the complex (social) structure [8], to the ubiquity of their bipartite structure in complex systems [9,10], and to the large databases available (arXiv and Science Index).

Relevant and remaining questions pertain, not only to the dynamics of properties on the network, but also to the network structure itself, e.g. clique formation and clustering. In this line of thought, the study of coupled networks (networks composed of several kinds of nodes) is of interest in order to account for the specificity of the nodes, for instance in social networks where such specificities prevent links to develop between any kind of nodes. In this paper, we analyze freely available data for collaboration networks, for which such competing specificities play an important role. To do so, we focus on the development of neighbouring scientific disciplines in the course of time, thereby eyeing the spreading of ideas in the scientific community. We will ask whether or not a co-author plays a role in the change of disciplines of interest to scientists. The data analysis will highlight that most contacts between the two disciplines are driven by inter-disciplinary collaborations, allowing interface propagation, and also reveals interesting time-dependent properties. We will show below that some behavior is similar to features arising at metastable equilibria, and can found its basic root in percolation ideas.

Let us stress that the identification of the mechanisms responsible for diffusion and, possibly leading to scientific avalanches, is primordial in order to understand the scientific response to external field (e.g. political) decisions, and to develop efficient policy recommendations. No need to say that most of the ideas here below developed can encompass many networks, not only formed by scientists,but by many other agents in which some clusters and interfaces can be intuitively imagined.

In this article, we propose a novel approach where the scientific system is seen as coupled networks, where the nodes (agents) themselves evolve in a ferro-electric-like way [13], i.e. the state of the nodes is defined through the link nature (with history) itself. The empirical features are reminiscent of dynamical phase transitions [14], like percolation or fracture [15,16], glass ageing [17] and other agglomeration-desaggregation processes [18].

---

[a] e-mail: `Marcel.Ausloos@ulg.ac.be`

[b] e-mail: `Renaud.Lambiotte@ulg.ac.be`

After a description of the description of the data acquisition (Sect. 2.1) we present the methodology (Sect. 2.2). The co-author role and the network co-evolution are found in Sections 2.3, 2.4. We (explain how to and) measure the probability of co-working and its role on the network evolution in Section 2.5. A brief conclusion is found in Section 3.

2 Data analysis

2.1 Overview

For the present purpose, let us concentrate empirically on data extracted from the arXiv database. To do so, we discriminate two sub-communities of physicists, those studying *"complex networks"* (N) and those studying *"granular media"* (G) . This choice is motivated by the relative closeness of these fields, that intuitively allows interactions between sub-communities (inter-disciplinarity collaboration), and the passage of a scientist from one field to the other (scientist mobility).

The data set contains all articles from arXiv in the time interval [1995–2005], that contain the word *network* or (exclusive "or") the word *granular* in their abstract and are classified as *"cond-mat"*. In the following, we assume that this simple semantic filter is sufficient to distinguish the specialty papers, whence that of scientists. We recognize that the method does not ensure a perfect characterization of the paper subject [19], but we accept such an approximation thereafter. In order to discriminate the authors and avoid spurious data, we checked the names and the first names of the authors. Moreover, in order to avoid multiple ways for an author to cosign a paper, we also took into account the initial notation of the pre-names (see [20] for details on the data acquisition). Given this identification method, we find 3297 scientists and 2305 articles. Among these scientists, 105 have written their articles by themselves, i.e. without co-author. As these people are excluded from the co-authorship network, we neglect them immediately in the following. There are 150 scientists who wrote articles in both fields. These authors are by definition multi-disciplinary scientists, and thereby ensure direct communication between the two scientific disciplines. For giving some scaling factor let us mention that among the 3192 remaining scientists, 2270 ones have written *at least* one *network* article, and 1072 ones have written *at least* one *granular* article.

The histogram of the total number of co-authors per author for the whole considered time interval is given in Figure 1. The plot reveals that some authors have many co-authors: one person worked with 42 colleagues, though most of the authors have only a few co-authors; e.g. about 1000 authors have at most 2 co-authors. The average number of co-authors/author in the dataset is 3.98.

2.2 Methodology

In order to build the co-authorship network, we apply Newman method [21], namely we consider a network of scientists placed at nodes, with a link between them if they co-authored a common paper. Consequently, a link between authors correspond to one article that they wrote together *there may be several links between two nodes*), each of the link being tagged *granular* or *network*. In addition, we also discriminate scientists as *granular* (G), *network* (N) or *multi-disciplinary* (M) scientists, depending on their collaborations. By convention, if more than 80% of the links are *granular* (*network*), the author is considered as a *granular* (*network*) *scientist*. Else, the author is a *multi-disciplinary* one, such that the system can be viewed as a network composed of different kinds of nodes and links.

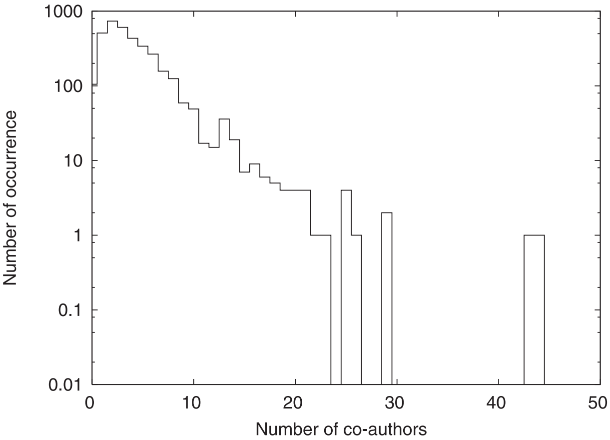

Fig. 1. Histogram of the total number of co-authors per author having written a N or G paper during the [1995–2005] epoch, having put their paper on the arXives, during the whole considered time interval for the fields of research hereby considered.

Having accumulated the data throughout the whole time interval, the system in 2005 is observed to be formed by a large "continent" with 1180 scientists belonging to the three genres, beside a multitude of disconnected "islands" (I) with various sizes — reminiscent of the clique habits of authors. The main island ($I = 1$) exhibits typical features of social networks, i.e. strongly connected scientists (recall Fig. 1) [22].

It is also very important to emphasize at once the existence of rather large regions of the network, made of mainly *granular* or *network* nodes, thereby confirming that authors collaborate primarily with others with whom their research focus is aligned [23]. This indicates the presence of homogeneously connected phases in a thermodynamic sense.

2.3 Role of the co-author

One may reasonably ask whether or not a co-author plays a role in the changes of disciplines of scientists. To answer this question, we focus on the scientists i (i = 1, 150) who wrote articles in *both* disciplines and are nodes characterized by L_G^i and L_N^i *granular*/*network* links. By definition they form the interface of the network. Data analysis shows that the total number of links

$L = \frac{1}{2}\sum_{i=1}^{150} L_G^i + L_N^i = 733$. In contrast, the total number of collaboration pairs (i, j) that are related by **both** kinds of links, i.e. number of pairs of scientists working in both fields together, is equal to 63, so that there are 126 such links. This suggests that when a scientist works in two fields, he works in each field with different persons [24]. This conclusion will be confirmed in a more quantitative analysis in Section 2.5.

2.4 Co-evolution

Let us now focus on the dynamical processes that take place in the system, i.e. the co-evolution of the network structure and the node/link nature. To do so, we find that it is of interest to study overlapping time windows of 3 years. E.g. we start in July 1996 and move the window forward in time by small intervals of 1 month. This method ensures a smooth time evolution of the different variables [25]. We decided to characterize time windows by the date at the center of the interval, i.e. we denote the interval [01/2002; 12/2004] by 07/2003, thus the time axis starts in 1998.

Moreover, it seems reasonable to focus our attention on the main island(s) only. Each island I is characterized by its proportion of *network* links, $p_N^I = L_N^I/L_T^I$, where L_N^I and L_T^I are its number of *network* links and its total number of links respectively (Fig. 2a), and by its number of nodes N^I (Fig. 2b). The time evolution of the proportion of *network*, *granular* or *multi-disciplinary* nodes in the main island is shown in Figure 3. The time evolution of these quantities is also rather smooth, except at four important dates (1)–(4) where sudden jumps take place.

- Between 1998 and Nov. 2000 (1), the system is stable and p_N in the main ($I = 1$) island, i.e. p_N^1 remains around 0.2. In other words, the continent is less N than G ($p_N^1 < 1/2$).
- Around Nov. 2000 (1), the system shows a strong increase of p_N^1, i.e. p_N^1 jumps to 0.3 and goes to 0.6 thereafter, i.e. the continent switches (makes a transition) from a *granular* state to a *network* state. Detailed analysis shows that, before this event, the largest island ($I = 1$) encompasses scientists like H.J. Herrmann (let us note this island I_H), and the next-to-largest island ($I = 2$) is centered around A.L. Barabási (I_B). It is also found (Fig. 2b) that the number of nodes N_{I_B} in I_B grows faster than in I_H, and that N_{I_B} exceeds N_{I_H} precisely at the moment (1), thereby leading to a sudden change in the properties of the largest island and a *discontinuity* in its so called (order) parameter p_N^1.
- After Nov. 2000 the number of network links much increases, as indicated till 0.65, approximately proportional to the increase in nodes of the (Barabasi) now new main island.
- Around May 2002 (2), a second drastic event takes place, associated to a negative jump in p_N^1. As seen in Figure 2b, the number of nodes in the largest island suddenly starts to increase at that time, and becomes roughly equal to the sum of N_{I_H} and N_{I_B}. We conclude that this sudden increase is due to the merging of the islands I_H and I_B that leads to a dilution [26] of the number of *network* links and to a drastic drop in p_N^1 (and, accordingly, to an increase in the proportion of *granular* links in the largest island).
- Near August 2002 (3) and October 2002 (4), negative and positive jumps (from 0.65 to 0.4) in p_N^1 are observed. From Figure 2b, one notes that, at (3), the two islands separate again while they re-collide two months later. This observation shows that the jumps in p_N^1 are due to fluctuations close to some sort of percolation point, i.e. during the merging of the two types of phases. This merging-rupture process is illustrated through Figure 4.
- After (4), the proportion of N links increases continuously, due to the overall more rapid growth of the *network* field as compared to the *granular* field.

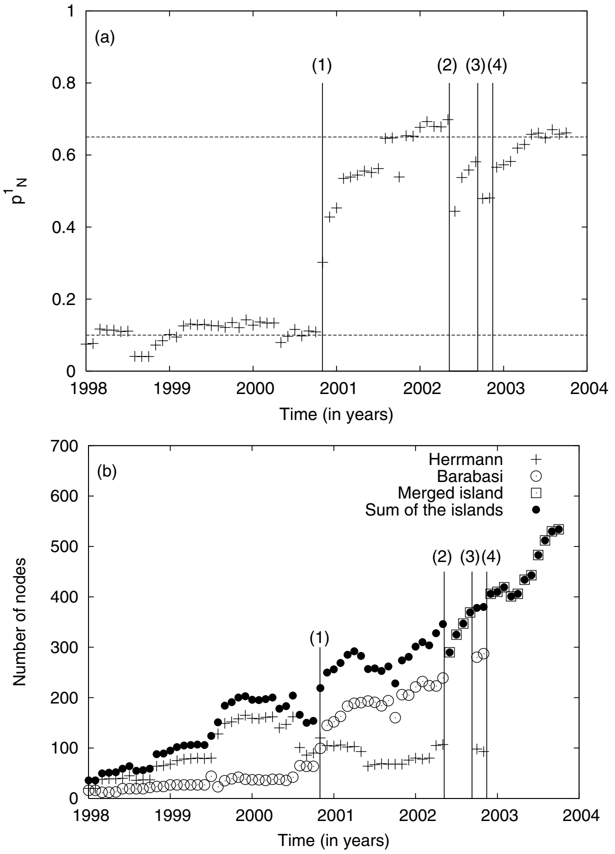

Fig. 2. In (a), time evolution of the proportion of *network* links p_N^1 in the main island. In (b), time evolution of the number of nodes (scientists) in the I_H, I_B and the merged island. The sum $I_H + I_B$ is also shown for comparison (see text). The vertical lines point to the key events occurring in the system: (1) the "positive" transition from a *granular* to a more *network* phase; (2) and (4) the collision between I_H and I_B and their merging; (3) the desagregation/rupture of the main island during the Aug.–Oct. 2002 months.

The above phenomena can also be observed as a function of time, in Figure 3a, where we plot the evolution of the proportion of *granular*, *network* or *multi-disciplinary*

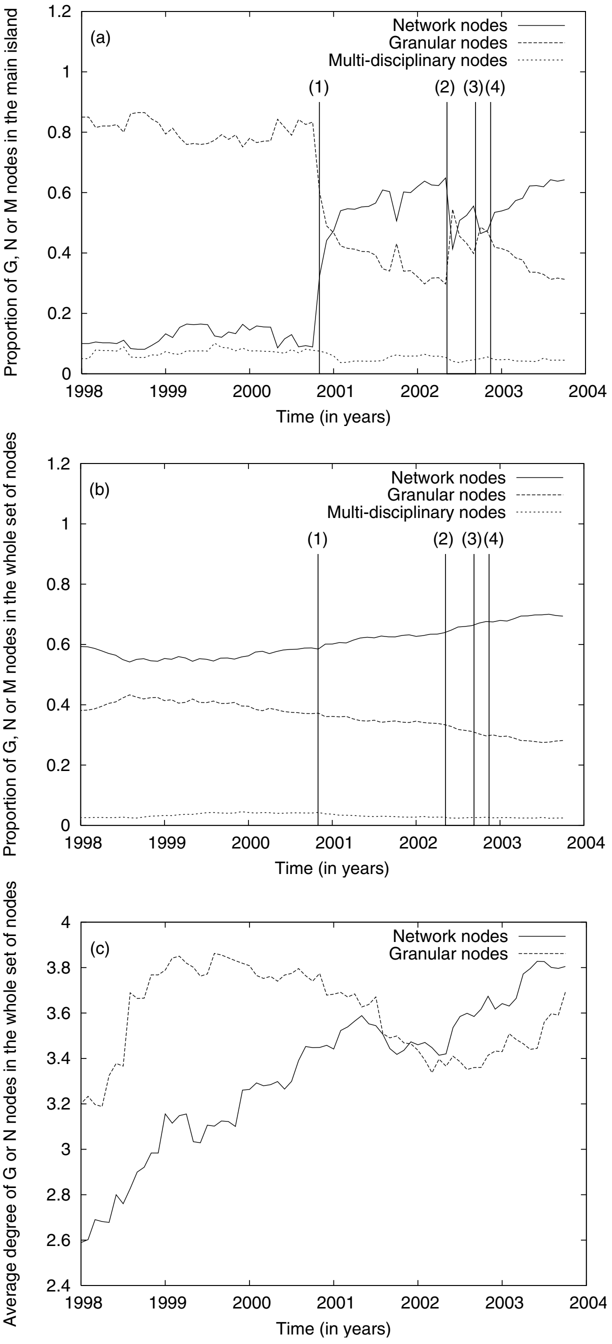

Fig. 3. Time evolution of the proportion of *network*, *granular* or *multi-disciplinary* nodes in the main island (a) and in the whole set of nodes (b). In (c), we plot the time evolution of the average degree of nodes of type G or N in the whole set of nodes.

nodes in the main island. Jumps similar to those of Figure 2a also take place for each of these quantities. In contrast (Fig. 3b), the evolution of these proportions is rather continuous when they are measured for the whole set of nodes. Moreover, it is observed from the latter figure that the proportion of authors belonging to the *network* kind increases rather monotonically since 1999, in contrast to the *granular* set. It is also important to stress that there are more *network* nodes than *granular* nodes before the the event (1), when the main island is composed of a majority of *granular* nodes. This suggests that the collaboration network of the *granular* authors was more dense during this time period (1999–2001), i.e. *granular* authors were more connected. This is verified by comparing the average degree of *granular* nodes with the average degree of *network* nodes, as shown in Figure 3c.

2.5 Interaction probabilities

Let us now focus more quantitatively on the processes taking place during the network evolution. To do so, we study how scientists are influenced by the people with whom they collaborate in order to discover new fields. We proceed as follows. We consider articles *co-authored by at least 2 scientists* (the other articles are not interesting in the present analysis) in a chronological way.

Let us first define two quantities $W_{\alpha,\beta}$ and $Y_{\alpha,\beta}$ with index running in the set (G, N, M) (*granular*, *network* or *multi-disciplinary*) such that $W_{\alpha,\beta}$ is the total number of articles co-authored by authors with kinds α and β, while $Y_{\alpha,\beta}$ is the total number of *network* articles co-authored by authors of kinds α and β. These quantities will be used in the following in order to evaluate the probability for two authors α and β to collaborate, as well as the nature of the resulting collaboration.

The counting begins by initializing the number of articles L_N^i and L_G^i for each author i and focusing on the first published article of the dataset. By construction, the authors have not written an article together before, so that one can not estimate whether the authors are G, N or M. Before focusing on the next article, we update the values of L_N^i and L_G^i (changes due to the first article) and define the author type (N, G, or M) according to the value of L_G^i/L^i where $L^i = L_N^i + L_G^i \neq 0$. One proceeds recurrently. If at any step τ, the two authors had already published papers in the past ($L^i \neq 0$ and $L^j \neq 0$), so that one can assign to the collaborating pair a kind $(\alpha, \beta) \in$ (G, N, M), one updates the quantity $W_{\alpha\beta}$ with 1 unit. One also updates $Y_{\alpha\beta}$ *if* the current article is of the *network* type. Let us insist on the fact that these increments are performed before updating the type (N, G, or M) of the authors due to article τ. At the end of the run, two probabilities can be extracted from the above quantities. After normalizing, the quantity $q_{\alpha\beta} = W_{\alpha\beta}/W$, where W is the sum over all values of $W_{\alpha\beta}$, is the probability that two agents α and β interact and write a paper together. Moreover, the quantity $p_{\alpha\beta} \equiv Y_{\alpha\beta}/W_{\alpha\beta}$ is the probability that the paper written by two authors of type α and β is a *network* paper; $(1 - p_{\alpha\beta})$ is their probability to write a granular paper together. In summary, the quantities $q_{\alpha,\beta}$ and $p_{\alpha,\beta}$ are respectively the probability for two authors (α, β) to interact and the probability that this collaboration is a network article.

The final results are summarized in Tables 1 and 2. One observes two important properties: (i) from Table 1, it is found that authors of different types have a lower probability to interact than authors of the same type. This

Fig. 4. Demonstration of the merging of the two main sub-islands in the system. The left and right figure correspond respectively to September 2002, between events (3) and (4) and January 2003, after event (4) snapshots.

Table 1. Values of $q_{\alpha\beta}$, i.e. probability that a pair of authors (α, β) of type N, M or G collaborates.

	NN	NM	MM	GN	GM	GG
$q_{\alpha\beta}$	0.44	0.09	0.02	0.03	0.07	0.35

Table 2. Probability $p_{\alpha\beta}$ that a pair of (α, β) authors writes a network paper, if they collaborate; $(1 - p_{\alpha\beta})$ is the probability that they write a granular paper together.

	NN	NM	MM	GN	GM	GG
$p_{\alpha\beta}$	0.99	0.97	0.56	0.46	0.10	0.02
$(1 - p_{\alpha\beta})$	0.01	0.03	0.44	0.54	0.90	0.98

is obvious by comparing q_{GG}, q_{NN} and q_{GN}; (ii) from Table 2, an expected feature is confirmed, since $p_{\alpha,\beta}$ is very large for identical pairs of author types, i.e. granular (network) authors, when working together have an almost certitude to work in granular (network) media, while a collaboration with scientists from another field network (granluar) triggers their probability to work in this other field.

3 Conclusion

Here above, we have empirically focused on data collected from the arXiv database, in order to observe the emergence of crises and trends in complex networks, hereby scientific avalanches so highlighting a rich and complex phenomenology. We have focused on an empirical case study in order to understand the way networks composed of many kinds of nodes and links evolve in a self-organized way. In the present case, the network consist of authors (nodes) and articles (links) drawn between authors if they have co-authored an article. The articles may be of two kinds (G, N) and the kind (G, N, M) of an author is found from his previously written articles.

We have observed that inter-connections between distinct scientific disciplines play a central role. It has been shown that attachment-detachment processes may lead to a percolation-like effect, seen as drastic jumps in the concentration of nodes of a certain kind in the system. It is also important to stress that these drastic jumps are observable at the level of the main island and but are hidden when examining the data over the whole set. For further work, necessarily accumulating much data, one could consider to describe the system evolution through some Langevin equation, and verify whether the system statistics obey a fluctuating mass process [27] — even going further than in the latter consideration, i.e. taking into account external mass (here nodes) addition as a function of time.

We have also shown that the community of G and N is smoothly time dependent, but that the G community is more tightly grouped than the N. Correlation functions in the main island and between clusters might be complementarily investigated in order to emphasize the jump mechanisms, the emergence of hierarchy and modularity in such evolving networks [28–31].

Finally, we have examined the way node kinds (G, N, M) interact with each other, thereby showing that the author evolution, or mobility (change of scientific field), is triggered by his collaborations. The above empirical results open perspectives for the modeling of author mobility that differ from classical approaches (e.g. master equations with auto-catalytic processes [11], epidemic models on static networks [12]...) by accounting for the interplay between the node specialisation and its surrounding collaboration network. We stress that features are *qualitatively* those of percolation systems, but within a ferro-electric models spirit, with dynamics driven by the collaboration links, and not by the "spin" attached to the nodes.

Figure 4 has been plotted with *visone* [32]. We would like to thank A. Scharnhorst, A. Fronczak, I. Hellsten, K. Suchecki and J.A. Hołyst for fruitful discussions. This work is part of the European Commission Project CREEN FP6-2003-NEST-Path-012864 which supports RL.

References

1. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
3. R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, 2004)
4. S.-M. Yoon, K. Kim, e-print `arXiv:physics/0503017`
5. R.J. Williams, N.D. Martinez, Nature **404**, 180 (2000)

6. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004)
7. M.E.J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001)
8. M.E.J. Newman, D.J. Watts, S.H. Strogatz, Proc. Natl. Acad. Sci. USA **99**, 2566 (2002)
9. A.L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T. Vicsek, Physica A **311**, 590 (2002)
10. J.J. Ramasco, S.N. Dorogovtsev, R. Pastor-Satorras, Phys. Rev. E **70**, 036106 (2004)
11. E. Bruckner, W. Ebeling, A. Scharnhorst, Scientometrics **18**, 21 (1990)
12. J.A. Holyst, K. Kacperski, F. Schweitzer, Ann. Rev. Comput. Phys. **9**, 253 (2001)
13. F.Y. Wu, Phys. Rev. Lett. **22**, 1174 (1969)
14. R.A. Blythe, J. Phys.: Conf. Ser. **40**, 1 (2006)
15. D. Stauffer, A. Aharony, *Introduction to Percolation Theory,* 2nd edn. (Taylor and Francis, London, 1994)
16. H. Auradou, M. Zei, E. Bouchaud, Eur. Phys. J. B **44**, 365 (2005)
17. C.J. Ellison, S.D. Kim, D.B. Hall, J.M. Torkelson, Eur. Phys. J. E **8**, 155 (2002)
18. J. Rubi, A. Gadomski, Physica A **326**, 333 (2003)
19. E.g. some *"network papers"* may not focus on complex networks, such as *The response function of a sphere in a viscoelastic two-fluid medium* by Levine and Lubensky (Phys. Rev. E **63**, 041510 (2001)) which focuses on microrheology of *polymer* networks
20. R. Lambiotte, M. Ausloos, Phys. Rev. E **72**, 066117 (2005)
21. M.E.J. Newman, Phys. Rev. E **64**, 016132 (2001)
22. M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E **64**, 026118 (2001)
23. M. Girvan , M.E.J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002)
24. However, there are notable exceptions, such as the triplet made by (F. Coppex, M. Droz, A. Lipowski) who work *together* on *both* subjects (see `arXiv:q-bio.PE/0312030` and `arXiv:cond-mat/0205058` for instance), and form a small island
25. Noise is intrinsic to this sort of data as anyone having put a paper on the arxives knows
26. It can be observed from Figure 2b that the number of nodes of type N is about 140 and G *ca.* 60, before event (2), and remains approximately the same after event (2), but since the island ($I = 2$) is essentially made of G, the proportion of N in the main island falls to 0.40 (Fig. 2a) after the event
27. M. Ausloos, R. Lambiotte, Phys. Rev. E **73**, 11105 (2006)
28. E.A. Variano, J.H. McKoy, H. Lipson,Phys. Rev. Lett. **92**, 188701 (2004)
29. R. Lambiotte, M. Ausloos, e-print `arXiv:physics/0703266`
30. G. Palla, A.-L. Barabási, T. Vicsek, Nature **446**, 664 (2007)
31. R. Lambiotte, M. Ausloos, J.A. Holyst, Phys. Rev. E **75**, 030101(R) (2007)
32. `http://www.visone.de/`