

Where Is Evolutionary Computation Going? A Temporal Analysis of the EC Community

Carlos Cotta¹ and Juan-Julián Merelo²

¹ Dept. Lenguajes y Ciencias de la Computación, Universidad de Málaga,
ETSI Informática, Campus de Teatinos, 29071 – Málaga – Spain
ccottap@lcc.uma.es

² Dept. Arquitectura y Tecnología de Computadores, University of Granada,
ETSI Informática y Telecomunicaciones, C/ Daniel Saucedo Aranda, s/n, 18071 – Granada – Spain
jj@merelo.net

Abstract. Studying an evolving complex system and drawing some conclusions from it is an integral part of nature-inspired computing; being a part of that complex system, some insight can also be gained from our knowledge of it. In this paper we study the evolution of the evolutionary computation co-authorship network using social network analysis tools, with the aim of extracting some conclusions on its mechanisms. In order to do this, we first examine the evolution of macroscopic properties of the EC co-authorship graph, and then we look at its community structure and its corresponding change along time. The EC network is shown to be in a strongly expansive phase, exhibiting distinctive growth patterns, both at the macroscopic and the mesoscopic level. *Keywords:* Complex networks, evolutionary computation, social networks analysis

1 Introduction

Science as done by humans is a complex system and, as such, is composed of individual agents that take informed decisions. These decisions, analyzed as a whole, create patterns that, unsurprisingly, have a lot in common with patterns created in other complex systems, such as predator-prey interactions or metabolic networks [15]. The evolution of complex systems is usually driven by optimization, but, in most cases, it is difficult to find out what exactly is being optimized. Besides, optimization is an emergent property, a consequence of individual behaviors, but has no mirror in the individual decisions. In the case at hand, it can hardly be said that every author sits down before writing a paper to choose topic, and then co-authors: co-authorship is obviously a conscious decision but, if there is some kind of optimization, it is quite personal and obviously overlooks the network as a whole (even if it was possible to look at it, which has not been until now). However, we would like to find out what kind of rules at the individual level create the observed patterns; so, as a first step for a model of creation and evolution of scientific authorship networks, it is necessary to understand how it evolves in time.

Studying co-authorship networks is also interesting *per se*, for several reasons: it allows us to identify *sociometric stars* within the field, which would correspond, if not to the most important scientists in the field, at least those that are in the middle of several groups with diverse interests and topics, and, thus, are able to deal with a wide range of topics within the field; second, since many funding efforts are aimed at creating networks of connected scientists (for instance, the so-called *Networks of Excellence* funded by the European Union's 5th and 6th framework program), social network analysis allows us to measure their real impact in the scientific community. Finally, an exhaustive mapping of the evolutionary computation social network could have some practical impact on conference or journal referee assignment by using it to select or vet reviewers for papers based on their degrees-of-separation.

In line with these considerations, previous work by the authors [4] analyzed the EC community from a static perspective, elucidating its global structure, and identifying the key actors in the network; very recently, an article has done a similar analysis restricted to the GP bibliography database [20]. We now focus on its dynamic properties, basing our study in two different aspects: first, the study of the macroscopic evolution of the network and, second, the evolution of communities within the network, with special emphasis on the macroscopic structure of these communities and its implication on how knowledge might be managed within them. To do so, we will firstly present in the next section an overview of related work, and a description of the materials used in the analysis.

2 Background and Materials

The study of all kind of networks has undergone an accelerated expansion in the last few years, following the introduction of models for power-law [1] and scale-free networks [21], which, in turn, has induced the study of many different phenomena under this new light. One of them have been co-authorship networks [13]: nodes in these networks are paper authors, joined by edges if they have written at least a paper in common. In this section we firstly describe the state of the art in the analysis of these networks. Subsequently, we will detail the procedure used to construct the EC co-authorship network.

2.1 State of the Art

The evolution of co-authorship networks has been studied mainly in the context of checking the validity of the preferential attachment hypothesis [10, 2], finding that creation of new links (that is, writing a paper with novel author combination) roughly follows preferential attachment, which means that co-authorship networks tend to be assortative; in fact, that was anticipated in the pioneering work by H. Kretschmer [11, 12], who discovered that co-authorship tends to happen more frequently between authors with similar number of publications if they are in different institutions (in what she called *invisible colleges*). A more thorough examination of possible models of network evolution and its consequence on the observed network structure is shown in Dorogovtsev and Mendes' book, [5]. In theory, it should be possible to look at the current state of the network and find out what has caused it; in practice, sampling problems and inadequate or incomplete information avoid this. Besides, co-authorship networks, as any social network, do not always follow the same laws along time. From a slightly different point of view, studies by Solé et al. [18] look at the inception of scale-free networks.

Temporal studies of co-authorship patterns [7] study the correlation co-authorship vs. productivity and citation impact, which is strikingly negative in the first case (at least after a co-authorship peak, whose value is characteristic of the field, is reached) and positive in the second. The general conclusion is that most fields aim at higher cooperativity values (increasing number of co-authorships, that is, increase of the number of links in the co-authorship graph), with those who have more links getting increasingly more with years. However, the underlying mechanism is unknown, although it points at the Matthew effect (those that have, will get more).

Studies on evolution of macroscopic network values have been done, for instance, on Mathematics by Grossman [8], which has uncovered a trend to increase the number of collaborations and productivity, as measured by the average number of papers per author.

However, the paper closest to this one has been a study on the Computer Support Cooperative Work community [9], which studies this community from its inception (as a spin-off of the Computer-Human Interface –CHI– field); however, being concentrated in this narrow field, its main emphasis was on volatility, but a new theme, the centrality of the group within the whole field (CHI) was also investigated. One of its findings was that the number of new members varied roughly with the size of the graph giant component. On the other hand, Yoshikane and Kageura [22] have looked at the evolution of co-authorship networks in different fields by looking at the number of co-authorships and the change in *loyalty*, or, as the authors say, *inequality*, that is, how different is the frequency of cooperation between both partners. They arrive at the conclusion that both quantities are related to the topic under study in each field, and research styles.

In general, a study in depth of a particular field such as EC, at the macroscopic (network properties) as well as mesoscopic (community formation and structure) which aims at uncovering the mechanism that a scientific career follows, has not been done so far. This paper is a first step in that direction.

2.2 Data Used in the Analysis

The bibliographical data used for the construction of the scientific-collaboration network in EC has been gathered from the DBLP³ –*Digital Bibliography & Library Project*– Computer Science bibliography server, maintained by Michael Ley at the University of Trier. This database provides bibliographic information on major computer science journals and proceedings, comprising more than 730,000 articles and several thousand computer scientists (as of March 2006). Besides this wide coverage of computer science, the database has the advantage of having a highly structured and regular web interface. Hence it is very amenable for automated parsing by a scraping program. In particular, hyperlinks are provided for every co-author of a paper, making navigation through the database very easy.

The process to obtain the raw data is the following: our scraping robot is firstly fed with a collection of DBLP author keys. These correspond to all authors that have published at least one paper since year 2000 in any of the following large EC conferences: GECCO, PPSN, EuroGP, EvoCOP, and EvoWorkshops. This information constitutes the seed for the search, which will be expanded to include EC authors publishing in other conferences or journals. To be precise, for each of these authors in the initial seed, the robot analyzes his/her publications, and identifies EC papers looking for certain patterns in the publication reference (e.g., acronyms of EC-specific conferences such as GECCO, PPSN, EuroGP, etc., or keywords such as “Evolutionary Computation” and “Genetic Programming”), or in the title (e.g., keywords such as “evolutionary algorithm” or “genetic algorithm”, or acronyms such as “EA”, “GA” or “GP”). If a paper triggers any of these criteria, it is classified as an EC paper and the co-authors are recursively explored⁴ (already processed authors are stored in a sorted tree, to avoid cycles). This system has turned out to be rather accurate in detecting EC papers. Actually, the visual inspection of the resulting network indicated that only a small fraction of false positives (well below 1% of the total number of papers) passed the filters. These were mostly computational biology papers, and were readily removed from the network.

³ <http://www.informatik.uni-trier.de/~ley/db/>

⁴ To give an idea of the breadth of the search, the initial seed contains 2,967 authors, and the final network comprises 6,555 authors.

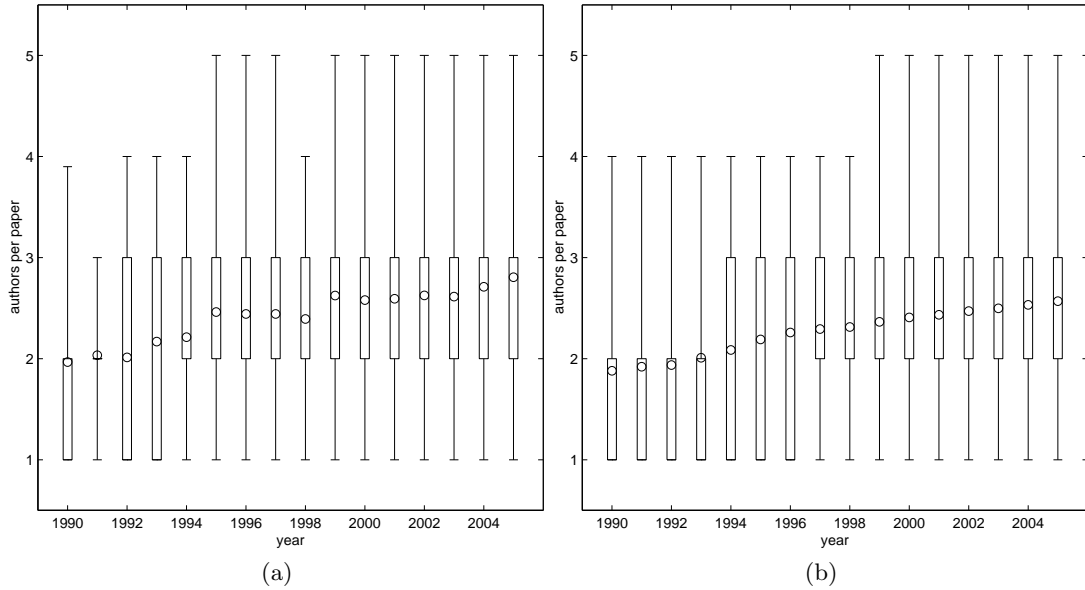


Fig. 1: Distribution of authors per paper. (a) Data by year. (b) Cumulative data. In this and all subsequent boxplots, the boxes comprise the second and third quartile, the inner circles indicate the mean, and the whiskers mark the 5% and 95% percentiles.

We took a snapshot of the EC-sphere in early February 2006, following the methodology described above. Although some EC papers in the database date back from as early as 1973, the coverage is sparse for these initial years. For this reason, we have centered our analysis in the 1990 – 2005 time period. Our time-window thus spans the latest 16 years in the field (we are referring here to the time window for analyzing the data on a year by year basis; we have also considered cumulative data, thus including all data before 1990), a interval long enough for many interesting phenomena to show up, as illustrated in the following. In any case, we do not pretend to have an exhaustive graph of the EC author community, even in the years we are covering. However, we expect that the behavior of the graph at large will be very similar to what is shown here, since, essentially, the authors will be the same. This remains to be proved, however, and we just claim in this paper that, by choosing an easily accessible bibliography-database, we have an essentially non-biased sample of the community.

3 Evolution of Macroscopic Properties of the Network

Complex networks seldom follow a random path: they have some internal logic and constraints that make its macroscopic properties evolve in a certain direction. In the case at hand, we have observed the macroscopic properties of the EC co-authorship graph (let us call it C) from 1990 up to 2005, 16 years in all, and also of the graph formed by authors that have published in a single year (which we will call C_i). We also consider the cumulative data up to a certain year, i.e., $\tilde{C}_y = \cup_{i < y} C_i$.

Looking at the data, we observe that although there is no drastic change in co-authorship patterns, it is clear that several interesting trends are present in the data. First of all, there exists a marked trend towards an increasing number of authors per paper (Fig. 1a).

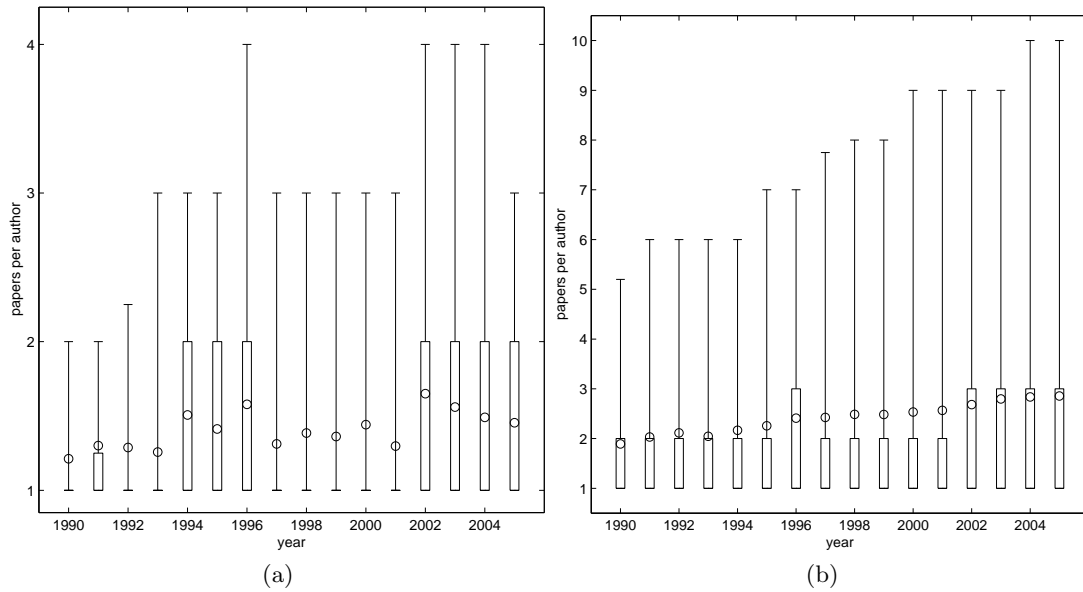


Fig. 2: Distribution of papers per author. (a) Data by year. (b) Cumulative data. The fact that the average is displaced (with respect to the bar center) is explained because the distribution follows a power law, with a few authors having many papers, and most authors having just a few papers [4].

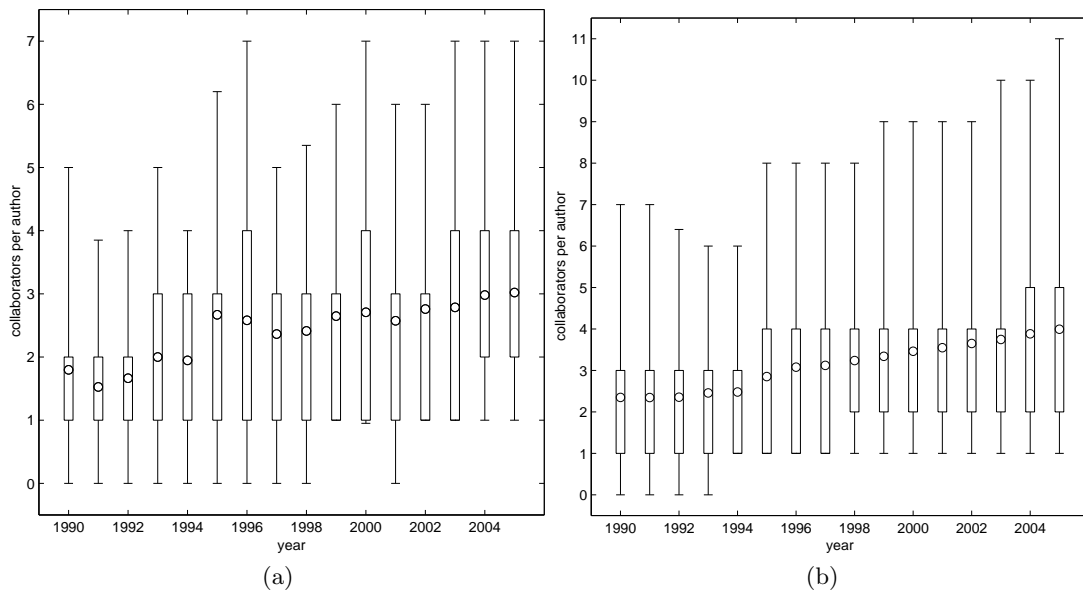


Fig. 3: Distribution of collaborators per author. (a) Data by year. (b) Cumulative data.

This pattern can be interpreted in terms of the maturity that our field is gaining, and the subsequent increased complexity of research papers. Many of the “easy” question were solved in the initial years of the paradigm, and current works are more specialized, thus requiring a higher effort and/or interdisciplinarity. Indeed, by assuming that the effort expendable by

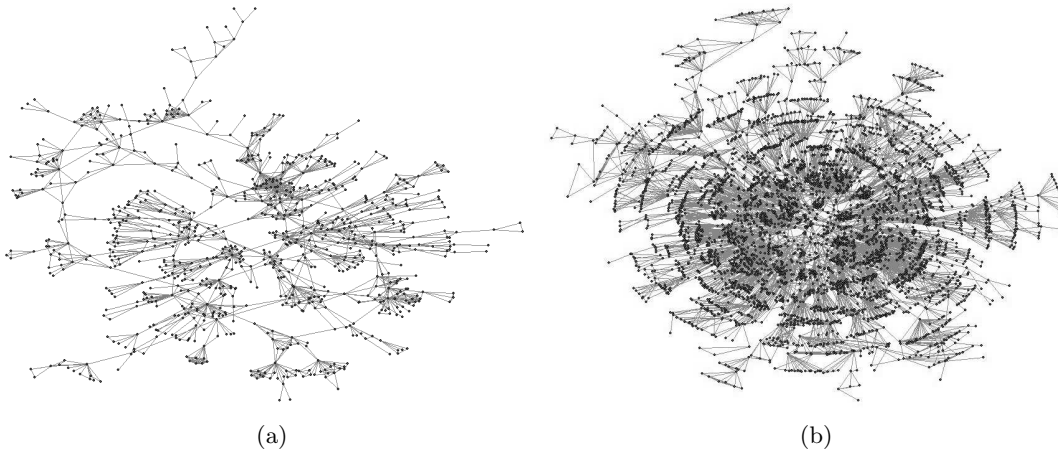


Fig. 4: The giant component of the EC network in (a) 2000, and (b) 2005.

a researcher per year is constant (something that is backed up by the fact that the average number of papers per author follows a rather flat trajectory, as shown in Fig. 2a), we can easily derive that the average paper requires much more effort now than 16 years ago. This is also reflected in the increasing trend of annual collaborators per author (Fig. 3a). Not only a single paper requires more effort, but also a single author is likely to collaborate with a higher number of people throughout the year. Again, this indicates the specialization and diversification of interests within the community.

An overview of the main macroscopic features of the network is shown in Table 1. The number of authors in C_y generally increases, which means that EC is still a vibrant field attracting new researchers, and generating new papers; this impression is augmented by the fact that half the authors that publish an EC paper each year are new, i.e., they had not published in EC before. This is a remarkable figure that indicates the health of the area. Furthermore, if we take a look at the volatility, that is, the number of authors that published in a certain year, but have not published any other EC paper since then, we can clearly see that there is a net amount of researchers that come to stay in EC⁵.

The existence of a giant component, i.e., a single connected component much larger than the remaining components (whose sizes typically follow a power-law), is another relevant feature of this kind of networks (see Fig. 4). This giant component started to rise in the mid 1990s (compare the sizes of the largest and 2nd largest component); at a certain point in time, around 2002, there is an abrupt increase in the size of this component (it grows from $\sim 30\%$ to $\sim 50\%$ of the existing authors) due to percolation. This merging of previously unconnected subgraphs, among other factors, has caused the network to follow the path of increasing decentralization, as can be seen in the *network centralization* row. This quantity reflects the *starness* of the network: the higher the centralization, the more it looks like a star (where every node is only connected to a central hub); since centralization decreases with time, it can be inferred that new groups coalesce around many researchers in the field, instead of being created around a small group of researchers. We will return to this point later.

⁵ Volatility figures for recent years have to be interpreted as upper bounds, since the time window for considering that an author has left the field gets narrower as we approach current year.

Table 1: Evolution of the macroscopic properties of the EC co-authorship network. (Top) Cumulative data. (Bottom) Data by year.

CUMULATIVE DATA

	2005	2004	2003	2002	2001	2000	1999	1998
number of authors	6,555	5,611	4,708	3,984	3,284	2,837	2,262	1,851
% new authors	14%	16%	15%	18%	14%	20%	18%	19%
% volatility	—	21%	17%	14%	12%	16%	11%	13%
largest component	4119	3276	2503	1917	1098	863	692	583
2 nd largest component	84	36	34	33	99	77	38	26
network centralization	0.98%	1.02%	1.09%	1.09%	1.05%	1.22%	1.40%	1.72%
clustering coefficient	0.809	0.810	0.807	0.814	0.820	0.821	0.819	0.822
average distance	8.31	7.93	8.70	8.39	6.96	6.62	7.35	8.24
diameter	22	20	25	21	22	19	21	23
	1997	1996	1995	1994	1993	1992	1991	1990
number of authors	1,497	1,204	916	663	423	282	214	166
% new authors	20%	24%	28%	36%	33%	24%	22%	41%
% volatility	11%	15%	17%	15%	11%	9%	12%	17%
largest component	384	265	194	151	62	36	31	24
2 nd largest component	69	41	23	15	54	34	28	19
network centralization	1.67%	1.83%	2.10%	2.65%	3.46%	4.53%	4.58%	5.31%
clustering coefficient	0.821	0.816	0.821	0.818	0.824	0.817	0.824	0.824
average distance	7.31	6.63	6.23	6.11	3.78	2.85	2.80	2.32
diameter	17	17	15	15	9	7	7	6

DATA BY YEAR

	2005	2004	2003	2002	2001	2000	1999	1998
number of authors	1,942	1,852	1,597	1,371	959	1,095	753	705
% new authors	49%	49%	45%	51%	47%	53%	55%	50%
% volatility	—	63%	49%	40%	39%	43%	33%	34%
largest component	61	45	37	55	35	37	21	15
2 nd largest component	39	44	36	36	10	34	16	12
network centralization	0.83%	0.82%	0.83%	1.05%	0.78%	1.13%	1.54%	1.66%
clustering coefficient	0.886	0.885	0.875	0.876	0.914	0.895	0.908	0.902
average distance	2.33	2.44	2.24	2.44	1.76	2.01	1.51	1.43
diameter	8	7	8	9	6	7	5	4
	1997	1996	1995	1994	1993	1992	1991	1990
number of authors	555	534	448	380	214	125	95	99
% new authors	53%	54%	56%	63%	66%	54%	51%	69%
% volatility	30%	33%	35%	27%	22%	20%	27%	28%
largest component	15	20	15	9	15	6	8	8
2 nd largest component	11	18	11	8	8	6	4	7
network centralization	1.03%	2.17%	2.11%	1.34%	2.84%	2.73%	6.08%	4.38%
clustering coefficient	0.911	0.890	0.910	0.862	0.903	0.890	0.913	0.922
average distance	1.43	1.60	1.38	1.33	1.49	1.25	1.19	1.28
diameter	6	4	4	3	4	2	2	3

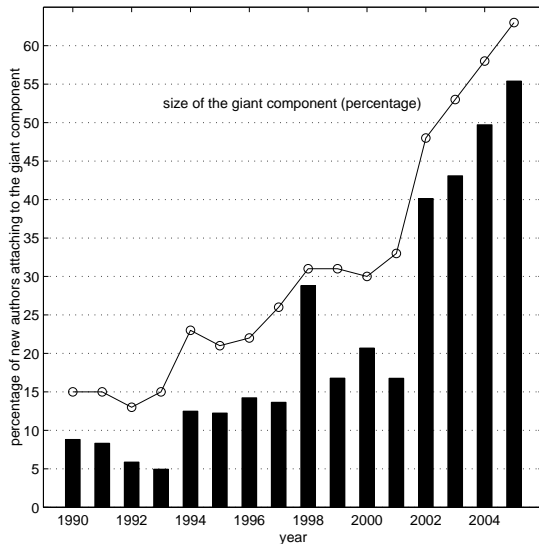


Fig. 5: Percentage of new authors per year attaching to the giant component (bars) and relative size of the giant component (line).

Since there is a continuous inflow of new researchers, and centralization decreases with time, it would follow that the network diameter should shoot up. The fact that it does not, with an average distance around 8, and a network diameter around 20, shows that the network has the *small world* property [21], with the creation of new nodes offset by the fact that they link with nodes that have a central position within the network. On average, only around 8 handshakes are needed to, e.g., move from a EC researcher in a Korean university to another in Argentina. The clustering coefficient is also quite high, as compared with a random network, and stays high through time, which also supports its small-world feature (already shown in [4]).

Another important conclusion we can extract from the evolution of the network is that, in general, preferential attachment does not happen (at least in the links created by addition of new nodes). As shown in figure 5, the number of new authors that attach to the giant component of the network grows roughly in the same way as the size of the latter. While it could still happen that they attach preferentially to those nodes that already have many coauthors, it must be noted that most such high-degree authors are in the giant component. Hence, it would be expected that new authors attached mostly to this giant component. The fact they do not reveals an interesting feature of the dynamics of \tilde{C} , that is, the cumulative network. Further analysis of this dynamics requires deepening into the internal structure of the network, as done in next section.

4 Space-Time Structure of the Network

The EC network, as an evolving system, can be analyzed from a spatio-temporal perspective. Taking the network state at a certain point as the spatial dimension, we can investigate its internal structure, and project this analysis into time. To do so, we interpret the whole network as an emergent system, formed by the aggregation of simpler structures. This is certainly the case since the EC network is not an undifferentiated mass of vertices, but it is distinctively inhomogeneous: some groups of authors are densely connected among them,

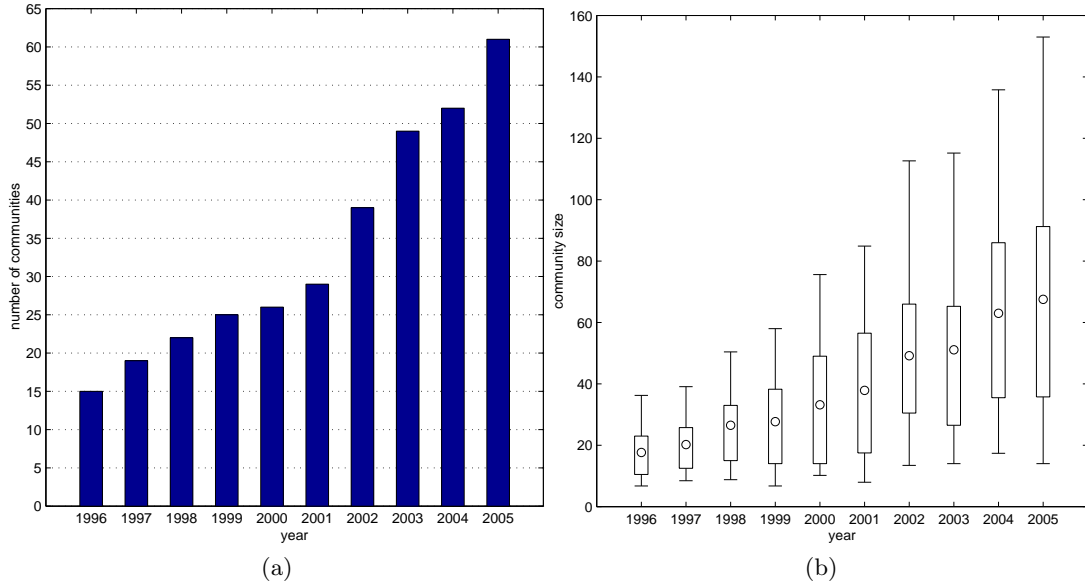


Fig. 6: (a) Evolution of the number of communities in the giant component of the network. (b) Evolution of community sizes in the giant component of the network.

but sparsely linked with authors outside the group. These groups are termed *communities* [6].

There are several approaches for identifying community structure in networks [16]. In this work, we have opted by a method based on geodesic information flow (that is, information flow following the shortest path), devised by Newman and Girvan [6], and that has been shown to perform well in several domains. This method is based on the iterative removal of edges. To be precise, edges are removed according to their *betweenness*, i.e., the number of geodesics (a geodesic is the shortest path between two nodes) that pass through it. Intuitively, if a network contains communities or groups that are only loosely connected by a few intergroup edges, then all shortest paths between different communities must go along one of these few edges. Hence, these high-betweenness edges can be interpreted as bridges among different subgroups in the network, their removal resulting in the isolation of such groups. By iterating this removal process until no edge remains, we end up with a dendrogram, each leaf corresponding to an author, and each internal node to a split event (the point at which a component divides into two smaller subcomponents). Newman and Girvan define a metric to measure to which extent a split results in more cohesive picture of the network than the original non-split group. We refer to [6, 17] for details.

We have applied this analysis to the giant component of the network in the last ten years (1996-2005). As mentioned before, the giant component is the result of the percolation of the community, and hence can provide interesting information about how communities evolve with time. First of all, Fig. 6 shows the evolution of the number of communities and their sizes. Not only the number of communities grows with time (from 15 communities in 1996 up to 61 communities in 2005 – Fig. 6a), but also their size (from an average of 17.7 authors in 1996 up to 67.5 authors in 2005 – Fig. 6b). This is related with the likely non-preferential attachment of new nodes mentioned in previous section. Communities do not simply get larger by aggregation but also arise as spin-offs from existing communities.

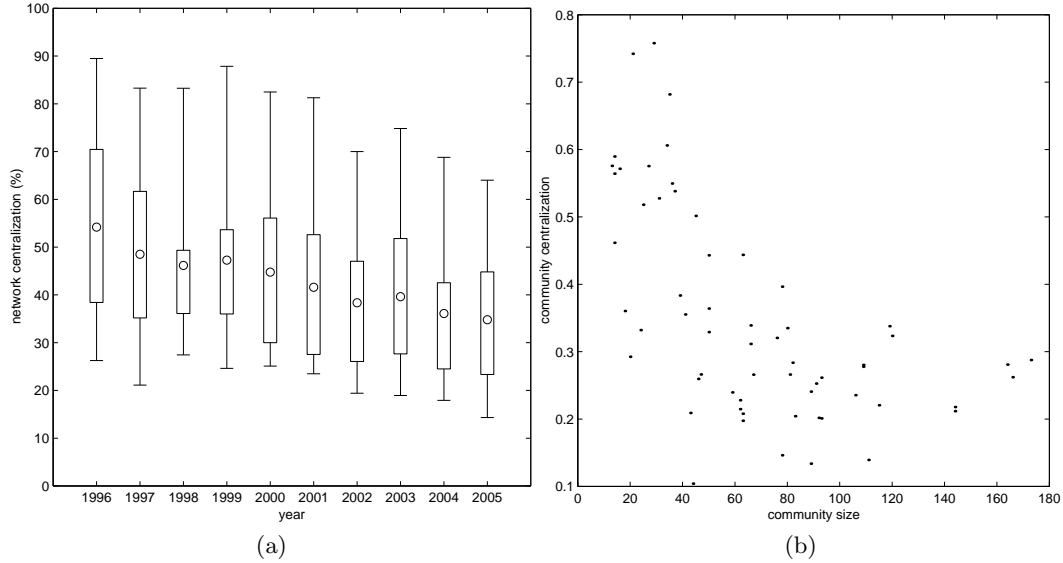


Fig. 7: (a) Evolution of network centralization in communities. (b) Correlation between community size and community centralization in 2005.

The spin-off effect is also backed up by the trend in decreasing centralization of communities with time. Two factors are responsible of this. Firstly, the size of a community plays an important role; as a community grows, it is more difficult to keep it centralized. It has been argued by Sparrowe *et al.* [19] that this decentralization is actually detrimental for the performance of the community. However, while this might be true in small-size, short-term groups in some contexts, we believe it is not the case in scientific research, where the goals are much more diverse, and collaborations are often mid- or long-term. Indeed, the trend of increasing decentralization is the natural result of an (mostly unconscious) resource allocation process in the EC community as a whole: centralization comes at the cost of increased overload in the central hub, something which is untenable in the long run for most researchers. In this sense, there is an interesting anomaly for larger community sizes: centralization starts to rise (or at the very least, ceases to decrease) at a certain point as it can be seen in Fig. 7b (a fit to a 2nd degree polynomial indicates a centralization minimum at 119 authors). Cohesiveness of such large communities can only be maintained if a strong research personality is present, and the community is structured more or less around him/her. As an example, Fig. 8 shows the largest community in 2005. The presence of D.E. Goldberg provides a center of mass that prevents disintegration of this community as currently structured.

The second effect that contributes to the pattern of decreasing centrality has to do with the internal dynamics of communities, as driven by the scientific progression of its members. Clearly, as maturity is gained by some scientist, he/she starts to (i) seek collaborations with other researchers other than his/her initial tutor, and (ii) take students under his/her direction, thus contributing both to the growth and decentralization of the network. Upon achieving a critical mass, these communities typically split into separate groups. In this sense, it may be interesting to take a look at the community interaction graph, where all authors within a community collapse into a single node, and edges indicate collaboration

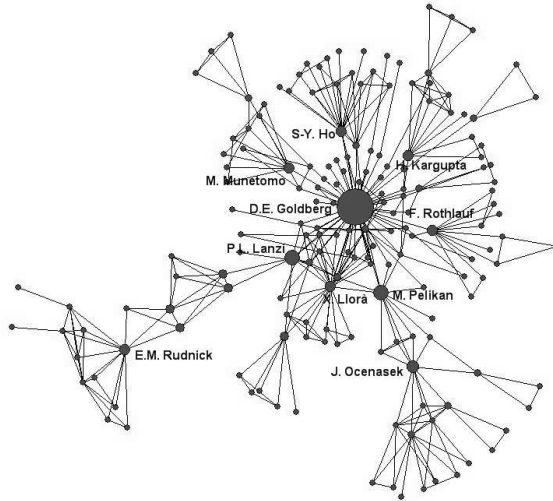


Fig. 8: The largest community in the giant component of the EC network. Node sizes are proportional to *internal* betweenness.

between authors from two separate communities. Fig. 9 shows these interaction graphs for three different time points.

It must be noted that inter-community edges can correspond to two different causes. On one hand, they can indicate community splits, i.e., they are the fossil record of a previously connected component that separates into two groups (notice that if further division took place in the latter groups, these edges would refine and indicate the historical links between the corresponding sub-subcommunities). On the other hand, they can indicate an actual collaboration between members of two communities that were never tied together. These collaboration events are becoming more common in recent years, due to the increasing number of authors required to publish a paper, and to the specialization of the field (and the subsequent growth in interdisciplinarity). For example, it is not unusual to have a paper co-authored by people specialized in algorithmic aspects and people working in applications.

The evolution of the community interaction graph is thus interesting as it provides a view on the changes in the large-scale structure of the network. An interesting feature of this evolution is the progressive formation of a core/periphery structure [3]. In 1996 (Fig. 9a), there are just a few communities, and they do not interact much, with a small proto-core formed by only three communities. In the next snapshot (year 2000, Fig. 9b), there is a small yet cohesive core formed by ten communities (denoted by black triangles), surrounded by an outer layer of the same size; a small periphery starts to show up. Finally, in 2005 (Fig. 9c), the core is very well defined, comprising 26 communities; a marked periphery is also present, encompassing 19 communities; finally, there are two intermediate layers of 10 and 6 communities. Overall, this seems to indicate an aggregation model where new communities rise in the periphery, and become integrated in the core with time. Community splits typically result in a reduction of the coreness of the resulting groups. Both phenomena contribute to the existence of thin intermediate layers. In the large picture, we take as a positive fact that the bulk of the network tends to create a core structure: it shows that specialization is not happening by isolation of “watertight compartments”, but by the creation of cross-disciplinary links [14]. Let us note *en passant* that, with respect to

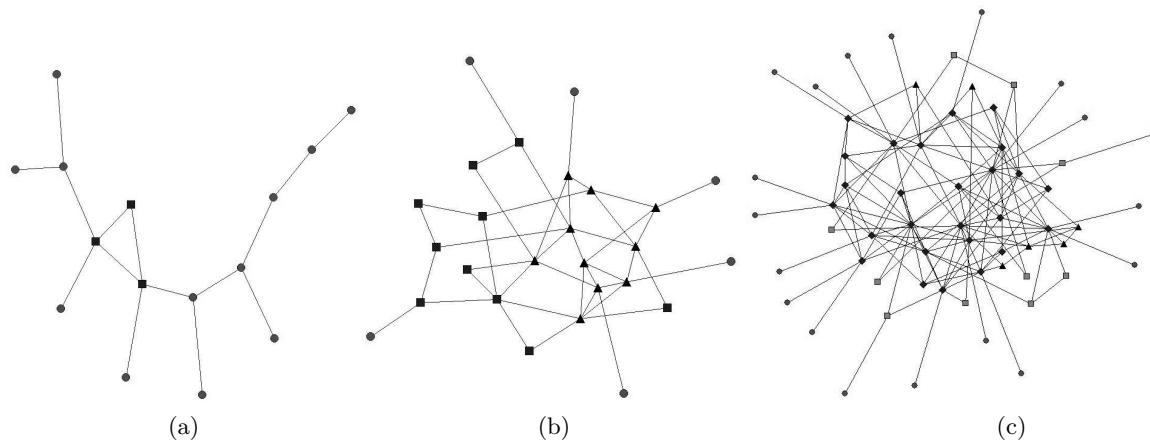


Fig. 9: Community interaction graph in the giant component. Node shapes indicate the *coreness* of each community (circle, square, triangle, diamond, from lower to higher). (a) Year 1996. (b) Year 2000. (c) Year 2005.

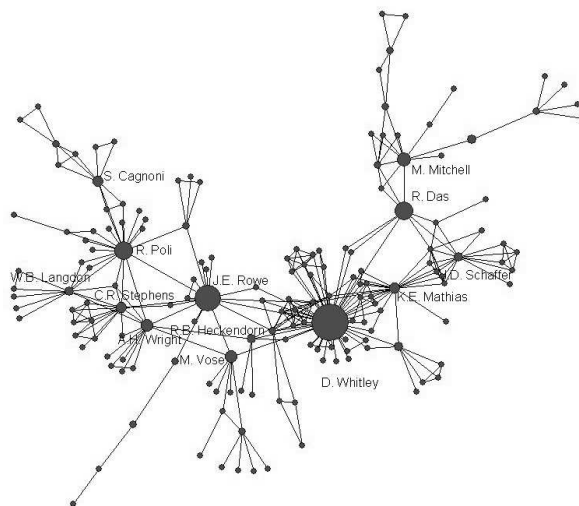


Fig. 10: The most central community in the giant component of the EC network. Node sizes are proportional to *internal* betweenness.

betweenness, the most central community in the current core (Fig. 10) is mostly theory-oriented. It turns out that theoretical developments still occupy a prominent place in our field.

5 Conclusions and future work

This work has given the first steps towards the understanding of the dynamics of the EC co-authorship system and its modelling using an individual-based model. The evolution of macroscopic properties shows that, even taking into account churn, there is an evolution towards higher connectivity and lower centralization, which implies that preferential attachment is probably not as strong as in other communities; still, some kind of Matthew effect

takes place, since the diameter of the network is kept small, both through long-distance links and attachment of new nodes to those that already have a central position. An analysis of the dynamics of communities within the main component of the EC co-authorship graph reveals a similar pattern: initially sparse communities finally coalesce around a core, which has Prof. Goldberg's community as its largest component, and a community composed mainly with scientists dealing with EC theory as the most central one.

Future work will include applying evolutionary techniques to this social network analysis, which, besides providing novel long-distance connections on the graph, will allow to obtain better results for some network properties. We will also try to check these results by including data coming from other sources such as IEEE proceedings and conferences; and, finally, we will perform an study on community genealogy to find out the micro-mechanisms of community formation and splitting.

Acknowledgments

This work has been supported in part by MCyT projects TIC2003-09481-C04-04, and TIN2005-08818-C04-01. Thanks are due to Prof. Mark Newman for providing us his code for the calculation of edge betweenness.

References

1. A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, October 1999.
2. A.L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311(3):590–614, 2002.
3. S.P. Borgatti and M.G. Everett. Models of core/periphery structure. *Social Networks*, 21:375–395, 1999.
4. C. Cotta and J.J. Merelo. The complex network of ec authors. *SIGEvolution*, 1(2):2–9, 2006.
5. S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, 2003.
6. M. Girvan and M.E.J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences USA*, 99:7821–7826, 2002.
7. W. Glanzel. Coauthorship patterns and trends in the sciences: a bibliometric study with implications for database indexing and search strategies 1980–1998. *Library Trends*, 50(3), 2002.
8. J.W. Grossman. Patterns of collaboration in mathematical research. *SIAM News*, 35(9):8–9, November 2002.
9. D.B. Horn, T.A. Finholt, J.P. Birnholtz, D. Motwani, and S. Jayaraman. Six degrees of Jonathan Grudin: a social network analysis of the evolution and impact of CSCW research. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 582–591, New York, NY, USA, 2004. ACM Press.
10. H. Jeong, Z. Neda, and A.L. Barabási. Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4):567–572, 2003.
11. H. Kretschmer. Coauthorship networks of invisible collees and institutional communities. *Scientometrics*, 31(1):363–369, 1994.
12. H. Kretschmer. Patterns of behaviour in coauthorship networks of invisible colleges. *Scientometrics*, 40(3):579–591, 1997.
13. T. Luukkonen, O. Persson, and G. Sivertsen. Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, 17(1):101–126, 1992.
14. J.J. Merelo and C. Cotta. Building bridges: The role of subfields in metaheuristics. *SIGEvolution*, 1(4):9–15, 2007.
15. M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.
16. M.E.J. Newman. Detecting community structure in networks. *European Physics Journal B*, 38:321–330, 2004.
17. M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physics Review E*, 69:026113, 2004.

18. R. Solé and S. Valverde. Information theory of complex networks. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, editors, *Complex Networks*, Lecture Notes in Physics, pages 169–190. Springer-Verlag, 2004.
19. R.T. Sparrowe, R.C. Liden, S.J. Wayne, and M. L. Kraimer. Social networks and the performance of individuals and groups. *Academy of Management Journal*, 44(2):316–325, 2001.
20. M. Tomassini, L. Luthi, M. Giacobini, and W.B. Langdon. The structure of the genetic programming collaboration network. *Genetic Programming and Evolvable Machines*, 8:97–103, 2007.
21. D.J. Watts and S.H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.
22. F. Yoshikane and K. Kageura. Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, 60(3):433–444, 2004.