An Analysis of the Structure and Evolution of the Scientific Collaboration Network of Computer Intelligence in Games

R. Lara-Cabrera, C. Cotta, A.J. Fernández-Leiva

Department of "Lenguajes y Ciencias de la Computación"
ETSI Informática, University of Málaga, Campus de Teatinos
29071 Málaga – Spain

Abstract

Games constitute a research domain that is attracting the interest of scientists from numerous disciplines. This is particularly true from the perspective of computational intelligence. In order to examine the growing importance of this area in the gaming domain, we present an analysis of the scientific collaboration network of researchers working on computational intelligence in games (CIG). This network has been constructed from bibliographical data obtained from the *Digital Bibliography & Library Project* (DBLP). We have analyzed from a temporal perspective several properties of the CIG network at the macroscopic, mesoscopic and microscopic levels, studying the large-scale structure, the growth mechanics, and collaboration patterns among other features. Overall, computational intelligence in games exhibits similarities with other collaboration networks such as for example a log-normal degree distribution and sub-linear preferential attachment for new authors. It also has distinctive features, e.g. the number of papers co-authored is exponentially distributed, the internal preferential attachment (new collaborations among existing authors) is linear, and fidelity rates (measured as the relative preference for publishing with previous collaborators) grow super-linearly. The macroscopic and mesoscopic evolution of the network indicates the field is very active and vibrant, but it is still at an early developmental stage. We have also analyzed communities and central nodes and how these are reflected in research topics, thus identifying active research subareas.

Keywords: Complex network, scientific collaboration, network evolution, computational intelligence, games

1. Introduction

Games have long been seen as an ideal test-bed for the study of artificial intelligence (AI) [1]. In the past, much of the academic work on AI and games focused on traditional board games, such as checkers and chess, and these games were in turn used

Email addresses: raul@lcc.uma.es (R. Lara-Cabrera), ccottap@lcc.uma.es (C. Cotta), afdez@lcc.uma.es (A.J. Fernández-Leiva)

to check the goodness/efficacy of AI techniques. Although this traditional focus is still being used, there is now also a growing body of work on applying computational intelligence (CI) to game development (including video games) with the aim of improving the game itself. CI comprises a collection of nature-inspired methods –such as evolutionary algorithms, artificial neural networks and fuzzy logic– for the resolution of complex problems, and can be applied to optimize game development from different perspectives: for instance, from the user's point of view, current players demand, not only outstanding graphics, but also other non-visual features that CI can help to significantly improve such as intelligent/adaptive behavior of non-player characters (NPC), interesting narrative, or more attractive contents (in the form of, for instance, levels, maps, weapons, armor, terrains, music or even the game rules themselves) and many more. In addition, from the industry's point of view, CI is starting to be viewed as a mechanism to improve the process of game development as well as a tool to extend the commercial life of games. Regarding the first issue, CI is being used to automatically generate game content, such as terrains, maps or even music [2], whereby the automation of this process would lead to a reduction in the production costs as the content would not be entirely hand-created and thus human intervention might be reduced. Moreover, CI techniques applied to achieve this objective in one particular game might presumably be adapted to a similar game belonging to the same genre and this would also reduce the expense of implementing new games. With regard the latter issue, the capacity of CI to produce automated NPC behaviors as well as multimedia content is interesting (and promising) as, in theory, it would be natural to extend a successful (i.e., a best seller) game by procedurally creating new content. This new content in turn would produce a new (commercial) version of the game (the recursive application of this idea sculptures the concept of infinite game) which represents a way of extending the earnings at a minimum cost. Moreover, the industry perceives CI as a tool from which it is possible to gain additional benefits as it can be used to extract interesting players' data that can be employed, for instance, to evaluate player satisfaction (e.g., via emotional analysis) or to construct specific games according to the profile of the users.

However, the application of CI techniques to game development goes beyond the idea of just generating pure entertainment and also covers the purpose of solving hard (and, in many cases, socially accepted) problems; this is for instance, the case of the so-called "serious games". These games prioritize additional objectives then that of simply fun such as training, learning, evaluation, management, marketing, or advertisement among others. Serious games have already been implemented in a wide number of heterogeneous areas such as education, health care, politics, or defense [3] and CI has already been applied successfully to the construction of serious games – see [4].

In general, one can be certain that CI will have a great impact on game development and for this reason the application of CI to games is currently attracting increasing interest from both the industry and academia. The consequence of this interest is that this field is a burning issue that seems to have undergone a huge transformation in recent years. To study this transformation, we have embarked on an analysis of the social network of researchers working on CI in games (CIG). Social network analysis – see [5] for a survey – although an old discipline, has recently been given new impetus and tools from the field of complex networks. A review of complex networks is presented in [6]. The study of all kinds of networks has been an extremely active research topic in the recent past, following the introduction of models for power-law [7] and scale-free networks [8], which, in turn, has induced the study of many different phenomena in this new light. The particular class of network the study of which is addressed in this work is a co-authorship network [9]: nodes in these networks are paper authors, joined by edges if they have written at least one paper together. Note that the co-authorship data is a social network since collaborating in a research study usually requires that the coauthors become personally acquainted. Thus, studying the ties, their structure, and their evolution enables a better understanding of the factors that shape scientific collaboration. In this sense, this network analysis has been done in many other fields, like mathematics [10, 11], evolutionary computation (EC) [12, 13, 14], computer-supported cooperative work (CSCW) [15], social sciences [16] and physics, biomedicine and computer science [17, 18], to cite some examples.

This work analyzes the collaboration network of the CIG community on three different scales –macroscopic, mesoscopic and microscopic– aiming to discern the status and distinctive features of the community, the rules governing scientific cooperation within it, and the direction in which it is heading as well as trying to identify which seems to be some of the most active research topics in the area.

2. Materials and methods

The bibliographical data used for the construction of the scientific-collaboration network has been obtained from the *Digital Bibliography & Library Project* (DBLP[1]) bibliography server. This database, maintained by the DBLP Team at the University of Trier, provides bibliographic information on major computer science journals and proceedings. It lists more than 1.9 million publications and several thousand computer scientists (as of March 2012). Besides this wide coverage of computer science, the DBLP database provides an easy to use application programming interface (API) to obtain filtered results. The results can be filtered by author, venue, date, keyword, etc. These results are encoded into XML or JSON files, which can be easily parsed by a scraping program.

We have built a program which queries the database and obtains the co-authoring network. The process of obtaining this data is as follows. The program reads a set of query terms from an external file. For each term, it queries the DBLP with this term, and processes the results returned. These results are sets of articles with the information associated with them, including the authors. The program marks as co-authors each pair of authors of every returned article. It avoids processing duplicated articles, as the DBLP API provides an internal ID number for each article. Once the data has been gathered, the program generates several files describing the co-authoring network. To obtain a suitable dataset we have chosen query terms which list all the articles published in some conferences and journals on computational intelligence in games (i.e. *CIG, TCAIG, AIIDE, GAMEON*, etc). In addition, the program also searches articles from other conferences and journals (i.e. *GECCO, PPSN, IJCAI*, etc.) filtered by some

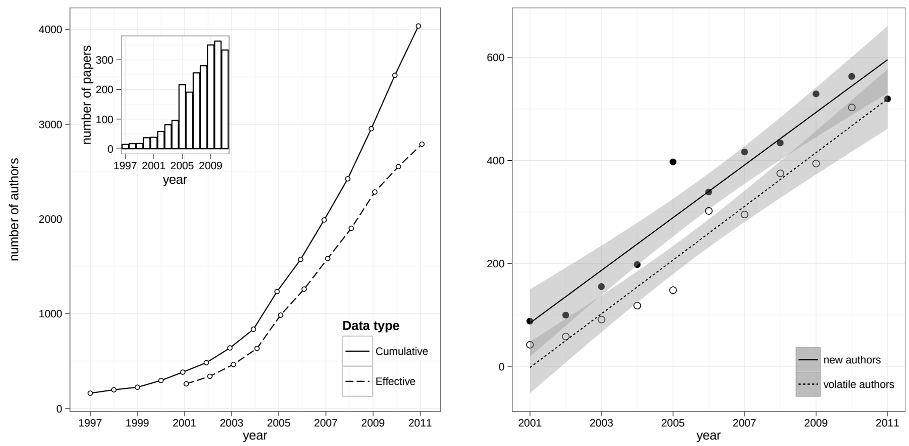
[1] `http://www.dblp.org/db/`

game-related keywords. Finally, a group of keywords are used to obtain related articles possibly missed from previous steps. There is a list with these conferences, journals and keywords in Appendix Appendix A. We have used Cytoscape [19] to draw and analyze the graphs obtained in addition to the *igraph* library, a software package for complex network research [20].

The data used in this paper were collected in late April 2012, using the program described. There are articles that date from 1971 to 2012, but the coverage is sparse for these initial years. For this reason, we have concentrated our analysis on the period between 1997-2011. Our time-window thus spans the last 15 years in the field and while we do not claim to have the full graph of the computational intelligence in games community, we expect that the behavior of this graph will be very similar to what is shown in this paper, since, essentially, the authors will be the same. We have obtained an essentially non-biased sample of the community by using an easily accessible bibliography database. Needless to say, the usual caveats regarding distinguishing authors with the same surname, first name and initials of middle names, or the existence of authors whose pen name changes apply. While mostly unavoidable, these situations are also the exception rather than the rule and therefore we do not expect them to have a significant impact on our results.

3. Analysis of the Collaboration Network of CI in Games

Using the data gathered from the DBLP database we can represent the collaboration network as a weighted graph $G(V, E, W)$ where each vertex $v \in V$ represents an author and each edge $(v, w) \in E \subseteq V \times V$ indicates that authors v and w have co-authored (possibly with other additional co-authors) at least one paper. The number of collaborations between two authors is captured by W. More precisely, W is a function with signature $W : E \to \mathbb{N}^+$ such that $W((v, w))$ is the number of papers authors v and w have co-authored. Since we are interested in the temporal evolution of this network, we define a series of weighted graphs G_t, each of them comprising the collaboration information up to year t (included). We have focused on the last 15 years as mentioned before, and thus $1997 \leqslant t \leqslant 2011$ (the data for 2012 is still incomplete and has therefore been left out of the study; we also note that publications in this field before 1997 are very scarce and – while those early links are included in the network– it does not make sense to extend the analysis much further back in time). This time span is large enough to observe interesting phenomena in the evolution of the network. It is also a long enough stretch of time for some authors to have become inactive in the field. Although the information provided by these "fossil" vertices and links is still interesting, they can also distort, in some cases, the short-term behavior of the network. This limitation has also been recognized by other authors, e.g. [14], who define the notion of *effective* network as one comprising only real (active) collaborations. Of course this network can only be approximated by the bibliographical data since other forms of social interaction are not available in our database. Following [14] we consider a moving frame of X years and build a network G_t^X comprising information for years t' such that $t - X < t' \leqslant t$. Authors and links can thus disappear if they become inactive in the area for X number of years or more. We have considered here a temporal frame of $X = 5$ years.



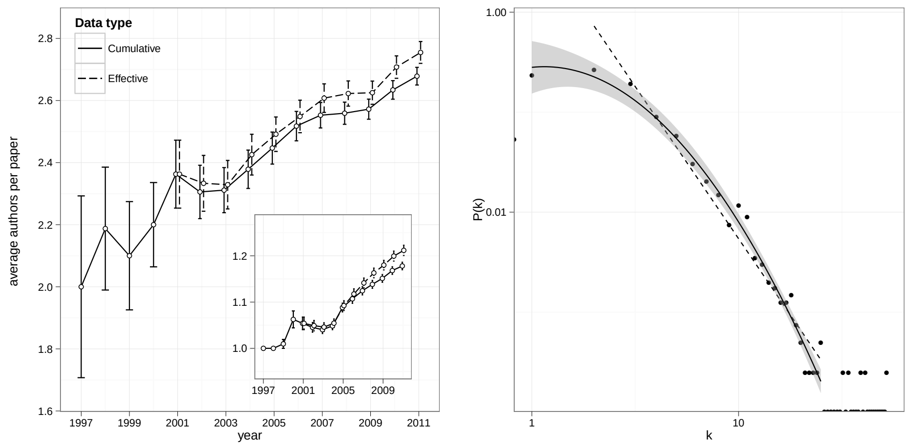
(a) (b)

Figure 1: (a) Evolution of the total number of authors. The inset shows the number of papers published by year. (b) Temporal evolution of the number of new authors and volatile authors (annual data). Straight lines are their linear fits ($\Delta_x = 51.15, R^2 = 0.92$ for new authors and $\Delta_x = 52.16, R^2 = 0.95$ for volatile authors).

We will commonly refer to G_t and G_t^5 networks as *cumulative* and *effective*, respectively. Next, we analyze the properties and structure of these networks.

3.1. Basic Macroscopic Properties

The internal logic of the scientific collaboration network, namely the customary publication behavior of authors in the field, makes its macroscopic properties evolve in a certain direction. Looking at the data, we observe that several interesting trends are present. First of all, Fig. 1a shows a glimpse of the size of the network and how it evolves with time. The curve depicting the number of authors has positive curvature in the initial years, indicating an accelerated growth in this number up to the mid-2000s, where the growth stabilizes (more clearly seen in the effective network). This indicates that CI in games is an active and vibrant field, attracting new researchers and generating new papers. As shown in the inset of Fig. 1a, there is actually a marked increase in the number of papers published per year from 2005 onwards, sustaining the steady growth of the network. Clearly, the linear trajectory of the size of the network indicates that in addition to a sustained influx of new authors there is also a constant flux of authors moving out of this field of research. If we define an author as *volatile* if (s)he does not publish after a given year t [12], we see in Fig. 1b that the rate at which volatile authors leave the network parallels that at which new authors arrive. Both magnitudes can be shown to grow linearly, the vertical offset providing a rather constant net result in favor of new authors. Actually, the linear coefficient governing the growth of volatile authors is slightly higher than that of new authors which might indicate a light deceleration of the community. However, the estimation of volatility is inherently pessimistic as we get closer in time to the present since authors have less time to publish again. The real

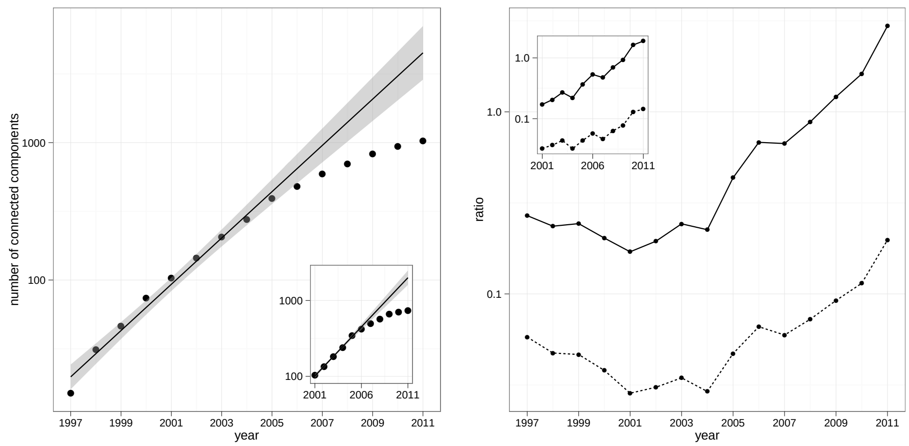


(a) (b)

Figure 2: (a) Distribution of authors per paper and papers per author (inset). In this and all subsequent mean plots, the points indicate the mean and the whiskers mark the mean standard error. (b) Degree distribution for the cumulative network G_{2011}. Note the use of a log-log scale. The dashed line is a power-law fit $P(k) \sim k^{-\gamma}$ with $\gamma = 3.27$ and $R^2 = 0.96$ ($2 \leqslant k \leqslant 25$). The solid line is the log-normal fit $P(k) \sim \exp(\alpha \ln k - \beta(\ln k)^2)$ with $\alpha = -10.5908$ and $\beta = -2.9923$ ($R^2 = 0.98$).

trend for volatility will thus be lower and more similar to that of new authors, hence supporting the estimation of constant growth of the network.

Another salient macroscopic pattern in the temporal evolution of the community is an increase in the mean number of authors per paper – see Fig. 2a. This pattern is consistent with that observed by [12] in an analysis of the network of evolutionary computation (EC) scientists but it cannot be attributed –at least not completely– to similar reasons, namely the maturity that this field is developing and the increasing complexity of research papers. This explanation made sense in the case of the EC given its longer temporal trajectory and the rather flat trend in the mean productivity of authors. However, the average number of papers per author –Fig 2a (inset)– has increased in the effective network by about 10% since 2005, roughly the same rate at which the average number of authors per paper grows for the same years, which suggests that the increasing productivity of authors is due to their involvement in multiple co-authored papers. Assuming authors try –consciously or not– to optimize their productivity at a constant available effort, this emerges as a natural strategy, if only as a means for individual authors to keep up with the pace of the community as a whole. Of course, the issue is much more complex due to the very mechanics of science and academic production (whether the combined effort of e.g. two papers written by the same two co-authors is higher/lower than that of two papers authored by each of those authors alone is open to much debate and cannot possibly be answered in a general way). At any rate –and independently of whether the trend under discussion is the result of a social strategy or a genuine requirement of increasingly complex papers (and note that these two options are not mutually exclusive)– it requires (i) the availability of a growing network of sci-



(a)

(b)

Figure 3: (a) Evolution of the number of connected components in the cumulative network. Y-axis in log-scale. The solid line shows the best fit ($y = \alpha(1+\beta)^x$, $\beta = 0.47$, $R^2 = 0.98$ with $1997 \leqslant x \leqslant 2005$). The inset shows the same fit for the effective network ($\beta = 0.35$, $R^2 = 0.99$ with $2001 \leq x \leq 2005$). (b) Solid line shows the ratio between the size of the largest component of the cumulative network and $N^{2/3}$, and the dashed line shows the ratio between the former and the size of the giant component size of a random graph with the same number of nodes and edges. The inset shows the same data for the effective network.

entists and (ii) a true confluence of research interests justifying joint work. Given these ingredients, different scenarios can arise depending on the laws governing the mechanics of co-authorship. Such laws can be analyzed by observing their global imprint on the network, as we will now do.

3.2. Large-Scale Structure

One of the most distinctive features of complex networks is the emergence of a giant component, namely a single connected component that comprises a non-trivial fraction of all nodes. This component arises from the coalescence of multiple smaller components as new links are added. Of course, new nodes are also continuously added and these may attach to existing nodes or not. Thus, new components may be continuously created but as the giant component grows chances are higher that new nodes will also attach to it as it will encompass most of the network. Fig. 3a shows the evolution of the number of connected components both in the cumulative and in the effective network. As expected, the number of components grows very fast, increasing by 47% (resp. 35%) by year in the cumulative (resp. effective) network until the mid 2000s. From around 2005 the growth markedly decelerates, pointing to the initial stages of the formation of a giant component. However, this component is not really large –customarily, we can say a component is large if it comprises at least $N^{2/3}$ nodes, where N is the total number of nodes [21]– until very recently (2009 for the cumulative network, 2010 for the effective network) as shown in Fig. 3b. Indeed, this largest component is much smaller than that of a random –Erdős-Rényi (ER) [22]– network with

(a) 2006 (b) 2011

Figure 4: The giant component of the computational intelligence in games network in (a) 2006 and (b) 2011. Nodes represent authors and edges represent a collaboration between them.

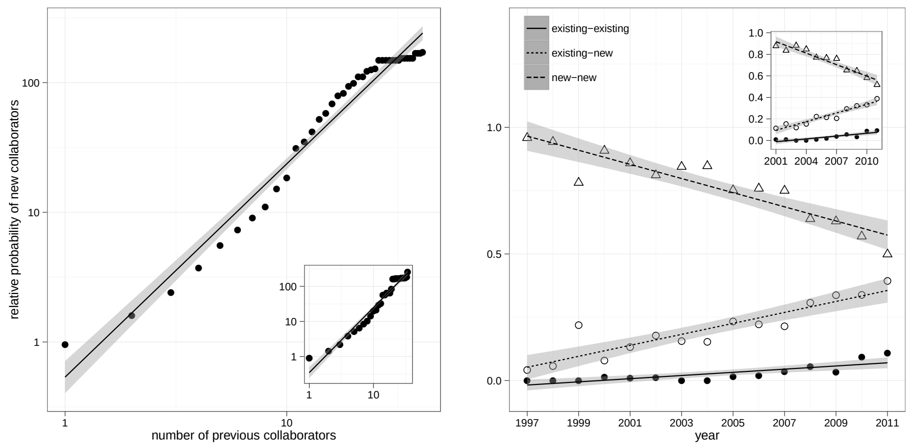
the same number of nodes and density (the fraction q of nodes contained in the giant component of a random network is given by the solution to $q = 1 - e^{q(N-1)p}$, where p is the network density and $N = |V|$ is the number of nodes [23]). Furthermore, its size is still relatively modest (about 18.6% of the cumulative network –and 13.5% of the effective network– in 2011) in comparison with say, 36.5% for a genetic programming (GP) network [14] or 62.8% for an evolutionary computation (EC) network [13] just to mention two related examples. This indicates the CIG network is still at an early stage of development, forming bonds and gaining cohesiveness but not yet fully formed. For illustration purposes, Fig. 4 shows the evolution of the largest component by providing two snapshots of its structure in 2006 and 2011.

Having shown how the macrostructure of the CIG network changes over time, we will turn our attention to the growth mechanics of the network in next subsection.

3.3. Growth Mechanics: Preferential Attachment

In order to understand how the network grows, let us start by observing the actual distribution $P(k)$ of node degrees. This is shown in Fig. 2b. Visual inspection suggests its tail follows a power-law $P(k) \sim k^{-\gamma}$ and is thus a scale-free network [7]. This is not uncommon in other scientific collaboration networks, e.g. [10, 12, 17, 14] and suggests the growth of the network is driven by preferential attachment [7]. Contrary to Erdős-Rényi networks [22] in which two vertices may link with constant uniform probability p (thus giving rise to a Poisson distribution of node degrees $P(k) \sim e^{-k}\lambda^k/k!$), in the presence of preferential attachment a new vertex will more likely connect to high-degree vertices than to low-degree ones. More precisely, let the probability of a new vertex linking to a vertex v be

$$\Pi_v = \frac{k_v^\alpha}{\sum_{w \in V} k_w^\alpha}, \tag{1}$$



(a) (b)

Figure 5: (a) Cumulated relative probability (computed over the last 5 years (2007-2011)) of new authors collaborating with existing authors depending on the number of previous collaborators of the latter. The data follows a power-law ($\alpha + 1 = 1.64$ and $R^2 = 0.97$). The inset shows the cumulated relative probability of new collaborations (i.e. existing authors starting collaborations for the first time among them) depending on the number of previous collaborators. The data follows a power-law ($\alpha + 1 = 2.01$ and $R^2 = 0.91$). In all cases the number of previous collaborators is computed using the cumulative network. (b) Links type percentage for cumulative data. The solid line shows a linear fit to the number of links between existing nodes ($\Delta_x = 0.0036$, $R^2 = 0.68$), the dotted line shows a linear fit to the number of links between an existing and a new node ($\Delta_x = 0.0217 R^2 = 0.83$) and the dashed line shows a linear fit to the number of links between new nodes ($\Delta_x = -0.0279$, $R^2 = 0.85$). The inset shows the same data for the effective network. Fit parameters are $\Delta_x = 0.0087$, $R^2 = 0.75$, $\Delta_x = 0.0269$, $R^2 = 0.0268$ and $\Delta_x = -0.0356$, $R^2 = 0.92$ respectively.

where k_w is the degree of a vertex $w \in V$ and $\alpha \geqslant 0$ is some constant. For $\alpha = 1$ we have linear preferential attachment which gives rise to a scale-free network. As a matter of fact, if nodes enter the network at a constant rate and link to a constant number m of existing nodes under linear preferential attachment an exponent $\gamma = 3$ is obtained [10, 24]. The exponent we obtain in the CIG network is somewhat larger ($\gamma = 3.27$). As a matter of fact, although the fit to the power-law is good ($R^2 = 0.96$), the distribution better adapts to a log-normal distribution. More precisely, we find that the tail of the distribution has a slope $\approx -3 \ln k$ on a log-log scale. A log-normal behavior has also been observed by [25] when analyzing Slovenia's scientific collaboration network and suggests it may arise from near-linear preferential attachment [26].

To analyze the existence of preferential attachment in the CIG network we follow the methodology suggested by [27]. The underlying idea is to measure to what extent the actual attachment rate of new nodes departs from a uniform distribution. In this sense, note that if we denote by $n_{k'} = |\{v \in V | k_v = k'\}|$ the number of nodes in the network with exactly k' neighbors, the probability $P_{k'}$ that a new node v attaches to a vertex $w \in V$ with degree $d_w = k'$ would be $P_{k'} = n_{k'}/N$ if links are uniformly random. If preferential attachment is at work, high-degree vertices will receive more links than indicated by the previous expression. A term $R_{k'}$ is introduced to account for this deviation, resulting in the expression:

$$P_{k'} = R_{k'} \frac{n_{k'}}{N} \tag{2}$$

$R_{k'}$ can thus be interpreted as the relative probability of attachment to a vertex with degree k'. It can be calculated by computing the actual probabilities $P_{k'}$ from the empirical data and solving for it in Eq. (2) using the known values n_k and N. Note also that all these quantities are time-dependent and should be understood as a function of time t. Fig. 5a shows the results for the CIG network. Given the moderate size of G_t and the fact that not all high degree values may be adequately represented in it, we use cumulative data ($R'_k = \sum_1^k R_k$) to improve statistics. The data fits to a power-law with exponent $\alpha + 1 = 1.64$. This indicates a sub-linear preferential attachment rate, much like [14] found for the GP network ($\alpha = 0.76$), and by Barabási et al. for mathematics ($\alpha = 0.8$) and neuroscience ($\alpha = 0.75$). We hypothesize that the relatively lower value in this case is due to the fact that the CIG network is still at an early developmental stage in comparison with the networks mentioned and hence it is more decentralized (i.e. prominent figures are emerging but they are not yet fully responsible for the macroscopic agglutination of new authors in the whole network). This is also borne out by inspecting the profile of new links in the network as shown in Fig. 5b. Note how the bulk of new links corresponds to so-called *external* links, i.e. links between new authors entering the network. There is however a marked decreasing trend in the proportion of such links, whereas *attachment* links (i.e. those connecting new authors to existing ones) and *internal* links (i.e. those between existing authors) are clearly increasing in importance. Regarding the former, it is interesting to note the trend similarity to the GP network [14] and the fact that in this larger community a crossover (attachment links surpassing external links) took place in the year 2000 whereas we will have to wait until 2014 (resp. 2015) to see a similar event in the effective (resp. cumulative) CIG network.

As for internal links, they have been recognized as having a huge influence on the topology and dynamics of the network [28]. As pointed out by [10] we have also observed these links to be subject to preferential attachment in the CIG network (see the inset in Fig. 5a). Indeed, this preferential attachment is linear, since the cumulative data follows a power law with exponent $\alpha + 1 = 2.01$. Hence, established authors are increasingly likely to collaborate for the first time, the more previous co-authors they have. This however does not explain the whole story since in principle authors working in disparate sub-areas are not as prone to collaborate if only because of a divergence of interests. Furthermore, once a collaboration is established it is interesting to analyze how long it lasts (or more precisely how productive it is). The following subsection will shed some light on how collaborators are picked and the extension of such collaborations.

3.4. Collaboration patterns and clustering

As mentioned before, the selection of a co-author is a process in which the *authoritativeness* of the actors (as measured by their number of previous co-authors) has

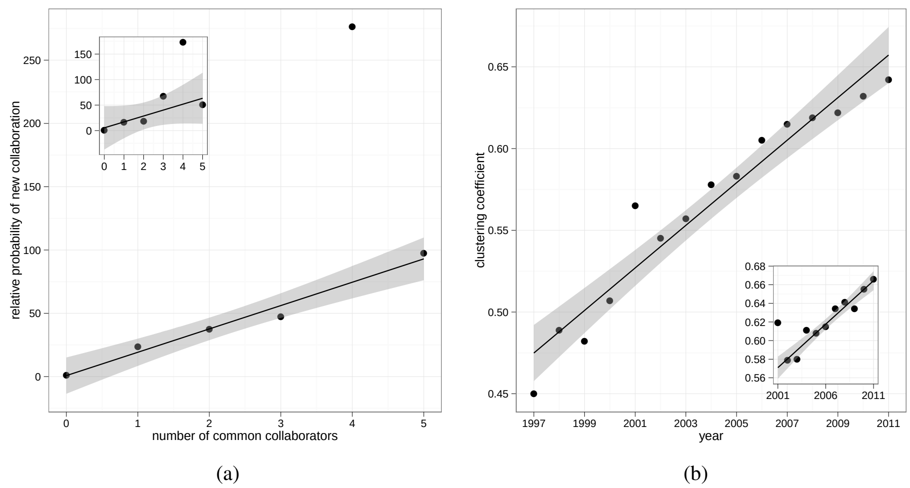
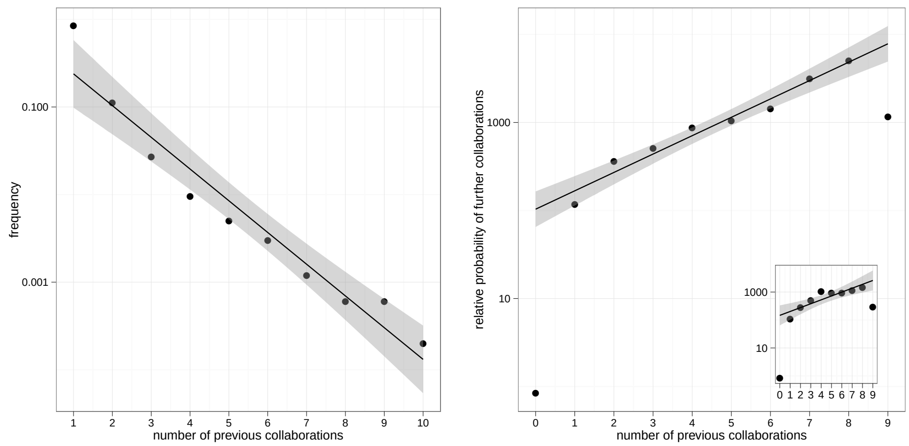


Figure 6: (a) Relative probability of new collaboration depending on the number of common collaborators. The solid line is the best lineal fit ($\Delta_x = 18.44$, $R^2 = 0.98$, without including $x = 4$ whose outlier status we attribute to a finite effect due to the small size of the network). The inset shows the same data for the effective network ($\Delta_x = 11.59$, $R^2 = 0.66$, without including $x = 4$). (b) Evolution of the clustering coefficient. The solid line is the best lineal fit ($\Delta_x = 0.013$, $R^2 = 0.93$). The inset shows the evolution of this coefficient grouped by 5-year frames and its lineal fit ($\Delta_x = 0.0093$, $R^2 = 0.94$)

an influence but it is not the only factor. A confluence of research interests is necessary too. To some extent, this confluence can be indirectly inferred by the existence of common co-authors: assuming each author is characterized by a collection of research topics and a collaboration implies a non-empty intersection of the corresponding topic collections, the higher the number of common co-authors that two non-connected authors have, the greater the chance that their research interests will have some overlap. Thus, it should be more likely that two authors collaborate if both of them have previously collaborated with the same people in the past. As in [27], this phenomenon has been measured in a similar way as above, namely computing the relative probability R_m that two authors with m previous co-authors start collaborating together. If we denote as n_m the number of non-connected author pairs with exactly m common neighbors, we have that the probability P_m of a new internal link connecting two such authors is

$$P_m = R_m \frac{2n_m}{N(N-1)} \tag{3}$$

The relative probabilities R_m for the CIG network are shown in Fig. 6a. As it can be seen, there is a linearly increasing relative probability of collaborating as the number of common collaborators increases. This collaboration pattern gives rise to clustering, namely the fact that one's neighbors are more likely to be neighbors themselves than pure random chance would otherwise indicate. This is quantitatively captured by the network's clustering coefficient. Mathematically, the local clustering coefficient C_i of a node i is given by $C_i = 2E_i/\left[k_i(k_i-1)\right]$ where E_i is the number of edges connecting



(a) (b)

Figure 7: (a) Distribution of the number of previous collaborations. Y-axis in log-scale. The solid line shows the best exponential fit ($P(y) \sim a^x, a = 0.4343, R^2 = 0.94$) (b) Relative probability of further collaboration depending on the number of previous collaborations in 2011. Y-axis in log-scale. The solid line shows the best exponential fit ($a = 1.6318, R^2 = 0.97$). The inset shows the same data grouped by 5 years and its exponential fit ($a = 1.3763, R^2 = 0.80$). The data corresponds in both cases to the cumulative network and the fitting excludes the first and last data points.

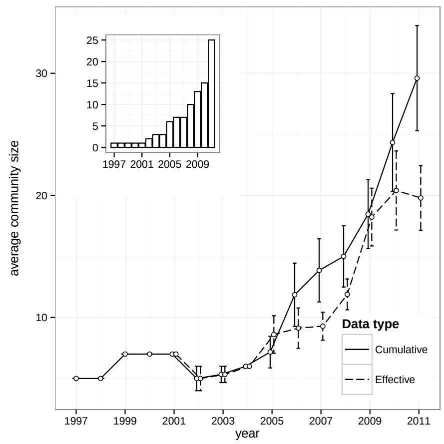
the immediate neighbors of node i and k_i is the degree of node i [29], i.e. the fraction of one's neighbors that are neighbors too. Once the local clustering coefficient has been computed for each node, it can be averaged to obtain the clustering coefficient of the whole network. The evolution of this coefficient is shown in Fig. 6b. quite interestingly, we see there is a linearly increasing trend in clustering. This contrasts with the empirical findings of [25] regarding Slovenia's scientific community, but is consistent with [14] results for the GP network. Similarly, [10] have analyzed a simple model for the evolution of co-authorship networks in which there is a constant net growth and preferential attachment rules the creation of both attachment and internal links (much as we have observed in the case of the CIG network) and have observed that for any positive rate of creation of internal links there should be an asymptotical increase in the clustering coefficient.

Having analyzed how internal links are created, let us now turn our attention to how long collaborations last. To this end, it is useful to analyze the distribution of edge weights in the network. Recall that each edge is weighted by the number of times the corresponding authors have collaborated. Therefore, this distribution indicates how likely it is that an established collaboration reaches a given productivity. This distribution is shown in Fig. 7a. Note how the number of co-authored papers that two authors have is exponentially distributed. This indicates a memoryless process in which the probability of getting a new paper is constant (about 43% from the observed empirical data). This decay rate can be attributed to the intrinsic difficulty of getting a paper accepted (i.e. related to the acceptance rates in the area) and/or to the fact that collaborations dissipate after a certain amount of time. We believe the former reason is more important than the latter in light of the data shown in Fig. 7b. This figure shows the relative probability of further collaborations depending on the number of previous collaborations (computed using the methodology described before). Excluding the first point (corresponding to no previous collaboration and hence beyond the scope of established collaborations) and the last point (where a decay due to the finite size of the network and the number of publications is present), this relative probability grows geometrically. As a comparison, [27] found a linear increase of this relative probability for two databases of Physics and Medical papers. Although the growth factor is not much larger than 1 – considering the last five years, it is about 38% more likely to get a new paper with a co-author with whom one has m papers than with another one with $m-1$ previous papers – this indicates a interesting trend of "fidelity" in research collaborations, which seem to be very durable in the CIG area.

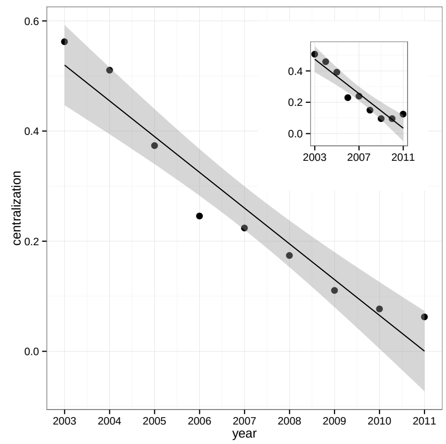
3.5. Mesoscopic structure of the network

The attachment and internal linking dynamics described in the previous subsections give rise to a particular structure of the network when analyzed on a larger scale. The CIG network is not homogeneous, since some groups of authors are densely connected between each other, but sparsely linked with authors outside the group. These groups are called *communities* [30]. There are several approaches for identifying community structure in networks. In this work we have opted for a method based on the greedy optimization of modularity [31]. This method merges individual nodes into communities in a way that greedily maximizes the modularity score of the network. Intuitively, modularity is a quality index of a partition of the network. Good scores are those in which the internal cohesiveness (measured by the number of internal connections inside groups) is high and maximally different from the average density of the network. We refer to the previous references and to [32] for more details on the modularity measure.

We have applied this analysis to obtain interesting information about how communities evolve over time. Fig. 8 shows the evolution of the number of communities and their sizes. Not only does the number of communities grow over time (since 2003, the growth is more marked – see the inset in Fig. 8a) but so does their size (Fig. 8a). Communities get larger by attachment of new authors and by aggregation of smaller communities, whereas they can get smaller by dividing into separate communities over time. As shown in Fig. 8b the increase in the number of communities is also macroscopically reflected in the decreasing degree centralization of the giant component (intuitively, degree centralization measures the *starness* of the network, i.e. the extent to which it resembles the most centralized structure, namely a star with a central hub to which all remaining nodes connect – see [33] for a detailed definition). This pattern was also observed by [13] for the EC network and can not be just attributed to the percolation of different communities into a giant component but also to the age of nodes (as a scientist evolves from PhD student to senior researcher (s)he begins to develop new connections and take more students under his supervision, thus giving rise to new hubs in the network). A preliminary analysis of the publication venues chosen by each community indicates that there is a significant overlap among them, hence suggesting that venue segregation is not the ultimate factor behind community formation.

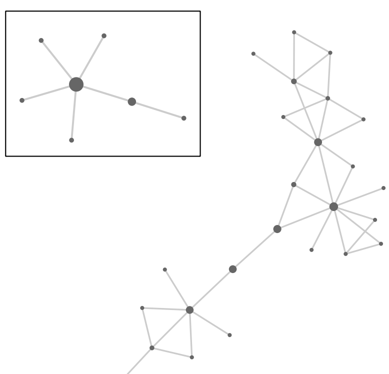


(a)

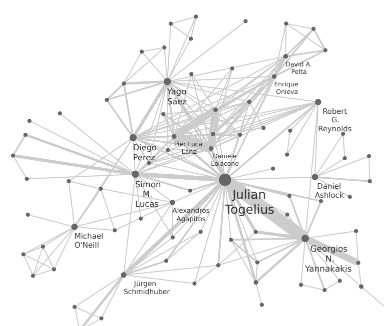


(b)

Figure 8: (a) Evolution of the average community size in the giant component of the cumulative network. The inset shows the number of communities in the giant component. (b) Evolution of degree centralization in the giant component of the cumulative network. A linear fit ($\Delta_x = -0.065, R^2 = 0.93$) is included. The inset shows this data for the effective network ($\Delta_x = -0.055, R^2 = 0.87$).



(a)



(b)

Figure 9: (a) Community interaction graph in the giant component until 2006 (sub-figure) and 2011, using cumulative data. Each node represents a community. Node size is proportional to the betweenness of the node. (b) Largest and most connected community in the giant component until 2011 using cumulative data. Node size is proportional to the betweenness of the node and edge size is proportional to the number of collaborations between the authors.

Once communities have been identified we can construct the community interaction graph, i.e., a graph where all authors within a community collapse into a single node and edges indicates collaboration between authors from two separate communities. Fig. 9a shows this interaction graph for the giant component in G_{2011}. Inside this

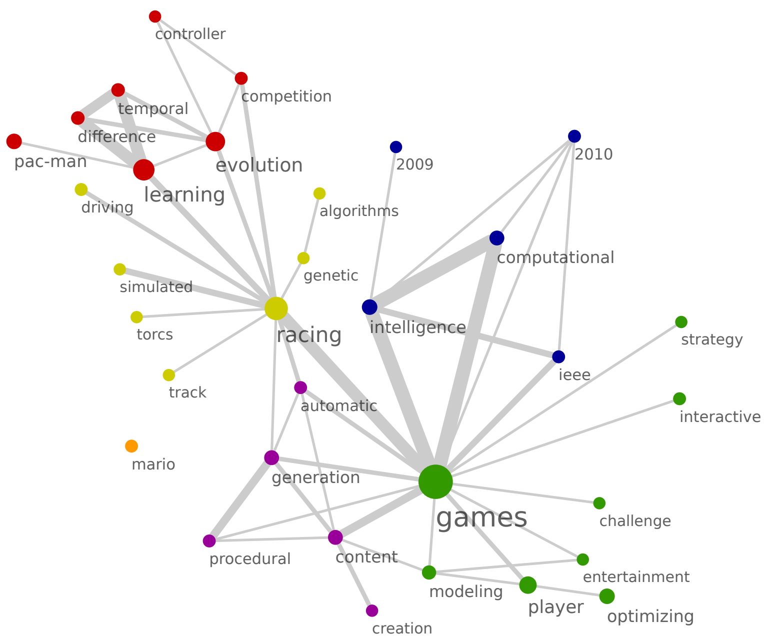


Figure 10: Graph representing the most-frequently used words in the title of articles written by at least one member of the largest and most central community. Each node represent a word and an edge between two nodes means that these two words appear in the same paper's title. Node sizes represent the frequency of the words, while edges sizes represent the frequency of the relationship. Node color represents the community of the node.

figure, there is a sub-figure that shows the community interaction graph for the giant component in G_{2006}. It can be clearly seen how the giant component has not just grown in size but has also developed an increasingly complex internal structure with a large number of interconnected communities. Note how this interaction graph has, however, an elongated shape, which is not the signature of a scale-free network from this perspective. Furthermore, there is a single community which can be identified as playing a prominent role both in the interaction graph (it is the most connected community and the one with higher betweenness – see Sect. 3.6) and in the original network as well (it is the largest one). This community is depicted in Fig. 9b using node sizes and edge widths to denote the centrality (a topic which will be tackled again later) of authors and the strength of collaborations respectively.

In order to analyze the scope and research area of this prominent community, we have to analyze not just the authors but the publication text [34]. To be precise, we have generated a graph representing the words used in the title of the articles written by at least one of its members, see Fig. 10. In this graph, each node represents a single word and edges between nodes mean that these words appear in a same article's title. The node size represents the frequency of the word, while the size of the edges represents the frequency of the relationship. Words such as articles or prepositions have been filtered because they do not offer any useful information. Additionally, we have only

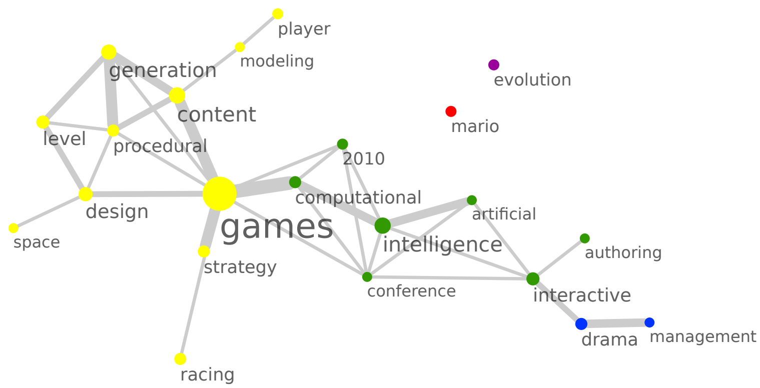


Figure 11: Graph representing words most used in the titles of articles written by at least one member of the most central authors in 2011's effective network. Each node represents a word and an edge between two nodes means that these two words appear in the same paper's title. Node sizes represent the frequency of the words, while edges sizes represent the frequency of the relationship.

included words summing up the top 10% of appearances to avoid clutter. For the same reasons, edges connecting words that only appear together once are not included in the graph. Once this graph has been obtained, we run a community detection algorithm as above, suggesting 6 topic areas in which this community is working. Apart from a group of keywords related to proceedings edition, there are three areas dealing with specific games: Super Mario (isolated node), Pac-Man (in the context of evolution and temporal difference learning) and car-racing games (involving TORCS and techniques such as genetic algorithms). There is also an indication of two topic areas, one centered on procedural content generation and the other one on interaction and player modeling. This indicates some of the topics in which a core part of the CIG community is working. A more global perspective of such topics requires a broader characterization of who the core researchers are in the CIG community. This will be tackled next.

3.6. Centrality analysis

Centrality measures are indicators of the importance of a node within the network. This importance can be assessed in different ways depending on the context and the meaning of connections and hence different centrality measures have been defined in the literature. Borgatti and Everett [35] have provided a cross-classification of centrality measures along two axes: radial vs medial and volume-based vs length-based. In order to capture the corresponding different perspectives on centrality provided by this classification we have considered four centrality measures, each of them belonging to a different cell of this four-fold characterization. Betweenness [36] has been chosen as the volume-based medial measure. Roughly speaking, high betweenness nodes are those that appear in many of the shortest paths between nodes (and hence are important intermediaries in the exchange of information). As for volume-based radial measures we have chosen Kleinberg's centrality [37] defined as the principal eigenvector of $A' \times A$, where A is the adjacency matrix of the graph. Intuitively, a node is here deemed influential if it is also connected to other influential nodes. Regarding the length-based dimension we have chosen closeness [38] as the radial measure. Nodes with high closeness have a low average distance to the rest of the nodes in the network (measured as the number of steps required to reach the latter). Thus, they can be considered privileged emitters as information originating in them will quickly reach the rest of the network. Finally, Borgatti's distance fragmentation [35] is chosen as the length-based medial measure. This index accounts for the change in the average distance between nodes once a certain node is removed. Highly central nodes under this measure are those which produce a large increase in this average distance and thus are important in maintaining the cohesion of the network.

We have computed these centrality measures in the effective network for 2011 for each node and treated this data as a multi-objective optimization problem as done by [39]. The Pareto front comprises four scientists. Subsequently, we conducted a similar keyword analysis as in the previous section, i.e. we picked the papers written by at least one of these authors and created the word graph depicted in Fig. 11. Now, a single community comprising procedural content generation, car racing and player modeling emerges. Two isolate nodes reflect the relevance of Super-Mario and evolution. Finally a topic area related to a conference edition bridges the gap with topics such as drama management, interaction and authoring. Overall, this graph provides an interesting overview of some of the most relevant research topics in the CIG community.

4. Conclusion and future work

This work has provided the first steps towards understanding the dynamics of the Computational Intelligence in games (CIG) co-authorship network and its modeling, using an individual-based model. CIG is an active and vibrant field, attracting new researchers and generating new papers, having a stable growth in the number of authors, which had an accelerated growth up to the early/mid-2000s. The number of published papers per year has been increasing since 2005, sustaining the steady growth of the network. In addition to a sustained influx of new authors, there is a constant flux of authors moving out this field of research that might indicate a light deceleration of the community. However, the real trend for volatility is more similar to that of new authors, supporting the estimation of constant growth of the community. The CIG network is still at an early stage of development, forming bonds and gaining cohesiveness but not yet fully formed, as we can observe from the size of its largest component.

Addressing growth mechanics, the growth of the network is driven by preferential attachment with a sub-linear rate. This rate is due to the fact that the CIG network is still gaining maturity and hence it is more decentralized. Preferential attachment supports the fact that new authors are increasingly more likely to enter the field by collaborating with established authors when the latter have a higher number of previous co-authors. There is also a significant fidelity in research collaborations, due to a geometric growth in the relative probability of a further collaboration depending on the number of previous collaborations.

With regard to the mesoscopic structure of the network, the increase in the number of communities is reflected in the decreasing degree of centralization of the giant component. The incorporation of smaller communities to the latter suggests that the CIG community is starting to percolate thus supporting the idea that the network is still at an early stage of development. Precisely this fact, and the heterogeneity of the field, should encourage researchers who have not yet made their presence felt in the CIG community, to participate in this area.

Future work will include an analysis of which sub-fields of computational intelligence in games are gaining importance in the community. This is an interesting (and hard to do) analysis, because it is not enough to process just the authors of the article, but the complete text [34]. We are working on this to obtain a classification of community authors based on the topics they research. We will also include additional data from other bibliographic databases, like Scopus[2] or Google Scholar[3].

Acknowledgements

This work is partially supported by Spanish MICINN under project ANYSELF[4] (TIN2011-28627-C04-01), and by Junta de Andalucía under project P10-TIC-6083 (DNEMESIS[5]).

References

[1] S. Lucas, Computational intelligence and games: Challenges and opportunities, International Journal of Automation and Computing 5 (1) (2008) 45–57.

[2] J. Togelius, G. Yannakakis, K. Stanley, C. Browne, Search-based Procedural Content Generation: A Taxonomy and Survey, IEEE Transactions on Computational Intelligence and AI in Games 3 (2011) 172–186.

[3] U. Ritterfeld, M. Cody, P. Vorderer (Eds.), Serious Games: mechanisms and effects, Routledge, Taylor and Francis, 2009.

[4] M. Pirovano, R. Mainetti, G. Baud-Bovy, P. L. Lanzi, N. A. Borghese, Self-adaptive games for rehabilitation at home, in: 2012 IEEE Conference on Computational Intelligence and Games, CIG 2012, IEEE, Granada, Spain, 2012, pp. 179–186.

[5] J. Scott, Social Network Analysis, Sociology 22 (1) (1988) 109–127.

[6] M. E. J. Newman, The Structure and Function of Complex Networks, SIAM Review 45 (2) (2003) 167–256.

[7] A.-L. Barabási, R. Albert, Emergence of Scaling in Random Networks, Science 286 (5439) (1999) 509–512.

[2]http://www.scopus.com
[3]http://scholar.google.com
[4]http://anyself.wordpress.com/
[5]http://dnemesis.lcc.uma.es/wordpress/

[8] D. J. Watts, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.

[9] T. Luukkonen, O. Persson, G. Sivertsen, Understanding Patterns of International Scientific Collaboration, Science, Technology & Human Values 17 (1) (1992) 101–126.

[10] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A: Statistical Mechanics and its Applications 311 (34) (2002) 590–614.

[11] J. W. Grossman, The evolution of the mathematical research collaboration graph, Congressus Numerantium (2002) 201–212.

[12] C. Cotta, J. J. Merelo, The complex network of EC authors, ACM SIGEVOlution 1 (2) (2006) 2–9.

[13] C. Cotta, J. J. Merelo, Where is evolutionary computation going? A temporal analysis of the EC community, Genetic Programming and Evolvable Machines 8 (3) (2007) 239–253.

[14] M. Tomassini, L. Luthi, Empirical analysis of the evolution of a scientific collaboration network, Physica A 385 (2007) 750–764.

[15] D. B. Horn, T. A. Finholt, J. P. Birnholtz, D. Motwani, S. Jayaraman, Six degrees of jonathan grudin: a social network analysis of the evolution and impact of CSCW research, in: Proceedings of the 2004 ACM conference on Computer supported cooperative work, CSCW '04, ACM, New York, NY, USA, 2004, pp. 582–591.

[16] H. Hou, H. Kretschmer, Z. Liu, The structure of scientific collaboration networks in Scientometrics, Scientometrics 75 (2) (2008) 189–202.

[17] M. E. J. Newman, Scientific collaboration networks. I. Network construction and fundamental results, Phys. Rev. E 64 (1) (2001) 16131.

[18] M. E. J. Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Phys. Rev. E 64 (1) (2001) 16132.

[19] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, Genome Research 13 (11) (2003) 2498–2504.

[20] G. Csárdi, T. Nepusz, The igraph software package for complex network research, InterJournal Complex Systems (1695) (2006) 1695.

[21] B. Bollobás, Random Graphs (Cambridge Studies in Advanced Mathematics), 2nd Edition, Cambridge University Press, 2001.

[22] P. Erdős, A. Rényi, On the evolution of random graphs, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5 (1960) 17–61.

[23] P. Erdős, A. Rényi, On random graphs, Publicationes Mathematicae 6 (1959) 290–297.

[24] M. Jackson, Social and Economic Networks, Princeton University Press, Princeton NJ, 2008.

[25] M. Perc, Growth and structure of Slovenia's scientific collaboration network, Journal of Informetrics 4 (2010) 475–482.

[26] S. Redner, Citation statistics from 110 years of Physical Review, Physics Today 58 (2005) 49.

[27] M. E. J. Newman, Clustering and preferential attachment in growing networks, Physical Review E 64 (2001) 025102(R).

[28] S. Dorogovtsev, J. Mendes, Scaling behaviour of developing and decaying networks, Europhys. Lett. 52 (2000) 33–39.

[29] B. Bollobás, R. Kozma, D. Miklos, Handbook of Large-Scale Random Networks: Bolyai Society Mathematical Studies, 1st Edition, Springer Publishing Company, Incorporated, 2009.

[30] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences 99 (12) (2002) 7821–7826.

[31] A. Clauset, M. E. J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004) 066111.

[32] M. E. J. Newman, M. Girvan, Finding community structure in very large networks, Phys. Rev. E 69 (2003) 026113.

[33] S. Wasserman, K. Faust, Social Network Analysis, Cambridge University Press, Cambridge, 1994.

[34] M. Perc, Self-organization of progress across the century of physics, Scientific Reports 3 (2013), 1720.

[35] S. P. Borgatti, M. G. Everett, A Graph-theoretic perspective on centrality, Social Networks 28 (4) (2006) 466–484.

[36] U. Brandes, A faster algorithm for betweenness centrality, The Journal of Mathematical Sociology 25 (2) (2001) 163–177.

[37] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1999) 604–632.

[38] L. C. Freeman, Centrality in networks: I. conceptual clarification, Social Networks 1 (3) (1978) 215–239.

[39] J. J. Merelo, C. Cotta, Who is the best connected EC researcher? centrality analysis of the complex network of authors in evolutionary computation, in: H. Lipson (Ed.), 2007 Genetic and Evolutionary Computation Conference, ACM Press, London, UK, 2007, p. 1526.

Appendix A. Conferences, journals and keywords used to query the DBLP database

- Artificial and computational intelligence in games:
  - IEEE Trans. on Computational Intelligence and AI in games (TCIAIG)
  - Artificial Intelligence and Interactive Digital Entertainment Conference (AI-IDE)
  - IEEE congress on Computational Intelligence and Games (CIG)
  - Agents for Games and Simulations (AGS)
  - Advanced Intelligent Paradigms in Computer Games
- Artificial intelligence and evolutionary:
  - IEEE congress on evolutionary computation (CEC)
  - Genetic and Evolutionary Computation Conference (GECCO)
  - Parallel Problem solving from Nature (PPSN)
  - National Conference on Artificial Intelligence
  - Applications of Evolutionary Computation (EvoApplications)
  - International Conference on Artificial Neural Networks (ICANN)
  - Canadian Conference on Artificial Intelligence
  - Artificial Intelligence and Applications (AIA)
  - Artificial Intelligence Applications and Innovations (AIAI)
  - Artificial Intelligence and Computational Intelligence (AICI)
  - Artificial Intelligence: Methodology, Systems, Applications (AIMSA)
  - Artificial Intelligence and Simulations (AIS)
  - Computational Intelligence
  - European Conference on Artificial Intelligence (ECAI)
  - International Conference on Agents and Artificial Intelligence (ICAART)
  - International Conference on Artificial Intelligence (IC-AI)
  - International Conference on Artificial Intelligence and Soft Computing (ICAISC)
  - International Conference on Evolutionary Computation (ICEC)
  - International Conference on Genetic Algorithms (ICGA)
  - International Conference on Genetic and Evolutionary Computing (ICGEC)
  - International Conference on Intelligent Computing (ICIC)
  - International Joint Conference on Artificial Intelligence (IJCAI)
  - International Joint Conference on Computational Intelligence (IJCCI)
  - International Joint Conference on Neural Network(IJCNN)
  - International Symposium on Computational Intelligence and Design

  - International Symposium on Neural Networks
  - International Work-Conference on Artificial and Natural Neural Networks (IWANN)
  - IITA International Joint Conference on Artificial Intelligence
  - Modeling Decisions for Artificial Intelligence
  - IEEE Transactions on Evolutionary Computation (TEC)
  - International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC)
  - International Conference on Adaptive and Natural Computing Algorithms (ICANNGA)
  - Artificial Life (ALIFE)
  - Simulation of Adaptive Behavior (SAB)
  - International Conference on Artificial Neural Networks (ICANN)
  - Simulated Evolution and Learning (SEAL)
  - European Conference on Advances in Artificial Life (ECAL)
  - International conference on Swarm Intelligence (ICSI)
  - Hybrid Intelligent Systems (HIS)
  - European Symposium on Artificial Neural Networks (ESANN)
  - IEEE Transactions on Neural Networks
  - IEEE Transactions on Fuzzy Systems

- Games:
  - GAME-ON Conference (GAMEON)
  - Advances in Computer Games (ACG)
  - Computers and Games
  - Computers Games Conference (CGAMES)
  - Conference of the Digital Games Research Association
  - International Conference on Entertainment Computing
  - Computers in Entertainment
  - Advances in computer entertainment technology

- Keywords:
  - Artificial and computational intelligence in games: no keywords. All articles selected.
  - Artificial intelligence and evolutionary: game/s, puzzle, player.
  - Games: fuzzy, evol*, genetic, swarm, agent, local search, neural, ant.
  - Other: all combinations from {fuzzy, evol*, genetic} and {game/s, puzzle, player}