ORIGINAL PAPER

On the analysis of the collaboration network of the Brazilian symposium on computer networks and distributed systems

30 Editions of history

**Guilherme Maia · Pedro O. S. Vaz de Melo ·
Daniel L. Guidoni · Fernanda S.H. Souza · Thiago H. Silva ·
Jussara M. Almeida · Antonio A. F. Loureiro**

Abstract The Brazilian symposium on computer networks and distributed systems (SBRC) reached its 30th edition as the paramount scientific event in the area of computer networks and distributed systems in Brazil. Faced with this opportune moment in the event's history, we here study the collaboration network established among authors who have jointly published in the symposium. Towards that end, we collected bibliographic data from all 30 editions, and built the co-authorship network of the event. We then analyzed the network structural features and evolution throughout its history. Our results reveal the main kind of co-author relationship among authors, show the most prominent communities within SBRC, the regions of Brazil that attracts the most authors, the researchers with central roles in the network as well as the importance of inter-state collaborations. Finally, we align our results with historical facts that may have had a key impact on the symposium success.

Keywords Collaboration networks · Scientific networks · Social networks · SBRC

G. Maia, P. O. S. Vaz de Melo, D. L. Guidoni, F. S. H. Souza and T. H. Silva have equally contributed to this research paper.

G. Maia (✉) · P. O. S. Vaz de Melo · T. H. Silva · J. M. Almeida · A. A. F. Loureiro
Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
e-mail: jgmm@dcc.ufmg.br

P. O. S. Vaz de Melo
e-mail: olmo@dcc.ufmg.br

T. H. Silva
e-mail: thiagohs@dcc.ufmg.br

J. M. Almeida
e-mail: jussara@dcc.ufmg.br

A. A. F. Loureiro
e-mail: loureiro@dcc.ufmg.br

D. L. Guidoni · F. S. H. Souza
Federal University of São João del-Rei,
São João del-Rei, MG, Brazil
e-mail: guidoni@ufsj.edu.br

F. S. H. Souza
e-mail: fsumika@ufsj.edu.br

1 Introduction

In 2012, the Brazilian symposium on computer networks and distributed systems (SBRC) reached its 30th edition as the paramount scientific event in the area of computer networks and distributed systems in Brazil. Its importance may be evidenced by the number of papers submitted and by the number of participants in the last editions of the event. For instance, in the last few editions, the symposium received between 250 and 300 papers from about 1,000 authors, including researchers, professionals and students. Due to its wide acceptance, SBRC assembles most of the work in the areas of computer networks and distributed systems from Brazil's academic and professional communities, besides international researchers. Scientific events play a central role in knowledge dissemination, since they are one of the few opportunities for researchers with common interests to gather together, present new ideas and establish new collaborations. SBRC is not different, as we shall present throughout this paper. Hence, given this opportune moment in the event's history, a broad investigation of such research community is timely.

We use social network analysis (SNA) to further investigate this well established research community. Because of the popularity of online social networks and the large availability of real social data, SNA has gained a lot of momentum in the last few years [22, 26, 36, 43]. Besides online social networks

[20,23,25], it is possible to apply SNA to discover knowledge in the most diverse systems, such as mobile operators [12,19,40], Internet websites [1,13], railroads [13], citation networks [17], movies and actors [21], sports leagues [28] and many others.

In summary, a social network is composed by a set of individuals or a group connected by different kinds of relationships. Individuals, also known as *actors*, may represent a single person, a group or even an organization. Their relationships, or *ties*, may indicate, for instance, a friendship, a professional relationship or a scientific collaboration. Actors and ties are defined according to the question of interest.

A scientific collaboration network is a special type of social network in which the *actors* represent authors and *ties* indicate that the authors have published at least one paper together. Collaboration networks have been widely analyzed [31–33,35], as these studies disclose several interesting features of the academic communities that comprise them. For instance, the analysis of topological features enables the identification of communities [2], the intensity of collaborations among authors [11] and how the network evolves over the years [25].

Therefore, in this paper we study the SBRC's collaboration network. Towards that goal, bibliographic data from all 30 editions of the event were collected and a series of features, obtained from the topological structure of the collaboration network, was analyzed. In particular, we here investigate the evolution of the largest connected component, number of communities, importance of nodes, their degree distribution and correlations, and network homophily.[1] Through this study, it is possible to better understand the behavior of such a vibrant community and part of the impact produced by some crucial collaborations established through the years. For example, we are interested in investigating the peculiarities of collaborations among researchers from a region with a historically very active and productive research community, and among researchers from a region with no such community. It is worth noticing that when compared to previous studies on collaboration networks, our work stands out for three main reasons. First because we analysed 30 years of data, which to the best of our knowledge is more than any other study available in the literature. Second, our analysis considers several features that are usually not present, such as the geographic location of the researchers, the institutions they work for, among others. Finally, we make a parallel of our findings with several historical facts that may have had a key impact on the symposium success and also may have changed the way research is done in Brazil.

The remainder of this paper is organized as follows. Section 2 presents the related work. Then, Sect. 3 describes how data used in this work was collected and how the network was built. Section 4 presents some statistics about the participation of authors from different regions of Brazil. Next, Sect. 5 describes the main kinds of collaborations among authors, whereas Sect. 6 presents a study of the connected components of the network. Section 7 discusses distance and clustering measures, and Sect. 8 analyzes the main communities within SBRC. Researchers with strategic positions in the network are identified in Sect. 9, and Sect. 10 analyzes homophily in the SBRC network. Section 11 presents a cross analysis among some evaluated metrics. Section 12 presents the conclusions of this work. Finally, Appendix presents the historical aspects that may have contributed to foster the research development in Brazil.

2 Related work

The analysis of collaboration networks is well explored in the literature. For instance, Newman [31,32] presents some of the pioneering studies in this area. The author analyzes three scientific communities—Computer Science, Physics and Biomedicine—and presents several structural and topological features of these communities, focusing on the main similarities and differences among them. Although these communities share some similarities, Newman shows that they also have substantial differences. In that direction, Menezes et al. [29] assess how the process of knowledge production in Computer Science happens in different geographic regions of the globe. The authors divide the globe into three main regions and evaluate how research is conducted in 30 different subfields of Computer Science for each of the considered regions, focusing on the structural and temporal features of the network. Among the main results, Menezes et al. show that the scientific production of Brazilian researchers is increasing in recent years, which they attribute to an increase in funding provided by Brazilian government agencies to foster research in the country.

Towards analyzing the Brazilian scientific production, Freire and Figueiredo [15] show the main similarities and differences between two co-authorship networks they propose: "Global", created from all publications of the DBLP database, and "Brazilian", which is a subset of the first network considering only researchers affiliated to Brazilian institutions. Moreover, they propose a new ranking metric to measure the importance of both an individual in the network and groups of individuals. This metric is applied to the Brazilian network and is compared with two existing ranking measurements in Brazil: the Research Fellowship Program of CNPq (an agency of the Brazilian Ministry of Science and Technology) and graduate programs in Computer Science provided by CAPES (an agency of the Brazilian Ministry of Education). The authors show that the proposed metric can

---

[1] All metrics needed to perform this investigation are described in Sect. 3.

accurately identify influential groups and well-established graduate programs in Brazil.

There are studies that analyze specific events and areas. Procópio et al. [38] create and analyze the co-authorship network of articles published during the first 25 years of the Brazilian Symposium on Databases (SBBD). The authors focus on the network's structural features and temporal evolution throughout the event's history. They present and study statistics such as average number of papers per author, average number of papers per edition of the symposium, average co-authors per paper, among others. Finally, the work shows that the studied network follows a well-known phenomenon called small world, typically found in other social networks. Silva et al. [41] create and analyze the co-authorship network of papers published in three international top conferences focused on Ubiquitous Computing (Ubicomp). They provide useful analysis for that network, such as representativeness of authors and institutions, and formation of communities. Finally, Nascimento et al. [30] analyze the co-authorship graph of the ACM Special Interest Group on Management of Data (SIGMOD) Conference. Among the main results, the authors observe that the SIGMOD community is also a small-world network. In comparison with these previous studies of co-authorship networks of specific research communities, we go further and analyze three fundamental aspects of researchers who publish in SBRC: geographic location, topological characteristics in the network and productivity statistics in the conference.

Finally, scientific collaboration networks are not limited to co-authorship networks. Bazzan and Argenta [4] create a social network of the PC (Program Committee) members of conferences sponsored by the Brazilian Computer Society (SBC). The relations among nodes of this network are established according to co-authorship data extracted from the DBLP. By using well-known network metrics, such as node degree, largest connected component and clustering coefficient, the authors show that the studied network does not fit any well-established pattern when compared to other networks studied in the literature. This is probably due to the fact that members of this network do not necessarily interact with one another in terms of co-authorship, once they belong to different sub-areas within Computer Science. One of the main findings was that the most connected nodes are non-Brazilian PC members, and they play an important role in the network by acting as connectors between Brazilian researchers. When compared to our work, we point out that SBRC includes both well-established authors and newcomers to the symposium, while the PC network is formed exclusively by members of senior character, which explains the difference in some of the metrics. Nevertheless, we observed that the SBRC network follows similar patterns to other previously analyzed scientific events and communities, such as the ones in [30,38] and [41].

3 The network of the SBRC symposia

3.1 Data acquisition

Our study is based on bibliographic data of the 30 editions of SBRC, which took place from 1983 until 2012. It is focused on the collaboration network established among authors of papers published in the main track of each edition of the symposium. Thus, we collected data of full papers published in the proceedings of the event, excluding lectures, tutorials and workshop papers. For each paper, we collected its title, year of publication, list of authors with their respective affiliations, geographic location of the authors' institutions, and the language the paper was written. The data comprises digital and non-digital sources, since the first editions of the event occurred before the existence of the Web. Part of the bibliographic data was obtained automatically through the website of the Brazilian Computer Society (SBC)[2], while the rest was collected manually from the proceedings of each edition. We manually disambiguated all author names to ensure data consistency.

3.2 Network creation

In this paper, the SBRC network is represented as a temporal graph $G_y = (V_y, E_y)$, where V_y is the set of vertices, E_y is the set of edges and y is the year the network refers to. The graph $G_y = (V_y, E_y)$ is an undirected weighted graph, where the vertices are authors and the edges indicate that two authors have published together in and before the year y. Moreover, each edge has a corresponding weight, which represents the number of papers the authors published together in and before the year y.

The complete SBRC collaboration network, built from all papers published in its 30 editions, has a total of 1,808 authors (vertices) and 4,066 collaborations (edges), comprising a total of 1,406 papers. The average number of papers per year is 46.8 (with a standard deviation of 20.66) and the average number of authors per year is 115.1 (with a standard deviation of 65.51). The reason behind this large variance is due to the constant growth of the conference throughout the years. For instance, in the first year, 1983, the number of authors was 22 and the number of papers was 12. In the last year, 2012, the number of authors was 174 (690 % higher) and the number of papers was 59 (391 % higher). Finally, the average number of papers per author is 2.31 (with a standard deviation of 4.25), while the average number of authors per paper is 1.97 (with a standard deviation of 1.37). Figure 1 shows the complete SBRC network as viewed in 2012, representing 30 years of history. Observe that the network contains clusters of nodes with the same color, which

---

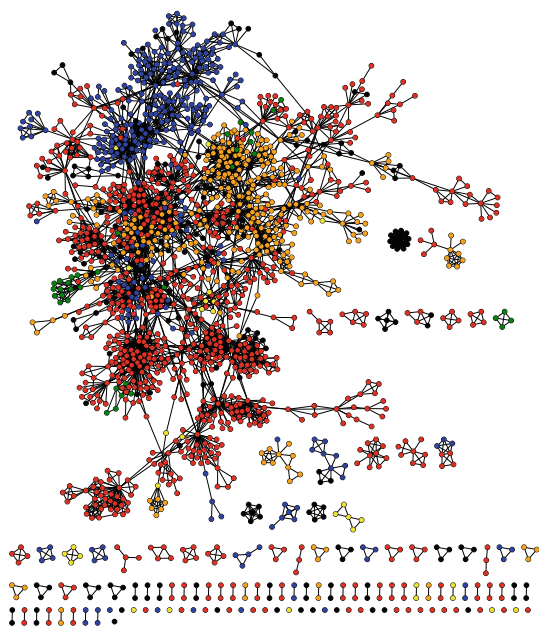[2] See http://bibliotecadigital.sbc.org.br.

Fig. 1 The complete network as viewed in 2012. *Nodes with the same color* represent authors affiliated to universities located in a given region of Brazil. *Green* represents authors affiliated to universities located in the North region of Brazil, *blue* for the South, *red* for South-east, *yellow* for Center-west, *orange* for North-east and finally, *black* for authors affiliated to foreign universities. Notice that the *colors* used in this figure are the same as the ones used in Fig. 4, where we show the Brazilian map

represent authors affiliated to universities located in a given region of Brazil. Green represents authors affiliated to universities located in the North region of Brazil, blue for the south, red for southeast, yellow for center-west, orange for northeast and, finally, black for authors affiliated to foreign universities.

3.3 Network metrics

Network metrics have great importance while investigating network representation, characterization and behavior. This section presents a summary of the key network measurements used in our analysis, which are discussed along the paper.

The order of G_y is the number of its vertices. The size of G_y is the number of its edges. The *degree* (k_i) of a vertex $i \in V_y$ is the number of edges incident to vertex i and the *degree distribution* $(P(k))$ expresses the fraction of vertices in the whole graph with degree k. The assortativity measures whether vertices of high degree tend to connect to vertices of high degree (assortative network) whereas the network is called disassortative when vertices of high degree tend to connect to vertices of low degree. A path connecting two vertices $i, j \in V_y$ is said to be minimal if there is no other path connecting i to j with less links. Accordingly, the *average path length* of G_y is the average number of links in all short-

est paths connecting all pairs of vertices in V_y. The graph *diameter* is the length of the longest shortest path between all pairs of vertices in V_y. The *clustering coefficient of a vertex i* is the ratio of the number of edges between neighbors of vertex i to the upper bound on the number of edges between them. For instance, given $i, j, k \in V_y$ and assuming that edges $(i, j), (i, k) \in E_y$, the clustering coefficient defines the probability that (j, k) also belongs to the set E_y. The *clustering coefficient of a graph* is the average value of the clustering coefficients of all vertices in G_y. The *betweenness centrality* of a vertex i is associated with an importance measure, based on the number of shortest paths between other pairs of vertices that include vertex i. The *closeness centrality* of a vertex i is defined as the inverse of *farness*, which in turn, is the sum of distances to all other nodes. *Homophily* is the tendency of people (in our case, researches) with similar features to interact with one another more than with people with dissimilar features. The indicator function $\mathbb{1}c_i = c_j$ assumes value 1 if the class c_i of node i is equal to the class c_j of node j, and 0 otherwise. Notice that the assortativity is the homophily when the class c_i of node i is its degree k_i. Table 1 summarizes the mathematical formulas for the main network metrics outlined above. Please refer to Costa et al. [10] for a complete review of measurements.

4 Statistics

In this section we present some statistics that give evidence of why SBRC is one of the most important scientific events in Computer Science in Brazil, with a growing community over the years. Figure 2 presents the aggregated number of distinct authors who published papers in SBRC, Fig. 2a, the aggregated number of distinct authors' affiliations, Fig. 2b, and also the aggregated number of published papers, Fig. 2c, over the years.[3] As can be observed, the number of new authors more than doubled between the years 2000 and 2012. The same increase also happened to the number of new universities and published papers. These results show that SBRC is attracting the participation of new researchers and new institutions over the years. Moreover, they clearly reflect the increase in the number of new graduate programs in Computer Science in Brazil, especially during the 2000s, as shown in Fig. 26 of the Appendix.

The previous results can be summarized in Fig. 3, that shows the SBRC network density over the years. The network density is calculated by dividing the number of edges by the number of nodes present in the graph. Observe that the density grows fast in the first years of the symposium, then

---

[3] The aggregated number of authors (universities and papers) for year y is the number of unique authors (universities and papers) in all years up to y. Henceforth, all aggregated results follow the same logic.

Table 1 Network metrics

Metric	Formula
Order	$N = |V_y|$
Size	$M = |E_y|$
Degree	$k_i = \sum_{j \in V_y} a_{ij}$
	where $a_{ij} = \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases}$
Degree distribution	$P(k) = \frac{n_k}{N}$
	where n_k is the number of vertices with degree k
Assortativity	$r = \dfrac{\frac{1}{M}\sum_{j>i} k_i k_j a_{ij} - [\frac{1}{M}\sum_{j>i}\frac{1}{2}(k_i+k_j)a_{ij}]^2}{\frac{1}{M}\sum_{j>i}\frac{1}{2}(k_i^2+k_j^2)a_{ij} - [\frac{1}{M}\sum_{j>i}\frac{1}{2}(k_i+k_j)a_{ij}]^2}$
Average path length	$L = \frac{1}{N(N-1)} \sum_{i,j \in V_y : i \neq j} d_{ij}$
	where d_{ij} is the distance between vertices i and j
Diameter	$D = \max\{d_{ij}\}, \ \forall i, j \in V_y, i \neq j$
Clustering coefficient of a vertex	$cc_i = \frac{2e_i}{r_i(r_i-1)}$
	where e_i is the number of edges between neighbors of i, and r_i is the number of neighbors of vertex i
Clustering coefficient of a graph	$CC = \frac{1}{N} \sum_{i \in V_y} cc_i$
Betweenness centrality of a vertex	$B_i = \sum_{s,t \in V_y : s \neq t} \frac{\sigma(s,i,t)}{\sigma(s,t)}, \ s \neq i, t \neq i$
	where $\sigma(s,i,t)$ is the number of shortest paths between vertices s and t that pass through vertex i and $\sigma(s,t)$ is the total number of shortest paths between s and t
Closeness centrality of a vertex	$C_i = \left[\sum_{j \in V_y} d_{ij} \right]^{-1}$
	where d_{ij} is the distance between vertices i and j
Homophily	$H = \frac{\sum_{\forall (i,j) \in E_y} \mathbb{1}_{[c_i = c_j]}}{2|E_y|}$
	where $\mathbb{1}_{[c_i = c_j]}$ is an indicator function that assumes value 1 if the class c_i of node i is equal to the class c_j of node j, and 0 otherwise

(a) Authors   **(b)** Institutions   **(c)** Papers
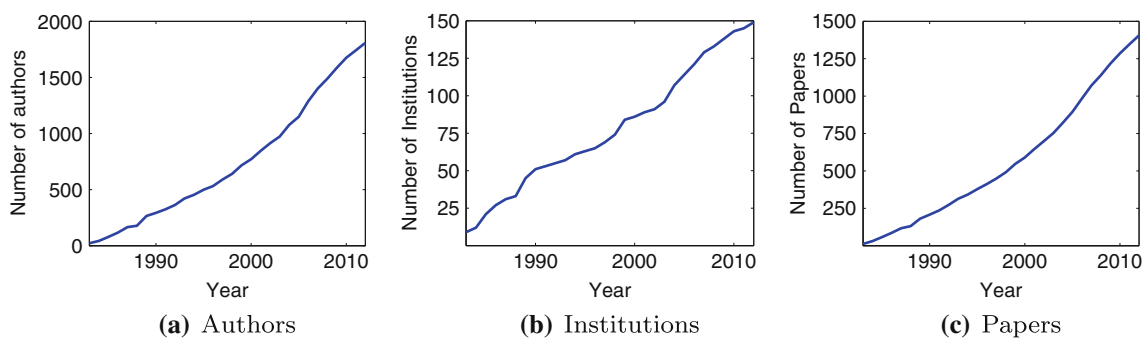
Fig. 2 Evolution of the number of distinct authors, institutions and published papers in all 30 editions of SBRC

it remained practically constant during the 1990s and grew again in the 2000s. Once more, this behavior is strongly correlated with the number of graduate programs in Computer Science in Brazil. In the 1990s, since the number of graduate programs remained practically constant and the means of communication were not as developed as in the 2000s, the papers were mostly composed either by repeated collaborators or by new authors, what explains the constant network density in this decade.

Figure 4 illustrates the participation of authors from different Brazilian states and regions in the symposium by showing the number of papers published with authors from each
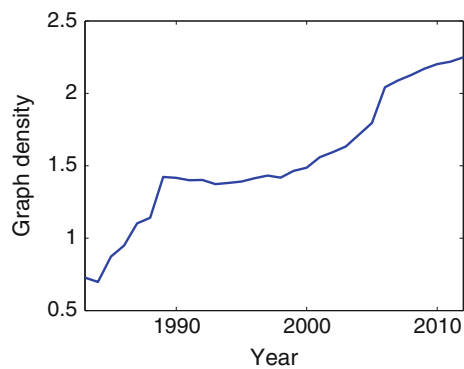
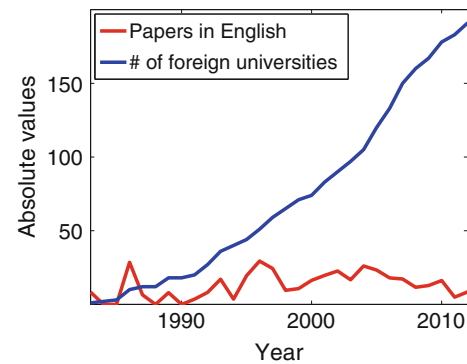**Fig. 3** The graph density (number of edges divided by the number of nodes) over time

Fig. 6 Aggregated evolution of collaborations with distinct foreign universities, and the number of papers written in English
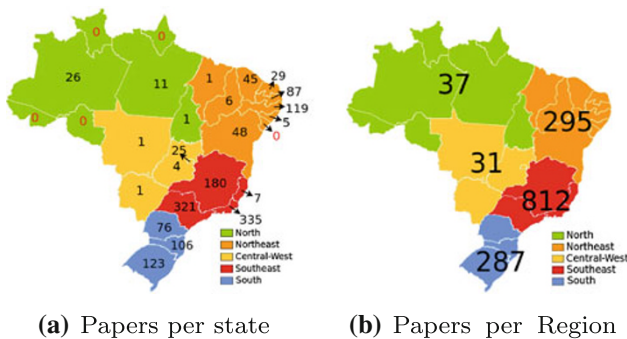

(a) Papers per state **(b)** Papers per Region

Fig. 4 Total number of publications with authors from each state and region

Fig. 5 Evolution of the number of publications per Brazilian's regions

of them belonging to the north region of Brazil, never published in SBRC. To better understand the participation of each region of Brazil in SBRC, Fig. 5 shows the evolution of the number of publications for each of the five regions. An interest fact in this figure is that it clearly reflects the evolution of the number of Computer Science graduate programs per region, as shown in Fig. 26 of Appendix. This shows that investments in educational initiatives, especially the opening of new graduate programs, leads to research advancements. These results also explain why the participation in SBRC is mostly concentrated in the northeast, southeast and south regions, while the north and center-west are under represented.

SBRC is a national symposium targeted at the Brazilian research community. However, the participation of authors with foreign affiliation is increasing over the years, as it can be observed in Fig. 6, which shows the aggregated number of foreign institutions with papers published in SBRC. In order to verify if such increase in the number of foreign institutions is solely a consequence of an increase in the number of foreign authors, Fig. 6 shows the number of papers published in English over the years. For our surprise, this number is actually decreasing in recent years. Intuitively, this result tells us that the number of active foreign authors publishing in SBRC is not increasing, but rather the number of Brazilian authors in foreign institutions is. This finding is consistent with Bazzan and Argenta [4], who suggest that more efforts are necessary to internationalize the Brazilian research community.

Finally, Tables 2 and 3 show the top 20 authors with the largest number of published papers from Brazilian and foreign institutions, respectively. Table 2 identifies several well-known researchers in the fields of computer networks and distributed systems. This is another indication of the paramount importance of the SBRC for the Brazilian community. Table 3 also identifies some Brazilian researchers with foreign affiliations at the time of publication. This reinforces the hypothesis that the number of active foreign authors publishing in SBRC is not increasing.

state (Fig. 4a), and from each of the five Brazilian regions (Fig. 4b). It is possible to see that the participation is mostly concentrated in the northeast, southeast and south regions of Brazil, summing up more than 95 % of the total published papers. Moreover, the top three states in numbers of papers (Rio de Janeiro, São Paulo and Minas Gerais) are in the same Brazilian region (southeast). Notice that of five states (Acre, Amapá, Rondônia, Roraima and Sergipe), four

Table 2 Top 20 Brazilian authors

Author	Number of publications
Antonio A. F. Loureiro	62
Otto C. M. B. Duarte	50
Nelson L. S. da Fonseca	50
José Ferreira de Rezende	46
Liane M. Rochenbach Tarouco	44
José Marcos Nogueira	40
Joni da Silva Fraga	38
Djamel Sadok	35
Edmundo R. M. Madeira	32
Paulo R. F. Cunha	32
Luci Pirmez	30
Luiz F. G. Soares	25
Lisandro Z. Granville	25
José A. S. Monteiro	25
Edmundo A. de S. e Silva	24
Maurício F. Magalhães	23
Jean-Marie Farines	23
Marinho P. Barcellos	23
Jussara Almeida	22
Luciano Paschoal Gaspary	20

Table 3 Top 20 foreign authors

Author	Number of publications
Guy Pujolle	11
Francisco Vasques	5
Alysson Neves Bessani	5
José Neuman de Souza	4
Serge Fdida	4
Aline Carneiro Viana	4
Emir Toktar	4
José Marcos Nogueira	4
Dominique Gaiti	3
Badri Nath	3
Luís Ferreira Pires	3
Pedro Braconnot Velloso	3
Azzedine Boukerche	3
Miguel Correia	3
Pierre de Saqui-Sannes	3
Don Towsley	3
Gregor von Bochmann	3
Jean-Pierre Courtiat	3
Marcelo Dias de Amorim	3
Michel Hurfin	2
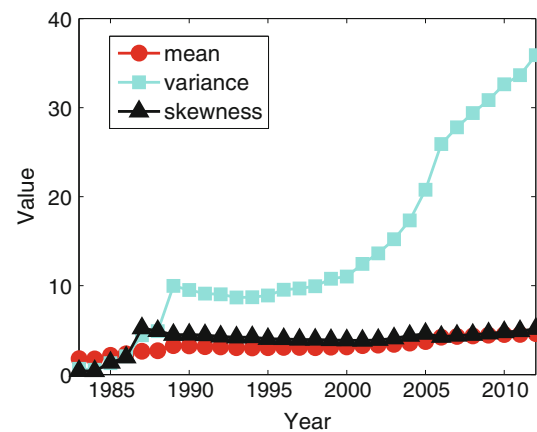
Fig. 7 First three moments of the degree distribution over the years

5 Collaborations

As stated before, an edge between two researchers indicates a scientific collaboration between them. Thus, the degree of a node i represents the number of collaborators of researcher i. The analysis of the node's degree in a collaboration network allows the assessment of the structure of co-authorship relationships among researchers in the communities of computer networks and distributed systems in Brazil.

Figure 7 shows the first three moments of the degree distribution over the years. We can observe that the average number of collaborations only increased from approximately 2 in the first year of the symposium to approximately 4 in the last year. However, both variance and skewness of the distribution are significantly large, indicating that a considerable number of researchers possess a high degree. Finally, we observe that the three moments of the distribution become reasonably steady in the late 1980's, and after that the network variance increases at the end of the 1990's.

Analyzing each year individually, we can observe that the node degree distribution is close to a power-law distribution [13], as shown in Fig. 8 for selected years. Mathematically, an amount x follows a power-law if it can be taken from a probability distribution $p(x) \propto x^{-\alpha}$, where α is a constant parameter known as exponent or scale parameter, and it is typically a value between $2 < \alpha < 3$. Graphically, α and $\alpha - 1$ represent the slopes of the lines that define the probability density function $\Pr(X = x)$ and the complementary cumulative distribution function $\Pr(X \geq x)$, respectively. The adjustments were made according to the method based on the maximum likelihood described in [9].

Figure 9 shows the evolution of the exponent α of the degree distributions over time. The points identified as "*biased fit*" represent biased fits and should not be considered good fits[4] [9]. It is worth noticing that there is a general trend

---

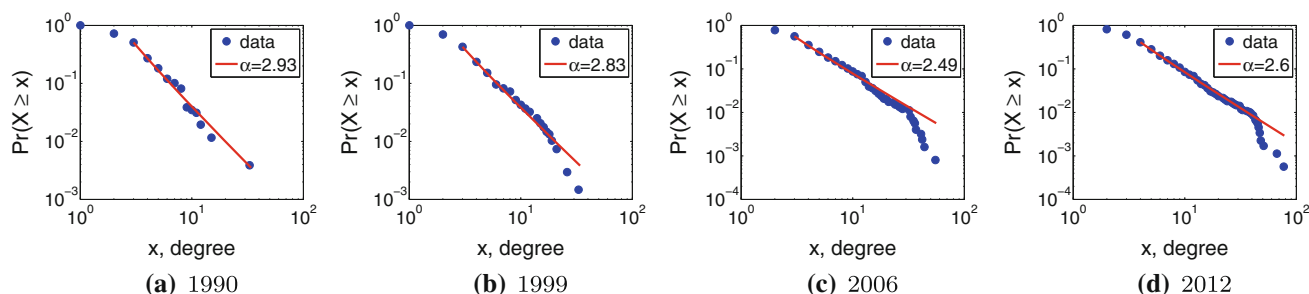[4] The number of points that can be explained by a power law is too small.

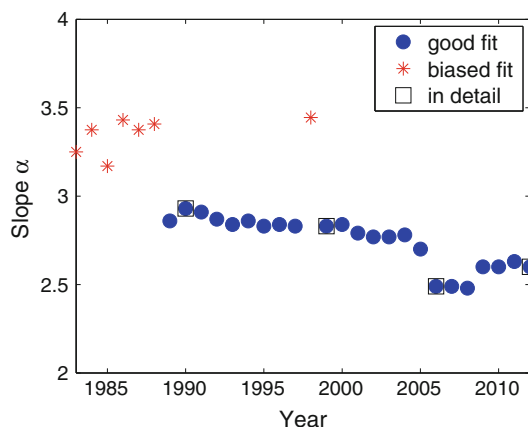**Fig. 8** Degree distribution at four specific years

Fig. 9 Slope α of the adjustment made in the degree distribution. We can observe the slope decreases over the years, which testifies the increase in the variance observed in Fig. 7. Points identified as *in detail* are the distributions presented in Fig. 8

towards α decreasing over the years, which indicates that the variance distribution increases as the number of nodes with a high degree in the network grows. For instance, in the first year of the SBRC network, all nodes have degrees of the first order of magnitude, i.e., lower than 10. In the last year, however, while several nodes have node degrees close to the third order of magnitude, the large majority still have degree lower than 10. This is an expected behavior in a collaborative network since, over time, researchers tend to consolidate and aggregate groups and communities that share the same interests. This shall be seen in more details hereafter.

An interesting way to identify the differences in the way senior researchers and newcomers connect among themselves is through a metric called $\langle k_{nn} \rangle_K$ [5], which indicates the average degree of neighbors of a given node with degree k. By using the $\langle k_{nn} \rangle_K$ metric, it is possible, for instance, to observe if high degree nodes tend to connect to each other or with low degree nodes. Figure 10 shows the function $\langle k_{nn} \rangle_K$ for four different years. While in 1989 there is a slight tendency of nodes with similar degrees to connect to each other (slightly increasing curve), in 1995 there is almost no correla-

tion (curve slightly negative). In 2003 and 2012, the tendency is to have high degree nodes connected to low degree nodes (descending curves).

In order to evaluate the behavior of $\langle k_{nn} \rangle_K$ over the years, the assortativity [34] is calculated for each network, over the years. The network assortativity measures the tendency of nodes with similar degrees to be connected. That is, in a *assortative network*, high degree nodes tend to connect with other high degree nodes, whereas in a *disassortative network*, high degree nodes tend to be connected to low degree nodes. The assortativity values range from -1, when the network is fully disassortative, to 1, when it is fully assortative. Figure 11 shows that the SBRC collaboration network becomes disassortative over the years. In 1983, the network is completely assortative due to the presence of cliques, i.e., each node is connected to nodes having the same degree. During the initial years, the network still presents an assortative feature, due to the large presence of isolated cliques or small connected components. However, from the end of the 1990s on, the network is consolidated as disassortative, where the tendency is that high degree nodes be connected to low degree nodes. This is the natural behavior in collaboration networks, as students or newcomers (low degree nodes) tend to connect with well-established and expert researchers (high degree nodes) to grow in their academic careers.

6 Connected components

In this section we show how the connected components of the network evolved over the years. Figure 12 shows the evolution of the number of network components. Notice that the increase in the number of network components is more significant during the first editions of the symposium. For instance, in 1983, the network had 11 components, while in 1989, after seven editions, the collaboration network had 78 components, an increase of more than 609 %. Thereafter, 21 editions later, in 2011, the network had 124 components, an increase of 58 % compared to 1989. This is explained by the fact that the collaborations among researchers in the early years of the conference were geographically constrained, i.e.,
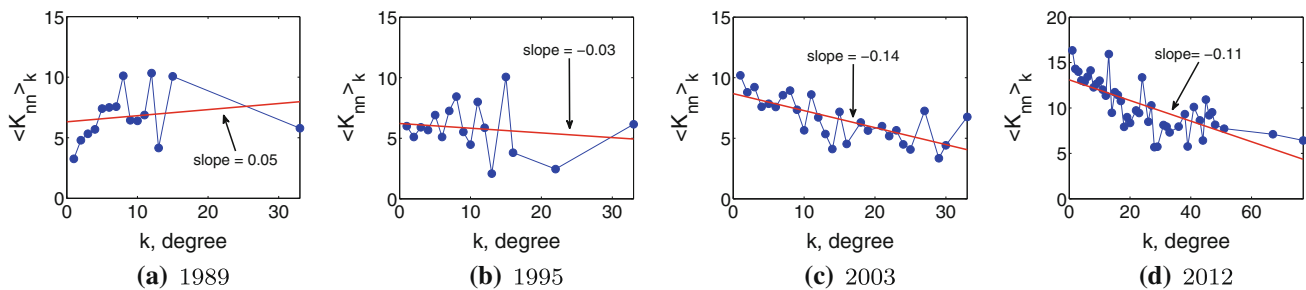
Fig. 10 Average $\langle k_{nn} \rangle_k$ degree of the neighbors of a given node with degree k
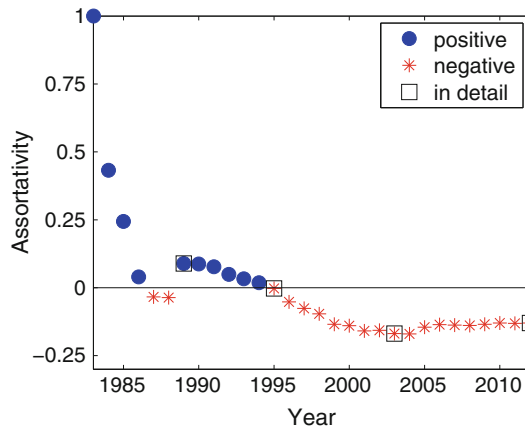

Fig. 11 The network assortativity over the years. It is possible to observe that the network becomes disassortative over the years, indicating that high degree nodes tend to be connected to low degree nodes. The behavior of $\langle k_{nn} \rangle_k$ for the networks, which is represented by dots marked as *in detail*, is shown in Fig. 10
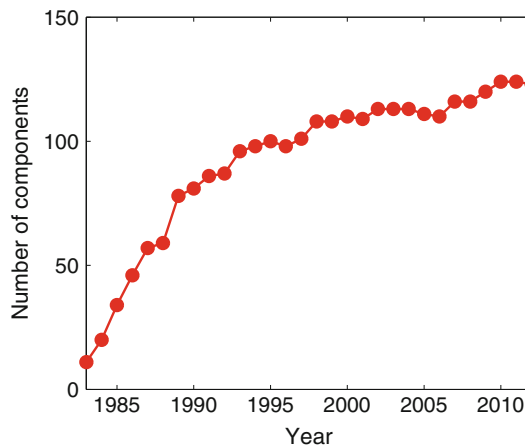

Fig. 12 Number of components

a collaboration between researchers of different institutions was rare. Recall from Fig. 2b of Sect. 4 that the number of new authors' affiliations more than doubled in the first seven editions of the event. Moreover, the means of communication in Brazil during this early period were not as developed. Therefore, collaborations among authors were restricted to

researchers working at the same institutions, leading to the creation of many network components or isolated groups of researchers (for a proper discussion of this fact, see Sect. 10 on homophily).

Table 4 shows the top five largest components for different years. We can observe that in the first editions of the symposium, the number of researches in each component was small, thus confirming the discussion above. In the first editions, essentially, each component was a representation of each published paper so far. In 1985 and 1986, we can observe the creation of research groups inside each university. This also reinforces the fact that in the first editions of the symposium the collaborations were geographically constrained. As the means of communication evolved during the mid-1990s and the number of graduate programs in Brazil started to increase, we can also observe an increase of the size of each component, since new collaborations among authors from different groups start to arise. From the last decade until today, we can also observe an increase in the size of the largest connected component. This happens because nowadays, the collaborations among researchers are not geographically constrained and the students from the 1980s and 1990s are, today, research leaders in different regions of Brazil with well-established communities (for a discussion on communities, see Sect. 8, and for a discussion on important nodes, see Sect. 9).

Figure 13 shows the evolution of the two largest connected components of the network. We can observe that, up to 1995, the largest connected component (LCC) and the second largest connected component (SLCC) represent about 21 and 10 % of the network, respectively. After 1995, the LCC increases over the years and the SLCC becomes steady until 2001, when it suddenly decreases considerably. This sudden decrease was caused by the previous SLCC merging with the LCC. An important issue when analyzing connected components is the collaboration between individual researchers. A collaboration which previously did not exist may drastically change the network structure.

To illustrate how important individual collaborations can impact the network structure, consider the year of 2001, when the SLCC merges with the LCC. This happened exclusively

Table 4 Top five largest components

Component	Size								
	1983	1984	1985	1986	1989	1995	2001	2007	2012
1	4	4	8	10	48	156	577	1,108	1,476
2	3	4	6	10	36	73	12	13	13
3	3	4	4	10	13	31	9	12	12
4	2	4	4	7	12	24	8	8	11
5	2	3	4	4	10	9	7	7	11

The number of components for each year is: 1983 (11), 1984 (20), 1985 (34), 1986 (46), 1989 (78), 1995 (100), 2001 (110), 2007 (116) and 2012 (123). Size is the number of researches inside the component
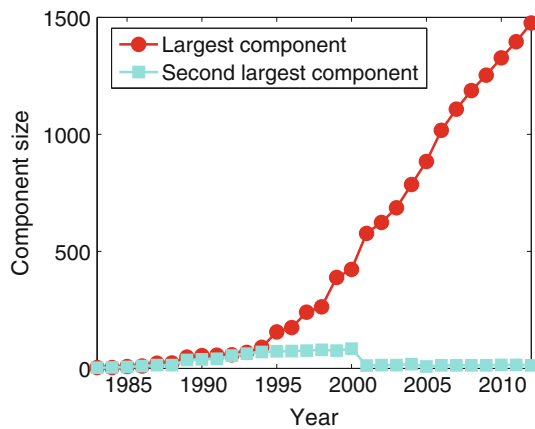
Fig. 13 Two largest components
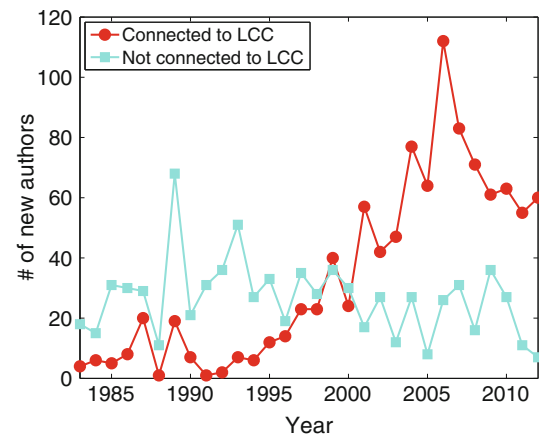
Fig. 14 Number of newcomers per year

because of the collaboration of two researchers from the SLCC with researchers belonging to the LCC. More specifically, in 2001, Michael A. Stanton, an author in the SLCC in 2000, was a co-author with Noemi de La Rocque Rodriguez, who belongs to the LCC in 2000. Similarly, also in 2001, José Neuman de Souza, who belongs to the SLCC in 2000, co-authored a paper with Nelson L. S. da Fonseca, who belongs to the LCC in 2000. These two collaborations illustrate a non-geographically constrained collaboration and a geographically constrained collaboration, respectively. For instance, in 2001, Michael A. Stanton was working at the Federal University Fluminense, located in Niterói, RJ, and Noemi de La Rocque Rodriguez was working at the Pontifical Catholic University of Rio de Janeiro, located in Rio de Janeiro, RJ. These two cities are about 20 km from one another. However, in 2001, José Neuman de Souza was working at the Federal University of Ceará, located in Fortaleza, CE, and Nelson L. S. da Fonseca was working at the State University of Campinas, located in Campinas, SP. These two cities are about 3,000 km far way from one another. It is important to notice that during the 2000s, collaborations like the one between Neuman and Fonseca start to become more common due to the many technological advancements in telecommunication and transportation, and also to the

expansion of Computer Science graduate programs in many regions of Brazil.

Figure 14 presents the number of newcomers to the symposium per year. Newcomers are the authors who are publishing in the SBRC for the first time. In Fig. 14, we classify them according to two categories: connected to the LCC and not connected to the LCC. Note that, in the early editions of the symposium, newcomers connected to the LCC are a minority, compared to the others. However, from 1995 on, the number of newcomers connected to the LCC starts to increase considerably, on a year basis, whereas the same is not observed for the number of newcomers not connected to the LCC. Indeed, from 2001 on, most of the newcomers are connected to the LCC. As the LCC becomes much larger than any other connected component starting at 1995, it is natural that the number of newcomers connected to it also increases from this year onwards. This result also corroborates the fact that until the mid-1990s, authors from the same paper would form a new connected component or connect to the smaller components already present in the network, thus leading to many isolated communities. However, from the mid 1990s onward, as new collaborations start to emerge, isolated components merge into one another, thus resulting in many larger communities.

7 Clustering and distance

The clustering coefficient (CC) and distance are important metrics to evaluate social networks. The clustering coefficient cc_i characterizes the density of connections close to vertex i. It measures the probability of two given neighbors of node i to be connected. The clustering coefficient of the network is the average cc_i, $\forall i \in V$.

Figure 15 shows the evolution of the network clustering coefficient and the clustering coefficient of the equivalent random network. The random network was generated using the model proposed in [3], that generates a random graph with the same number of vertices, edges and degree distribution. In the first edition of the symposium, in 1983, the clustering coefficient was 0.45. In that year, the authors had a CC equal to 0 or 1. A CC equal to 0 indicates that an article has one or two authors while a CC equal to 1 indicates that an article has three or more authors. In the first edition of the symposium, there were only collaborations among authors of the same article. In 1984, the CC of the network is significantly reduced, decreasing to 0.34. This is due to an increase in the number of authors with a CC equal to zero, i.e., articles with one or two authors. For instance, from the 27 authors in that edition, 20 authors have a CC equal to zero. In the most recent years, the CC tends to stabilize, due to an increase in the collaborations among authors. In 2012, the CC is 0.67, similar to other collaboration networks studied in the literature [18,33]. We also observe the SBRC clustering coefficient is, on average, one order of magnitude higher than the clustering coefficient of its equivalent random network (from late 1980s).

An important construction of social networks is the small-world network concept [44]. It is characterized by having a clustering coefficient significantly higher than the one of its equivalent random network and an average shortest path length (SP) as low as the one of the equivalent random network. The SP measures the average shortest distance (in
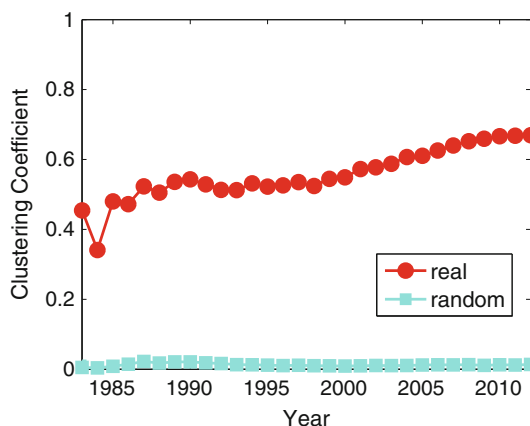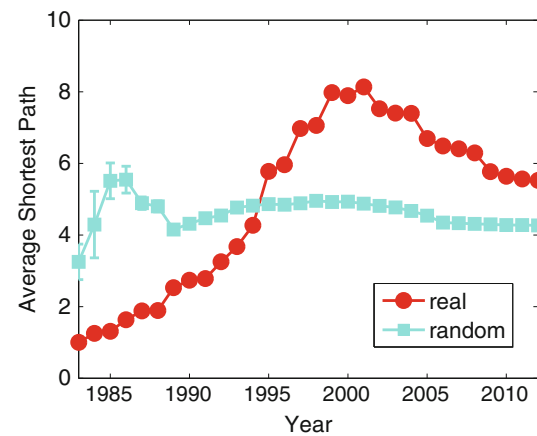

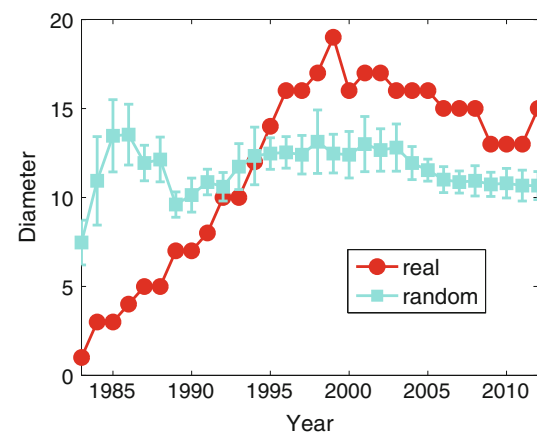**Fig. 16** Average shortest path over the years

Fig. 17 Network diameter over the years

hops) between every pair of nodes in the network. Figure 16 shows the evolution of the average shortest path of the historical SBRC network in comparison to the average shortest path of the random network. We observe that the SBRC network SP increases until the late 1990's, when the SP starts to decrease. This can be attributed to the advancements in telecommunications and technology as well as the creation of graduates programs, which resulted in an increase of the collaboration among researchers from different groups. During the last editions of the SBRC, the SP of the SBRC is 1.29 times greater compared to the random network. The high clustering coefficient, combined with the small shortest path, characterizes the SBRC network as a small-world network. In 2012, the average shortest path between authors was around 5.5, which follows the six degree of separation theory [42]. As a practical consequence, the short paths between SBRC researchers means that new hot topics on computer networks and distributed systems may propagate quickly among SBRC researchers.

The behavior of the network diameter is illustrated in Fig. 17. The network diameter measures the largest shortest

Fig. 15 Clustering coefficient over the years

path in the network. In the first two decades, the shortest paths among researchers increase, which leads to an increase in the network diameter. However, after 1999, due to an increase in new collaborations among authors and the network densification (see Fig. 3 of Sect. 4), the network diameter starts to decrease. In 1999, it was 19 hops, but diminished to 15 hops in 2012.

8 Communities

One of the most relevant characteristics of graphs representing real systems is the structure of communities, i.e., the organization of vertices into clusters, with many edges between the vertices of the same cluster and relatively few edges connecting vertices of different clusters. In order to identify communities in the collaboration network, we used the k-clique community identification algorithm. A community is defined as the union of all cliques of size k that can be achieved through adjacent k-cliques (two k-cliques are considered adjacent if they share $k - 1$ vertices). In other words, a k-clique community is the largest connected subgraph obtained by the union of a k-clique and of all k-cliques which are connected to it. The implementation of this algorithm was based on Palla et al. [37].

Our main goal is to evaluate how distributed and clustered are the collaborations among authors in the SBRC network. This justifies the choice for the k-clique community algorithm, since it is a good measure to select sub-communities and also overlapping communities [14]. In order to achieve our goal, we use the lowest bound value of $k = 3$, since it is the most favorable value to capture the largest group of

authors (largest connected sub-graph) that forms a community, according to the algorithm specification. When executing the k-clique community algorithm with $k = 3$, assuming a network with high collaboration between nodes, it is expected to find very few communities. However, as discussed hereafter, this is not the case for the SBRC network.

8.1 View of communities

In this section we present two visualizations of communities: one observing the university the author has worked for and the other observing the state in which this university is located (a more detailed discussion about communities shall be presented in Sect. 8.2). Thus, each node in the network is associated with one or more states and universities, given that an author may be affiliated to more than one university during his career. Figure 18 presents a view of the four largest communities by Brazilian states, while Fig. 19 shows the four largest communities by university. These communities have 182, 87, 79 and 69 authors, respectively. In both figures, the size of the word indicates its popularity within the community. This means that in the largest identified community, shown in Fig. 18, the states of Rio de Janeiro (RJ) and Rio Grande do Sul (RS) are the most representative ones. It is worth noting that the word *FOREIGN* represents researchers from institutions located outside Brazil.

After executing the k-clique community algorithm (with $k = 3$), we would expect to find a small number of communities. But, as we can see, we identified many different communities. Obviously, with higher values of k we find communities that have authors more connected among them-
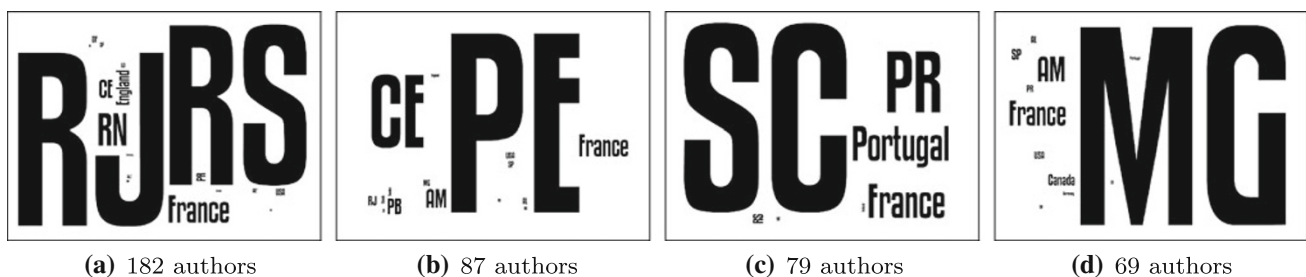

| (a) 182 authors | (b) 87 authors | (c) 79 authors | (d) 69 authors |

Fig. 18 The four largest communities: visualization by state

| (a) 182 authors | (b) 87 authors | (c) 79 authors | (d) 69 authors |

Fig. 19 The four largest communities: visualization by university

selves. Considering $k = 4$, for example, the largest, second largest, and third largest communities have 42, 39, 31 authors, respectively. If we consider $k = 5$, the number of authors in the largest, second largest, and third largest communities drops to 16, 16, 15, respectively.

A value of $k = 3$ is particularly interesting for visualizing the general interaction among the authors of the SBRC network, but on the other hand this may not find very strong communities. This is what happened for the community consisting mainly of authors from RS and RJ (largest 3-clique community). After a closer look, we can see that the number of collaborations between these groups of authors is not as large as the number of collaborations within the groups. For instance, when we execute the algorithm considering $k = 4$, we notice that this community is divided into two communities, one formed mostly by authors from RS, and the other formed by authors from RJ. This shows that RJ and RS together as the largest 3-clique community do not represent a very strong connected community.

In general, we observe that most of the interactions tend to happen among authors from particular regions and institutions. This information might be particularly interesting to support decisions towards the improvement of collaborations among researchers from different universities and regions of Brazil.

8.2 Community evolution over time

In this section we present a more detailed analysis of the identified communities. Figure 20a shows the number of communities over the years. We can see that the number of communities increases over time, reaching more than 250 communities in 2012. The choice of $k = 3$ also has implications in this result. For the SBRC network, higher values of k may imply into a smaller number of communities. For example, three authors of the same paper, that published just this paper in the entire history of SBRC, are considered a community when using $k = 3$, but not when using $k = 4$.

Figure 20b shows the cumulative distribution function (CDF) of the number of authors in the communities, considering the years of 1983, 1993, 2003, and 2012. A high number of communities, as observed in Fig. 20a, does not mean that there are many authors in all these communities. Figure 20b shows that communities with a small number of authors represent a considerable subset of all communities. Around 90 % of all communities have less than 10 authors, and approximately 55 % have only three authors. However, we can notice that over the years, due to an increase in the number of collaborations, communities with a higher number of authors start to arise. For example, in 1983 the largest community had only four authors, whereas in 2012 six communities had more than 30 authors.
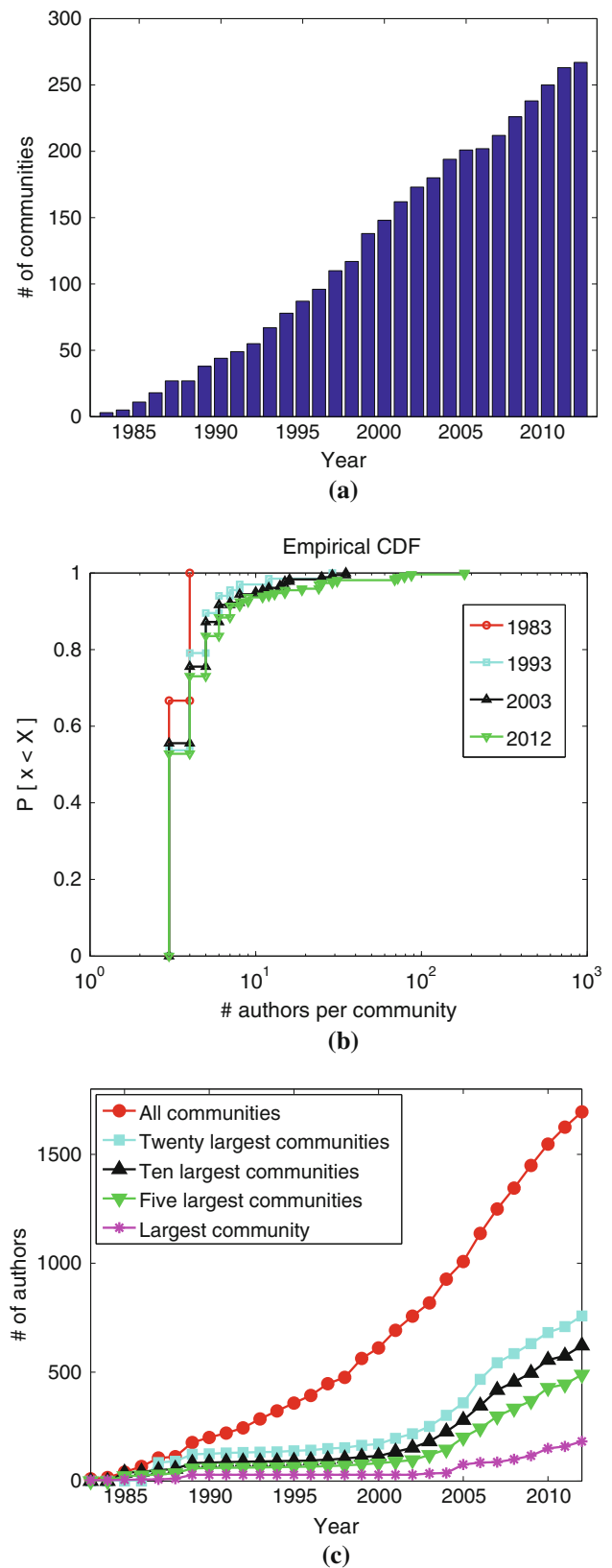

Fig. 20 Communities. **a** Number of communities per year. **b** CDF of the number of authors in communities of the years 1983, 1993, 2003 and 2012. **c** Number of authors in specific communities

Figure 20c shows the number of authors over the years for the following groups of communities: all communities, 20, 10 and 5 largest communities, and the largest community. We observe that from 2004 to 2012 the number of authors per community increases considerably. As stated before, such increase is due to the growth of a few communities with a large number of authors. In this way, we observe that in 2004, the 5 largest communities represent approximately 64 % of the top 10 communities and approximately 48 % of the top 20 largest communities. Considering the year 2012, these values are 79 and 65 %, respectively. We also observe that the top 5 communities represent a significant amount (29 %) of all considered authors. This result indicates that authors in the largest communities interact with researchers outside their communities, thus increasing it over time.

Finally, someone may attribute the change in the communities dynamics during the 2000s, as shown in Fig. 20, to the merge of the LCC and SLCC in 2001, as previously described in Sect. 6. However, this event alone does not totally explain such a change. It is worth noting that it is during the 2000s that significant historical events start to happen in Brazil (see Appendix). For instance, we can outline the developments in the telecommunications and transportation sectors. Moreover, Brazil witnessed a rapid growing in the number of Computer Science graduate programs all over the country. Therefore, we can conclude that the combination of these events changed the way researchers used to collaborate, thus better explaining the change in the communities dynamics during this decade.

9 Important nodes

The identification of important nodes within a social network structure is a common activity in SNA. Usually, the identification of such nodes is performed by using centrality metrics, such as the closeness and betweenness [6]. These metrics aim to identify nodes that possess strategic locations within the social network structure. A strategic location may indicate that a node has a high influence over other nodes, or it hold the attention of nodes whose positions are not as convenient in the social context.

The main idea behind the closeness centrality metric is to show how close a node is to all other nodes in the network, i.e., how many edges separate a node from other nodes. On the other hand, the main idea behind the betweenness centrality is to show how often a node is in the shortest path between any two other nodes. In the perspective of a co-authorship network, the closeness centrality may indicate the authors with a favorable location in the network structure to start the dissemination of new scientific findings or research directions to the whole network. For instance, if an author with a high closeness disseminates a new scientific finding, the probability for this new finding reaching the whole network

Table 5 Top 10 betweenness authors

Name	Betweenness
José Neuman de Souza	0.186
Nelson L. S. da Fonseca	0.124
Paulo Roberto Freire Cunha	0.109
José Ferreira de Rezende	0.095
Maurício Ferreira Magalhães	0.086
Marcos Rogério Salvador	0.077
José Marcos Nogueira	0.070
Artur Ziviani	0.065
Liane M. R. Tarouco	0.064
Luci Pirmez	0.055

Table 6 Top 10 closeness authors

Name	Closeness
José Neuman de Souza	0.243
Nelson L. S. da Fonseca	0.229
José Ferreira de Rezende	0.223
Luci Pirmez	0.221
Jorge Luiz de Castro e Silva	0.216
Paulo Roberto Freire Cunha	0.215
Alexandre Lages	0.214
Elias Procópio Duarte Jr.	0.212
Flávia Coimbra Delicato	0.212
Rossana M. C. Andrade	0.211

in the least amount of time is higher than if the dissemination started at an author with a lower closeness.

In the case of the betweenness centrality, it may indicate the most efficient authors to act as bridges to carry information among different authors or communities. For instance, if an author has a high betweenness, the probability that a given piece of information being disseminated passes through this researcher is higher than for an author with a lower betweenness. Therefore, we hope that these metrics are able to identify not only strategically located authors in the co-authorship network, but also distinguished researchers in the scientific community of computer networks and distributed systems.

Table 5 shows the top 10 authors with the largest betweenness values, and Table 6 shows the top 10 authors with the largest closeness values. Indeed, we can note by looking at both tables that authors identified by both metrics are researchers that are widely known within the SBRC community, and even within the international scientific community. Conversely, we also can note that some prolific authors (e.g., Antonio A. F. Loureiro and Otto C. M. B. Duarte shown in Table 2) are not listed by both tables. Hence, one may wonder whether these metrics are actually accurate in capturing influential authors in the co-authorship network and also dis-

tinguished researchers. On the other hand, these authors may have a high impact in their research field but a not as high impact considering the interaction among research topics.

For instance, the researcher *Alexandre Lages* is in the top 10 authors for the closeness, but this author has only four publications in the SBRC and his last work was in 2007. However, a careful analysis of the collaborations of this author explains why such a fact occurs. It also highlights that the importance of an author in the co-authorship network, as identified by the centrality metrics, is strongly influenced by the pattern of his collaborations. That is, despite Lages' small number of publications, they were in collaboration with very influential and central authors. For instance, in 2004, Lages' work has as collaborators the following influential authors: *Flávia Coimbra Delicato* (16 publications in SBRC), *Luci Pirmez* (30 publications in SBRC) and *José Ferreira de Rezende* (46 publications in SBRC). Lages also has collaborations with *José Neuman de Souza* (17 publications in SBRC), *Lisandro Granville Zambenedetti*, (25 publications in SBRC) and *Liane Margarida Rochenbach Tarouco* (44 publications in SBRC). It can be observed that these authors are identified by one or both metrics as influential within the SBRC community (despite the author *Lisandro Granville Zambenedetti* does not appear in both tables, he is in the top 20 for both centrality metrics).

From this result we can conclude that when an author collaborates with central authors with a high closeness, then this researcher also increases his own closeness to all other authors in the network. For instance, in 2004, when Lages published together with *José Ferreira de Rezende*, his distance to *Otto C. M. B. Duarte* went from *not possible to reach* to two edges. Therefore, a collaboration with a central author made Lages closer to another author that was not his direct collaborator. Notice that the same may also happen to the betweenness, i.e., when two or more authors publish a paper together, these authors may create a new "bridge" connecting different groups of researchers, thus increasing the betweenness for these authors.

Looking at Tables 5 and 6 in this section and Table 2 in Sect. 4, we can notice two interesting facts. First, the top two publishers in SBRC, *Antonio A. F. Loureiro* and *Otto C. M. B. Duarte*, do not appear in the top 10 of both centrality metrics. Second, an author that is not in the top 30 publishers in SBRC, *José Neuman de Souza* (17 publications in SBRC), is the most central author according to both centrality metrics. For instance, if we look into the history of both Loureiro and Souza we can notice similar aspects. They are constantly publishing in SBRC since 1995, they appear in almost the same number of communities (Loureiro appears in 7, while Souza in 6), they collaborate with almost the same number of universities (Loureiro has collaborators in 14 universities, while Souza has collaborators in 15) and also states (Loureiro has collaborators in 11 states, while Souza has in 10).

However, once again, a careful analysis of the collaboration of these authors might explain why such facts occur. Using the same $\langle k_{nn} \rangle$ metric as in Sect. 5 we find that the average degree of Loureiro's collaborators is 6.42, while for Souza it is 14.28. Therefore, we can assume that while Loureiro usually publishes with his students, Souza usually publishes with senior researchers, probably acting as a "bridge" among prominent groups within the SBRC community. In particular, Souza is a collaborator to 5 authors in the top 10 betweenness and to 8 authors in the top 10 closeness. As an experiment, let us assume that Loureiro and Souza published a paper together at some point in the history of SBRC, resulting in an edge between the two authors. By adding this single collaboration, Loureiro goes from the 51st largest closeness in the network to the 13th largest closeness. Considering the betweenness, Loureiro goes from the 11th largest betweenness to the 6th largest betweenness. Actually, Loureiro's betweenness suffers an increase of about 60 %. Therefore, we can conclude that in a co-authorship collaboration network, the number of publications alone does not dictate the importance of an author within the community, but rather the pattern of his collaborations.

Furthermore, it is important to notice that centrality metrics are important tools in identifying strategic nodes in a network structure. Nevertheless, these metrics alone do not hold the final word on which nodes are actually important or not. For instance, we showed that using these metrics alone we were able to identify a central author that, apparently, is not active in the community anymore, and also active and prolific authors that are not considered as central authors.

Figure 21a and b show the evolution of the betweenness and closeness over the years for authors owning the five largest values in all SBRC history, as previously presented in Tables 5 and 6. For both metrics, the authors alternate their positions for the highest value throughout the years. For instance, *Maurício F. Magelhães*, *Paulo R. F. Cunha*, *Nelson L. S. da Fonseca* and *José Neuman de Souza* had the largest value of betweenness in different years, with the latter holding the top position since 2004. Notice that the values of closeness follow a similar behavior, which is mainly due to the arrival of new authors in the network and the emergence of new collaborations, especially after 1995. In particular, we can see that both metrics drastically increased in 2001 for the authors *José Neuman de Souza* and *Nelson L. S. da Fonseca* due to a new collaboration between them. Recall from Sect. 6 that this collaboration was responsible for merging the two largest connected components at the time. Figure 21c shows the degree evolution for the five researchers with the highest degrees in the network. It is worth noticing that four of the five researchers have little collaborations until 1995, but then experience a dramatic increase in their degrees afterwards.

Figure 22a, b show the first three moments of the betweenness and closeness distributions. Regarding the betweenness,
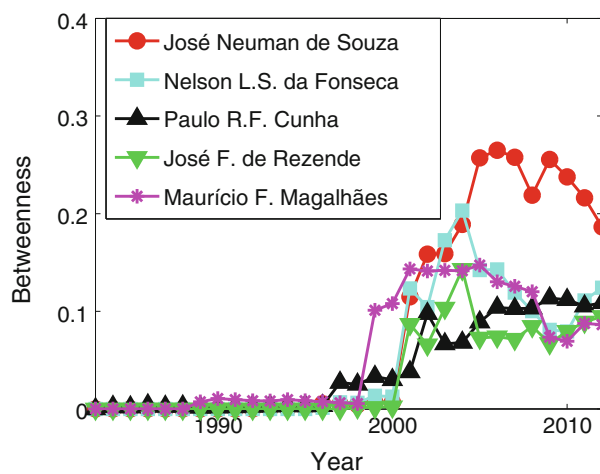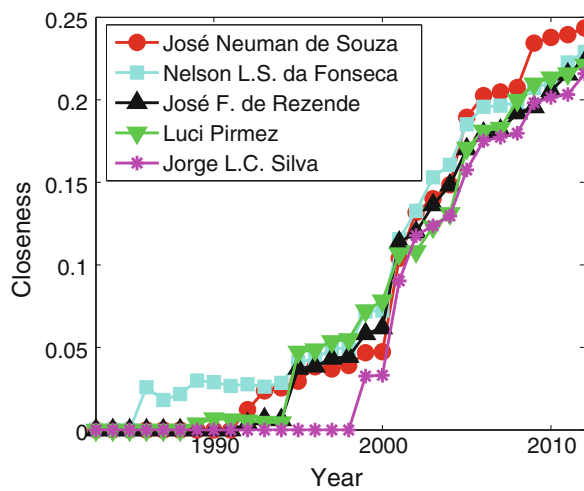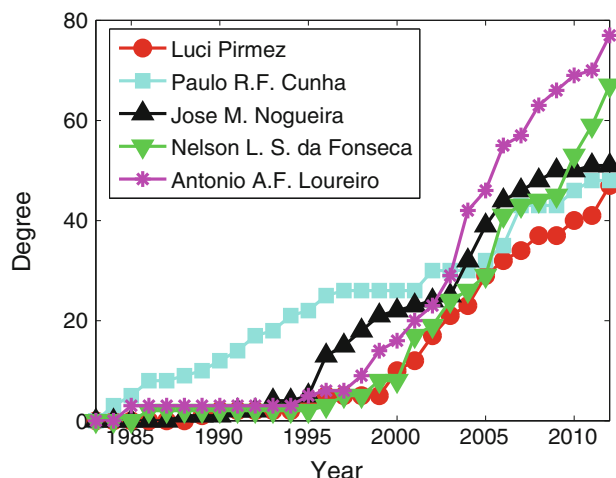
(a) *Betweenness*

(b) *Closeness*

(c) Degree

Fig. 21 Evolution of the betweenness, closeness and degree over the years for the top-5 prominent authors
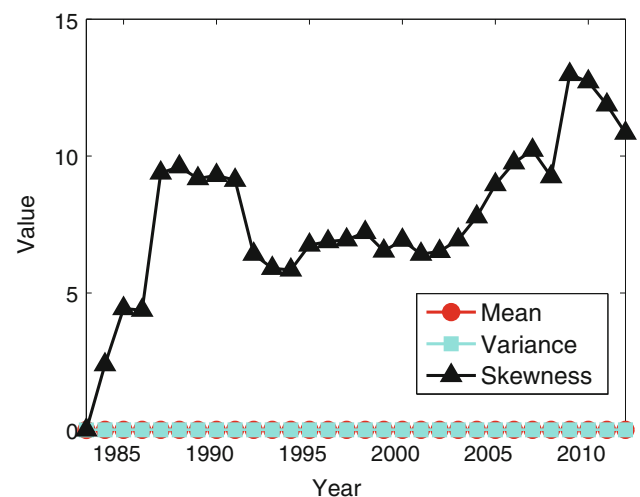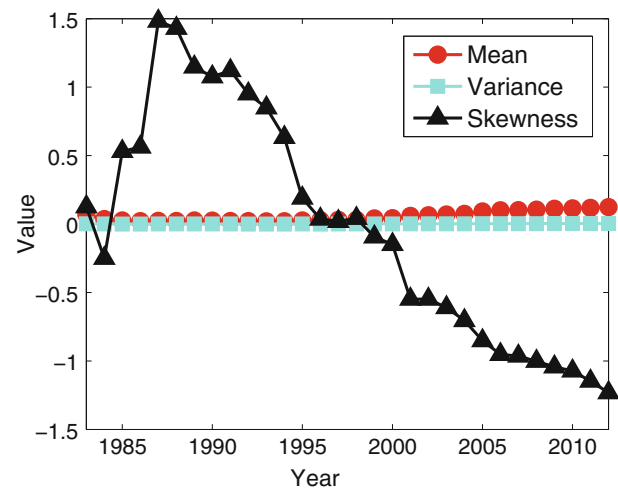

(a) *Betweenness*

(b) *Closeness*

Fig. 22 The first three moments of the betweenness and closeness distributions

the skewness value remains at 1, indicating that the betweenness distribution follows a power-law distribution. For the closeness, a signal change is observed for the skewness, indicating a move in the closeness distribution. During late 1980s and early 1990s, there is a small number of authors with high closeness values. Around 1997 there is a balance, and, in 2012, there is a high number of authors with high closeness values. The main observations we can draw from these results are that the SBRC network has a small set of highly influential nodes. Moreover, these nodes can easily spread information to all nodes in the network, due to the "proximity" among nodes in the network. Indeed, such characteristic is very desirable for a scientific network, especially if we consider the easiness in spreading new research directions or research findings.

10 Homophily and its impact

In SNA, the homophily principle states that similar nodes are more likely to connect than non-similar ones [24]. Consider similar nodes that share, for instance, the same gender, age, social status, religion, education, geographic location, and other types of attributes. Homophily has powerful implications in our world, limiting the information people receive, the attitudes they take, and the interactions they experience [27]. Thus, in this section we analyze homophily in the SBRC network, using the geographic location of the corresponding author as the node attribute that determines similarity, i.e., the state where the author's institution is located. It is natural to think that researchers who are geographically closer are more likely to publish together. However, here we also show the impact of this expected geographic segregation in the spread of research information in a large country as Brazil.

The calculation of the network homophily we use here is very intuitive. Consider a node i and its class c_i, which, in the present case, can be its geographical region (e.g. southeast), state or university. The homophily of the network $G(V, E)$ is calculated in the following as

$$\text{Homophily} = \frac{\sum_{\forall (i,j) \in E} \mathbb{1}_{c_i = c_j}}{2|E|}, \tag{1}$$

where $\mathbb{1}_{[c_i = c_j]}$ is an indicator function that assumes value 1 if the class, or state, c_i of node i is equal to the class c_j of

node j, and 0 otherwise. In other words, the homophily is calculated by counting the number of edges between collaborators of the same state and dividing it by the total number of edges.

In Figs. 23a–c, we show the evolution of the homophily in the SBRC network. We show homophily results computed yearly, i.e., computed considering the papers published during each edition of the event, as well as homophily results computed over the aggregated network built from all publications up to a given year. In the first year, the network homophily is 1 for all node classes (regions, states and universities), indicating that researchers only collaborated with others from the same university. However, the aggregated homophily drops very sharply in the first 4 years for all four node classes, with a smooth decay in the following years. Considering the yearly homophily measures, we note a very irregular behavior, with peaks and valleys, although, in general, a decreasing trend can be noticed. Finally, observe that the homophily in general grows when the granularity of the node class moves from "university" to "region", indicating that the geographic aspect plays an important role in the formation of collaborations.

After verifying that homophily decreases over time in the SBRC network, a natural step is to analyze if it brings any impact to research. As we have seen previously, the distribution of publications among the states is concentrated into a few states. However, recently a few states, which were completely inactive, showed a small but significant
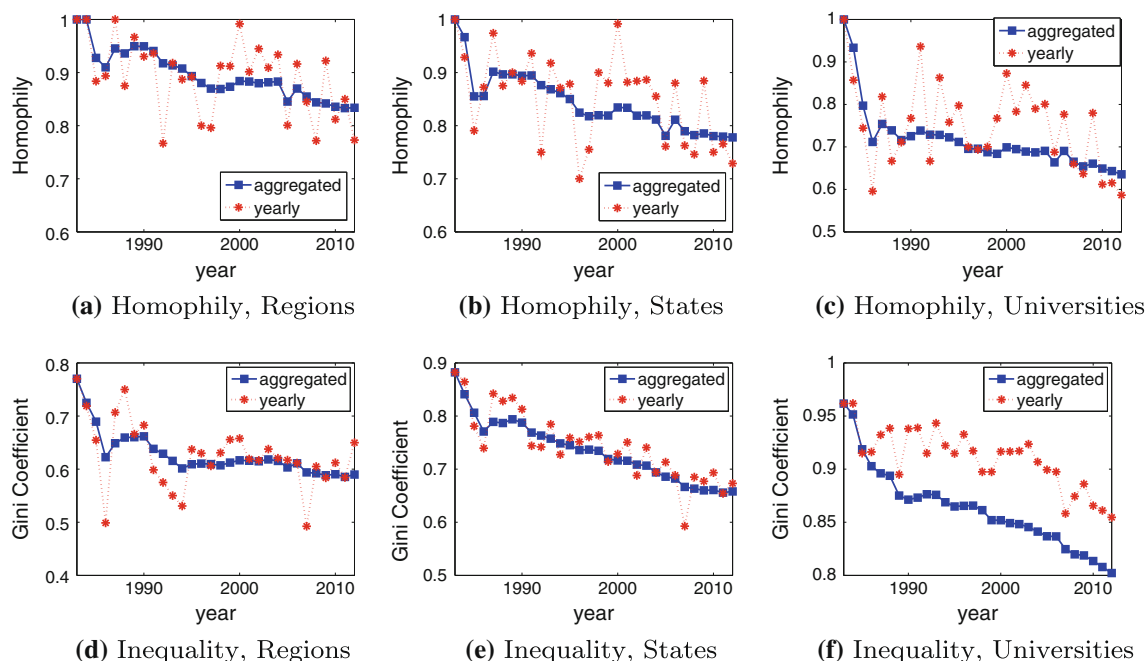

Fig. 23 The growing rate of collaborations between researchers of different backgrounds, and the decrease in the network's homophily, occurs together with a more evenly distribution of publications among the regions, states and universities of Brazil

progress. For instance, the state of Pará had only two publications in the first 20 years of SBRC, in the years of 1997 and 1998. In the last 10 years, researchers from the state of Pará had published a total of nine papers in six distinct years.

In order to formalize this, we use the Gini coefficient [8,16] to measure the inequality in the number of publications over the regions, states and universities of Brazil. The Gini coefficient was initially proposed to describe the income inequality in a population, commonly between countries and within countries [8,16]. It has found application in the study of inequalities in several other disciplines [39] and here we apply it to measure how the publications are distributed among the states of Brazil. It assumes values from 0, which expresses perfect equality, where all values are the same, to 1, which expresses maximal inequality among values, where all publications are concentrated in a single state.

In Figs. 23d–f, we show the Gini coefficient for the SBRC network computed on an year basis as well as over the aggregated network, considering the distribution of the publications among the geographical regions, states and universities. Like the homophily, observe the Gini coefficient decreases over the years, indicating that the distribution of the number of publications is becoming more equal. In fact, it decreases practically at the same rate as the homophily decreases. The Pearson's correlation coefficient between the homophily and the Gini coefficient in the SBRC network is 0.90 and, yearly, 0.45, among regions; 0.95 and, yearly, 0.54, among states; and 0.92 and, yearly, 0.70, among universities. This fact strongly suggests that the increase in the collaborations between researchers from different backgrounds significantly contributes to diminishing the inequality in the number of publications in Brazil, indicating that the network is becoming more heterogeneous.
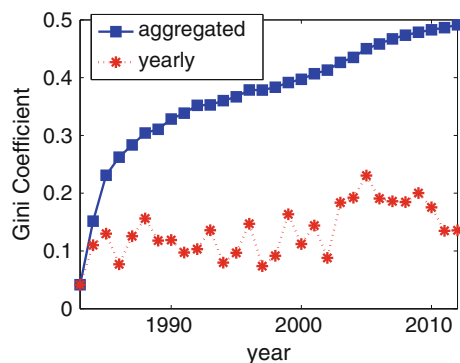

Fig. 24 The inequality concerning publications in SBRC among researchers in Brazil is increasing over time, probably due to the "rich get richer" phenomenon

Although researchers from different parts of Brazil are publishing more and more in SBRC, when we compute the Gini coefficient considering the number of publications per author, instead of per locality, we see that the inequality is increasing. First, observe in Fig. 24 that the Gini coefficient of the yearly SBRC network is considerably low through the 30 years of the symposium, varying from ≈ 0.04 to ≈ 0.23 and having mean 0.13. This suggests that researchers publish equally in SBRC each year. However, observe in Fig. 24 that the Gini coefficient of the aggregated SBRC network grows approximately 0.5 points in 30 years. Conversely, this suggests that while new researchers are constantly publishing in SBRC every year, there is also a group of researchers who are always publishing in the conference, increasing significantly their number of publications compared to the others. This, in fact, is not a surprise, since it is common to have in social networks a few "super nodes" while the majority are "ordinary" nodes, a consequence of the "rich gets richer" phenomenon.

This conclusion shows the importance of inter-state and inter-country collaboration programs, such as the recently created "Ciência Sem Fronteiras"[5] Brazilian program and the creation of graduate programs over the years. Such programs and other incentive mechanisms allow that regions with low research activity develop, mirroring their more productive partners.

11 Cross analysis

In Fig. 25, we show the Pearson's correlation coefficient between network metrics, namely the degree k, clustering coefficient cc, betweenness B and closeness C centralities, and the number of papers p an author had published. We consider three snapshots of the SBRC network that divide time into three periods of 10 years. First, we observe that for some metrics, the correlation changes over the years, while for others, it remains constant. Note that the correlation between the number of papers published p and the network metrics degree k and the betweenness centrality B grows over time. In the year of 2012, for instance, the correlation between the degree and the number of papers published is 0.89, a very high correlation that strongly corroborates with the fact that the "rich gets richer" phenomenon is present in co-authorship networks, since high degree nodes tend to "attract" a higher number of publications. On the other hand, observe that the clustering coefficient is always negatively correlated (−0.27 in 2012) with the number of papers published. This indicates that researchers who do not expand their collaborations, i.e., whose circle of collaborations remains constant over the years, tend to publish less in the SBRC.
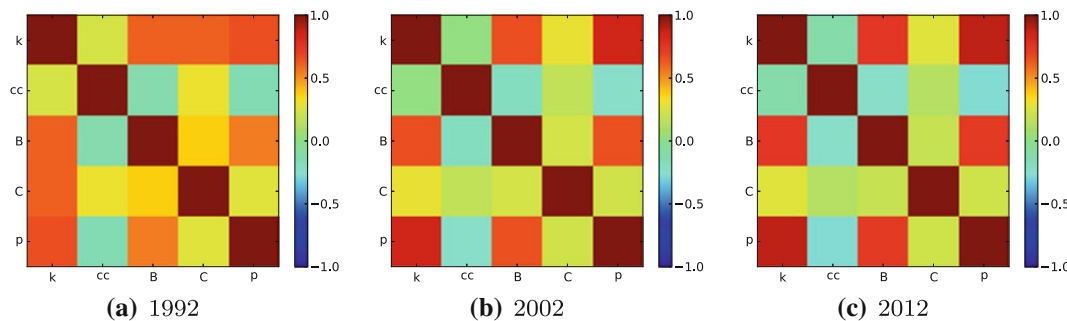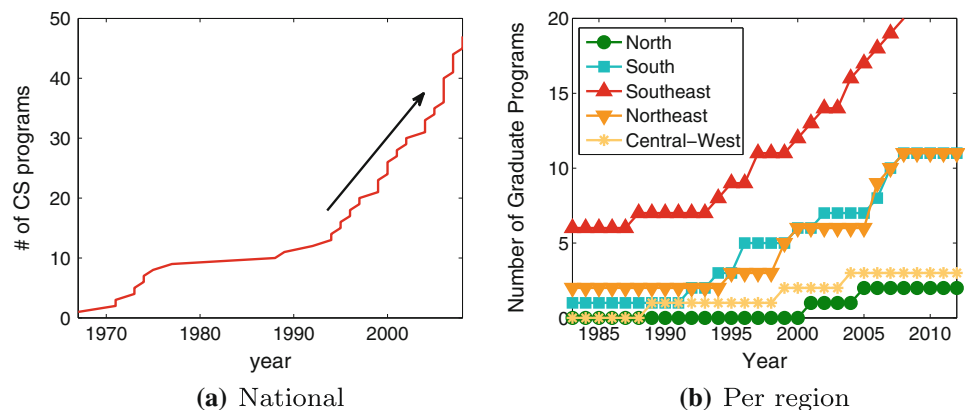

Fig. 25 The Pearson's correlation coefficient between network metrics (degree k, clustering coefficient cc, betweenness B and closeness C centralities) and the number of papers p an author had published

Fig. 26 Number of CS graduate programs in Brazil over the years

(a) National

(b) Per region

Finally, it is interesting to observe that the closeness centrality is becoming a more independent feature over the years, having in 2012 very small correlations with all the other metrics.

12 Conclusions

In this paper we made an analysis of the collaboration network between authors who have published in the editions of the Brazilian Symposium on Computer Networks and Distributed Systems. From this analysis, we have shown why the symposium is so relevant for the Brazilian research community and the regions with the highest number of participations. Moreover, we showed that the main kind of co-authorship is between well-established authors and newcomers to the symposium, which represents the natural kind of co-authorship between student and advisor. The most prominent communities were presented in two visualizations, one by universities and another by the Brazilian states. Furthermore, we identified the researchers who have a strategic position within the collaboration network and, thus, the power to influence others. Finally, we presented some Brazilian historical aspects that may have had a great impact on the symposium success, by allowing the collaboration of geographically distant researchers, thus strengthening the creation and

establishment of new communities. As future work, it would be interesting to analyse other Brazilian Symposiums, such as the SBBD, SBES and SIBGRAPI. By analysing these communities at the same level of detail as the study here performed, it would be possible to draw a bigger picture of the research community in Computer Science in Brazil.

Appendix: Historical aspects

To better understand the collaboration network of the SBRC, it is important to point out some relevant events and technological advancements in the Brazilian history. For instance, the development of the Internet in the late 1980s and its public availability in 1995, the privatization of the telephony sector in the late 1990's and the increase on the number of graduate programs during the 2000s. Therefore, in the next few paragraphs we outline the main events that may have contributed to foster the research development in Brazil, with a special focus in the computer networks and distributed systems communities comprising the SBRC. Initially,

we highlight historical events regarding the Internet development in Brazil from the 1980's until nowadays, divided in three periods [7]. Thereafter, we highlight key events in the telephony and transportation sectors. Finally, we present the main advancements in the educational front.

1980s: In the beginning of this decade, there were experimental computer networks inside universities, mainly connecting workstations. In 1985, the Embratel (Empresa Brasileira de Telecomunicações—Brazilian Telecommunications Company) released the RENPAC (Rede Nacional de Comunicação de Dados por Comutação de Pacotes—National Packet Switched Network) to interconnect workstations and mainframe computers located anywhere in the country and abroad. However, the research community wanted to interconnect the academic Brazilian network to some academic network in USA using the Bitnet, the predecessor to the Internet in Brazil. In the 1988, there were three links between Brazil and USA. The first link was created between the Federal University of Rio de Janeiro and the University of California, Los Angeles. The second link was created between the LNCC (Laboratório Nacional de Computação Científica—National Scientific Computation Laboratory) and University of Maryland. Later, the Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo—Foundation for Research Support for the State of São Paulo) created a link to the Fermi National Laboratory. At the end of this decade, Fapesp deployed the Academic Network at São Paulo (ANSP), the first Brazilian academic network to connect universities in the São Paulo state to the Bitnet network using a 4,800 bps link. Em 1989, the Ministério da Ciência e Tecnologia (Science and Technology Ministry) created the RNP (Rede Nacional de Pesquisa—National Research Network) in order to create an Internet infrastructure to connect the academic community in Brazil. In summary, we can note that in the early development of the Internet in Brazil access was restricted to universities and research institutions, mainly in the South-east region of the country.

1990s: This decade was very important to increase the Internet usage among universities, researchers and the general public. In the beginning of this decade, the Brazilian government created incentives to foster the acquisition of computers, peripherals and telecommunications equipment by allowing the importation of electronics devices at lower taxation rates. In 1991, the RNP started to expand and in 1992 its backbone covered 11 cities with 9.6 and 64 kbps links. In 1994, the RNP backbone covered all the Brazilian regions. Due to the Internet expansion during the beginning of this decade, the Brazilian Government created the Internet Steering Committee in Brazil (CGI.br—Comitê Gestor da Internet no Brasil) to coordinate and integrate all Internet initiative services in the country, promoting technical quality, innova-

tion and dissemination of the services offered. The year 1995 was a milestone for the development of the commercial Internet. The Ministries of Communications, Science and Technology allowed the establishment of private Internet Service Providers (ISP), thus enabling the first commercial operations in Brazil. In 1999 the UOL launched the first Brazilian instant messenger software, called ComVC. In the end of this decade, the number of Internet users was more than 2.5 million. We can note that during this decade the Internet access started to become more democratic in Brazil, when all regions of the country became covered by the main backbone and also due to the development of the commercial Internet.

2000s: In late 2000, Brazil had more than 150 ISPs. Due to the shortcomings of the current Internet, a new Internet, called Internet 2, with a higher performance was developed. During this period, the RNP network was fully updated to support advanced applications. Since then, the RNP backbone has points of presence in all Brazilian states. In 2005, the backbone was updated with optical links operating at multi-gigabits per second. Nowadays, there are some Internet providers that offer a link up to 100 Mbps. We can note that it was during the 2000s that the Internet reached all parts of Brazil, thus becoming fully democratic. Moreover, the increase in links bandwidth enabled the use of advanced applications, e.g., collaboration tools and VoIP telephony applications. These aspects combined certainly contributed to diminish the barriers imposed by geographical distances, thus enabling the interaction/collaboration of geographically apart groups of people.

There are other important events that happened during these three decades that may have had a great impact on how people interact. For instance, in 1998 the Brazilian government privatized the phone sector. As a consequence, the price for making phone calls decreased substantially and the number of phones increased. Moreover, in the aviation sector we can highlight the following points: in 1996 the TAM (Transportes Aéreos Marilha) airline began to operate nationwide flights.[6] The GOL airline was established and started operations in 2001 with affordable ticket fares when compared to existing airline companies.[7] In a country with continental dimensions such as Brazil, this certainly contributed to attract researchers from all regions of Brazil to gather together every year at the symposium, thus increasing the chances of new collaborations.

Another important historical aspect that may have had a significant influence on the temporal behavior of the SBRC network is the growth of the number of computer science graduate programs over the years, shown in Fig. 26a.

---

[6] www.tam.com.br.

[7] www.voegol.com.br.

Observe how the behavior of the curve corroborates all the temporal results we have shown in this paper. From the 1990s onward, the number of programs have grown significantly, approximately one new program each year. Moreover, Fig. 26b shows this growth for each region of Brazil. It is interesting to notice that the evolution on the number of publications in SBRC for each region, as shown in Fig. 5, follows almost the same pattern as the evolution of graduate programs for each region. This clearly reflects the fact that investments in educational development, especially graduate programs, leads to an increase in knowledge production.

References

1. Albert R, Jeong H, Barabási AL (1999) Diameter of the world wide web. Nature 401:130–131
2. Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: KDD '06: proceedings of the 12th ACM SIGKDD, pp 44–54
3. Bayati M, Kim JH, Saberi A (2010) A sequential algorithm for generating random graphs. Algorithmica 58(4):860–910
4. Bazzan A, Argenta V (2011) Network of collaboration among pc members of Brazilian computer science conferences. J Braz Comput Soc 17:133–139
5. Ben-Naim E, Frauenfelder H, Toroczkai Z (2004) Complex networks. Lecture notes in physics. Springer, Berlin
6. Bonacich P (1987) Power and centrality: a family of measures. Am J Sociol 95(5):1170–1182
7. de Carvalho MSRM (2006) A trajetria da internet no brasil: do surgimento as redes de computadores ã instituicao dos mecanismos de governança. Coope, Federal University of Rio de Janeiro, Master's Thesis
8. Ceriani L, Verme P (2012) The origins of the gini index: extracts from variabilità e mutabilità (1912) by corrado gini. J Econ Inequal 10(3):421–443
9. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. SIAM Rev 51(4):661–703
10. Costa LF, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56:167–242
11. Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. In: Proceedings of 14th ACM SIGKDD, pp 160–168
12. Du N, Faloutsos C, Wang B, Akoglu L (2009) Large human communication networks: patterns and a utility-driven generator. In: KDD '09: Proceedings of the 15th ACM SIGKDD, pp 269–278
13. Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. In: SIGCOMM, pp 251–262
14. Fortunato S, Lancichinetti A (2009) Community detection algorithms: a comparative analysis: invited presentation, extended abstract. In: Proceedings of the fourth international ICST conference on performance evaluation methodologies and tools, VALUETOOLS '09, pp 27:1–27:2
15. Freire V, Figueiredo D (2011) Ranking in collaboration networks using a group based metric. J Braz Comput Soc 17:255–266
16. Gini C (1912) Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. pt. 1. Tipogr. di P. Cuppini
17. Guo Z, Zhang Z, Zhu S, Chi Y, Gong Y (2009) Knowledge discovery from citation networks. In: 2009 Ninth IEEE international conference on data mining, pp 800–805
18. Hassan AE, Holt RC (2004) The small world of software reverse engineering. In: Proceedings of the 11th working conference on reverse, engineering, pp 278–283
19. Hidalgo CA, Rodriguez-Sickert C (2008) The dynamics of a mobile phone network. Phys A Stat Mech Appl 387(12):3017–3024
20. Hill S, Nagle A (2009) Social network signatures: a framework for re-identification in networked data and experimental results. In: CASON '09: Proceedings of the 2009 international conference on computational aspects of social networks, pp 88–97
21. Jensen DD, Fast AS, Taylor BJ, Maier ME (2008) Automatic identification of quasi-experimental designs for discovering causal knowledge. In: KDD '08: Proceeding of the 14th ACM SIGKDD, pp 372–380
22. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. Science 311(5757):88–90
23. Kumar R, Novak J, Tomkins A (2006) Structure and evolution of online social networks. In: KDD '06: Proceedings of the 12th ACM SIGKDD, pp 611–617
24. Lazarsfeld PF, Merton RK (1954) Friendship as a social process: a substantive and methodological analysis. In freedom and control in modern society. 18(1):18–66
25. Leskovec J, Kleinberg JM, Faloutsos C (2007) Graph evolution: densification and shrinking diameters. TKDD 1(1):2
26. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: a new social network dataset using Facebook.com. Soc Netw 30(4):330–342
27. McPherson M, Lovin LS, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Sociol 27(1):415–444
28. Vaz de Melo POS, Almeida VAF, Loureiro AAF (2008) Can complex network metrics predict the behavior of nba teams? In: KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 695–703
29. Menezes GV, Ziviani N, Laender AH, Almeida V (2009) A geographical analysis of knowledge production in computer science. In: Proceedings of the 18th international conference on world wide web, WWW '09, pp 1041–1050
30. Nascimento MA, Sander J, Pound J (2003) Analysis of sigmod's co-authorship graph. SIGMOD Rec 32(3):8–10
31. Newman MEJ (2001) Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E 64(1):016131+
32. Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E 64(1):7
33. Newman MEJ (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci USA 98(2):404–409
34. Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256
35. Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration. In: Proceedings of the national academy of sciences, pp 5200–5205
36. Newman MEJ (2011) Complex systems: a survey. Am J Phys 79(8):800–809
37. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818
38. Procópio PS, Laender AHF, Moro MM (2011) Anãlise da rede de coautoria do simpósio brasileiro de bancos de dados. In: Simpósio Brasileiro de Banco de, Dados, pp 050–1-050-8
39. Sadras V, Bongiovanni R (2004) Use of Lorenz curves and gini coefficients to assess yield inequality within paddocks. Field Crops Res 90:303–310
40. Seshadri M, Machiraju S, Sridharan A, Bolot J, Faloutsos C, Leskove J (2008) Mobile call graphs: beyond power-law and

lognormal distributions. In: KDD '08: Proceeding of the 14th ACM SIGKDD, pp 596–604

41. Silva TH, Celes CSFS, Mota VFS, Loureiro AAF (2012) Overview of ubicomp research based on scientific publications. In: Proceedings of IV Simpósio Brasileiro de Computação Ubíqua e Pervasiva, SBCUP 2012

42. Watts DJ (2004) Six degrees: the science of a connected age. W. W. Norton & Company, New York

43. Watts DJ, Dodds PS, Newman MEJ (2002) Identity and search in social networks. Science 296(1):1302–1305

44. Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393:440–442