XL CILAMCE
IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING
NOVEMBER 11-14, 2019
Praiamar Natal Hotel & Convention Natal, RN-BRAZIL

A CO-AUTHORSHIP NETWORK ANALYSIS OF RESEARCH FROM GRADUATE PROGRAMS OF THE INTERDISCIPLINARY AREA IN BRAZIL

Solon M. Soares
Pedro H. T. Schimit
soares.solonm@gmail.com
schimit@uni9.pro.br
Universidade Nove de Julho – Programa de Pós-Graduação em Informática e Gestão do Conhecimento
Rua Vergueiro, 235/249, 01504-001, São Paulo/SP, Brazil

Abstract. Scientific collaboration has an increasing acknowledgment as one of the keys to research quality, which impacts the academic performance of graduate programs (GP). This investigation assesses the researchers' collaboration in terms of the co-authorship network formed by scientific papers published by permanent professors in Brazilian GPs from the Interdisciplinary area in 2017. In general, the co-authorship network is formed by permanent professors as the vertices, and the joint publication of a paper identifies the edges. We obtained data from the Sucupira and Lattes platform of a sample set of 1040 permanent professors, in 56 graduate programs, in 48 universities for the 2017 year. For each GP, a collaborative network was created based on published articles, and topological parameters, like cluster coefficient, density, diameter, degree, and shortest path were calculated. Other variables have been taken into consideration to help the analysis, such as the GPs number of permanent professors, the GPs evaluation grade applied by CAPES in the 2013-2016 period analysis, the university management type (public or private), and the region from Brazil where it is located. Therefore, the objective is to verify the variables which may indicate the GP quality, and although the properties of the network are not statistically related to the GP quality, they help to analyze some characteristics of GPs according to the management type of the institution.

Keywords: academic performance, scientific collaboration network, social network.

1 Introduction

The impact of scientific collaboration on research quality has been an object of study over the last decades. These studies aim to understand what is the impact of collaboration in the construction of scientific knowledge in terms of the establishment of social links [1], impacting the professionalization of scientific research [2], the knowledge diffusion [3], positively influencing the development of build knowledge and academic performance [4].

In general, the study on these social networks is limited to rather small networks. This investigation views a small community restricted to permanent professors attached to Brazilians graduate programs that propose to construction scientific knowledge. Said that this is institutional organization studies about a social process that considers action between researchers at where socially features establish limits and possibilities for scientific activities [4].

One of the ways to measure the relationship between researchers and measure its effects is through scientific research performed by co-authorship works [5-6]. The mapping of this scientific collaboration network is established by co-authorship in one or more papers [7]. This social network of researchers can be modeled by graphs, where the vertices represent the researchers and the connection between the nodes, the edges, represents the publication of one or more scientific papers produced and published in co-authorship with other researchers [8-9]. The modeling of these scientific collaboration networks through graphs allows the observation of social characteristics such as the influence exerted by a researcher or research group [10-11].

It is important to note that knowledge areas can group scientific research, and this delineation into smaller groups can also be restricted to geographical locations or entities such as universities or journals, revealing distinct characteristics of specific groups. In Brazil, scientific research is conducted and produced from universities and their graduate programs [13-14]. These programs are assigned to knowledge areas, and the related networks have formed a solid basis for the study and understanding of the research quality in Brazil [15-16].

Concerning this scope, this paper studies the distinct characteristics of a set of GPs, analyzing the co-authorship networks between researchers for the Interdisciplinary courses in the Engineering, Technology and Management subarea, for 2017.

2 Methodology

The Brazilian scientific scenario is conducted within universities and their GPs that are related to knowledge areas and can be subdivided into even more specific subareas. The formatting of these programs is also covered by two modalities where the objectives differ in the approach in the formation of new researchers. The definition of these areas and modalities are coordinated by the Higher Education Personnel Improvement Coordination (direct translation to CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

CAPES has a methodology for measuring the quality of the programs, called the CAPES concept, ranging from 1 to 7. Programs with evaluations 1 and 2 do not have the minimum qualifications for training researchers, being discontinued. Programs with a grade of 3 or higher have the minimum quality required for training researchers with a master's degree. Programs with grade 4 or higher have the minimum qualification required for the training of researchers with doctoral degrees. The programs with grade 5 or higher represent programs with national excellence quality and internationalization activities. The programs with grade 6 or higher represent programs with national excellence quality and a representative international insertion. The programs with grade 7 represent national and international excellence. The discontinuation of programs does not represent the termination of the program, since it may be restructured and resubmitted to CAPES to obtain a new authorization to operate.

In this paper, we will consider 1040 permanent professors and 56 academic programs from 48 universities in the Interdisciplinary area, Engineering, Technology, and Management subarea for the quadriennium 2013-2016, when the last evaluation was carried out by CAPES. The data was acquired from Lattes and Sucupira public data platforms between April and May, 2019. Lattes [17] is an open

public platform with information provided by researchers to make their curriculum available online. Sucupira [18] is a public platform where PGs submit the data required by CAPES to evaluate them. It is worth noting that we considered only the permanent professors of the programs, excluding the collaborators and visiting positions.

The total data analysis took an average of 727 seconds and has an error in identifying article authorship below 1%, with the co-authorship identification error close to 0.1%. The GP structure analysis was performed in 203 seconds and showed no measurable error. The co-authoring network mapping is accomplished per GP and the analysis of its characteristics is done by the iGraph tool [21]. The computer used in the analysis is a Intel i5 1.8GHz and 4GB RAM.

The data will be examined by using statistical analysis, and the topological parameters of the networks. The topological parameters to be calculated for each network are the average values of the clustering coefficient (cc), density (den), diameter (diam), average degree (deg), and shortest path length (spl). For one node of the network, the clustering coefficient is the percentage of connections between its neighbors, and the degree is its number of connections. We consider only papers published in international journals.

The density represents a fraction of possible edges between all nodes, the shortest path length is the to travel minimum path between two nodes, and the diameter is the higher shortest path length. We consider the average values of these parameters for all network nodes [1, 8, 9]. The clustering coefficient describes the social interaction between nodes, were three nodes connected form a triangle, in a sense, a trivial form of fractional transitive communication between these nodes, average degreed represents the connection average quantity that node has.

3 Result

The first results to be presented are related to macro data of the GPs analyzed. Therefore, Figures 1, 2 and 3 have the distribution of the GP in Brazil regions, the number of GPs for each management type (state, federal and private), and the distribution of GPs according to the CAPES concept, respectively. Note that the GPs are concentrated on the Southeast region, with more federal universities with the CAPES concepts 3 and 4.

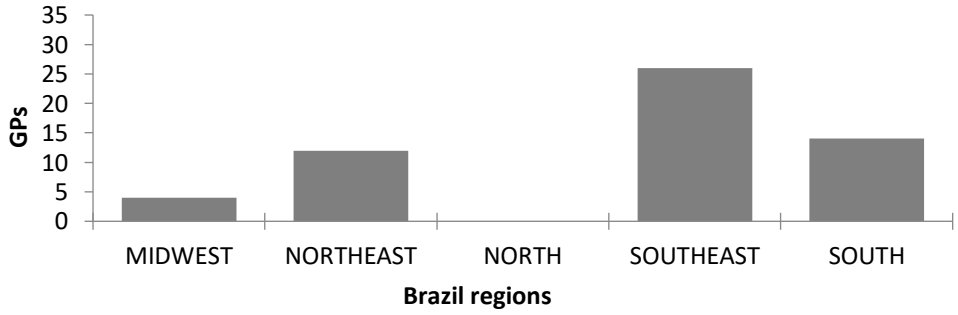

Figure 1. Distribution of the GP in Brazil regions.

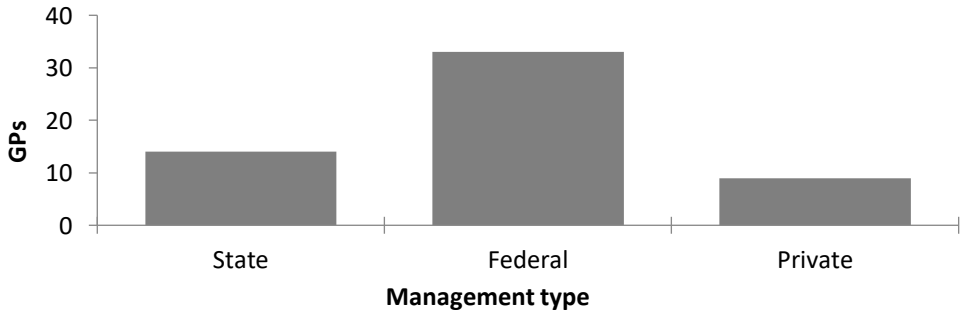

Figure 2. Distribution of the GP in Brazil at management type.

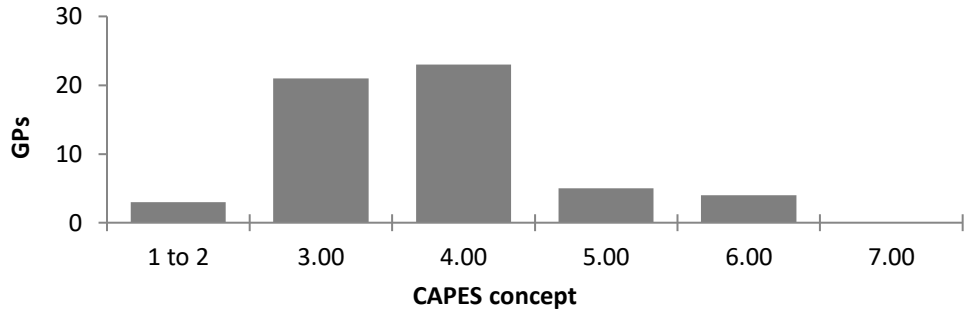

Figure 3. Distribution of the GP in Brazil by the CAPES concept.

The distribution of the CAPES concept in the Brazilian region is in Table 1. Note that the Southeast region contains the GPs with a higher concept and has a higher CAPES concept average for state universities. Moreover, the number of universities per management type and their average CAPES concept are in Table 2. Note that the Southeast region contains a higher average concept and the private management type has a more high-value average CAPES concept.

Table 1: Distribution of regional CAPES concept at the graduate-programs in 2017.

Qt GP	Region	Concept CAPES					
		1 to 2	3	4	5	6	7
4	MIDWEST	1	1	1	1	0	0
12	NORTHEAST	1	6	3	2	0	0
0	NORTH	0	0	0	0	0	0
26	SOUTHEAST	1	6	14	2	3	0
14	SOUTH	0	8	5	0	1	0
56	BRAZIL	3	21	23	5	4	0

Table 2: Distribution of regional by university management type at the graduate programs in 2017.

Qt GP	Region	Management type			Average CAPES concept			Average concept
		State	Federal	Private	State	Federal	Private	
4	MIDWEST	0	4	0	-	3.50	-	3.50
12	NORTHEAST	3	7	2	2.67	3.57	4.50	3.50
26	SOUTHEAST	8	14	4	4.38	3.86	3.75	4.00
14	SOUTH	3	8	3	3.00	3.75	3.67	3.57
56	BRAZIL	14	33	9	3.71	3.73	3.89	3.75

A general picture of the professors in the GPs is in Table 3, with the number of professors, average experience time, and an average number of professors per GP according to management type. Note that the Southeast region has a large professor quantity and has a high average experience time. Moreover, when relating this data with Table 2 we can see that the average experience time bethink the average concept of the GP by regions. The Professor's experience time is the average time, in years, since the Professor's doctoral conclusion.

Table 3: Distribution of regional professors at the graduate programs in 2017.

Region	Professors	Professors' experience time	Average Professor by management type		
			State	Federal	Private
MIDWEST	79	12.04	0.00	19.75	0.00
NORTHEAST	190	12.58	15.67	15.86	16.00
SOUTHEAST	528	15.58	23.50	20.64	12.75
SOUTH	243	12.29	22.67	16.50	14.33
BRAZIL	1040	14.24	21.64	18.52	14.00

Figures 4, 5, and 6 show the histograms for the distribution of GPs according to density, degree, and clustering coefficient, respectively. In general, the GPs networks have a low density, average degree, and clustering coefficient, indicating that Professors do not have a high-quality paper published together. Moreover, due to this low level of edges, usually the networks are not connected, and the shortest path length and the diameter do not return meaningful values for the analysis.

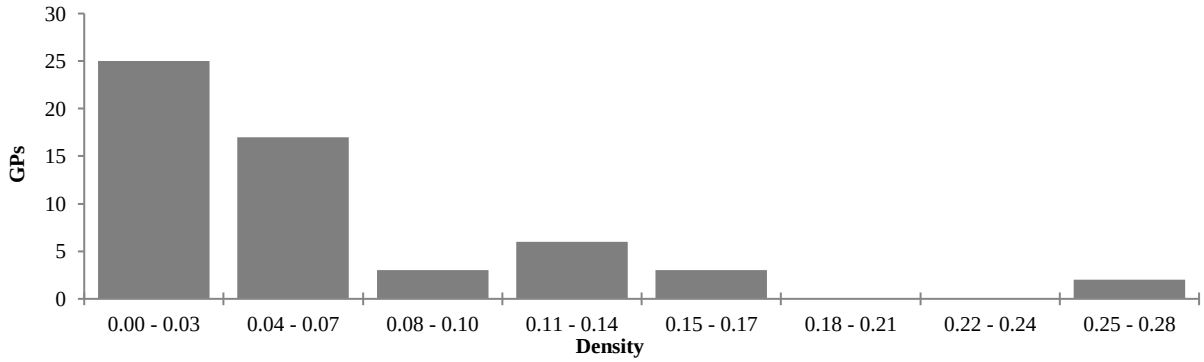

Figure 4. Distribution of the co-authorship network density results.

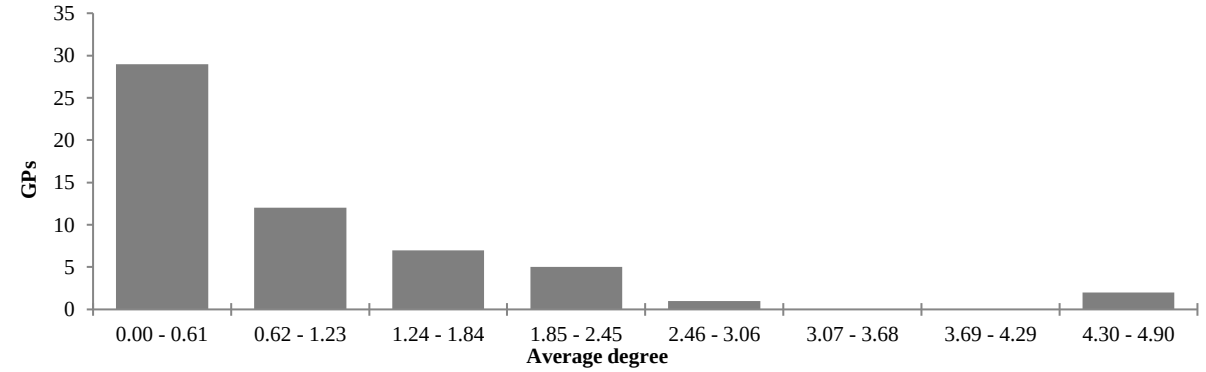

Figure 5. Distribution of the co-authorship network average degree.

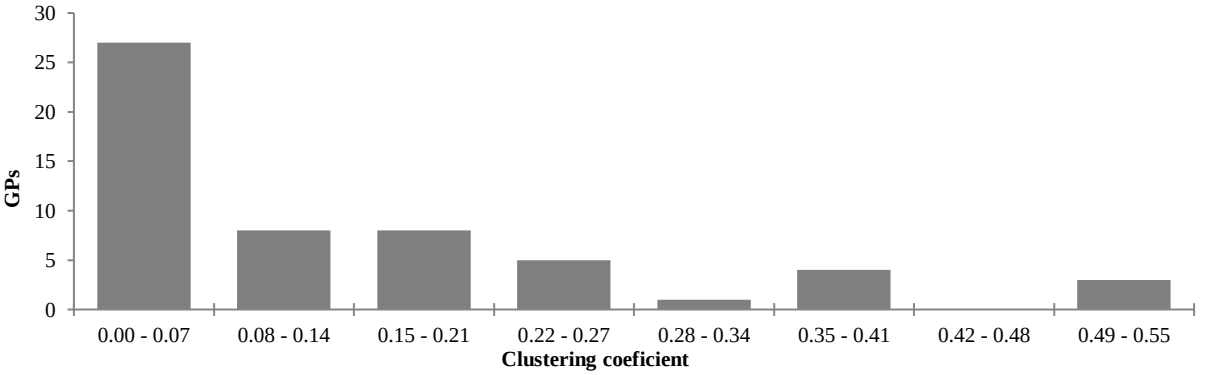

Figure 6. Distribution of the co-authorship network cluster coefficient.

However, these data are different when we split the GPs according to the university management type. A summary of the average values for the clustering coefficient, degree, density, Professor's experience time, and CAPES concept is in Table 4. Here, it is possible to see that GPs from private universities have networks with more connections, indicating that Professors have more papers published conjointly with other Professors of the same program and have the highest CAPES concept, though the average experience time is the lowest value of the table. The other data for public universities are similar, except for the Professor's experience time, which is lower for Federal universities.

Table 4. Average co-authorship network features by type of university management.

Management type	cc	deg	den	Professors' experience time	Average Professor	CAPES concept
Private	0.2385	1.6139	0.1147	12.88	14.00	3.89
State	0.0965	0.5316	0.0279	15.32	20.20	3.71
Federal	0.1048	0.9382	0.0516	13.56	18.52	3.73
Public	0.1023	0.8171	0.0445	14.15	19.04	3.72
Brazil	0.1242	0.9451	0.0558	13.99	18.25	3.75

4 Discussion

The objective of the paper was to verify collaboration networks of Graduate Programs and analyze the topological parameters to identify relations with the quality of the program, expressed by the CAPES concept. Therefore, the macro data of the GP from the Interdisciplinary area, Engineering, Technology, and Management subarea were presented. Also, the histogram of distributions of density, average degree, and clustering coefficient showed that, in general, the networks of these programs have few

connections. However, by separating the GPs according to the management type, some interesting differences appeared between private and public institutions.

Of course, the small sample is not ideal for this type of study. However, to explore a small collaborative network a first analysis must be concentrated in one knowledge area due to different aspects of publication and researcher cooperation activities. The Sucupira platform has proven to be an excellent source of data for identifying these small collaborative networks and when worked in conjunction with the Lattes platform the amount of data available helps to clarify the results obtained [19-20]. Note that even considering only 56 GPs, the results show meaningful differences between the Brazilian regions and the universities with different management types. Midwest, North, and Northeast regions have a low number of GPs, though three of the nine high-quality GPs (CAPES concepts 5 and 6) are present in these regions. Also, in the Southeast region that has a high number of GP and the Professors have more experience than the rest of the country what can justify the high average CAPES concept.

The topological parameters analysis showed networks with a low number of connections, indicating low cooperation among members of the same GP. It cannot be concluded that Brazilian researchers of this knowledge area do not have a cooperation network, but only that it is not so usual the cooperation with the closest pairs. Previous studies support that scientific cooperation reflects professionalization in scientific investigations, this would allow better access to resources and knowledge to rationalize the work and time of the researcher [2, 3]. That said when observing the results of the topographic parameters of collaboration networks at private universities conclude that there is a practice differentiated in the distribution of resources and knowledge to take more cooperation that other type management [13].

Although cooperation networks are analyzed using many researchers [1, 8, 9], the small network analysis is becoming increasingly common. This small world has allowed the verification of quality, aspects, and differences unique to one or more areas of knowledge [15, 16], exposing their differences in universities in a country [12] or their regionalities [14] or cooperative behavior [4]. The number of articles published, co-authorship, citations or the quality of these publications were not taken into consideration, only the existence of social interaction between peers and what are the characteristics that may influence cooperation [5]. These investigations it was possible to identify that in small collaborative networks topographic parameter such as diameter and shortest path length return poor results [7], perhaps in this case, the larger temporal sampling allows to observe these characteristics.

Therefore, the next steps of this work will deal with more network properties, consider the impact of the paper in terms of quality and citations, extend the period analyzed and use the data for other sub-areas of the Interdisciplinary area. It was realized that for the evaluation of small social networks the number of members can mask the real connectivity of their members, so there is a need for the new formulation to understand the average degree.

Acknowledgments

PHTS is supported by grants #303743/2016-6 and #402874/2016-1 of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and grant #2017/12671-8, São Paulo Research Foundation (FAPESP).

References

[1] A. L. Barabási; et al. Evolution of the social network of scientific collaborations. *Physical A*, n. 42, pp. 590-614, 2002.

[2] D. D. Beaver; R. Rosen. Studies in scientific collaboration Part I. The professional origins of scientific co-authorship. *Scientometrics*, vol. 1, n1, pp. 65-84, 1978.

[3] Z. Xie; J. Li; M. Li. Exploring Cooperative Game Mechanisms of Scientific Coauthorship Networks. *Complexity*, vol. 2018, pp. 1-11, 2018.

[4] E. R. Guarido Filho; C. L. Machado-da-Silva; L. Rossoni. The social and intellectual dimensions in the construction of scientific knowledge: the institutional theory in organization studies in Brazil. *Bar - Brazilian Administration Review*, vol. 7, n. 2, pp. 136-154, 2010.
[5] J. P. Mena-Chalco; et al. Brazilian bibliometric coauthorship networks. *Journal Of The Association For Information Science And Technology*, vol. 65, n. 7, pp. 1424-1445, 2014.
[6] L. Aldieri; M. Kotsemi; C. P. Vinci. The impact of research collaboration on academic performance: An empirical analysis for some European countries. *Socio-economic Planning Sciences*, vol. 62, pp. 13-30, 2018.
[7] M. E. J. Newman. Who Is the Best Connected Scientist? A Study of Scientific Coauthorship Networks. *Complex Networks*, pp. 337-370, 2004
[8] M. E. J. Newman. Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review e*, vol. 64, n. 1, pp. 1-8, 2001.
[9] M. E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review e*, vol. 64, n. 1, pp. 1-7, 2001.
[10] P. Zhang; et al. Proactive rumor control in online network. *World Wide Web*, vol. 22, n. 4, pp. 1799-1819, 2018.
[11] E. Perez-Cervantes; et al. Using Link Prediction to Estimate the Collaborative Influence of Researchers. *2013 Ieee 9th International Conference On E-science*, pp. 293-300, 2013.
[12] H. W. Choi; Y. J. Choi; S. Kim. Network analysis of scientific collaboration in North Korea. *Science Editing*, vol. 6, n. 1, pp. 25-34, 2019.
[13] R. C. Ribeiro; A. M. Magalhães. Política de responsabilidade social na universidade: Conceitos e desafios. *Educação, Sociedade e Cultura*, n. 42, pp. 133-156, 2014.
[14] C. G. Souza; M. L. A. Ferreira. Researchers profile, co-authorship pattern and knowledge organization in information science in Brazil. *Scientometrics*, vol. 95, n. 2, pp. 673-687, 2012.
[15] M. M. M. Autran; et al. Perfil de produção acadêmica dos programas brasileiros de pós-graduação em Ciência da Informação 2008-2012. *Perspectivas em Ciência da Informação*, vol. 20, n. 4, pp. 57-78, 2015.
[16] E. A. Haddad; J. P. Mena-Chalco; O. Sidone. Produção científica e redes de colaboração dos docentes vinculados aos programas de pós-graduação em Economia no Brasil. *Estudos Econômicos (São Paulo)*, vol. 47, n. 4, pp. 617-679, 2017.
[17] CAPES. Plataforma Lattes [internet]. Brazil [access at: 31 jul. 2019]. Available in: http://lattes.cnpq.br/
[18] CAPES. Plataforma Sucupira [internet]. Brazil [access at: 31 jul. 2019]. Available in: https://sucupira.capes.gov.br/sucupira
[19] J. P. Mena-Chalco; R. M. Cesar Junior. Towards Automatic Discovery of co-authorship Networks in the Brazilian Academic Areas. *2011 Ieee Seventh International Conference On E-science Workshops*, pp. 1-8, 2011.
[20] T. M. R. Dias; G. F. Motta; P. M. Dias. Um estudo sobre a rede de colaboração científica dos pesquisadores brasileiros com currículos cadastrados na Plataforma Lattes. *em Questão*, vol. 25, n. 1, pp. 63-86, 2019.
[21] G. Csardi; T. Nepusz. The igraph software package for complex network research, *InterJournal, Complex Systems*, 2006.