Revista de Sistemas de Informação da FSMA
n. 14 (2014) pp. 2-10

SISTEMAS DE INFORMAÇÃO
http://www.fsma.edu.br/si/sistemas.html

TRILHA ESTUDANTIL

Trend Analysis of the Brazilian Scientific Production in Computer Science

Caio Cesar Trucolo, *Student at the Masters Program in Information Systems, USP*,
Luciano Antonio Digiampietri, *Ph.D. Prof. at the undergraduate and graduate programs in Information Systems at USP*

***Abstract*—The growth of scientific information volume and diversity brings new challenges in order to understand the reasons, the process and the real essence that propel this growth. This information can be used as the basis for the development of strategies and public politics to improve the education and innovation services. Trend analysis is one of the steps in this way. In this work, trend analysis of Brazilian scientific production of graduate programs in the computer science area is made to identify the main subjects being studied by these programs in general and individual ways.**

***Keywords*—Trend Analysis, Academic Social Networks, Computer Science**

I. Introdução

IN order to understand the behavior of individuals or of a group of individuals that belong to a social circle and create any kind of knowledge, such as opinions in Internet portals or papers in scientific journals, we need more than one single type of analysis.

The change of the behavior with time can be considered as one of the topics to be analyzed. Since the behavior of groups and individuals changes with time, several services try to identify or forecast trends in order to increase business competitiveness or to establish adequate policies for those variations.

The analysis of the behavior of the Brazilian scientific production may be considered a challenge due to the fact that Brazil is the fifth largest country in the world, has the fifth largest population and a huge cultural diversity. In the last decades, the number of Brazilian papers increased exponentially (close to 12% per year), growing from little more than 10,000 full papers published in 1980 to almost 400,000 in 2010, according to data obtained from the analysis of more than one million Lattes curricula [1].

The understanding of the characteristics of this production allows the creation of public strategies and policies that go together with the scientific trends, increasing even more the Brazilian potential in this context.

This paper aims to analyze part of the Brazilian scientific production, most specifically, the one in the area of computer science. The goal of this work is to develop and apply a methodology to identify trends in subjects and research branches in the *stricto sensu* graduate programs evaluated by CAPES, based on the information included in the curricula of the advisor professors that are stored in the Lattes platform. Besides, we performed a network analysis of co-authorship in order to verify if there are strong relations between the verified trends and the co-authorship networks. It is important to point out that this paper intends to verify current trends based on a historical analysis. It is out of its scope to verify the diffusion and popularization speed of these trends within the programs. This analysis helps identifying the directions taken by research in Computer Science in Brazil, which may help researchers identify growing fields and helping the process of evaluating new research projects.

Corresponding author: Luciano Digiampietri, *digiampietri@usp.br*

The rest of this paper is organized as follows. Section 2 summarizes related work, section 3 presents the Lattes platform and its importance as a guideline for the national scientific production. Methodology is described in section 4, section 5 presents and analyzes the results and, at last, section 6 presents the final considerations.

II. Related work

FOR the last years, several papers analyzed trends for different types of applications, which span from social networks to stock markets. Among those works the ones closer to this are those that study historical text documents.

In the work of Bolelli et al. [2], *Lattent Dirichilet Allocation* and *Gibbs Sampling* were used together with the time order of the documents to create a generative model that learn author, topic and word distributions. In a synthetic application, the accuracy was approximately 72%.

The work of Kawamae [3] consisted of trying to forecast the topic distribution in scientific articles taking the time distribution into consideration. Based on the same idea, a new work [4] established the difference between stable topics (that do not have a significant variation with time) and dynamics ones, trying to refute other models that only take into consideration the topic explosion (sudden increases of topic presence during some periods of time). For the evaluation, the author compared the proposed model with two others using the perplexity and the L1 error rate. At last, the model presents those two rates smaller than the other two models, achieving an average L1 of 2.44.

Jayashri and Chitra [5] proposed a model with an ART network (*Adaptive Resonance Theory*) to identify topics in scientific documents from different databases and to detect trends using the peaks of those extracted topics. The approach used was able to detect the topics in vogue for different databases.

Park et al. [6] used a trend detection approach using characteristic selection based on IG-I (*Improved Gini - Index*). For each topic given as input, subtopics were extracted to analyze its time behavior and identify it as growing or decreasing. For the four topics given as input to the model, F1 measurement tests were perform together with SVM (*Support Vector Machine*) and kNN (*k - nearest neighbors*). The F1 result for SVM was 0.982 and for kNN was 0,916. One limitation of this method is that the process is not fully automatic.

Abe e Tsumoto [7] selected important terms using TF-IDF (*Term Frequency - Inverse Document Frequency*) and the *Jaccard coefficient* using linear regression *a posteriori* to detect emerging trends. All the detected trends were confirmed as real ones by domain experts.

Besides the textual documents used as basis for trend analysis, in the last years social networks have also been used to help this analysis. The work of Cimenler, Reeves and Skvoretz [8] analysis using mostly centrality metrics in order to understand how significant these metrics are to forecast the performance of researchers using citation indices such as the h-index. The Poisson regression method is used to analyze the importance of the metrics for some types of networks formed by researchers from an Engineering college.

A systematic review on trend identification and analysis techniques for other applications besides historic text documents can be found in [9].

This work deepens the application in an interesting way. The data extracted from the Lattes platform allow for a fairly rich analysis on the national scientific condition. We can say, therefore, that this work differs from the previous works from the good approximation of the analysis to the national scientific reality. Another important characteristic is the fact that this analysis does not require human effort, that is, the trend analysis is perform based on terms and expressions automatically extracted from the database without the need to manually establish the importance of terms or limit values to the trends.

III. Lattes Platform

The Lattes platform is a Brazilian database that stores curricula from researchers and whose rule is of the utmost importance to academia. There are more than three million curricula stored with important information to the analysis of researchers and academic networks. In this platform it is possible to obtain data from researchers such as scholar information, area of work, bibliographic production, participation in thesis committees, advisor work and much more.

For this work, we selected 57.501 different papers, without the repetitions and redundancies caused by the characteristics of the platform. This number is distributed among the 45 Brazilian graduate programs on Computer Science and contains published papers from 1991 up to 2011. We decided to use the papers up to 2011 because we verified that the number of papers in general falls dramatically after 2011 because, in average, the users take more than a year to update their curricula [1]. Due to the fact that there was no dramatic change in national scientific policies in the last years, we assume that using data up to 2011 we can identify trends in a sufficiently precise and current way.

IV. Methodology

The entire analysis process was performed in three steps: data gathering, automatic term extraction and trend analysis of the extracted terms.

A. Data gathering

In order to gather the data, we first identified the advisors of the 45 *stricto sensu* graduate programs in computer science[1]. The information of the researchers curricula were tabulated and stored into a database according to the methodology described in [10]. In order to perform the tests, we extract the terms from 57,501 publication titles distributed from the 45 programs and published from 1991 up to 2011. These terms were then submitted to trend analysis considering publications from a single program and, afterwards, from all the graduate programs.

B. Automatic term extraction

The automatic term extraction technique used consists in determining the most important terms in the set of document by the adjacent frequency of the words that compose these terms. The formula used to calculate the weights for each candidate term is the following:

$$FED\,(TC) = f\,(TC) \times \left(\prod_{i=1}^{T} (FE\,(N_i) + 1) \times (FD\,(N_i) + 1) \right)^{1 \div T} > 1$$

where $f(TC)$ is the frequency of the candidate term, TC, and $FE(Ni)$ and $FD(Ni)$ indicate the frequency of the candidates to the left and to the right, respectively. This formula is described in details in [11].

A list of *stop words* in Portuguese, Spanish and English was used to filter the candidate terms.

We observed empirically that the composite terms were more meaningful than the simple terms in relation to the subject matters approached by the publications. Hence, the terms used in the trend analysis phase were the composite terms and the ones with higher weights.

[1]This information was obtained from the triennium report referring to the period 2007-2009.

C. *Trend Analysis*

Based on the extracted terms, we calculated the importance indices of the terms for each period (year). The importance index used in this paper is the TF-IDF (*Term Frequency divided by Inverse Document Frequency*), which is one of the most used indices to infer the importance of terms. With the calculated indices, we performed linear and non-linear regression in order to determine the trend curves that most fit the time series for each term. The regression types used were linear, exponential, logarithmic, power law and polynomial, with degrees from two to five. Afterwards, we calculated the quadratic error for each trend curve generated in order to determine the curve that best fits the time series for each term.

The classification of terms as trends was based on the forecasting, based on the best fit trend curve (that is, the one with the smallest quadratic error), for the first year after the historical analysis interval: the year of 2011. In spite of the fact that trend analysis approach used allows for forecasting of posterior years, we decided to use only the year after the historical analysis because this index already indicates the trend pointed out by the regression curve.

V. Results

FIGURE 1 presents the time behavior of three terms extracted based on the TF-IDF indices which were calculated with the trend analysis techniques. It is possible to see the big difference between the behaviors of these three terms. While the term *sensor networks* had a huge increase in the last few years, *object oriented* decreased significantly. Meanwhile, the term *neural networks* has been used in research for a while, and its time series has some ups and downs in the period under analysis, but has shown a quite stable behavior.

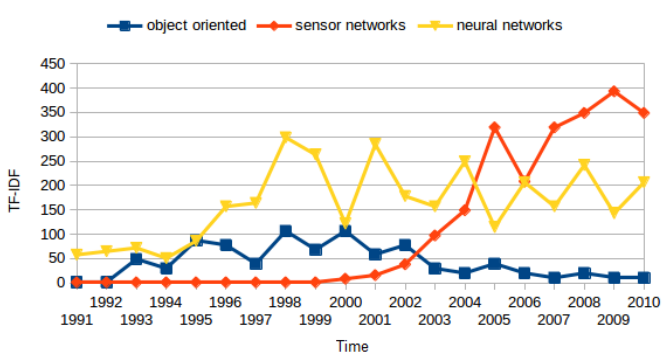

Fig. 1. Time behavior of three terms

Figures 2 and 3 show the trend curves based on nonlinear 3-degree polynomial regression and *power law*, respectively, of the terms *sensor networks* and *object oriented*. We can see in figure 2 that the term *sensor networks* growth started at the year 2000 while the term *object oriented* was increasing between the years 1998 and 2000, when it started to decrease. Comparing both behaviors, it becomes clear in this context that the term *sensor networks* can be identified as a trend in current research while *object oriented* cannot.

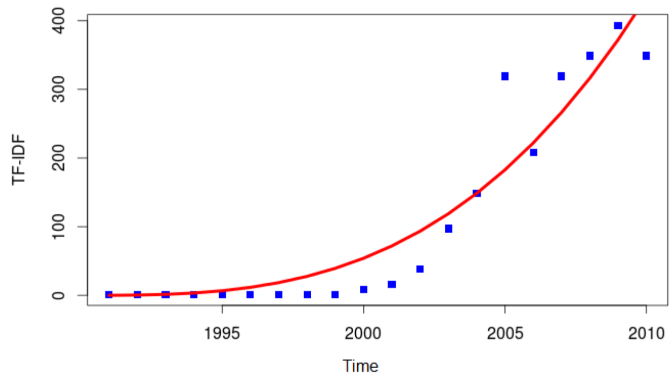

Fig. 2. Trend curve generated by nonlinear *power law* regression for the term *sensor networks*

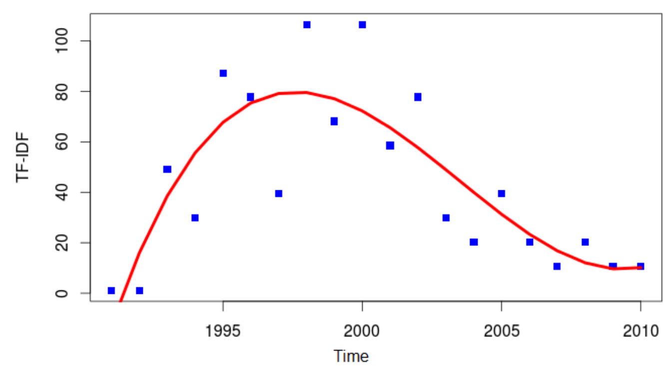

Fig. 3. Trend curve generated by nonlinear 3-degree polynomial regression for the term *object oriented*

In a global analysis, the table I shows the 20 term with the highest TF-IDF in the forecast for 2011, that is, the terms with the highest popularity trends. Comparing the forecast values with the real values for 2011, we observed an error of approximately 26.5% and a high correlation between the forecast values and the real ones (about 0.68).

Besides a global analysis, we identified the individual trends for each program. Based on that information, it is possible to see some characteristics of the research works for each program and whether they follow a national trend. Table II presents the main trends for one of the analyzed programs. It is clear that this program is highly interested in topics related to artificial intelligence, with a focus on the image retrieval application.

For a global view of the Brazilian graduate programs in Computer Science, table III shows the main trends for each one of the 45 programs evaluated by CAPES.

At last, we performed an analysis of co-authorship networks among professors in these graduate programs in order to verify if the trends we found are strongly correlated to the co-authorship networks.

Figure 4 contains the co-authorship network for full papers published by the professors of the 45 programs under analysis. The numbers presented in the caption

TABLE I
Main trends of the extracted terms

Term	Forecast TF-IDF for 2011
product line	413.57
wireless sensor	402.99
sensor networks	321.47
wireless sensor networks	320.69
neural networks	277.29
software product	255.62
product lines	243.05
software development	238.76
particle swarm optimization	227.22
swarm optimization	227.22
particle swarm	224.63
optimum-path forest	219.73
augmented reality	209.51
time series	208.92
genetic algorithm	207.47
case study	204.01
scheduling problem	181.71
social networks	181.41
infocomp ufla	176.29
genetic programming	173.48

TABLE II
Example of trends in one of the programs

Term	Forecast TF-IDF for 2011
neural networks	63.36
time series	56.65
artificial neural	51.89
artificial neural networks	51.84
product lines	30.05
image retrieval	24.86
access methods	24.33
feature selection	21.57

correspond to the same numbers used in tables III and IV. Each node in this network corresponds to a professor and each edge indicates co-authorship between the two professors connected by it. It is possible to observe in the graph the existence of a big connected component in the center of the image and the prevalence of colored edges (which indicate co-authorship between professors from the same graduate program). On the other hand, edges in gray represent collaborations between professors from different programs[2].

Table 4 presents for each graduate program the two programs most related to it, using the number of co-authorship relationships as metric. It is possible to realize that when crossing the information on tables 3 and 4 that the co-authorship relationship between programs is not very much related to the research tendencies in each pair of programs. Taking into consideration the information in tables 3 and 4, that is, the main research terms for each program and the two programs most related to each one, less than 9% have equal terms. One of these cases is, for instance, the case for programs 6 (UNICAMP - Computer Science) and 40 (UNESP/SJRP - Computer Science), which are programs that have some co-authorship relationships and share the term *optimum-path forest* which, as verified manually, is part of the title of papers published in co-authorship between professors in both programs. This can be an indication that the trends developed in each program usually are a result from internal collaborations, while the collaborations between programs are related to wider works (potentially in consolidated subjects). We intend to study this hypothesis deeper in future works. A more specific discussion on productivity and co-authorship between Brazilian graduate programs in Computer Science can be found in [12], [13].

VI. Final Considerations

THIS work presented general information on the trends in the Brazilian scientific production in the Computer Science field using an automatic term and expression identification technique. We performed an analysis of the general trends and individual analysis for each program. Thus, it was possible to identify which subjects are in vogue and which are disregarded. Besides, we perform an analysis of the co-authorship networks of the advisor professors in the graduate programs in the field of Computer Science in which we observed that there is no strong correlation between the trends we found in each program and the co-authorship between them.

The results found in this paper are an initial step if we consider the full potential of the trend analysis of the national scientific production. In future works, the structure of the sources of information will be included as a factor for analysis. This way, the academic social networks will be treated as a new variable to increase the accuracy of the new trend forecasting model.

Besides, we intend to group professors and terms according to the subareas of Computer Science in order to identify the dynamics and the publication trends in those subareas.

[2]A figure containing only the information from the 13 programs with the best evaluations from CAPES can be found in: http://www.each.usp.br/digiampietri/FSMA/maioresprogramas.png

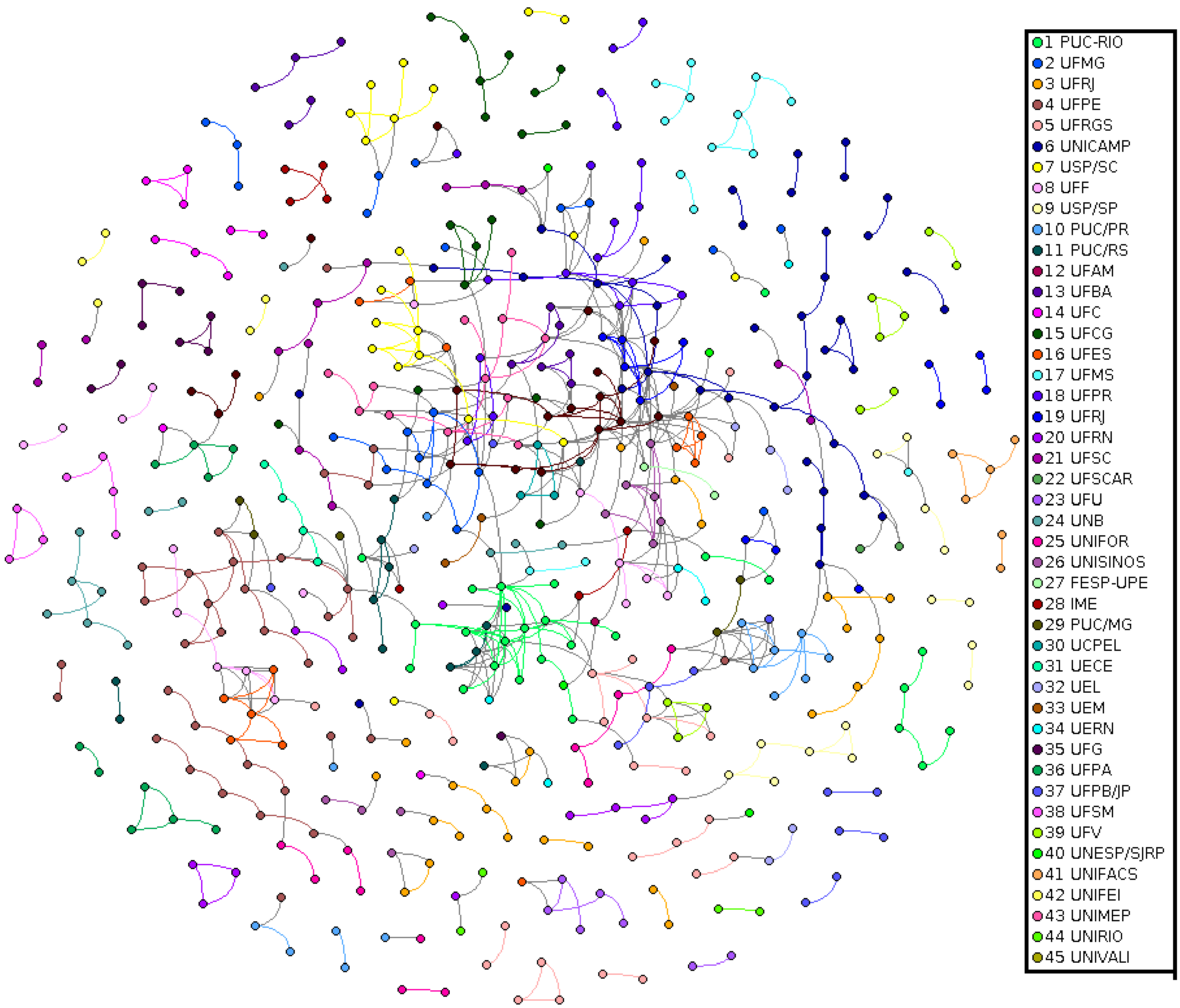

Fig. 4. Co-authorship networks from the analyzed programs

TABLE III
MAIN TRENDS IN TERMS FOR EACH PROGRAM

	Programa	Primeira tendência	Segunda tendência
1	PUC-RIO - INFORMATICS	product line	microscopy images
2	UFMG - COMPUTER SCIENCE	genetic programming	name disambiguation
3	UFRJ - COMPUTER AND SYSTEMS ENGINEERING	hyperbolic smoothing	clustering method
4	UFPE - COMPUTER SCIENCE	software development	time series
5	UFRGS - COMPUTING	sensor networks	eye fundus images
6	UNICAMP - COMPUTER SCIENCE	optimum-path forest	foresting transform
7	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS	neural networks	time series
8	UFF - COMPUTING	wave propagation	cellular automata
9	USP - COMPUTER SCIENCE	oriented relational	field-research oriented relational database
10	PUC / PR - INFORMATICS	arq scheme	music genre
11	PUC / RS - COMPUTER SCIENCE	promising protein receptor snapshots	infocomp (ufla
12	UFAM - INFORMATICS	grounded theory	medical devices
13	UFBA - COMPUTER SCIENCE - UFBA - UNIFACS	real-time systems	failure detectors
14	UFC - COMPUTER SCIENCE	software product	data integration
15	UFCG - COMPUTER SCIENCE	power management	application user interfaces
16	UFES - INFORMATICS	fault diagnosis	matrices reordering algorithms
17	UFMS - COMPUTER SCIENCE	shuffling experiments	e-sapi bovis
18	UFPR - INFORMATICS	swarm optimization	particle swarm optimization
19	UFRJ - INFORMATICS	three protagonist businessmen	visual impairment
20	UFRN - SISTEMAS E COMPUTAÇÃO	product line	wireless sensor
21	UFSC - COMPUTER SCIENCE	image segmentation	cronic non-transmissible
22	UFSCAR - COMPUTER SCIENCE	production scheduling	engenharia elétrica
23	UFU - COMPUTER SCIENCE	trace alignment	diferenciais parciais
24	UNB - INFORMATICS	particle swarm optimization	particle swarm
25	UNIFOR - APPLIED INFORMATICS	wireless sensor	decision analysis
26	UNISINOS - APPLIED COMPUTING	multilevel approach	composite structure
27	FESP / UPE - COMPUTER ENGINEERING	time series	particle swarm optimization
28	IME - SYSTEMS AND COMPUTING	upper bound	web services
29	PUC / MG - INFORMATICS	sensor networks	graph matching
30	UCPEL - INFORMATICS	simulação quântica	architecture using
31	UECE - COMPUTER SCIENCE	test case	release planning
32	UEL - COMPUTER SCIENCE	spectral analysis	detection using
33	UEM - COMPUTER SCIENCE	users personomy	personomy using
34	UERN - COMPUTER SCIENCE - UERN - UFERSA	sensor networks	transcoded videos
35	UFG - COMPUTER SCIENCE	capacitated arc	capacitated arc routing
36	UFPA - COMPUTER SCIENCE	iso/iec 12207	neurais artificiais
37	UFPB / J.P. - INFORMATICS	naive bayes	training assessment
38	UFSM - INFORMATICS	cognitive style	adaptive hypermedia
39	UFV - COMPUTER SCIENCE	scheduling problem	sequence dependent setup
40	UNESP / SJRP - COMPUTER SCIENCE	optimum-path forest	contours initialized
41	UNIFACS - SYSTEMS AND COMPUTING	developers context-specific preferred representational	preferred representational
42	UNIFEI - COMPUTER SCIENCE AND TECHNOLOGY	rough sets	self-organizing map model
43	UNIMEP - COMPUTER SCIENCE	realidade aumentada	augmented reality
44	UNIRIO - INFORMATICS	business models	case study
45	UNIVALI - COMPUTING	process capability models	neurais artificiais

TABLE IV
MOST RELATED PROGRAMS IN TERMS OF CO-AUTHORSHIP

	Program	Most related program	Second most related program
1	PUC-RIO - INFORMATICS	UNIRIO - INFORMATICS	UFF - COMPUTING
2	UFMG - COMPUTER SCIENCE	UFAM - INFORMATICS	PUC / MG - INFORMATICS
3	UFRJ - COMPUTER AND SYSTEMS ENGINEERING	UFRJ - INFORMATICS	UFF - COMPUTAÇÃO
4	UFPE - COMPUTER SCIENCE	FESP / UPE - COMPUTER ENGINEERING O	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS
5	UFRGS - COMPUTAÇÃO	PUC / RS - COMPUTER SCIENCE	UCPEL - INFORMATICS
6	UNICAMP - COMPUTER SCIENCE	UNESP / SJRP - COMPUTER SCIENCE	UFMG - COMPUTER SCIENCE
7	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS	UFRN - SYSTEMS AND COMPUTING	UFSCAR - COMPUTER SCIENCE
8	UFF - COMPUTING	UFRJ - COMPUTER AND SYSTEMS ENGINEERING	PUC-RIO - INFORMATICS
9	USP - COMPUTER SCIENCE	UFMS - COMPUTER SCIENCE	FESP / UPE - COMPUTER ENGINEERING
10	PUC / PR - INFORMATICS	UFPR - INFORMATICS	UFSC - COMPUTER SCIENCE
11	PUC / RS - COMPUTER SCIENCE	UFRGS - COMPUTING	UCPEL - INFORMATICS
12	UFAM - INFORMATICS	UFMG - COMPUTER SCIENCE	PUC-RIO - INFORMATICS
13	UFBA - COMPUTER SCIENCE - UFBA - UNIFACS	USP - COMPUTER SCIENCE	UNIFACS - SYSTEMS AND COMPUTING
14	UFC - COMPUTER SCIENCE	PUC-RIO - INFORMATICS	UFRJ - COMPUTER AND SYSTEMS ENGINEERING
15	UFCG - COMPUTER SCIENCE	UFPB / J.P. - INFORMATICS	UFPE - COMPUTER SCIENCE
16	UFES - INFORMATICS	UNIRIO - INFORMATICS	PUC-RIO - INFORMATICS
17	UFMS - COMPUTER SCIENCE	USP - COMPUTER SCIENCE	PUC-RIO - INFORMATICS
18	UFPR - INFORMATICS	PUC / PR - INFORMATICS	UFMG - COMPUTER SCIENCE
19	UFRJ - INFORMATICS	UFRJ - COMPUTER AND SYSTEMS ENGINEERING	UFRN - SYSTEMS AND COMPUTING
20	UFRN - SISTEMAS E COMPUTING	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS	UFRJ - INFORMATICS
21	UFSC - COMPUTER SCIENCE	UNIVALI - COMPUTING	UFRGS - COMPUTING
22	UFSCAR - COMPUTER SCIENCE	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS	UFMG - COMPUTER SCIENCE
23	UFU - COMPUTER SCIENCE	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS	
24	UNB - INFORMATICS	UFMS - COMPUTER SCIENCE	
25	UNIFOR - INFORMATICS APLICADA	PUC-RIO - INFORMATICS	UFPR - INFORMATICS
26	UNISINOS - COMPUTING APLICADA	UFRGS - COMPUTING	PUC / RS - COMPUTER SCIENCE
27	FESP / UPE - COMPUTER ENGINEERING	UFPE - COMPUTER SCIENCE	USP - COMPUTER SCIENCE
28	IME - SYSTEMS AND COMPUTING	PUC-RIO - INFORMATICS	
29	PUC / MG - INFORMATICS	UFMG - COMPUTER SCIENCE	UFAM - INFORMATICS
30	UCPEL - INFORMATICS	UFRGS - COMPUTING	PUC / RS - COMPUTER SCIENCE
31	UECE - COMPUTER SCIENCE	UFF - COMPUTING	UFC - COMPUTER SCIENCE
32	UEL - COMPUTER SCIENCE	UFRGS - COMPUTING	
33	UEM - COMPUTER SCIENCE	UNICAMP - COMPUTER SCIENCE	USP / SC - COMPUTER SCIENCE AND COMPUTATIONAL MATHEMATICS
34	UERN - COMPUTER SCIENCE - UERN - UFERSA	UFRN - SYSTEMS AND COMPUTING	FESP / UPE - COMPUTER ENGINEERING
35	UFG - COMPUTER SCIENCE	USP - COMPUTER SCIENCE	UFMG - COMPUTER SCIENCE
36	UFPA - COMPUTER SCIENCE	UFPE - COMPUTER SCIENCE	
37	UFPB / J.P. - INFORMATICS	UFCG - COMPUTER SCIENCE	PUC-RIO - INFORMATICS
38	UFSM - INFORMATICS	PUC / RS - COMPUTER SCIENCE	UCPEL - INFORMATICS
39	UFV - COMPUTER SCIENCE		
40	UNESP / SJRP - COMPUTER SCIENCE	UNICAMP - COMPUTER SCIENCE	
41	UNIFACS - SYSTEMS AND COMPUTING	PUC-RIO - INFORMATICS	UNICAMP - COMPUTER SCIENCE
42	UNIFEI - COMPUTER SCIENCE AND TECHNOLOGY		
43	UNIMEP - COMPUTER SCIENCE	UFSCAR - COMPUTER SCIENCE	
44	UNIRIO - INFORMATICS	PUC-RIO - INFORMATICS	UFRJ - COMPUTER AND SYSTEMS ENGINEERING
45	UNIVALI - COMPUTING	UFSC - COMPUTER SCIENCE	

References

[1] L. Digiampietri, J. Mena-Chalco, J. de Jésus Pérez-Alcázar, E. F. Tuesta, K. Delgado, and R. Mugnaini, “Minerando e caracterizando dados de currículos Lattes,” in *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, 2012.

[2] L. Bolelli, S. Ertekin, D. Zhou, and C. L. Giles, “Finding topic trends in digital libraries,” in *Proceedings of the 2009 joint international conference on Digital libraries - JCDL ’09*, New York, New York, USA: ACM Press, Jun. 2009, p. 69, ISBN: 9781605583228. DOI: 10 . 1145 / 1555400 . 1555411. [Online]. Available: http://dl.acm.org/citation.cfm?id=1555400.1555411.

[3] N. Kawamae and R. Higashinaka, “Trend detection model,” in *Proceedings of the 19th international conference on World wide web - WWW ’10*, New York, New York, USA: ACM Press, Apr. 2010, p. 1129, ISBN: 9781605587998. DOI: 10 . 1145 / 1772690 . 1772838. [Online]. Available: http://dl.acm.org/citation.cfm?id=1772690.1772838.

[4] N. Kawamae, “Theme Chronicle Model: Chronicle Consists of Timestamp and TopicalWords over

Each Theme," in *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12*, New York, New York, USA: ACM Press, Oct. 2012, p. 2065, ISBN: 9781450311564. DOI: 10.1145/2396761.2398573. [Online]. Available: http://dl.acm.org/citation.cfm?id=2396761.2398573.

[5] M. Jayashri and P. Chitra, "Topic clustering and topic evolution based on temporal parameters," in *2012 International Conference on Recent Trends in Information Technology*, IEEE, Apr. 2012, pp. 559–564, ISBN: 978-1-4673-1601-9. DOI: 10.1109/ICRTIT.2012.6206816. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6206816.

[6] H. Park, E. Kim, K.-J. Bae, H. Hahn, T.-E. Sung, and H.-C. Kwon, "Detection and Analysis of Trend Topics for Global Scientific Literature Using Feature Selection Based on Gini-Index," in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, IEEE, Nov. 2011, pp. 965–969, ISBN: 978-1-4577-2068-0. DOI: 10.1109/ICTAI.2011.166. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6103457.

[7] H. Abe and S. Tsumoto, "Evaluating a method to detect temporal trends of phrases in research documents," in *2009 8th IEEE International Conference on Cognitive Informatics*, IEEE, Jun. 2009, pp. 378–383, ISBN: 978-1-4244-4642-1. DOI: 10.1109/COGINF.2009.5250711. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5250711.

[8] O. Cimenler, K. a. Reeves, and J. Skvoretz, "A regression analysis of researchers social network metrics on their citation performance in a college of engineering," *Journal of Informetrics*, vol. 8, no. 3, pp. 667–682, Jul. 2014, ISSN: 17511577. DOI: 10.1016/j.joi.2014.06.004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1751157714000571.

[9] C. Trucolo and L. Digiampietri, "Uma revisão sistemática acerca das técnicas de identificação e análise de tendências," in *X Simpósio Brasileiro de Sistemas de Informação (SBSI 2014)*, Londrina, 2014, pp. 639–650.

[10] L. Digiampietri, J. Mena-Chalco, G. S. Silva, L. Oliveira, A. Malheiro, and D. Meira, "Dinâmica das relações de coautoria nos programas de pós-graduação em computação no brasil," in *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, 2012.

[11] H. Nakagawa and T. Mori, "A simple but powerful automatic term extraction method," in *COLING-02 on COMPUTERM 2002: Second International Workshop on Computational Terminology - Volume 14*, ser. COMPUTERM '02, Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 1–7. DOI: 10.3115/1118771.1118778. [Online]. Available: http://dx.doi.org/10.3115/1118771.1118778.

[12] L. A. Digiampietri, J. P. Mena-Chalco, P. O. S. Vaz de Melo, A. P. R. Malheiro, D. N. O. Meira, L. F. Franco, and L. B. Oliveira, "Brax-ray: an x-ray of the brazilian computer science graduate programs," *PLoS ONE*, vol. 9, no. 4, e94541, Apr. 2014. DOI: 10.1371/journal.pone.0094541. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0094541.

[13] A. Laender, C. de Lucena, J. Maldonado, E. de Souza e Silva, and N. Ziviani, "Assessing the research and education quality of the top Brazilian Computer Science graduate programs," *ACM Special Interest Group on Computer Science Education*, vol. 40, no. 2, pp. 135–145, 2008.

Caio Cesar Trucolo is a the bachelor in Information Systems at the University of São Paulo (2011). Nowadays he is a master student in Information Systems at the School of Arts, Sciences and Humanities at USP (EACH - USP). He has experience in the area of Computer Science, with emphasis in Information Systems, working mainly in the area of social network analysis and data mining.

Luciano Antonio Digiampietri (corresponding author) is a bachelor (2002) and a PhD (2007) in Computer Science at the Campinas State University (2002). Since April/2008 he has been a researcher professor in Information Systems at the School of Arts, Sciences and Humanities at USP (EACH - USP) and since 2010 he has been a professor at the graduate program in Information Systems at the University of São Paulo. He has experience in the area of Computer Science with emphasis in Computational Biology, Databases and Artificial Intelligence, working mainly on the following topics: scientific workflows, Bioinformatics, data provenance, automatic service composition, data mining and social network analysis.
E-mail: *digiampietri@usp.br*.