A Collaboration and Productiveness Analysis of the BPM Community

Hajo A. Reijers[1], Minseok Song[1], Heidi Romero[1], Umeshwar Dayal[2], Johann Eder[3], and Jana Koehler[4]

[1] Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
{h.a.reijers,m.s.song,h.l.romero}@tue.nl
[2] HP Laboratories, 1501 Page Mill Road, Palo Alto, 94304, USA
umeshwar.dayal@hp.com
[3] University of Klagenfurt, Universittsstrae 65-67, 9020 Klagenfurt, Austria
johann.eder@uni-klu.ac.at
[4] IBM Zurich Research Laboratory CH-8803 Rueschlikon, Switzerland
koe@zurich.ibm.com

Abstract. The main scientific event for academics working in the field of Business Process Management is the International BPM Conference. In this paper, social network analysis techniques are used to unveil the *co-authorship networks* that can be derived from the papers presented at this conference. Links between two researchers are established by their co-authorship of a paper at one of the conference editions throughout the years 2003-2008. Beyond the relations between individual authors, aggregated analyses are presented of the interactions between the institutes that the authors are affiliated with as well as their country of residence. Additionally, the output of individual authors is measured. All analyses are carried out for the individual conference years and at cumulative levels. In this way, this paper identifies the hotbeds of BPM research and maps the progressive collaboration patterns within the BPM community.

1 Introduction

In the introduction of the first proceedings of the BPM conference series[1], Business Process Management (BPM) has been characterized as the study of those methods, techniques, and software that can be used to design, enact, control, and analyze operational processes involving humans, organizations, applications, documents and other sources of information [1]. In line with this view, the technological perspective and the attention for formal methods have consistently been important ingredients of the papers presented in the series. Recently, an inflow of papers and keynotes have appeared that also deal with BPM as a management philosophy instead of a pure technological approach [2,3]. Additionally, papers have been included that concentrate not so much on the design but rather on the empirical evaluation of methods and techniques [4,5].

[1] For an overview of all its editions, see http://www.bpm-conference.org/

U. Dayal et al. (Eds.): BPM 2009, LNCS 5701, pp. 1–14, 2009.

Consequently, the character of the BPM conference extends over a diverse array of subjects and draws from disciplines such as computer science, information systems, management science, artificial intelligence, industrial engineering, software engineering, and economics. In this multidisciplinary context, it is evidently important for BPM researchers to foster and extend collaborations to stay up to date of state-of-the-art developments in their own field and acquire access to complementary fields of expertise.

This paper aims at analyzing the collaborations between individual researchers, their institutes, and the hosting countries as tied to contributions to the BPM conference series. For this purpose, all papers that have been included in the proceedings of the six conference editions between 2003 and 2008 have been considered. The biographical attributes of these papers have been subjected to various Social Network Analysis techniques [6,7,8]. The social network miner and analyzer [9], as part of the ProM framework [10], have been applied to carry out this analysis. The collaboration networks that are the main results identify the hotbeds of BPM research activity and disclose the interconnections between them. The presented analysis of these networks over time also gives an indication of how the BPM community extends and becomes more interwoven.

The structure of this paper is now as follows. In Section 2, we will describe the methodology to collect, conceptualize, analyze and verify the used data. Section 3 will present the results from this analysis. The paper ends with a discussion and a conclusion.

2 Methodology

The methodology employed to identify the underlying relationship between researchers followed several steps that will be detailed in this section. Those steps are described in four subsections: (1) Data collection, (2) Conceptualization of the event logs, (3) Log verification, and (4) Network generation using ProM.

2.1 Data Collection

Our data collection has consisted of the compilation of all the references from the conferences publications. The proceedings of these conferences were published by Springer as volumes in the Lecture Notes in Computer Science series (See Table 1).

The references were obtained from SpringerLink[2], the online access point to Springer's Lecture Notes in Computer Science series, and stored in Reference Manager[3], a bibliographical database system. Entries were extracted from SpringerLink in the RIS format using the "RIS FORMAT (Include ID)" filter available in the ReferenceManager software system. All references were classified as Book Chapters and included the chapter title, authors, start and end page, abstract and a unique ID (DOI). After all entries had been entered, the entire

[2] `http://www.springerlink.com/`

[3] `http://www.refman.com/`

Table 1. BPM Conference Proceedings

Conference	LNCS Volume	Venue	Dates	ISBN
BPM 2003	2678	Eindhoven	June 26-27, 2003	3-540-40318-3
BPM 2004	3080	Potsdam	June 17-18, 2004	3-540-22235-9
BPM 2005	3649	Nancy	September 5-8, 2005	3-540-28238-6
BPM 2006	4102	Vienna	September 5-7, 2006	3-540-38901-6
BPM 2007	4714	Brisbane	September 24-28, 2007	978-3-540-75182-3
BPM 2008	5240	Milan	September 2-4, 2008	978-3-540-85757-0

database has been exported to an Excel Sheet, which was a convenient format to translate the data into the form of a MXML file, which is readable by ProM.

2.2 Conceptualization

To arrive at an MXML file that could be analyzed with the social network analysis toolset [9], a decision needed to be made on how to translate the previously described information into a form that is both convenient and meaningful for analysis. It should be noted that the main purpose of the ProM framework, which was selected for the derivation and analysis of the collaboration networks because of its ease of use and analytical power, is to support the analysis of an *event log*. This is a collection of events that is stored by any kind of transactional information system, e.g., ERP, CRM, or workflow management system [11]. Clearly, the occurrences of successive editions of a conference series is not a business process in the traditional sense, so a mapping of the domain concepts on the business process concepts that are common to ProM had to be established.

In essence, each publication has been considered as a separate case instance that could be processed by one or more authors. Each author performs a unique activity in dealing with a case, as to avoid conflicts in situations where one author participates in more than one publication in the same year. Because we only have the year information and no more fine-grained time notion, it would otherwise be difficult to distinguish these. This simple conceptualization leads to a total of 383 authors and the same amount of activities.

As mentioned, ProM can read files in the MXML format [12], a generic XML format appropriate to representing event log data (see Figure 1). It includes the definition of the process instance with the audit trails including attributes like Workflow Model Element, Event type, Timestamp and Originators. The exact mapping of bibliographical elements to the elements of the MXML format that is used by ProM can now be specified as follows.

The process instance is defined as a publication (book chapter), and each of these has an identical Id which is the DOI, related to the book that this book chapter belongs to. The description linked to every Process Instance is the name of the publication, for instance "Modeling Medical e-Services". For each Process Instance an attribute was defined as a Data item called "ConferenceYear", to facilitate filtering the publications related only to a specific year. For each process

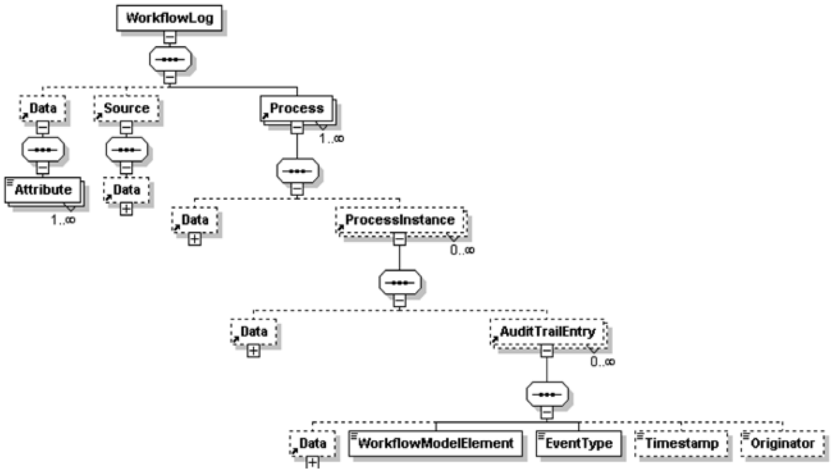

Fig. 1. The MXML format

instance there is an Audit Trail Entry related to each activity followed in the process. We define an activity as the contribution of every author to the publication. This means that a publication performed by two authors will contain two audit trails, each one with a Workflow Model Element that contains the author's name, and only containing "Complete" event types. The timestamp was not defined in the log because the information needed is the year related to the conference in which the paper was published. The latter information is already provided by the Data attribute mentioned earlier. Finally, the Originator is the name of the author.

In a similar way an organizational model file [9] (without log) was created, to relate the authors to their university and countries. The university was identified as the Role and the country as the Organizational Unit. The university or institution associated to each author is where the author works or is affiliated with. The country being considered is the country where the university resides. We found authors that changed their working place during the time of the study. In those few cases, the author was assigned to the university where they published the most and, in case of a tie, was assigned arbitrarily to one.

2.3 Verification

To verify whether no errors have been introduced in the creation of the MXML file, the basic ProM feedback was used to assess high level errors. ProM reports that the log contains one process with 190 cases (publications) and 383 originators (authors). It further indicates a minimum number of events that equals 1, a mean number of 2 and a maximum of 7, representing the number of authors per publication. This information is consistent with the figures that could be obtained in the Excel sheet. Next, we used the "ConferenceYear" attribute to check for potential mistakes in the assignment of the year to cases. We manually inspected each entry, for example, by applying the filter using the ConferenceYear 2003, it is expected to see only the instances with the Id 3-540-44895-0. In case that a different code appears it means that there is a publication from another year misclassified. For this study every filter was manually verified to avoid this error, before further analysis.

Another important point of verification is whether the organizational model links every author to exactly one country and one university. To inspect this aspect we used the Filter / Advanced and selected Replacement Filters to generate separate lists of originators, representing the countries and universities. In this way it was possible to detect for each author a missing assignment or an undesirable swap, e.g. a country that was used as a university.

2.4 Network Generation

The social networks that were generated are all based on the working together metric [13, Definition 4.8]. Informally stated, the metric expresses for each pair of performers whether or not they have performed activities for the same case. In our domain, this means that two authors have worked on the same paper. Clearly, the relation is symmetrical. While our analysis considers only undirected graphs, the figures in this paper show graphs where the relation is show as two arrows with opposed directions. It should be noted that such a visual pair of arrows will account for only a single link. We have established the network on the basis of the working together metric for each individual conference year (leading to six networks) and for each of the cumulative increments for the conferences (leading to another six networks, e.g. for 2003, 2003+2004, 2003+2004+2005, etc.). Additionally, the results on the working together metric between individuals were exported in a matrix representation to NetMiner [14], which is another tool to analyze the social networks. With NetMiner the following analysis measurements were derived for each network:

1. *Density* measures the level of connectedness among the nodes in a network. This measurement is important to compare networks of different sizes. It defined as follows.
   $D = \frac{2|A|}{|N|(|N|-1)}$, where $|A|$ denotes number of arcs and $|N|$ denotes number of nodes.

2. *Inclusiveness* represents a measurement of the connectivity between nodes in the network. It defined as follows.
   $I = \frac{|N|-|N_i|}{|N|} * 100$, where $|N|$ denotes number of nodes and $|N_i|$ denotes number of isolated nodes.
   The isolated nodes are those with a connection degree equals to zero.

Finally, these networks and analysis measures were also established by aggregating the authors to the level of *research institutes* and *countries*. For example, a co-publication at the BPM 2005 conference between an author from Macquarie University in Australia and another author from Tilburg University in the

Netherlands leads to a relation between the respective universities in the institute network for 2005, as well as a relation between Australia and the Netherlands in the countries network of 2005.

It should be noted that many more relations exist to create social networks, as well as many more analysis measurements to evaluate these. For an overview, see [13]. However, the working together metric is the most useful for our purpose. The networks resulting from our analysis will be presented in the next section.

3 Results

The results in this section will be successively presented on the individual (author) level, the institute level, and the country level.

3.1 Individual Level

To provide an initial idea of the networks that were created, Figure 2 represents the social networks on the level of co-authors for publications in the year 2003, while Figure 3 represents the same type of social network considering all publications of the years 2003 to 2008 combined. While it is clear that the network has expanded considerably over the years, the data in Table 2 gives a better insight into this development.

Over the years, the social network has grown as a whole, both with respect to the publications (going up from 23 in 2003 to 190 accumulated over all the years)

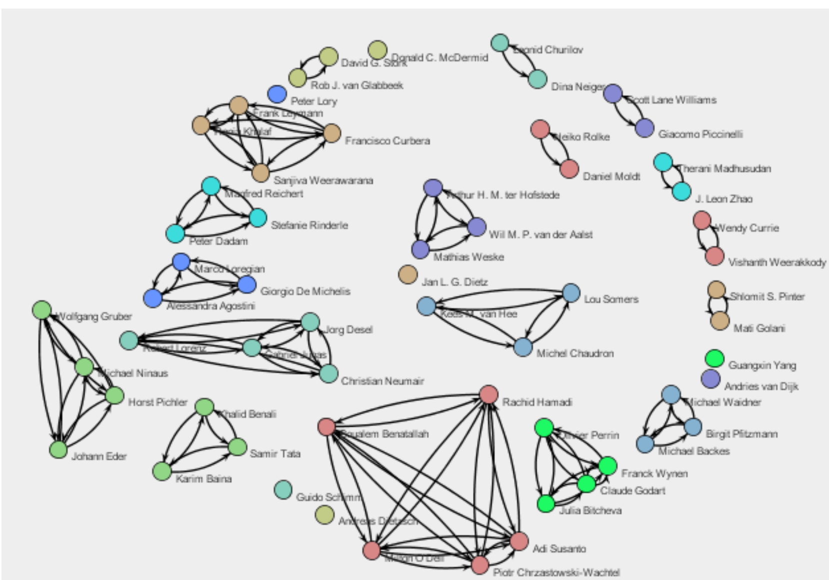

Fig. 2. Social network on individual level for 2003 (working together metric)

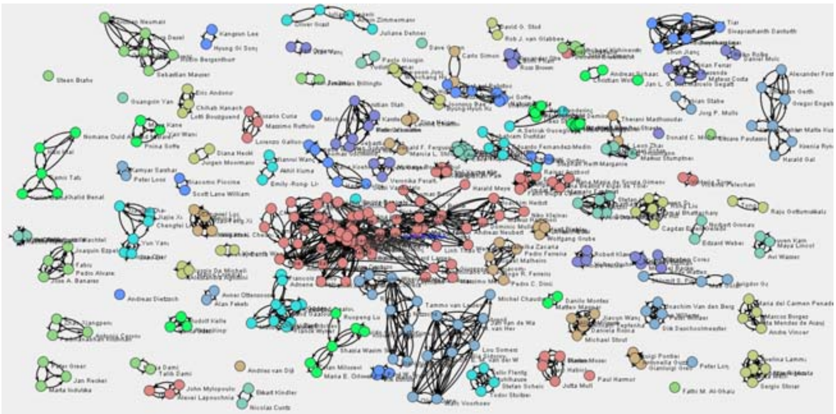

Fig. 3. Social network on individual level for 2003-2008 (working together metric)

Table 2. Cumulative statistics individual level (working together metric)

	2003	2003-2004	2003-2005	2003-2006	2003-2007	2003-2008
Number of publications	23	45	86	128	158	190
Number of authors	60	99	190	271	335	383
Number of links	59	103	217	372	463	555
Inclusiveness (%)	88.3	91.9	95.8	96.3	96.7	96.3
Network Density	0.032	0.021	0.012	0.010	0.008	0.008

and individual authors (going up from 60 in 2003 to 383 authors accumulated over all the years). The factors of growth are over 8 and 6 respectively. The inclusiveness of the cumulative networks also grows, from 88.3% to around 96%. In other words, the relative number of authors purely publishing by themselves drops as a relative measure, but it should be noted that this was already a small minority from the start. At the same time, the density of the network can be seen to drop, from 0.032 to 0.008. This means that the inflow of papers by new authors is not matched by a corresponding increase of new collaborations that become possible.

From all the 190 authors that have published one or more papers in the conference series, it can be determined how many papers they contributed overall. The authors that published 4 papers or more are shown in Table 5. It can be seen that the top contributors generate a considerable but not excessive share of the overall number of papers, which can be seen as a sign of academic health for the conference series.

3.2 Institute Level

To provide an understanding of the evolution in collaboration patterns between the institutes, several cumulative networks are now presented in succession.

Table 3. Contributions individual authors

Originator	Publications (absolute)	Publications (relative to total)
Wil M.P. van der Aalst	12	6.32%
Manfred Reichert	9	4.73%
Hajo A. Reijers	8	4.21%
Mathias Weske	8	4.21%
Gero Decker	6	3.16%
Marlon Dumas	6	3.16%
Claude Godart	6	3.16%
Arthur H.M. ter Hofstede	6	3.16%
Jan Mendling	6	3.16%
Stefanie Rinderle	6	3.16%
Peter Dadam	5	2.63%
Kees M. van Hee	5	2.63%
Monique H. Jansen-Vullers	4	2.11%
Chengfei Liu	4	2.11%
Olivier Perrin	4	2.11%
Natalia Sidorova	4	2.11%
Marc Voorhoeve	4	2.11%
Xiaohui Zhao	4	2.11%

Table 4. Cumulative statistics institute level (working together metric)

	2003	2003-2004	2003-2005	2003-2006	2003-2007	2003-2008
Number of institutes	33	54	91	114	140	154
Number of links	13	29	49	79	104	128
Inclusiveness (%)	51.5	66.7	71.4	72.8	75.7	76.6
Network Density	0.024	0.020	0.011	0.012	0.010	0.010

Figures 4, 5, 6, and 7 respectively show the network in 2003, the cumulative network over the years 2003 to 2006, the cumulative network over the years 2003 to 2007, and the cumulative network over the years 2003 to 2008.

In 2003, after the first conference, 7 collaboration groups could be identified (see Figure 4). By 2006, the number of groups had increased to 27 groups (see Figure 5). Among them, five big groups can be identified: two Dutch-German groups, one French group, one German-USA group, and one multi-continental group. By 2007, one of the Dutch-German group was notably extended (see Figure 6). The situation by 2008 can be seen in Figure 7. By that time, the two Dutch-German groups were merged and Eindhoven University of Technology became the bridge between these two formerly separate groups. The German-USA group had also grown by this time and the IBM JT Watson research center had become the center of it.

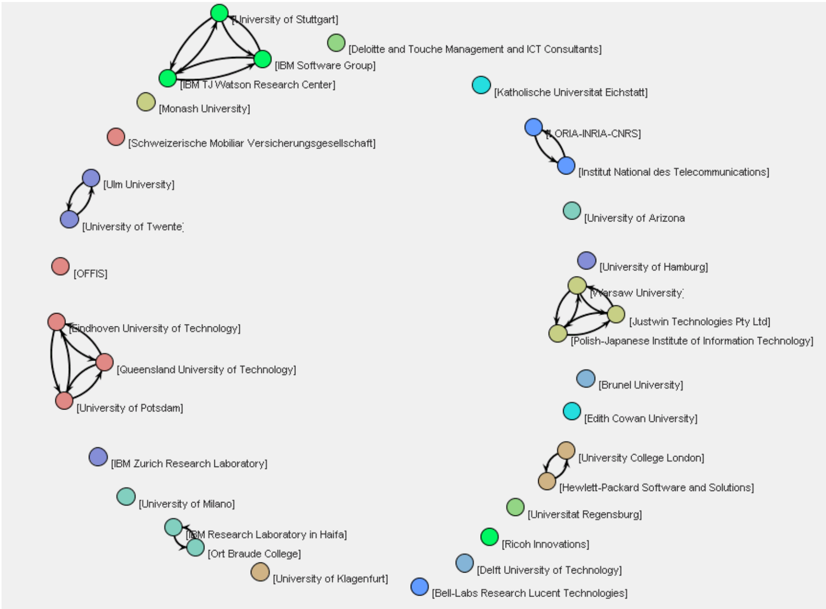

Fig. 4. Social network on institute level for 2003 (working together metric)

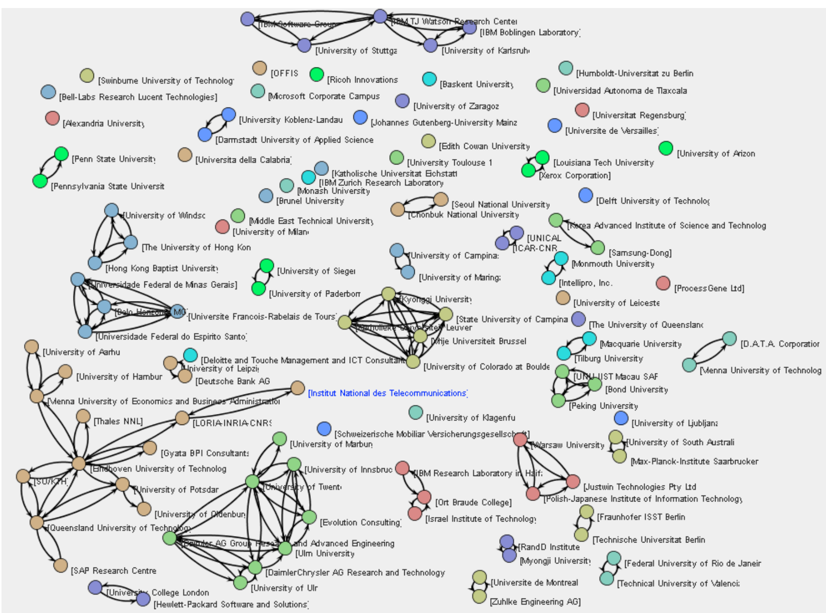

Fig. 5. Social network on institute level for 2003 - 2006 (working together metric)

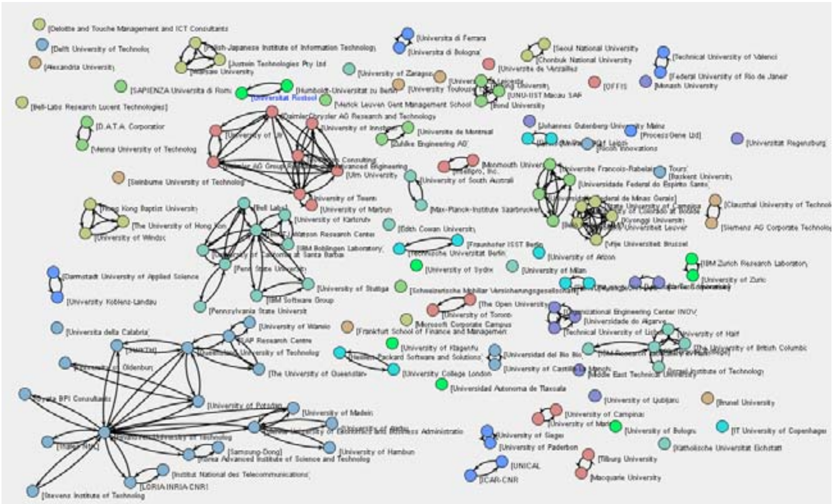

Fig. 6. Social network on institute level for 2003 - 2007 (working together metric)

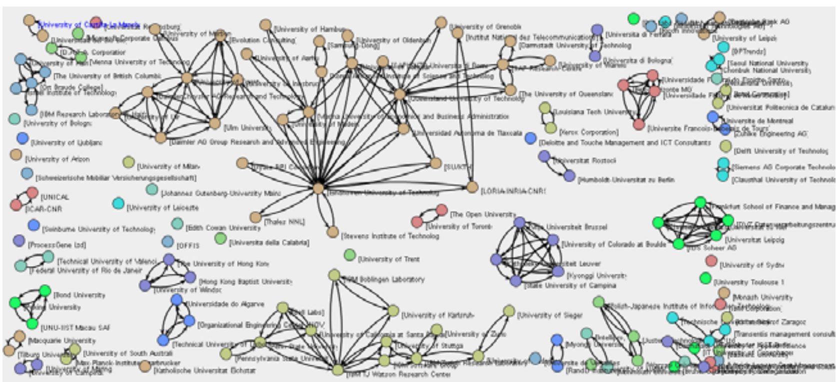

Fig. 7. Social network on institute level for 2003-2008 (working together metric)

Table 4 gives the data for the cumulative years. What can be noted is that the number of institutes grows considerably over the years (from 33 to 154), a factor greater than 4, but this is not as strong an increase as the number of papers or new authors over this period (see the previous subsection). Yet, the growth of the inclusiveness of the network, from 51.5% to 76.6%, is more impressive than the network on the individual level discussed previously. This means that the cooperation between people from different institutes grows stronger than the cooperation between individuals. Similar to the analysis level of individual authors, these increasing figures are accompanied by a drop of the network density throughout this development (from 0.024 to 0.010). This also suggests that the

Table 5. Contributions institutes

Originator	Publications (absolute)	Publications (relative to total)
Eindhoven University of Technology	72	37.89%
Queensland University of Technology	27	14.21%
LORIA-INRIA-CNRS	23	12.10%
University of Potsdam	21	11.05%
University of Ulm	20	10.53%
Swinburne University of Technology	15	7.89%
IBM Zurich Research Laboratory	13	6.84%
Humboldt-Universitaet zu Berlin	11	5.79%
University of Stuttgart	11	5.79%
SAP Research Centre	10	5.25%
The University of Queensland	10	5.26%

entry of new institutes in the BPM field is not matched with a corresponding growth of new collaborations on this level.

Finally, Table 5 shows the institutes that have generated 10 or more contributions to the various editions of the conference series. Here, it can be seen that the institute leading the table, Eindhoven University of Technology, has a very large part in the overall production of papers with a production that is over twice as big as that of the number two, Queensland University of Technology, and over three times as big of the number three, LORIA-INRIA-CNRS.

3.3 Country Level

Collaborations between authors from institutes in different countries are visualized with the network in Figure 8, while the quantitative data of the various cumulative networks are given in Table 6.

Over the years, the number of participating countries has more than doubled, growing from 11 countries in 2003 to 25 accumulated over the whole conference series. The inclusiveness of the network has also grown, from 54.5% to 88.0%, which is comparable to the growth of the network at the institute level over the same period (see the previous subsection). Interestingly, and in contrast to the networks that have been discussed previously for the other analysis levels,

Table 6. Cumulative statistics country level (working together metric)

	2003	2003-2004	2003-2005	2003-2006	2003-2007	2003-2008
Number of countries	11	15	17	23	25	25
Number of links	6	8	11	24	32	39
Inclusiveness (%)	54.5	66.7	70.6	73.9	80.8	88.0
Network Density	0.109	0.076	0.081	0.094	0.106	0.130

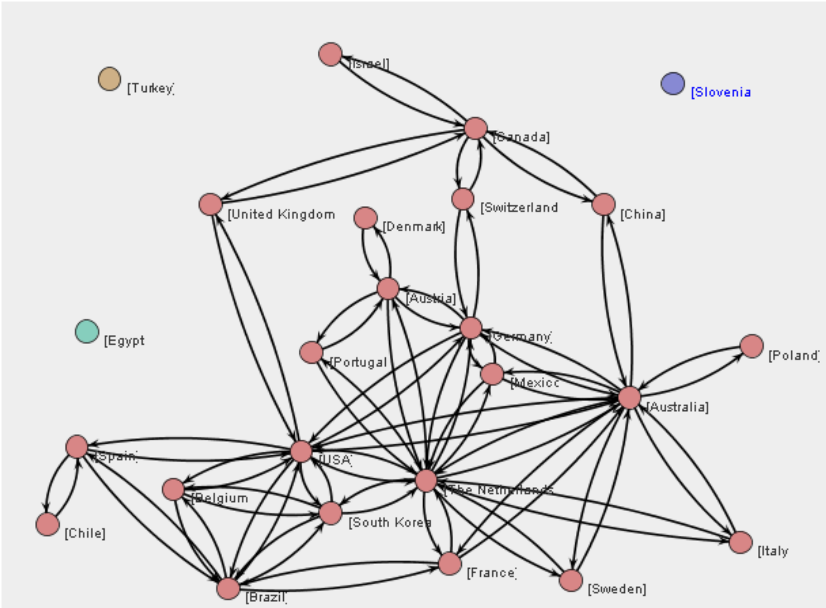

Fig. 8. Social network on country level for 2003-2008 (working together metric)

Table 7. Contributions countries

Originator	**Publications (absolute)**	**Publications (relative to total)**
Germany	127	66.84%
The Netherlands	87	45.78%
Australia	75	39.47%
USA	46	24.21%
France	33	17.37%
Italy	22	11.58%
Switzerland	18	9.47%
Austria	16	8.42%
Israel	15	7.89%
South Korea	14	7.37%
Belgium	12	6.32%
Spain	11	5.79%
Brazil	10	5.26%

the density of the network is rather stable over the years. In other words, the inflow of contributions from authors of 'new' countries goes along with an actual exploitation of such inter-country collaborations opportunities.

As a final analysis result, the countries are shown in Table 7 of which the hosted BPM researchers have contributed the most papers to the conference series. Note that only countries with 10 or more generated publications are shown. From this table, Germany, The Netherlands, and Australia emerge as dominant providers of content for the BPM conference series. This could well be expected from the results shown previously in Table 5, which mainly shows institutes located in these countries. What is somewhat surprising is the position of the USA in fourth place, while no American institute appears in Table 5. This seems to suggest that the BPM research in the USA is much more scattered over various institutes than is the case in Europe or Australia.

4 Discussion and Conclusion

The analysis of the collaboration patterns behind the papers on the BPM conference series shows a growth of collaborations at all the levels of analysis: author, institute, and country level. On the country level the increase of entrants result in a utilization of the new research ties that potentially become available with the new entrants. This can be taken as a sign that the BPM conference series appears as a genuinely international forum. At the individual and institute level this type of expansion is not so apparent. This can perhaps be more or less expected on an individual level, where young researchers will be entering this field with a small collaboration network. However, the lack of growth in cooperation patterns between the institutes over the years can be seen as somewhat worrisome if this is taken as an indication of researchers at the various institutes of favoring "in-house" research.

From the analysis of the numbers of contributions, one can establish that Germany, The Netherlands and Australia are the leading countries in their participation in this research area. Several universities are actively participating in Germany including the University of Potsdam, Ulm University, Humboldt-University of Berlin, and the University of Stuttgart, with shares of 5 up to 11% of the total number of papers. In the Netherlands, the leading university is Eindhoven University of Technology, with strong contributions of several authors amounting to a share of publications in the range of 38%. In Australia, a broader array of universities is contributing to the field, notably Queensland University of Technology, Swinburne University of Technology, and The University of Queensland, with respective shares of 14.21%, 7.89% and 5.26%.

To conclude, the co-authorship networks derived from the publications in the BPM international conferences 2003-2008 help to recognize how the community of researchers is built around this research area, and what influence is exerted by their universities and the countries that they are affiliated with. The paper is also a nice illustration of the versatility of the BPM tools that are developed by the same community. After all, it is the ProM tool that was developed for supporting various process mining techniques that was used for this bibliographic study. It is our hope that the presented results provide the members of the BPM community with some new insights and will encourage them to link up with others, in pursuit of breakthrough research.

References

1. van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M.: Business process management: A survey. In: van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003)
2. Willaert, P., Van den Bergh, J., Willems, J., Deschoolmeester, D.: The process-oriented organisation: A holistic view developing a framework for business process orientation maturity. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 1–15. Springer, Heidelberg (2007)
3. Harmon, P.: Business process management: Today and tomorrow. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 1–1. Springer, Heidelberg (2008)
4. Reijers, H.A., Song, M., Jeong, B.: On the performance of workflow processes with distributed actors: Does place matter? In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 32–47. Springer, Heidelberg (2007)
5. Siegeris, J., Grasl, O.: Model driven business transformation – an experience report. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 36–50. Springer, Heidelberg (2008)
6. Burt, R., Minor, M.: Applied Network Analysis: A Methodological Introduction. Sage, Newbury Park (1983)
7. Scott, J.: Social Network Analysis. Sage, Newbury Park (1992)
8. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
9. Song, M., van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. Decision Support Systems 46(1), 300–317 (2008)
10. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W(E.), Weijters, A.J.M.M.T., van der Aalst, W.M.P.: The proM framework: A new era in process mining tool support. In: Ciardo, G., Darondeau, P. (eds.) ICATPN 2005. LNCS, vol. 3536, pp. 444–454. Springer, Heidelberg (2005)
11. van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Medeiros, A., Song, M., Verbeek, H.M.W.: Business Process Mining: An Industrial Application. Information Systems 32(5), 713–732 (2007)
12. Guenther, C., van der Aalst, W.M.P.: A Generic Import Framework For Process Event Logs. BPM report (2006), `http://is.tm.tue.nl/staff/wvdaalst/BPMcenter/reports.htm`
13. van der Aalst, W.M.P., Reijers, H.A., Song, M.: Discovering Social Networks from Event Logs. Computer Supported Cooperative work 14(6), 549–593 (2005)
14. NetMiner: NetMiner Tool User's Guide (2009), `http://www.netminer.com`