A Retrospective Socio-Semantic Analysis of the PRO-VE Conferences

António Lucas Soares, Maria Almeida, Filipa Ramalho

A Retrospective Socio-Semantic Analysis of the PRO-VE Conferences

António Lucas Soares[1,2], Maria Almeida[1] and Filipa Ramalho[1]

[1] INESC TEC Porto, Portugal
[2] Faculdade de Engenharia da Universidade do Porto, Portugal
{asoares@inescporto.pt, filiparamalho@gmail.com, mi.almeida88@gmail.com}

Abstract. The IFIP WG 5.5 Working Conferences on Virtual Enterprises (PRO-VE) has created, in 14 editions, a remarkable scientific and professional community intersecting several disciplines and resulting in a new one: collaborative networks. In this paper we present the results of a retrospective study of the conceptual system evolution of PRO-VE providing an instrument to reflect about the field's past and future. Our approach was socio-semantic trying to devise the influence between the researchers social network and the evolution of the conceptual system. Firstly, we made a terminological analysis of every PRO-VE proceeding resulting in a picture of the main concepts used in each edition and their relative importance. Then, we used social network analysis techniques to conclude about the influence of the researchers on the conceptual system evolution. The results suggest a relatively stable set of concepts influenced by a network of core researchers. However, some marked evolution in the relative importance of the concepts can be identified.

Keywords: socio-semantic network; social-networks analysis; PRO-VE conference

1 Introduction

Scientific conferences series are somehow archetypes of researchers communities creation, evolution and death. If we envisage conference research communities as strong knowledge communities [1], we assist at knowledge creation and dissemination in varying degrees according to its (time-based) relevance. If we look at them as collaborative networks [2] we assist at social capital creation and dismissal and trust building and destruction influencing directly and indirectly single or collective research activities. It seems then clear that the dyad knowledge-collaboration is not separable in the studies of conference research communities even if the majority of the published research on the subject focus on one dimension or the other. The PRO-VE conference series reaches this year its 15th edition. This round number was the spark that motivated us to study the PRO-VE community evolution in two intertwined dimensions: the conceptual system and the social system. This is not "just another bibliometric or scientometric study" of conference proceedings. Although the later seem to be much less frequent than the studies based on scientific journals, our goal was to characterise qualitatively and in detail the evolution of the conceptual system of PRO-VE. Yet, the conceptual system evolution cannot be studied detached from the community of

researchers, as the relationships among them and the patterns of use of the concepts are fundamental for the understanding of such evolution. Our stance is thus socio-semantic, and the object of study is the socio-semantic network. The approach followed by us is original in several ways. The study of knowledge networks as socio-semantic networks is not new although not widespread [1]. The study of the conceptual system within these studies is normally restricted to keyword identification or extraction. Our approach relied in a comprehensive terminological and conceptual semi-automated analysis that enabled the re-construction of the de facto conceptual system of each conference edition. The study of the social network usually rely on the identification of co-authorship and citation analysis. Our approach relies on 2-mode (researchers–concepts) social networks analysis (SNA), enabling the study of the socio-semantic network. The rest of the paper is organised as follows: in the next section we briefly review some important related work, then we describe in detail the research design followed by the presentation of the results. Finally, the results are discussed, the limitations of the study pointed out and the paper is concluded. It is also important refer the authors motivation for this study. One of the co-authors is an active member of the PRO-VE community having participated in all but three conference editions. The other co-authors are young information science professionals, interested in understanding better the study of scientific communities.

2   Related Work

There has been recently a growing body of work on the analysis of communities and their temporal evolution in dynamic networks [3]. In the literature we can find some research studying scientific conferences, namely bibliometric and scientometric studies. [4] analysed semantically the proceedings of a conference in the area of international business. The study was done with a corpus that joined three years of proceedings. The main objective was to represent the current academic interest in the area, grouping keywords and analysing the most studied areas in the articles. Another bibliometric study is decribed in [5] analysing one year of a conference (DESIGN 2012). This study used the citations in the papers to describe citation trends by field, type of work and its distribution.

Within the PRO-VE conferences two papers addressed the social network and the conceptual evolution of the PRO-VE community. In [6] the authors exploited the concept of eigenvector centrality starting from the papers proceedings (2005-2009) and propose a weighted multi-hypergraph model to study the (eigenvector) centrally of PRO-VE authors and research topics. The model is roughly equivalent to our 2-mode network being the concepts taken from reference models and some text processing made to identify concepts through the papers "keywords". As we will see bellow, our research go beyond this model by extracting terms referring to domain concepts in the form of multi-word phrases that constitute a substantial majority of all technical vocabulary. We also propose to reuse a unified socio-semantic model borrowed from [1]. The very same approach to the domain concepts identification from the proceedings corpus, distinguishes our research from [7] whose main limitation is to consider only single-word terms.

The empirical and computational study reported in [1] suggests that the dynamics of a community, i.e. communities of scientists, software developers, wiki contributors and

others, can be amply described as the coevolution of a social and a socio-semantic network. The authors presents a theoretical framework based on a social network and a socio-semantic network. In this study we can see a description of the dynamics of a community comprising the social structure and socio-semantic structure. It is exactly in the work of Roth that we found our theoretical reference. A *social network* is denoted by $\mathbf{G}=(\mathbf{S},R^S)$ where $\mathbf{S}$ is the agent set and $R^S = R \subseteq S \times S \times N$ denotes the set of dated links: a link $l=(s,s',t) \in R^S$ means that s is related to s' at t. Considering now $\mathbf{C}$ a set of semantic objects which we call "concepts" and which correspond here to terms or noun phrases considered as atomic units, a *socio-semantic network* is thus defined formally as a network $\mathbf{G^C}$, made of actors of $\mathbf{S}$, concepts of $\mathbf{C}$ and links between these elements: R^C thus denotes the use of concepts by actors: an actor is linked to concepts he/she mentioned (e.g., in a paper from a proceeding). Thus, $R^C \subseteq S \times C \times N$, and a link $l^C =(s,c,t) \in R^C$ means that s used c at t. Although Roth studies computationally the coevolution of social and socio-semantic networks, in this paper we use only the socio-semantic network to study the evolution of the PRO-VE conceptual system.

3   Research Design

Recalling the research relevance and goals introduced in the beginning of this paper, our aim was to know more on how the PRO-VE community evolved regarding the technical and scientific topics (concepts) addressed in 14 editions of existence of the conference. Furthermore, we wanted to understand if there were traces of influence from the structures of collaboration of the researchers in the number and intensity of the addressed concepts. The specific research questions were:

RQ1. What were the main technical and scientific concepts addressed *de facto* in each of the PRO-VE conference editions? Are there any patterns of evolution?

RQ2. Is it possible to identify sub-sets of researchers having in common the use of similar sets of concepts (socio-semantic structures)? How did these structures evolved?

This research was fundamentally exploratory, and being so there were not strong hypothesis to advance. Nevertheless, a couple of them were advanced as preliminary answers to the research questions:

H1.1. There is a core set of concepts used in all the editions of the conference;

H1.2. The variability of the first ranked concepts after the core is high;

H2.1. A core set of concepts is shared by the majority of the researchers;

H2.2. There are groups of researchers that are likely to influence the adoption of new concepts.

The research strategy to answer the above questions recurred to two perspectives on the PRO-VE conferences: a conceptual (semantic) perspective and a social network perspective. The idea was to firstly identify and analyse the concepts that the researchers *de facto* used in the conference papers using terminology and natural language processing tools and then to identify and analyse the socio-semantic structures (sets of researchers and concepts they used) emerging from the conceptual analysis. This approach has several dimensions. Regarding the epistemology, our study is essentially interpretative, although strongly based on fact discovering through quantitative methods. Regarding the methods, the approach is multi-disciplinary as it was informed by and tools were used from terminology and natural language

processing, along with social networks analysis techniques. The data sources used were the 14 PRO-VE proceedings in electronic (pdf) format. From these data sources, three basic data-sets (per year) were extracted: authors, papers/authors, conference-title/table-of-contents/editorial. The overall research design is depicted in the figure 1.

3.1   Terminological and Conceptual Analysis

The goal of this phase was to identify and characterise the conceptual system suggested by each of the fourteen PRO-VE conference proceedings and to identify some patterns of evolution. For that, a textual corpus of the conference proceedings was created and organised by year (for the sake of economy of space we will not describe the data-sets preparation methods). A twofold approach was followed: first, a NLP tool was used to extract the candidate terms[1] from each proceeding; second, a terminological and conceptual analysis of the candidate terms was performed to result in a ranked list of terms (concepts) for each conference year. The term identification was made with the help of an automatic term recognition (ATR) service: the TerMine service provided by NaCTeM[2]. TerMine is based on a domain-independent method (*C-Value*) for the automatic extraction of multi-word terms, from machine-readable special language corpora. As described in [8], *C-value* combines linguistic and statistical analyses with an emphasis placed on the statistical part. The linguistic analysis enumerates all candidate terms in a given text by applying part-of-speech tagging, extracting word sequences of adjectives/nouns based, and stop-list. The statistical analysis assigns a termhood to a candidate term by using the following four characteristics: (i) the occurrence frequency of the candidate term, (ii) the frequency of the candidate term as part of other longer candidate terms, (iii) the number of these longer candidate terms, and (iv) the length of the candidate term.
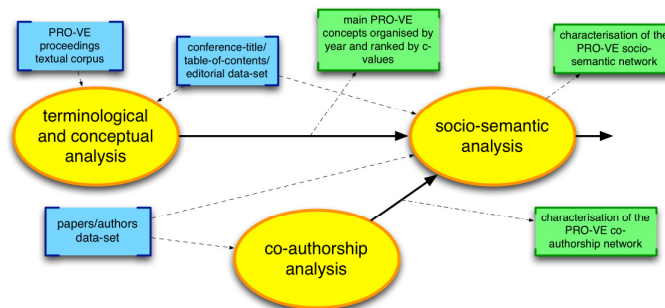
Figure 1 – Overall research design

   From the several ATR tools we tested, TerMine showed the best performance by maximizing precision and recall, while minimizing noise. Most of the other tools available are optimized for recognizing single word terms, which is not what was

---

[1] In terminology, a (technical/scientific) term is a possible designation for a concept considered to be part of a given domain. A term must have the explicit or tacit agreement of the technical/scientific community dealing with the concept. A candidate term is a term that was identified in some source and that needs the validation of a domain specialist to be considered a domain term.

[2] The National Centre for Text Mining (NaCTeM) - http://www.nactem.ac.uk/

required for analysing the PRO-VE proceedings. In fact, technical terms consist mostly of noun phrases containing adjectives, nouns, and occasionally prepositions. The discourse properties of the PRO-VE community are patterns of repetition that distinguish noun phrases that are technical terms, especially multi-word phrases that constitute a substantial majority of all technical vocabulary, from other types of noun phrase [10]. The TerMine service only provided a ranked list of candidate terms. The next and crucial step was to analyse the lists by specialists. Here, specialists are PRO-VE participants. We chose three specialists, actively working in the PRO-VE topics, with different degrees of participation in the fourteen editions (12, 8, and 3 participations). The conceptual analysis involved the following tasks performed by a specialist in the PRO-VE domains: (i) to go through the first 100 candidate terms returned by TerMine for each year, to clean non-terms and discard non-relevant terms; (ii) to make two rankings for each year, one of the first 10 and other of the first 20 terms using the C-Value termhood returned by TerMine; (iii) to fine-tune the rankings for each year by looking over the proceedings chapters and editorials[3]. This phase finished with the validation of each ranking by the above mentioned specialists.

3.2 Social network analysis

The goal of the social network analysis phase in our approach was to provide a description of the socio-semantic structure of the PRO-VE scientific community. SNA techniques were used to (i) to identify the network of relationships between authors and the concepts used in each conference proceeding, (ii) to identify meaningful sub-structures of concept-author relationships and (iii) to relate the co-authorship network with the socio-semantic network. For this we used mostly two-mode (2-mode) network analysis techniques. Due to the effort required to prepare the data-sets and to perform 2-mode network analysis, we selected a sample of 4 PRO-VE proceedings: 1999, 2005, 2009 and 2013. For each of the selected years, a rectangular data matrix of authors (rows) by concepts (columns) was built[4], according to the following protocol: (i) the set of concepts to be used is the top 10 ranking for the year in consideration; (ii) for each of the concepts, the full text of the proceedings was searched to count how many times the concept appears in each paper; (iii) this number was cumulatively added to intersection of the paper authors (row) with the concept (column). For each year, the resulting authors-by-concepts matrix is calculated as:

$$v_{i,j} = \sum_{p_{k,i} \in P} f(p_{k,i}, c_j)$$

where $v_{i,j}$ is the i,j cell value, $p_{k,i}$ is a paper co-authored by author a_i and included in the set of papers P of the proceedings, c_j is the concept being look up and f a function that returns the frequency of presence of c_j in paper p_k. Exclusion conditions were defined so as to discard papers were a term appears in a non relevant way. A

---

[3] For the conceptual analysis phase it would be expectable to use the keywords assigned by the authors to their papers. In PRO-VE, only from 2009 on keywords where used to describe the paper subject. For the sake of consistency, we decided not to use the keywords in the conceptual analysis.

[4] In our case the 2-mode network is represented by a "author-by-concept" (rectangular) matrix. This 2-mode network can be transformed in two 1-mode networks: a "author-by-author" and a "concept-by-concept". These are represented by square matrices.

$\min \left(f\left(p_{k,i}, c_j\right)\right) = 3$ was established. Also, if the concept appeared only in the title or abstract and not in the paper body, it would be excluded i.e, $f\left(p_{k,i}, c_j\right) = 0$.

Usually, 2-mode data in SNA can be analysed either as 1-mode or 2-mode. We used both, but focusing on algorithms and representations that highlight "sub-structures" instead of the characteristics of nodes (authors or concepts) within the network [9]. For our purposes, we selected the 1-mode networks to be run by the *k-core* analysis and the *core-periphery* analysis. The 2-mode network was analysed in terms of *degree centrality*. A *k-core* is a maximal group of actors, all of whom are connected to some number (k) of other members of the group. It allows actors to be included in the group if they are connected to other k members, regardless of how many other members they may not be connected to. By varying the value of k (that is, how many members of the group do you have to be connected to), different pictures can emerge [9]. In the case of our 2-mode network of authors-by-concepts, the k-core algorithm finds the maximal group of authors that use k concepts and concepts that are used by k authors. The *core-periphery* structure is an ideal typical pattern that divides both the rows and the columns into two classes. One of the blocks on the main diagonal (the core) is a high-density block; the other block on the main diagonal (the periphery) is a low-density block. In our case, when applying the core-periphery model to the author-by-author data, the model seeks to identify a set of authors who have high density of ties among themselves (the core) by sharing many concepts in common, and another set of authors who have very low density of ties among themselves (the periphery) by having few concepts in common. We chose degree centrality in terms of socio-semantic relations because, as suggested by [1] it may be loosely interpreted as semantic capital: in other words, the number of concepts an actor has used is likely to render the variety of topics s/he has dealt with in the epistemic network.

The data sets were analysed by two software packages: NetMiner 3 from Cyram Co. Ltd. and UCINET 6 from Analytical Technologies.

4   Results

4.1   The PRO-VE Conceptual System Evolution

From the terminological and conceptual analysis described in the previous section resulted the top-10 and top-20 rankings of concepts for each of the PRO-VE proceedings. An extract of them is shown in figure 2. In the top-10 ranking, 43 distinct concepts were identified.

concepts 2009	c-value	concepts 2010	c-value	concepts 2011	c-value	concepts 2012	c-value
virtual breeding environment	484,5	collaborative network	586,8	virtual organization	430,1	collaborative network	426,6
collaborative network	340,1	virtual breeding environment	303,5	collaborative network	404,3	virtual breeding environment	310,6
virtual organization	336,1	virtual organization	256,3	social capital	184,5	virtual enterprise	245,9
virtual enterprise	268,8	supply chain	236,0	virtual breeding environment	175,8	business process	157,1
senior professional	175,3	business process	180,9	business process	133,3	business model	104,3
supply chain	173,3	supply network	89,2	supply chain	108,4	virtual organization	95,3
social network	101,7	information system	87,6	social network	87,1	cloud computing	76,3
business process	97,6	senior professional	76,0	virtual enterprise	84,3	supply chain	70,2
project management	76,6	collaborative network member	75,9	supply network	64,8	service composition	67,6
virtual community	57,3	collaborative process	72,5	collaborative network member	55,7	collaborative system	54,0
knowledge management	56,8	business model	72,3	process model	54,3	process model	47,0
collaborative process	56,6	service provider	67,9	performance management	53,6	information system	45,7
reference model	55,9	virtual enterprise	65,5	performance measurement	52,6	service dominant strategy	45,5
business model	53,6	business service	64,8	business opportunity	52,0	collaborative process	45,4
web service	52,7	performance measurement	63,1	collaborative relationship	42,0	service provider	43,9
virtual team	50,6	network member	60,3	network member	41,1	collaborative relationship	43,0
epal environment	50,0	social capital	57,3	collaborative networks	39,8	reference model	42,0
VBE member	47,0	business opportunity	53,6	knowledge management	39,2	VBE ontology	41,8
innovation process	44,0	reference model	52,5	maturity model	38,1	software service	40,7
information system	43,9	security awareness	51,2	open supply web	35,5	business case	38,1
295000		290000		199000		215000	

Figure 2 – Extract of the rankings of concepts for each PRO-VE proceeding

Analysing the rankings for the whole existence of PRO-VE, it is possible to extract a conceptual core i.e., the *foundational concepts* of PRO-VE defined as the maximal sets of concepts that span a maximum number of years. The results are the following:

$C_{max1} = \{virtual\ organization, virtual\ enterprise, business\ process, supply\ chain\}$
(span: 14 years)
$C_{max2} = \{collaborative\ network\}$ (span: 11 years)
$C_{max3} = \{virtual\ breeding\ environment, business\ model\}$ (span: 10 years)

The foundational concepts represent approximately 16% of the total distinct top 10 concepts. We call *contingent concepts* to the concepts that were included in the top-10 ranking for at most 2 consecutive years and then disappeared. These concepts are likely to appear and disappear due to combinations of factors such as the "fashionable" research topics and technologies, major on-going projects or even the suggested topics from the "Call for Papers". The contingent category represents 43% of the total. A third class of concepts are those who are neither foundational nor contingent. We call these *regular concepts*. These are:

C_{reg}={*knowledge management, information system, business opportunity, virtual team, virtual community, web service, VBE member, VO creation, social capital, social network, CN member, VBE member, VE member, collaborative process, supply network, service design*}

Within the foundational concepts, it is interesting to analyse the case of *collaborative network*. This concept is mentioned for the first time in 2000 by Devine and Filos and more consequently used in 2002 by de Joode. In 2003 papers referring to this concept continued to grow and finally in 2004 *collaborative network* was definitely consolidated (top-10). From 2005 on, the *collaborative network* concept acquired a status of the most important concept within the PRO-VE community by being proposed to name a new scientific discipline [2]. Another interesting result comes from looking at the average of the C-values for each of the core concepts (see figure 3). *Virtual Breeding Environment* has the highest C-value average which is indicative of the importance of this concept within PRO-VE. Figure 3 represent the PRO-VE conceptual

core together with a set of relationships between concepts. These relationships were established by the three specialists after debate. Of course this is just a possible set. Other specialists could come out with a different set. It would be not odd to look at this concept map as an upper-level ontology of PRO-VE.
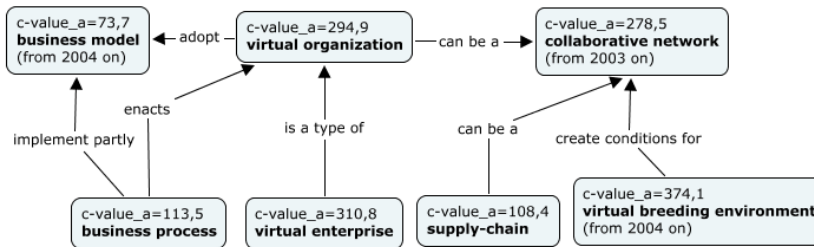

Figure 3 – A possible upper-level ontology created from the PRO-VE foundational concepts (c_value_a=average C-value)

4.2   The PRO-VE Socio-Semantic Network Evolution

As mentioned in the previous section a sample of four years was selected to analyse the PRO-VE socio-semantic network: 1999, 2005, 2009 and 2013. For each of these years the *k-core* and the *core-periphery* analysis were run respectively on the actors-by-concepts and actors-by-actors data sets. *Degree centrality* was calculated for the 1-mode networks (authors-by-authors and concepts-by-concepts) generated from the 2-mode one. Figure 4 shows a spring-like representation of the socio-semantic network. We can qualitatively observe which concepts are more used and, conversely, which authors use more concepts. The representation of the network follows an algorithm that group spatially the nodes having more connections. This can be also observed by looking at the size of the nodes which is proportional to the degree centrality of the actor or concept.

   An extract of the degree centrality values for the first five authors and concepts is shown in table 2. This representation enables a qualitative appreciation of the socio-semantic network. Using simultaneously the scores of degree centrality both for the authors and the concepts, a more detailed assessment can be made. The centrality of some authors and some concepts is immediately evident. Although not shown in the network representation, each edge linking an author to a concept is weighted. This explains for example the relatively high degree of the concept near the top left corner (actually *senior professional*) even if it has fewer authors talking about it than the concepts more centrally located. This can for example indicate a couple of papers intensively referring the concept. As defined above, a *k-core* is a maximal group of actors, all of whom are connected to some number (k) of other members of the group. The k-core algorithm was applied successively to the 1-mode bipartite matrix of each year. We selected the highest k from each year resulting in the groups represented in the next table. The k-core algorithm tries to find the maximum k for which there is a maximal group of authors connected to k concepts and vice-versa.
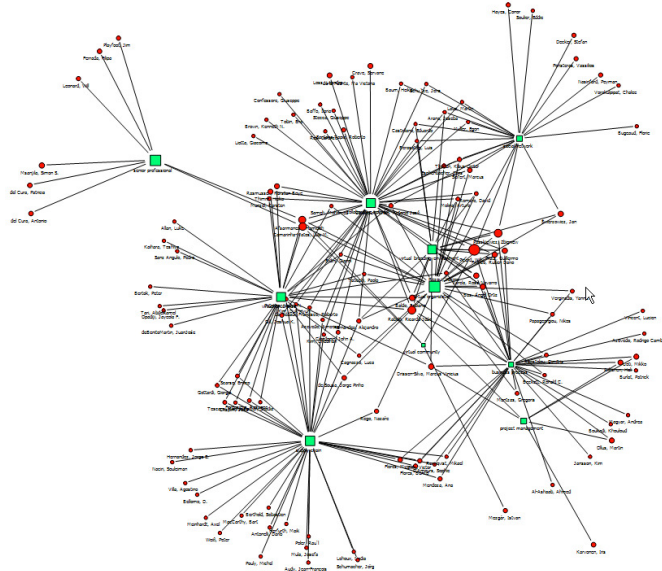
Figure 4 – PRO-VE 2009 socio-semantic network (node size~degree centrality)

Table 2 – The five highest scores of the degree centrality for each year (NetMiner).

	authors				concepts	
1999			**1**			**1**
			Degree			Degree
	1	H. Afsarmanesh	432,000000	1	virtual enterprise	1.071,000000
	2	L. M. Camarinha-Matos	421,000000	2	information management	303,000000
	3	C. Garita	171,000000	3	business process	244,000000
	4	Y. Ugur	162,000000	4	virtual organization	221,000000
	5	L.O. Hertzberger	154,000000	5	virtual enterprise coordinator	215,000000
2005			**1**			**1**
			Degree			Degree
	1	Hamideh Afsarmanesh	593,000000	1	virtual organization	2.641,000000
	2	Luis M. Camarinha-Matos	400,000000	2	virtual breeding environment	786,000000
	3	Martin Ollus	375,000000	3	collaborative network	694,000000
	4	Iire Salkari	151,000000	4	virtual enterprise	566,000000
	5	Iris Karvonen	151,000000	5	business process	453,000000
2009			**1**			**1**
			Degree			Degree
	1	Picard, Willy	323,000000	1	virtual organization	2.641,000000
	2	Paszkiewicz, Zbigniew	209,000000	2	virtual breeding environment	786,000000
	3	Rabelo, Ricardo José	197,000000	3	collaborative network	694,000000
	4	Camarinha-Matos, Luis M.	185,000000	4	virtual enterprise	566,000000
	5	Afsarmanesh, Hamideh	173,000000	5	business process	453,000000
2013			**1**			**1**
			Degree			Degree
	1	Luis M. Camarinha-Matos	219,000000	1	virtual organization	992,000000
	2	Fábio Müller Guerrini	130,000000	2	collaborative network	952,000000
	3	Filipa Ferrada	118,000000	3	business process	493,000000
	4	Ana Inês Oliveira	118,000000	4	virtual breeding environment	414,000000
	5	Ricardo J. Rabelo	117,000000	5	process model	342,000000

The results (see table 3) suggest that in PRO-VE 2005 there was a relatively big group of authors addressing 4 concepts and also a relatively big group of concepts addressed by 4 authors. The other years, regardless the value of k, this group is quite small. Table gives an idea of the connectedness of the 5-core and 3-core groups for 2013. A categorical core-periphery model enables to subdivide the authors-by-authors 1-mode network in two groups of authors: one exhibiting a very strong resemblance in the concepts and number of concepts addressed (the "core"), and other with very few communalities (the "periphery").

Table 4 shows the evolution of the "cores" for the four years in analysis. Each "core" of authors has an high density of ties among themselves because they addressed many concepts in common. We can observe a small "core" in 1999 and "cores" with almost the double of members in 2009 and 2013.

Table 3 – k-core results for the maximum k in each year (concepts in bold) (NetMiner)

1999	2005	2009	2013
5-core	**4-core**	**4-core**	**5-core**
A. Frenkel	Ana Ines Oliveira	Bas, Ángel Ortiz	Américo Azevedo
Alexandra Augusta Pereira Klen	Arturo Molina	Franco, Ruben Daño	Andrea Zangiacomi
C. Garita	Falk Graser	Paszkiewicz, Zbigniew	Eva Coscia
C. P. Lima	Hamideh Afsarmanesh	Picard, Willy	João Bastos
H. Afsarmanesh	Ingo Westphal	Prats, Guillermo	Rosanna Fornasiero
L. M. Camarinha-Matos	Iris Karvonen	Varela, Rosa Navarro	**collaborative network**
L.O. Hertzberger	Ivan Silveri	**virtual breeding environment**	**virtual organization**
Luiz Marcio Spinosa	Jens Eschenbacher	**collaborative network**	**business process**
Ricardo. J. Rabelo	Kim Jansson	**virtual organization**	**business model**
V. Santos Silva	lire Salkari	**business process**	**supply chain**
Y. Ugur	Luis M. Camarinha-Matos		
virtual enterprise	Martin Ollus		
business process	Nathalie Galeano		
virtual enterprise member	Toni Jarimo		
virtual enterprise coordinator	Ugo Negretto		
virtual enterprise environment	Urho Pulkkinen		
information management	**virtual organization**		
	virtual breeding environment		
	collaborative network		
	virtual enterprise		
	supply chain		
	virtual breeding environment member		
	virtual organization creation		
	virtual organization management		

Table 4 – Categorial "cores" resulting from the core-periphery analysis of the authors-by-authors network

1999	2005	2009	2013
A. Frenkel	Arturo Molina	Ruben Dario Franco	Peter Weiss
L. M. Camarinha-Matos	David Guerra	Angel Ortiz Bas	Simon Biggs
L.O. Hertzberger	Nathalie Galeano	Jan Swierzowicz	Adriano Fiorese
H. Afsarmanesh	Iris Karvonen	Fabiano Baldo	Fabio Miller Guerrini
C. Garita	lire Salkari	Ricardo Jose Rabelo	Filipa Ferrada
Y. Ugur	Martin Ollus	Rosa Navarro Varela	Lorenzo Tiacci
	Hamideh Afsarmanesh	Zbigniew Paszkiewicz	Reza Vatankhah Barenji
	Luis M. Camarinha-Matos	Willy Picard	Mohammad Safahi
	Ricardo Camacho	Guillermo Prats	Ana Ines Oliveira
		Luis M. Camarinha-Matos	Olaolu Sofela
		Simon S.Msanjila	Angelita Segoria Gasparotto

5   Discussion

We will discuss now the results presented in the previous section, using for that the hypothesis formulated in section 3.

H1.1. There is a core set of concepts used in all the editions of the conference

This hypothesis is confirmed as we identified, through the longitudinal terminological and conceptual analysis, a common set of concepts present in the top-10 ranking. These are the *foundational concepts* of the PRO-VE research community. The degree centrality measure of the concepts and the concepts returned by the k-core analysis seems to further confirm the hypothesis.

H1.2. The variability of the concepts not in the core is high;

Approximately 53% of the concepts in the top 10 are contingent, meaning that they are addressed in at most two consecutive editions of the conference. This suggests an high variability of the non foundational concepts. However, if we take the top 20, the variability ceases to be high as very few concepts entered anew for the top 20 each year.

H2.1. A core set of concepts is shared by the majority of the researchers

The results from the k-core analysis seem to suggest the contrary. Neither the foundational concepts all appear in the k-core nor the group of authors in the core is the majority. Even if we consider a lower k this is nor confirmed. For example, in 2009, the 2-core — the set of authors using two concepts — includes only 57% of that year's proceedings. This way, we cannot talk about epistemic communities, as defined by [1], within the PRO-VE community. To be considered an epistemic community, a group of authors should all share all the concepts in a given set.

H2.2. There are groups of researchers that are likely to influence the adoption of new concepts.

Although this is perceived empirically, it doesn't follow from the results of our study. From the degree centrality measures, the authors Camarinha-Matos and Afsarmanesh are likely to be the main influencers, given their centrality in the network. However, this is not confirmed neither the k-cores nor in the core-periphery analysis.

The conceptual system of PRO-VE evolved, but not as much as we expected. There is a core set of seven concepts in the top 10, which makes the concepts used in the majority of the papers very stable: the last foundational concepts appearing by the first time in the top 10 backs to 2004. On the contrary, the groups of authors sharing more concepts is not very stable throughout the years, as can be observed in the results of the k-cores and core-periphery analysis. On the other side, the more central actors are also more stable. The limitations of this study lie mainly in using a derived social network. In fact, it is formed by linking the authors that co-use of the same concepts in their papers. This prevents an orthogonal analysis between the social network and the semantic network (conceptual system).

6   Conclusions

It was a long way for the PRO-VE community since the dominance of the concept *virtual enterprise* in 1999 to the relevance and centrality of *collaborative network* in 2013. It was also a long way from 1999 where Camarinha-Matos and Afsarmanesh had a degree centrality more that the double of any other to 2013 where the centrality is more distributed between the more central researchers.

In spite of this evolution, it is our opinion that the evolution of the PRO-VE socio-semantic network is not enough to produce a more visible evolution in the knowledge communicated in the PRO-VE conferences. Apparently, the newcomers largely reproduce the "official" conceptual system (*foundational concepts*).

This study can be greatly improved, opening new possibilities for research. The more important improvement is to develop the social network model in order to include a temporal dimension. Besides building the usual co-authoring network for each year, the dynamics of this network could be captured by creating a cumulative function in the form $N_{ca}(t_k) = f(N_{ca}(t_{k-1}))$, where $N_{ca}(t)$ is a co-authorship network in the year t. This would provide a more detailed ways to analyse the evolution of the PRO-VE socio-semantic-network.

7   Acknowledgements

8   References

[1] Roth, Camille, and Jean-Philippe Cointet . "Social and Semantic Coevolution in Knowledge Networks." Social Networks 32 (1): 16–29. (2010)

[2] Camarinha-Matos, Luis M, and Hamideh Afsarmanesh. "Collaborative Networks: a New Scientific Discipline." *Journal of Intelligent Manufacturing* 16: 439_452. (2005)

[3] Greene, Derek. "Tracking the Evolution of Communities in Dynamic Social Networks." In Advances in Social Networks Analysis and Mining (ASONAM), International Conference On, 1–18 (2010).

[4] Hofer, Katharina Maria, Angela Elisabeth Smejkal, F Zeynep Bilgin, and Gerhard A Wuehrer.. "Conference Proceedings as a Matter of Bibliometric Studies: the Academy of International Business 2006–2008." Scientometrics 84 (3): 845–62. doi:10.1007/s11192-010-0216-6 (2010)

[5] Cash, Philip, Stanko Škec, and Mario Štorga. "A Bibliometric Analysis of the Design 2012 Conference": Proceedings of the INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13. 1–10 (2013).

[6] Volpentesta, A., and Felicetti, A. Eigenvector centrality based on shared research topics in a scientific community, in Camarinha-Matos L. M., Afsarmanesh H., Boucher X., Collaborative Networks for a Sustainable World. IFIP Advances in Information and Communication Technology, pp 626-633. Springer Berlin Heidelberg. (2010).

[7] Picard, W, Extracting the Dynamic Popularity of Concepts from a Corpus of Short-Sentence Documents. in Camarinha-Matos L. M., Afsarmanesh H., Lai Xu, Collaborative Networks in the Internet of Services. IFIP Advances in Information and Communication Technology 582-591 Springer Berlin Heidelberg (2012).

[8] Frantzi, K., Ananiadou, S. and Mima, H. Automatic recognition of multi-word terms. International Journal of Digital Libraries 3(2), pp.117-132. (2000)

[9] Hanneman, Robert A. and Mark Riddle. Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/). (2005)

[10] Justeson, J. and Katz, S. Technical terminology: some linguistic properties and an algorithm for identification in text. Natural Language Engineering, 1, pp 9-27. doi:10.1017/S1351324900000048. (1995)