Dynamic Analysis of the Development of Scientific Communities in the Field of Soft Computing

Ekaterina Kutynina[1] and Alexander Lepskiy[1]

Abstract This paper is dedicated to the research of the dynamics of development and interactions among several scientific communities in the field of fuzzy logic and soft computing. This analysis was performed with the help of the following characteristics: conferences participants' renewal, the level of cooperation in scientific communities, participation of one community's key players in activities of the other ones, comparative number of most active participants in each community, uniformity of key players' participation in different conferences.

Key words: scientific communities, key participants of communities, interaction between scientific communities

1 Introduction

At present scientific communities are an essential part and an important form of the scientific process organizing. In recent years, scientific communities are often studied by methods of network analysis. In particular, the co-authorship networks and citing networks [7] are popular. However, the scientific community tends to develop: there are new communities; some communities degrade, while others are combined, etc. The life cycle of scientific communities is considered in a number of works (see [1]). The interactions among the scientific communities in the field of artificial intelligence for the last 19 years were investigated in [2], and [4] is a similar study that was carried out for scientific communities in the field of computer science. In [10] the dynamic changes in the co-authorship network of conference ISIPTA [9] were analyzed.

The given work is devoted to the investigation of the development and interactions of scientific communities in the field of fuzzy mathematics

Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia {ekytinina, alex.lepskiy}@gmail.com

(EUSFLAT, NAFIPS), imprecise probability (SIPTA, BFAS) and soft computing (SMPS) during the period 1999-2014. The database of this study is based on the materials of the conferences held by the above mentioned scientific communities.

2 Dataset description

Following scientific communities were considered:

- BFAS (Belief Functions and Applications Society) [3]. BFAS was formed in 2010. Conference – BELIEF.
- EUFSLAT (European Society for Fuzzy Logic and Technology) [8] was founded in 1998. Conference – EUFSLAT.
- NAFIPS (North American Fuzzy Information Processing Society) [6]. NAFIPS was established in 1981. Conference – NAFIPS.
- SIPTA (The Society for Imprecise Probability: Theories and Applications) [9] was formed in 2002. Conference – ISIPTA (International Symposium on Imprecise Probability: Theories and Application).
- SMPS (International Conferences on Soft Methods in Probability and Statistics) [5]. Conference SMPS has been held since 2002.

Conferences EUFSLAT, ISPITA, SMPS, BELIEF (for brevity they are designated with letters E, I, S, B respectively) are held once every two years, and conference NAFIPS (symbol N) – every year.

The Fig. 1 provides the visualization of the intersection of the conferences' themes. It could be expected, that connection within groups of communities EUSFLAT, NAFIPS, SMPS on the one hand, and BFAS, ISIPTA on the other hand would be tighter within groups than between them.

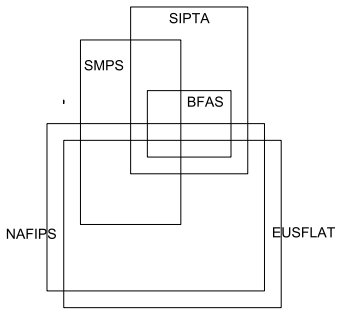

Fig. 1 The schematic visualization of the intersection of communities' themes

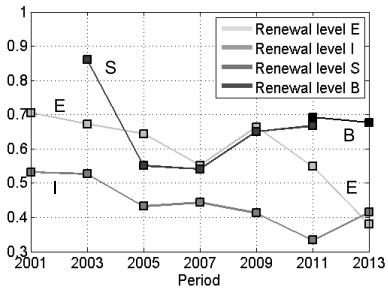

Fig. 2 The renewal level of participation in conferences

The data about the authors of the papers presented at conferences during the period 1999-2014 was collected. There are $N = 3377$ participants in total. Since almost all the conferences are held every two years, the entire time interval 1999-2014 was divided into 8 equal subintervals [1999,2001),

..., [2013, 2015). On the figures below the left boundaries of all subintervals are indicated on the horizontal axes. Data on the conferences held by the NAFIPS community have been combined for the 2009 and the 2010, 2011 and 2012 and then considered as a single event.

3 Analysis of the development dynamic of scientific communities

3.1 Renewal of conferences' participants

One of the main indicators that can characterize the internal development of the scientific community can be a number of new conference participants, who did not participate in previous conferences of the community. Let All_i^j is a set of all participants of the conference j in the period i. The coefficient of renewability for j -th conference in the period i can be considered as a value $U_i^j = \frac{\left|All_i^j \setminus \left(All_1^j \cup \ldots \cup All_{i-1}^j\right)\right|}{\left|All_i^j\right|}$, which is the ratio of the number of new entrants to the number of all participants of this conference in the considered period.

The Fig. 2 shows that the average renewal of ISIPTA conference participants is significantly lower than renewability for other conferences. Almost all the conferences (except for SMPS) tend to decrease renewal of participants. At the same time it should be noticed that the total number of participants in each conference on average varies slightly.

Experienced researchers take part in the conferences as well as their young colleagues. However, in terms of development and interaction of the scientific communities it seems more meaningful to consider the information about experienced researchers, in other words those who took part in several conferences and presented several papers at the same conference. Let's call these researchers key participants.

3.2 Key participants

The significance of a participant s is defined as $Val_i(s)$, the sum of the researchers contributions in the creation of all publications for the period i, where $s = 1, ..., 3377$, $i = 1, ..., 8$. It will be assumed that the total value of the publication is equal to 1 and is divided among all co-authors equally. If the participant s took part in several conferences during the period i his total contribution is calculated as $Val_i(s) = \sum_j Val_i^j(s)$ (the total contribution was calculated separately for two periods 1999-2007 and 2009-2015, since the number of conferences which were held during this periods was different). Those conference participants are called key participants, whose total contributions exceed a certain limit p. Below are the results for the cut-off threshold of the key participants $p = 2$, in other words, those who totally

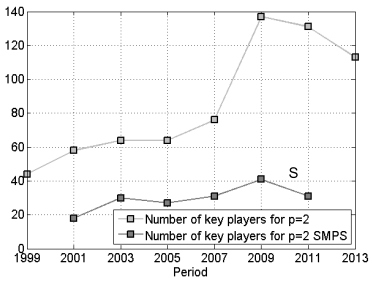

Fig. 3 Dynamics of changes in the total number of communities' key participants

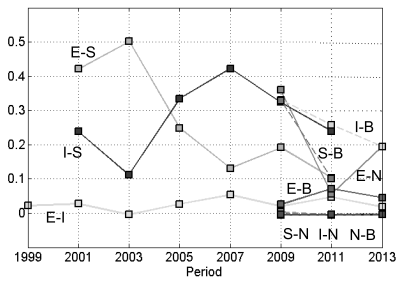

Fig. 4 The correlation between the key participants of the conferences

wrote not less than 2 works over 8 years. The set of key participants in the period i is indicated K_i.

Suppose that $K_i^j = K_i \cap All_i^j$ is a set of key participants of the community j, which held the conference in the period i. In this case there is an opportunity to study the dynamics of the K_i^j sets structural changes and characteristics of their interactions. The Fig. 3 shows the changes in the total numbers of the key participants in all communities K_i. As it can be seen from the graph, the rate tends to increase in the considered time interval. This suggests that interest in research in the field of imprecise probability, fuzzy sets and soft computing increases with time, the significance of this research area is growing.

3.3 The level of communication among communities in relation to the key participants

One of the key issues is to determine the level of cooperation in scientific communities in connection with the common key participants of the conferences organized by thematically close communities. This level can be defined as the correlation between the vectors of the significances of the key participants in couples of communities. Thus, $n_i = |K_i|$ is the number of the key participants in the period i. For each period i, the vector $\mathbf{w}_j^i = (w_{1j}^i, ..., w_{n_i j}^i)$ is put in correspondence to conference j, where $w_{sj}^i = Val_i^j(s)$ is the significance of scientist s just for the conference of j-th community in the period i. Then, the level of cooperation among communities k and j in the period i can be considered as a selective linear Pearson correlation coefficient r_{kj}^i between vectors $\mathbf{w}_k^i$ and $\mathbf{w}_j^i$. The Fig. 4 shows the variation of the selective correlation coefficient for all pairs of communities. It is evident that, as a rule, the level of cooperation among communities in terms of the common key participants of conferences either are initially small (for example, among E and I, E and B), or have a tendency to a decrease (for example, between E and S, S and B, I

and B). All this suggests a trend to isolate these communities. The exception here is a pair of I and S.

3.4 The participation of the key participants in other communities

If, however, the key participants of the community are involved in other conferences, the extent of such participation for the community j in the period i can be estimated with the formula $k_j^i = \frac{1}{l \cdot n_j^i} \sum_{k=1,k\neq j}^{5} m_{jk}^i$, where l is the number of non-empty sets K_i^j in the i-th period, $m_{jk}^i = \left|K_i^j \cap K_i^k\right|$ is the number of common key members of communities j and k in the period i, $n_j^i = \left|K_i^j\right|$ is the number of key conference participants of j-th community in the period i. The higher this ratio, the more actively the key participants of the particular community are involved in other communities' activities. The high value of this factor could mean that the key participants do not regard the community as a key community in the considered field of knowledge.

The Fig. 5 is a visualization of this ratio dynamics. One can see that the least "key" one was the community of S until 2009. The community N turned to be the most "closed", in other words, the key members of this community rarely visit other conferences. But this can be explained by "regional" separateness of this community. The most stable is the community E, for which the rate of participation of key scientists in the other communities does not change much and remains quite small.

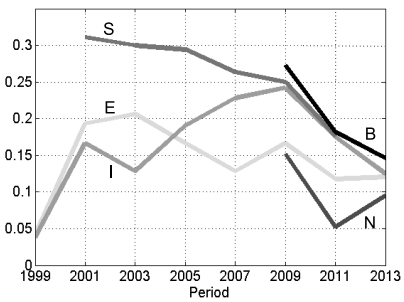

Fig. 5 The dynamics of changes in the participation rate in other communities

Table 1 The list of the most active participants of all communities

Key participant	Centrality	Participation in communities
Dubois D.	0.260	E(7), I(5), S(5), B(2)
Kacprzyk J.	0.256	E(8), S(5), N(2)
Grzegorzewski P.	0.229	E(6), S(6), N(1)
de Baets B.	0.211	E(8), I(1), S(4)
Trillas E.	0.183	E(7), N(3)
Prade H.	0.181	E(6), I(1), S(3)
Novák V.	0.175	E(8), N(2)
Recasens J.	0.172	E(7), S(1), N(2)

3.5 The most active community members and most active communities

Suppose K_i is a set of the key participants of all conferences for the period i, $K = \bigcup_i K_i$ is a set of the key participants of all conferences on record.

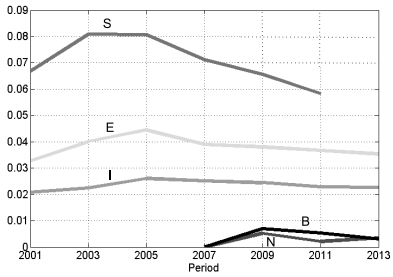

Fig. 6 The dynamics of changes in the average value of activity for the all communities

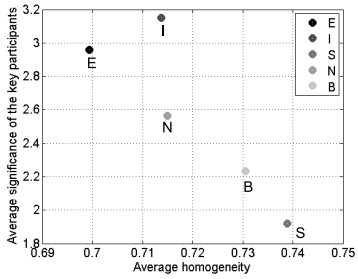

Fig. 7 The average homogeneity and average significance of the key communities' participants

Consider the "friendship" graph for the communities participants (conferences) $G_i = (K_i, E_i)$, $i = 1, ..., 8$ and $G = (K, E)$, where E_i (E) is the set of edges with weights e_{st}, which is equal to the number of the joint participation of key participants s and t in the same conferences for the period i (for all periods).

In this connection we can rise the problem of determining those members who are "friends" with the greatest number of other key participants, taking into account not only direct relations (participation in one conference), but also indirect (i.e. a "sign through a friend"). Such participants can be considered as the most active members of the communities. This problem is solved in the analysis of network structures with the help of the eigenvector centrality. The calculation of the measure of centrality for each node is connected with the solution of eigenvalue finding problem regarding the adjacency matrix A of the network graph [7]: the vector of the relative centralities $\mathbf{x}$ is an eigenvector of the adjacency matrix that corresponds to the largest eigenvalue $\lambda_{\max}$.

On the basis of the eigenvector centrality indicators such as the average value of activities of key participants in each community were introduced: $act_j = \frac{1}{m_j N_j} \sum_{s=1}^{N} n_{sj} x_s$, where x_s is s-th component of the relative centralities vector $\mathbf{x} = (x_1, ..., x_N)$ of the "friendship" graph of key community members; n_{sj} is the number of times that the participant s took part in the conference j, $j = 1, ..., 5$; N_j is the total number of key participants of the community j at the moment of the index calculating; m_j is the number of conferences that had been held by the community j by the moment of index calculating.

The Table. 1 provides the list of most active members of all communities. The number of the considered scientists' participation in the conferences of each community is shown in brackets in the last column.

Dynamics of changes in the average value of activity for all communities is represented in Fig. 6. The most active key players are participants of SMPS community, the lowest average activity is observed among the participants of NAFIPS and BFAS communities.

3.6 Analysis of participation uniformity of the key participants in different communities

Each participant s was assigned with the vector $\mathbf{n}_s = (n_{s1}, ..., n_{s5})$, where n_{sj} is the number of times, which the participant s took part in the conference j, $j = 1, ..., 5$. The vector $\mathbf{n}_s = (n_{s1}, ..., n_{s5})$ was put to correspondence to the vector of relative frequencies $\mathbf{p}_s = (p_{s1}, ..., p_{s5})$, where $p_{sj} = \frac{n_{sj}}{\sum_{k=1}^{5} n_{sk}}$. Then $\mathbf{p}_s = (p_{s1}, ..., p_{s5})$ is some probability distribution. Pose the question of the non-uniformity degree of this distribution, which characterizes the degree of heterogeneity of participation in conferences of different communities for the s-th scientist. This degree can be estimated with the help of the Shannon entropy $S(\mathbf{p}_s) = -\sum_{j=1}^{5} p_{sj} \log_2 p_{sj}$. For the uniform distribution this function reaches its maximum $S\left(\frac{1}{5}, ..., \frac{1}{5}\right) = \log_2 5$. The entropy achieves the minimum $S(\mathbf{p}) = 0$ when exactly one of the p_j is one and all the rest are zero.

Now, for each set of key participants K^j of community j the average homogeneity was calculated by the formula $unif_j = \frac{1}{|K^j|} \sum_{s \in K^j} S(\mathbf{p}_s)$.

Great value of $unif_j$ indicates that the key members of this community are also actively involved in the work of other communities. A small value of $unif_j$ demonstrates a certain "isolation" degree of the community. The Fig. 7 is a graphical representation of communities by points on the plane, where the first coordinate on the horizontal axis is the average homogeneity of the community, and the second one on the vertical axis is the average value $\overline{Val^j}$ of the aggregate contributions of the key participants in community j for the entire considered period. It can be seen, that the most "closed" communities are EUSFLAT and ISIPTA, thus the average contribution of key participants of these conferences is the highest. The most open community is SMPS, which can be explained by the variety of scientific papers themes presented at conferences of this community. On the other hand it is evident that these two characteristics – average contribution and uniformity are strongly correlated. Again, the "outlier" here is ISIPTA community.

4 Conclusions

The main conclusions of this research are as follows:

- almost all of the conferences (except for SMPS) have a tendency to reduce the renewal of its members (at a fairly constant total number of conference participants); on average the renewal of ISIPTA conference participants is significantly lower than other conferences renewal;
- the level of cooperation in scientific communities in relation to the common key participants of the conferences either is initially small (for example, between EUSFLAT and ISIPTA, EUSFLAT and the BELIEF), or have a tendency to a decrease (for example, between EUSFLAT and SMPS,

SMPS and BELIEF, ISIPTA and BELIEF). All this says about the trend to isolate these communities; exceptions here are ISIPTA and SMPS conferences;

- in terms of the participation of the key participants of a particular community in the activities of other communities, until 2009 the most "open" was a conference SMPS; as far as this characteristics is concerned the most stable community is EUSFLAT, for which the participation rate of key scientists in other communities does not change much and remains quite small;
- the most active participants of the communities were emphasized; It shows that the most active participants are key participants of SMPS community; the lowest activity was observed among the participants of NAFIPS and BFAS communities;
- in terms of uniformity of participation of key participants of a particular community in other communities, the most "closed" communities are EUSFLAT and ISIPTA, thus the average contribution of key participants of these conferences is the highest; the most open community on this indicator is SMPS.

Acknowledgements. The financial support from the Government of the Russian Federation within the framework of the implementation of the 5-100 Programme Roadmap of the National Research University Higher School of Economics is acknowledged.

References

1. Belák V, Karnstedt M, Hayes C (2011) Life-Cycles and Mutual Effects of Scientific Communities. Procedia – Procedia Soc and Behav Sci 00: 36–47
2. Belák V, Hayes C (2015) The Risks of Introspection: A Quantitative Analysis of Influence Between Scientific Communities. In: Proc of the 28th Intern Florida Artif Intel Res Soc Conf 8–13
3. Belief Functions and Applications Society. url: http://www.bfasociety.org/.
4. Biryukov M, Dong C (2010) Analysis of Computer Science Communities Based on DBLP. In: Res and Advan Technol for Dig Libr. Springer, Berlin 228–235
5. Conference on Soft Methods in Probability and Statistics. url:http://smps2014.ibspan.waw.pl
6. North American Fuzzy Information Processing Society.url:http://nafips.ece.ualberta.ca
7. Newman MEJ (2010) Networks: An Introduction. Oxford Univ Press, Oxford, UK
8. The European Society for Fuzzy Logic and Technology. url:http://www.eusflat.org/
9. The Society for Imprecise Probability: Theories and Applications. url:http://www.sipta.org
10. Walter G, Jansen C, Augustin T (2015) Updated Network Analysis of the Imprecise Probability Community Based on ISIPTA Electronic Proceedings. In: Proc of the 9th ISIPTA, Pescara Italy, p.351