DYNAMIC ANALYSIS OF THE DEVELOPMENT OF SCIENTIFIC COMMUNITIES IN THE FIELD OF SOFT COMPUTING

Ekaterina Kutinina

Higher School of Economics, Moscow, Russia, aekytinina@gmail.com

Alexander Lepskiy

Higher School of Economics, Moscow, Russia, alex.lepskiy@gmail.com

The 8th International Conference on Soft Methods in Probability and Statistics - SMPS 2016, September 12 - 14, 2016, Rome, Italy

Contents

References

- Newman M.E.J. (2004), research on the co-authorship networks.
- Hirsch J.E. (2005), Egghe L. (2006), measurement of individual's scientific research output.
- Belák V., Karnstedt M., Hayes C. (2011), models of communities life-cycles.

Motivation and aims of the research

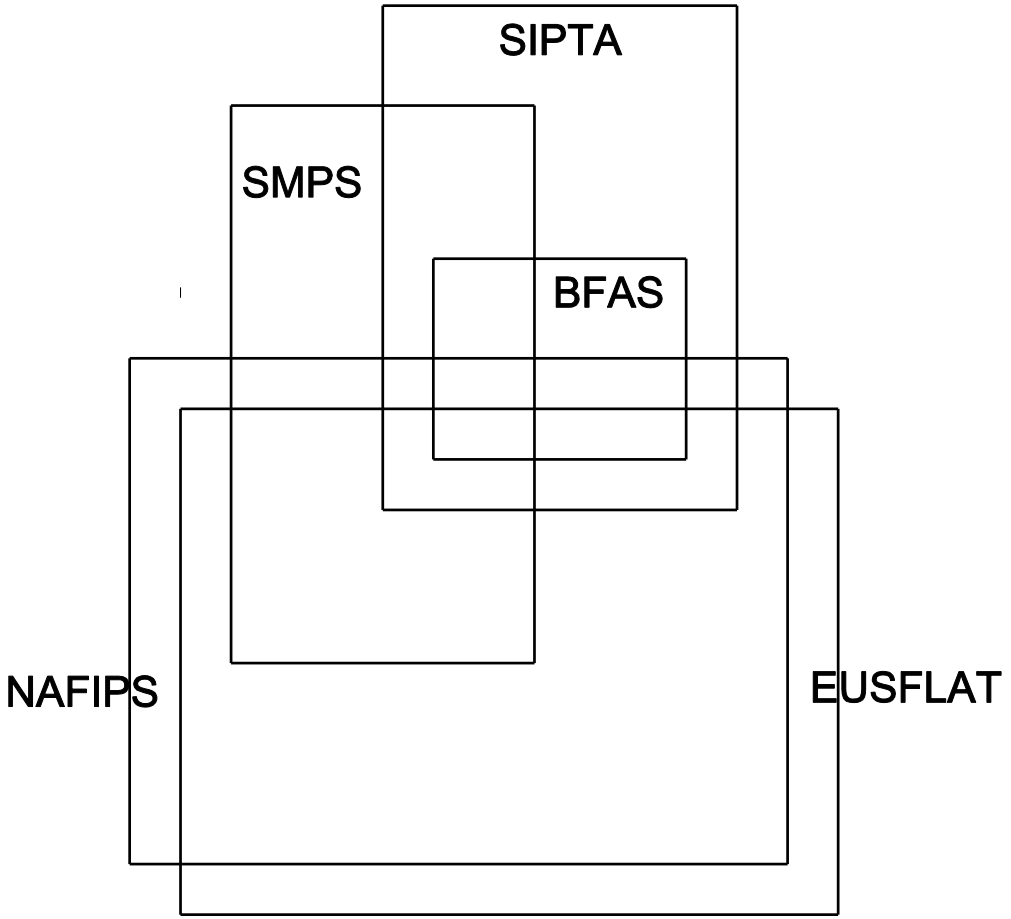
- The scientific communities is essential part of the scientific activity and the development of the communities can be considered as a reflection of the trends in the certain scientific field.
- The main hypothesis of the research: the progress in some scientific area is followed by appearance of narrow specialized branches and interactions among scientific communities become weaker.
- The aim of the research is to investigate the development and interactions of scientific communities in the field of fuzzy mathematics (EUSFLAT, NAFIPS), imprecise probabilities (SIPTA, BFAS) and soft computing (SMPS).

Data

- The data about the authors of the papers presented at conferences of the several scientific communities during the period 1999-2014 was considered.
- Scientific communities:
  1. **BFAS** (Belief Functions and Applications Society). BFAS was formed in 2010 to promote research of trust functions and their application. Conference - BELIEF.
  2. **EUFSLAT** (European Society for Fuzzy Logic and Technology) was founded in 1998. Conference - EUFSLAT.
  3. **NAFIPS** (North American Fuzzy Information Processing Society) . NAFIPS was established in 1981 and specializes in the modeling and management of uncertainty, with an emphasis on the fuzzy system theory and application. Conference - NAFIPS.
  4. **SIPTA** (The Society for Imprecise Probability: Theories and Applications was formed in 2002. Conference - ISIPTA (International Symposium on Imprecise Probability: Theories and Application).
  5. **SMPS** (International Conferences on Soft Methods in Probability and Statistics). Conference SMPS has been held since 2002.

Data

The visualization of the intersection of the themes presented at the conferences.



Main definitions

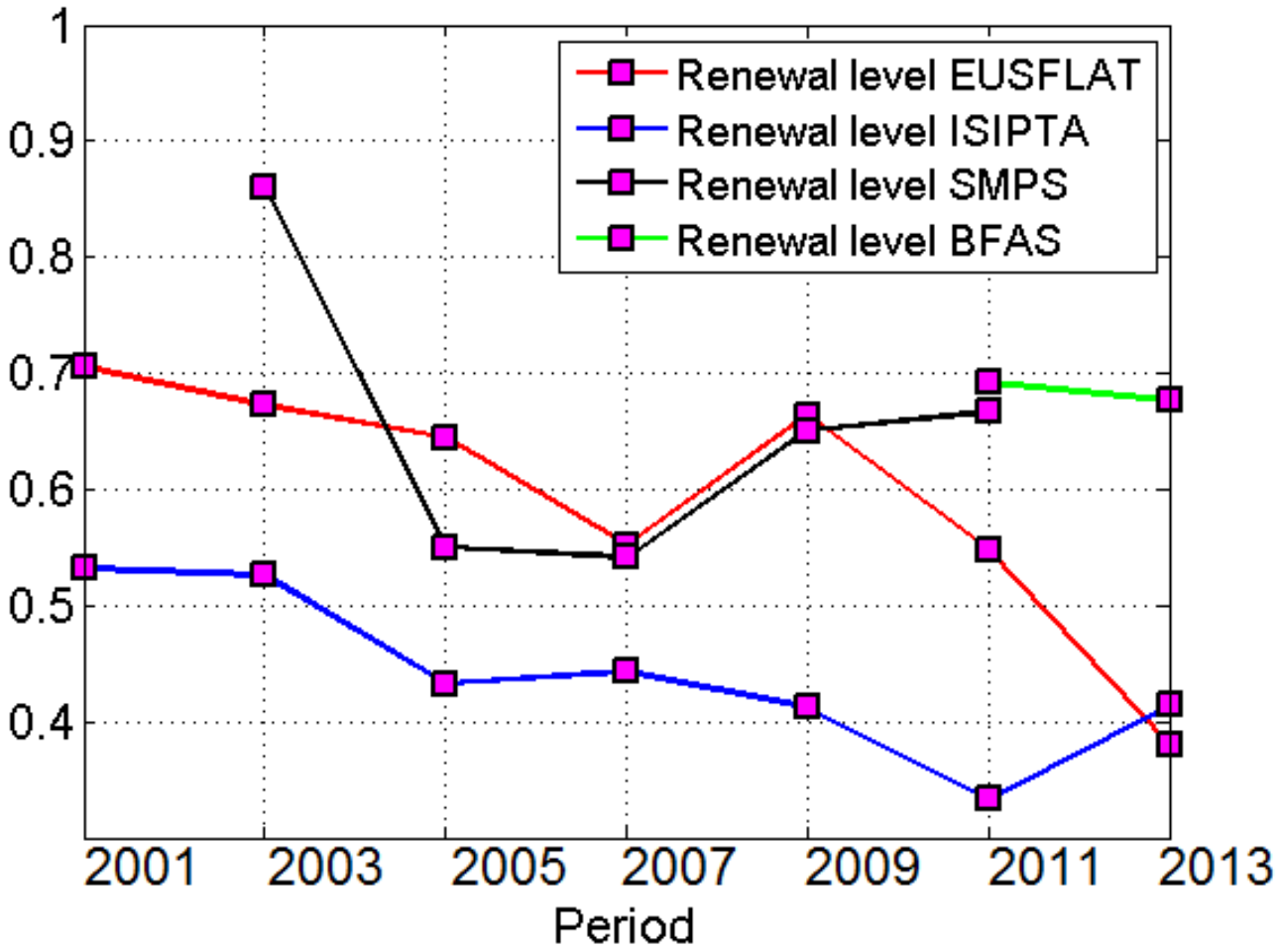
- All_i^j - a set of all participants of the conference j in the period i
- Pr_i^j - a set of all participants of the conference j who has participated in the previous periods $(1, 2, \ldots, i-1)$ at list once.
- $Val_i(s)$ (the significance of a participant s) - the sum of the researcher's contributions in the creation of all publications for the period i, where s = 1, ..., 3377, i = 1, ..., 8.
- Key participant – participant of the conferences, whose aggregate significance exceeds or equals to a certain limit.

Methods: Renewal of conferences' participants

Dynamics of the renewal level of participation in conferences

Number of new conference participants, who did not participate in previous conferences of the community

$$U_i^j = \frac{|\mathrm{All}_i^j \backslash \mathrm{Pr}_i^j|}{|\mathrm{All}_i^j|}$$



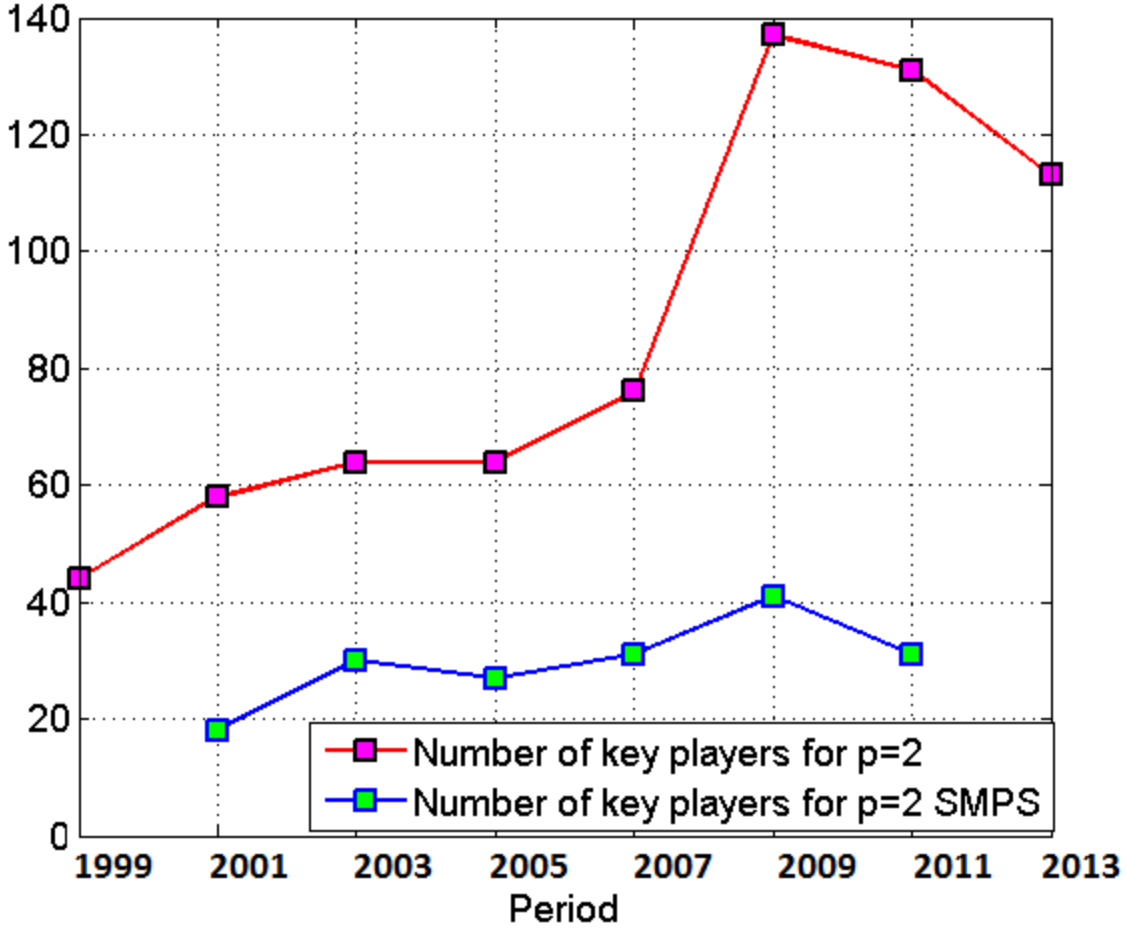
Methods: Key players

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

Key players are the most stable elements of the communities' structure.

$$K_i^j = K_i \cap All_i^j$$

$$K_i = \cup_j K_i^j$$



Methods: Key participants and the level of cooperation

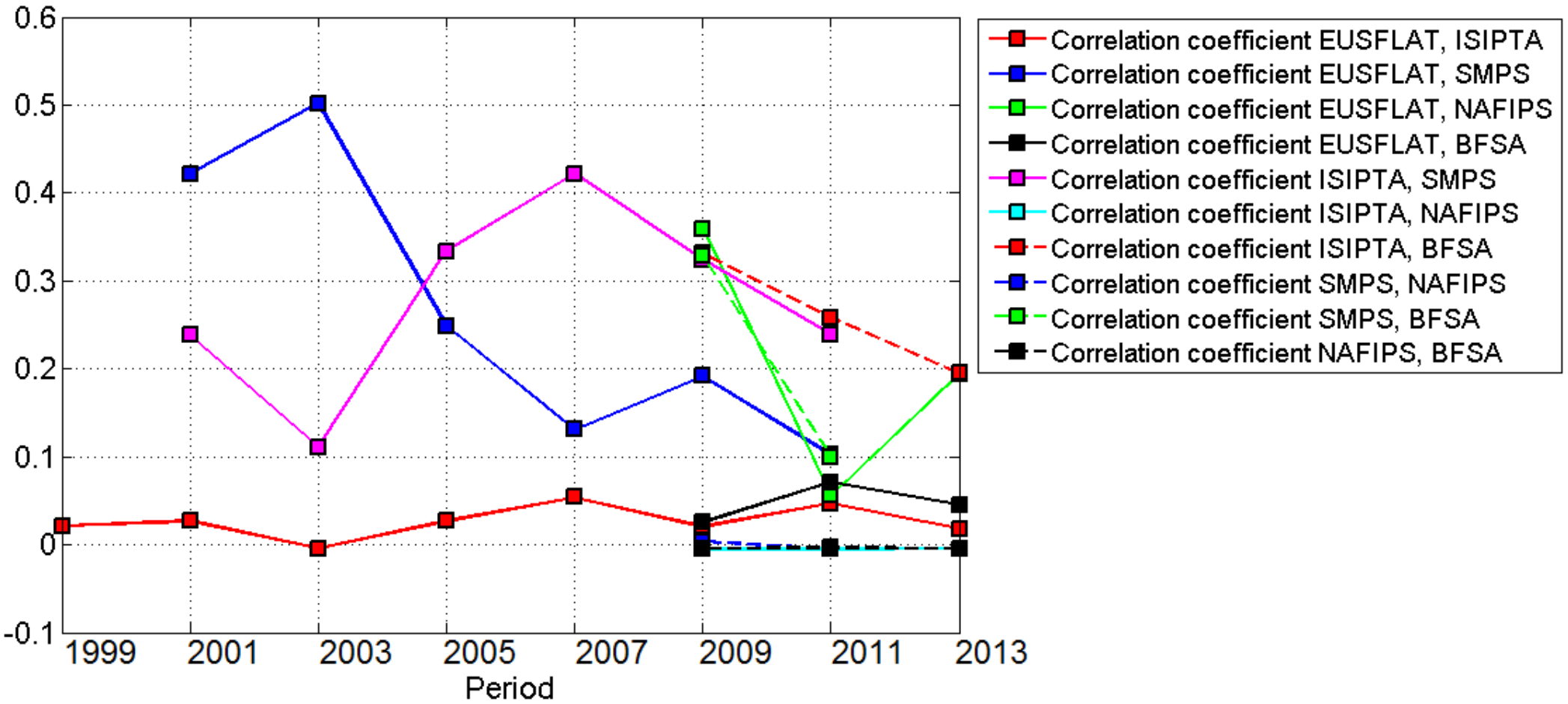
The level of communication among communities in relation to the key participants.

For every period i and each conference j the vector $\mathbf{w}_j^i = \left(w_{1j}^i, \dots, w_{n_i j}^i\right)$ was constructed, where $w_{sj}^i = Val_i^j(s)$.

The level of cooperation among communities k and j in the period i can be considered as a selective linear Pearson correlation coefficient between vectors of the significances of the key players for communities k and j .

$$r_{kj}^i = \frac{\sum_{s=1}^{n_i}(w_{sk}^i - \overline{w}_k^i)(w_{sj}^i - \overline{w}_j^i)}{\sqrt{\sum_{s=1}^{n_i}\left(w_{sk}^i - \overline{w}_k^i\right)^2 \sum_{s=1}^{n_i}\left(w_{sj}^i - \overline{w}_j^i\right)^2}}$$

Methods: Key participants and the level of cooperation



Methods: Participation of the key players in other communities

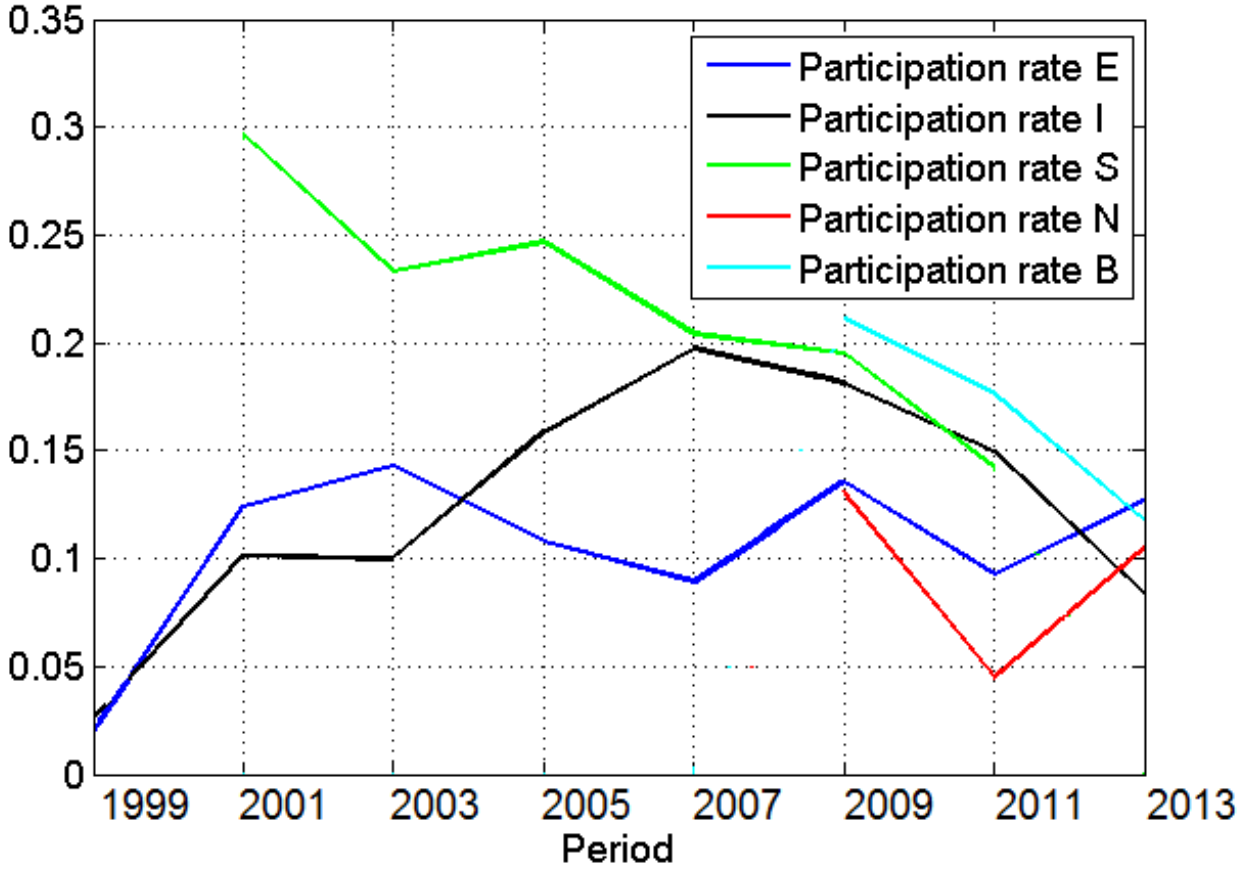
The more this ratio, the more actively key participants of a considered community take part in the other communities

$$k_j^i = \frac{\sum_{s=1}^{5} m_{s,j}^i}{l \cdot n_j^i}$$

$m_{s,k}^i = |K_i^s \cap K_i^k|$

l - number of non-empty sets K_i^j

$n_j^i = \left|K_i^j\right|$



Methods: The most active community members and most active communities

Consider the "friendship" graph for the communities participants (conferences) $G_i = (K_i, E_i)$

Eigenvector centrality: $A\mathbf{x} = \lambda_{max}\mathbf{x} \Rightarrow$ the relative centralities vector $\mathbf{x}$

This approach allows to determine the most “active” key players, those who are “friends” (directly and indirectly) with the greatest number of other key participants.

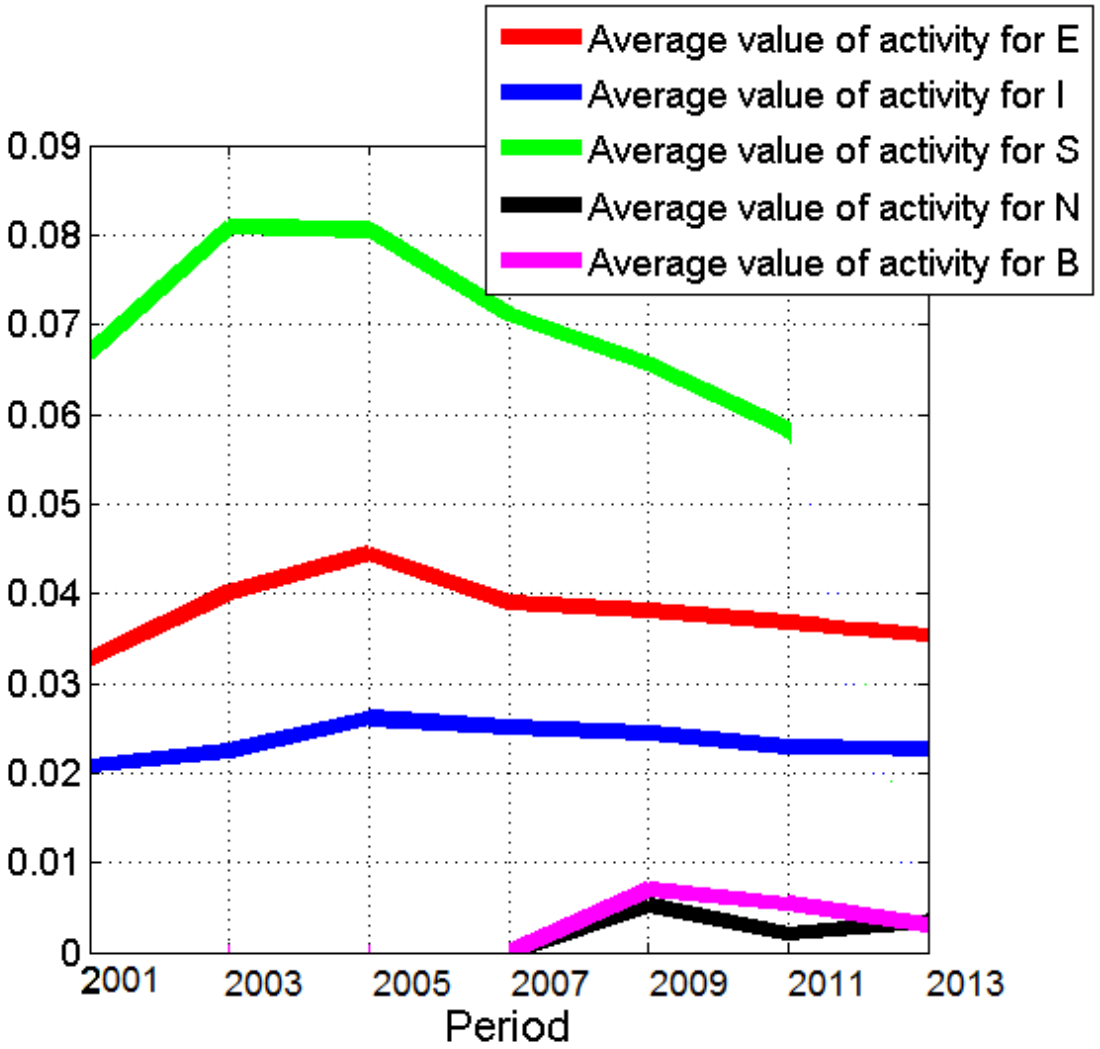
The average value of activities of key players

$$act_j = \frac{1}{m_j N_j} \sum_{i=1}^{N} n_{ij} x_i$$

x_i - i-th component of the relative centralities vector $\mathbf{x} = (x_1, \dots, x_N)$
n_{ij} - number of times than the participant i took part in the conference j.
N_j - total number of key participants of the community j at the moment of index calculating
m_j - number of the conferences that had been held by the community j by the moment of index calculating

Methods: The most active community members and most active communities

№	Key participant	Centrality	Participation in communities
1	Dubois D.	0.260	E(7), I(5), S(5), B(2)
2	Kacprzyk J.	0.256	E(8), S(5), N(2)
3	Grzegorzewski P.	0.229	E(6), S(6), N(1)
4	Baets B.	0.211	E(8), I(1), S(4)
5	Trillas E.	0.183	E(7), N(3)
6	Prade H.	0.181	E(6), I(1), S(3)
7	Novák V.	0.175	E(8), N(2)
8	Recasens J.	0.172	E(7), S(1), N(2)
9	Perfilieva I.	0.170	E(7), N(2)
10	Grabisch M.	0.163	E(6), I(1), S(2)



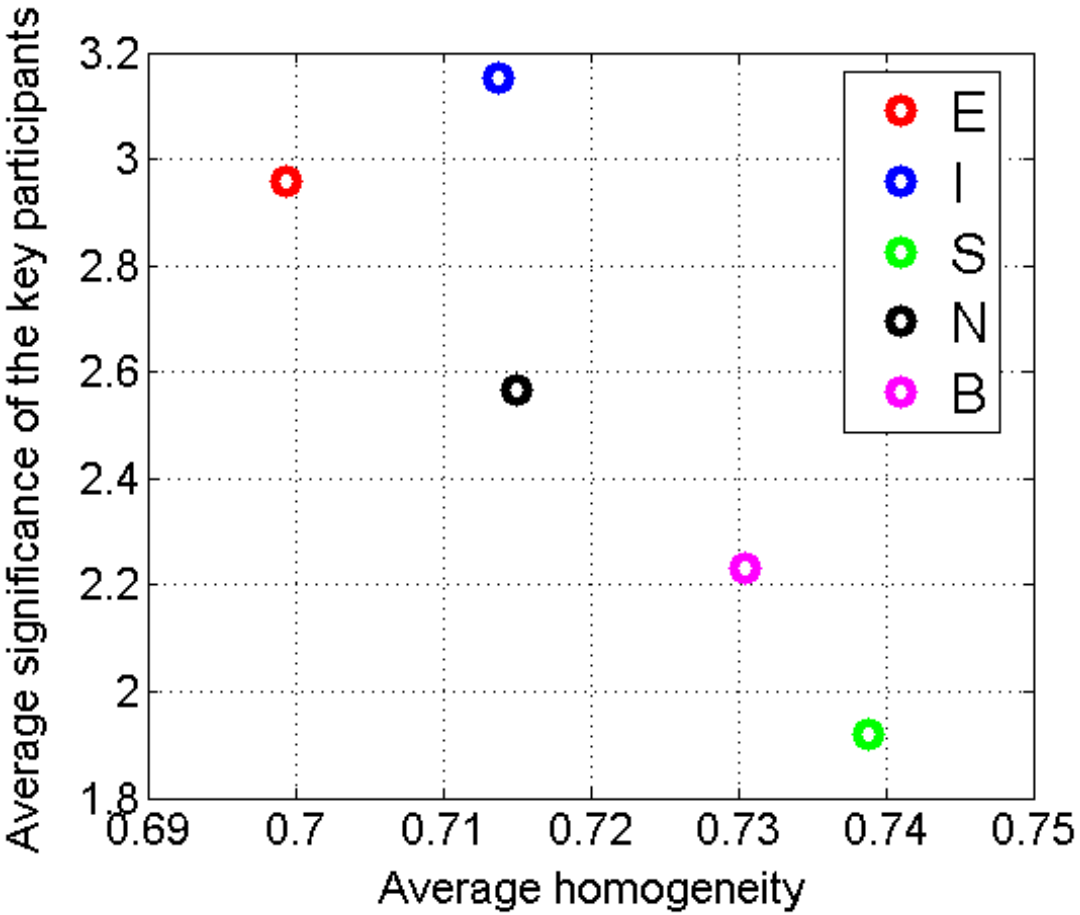
Methods: Analysis of participation uniformity of the key players in different communities

$p_i = (p_{i1}, \dots, p_{i5})$ - the vector of relative participation frequencies for i-th key player

Average homogeneity: $unif_j = \frac{1}{|K^j|} \sum_{i \epsilon K^j} S(\mathbf{p}_i)$.

Where $S(v)$ – Shannon entropy function.
It reaches maximum ($\log_2 5$)when v is a uniform distribution vector, and takes minimum when v is completely non-uniform.

The higher this indicator, the more key participants of the considered community are involved in the live of the other communities.



Results and conclusions

- Almost all of the conferences (except for SMPS) have a tendency to reduce the renewal of its members
- There are evidence about the trend to isolate these communities
- Key players 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
- The most active participants are key players of SMPS community.
- The “closer” community – the higher the significance of its key players