Evolution of Physics Sub-fields

Murali Krishna Enduri[1][a], I. Vinod Reddy[2][b] and Shivakumar Jolad[3][c]
[1]*SRM University-AP, Andhra Pradesh 522502, India*
[2]*Indian Institute of Technology Bhilai, Chhattisgarh 492015, India*
[3]*FLAME University, Pune, Maharashtra 412115, India*

Keywords: PACS Codes, Sub-fields, Physics.

Abstract: We study the evolution and relationships between sub-fields of Physics using the large data set of articles published in the various physical review journals from 1985-2010. Each article is assigned to some PACS codes by their authors which represent specific sub-fields of Physics. We construct a weighted network with nodes as PACS codes and there is a link between two PACS codes if there is an article assigned to both these codes. The weight of a link represents the number of articles in which both PACS codes appears. We study the time evolution of PACS network at various hierarchy levels of PACS codes. We observe that sub-fields Physics of elementary particles and fields, Nuclear Physics and Condensed matter physics have stronger connections inside the field compared to connections to other sub-fields. We also observe that both condensed matter physics sub-fields are strongly related compared to any other pair of sub-fields.

1 INTRODUCTION

Scientific disciplines are becoming increasingly interconnected. New fields of research have emerged by integrating concepts, ideas and tools different disciplines. Tracking the emergence of new areas can give insights into working of scientific enterprise. Among the natural sciences, Physics once considered most rigid has become increasingly fluid. Not only different sub fields of Physics are being merged/mixed to create new ares of physics, but are being widely different areas such as study of system biology, social dynamics, complex networks, etc.

With the emergence of Network Science to understand the study of complex networks using tools from statistical physics, dynamical systems and computer science, and availability of large data sets we can approach the evolution from Network Dynamics perspective (Newman, 2001), (Palla et al., 2007),(Palla et al., 2005), (Kumpula et al., 2008). Especially for physics there are many studies on Evolution of Physics Sub-Fields such as (Battiston et al., 2019),(Liu et al., 2017), (Jia et al., 2017),(Sinatra et al., 2015) (Dias et al., 2018) and (Pan et al., 2012).

Palla et al.(Palla et al., 2005) studied the evolution of communities in weighted networks using Clique Percolation Method (CPM) characterizing the community growth, merge split, and extinction.

The publications in American Physical Society (APS) Journals covers all branches of physics from 1900's. Articles published in APS journals from 1985 on wards carry subject classification codes from Physics and Astronomy Classification System (PACS). In past many authors have constructed PACS network representing the interconnection between the sub-fields and studied their dynamics.

Herrera et al. (Herrera et al., 2010) studied the evolution of communities (physics fields) and found that size of communities tend to increase with age of communities and these communities will have higher number of papers. Their analysis was restricted to third level in PACS hierarchy. In Omodei et al. (Omodei et al., 2013) analyzed the epistemic networks of PACS codes and socio-epistemic network of authors and PACS codes influence the dynamics of each other. Martin et al. (Martin et al., 2013), Redner (REDNER, 2005) studied the co-authorship and citations networks, observed that the tendency of authors cite themselves or their collaborators is higher than others. Pan et al. (Pan et al., 2012) studied PACS networks evolution through k-shell decomposition and observed that over the time there has been an increase in the interdisciplinarity inside physics.

[a] https://orcid.org/0000-0002-9029-2187
[b] https://orcid.org/0000-0002-0365-1138
[c] https://orcid.org/0000-0001-9523-1693

Enduri, M., Reddy, I. and Jolad, S.
Evolution of Physics Sub-fields.
DOI: 10.5220/0009424900880095
In *Proceedings of the 5th International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2020)*, pages 88-95
ISBN: 978-989-758-427-5

However, their work did not include the community evolution.

In this work we study the evolution and relationships between the sub-fields of physics using the physical review journals data. The rest of the paper organized as follows. In Section 2, we give a short description of our data set and PACS codes. In Section 3, we describe the network construction by using the PACS codes from prepared data set. In section 4, we analyse the trend of *single field* papers in a field for every five years with one and two level of PACS hierarchy. In the results section, we study correlation between the different sub-fields of physics using PACS network at different hierarchy levels. Finally, in the last section we conclude this paper and present some future research direction.

2 DATA SET

The American Physical Society (APS) started publishing Physical Review journal from 1893. APS added other journals like Reviews of Modern Physics (1929), Physical Review Letters (1958), Physical Review A,B,C and D (1970), Physical Review E (1993) and most recently Physical Review X in 2011. In our analysis we use all scientific papers published in APS Physical Review (PR) journals from 1985 to 2010. Metadata of each journal article contains an unique digital object identifier (DOI), title, authors of paper, date of publication, affiliations of each author, PACS codes, references to other cited Physical review articles. We primarily use PACS codes, described below, to identify the different sub-fields the articles belong to. PACS is a hierarchical classification scheme developed by American Institute of Physics (AIP), representing different fields and sub-fields of Physics up to five levels. The first level in the PACS hierarchy classifies the ten main fields of Physics such as general Particle Physics, Nuclear Physics, Condensed Matter, Atomic and Molecular Physics etc, as shown in the Figure 1. A journal article may contain PACS codes one or many of these main fields. A PACS code consists of two pairs of numbers followed by a pair of non numeric characters, separated by dots. For example in PACS code 04.25.*dg*, the first digit 0 represents *General* Physics, 4 - *General relativity and gravitation*, 25 - *Approximation methods; equations of motion* and *d* represents *Numerical relativity* and *g* represents *Numerical studies of black holes and black-hole binaries*. PACS codes are regularly revised and updated overtime by American Institute of Physics (AIP), new codes are introduced and some codes are deleted. In our analysis we consider PACS codes up to third level (first four digits) of hierarchy as they are reasonably stable upto this level and represents all sub-fields of physics. We ignore the higher level hierarchy to maintain consistency PACS codes of all papers in our analysis. PACS codes were introduced 1975 but large fraction of papers published between 1975 and 1984 have not assigned any PACS codes. Therefore we confined our analysis to the papers published between 1985 to 2010, as the usage towards PACS code jumped to more than 90% and have been consistently high since then. There is variation in the PACS codes after 2010. So we consider only data in between 1985 to 2010. Basic statistics of our data set is given in Table 1.

Table 1: Basic statistics of the data set 1985-2012.

Number of authors	343055
Number of papers	399713
Average number of papers by an author	9.07
Average number of authors per paper	7.59
Average number of PACS codes per author	10.04
Average number of PACS codes per paper	2.92

3 NETWORK CONSTRUCTION

We construct a PACS network, where the nodes represents the PACS codes and there is an edge between two PACS codes if they have appeared in the same article (Herrera et al., 2010).

The weight of an edge between two nodes i and j is defined as

$$s_{ij} = \sum_{p \in S} \frac{1}{n_p - 1} \tag{1}$$

Where S is the set of papers which belongs to both fields i and j and n_p is the number of scientific fields contained in PACS codes of paper p.

We construct two different PACS networks. The first network contains 10 nodes representing the broad research areas (shown in Figure 1) in physics. To construct this network we treat all the PACS codes with same first digit as a single node. The second network contains 100 nodes representing sub-fields of physics up to the second level of hierarchy. In this network the PACS codes with same first two digits treated as a single node. We use this PACS networks to study the evolution of relationship between different sub-fields of physics with time.

00 General Physics

10 Physics of Elementary Particles and Fields

20 Nuclear Physics

30 Atomic and Molecular Physics

40 Electromagnetism,Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics

50 Physics of Gases, Plasmas, and Electric Discharges

60 Condensed Matter: Structural, Mechanical and Thermal Properties

70 Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties

80 Interdisciplinary Physics and Related Areas of Science and Technology

90 Geophysics, Astronomy and Astrophysics

Figure 1: Sub-Fields of Physics based on PACS codes.

4 RESULTS

In the next subsections we present observation made by analyzing the PACS network at different hierarchy levels.

4.1 Descriptive Analysis of Single Field Articles

We call an article as single field article if it is assigned only one PACS code. These articles are highly specialized in one field and does not contribute to the weights of the PACS network. However they are helpful to define the extent to which a field is *intra-disciplinary* as opposed to *interdisciplinary*. We analyze the single field articles at first and second levels of hierarchy of PACS codes.

Example 1. *The set of PACS codes* $S = \{71.10.Ca, 71.15.Dx, 72.15.Rn\}$ *represents single field, two fields and three fields when restricted to first, second and third level of hierarchy respectively.*

The fraction of single field papers in a field is defined as the ratio of number of single field papers to total number of papers published in that field. For both first and second level the fraction of single field papers for each field are shown in Figure 2 for papers published between 1985-2010. In both the cases we observe that high fraction of single field papers are from the fields Physics of elementary particles and fields, Nuclear Physics and Condensed matter physics. In order to verify whether this trend is uniform over all the years we performed similar analysis

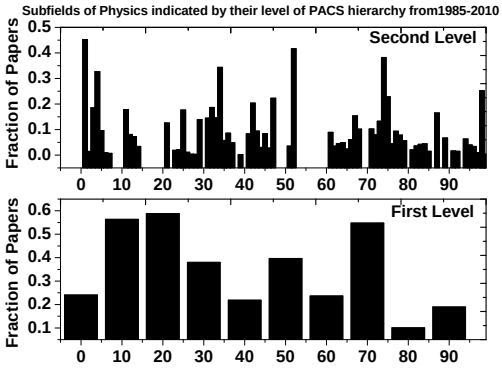

Figure 2: Fraction of single field articles at first and second levels of PACS hierarchy respectively for the articles published between 1985-2010.

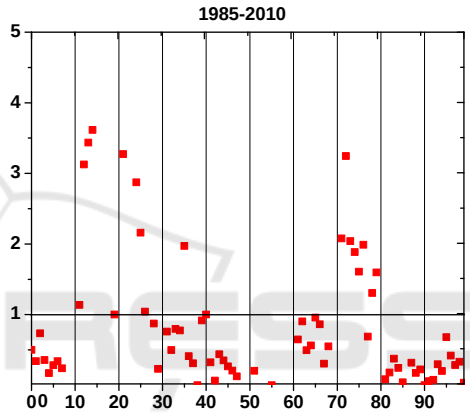

Figure 3: In and out ratio 1985-2010.

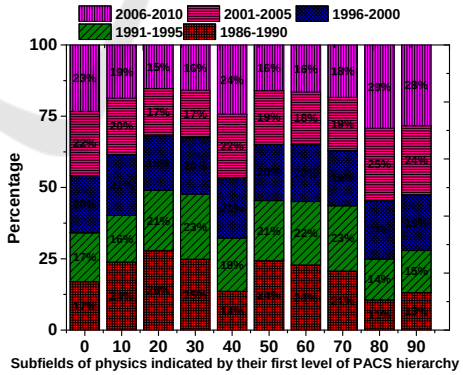

Figure 4: Evolution of fraction of single field articles over the time at first level of PACS hierarchy over the time.

(see Figure 9) for every five years and observed the similar trend. The reason for these fields to have high fraction of single field papers might be they are well established and researchers are able to write papers without using the knowledge from other fields. For a specific field, we call the ratio of the number of single field articles in that field to the total number of

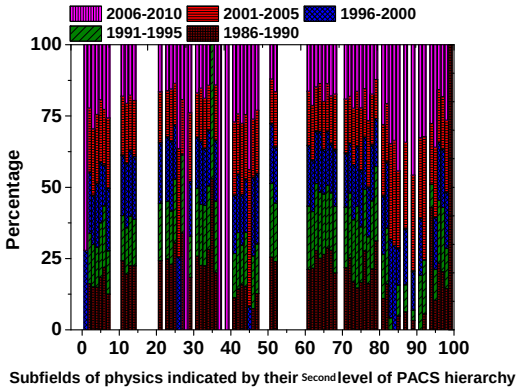

Figure 5: Evolution of fraction of single field articles over the time at second level of PACS hierarchy over the time.

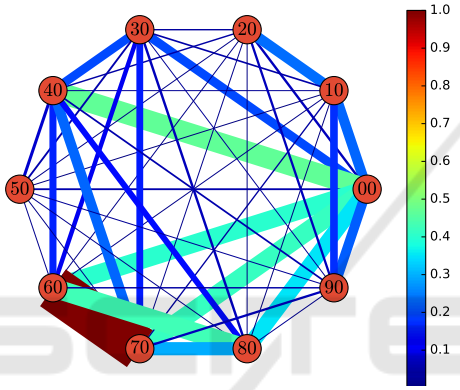

Figure 6: First level PACS network.

articles published in that field as in/out ratio of that field. The in/out ratio of top ten level sub-fields of physics is shown in the Figure 3. The in/out ratio for fields Physics of elementary particle and fields, Nuclear Physics and Condensed matter physics is greater than one indicating that these fields have stronger connections inside the fields compared to outside the fields.

We analyzed the evolution of fraction of single field articles over the time. Figure 4 and 5 shows fraction of papers with single field over every five years for first and second level of PACS hierarchy. We observe that over the time fraction of single fields are decreasing. This might be due to increase in interdisciplinary research.

5 EVOLUTION OF SUB-FIELDS OF PHYSICS

In this section we study the evolution of relation between sub fields of physics from 1985-2010 at different hierarchy levels.

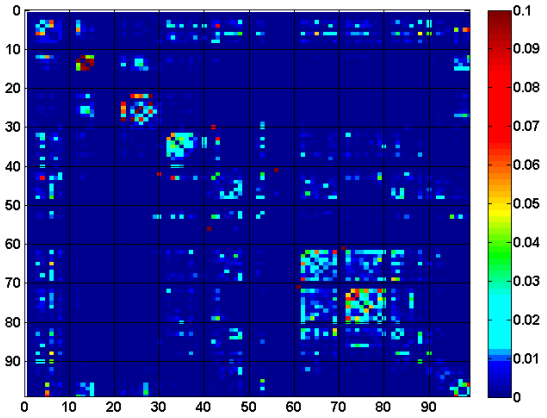

Figure 7: Intensity plot showing the strength between sub-fields of Physics at second level of PACS hierarchy.

5.1 First Level

We build a weighted PACS network by considering PACS codes at first level of hierarchy. This network contains ten nodes which represents the broad sub-fields of physics. Figure 6 shows the weighted PACS network for time period 1985-2010. We observe that two condensed matter physics fields are strongly related. We can also see there are strong connections from General physics to Condensed matter physics, Interdisciplinary Physics (80), Electromagnetism (40) and Nuclear Physics (20). To test whether this trend is uniform over the period 1985-2010, we constructed the same network for every five year interval (see appendix Figure 11). In all five-intervals we observe the same trend showing the strong relationship between the two condensed matter sub-fields.

5.2 Second Level

In this section, we study evolution of relation between the different sub-fields of physics using PACS network by considering the PACS codes up to second level of hierarchy. We have constructed a network whose nodes are sub-fields of physics and there is a edge between the two sub-fields if there is at least one paper containing both sub-fields in its PACS codes. Therefore we have at most 100 nodes denoted as $\{00,01,\cdots,99\}$. The edge weight w_{ij} between the two nodes i and j is defined as in equation (1).

Previously Pan et al.(Pan et al., 2012) have performed k-shell analysis to understand the evolution of PACS network. High and low k-shell indices indicates the position of a node in core and periphery of the network respectively. This does not captures the strength of outside connection of a node, since a node may be in core due to high internal connections. We define

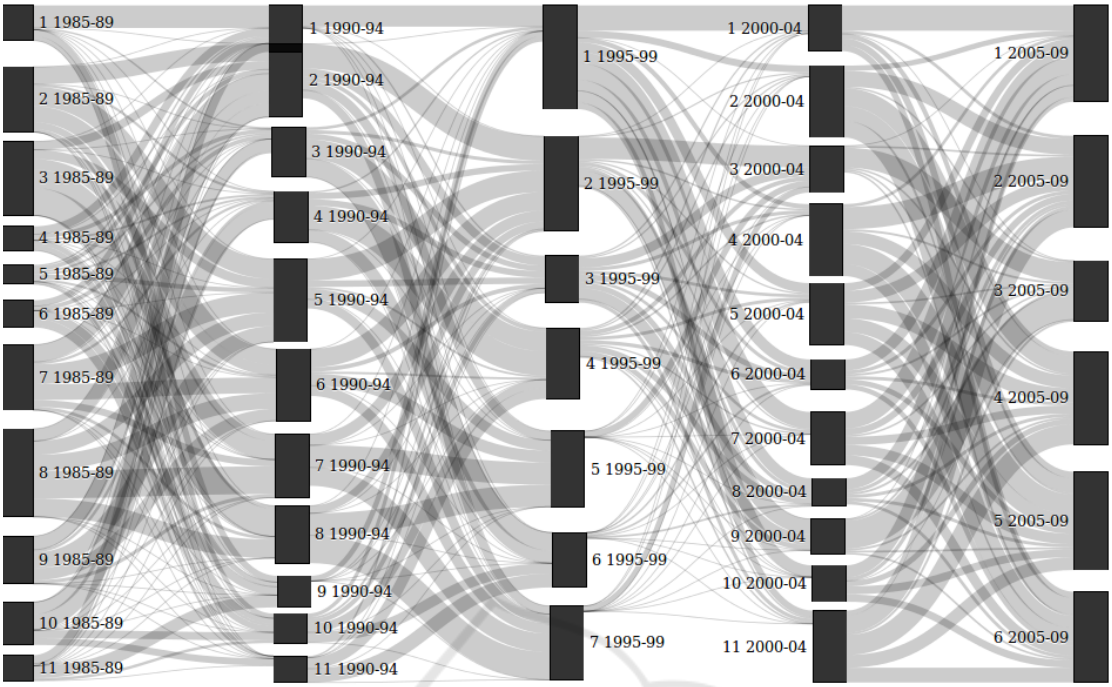

Figure 8: Evolution of communities in the PACS network for every five year intervals. The size of a block indicates the size of a PACS community. The width of shaded flows corresponds to the fraction of PACS codes moving from one community to another.

the in-strength and out-strength of node i (sub-field) as follows.

$$i_{in} = \sum_{j \in V(G) \,:\, \text{i and j have same first digit}} w_{ij} \qquad (2)$$

$$i_{out} = \sum_{j \in V(G) \,:\, \text{i and j have different first digit}} w_{ij} \qquad (3)$$

Figure 3 shows that ratio of in and out strength for each node for the time period 1985-2010. We observe that for nodes $\{10, \cdots 29, 70, \cdots 79\}$ which represent the fields Physics of Elementary Particle and Fields, Nuclear Physics and Condensed matter physics have the in out ratio is more than one indicating that these fields more interconnected and less interdisciplinary.

The intensity plot of strength between the sub-fields of physics for second level of hierarchy are shown in Figure 7. We observe that there is a high cross-discipline interactions between Condensed matter Physics (60,70) and Interdisciplinary Physics (80). From the both the time periods we observe that most of the papers published in Nuclear Physics (20), Atomic and Molecular Physics (30), are either single PAC papers or contains more than one PAC from same disciplines.

5.3 Third Level

We use the CFinder algorithm to study the evolution of scientific fields which is based on the Palla et al. (Palla et al., 2005),(Kumpula et al., 2008) clique percolation method (CPM). A *k-clique community* is defined as the union of all k-cliques that can be reached from each other through a series of adjacent k-cliques. Two k-cliques are said to be *adjacent* if they share k-1 nodes.

Before applying the CPM method to PACS network we preprocess the data by removing links weaker than a fixed threshold W. We take the threshold as the geometric mean of all its edge weights. We remove all links whose weight is less than threshold value.

The value of the k is chosen based on the following properties (a) The number of communities is as

Table 2: Basic statistics of PACS network 1985-2010.

Number of nodes	914
Number of edges	51015
Average degree	111.63
Network diameter	4.0
Average Path Length	2.031
Modularity	0.515
Graph density	0.122

large as possible (b) Avoid the excessively large communities. We consider value of k is 10 and shown the evolution of communities over every 5 years span (see Figure 8). We observe that from this Figure 8, very few cross links in the 1985 to 1989 and very high cross links in 2005 to 2009. This indicate the interdisciplinary increase over the years.

6 DISCUSSION AND CONCLUSION

We have studied the evolution of sub fields of physics with the help of PACS codes used in all published articles in Physical Review journal from 1985 to 2010. We have constructed network using PACS codes. First we have analyzed single field articles at different hierarchy levels of PACS codes and observed that high fraction of single field articles are from the fields of Physics of elementary particles and fields, Nuclear Physics and Condensed matter physics. We have observed that this trend is uniform over all the years by performing similar analysis for every five years. We studied the inter and intra strength of PACS codes and showed that Physics of elementary particle and fields, Nuclear Physics and Condensed matter physics is greater than one indicating that these fields have stronger intra connections compared to inter connections.

Furthermore we studied the correlation between the different fields of physics at different levels of PACS hierarchy starting form first level to third level. From the first level analysis we observed that both structural and electronic condensed matter physics are strongly related in the sense that many published articles cite PACS codes from both these fields. In the second level we analyzed strength of connections between the fields of physics using intensity plot. The analysis shows that Nuclear Physics (20) and Molecular Physics (30) have weak connections to the other fields. Whereas there is a high cross-discipline interactions between condensed matter physics (60, 70) and interdisciplinary physics (80).

In future direction, we can quantify the interdisciplinary by using this PACS code network over years. To do this we need to define a measure which measures the interdisciplinary. We constructed a PACS network, where the nodes represents the PACS codes and there is an edge between two PACS codes if they have appeared in the same article. We can also create network, there is an edge between two PACS codes if both PACS codes used by single author. This also gives us one more direction to measure the interdisciplinary.

ACKNOWLEDGEMENTS

The authors would like to thank American Physical Society for providing us with the meta data of all Physical Review journal articles with PACS codes.

REFERENCES

Battiston, F., Musciotto, F., Wang, D., Barabási, A.-L., Szell, M., and Sinatra, R. (2019). Taking census of physics. *Nature Reviews Physics*, 1(1):89.

Dias, L., Gerlach, M., Scharloth, J., and Altmann, E. G. (2018). Using text analysis to quantify the similarity and evolution of scientific disciplines. *Royal Society open science*, 5(1):171545.

Herrera, M., Roberts, D. C., and Gulbahce, N. (2010). Mapping the evolution of scientific fields. *PloS one*, 5(5):e10355.

Jia, T., Wang, D., and Szymanski, B. K. (2017). Quantifying patterns of research-interest evolution. *Nature Human Behaviour*, 1(4):0078.

Kumpula, J. M., Kivelä, M., Kaski, K., and Saramäki, J. (2008). Sequential algorithm for fast clique percolation. *Physical Review E*, 78(2):026109.

Liu, W., Nanetti, A., and Cheong, S. A. (2017). Knowledge evolution in physics research: An analysis of bibliographic coupling networks. *PloS one*, 12(9):e0184821.

Martin, T., Ball, B., Karrer, B., and Newman, M. E. J. (2013). Coauthorship and citation patterns in the physical review. *Phys. Rev. E*, 88:012814.

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409.

Omodei, E., Poibeau, T., and Cointet, J.-P. (2013). A symmetric approach to understand the dynamics of scientific collaborations and knowledge production. In *4e conférence sur les modèles et l'analyse des réseaux: Approches mathématiques et informatiques*, pages 10–p.

Palla, G., Ábel, D., Derényi, I., Farkas, I., Pollner, P., and Vicsek, T. (2005). K-clique percolation and clustering in directed and weighted networks. *Bolayai Society Mathematical Studies*.

Palla, G., Barabási, A.-L., and Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(7136):664–667.

Pan, R. K., Sinha, S., Kaski, K., and Saramäki, J. (2012). The evolution of interdisciplinarity in physics research. *Scientific reports*, 2.

REDNER, S. (2005). Citation statistics from 110 years of physical review. *Physics today*, 58(6):49–54.

Sinatra, R., Deville, P., Szell, M., Wang, D., and Barabási, A.-L. (2015). A century of physics. *Nature Physics*, 11(10):791.

APPENDIX

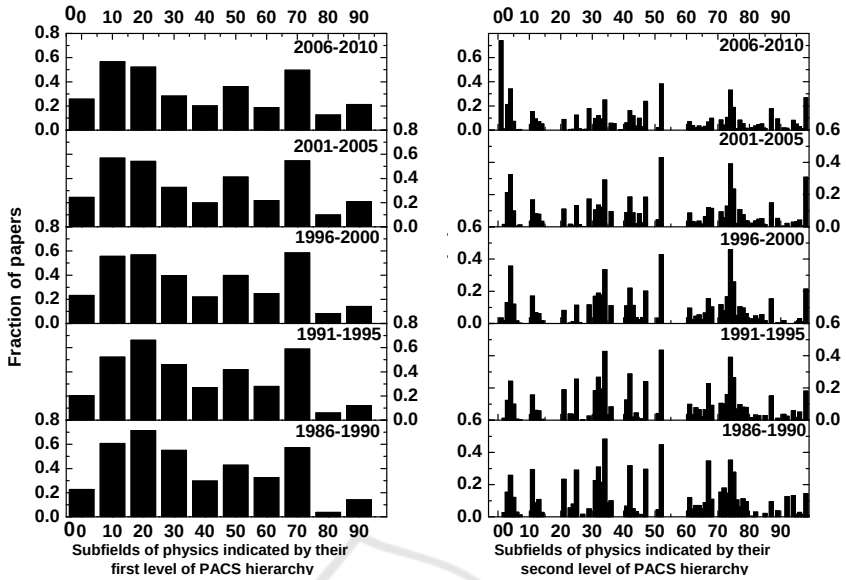

Figure 9: Fraction of single field papers for both first and second level for five year interval.

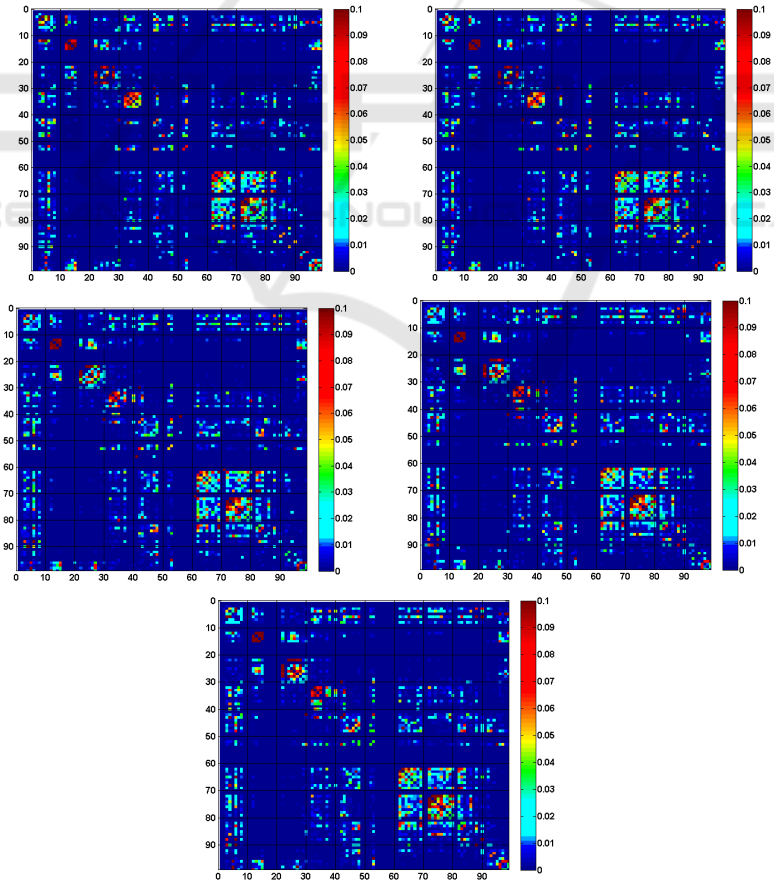

Figure 10: Intensity of two subfield for every 5 years span((a) 1986-1990 (b)1991-1995 (c) 1996-2000 (d) 2001-2005 (e) 2006-2010).

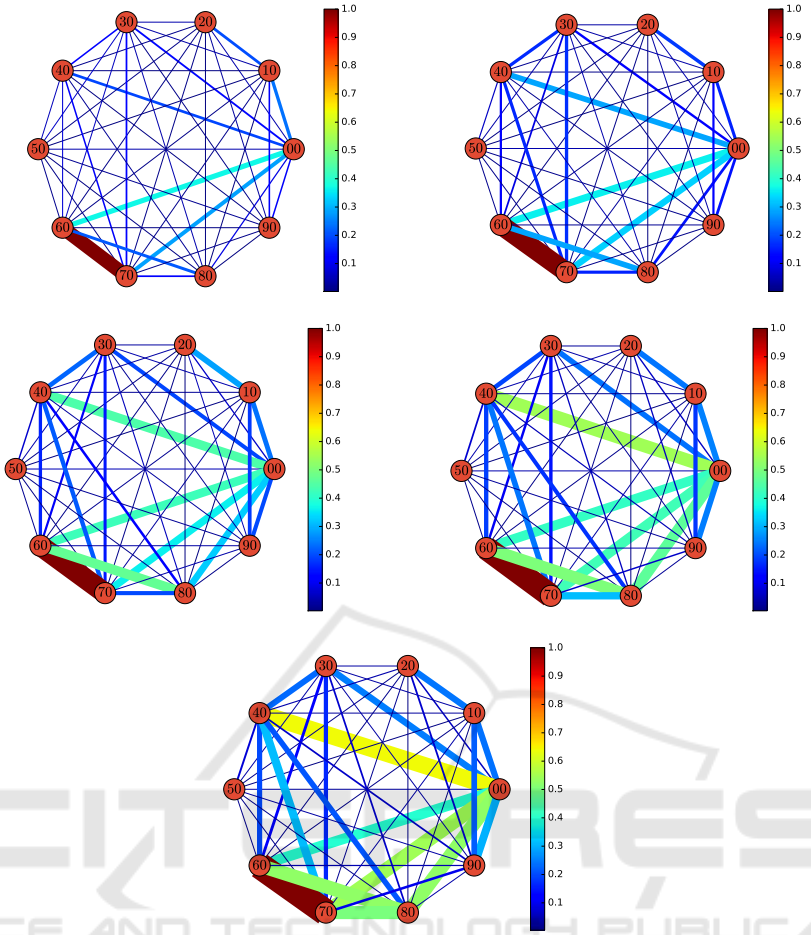

Figure 11: Strength of network for every 5 years span ((a) 1986-1990 (b)1991-1995 (c) 1996-2000 (d) 2001-2005 (e) 2006-2010).