Evolution of subfields in Physics

Soma Sanyal
Postdoctoral Research Associate
School of Library and Information Science
Indiana University

Katy Borner
Associate Professor
School of Library and Information Science
Indiana University

Shashikant Penumarthy
Doctoral Student
School of Library and Information Science
Indiana University

Robert Goldstone
Full Professor
Department of Psychology & Brain Sciences
Indiana University

Abstract

The Physics and Astronomy Classification Scheme (PACS) has been introduced by the American Institute of Physics (AIP) to identify fields and sub-fields of physics. This work analyses the co-occurence of PACS numbers in 263,300 documents published by the AIP between 1985-2005. We find that the network of PACS co- occurences is an extremely dense network with small world properties. We also use the clique percolation method to study the communities of PACS numbers for each year. The communities formed strongly overlap with each other.

We are able to identify the major fields in physics, as well as smaller sub-fields. We find that though different sub-fields change with time, the average path length of the network decreases due to the presence of strongly overlapping communities that connect most of the sub-fields to one another. The PACS codes in these communities have high betweenness centrality and correspond to basic theories and techniques in physics. The autocorrelation of the PACS over the 21 years show that there is a small positive correlation of the growth of a field with its recent growth.

The Network and Its Properties

The Data

In the field of physics, most journals use author supplied keywords or the Physics and Astronomy Classification Scheme (PACS) supported by the American Institute of Physics (AIP) for classifying articles in various sub-fields. The PACS is a hierarchical subject classification scheme introduced by the AIP containing 10 broad subject categories subdivided into narrower sub-categories. Authors are required to specify one or more PACS codes for their articles. Articles published in AIP journals have multiple PACS codes assigned to them.


PHYSICAL REVIEW E 71, 056104 (2005)

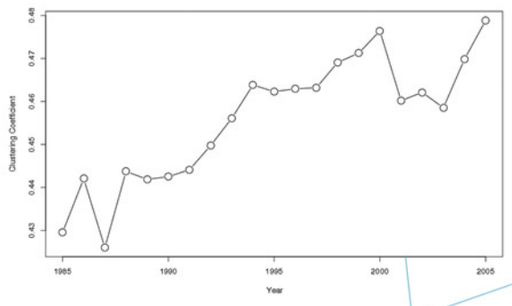
Diffusion-annihilation processes in complex networks

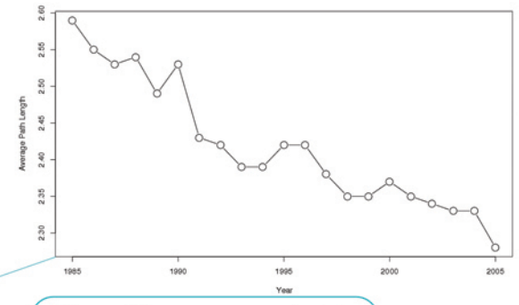
Michele Catanzaro,[1] Marián Boguñá,[1] and Romualdo Pastor-Satorras[2]

We generate a detailed analytical study of the $A+A \rightarrow \emptyset$ diffusion-annihilation process in complex networks. By means of microscopic arguments, we derive a set of rate equations for the density of A particles as vectors of a given degree, valid for any generic degree distribution, and which we solve for uncorrelated networks. For homogeneous networks (with bounded fluctuations), we recover the standard mean-field solution, i.e., a power-scale density decreasing as the inverse of time. For heterogeneous (scale-free networks) as the infinite network size limit, we obtain instead a density decreasing as a power law, with an exponent depending on the degree distribution. We also analyze the role of finite size effects, showing that any finite scale-free network tends to the mean-field behavior, with a prefactor depending on the network size. We check our analytical predictions with extensive numerical simulations on homogeneous networks with Poisson degree distribution and scale-free networks with different degree exponents.

DOI: 10.1103/PhysRevE.71.056104 PACS number(s): 89.75.-k, 87.23.Ge, 05.70.Ln

Generating the Network





Small-World Network

Network becomes more cliquish over time, but there are also shorter paths connecting subfields

The Algorithms

We used the clustering coefficient and the average path length to determine whether the network is a small world structure or not. This is because it is well known that information flow is most efficient in a small world structure.

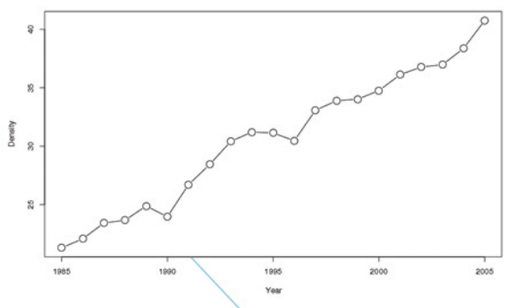
Clustering Coefficient is defined as the ratio of triangles to the number of connected triples in the network.

Average Path Length is defined as the average value of the shortest paths connecting all possible pairs of nodes in the network.

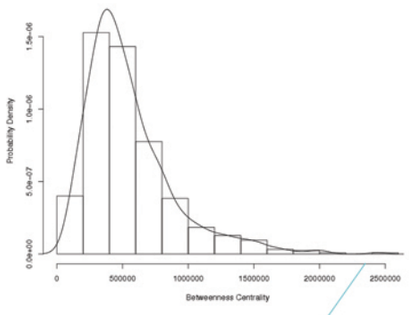
Density of a network is defined as the ratio of the number of edges of a network to the square of the total number of nodes.

A Small-World Network has a higher clustering coefficient and shorter average path length than a random network of similar density.

Betweenness Centrality is a measure that describes a nodes position in a network in terms of the flow it is able to control. Mathematically, it is defined as the number of shortest paths between pairs of nodes that pass through a given node.



Density is increasing over time



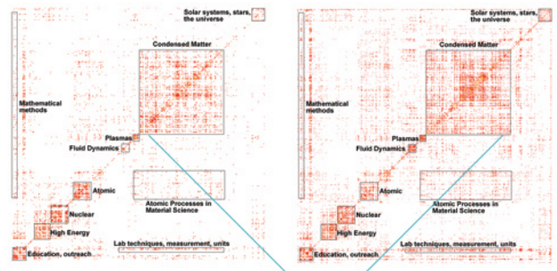
The Betweenness Centrality distributions of PACS codes

Subfields and Their Evolution

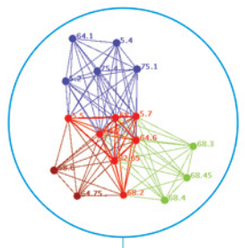
1985 - 2005 Co-occurence Matrices

The co-occurence matirx indicated the presence of clusters and communities, which is why we used clustering techniques to identify the communities. K-means clustering, agglomerative and divisive clustering techniques work for networks with well defined non-overlapping clusters. However, in our case all the communities overlap considerably with each other so these techniques did not work. For this reason, we have used the clique percolation method which works better for overlapping communities.

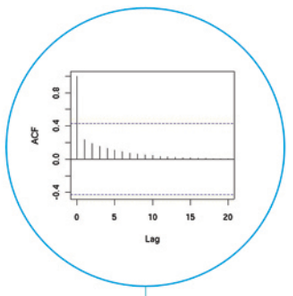
The clique percolation method is based on the following definitions. A k-clique is defined as a complete (fully connected) subgraph of k vertices. Two k-cliques are adjacent if they share k - 1 vertices and they are a connected community if they are a union of adjacent k - cliques. The clique percolation method finds the maximal k - clique connected subgraph for different values of k. So k = 3 gives the largest possible community while higher values of k give smaller communities.



Larger subfields have remained the same



Smaller communities overlap and can be identified by the clique percolation method



It is possible to obtain the auto correlation of the PACS codes. Subfields show momentum. The growth of a field is modestly positively correlated with its recent growth.

Summary & Interpretation of Results

1) The network of PACS codes has a small world structure.
A small world structure means there is very good communication within the subfields in physics.

2) With time the network density increases, the network becomes more cliquish but there are also shorter paths connecting subfields.
Densification of the network implies presence of community structure. Physics has evolved to have increasingly specialized subfields but also has bridging fields to link these specialties.

3) Some of the PACS codes have higher betweenness centrality than others.
These represent techniques and concepts that are common to various subfields in physics mostly related to condensed matter physics as that is one of the largest subfield in physics.

4) The subfields of physics overlap with each other considerably and it is possible to obtain an average autocorrelation for the fields over time.
Subfields show momentum: growth of a field is modestly positively correlated with its recent growth.

Future Work

Based on this analysis, we plan to have a model of science that reflects the following:

- Fields recursively split into sub-subfields
- Once created, fields become increasingly entrenched
- Subfields also emerge that bridge/fuse otherwise distant fields
- The density of different subfields is highly uneven, due to "rich-get-richer" effects.

References & Acknowledgements

1) Bhattacharya, S. (1997). "Cross-national comparison of frontier areas of research in physics using bibliometric indicators." Scientometrics 40(3) pp 385-405

2) Derenyi, I., Palla, G., and Viscek, T. (2005). " Clique percolation in random networks." Phys. Rev. Lett. 94, pp 160202

3) Palla, G., Derenyi, I., Farkas, I., and Vicsek, T. (2005). "Uncovering the overlapping community structure of complex networks in nature and society." Nature 435(7043) pp 814-8.

4) Börner, K., Sanyal, S. and Vespignani, A. (2007). "Network Science." In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Volume 41, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, chapter 12, pp. 537-607.

5) Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005). "Graphs over time: densification laws, shrinking diameters and possible explanations." Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp 177 - 187

James S. McDonnell Foundation, "Studying Complex Systems: 2005 Research Awards"; Cyberinfrastructure for Network Science Center

For more information, please visit the ModSci website: http://modsci.slis.indiana.edu/

ModSci — Scholarly Knowledge Modeling the Structure and Evolution of Science

cyberinfrastructure for NETWORK SCIENCE CENTER

Graphic Design by Elisha Hardy