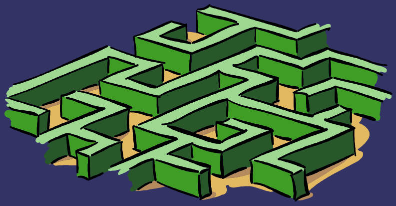
Analyzing research fields within Physics using network science

Soma Sanyal

Cyberinfrastrcture for Network Science Center
School of Library and Information Science
Indiana University, Bloomington

Shashikant Penumarthy
School of Library and Information Science
Indiana University, Bloomington



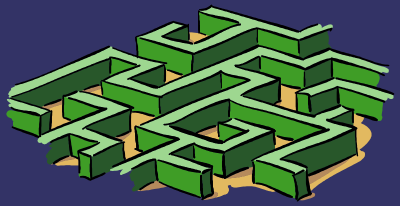
Brief Introduction on the PACS classification scheme

The small-world network of PACS numbers

Community detection and identifying fields in Physics

Overlapping communities

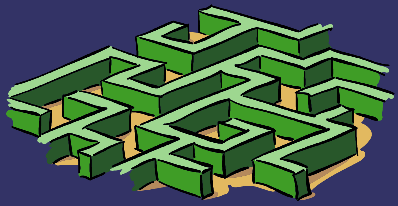
Summary and conclusions



The American Institute of Physics (AIP) in collaboration with other members of the International Council on Scientific and technical Information have prepared a subject classification scheme for physics and astronomy known as the Physics and Astronomy Classification Scheme® (PACS®).

It is a hierarchical classification scheme used extensively by journals and publications in physics and other related subjects.

It is divided into 10 broad categories which were initially subdivided into 66 major topics. In 2003 the scheme was revised to increase the number of topics to 73.



A typical PACS code consists of four numbers followed by a character that can be an uppercase letter or a “+” or ” -” This is followed by a lower case letter.

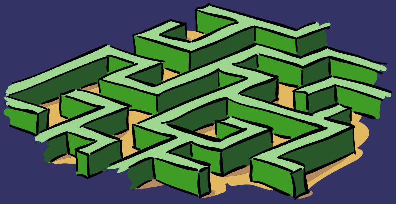
74.50.+r 44.18.-i 11.20.Bk

The first digit denotes the main category out of the 10 broad categories specified in the first level.

The second digit gives the more specific field within that category.

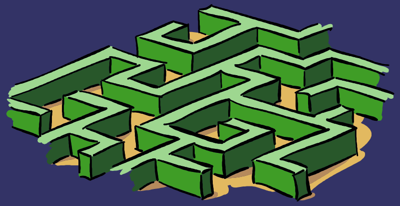
The other two digits denote the third level of hierarchy and specify topics within that field.

The “+” or ” - “ characters indicate whether there are any more subdivisions.



http://www.aip.org/pacs/

- 00 General
- 10 The Physics of Elementary Particles and Fields
- 20 Nuclear Physics
- 30 Atomic and Molecular Physics
- 40 Electromagnetism,Optics, Acoustics, Fluid dynamics etc.
- 50 Physics of Gases plasmas etc
- 60 Condensed Matter: Structure, mechanical and thermal properties
- 70 Condensed Matter: Electronic structure, electrical, magnetic and optical properties.
- 80 Interdisciplinary physics and others
- 90 Geophysics, Astronomy, Astrophysics.



An example: The code -------- 21.10.Dr

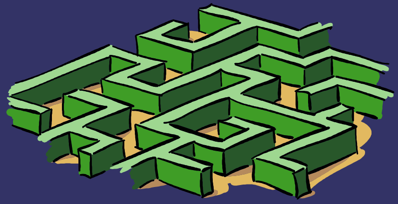
20 -- Nuclear Physics

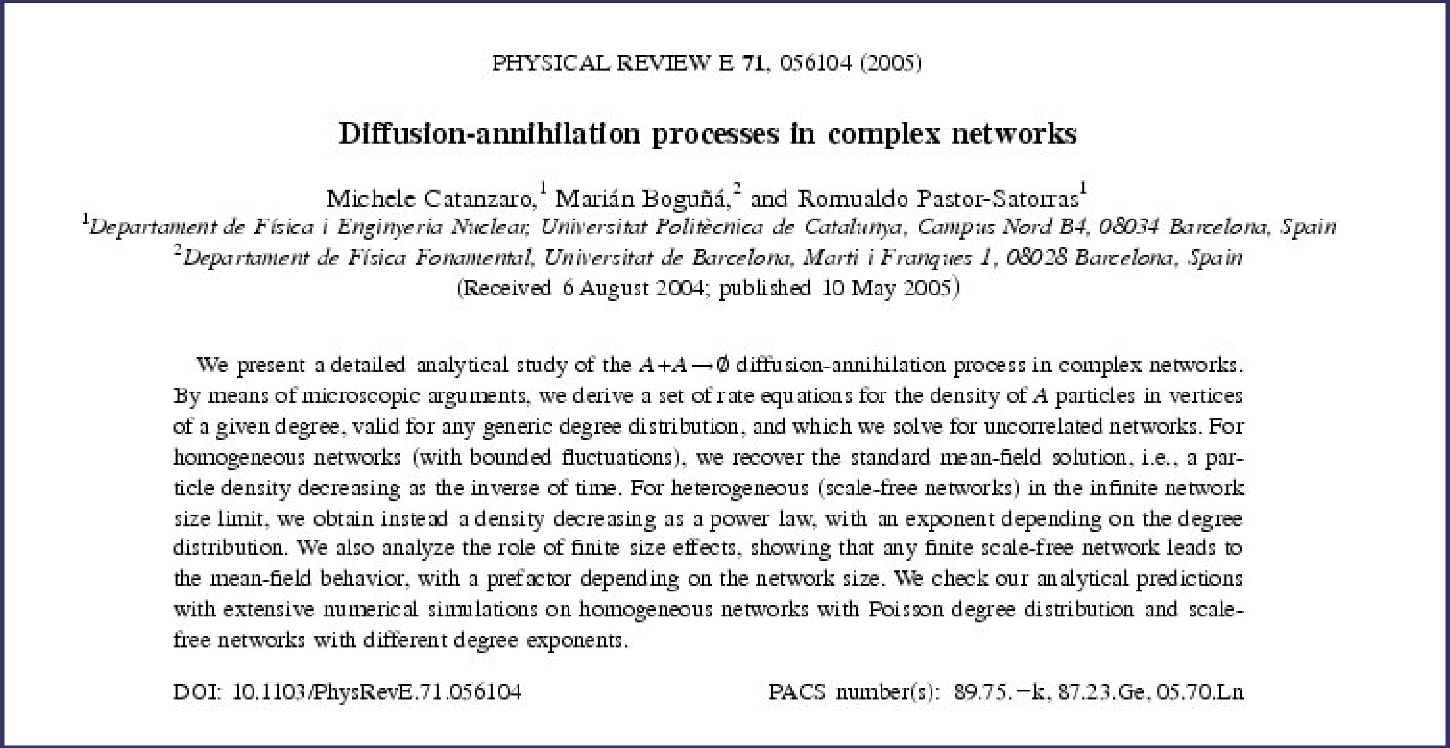
21. -- Nuclear Structure

21.10 -- Properties of nuclei; nuclear energy levels

21.10.Dr -- Binding energies and masses of nuclei

Papers published in the Physical Review journals have one or more of these codes to indicate the fields it corresponds to.





PHYSICAL REVIEW E **71**, 056104 (2005)

Diffusion-annihilation processes in complex networks

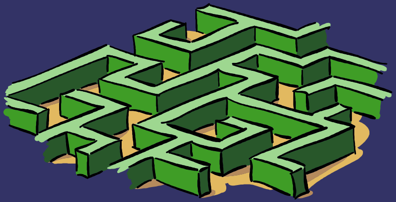
Michele Catanzaro,[1] Marián Boguñá,[2] and Romualdo Pastor-Satorras[1]
[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*
[2]*Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

(Received 6 August 2004; published 10 May 2005)

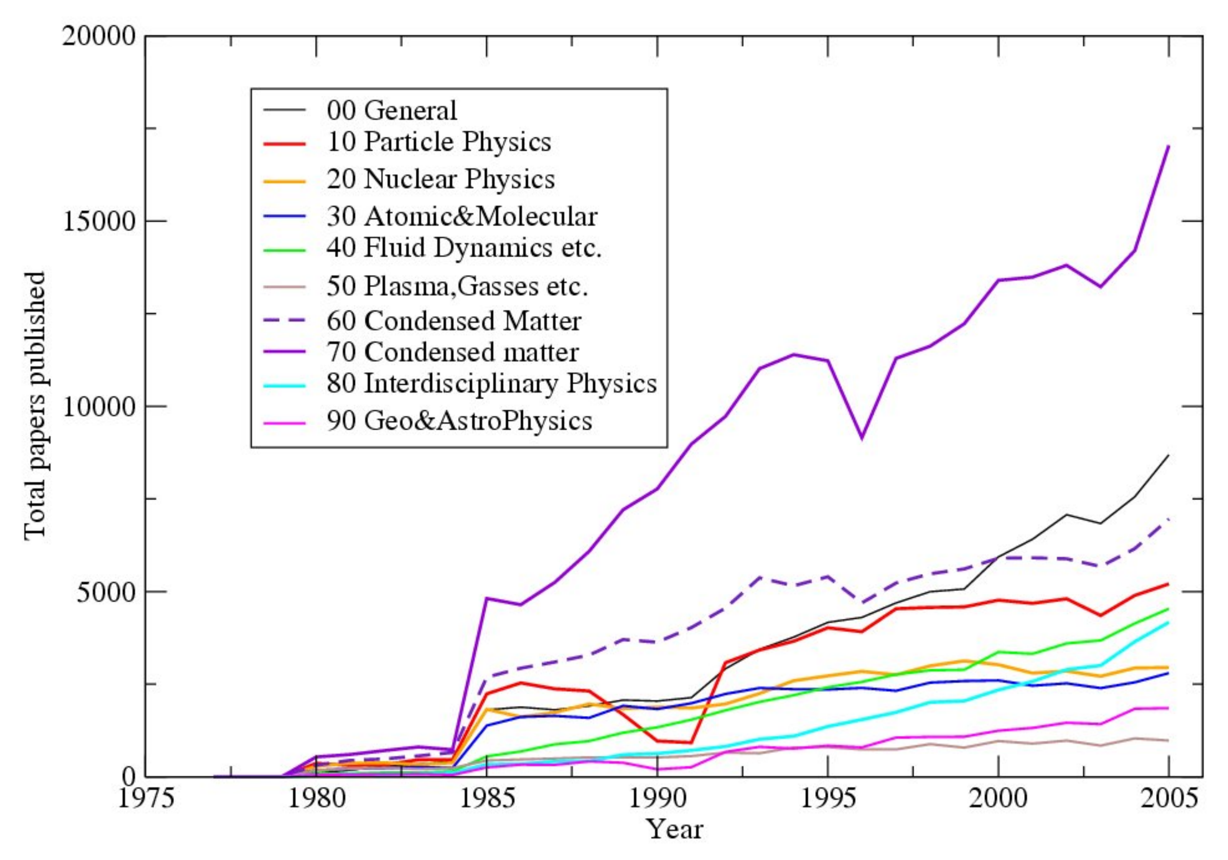
We present a detailed analytical study of the $A+A\to\emptyset$ diffusion-annihilation process in complex networks. By means of microscopic arguments, we derive a set of rate equations for the density of A particles in vertices of a given degree, valid for any generic degree distribution, and which we solve for uncorrelated networks. For homogeneous networks (with bounded fluctuations), we recover the standard mean-field solution, i.e., a particle density decreasing as the inverse of time. For heterogeneous (scale-free networks) in 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 leads 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

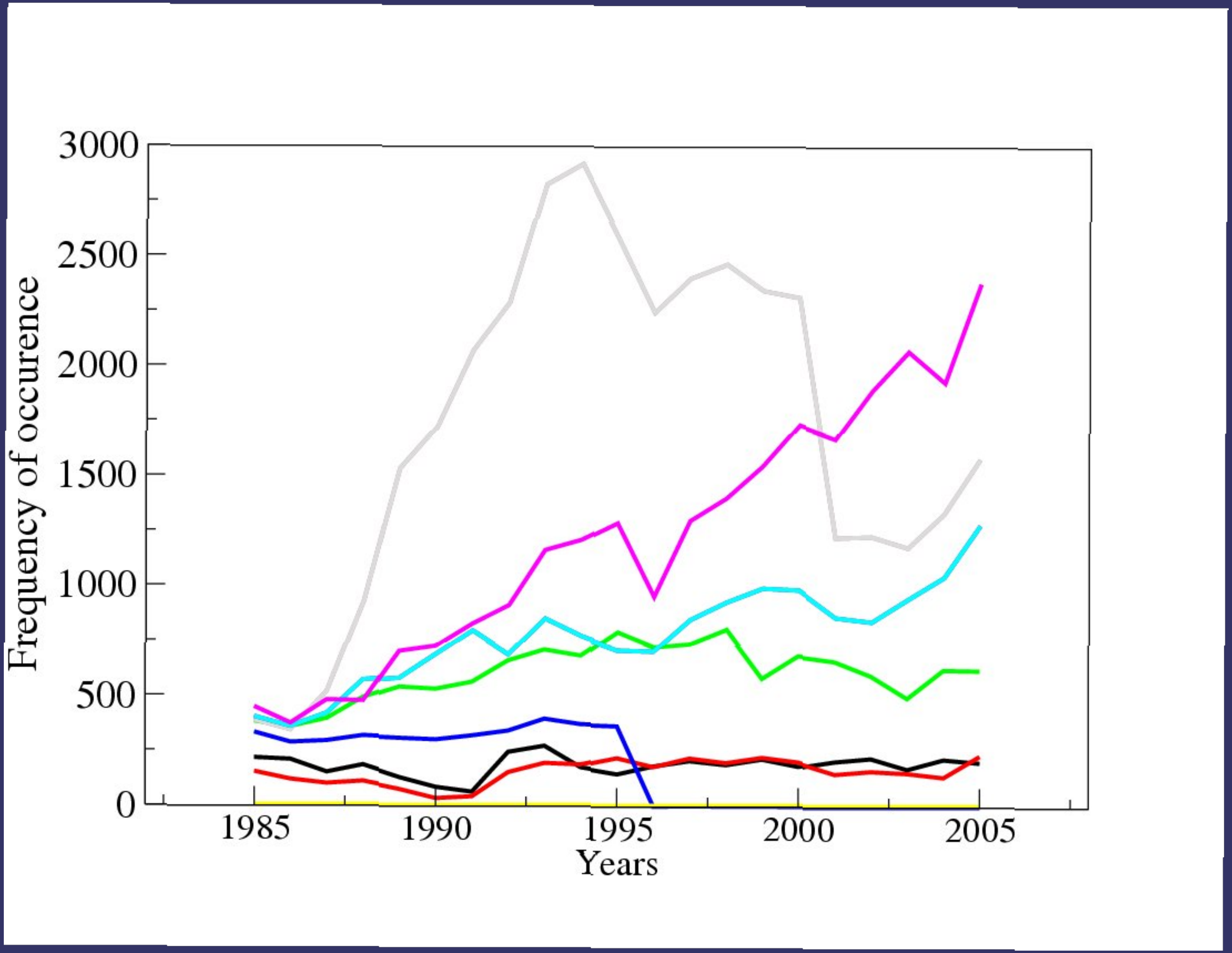
So analyzing the occurrence of PACS codes from the time they were incorporated in 1975 would indicate the activity in different fields of physics within that period.

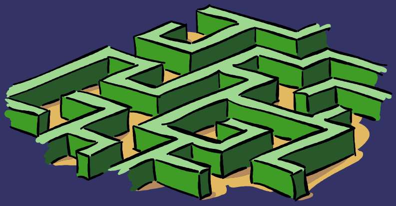


Frequency of occurrence of PACS codes in the first level

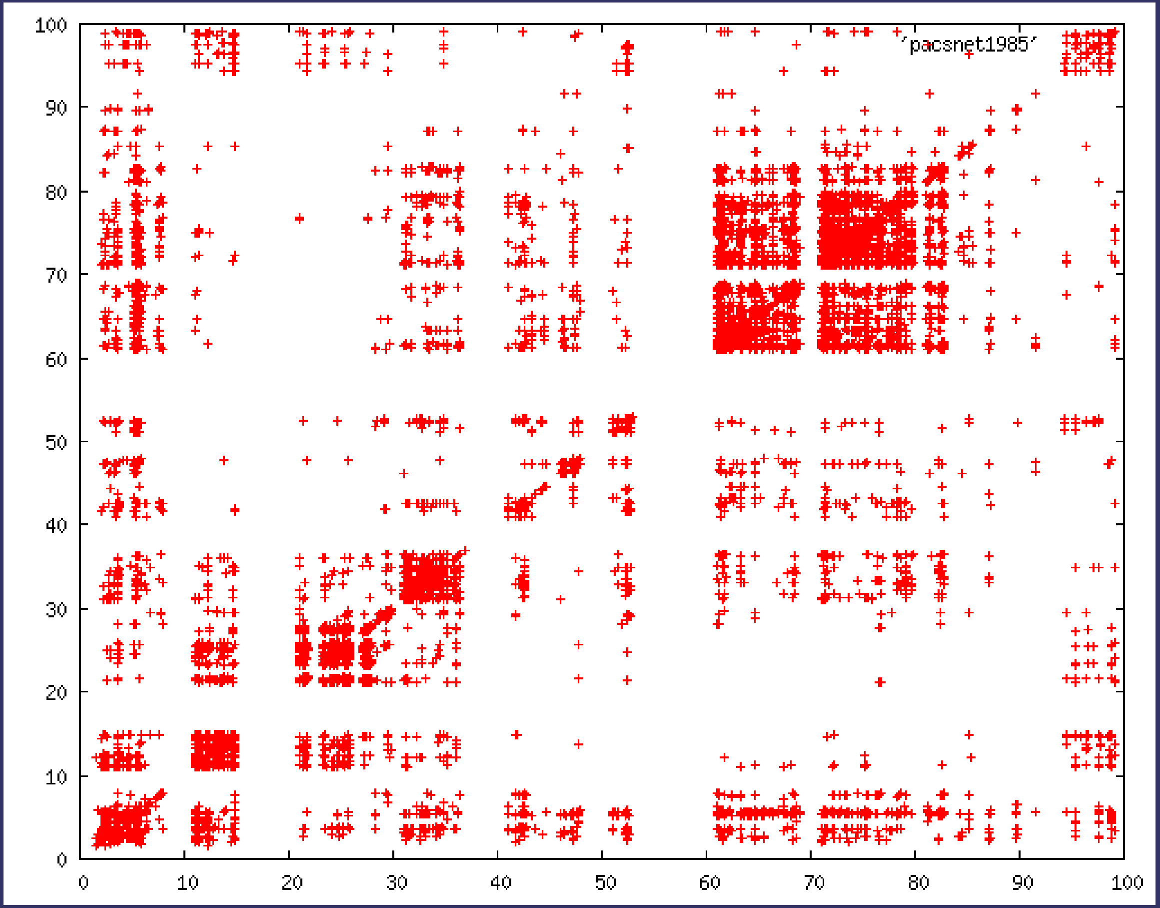


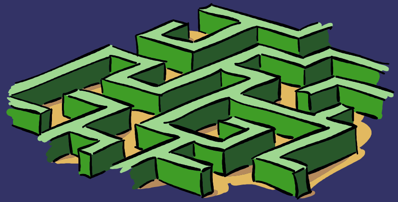
Frequency of occurrence of individual PACS codes





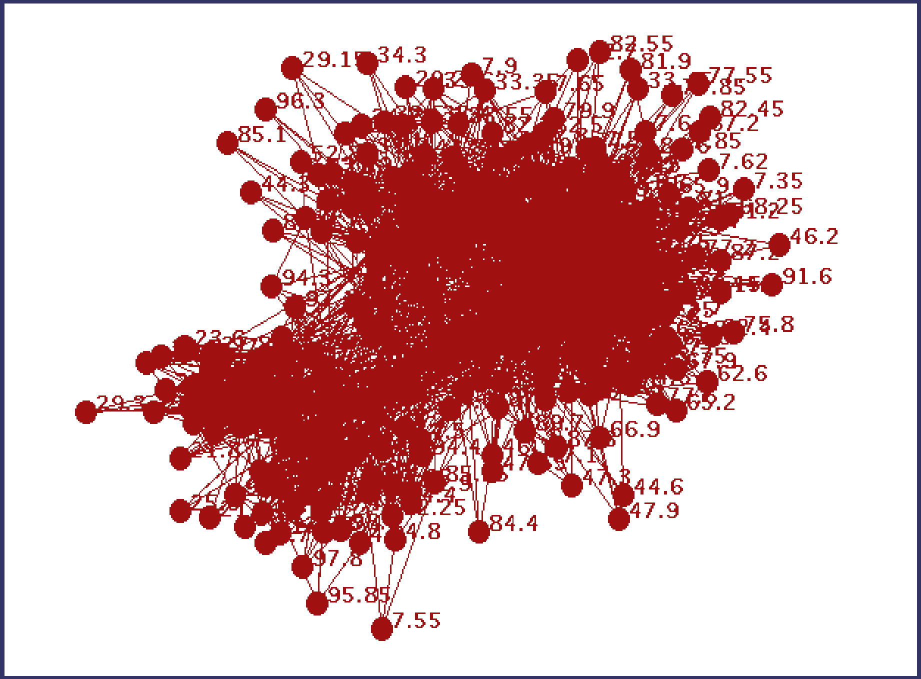
We want to identify how frequently different PACS codes co-occur with each other. So we obtain the co-occurrence matrix for the PACS codes for the time period 1985 – 2005 .



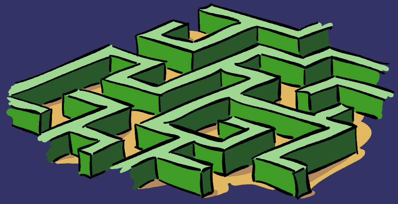


The small world network of PACS codes

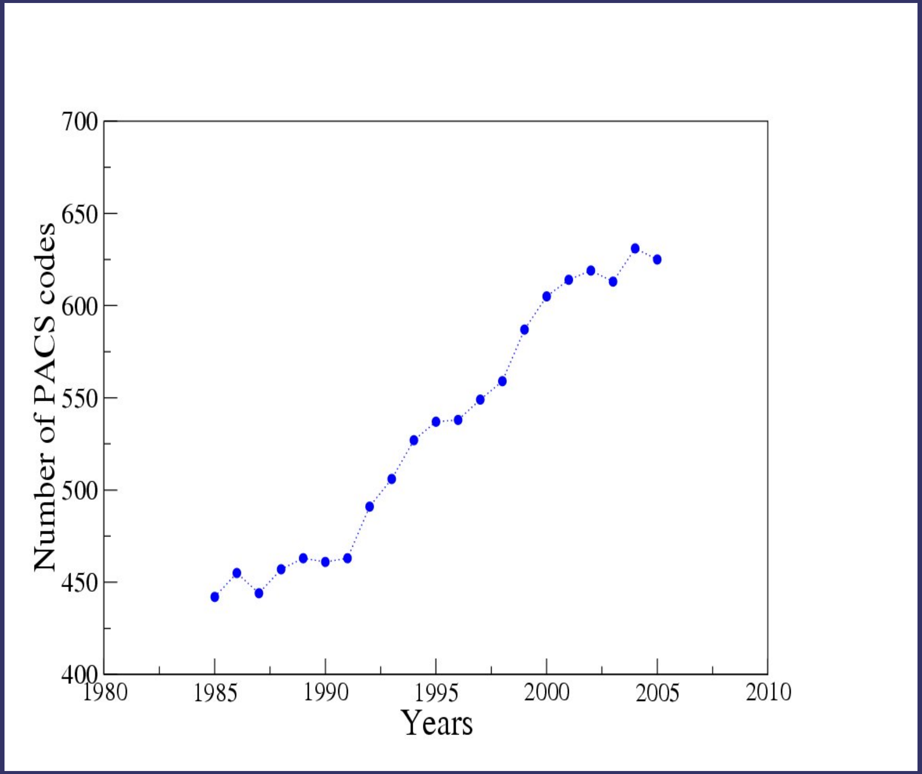
We generate a network of PACS codes. The nodes are the codes used in papers and an edge is established when two PACS codes co-occur in a paper.

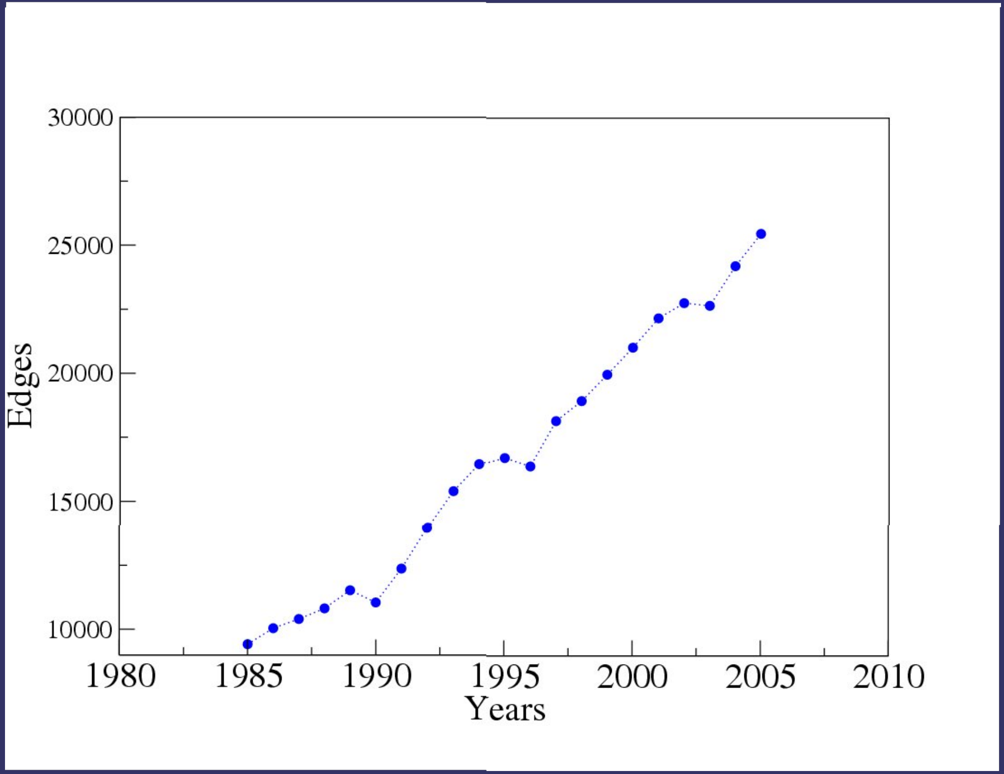


We find that the network consists of a fully connected component with a diameter of 5 for each year.

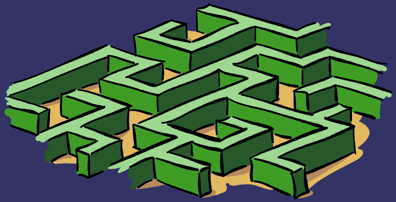


Network properties

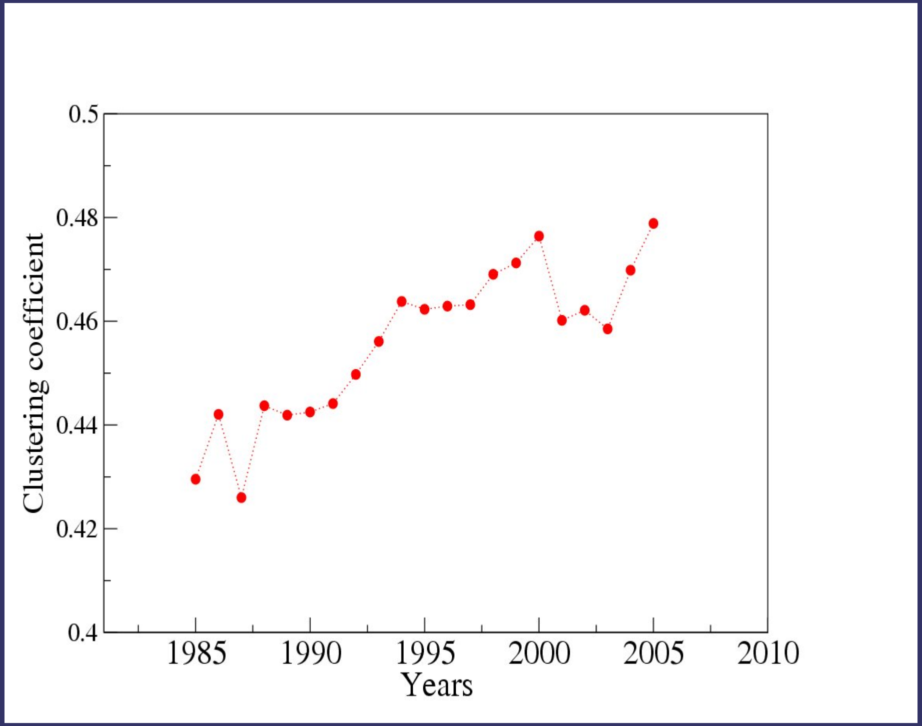


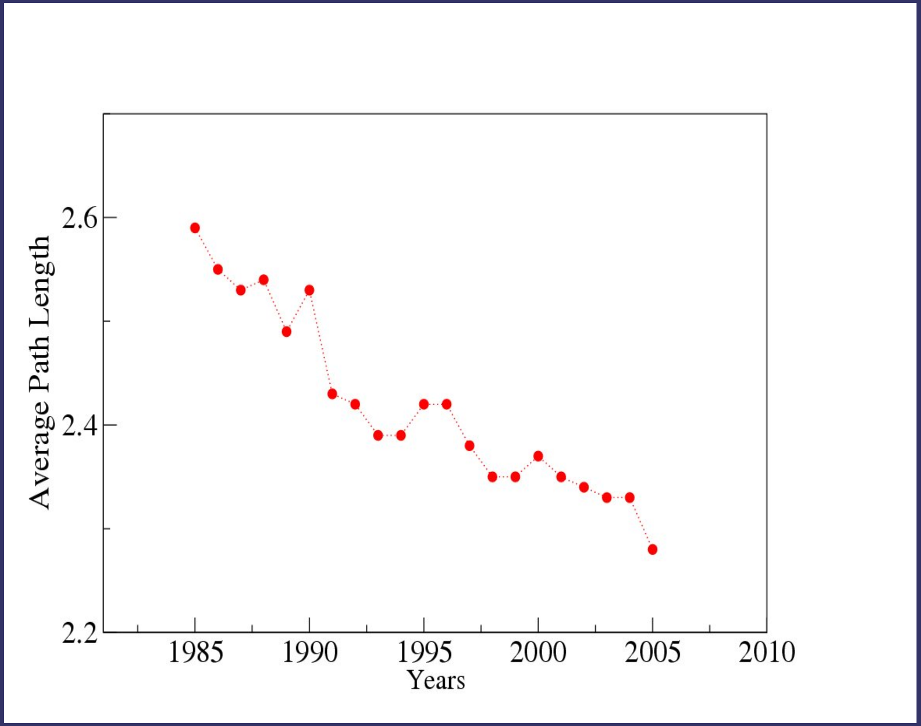


Increase in nodes and edges but diameter remains constant

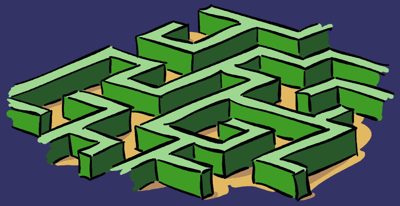


Other properties

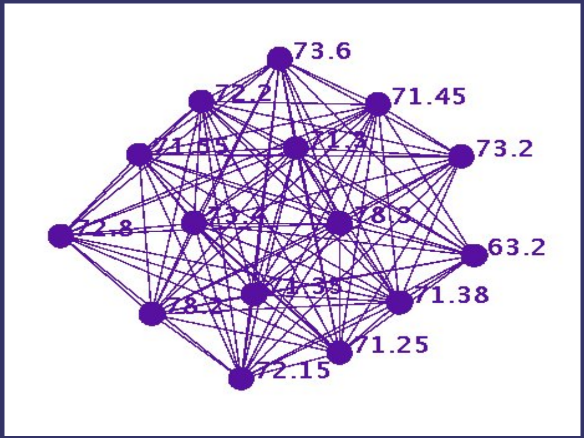




Results seem to indicate not only increased cliques but more “bridges” between these cliques



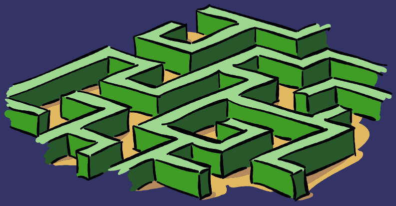
Communities or fields in Physics



The PACS hierarchy is based on how fields in physics are related to each other.

Frequent co-occurrence of multiple PACS in different papers must indicate they belong to the same field or “community”.

This is reflected by the high frequency of co-occurrence in and around the diagonal of our matrix.

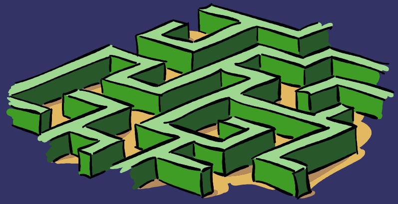


The off-diagonal elements indicate that there is a lot of overlap between the different communities.

Finding overlapping communities using hierarchical or partitioning methods for such networks is difficult.

Overlapping communities can be detected using the *k-clique community* finding algorithm.

Ref: Uncovering the overlapping community structure of
complex networks in nature and society.
Palla G, Derenyi I, Farkas I, Vicsek T.
Nature, 2005 Jun 9;435(7043):814-8



k-clique community detection

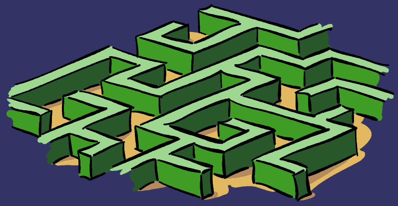
Community --------

Members of a community are linked to most of the other members in the community but not necessarily to all of them.

links between members greater than links to others in network

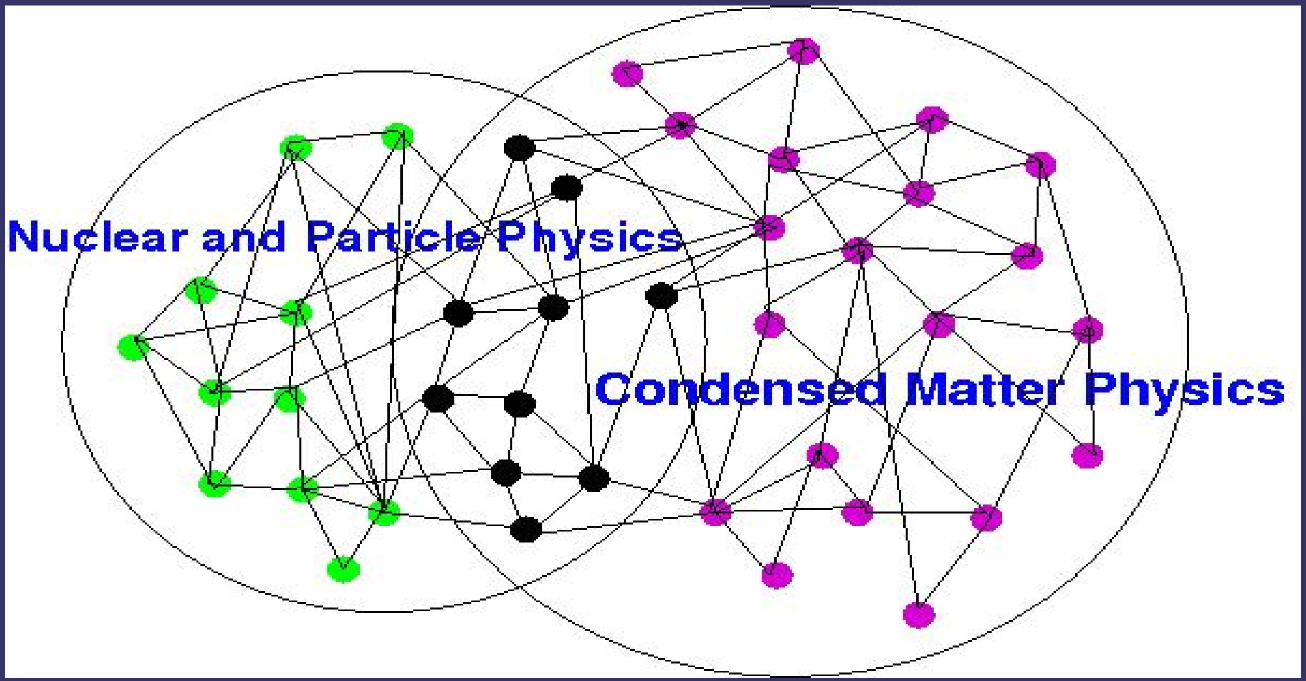
k – clique community ---------

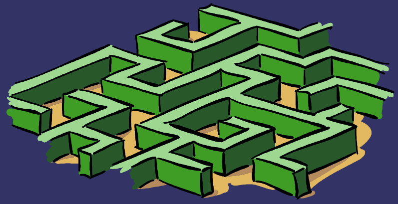
Union of all k-cliques that can be reached from each other through a series of adjacent k-cliques, where two cliques are adjacent if they share k-1 nodes.



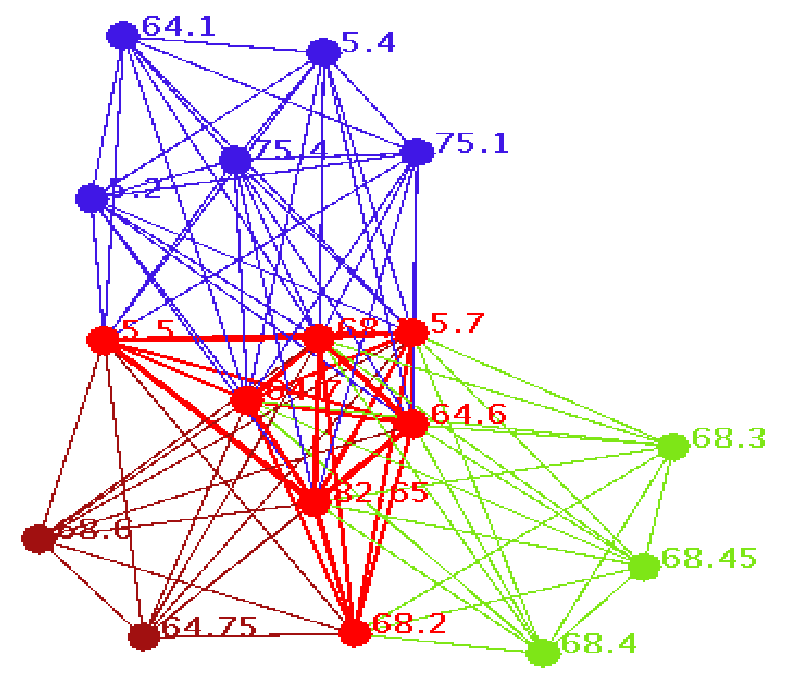
With this approach we can find communities of different sizes.

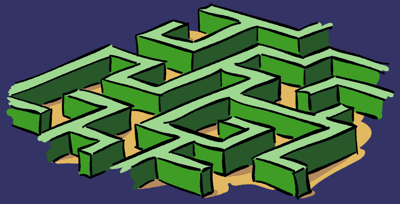
Two giant communities can be identified which have considerable overlap.



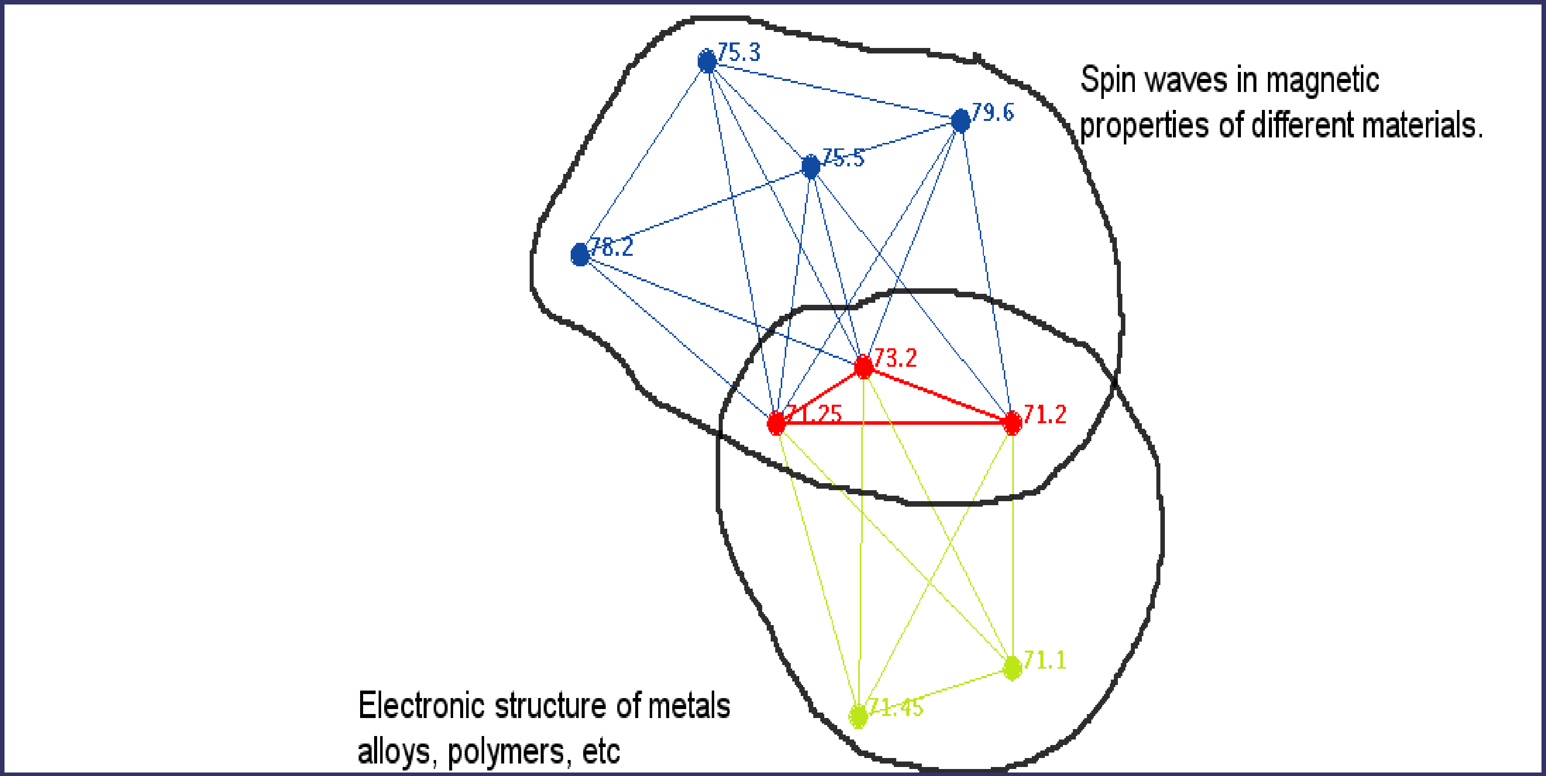


But we also obtain smaller communities pertaining to more specific fields in Physics.



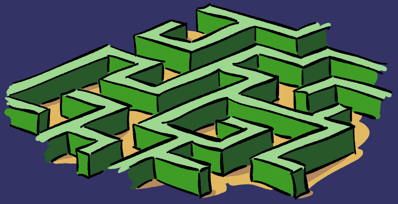


Expected overlaps

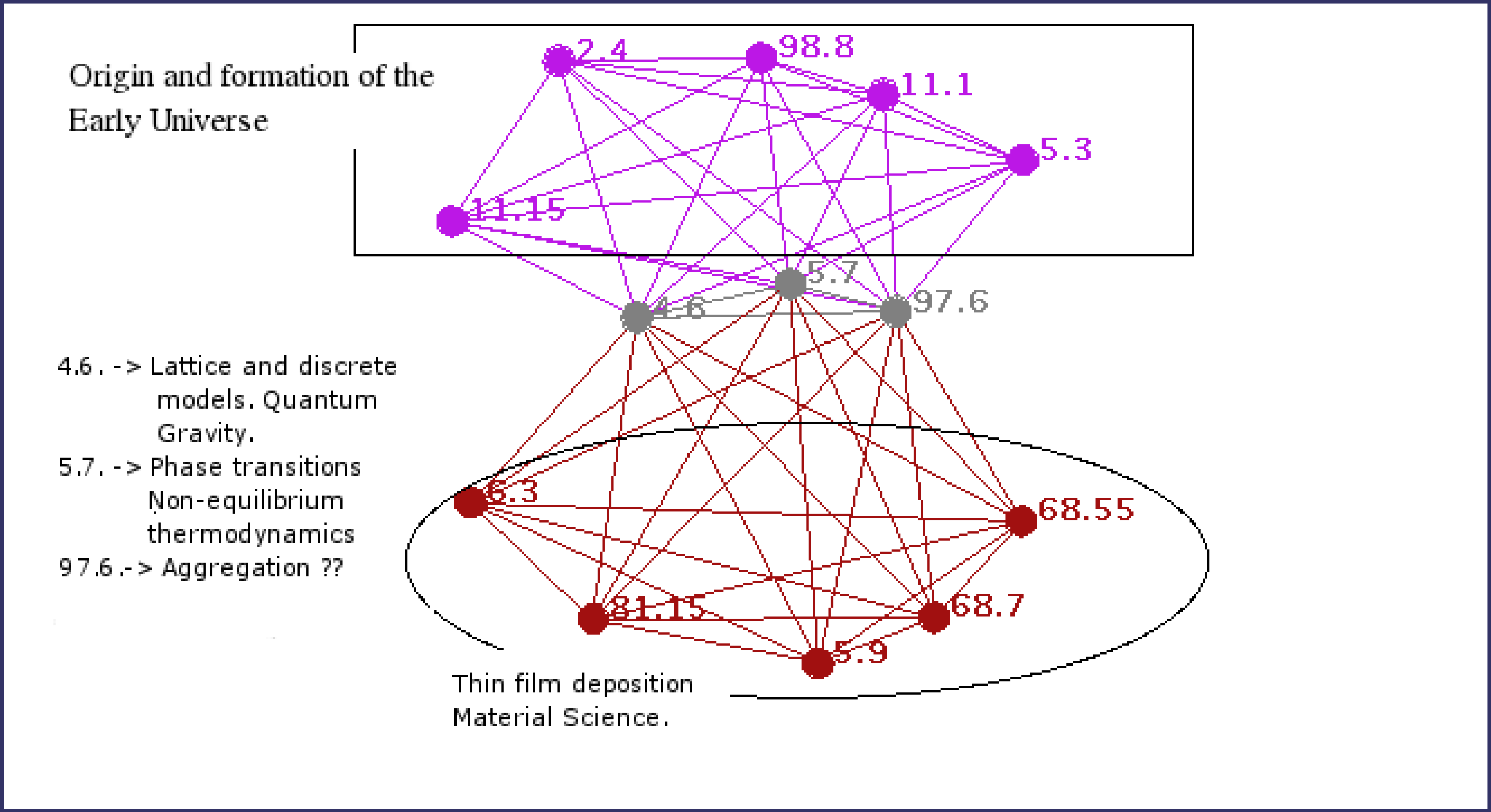


Spin waves in magnetic properties of different materials.

Electronic structure of metals alloys, polymers, etc



Unexpected overlaps



Origin and formation of the Early Universe

4.6 . -> Lattice and discrete models. Quantum Gravity.

5.7. -> Phase transitions Non-equilibrium thermodynamics

97.6 .-> Aggregation ??

Thin film deposition
Material Science.

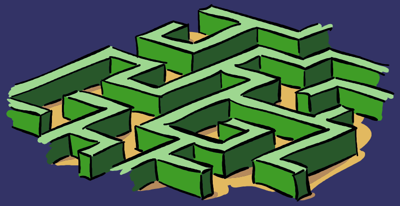
Why do we have these unexpected overlaps ?

Do they reflect the underlying universality in physics ?

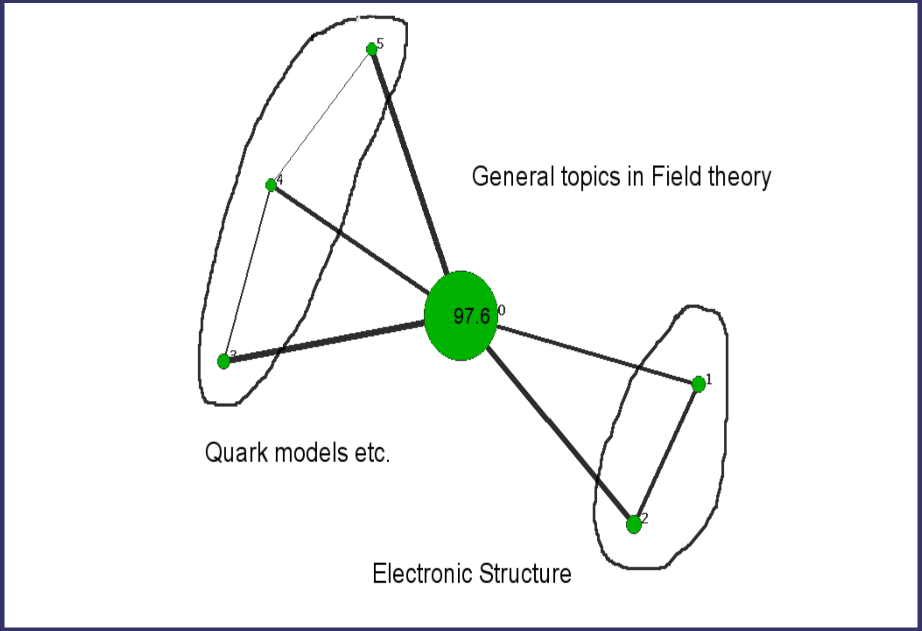
Do physicists apply their knowledge in some areas of physics to other (apparently unrelated) areas ?

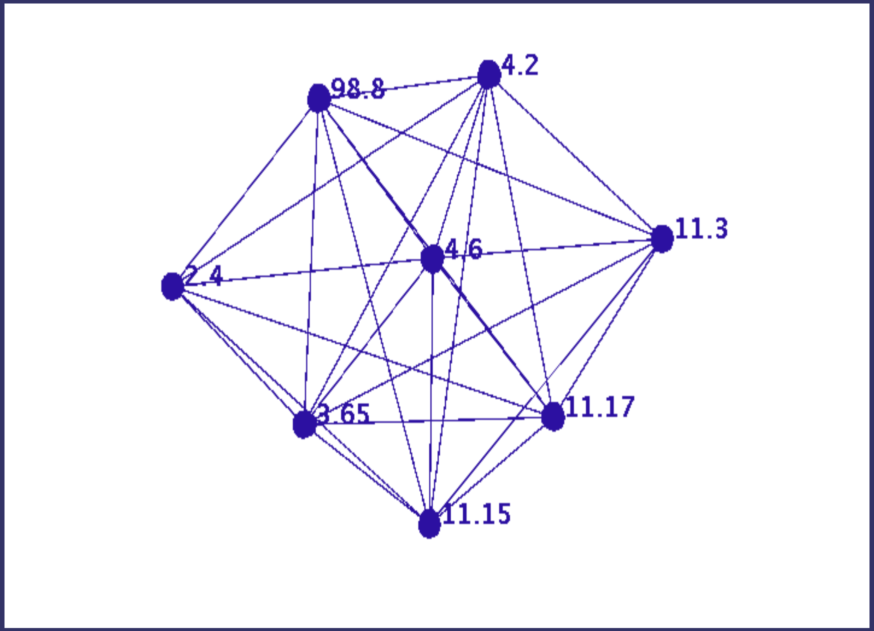
OR

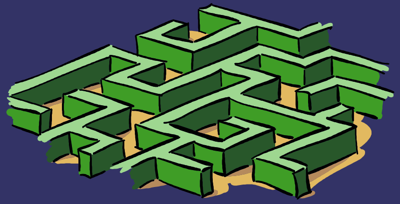
Do they reflect the lack of proper codes to describe specific topics in physics ?



PACS codes are revised periodically. So one way would be to look for these clusters a few years down the line.....





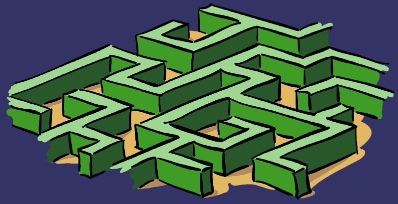


Recall :

Though there was an increase in number of nodes and edges, the average path length was shorter than what would *generally* be expected for a small-world network.

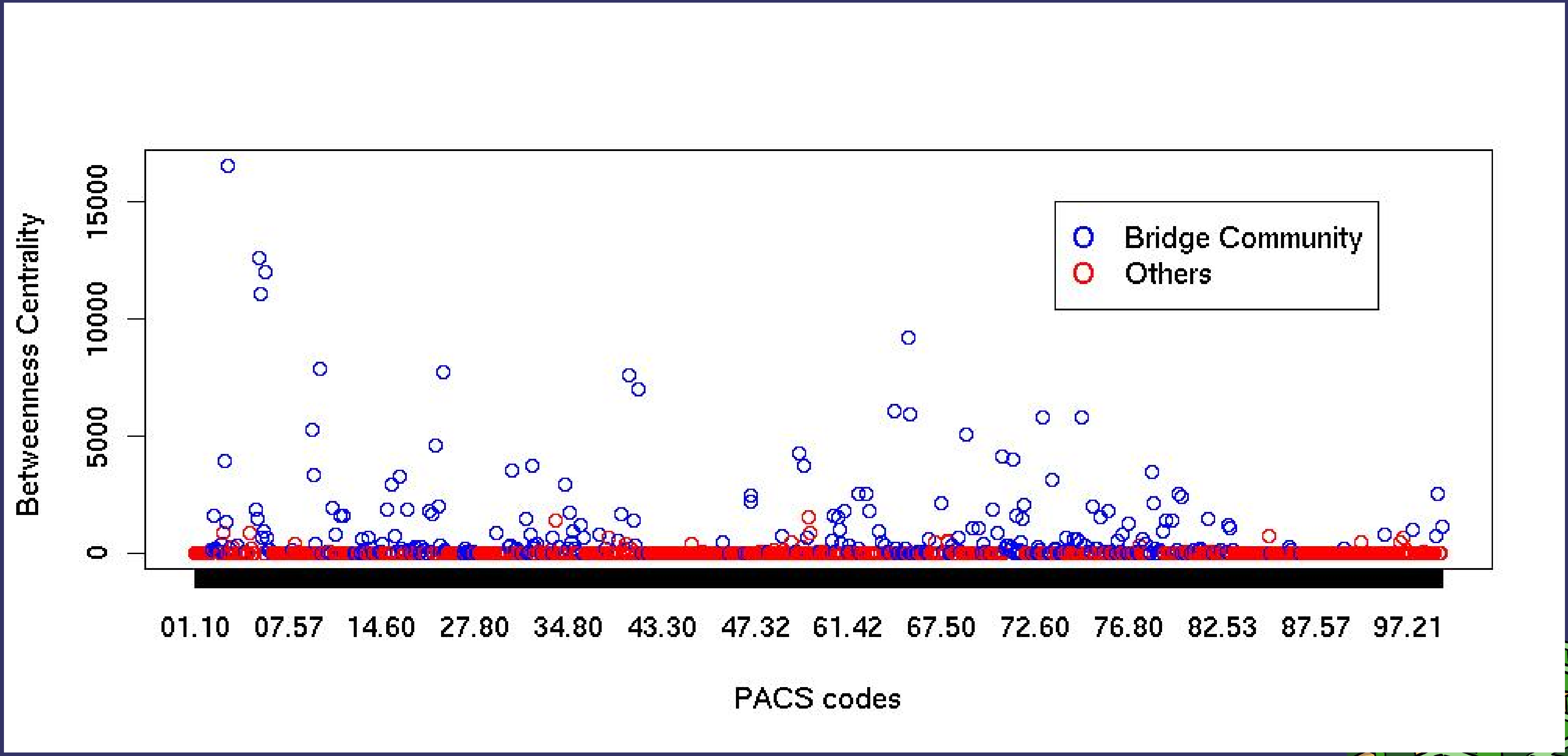
Reason --------- these "bridging" communities.

Nodes which act as bridges in the network have high betweenness centrality



Betweenness Centrality

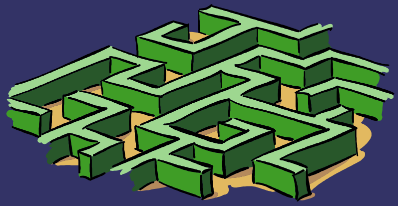
We obtain the betweenness centrality for all nodes in the network and compare them with betweenness centrality of the nodes which belong to our “bridge” community.



A comparison between the betweenness centrality of all the nodes in the network and the nodes in a bridge community indicates that most of the nodes with high betweenness centrality are also members of the bridge community.

So as the network evolves, nodes with high betweenness centrality tend to form cliques that form the “bridge” communities.

This is why we have the average path length decreasing even though the number of nodes increase.

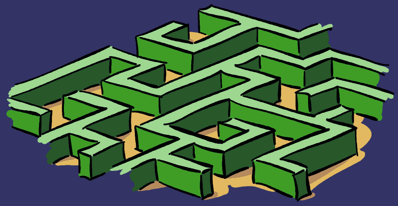


Summarizing:

Frequency of occurrence of top-level PACS codes are directly proportional to the total number of papers published in the particular fields.

Individual PACS codes denote specific concepts within different fields of physics. Their frequency of occurrence denote the number papers published which are related to these concepts.

Frequent co-occurrence of multiple PACS codes denote a community or field in physics.

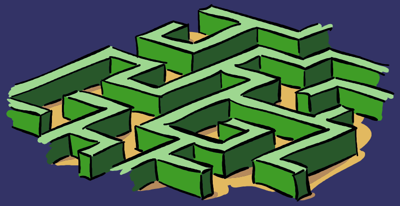


The co-occurrence of PACS codes with each other can be looked as a network of concepts within the field of physics.

The network is a single giant component with small world properties.

The average clustering co-efficient increases with time and the average path length decreases.

Though concepts within physics are forming more cliques, some of these cliques become “bridge” communities linking other large communities.



Acknowledgments:

Katy Börner and Robert Goldstone.
All the members of the IVLab and Cyberinfrastructure for Network Science Center.
Bonnie (Weixia) Huang and all the members of the NWB project.
Please visit http://nwb.slis.indiana.edu/

The R-project for Statistical Computing.
http://www.r-project.org/
The k-clique community detection algorithm (Cfinder)
http://www.cfinder.org/

The James S. McDonnell Foundation Grant in the field of studying complex systems.