Comprehensive journalometry based on complex networks of different kinds: a case study

O. Mryglod[1,2] and Yu. Holovatch[1,3]

[1] Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine, 79011 Lviv, Ukraine
[2] Lviv Polytechnic National University, 79013 Lviv, Ukraine
[3] Institut für Theoretische Physik, Johannes Kepler Universität Linz, 4040 Linz, Austria

Institute for Condensed Matter Physics

Lviv Polytechnic National University

Johannes Kepler Universität Linz

Abstract

Scientometry may serve as a good example of successful application of the complex network theory: to study phenomenon of science it is necessary to analyze a huge amount of interconnected data and it is convenient to present the data in the form of complex networks. In particular, an analysis of a scientific journal can involve complex networks of several kinds: co-authorship, citation, co-citation, network of bibliographic couples. There, the interconnections between separate papers, authors, authors affiliations, etc., are considered. Very often only one type of network is used for evaluation purposes, the other being discarded for no obvious reasons. The goal of our study is to offer a comprehensive analysis of a single periodical (as a subject of the analysis we have chosen the journal *Condensed Matter Physics*, CMP, http://www.icmp.lviv.ua/journal/) based on different possible complex network interpretation of its data. It is the comparison of the data obtained from networks of different origin that enables us to make conclusions about different features of the journal and their correlations.

Co-authorship network

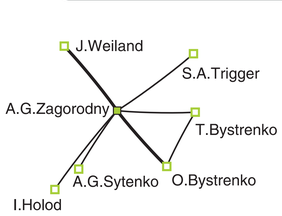


Fig.1. **NODES**: *authors of papers in CMP;* **LINKS**: *coauthorship of a paper in CMP.*

Numerical characteristics of the co-authorship network of the CMP journal (period of 1993–2007): N: number of nodes; L: number of links; $\langle k\rangle$, $k_{\max}$: the mean and maximal node degree, respectively; $\langle C\rangle$, $\langle C\rangle/C_r$: the mean clustering coefficient and its ratio to that of the equivalent random graph; $\langle l\rangle$, $l_{\max}$: the mean and maximal shortest path length.

N	L	$k_{\max}$	$\langle k\rangle$	$\langle C\rangle$	$\langle C\rangle/C_r$	$\langle l\rangle$	$l_{\max}$
944	1370	26	2.9	0.605	196.8	4.82	10
Maximal number of coauthors in one paper $= 8$							
Average number of coauthors per one paper ≈ 2							

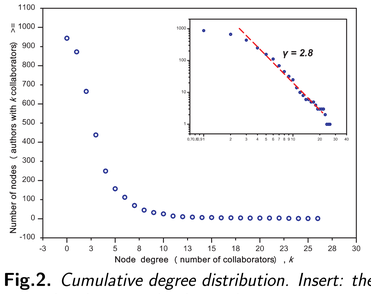
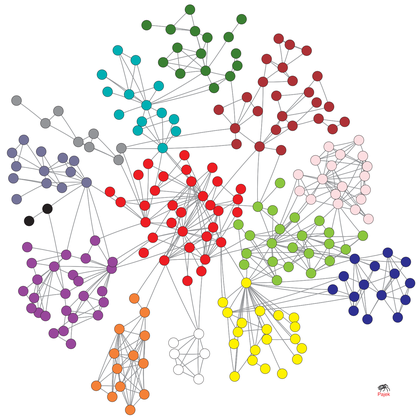


Fig.2. *Cumulative degree distribution. Insert: the same distribution in log-log scale.*

The network has features of a scale-free small world: (i) short mean distance between any two nodes; (ii) large clustering coefficient and (iii) node-degree distributions close to a power law [1].



The co-authorship network consists of one giant component (strongly connected cluster) with 233 nodes, numerous small clusters (the next-largest clusters contain only 15 nodes!) and 72 isolated nodes.

Fig.3. *Community structure (obtained by the Girvan-Newman algorithm [2]) in the largest connected cluster. Communities in the coauthorship network correspond to different research groups from one institution or authors with common scientific interests from different institutions. The network visualization is performed by the Pajek and NetDraw softwear [3].*

Citation network

Citation network is a bipartite network containing nodes of two kinds: the citing and the cited nodes. Depending on available data and purposes of study the citation networks can represent citations between authors, papers, journals, countries etc. Due to several reasons it was convenient for us to consider the following citation network of CMP.

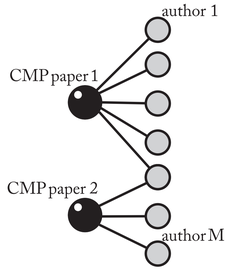


Fig.4. *Schematic figure of the bipartite citation graph.*
BLACK nodes: *citing papers;* **GRAY nodes**: *cited authors.*

N_{citing}	N_{cited} *	L	$k_{\max}$	$\langle k\rangle$
470	10644	17555	277	37.35
Average number of coauthors in cited papers $= 2$				

***NOTE:** the most cited authors, i.e. "Unknown" and "et al.", were ignored ☺.

Bibliographic coupling network

NODES: papers in CMP **LINKS**: presence of the citation of the same author

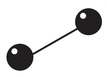


Fig.5. *One-mode projection of the bipartite citation graph of Fig.3.*

N	L	$k_{\max}$	$\langle k\rangle$	$\langle C\rangle$	$\langle C\rangle/C_r$	$\langle l\rangle$	$l_{\max}$
470	14069	211	59.87	0.54	4.27	2.16	6

Some specific features:

- the bibliographic coupling network is relatively weak correlated.
- 3 isolated nodes were detected (papers of separate or unusual for CMP directions):
  1. "Banach's space: Lviv and the Scottish Cafe", D. Henderson, 2004, **7**, No. 4: history of science;
  2. "Lagrangian vector field and Lagrangian formulation of partial differential equations", M. Chen, 2005, **8**, No. 2: mathematics;
  3. "Phase transitions in the coal-water-methane system", A.D. Alexeev, E.V. Ulyanova, N.A. Kalugina, S.E. Degtyar, 2006, **9**, No. 1: experimental physics.
- the nodes with high degree ("hubs"): interdisciplinary papers, reviews or papers which cite "authoritative" persons (hubs in the co-citation network).

Co-citation network

NODES: authors, cited in CMP **LINKS**: presence in the same list of references

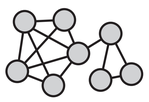


Fig.6. *One-mode projection of the bipartite citation graph of Fig.3.*

N	L	$k_{\max}$	$\langle k\rangle$	$\langle C\rangle$	$\langle C\rangle/C_r$	$\langle l\rangle$	$l_{\max}$
10644	498945	1544	93.75	0.892	101.28	2.9	7

Some possible interpretations:

- the nodes with high degree ("hubs") – influential authors, scientists with numerous pupils or authors who work in different thematic directions; top 5 of such authors:. I.V. Stasyuk, I.R. Yukhnovskii, G. Stell, L.D. Landau, D.N. Zubarev. Note: the set of hubs changes when self-citations are ignored.

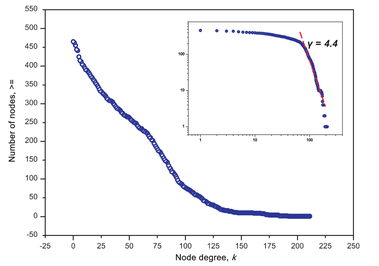


Fig.7. *Cumulative node-degree distribution of CMP bibliographic couples network. Insert: the same distribution in log-log scale.*

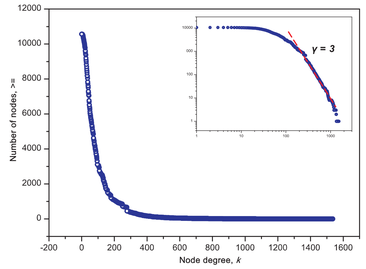
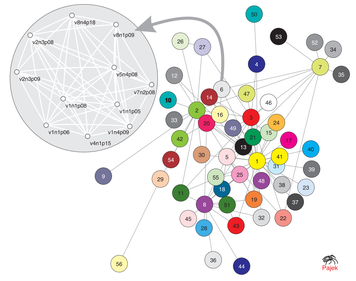


Fig.8. *Cumulative node-degree distribution of CMP co-citation network. Insert: the same distribution in log-log scale.*

Network of papers with common PACS numbers

NODES: papers in CMP



LINKS: common PACS numbers

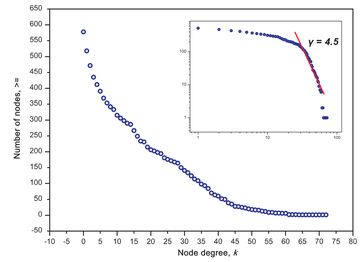


Fig.9. *Main communities of the PACS-related paper network. Each community contains from 1 to 86 associated papers. Papers of the community No. 6 are shown in detail (the most frequently used PACS numbers here: 74.72.-h, 74.20.-z, 74.20.Mn).*

Fig.10. *Cumulative node-degree distribution. Insert: the same distribution in log-log scale.*

N	L	$k_{\max}$	$\langle k\rangle$	$\langle C\rangle$	$\langle C\rangle/C_r$	$\langle l\rangle$	$l_{\max}$
470	3218	64	13.7	0.56	19.2	3.4	9

Some conclusions

Comparative analysis of data obtained from various networks considered above allows one to draw some conclusions about specific features of the journal as a constituting part of a scientific enterprise. Examples are given by:

- Community structure of the bibliographic couplings and PACS-related paper networks defines the groups of papers connected by the subject area;
- Community structure of the co-authorship network defines the groups of scientists with common research interests;
- The most influential ("classic") authors for particular journal can be identified as hubs in the co-citation network;
- The papers denoted by isolated or weakly connected nodes in the bibliographic coupling network can be recognized as exotic ones in a frame of a given journal.

References

[1] O. Mryglod and Yu. Holovatch, Condens. Matter Phys. **10**, 129 (2007).

[2] M. Girvan and M.E.J. Newman, PNAS **99**, 7821 (2002).

[3] A. Vlado. Pajek: Program for large network analysis; S.P. Borgatti, NetDraw: Graph Visualization Software. Harvard: Analytic Technologies (2002).