Journal of Informetrics 11 (2017) 1114–1127

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Hierarchical organization of H. Eugene Stanley scientific collaboration community in weighted network representation

Stanisław Drożdż[a,b,*], Andrzej Kulig[a], Jarosław Kwapień[a], Artur Niewiarowski[b], Marek Stanuszek[b]

[a] Complex Systems Theory Department, Institute of Nuclear Physics, Polish Academy of Sciences, Kraków, Poland
[b] Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Kraków, Poland

ARTICLE INFO

Article history:
Received 22 May 2017
Received in revised form
21 September 2017
Accepted 25 September 2017

Keywords:
Complex weighted networks
Scientific collaboration
Communities
Scientometrics
Erdős number generalised

ABSTRACT

By mapping the most advanced elements of the contemporary social interactions, the world scientific collaboration network develops an extremely involved and heterogeneous organization. Selected characteristics of this heterogeneity are studied here and identified by focusing on the scientific collaboration community of H. Eugene Stanley – one of the most prolific world scholars at the present time. Based on the *Web of Science* records as of March 28, 2016, several variants of networks are constructed. It is found that the Stanley # 1 network – this in analogy to the Erdős # – develops a largely consistent hierarchical organization and Stanley himself obeys rules of the same hierarchy. However, this is seen exclusively in the weighted network representation. When such a weighted network is evolving, an existing relevant model indicates that the spread of weight gets stimulation to the multiplicative bursts over the neighbouring nodes, which leads to a balanced growth of interconnections among them. While not exclusive to Stanley, such a behaviour is not a rule, however. Networks of other outstanding scholars studied here more often develop a star-like form and the central hubs constitute the outliers. This study is complemented by a spectral analysis of the normalised Laplacian matrices derived from the weighted variants of the corresponding networks and, among others, it points to the efficiency of such a procedure for identifying the component communities and relations among them in the complex weighted networks.

1. Introduction

The accelerating process of world globalization embraces and pervades all aspects of the human activity. Contemporary means and standards of conducting the scientific investigations deserve a special attention in this context as their progress at the same time constitutes both the condition and the result of this world globalization process. Indeed, the world most advanced scientific contemporary initiatives are based on multinational and often even on highly multidisciplinary collaborations. Some of them, like the ones carrying out the high energy physics experiments at CERN and at DESY in Europe, at Fermilab and at Brookhaven in the US, at KEK in Japan or the ones conducting the global astronomical sky-observations,

* Corresponding author at: Complex Systems Theory Department, Institute of Nuclear Physics, Polish Academy of Sciences, Kraków, Poland.
E-mail address: stanislaw.drozdz@ifj.edu.pl (S. Drożdż).

https://doi.org/10.1016/j.joi.2017.09.009
1751-1577/

are largely administratively arranged as far as their organization and staff involved is concerned. Typically this predetermines the co-authorship composition, usually very numerous, of the resulting, also numerous, publications. However, there recently emerge more spontaneous and at the same time more dynamical forms of the scientific cooperation. In most cases they are driven by the contemporary interdisciplinary trends in research, such that they involve a group of renown scientists (or even a single one) who, by their ability to create a scientifically stimulating environment, attract others to a productive collaboration, which proliferates further through various disciplines and diversified co-authorship compositions (Adams, 2012).

Paul Erdős, the famous Hungarian mathematician (De Castro & Grossman, 1999), who has written over 1400 papers with over 500 co-authors and who thus inspired the concept of the Erdős number, can be considered a forerunner. At present an even more spectacular cascading of scientific collaboration of this kind can be observed. In this regard H. Eugene Stanley, professor at the Boston University, whose scientific activity comprises a broad range of areas such as *Aggregation, Viscous Fingering, Statistical Physics, Phase Transitions, Critical Phenomena, Granular Materials, Surface Physics, Econophysics, Chemistry, Water, Social Networks, Physiology, Medicine, and Neuroscience*, and his constantly increasing number of collaborators create a particularly interesting phenomenon to study. H.E. Stanley's $h = 125$ index due to $N = 1208$ published articles co-authored in total by 738 scientists, as on March 28, 2016, listed by the *Web of Science (WoS)*, with all these figures constantly increasing (currently $h = 134$ and $N = 1301$) provides a formal evidence of this great success and his scientific collaboration network (SCN) deserves thus a particular attention.

Studying characteristics of various aspects of the scientific collaboration potentially constitutes a significant contribution towards understanding the structure and dynamics of the social interactions (Grossman & Ion, 1995; Jiang et al., 2013; Jin, Girvan, & Newman, 2001; Katz, 1994; Liljeros, Edling, Amaral, Stanley, & Åberg, 2001; Luukkonen, Persson, & Sivertsen, 1992) but, first of all, it is of great importance for an efficient stimulation of the future science development (Ausloos, 2013, 2014a; Bougrine, 2014; Miśkiewicz, 2013; Rotundo, 2014; Wilsdon, 2011). Quantifying properties of the scientific collaboration networks in an informative and transparent way becomes highly facilitated (Barabási et al., 2002; Lee, Goh, Kahng, & Kim, 2010; Li et al., 2007; Liu, Xu, Small, & Chi, 2011; Palla, Barabási, & Vicsek, 2007) thanks to the great advances in the field of network theory (Albert & Barabási, 2002). Most of the existing related works study the global properties of the collaboration networks (de Solla Price, 1965; Freeman, Ganguli, & Murciano-Goroff, 2014; Wagner & Leydesdorff, 2005; Wuchty, Jones, & Uzzi, 2007), including their evolutionary aspects (Newman, 2001a, 2001b, 2004; Newman, Strogatz, & Watts, 2001; Tomassini & Luthi, 2007), or occasionally point to the individual country contribution (He, 2009; Perc, 2010). Fewer works focus on characteristics of the selected scientists (Ding, 2011) in their creative role and of the range of their influence in the collaboration network. In order to make this issue and the related characteristics even more exposed, here, for several most outstanding scholar figures working in the domain of exact sciences, with a particular focus on H. Eugene Stanley, we generate their collaboration networks based exclusively on all the publications involving that particular scholar. Nodes then represent all the authors who appeared in any of the common publications and the links among them are assigned when their names appear together in the same publication. By construction, a node representing the author X defining such a network constitutes the central hub and all the other nodes in such a network have the collaboration number 1 relative to X, which by analogy to the Erdős number can be termed the X number 1 (X # 1).

2. Network construction and description

All the results presented in this work have been obtained using the data downloaded from the *Web of Science*. This website provides one of the most reliable and complete scientometrics sources. It covers many scientific disciplines belonging both to the exact sciences, to engineering as well as to the life sciences. Still, ensuring that all scientists are clearly identifiable and distinguishable, as needed in the present analysis, appears a highly non-trivial task. There are several elements that demand a special care. One particularly important is a proper distinction of different scientists. As it has already been estimated (Newman, 2001a; Newman et al., 2001; Perc, 2010), about 5% of all scientists have the same initials and surnames. What is even more troublesome is that there exist different scientists having the same name and the same surname as well. In order to overcome such an equivocation, an additional criterion of the scientific affiliation has been applied. This of course helps, but does not resolve the problem entirely due to the significant mobility of scientists. Another problem is the presence of typos in the names and surnames. Such possible errors have been taken care of by using the *Levenshtein measure* (Levenshtein, 1966) to strings of letters, here representing the names and surnames.

As in essentially all the network cases, the topology of SCN can be expressed by its adjacency matrix **A** whose elements a_{ij} assume the value 1, thus express existence of the resulting link, if the authors i and j co-author at least one publication. Otherwise a_{ij} equals 0. The corresponding ith node degree $k_i = \sum_{j=1}^{N} a_{ij}$, where N is the total number of authors (nodes) within the network. Complete description of SCN requires, however, taking into account not only its topology but also the weights of the links among the nodes (Boccaletti, Latora, Morenod, Chavez, & Hwang, 2006; Newman, 2001b). In SCN the weight of a given link is determined by the number n_{ij} of publications co-authored by the ith and jth authors. The so-weighted ith node degree, denoted as k^w, can be written as $k_i^w = \sum_{j=1}^{N} a_{ij} n_{ij}$.