Journal of Informetrics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Informetrics

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

Who is collaborating with whom? Part I. Mathematical model and methods for empirical testing

Hildrun Kretschmer[a,*], Donald deB. Beaver[b], Bulent Ozel[c], Theo Kretschmer[d]

[a] COLLNET Center, Germany
[b] Williams College, 117 Bronfman Science Center, 18 Hoxsey St., Williamstown, MA 01267, USA
[c] University of Jaume I, Department of Economy, Castellon De La Plana, Spain
[d] COLLNET Center, Borgsdorfer Str. 5, D-16540 Hohen Neuendorf, Germany

ARTICLE INFO

Article history:
Received 5 February 2013
Received in revised form
30 December 2014
Accepted 5 January 2015
Available online xxx

Keywords:
Social network analysis
Self-organization
Mathematical model
Co-authorship
3-D computer graphs
Animation

ABSTRACT

There are two versions in the literature of counting co-author pairs. Whereas the first version leads to a two-dimensional (2-D) power function distribution; the other version shows three-dimensional (3-D) graphs, totally rotatable around and their shapes are visible in space from all possible points of view. As a result, these new 3-D computer graphs, called "Social Gestalts" deliver more comprehensive information about social network structures than simple 2-D power function distributions. The mathematical model of Social Gestalts and the corresponding methods for the 3-D visualization and animation of collaboration networks are presented in Part I of this paper. Fundamental findings in psychology/sociology and physics are used as a basis for the development of this model.

The application of these new methods to male and to female networks is shown in Part II. After regression analysis the visualized Social Gestalts are rather identical with the corresponding empirical distributions ($R^2 > 0.99$). The structures of female co-authorship networks differ markedly from the structures of the male co-authorship networks. For female co-author pairs' networks, accentuation of productivity dissimilarities of the pairs is becoming visible but on the contrary, for male co-author pairs' networks, accentuation of productivity similarities of the pairs is expressed.

1. Introduction

The rise in collaboration in science and technology experienced world-wide at both national and international levels, has assumed such overriding importance that there is now a perceptibly urgent need to study such processes with a view to acquiring fundamental knowledge for organizing future research and its application to science and technology policies. New concepts have emerged in order to understand pattern formation in interactional processes of collaboration (Yin, Kretschmer, Hanneman, & Liu, 2006). Some of these concepts are self-similarity, self-organization, power laws, complex networks of interactions and others.

Two different bibliometric analysis techniques are usually used for gender and for collaboration studies:

* Corresponding author. Permanent address: Borgsdorfer Str. 5, D-16540 Hohen Neuendorf, Germany. Tel.: +49 3303 500 866; fax: +49 3303 50 4838.
E-mail addresses: kretschmer.h@onlinehome.de (H. Kretschmer), dbeaver@williams.edu (D.deB. Beaver), ozel@uji.es (B. Ozel), kretschmer.h@onlinehome.de (T. Kretschmer).

http://dx.doi.org/10.1016/j.joi.2015.01.004

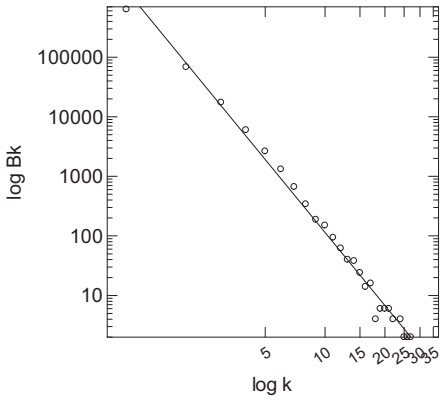

Fig. 1. Power law distributions of co-author pairs' frequencies B_k. Journal of Biochemistry, $B_k = c/k^b$, with c = constant $n = 26$; $R = 0.987$, $R^2 = 0.975$; F = Ratio = 508.65; $B_k = 5.39–4.539 \log k$.

- Descriptive analysis methods (standard bibliometric indicators, social network analysis, etc.), for example Naldi, Luzi, Valente, and Parenti (2004), Melin (2000), Carr, Pololi, Knight, and Conrad (2009), Kyvik and Teigen (1996), Pepe and Rodriguez (2009), etc.
- Parametric models or laws, for example Lotka (1926), Bradford (1934), Price (1963), Egghe (2008), Newman (2005), and Morris and Goldstein (2007).

The Social Gestalt model is a new parametric model with four parameters. First we refer to the version of counting co-author pairs leading to power laws and second to the new version of counting co-author pairs leading to 3-D graphs. In contrast to a single power function distribution (2-D graphs), the mathematical model of "Social Gestalts" (Kretschmer, 1999, 2002; Kretschmer & Kretschmer, 2007) visualizes the 3-D graphs, using animation in the form of rotating these graphs.

McGrath (2002, p. 314) has added the following remarks to the model of "Social Gestalts":

> "The social Gestalt can be defined as a holistic configuration of the parts of a relationship. Each Gestalt can be graphed as a 3-dimensional array of co-authorships. Though the interrelationships may vary, they can always be represented in a single holistic graph that, when stable exemplifies the conciseness principle."

The term "Social Gestalt" is selected in honor of both Wolfgang Metzger's deliberations in 1967 about the formation of social groups on the basis of the conciseness principle and in honor of the famous Berlin "Gestalt Psychology" at the beginning of the 20th century Metzger (1986).

1.1. First version of counting co-author pairs leading to power laws

Counting the number of publications of co-author pairs: The pairs are counted as units (*P*,*Q*) in analogy to single authors *P* in Lotka's Law, where *k* is the number of joint publications of the pair *P*, *Q* (For example: Smith & Miller). We assume there is a regularity for the distribution of coauthor pairs' frequencies B_k with *k* publications per co-author pair (*P*,*Q*) in form of a power law distribution:

Morris and Goldstein (2007) have already shown this kind of regularity in one of their empirical studies. In this connection Egghe (2008) has presented a theoretical model for the size-frequency function of co-author pairs.

We have studied the regularities for distributions if one counts the number of publications of the co-author pairs in the Journal of Biochemistry. We found a power law distribution, cf. Fig. 1.

Special collaboration structures, for example scale-free network models, self-similarity, power laws and others (Egghe, 2008; Morris & Goldstein, 2007; Newman, 2005) could be found in many larger networks. However, the investigation in large networks often rely on a wealth of bibliographic data, but very little or no other information about the individuals in the network (Pepe & Rodriguez, 2009).

1.2. Second version of counting co-author pairs leading to 3-D graphs

Because of the former missing information about the individuals in the network, the present paper is focused on social network analysis (SNA) applied to collaboration structures in co-authorship networks with special focus on the topic "Who is collaborating with whom". A developed procedure for visualizing a bivariate distribution of co-author pairs' frequencies hence producing 3-D graphs is presented. This distribution is described by a mathematical model, presented in Section 2. The detailed methods of counting the co-author pairs are shown in Section 3.

Please cite this article in press as: Kretschmer, H., et al. Who is collaborating with whom? Part I. Mathematical model and methods for empirical testing. *Journal of Informetrics* (2015), http://dx.doi.org/10.1016/j.joi.2015.01.004