PHYSA: 17228 Model 3Gsc pp. 1–13 (col. fig: NIL)

ARTICLE IN PRESS

Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Q1 # Directed networks' different link formation mechanisms causing degree distribution distinction

Q2 Stefan Kambiz Behfar [a,b,c,*], Ekaterina Turkina [d], Patrick Cohendet [d], Thierry Burger-Helmchen [c]

[a] Chair of Entrepreneurial Risk, Department of Management, Technology and Economics, ETH Zurich, Switzerland
[b] Institute for Solid State Physics, Department of Physics, ETH Zurich, Switzerland
[c] BETA — CNRS, University of Strasbourg, France
[d] Department of International Business, HEC Montreal, Canada

HIGHLIGHTS

- Outlinks feature different degree distributions than inlinks.
- Different link formation mechanisms cause the degree distribution distinctions.
- In/outdegree distribution distinction holds for different levels of system decomposition.

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form 31 March 2016
Available online xxxx

Keywords:
Directed Network
Link formation mechanism
Inlink and Outlink
Complex Network
Open-Source-Software (OSS)
System decomposition level

ABSTRACT

Within undirected networks, scientists have shown much interest in presenting power-law features within complex networks. For instance, Barabási and Albert (1999) claimed that a common property of many large networks was that vertex connectivity follows scale-free power-law distribution, and in another study Barabási et al. (2002) showed power law evolution in the social network of scientific collaboration. At the same time, Jiang et al. (2011) discussed deviation from power-law distribution; others indicated that size effect (Bagrow et al., 2008), information filtering mechanism (Mossa et al., 2002), and birth and death process (Shi et al., 2005) could account for this deviation. Within directed networks, many authors have considered that outlinks follow a similar mechanism of creation as inlinks' formation (Faloutsos et al., 1999; Krapivsky et al., 2001; Tanimoto, 2009) with link creation rate being the linear function of node degree, and a resulting power-law shape for both indegree and outdegree distribution. Some other authors have made an assumption that directed networks, such as scientific collaboration or citation, behave as undirected, resulting in a power-law degree distribution accordingly (Barabási et al., 2002). At the same time, we claim (1) Outlinks feature different degree distributions from inlinks; where different link formation mechanisms cause the distribution distinctions, (2) in/outdegree distribution distinction holds for different levels of system decomposition; therefore this distribution distinction is a property of directed networks. First, we emphasize in/outlink formation mechanisms as causal factors for distinction between indegree and outdegree distributions (where this distinction has already been noticed in Barker et al. (2010) and Baxter et al. (2006)) within a sample network of OSS projects as well as Java software corpus as a network. Second, we analyze whether this distribution distinction holds for different levels of system decomposition: open-source-software (OSS) project–project dependency within a

* Corresponding author at: Chair of Entrepreneurial Risk, Department of Management, Technology and Economics, ETH Zurich, Switzerland.
E-mail addresses: behfarka@phys.ethz.ch, kambiz.behfar@gmail.com (S.K. Behfar).

http://dx.doi.org/10.1016/j.physa.2016.06.035

Please cite this article in press as: S.K. Behfar, et al., Directed networks' different link formation mechanisms causing degree distribution distinction, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.06.035

cluster, package–package dependency within a project and class–class dependency within a package. We conclude that indegree and outdegree dependencies do not lead to similar type of degree distributions, implying that indegree dependencies follow overall power-law distribution (or power-law with flat-top or exponential cut-off in some cases), while outdegree dependencies do not follow heavy-tailed distribution.

1. Introduction

Q3

Among network models Erdős–Rényi (ER) [1] proposed a non-growing randomly connected model, Watts and Strogatz (WS) [2] proposed a non-growing randomly re-connected network model (so called small world) and Barabási–Albert (BA) [3] proposed a growing network with the probability of addition of new nodes proportional to the number of incoming links (so-called preferential attachment model or rich-get-richer). In ER and WS models, number of nodes in the network is fixed, and linkages among existing link formation nodes are formed, while BA model assumes time-homogeneous network growth with a mechanism for preferential attachment link formation. There are also other growth models such as fitness model (Bianconi et al. [4]) attractiveness model (Dorogovtesev et al. [5]), accelerating growth model (Dorogovtesev et al. [6]), logarithmic growth model (Shi et al. [7]), and random preferential attachment model (Liu et al. [8]).

Preferential attachment does not always explain network evolution, e.g. where the innovation of an article rather than the number of its citations causes a new attachment. Scientists such as Ergun et al. [9] and Xu et al. [10] have discussed a methodology of fit-get-richer, implying that new vertices connect to highly fitted vertices. This explains attachment to a new network based on its intrinsic physical property or quality. In this area, Caldarelli et al. [11] introduced a varying vertex fitness model. As far as link formation mechanisms are concerned, Newman [12] defined assortativity mixing for *undirected* networks as a node tendency to connect to other nodes with similar degree. Piraveenan [13,14] defined this for *directed* networks as: in (out)-assortativity is the tendency whereby nodes tend to connect to other nodes with similar in (out)-degrees. Jackson and Rogers [15] have also presented a dynamic model of link formation based on random as well as searching through the current structure.

Scientists have shown much interest in presenting power-law features within complex networks. For instance, Barabási and Albert [3] claimed a common property of many large networks was that vertex connectivity follows scale-free power-law distribution, and concluded that development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems. Barabási et al. [16] have also shown power law evolution in the social network of scientific collaboration. Furthermore, Faloutsos et al. [17] showed power-law features existing in the internet topology, implying its benefits in designing efficient protocols, creating accurate artificial models and speculating on the internet topology in the future.

Authors such as Jiang et al. [18] discussed a network model of *deviation* from power-law distribution. Some other authors had previously addressed this deviation and indicated that size effect (Bagrow et al. [19]), information filtering mechanism (Mossa et al. [20]), and birth and death process (Shi et al. [21]) accounted for this deviation. Maillart et al. [22] tested Zipf's degree distribution via link creation and deletion mechanism in open source Linux distribution. In another work, Maillart et al. [23] used data collected by Google to identify the existence of power-law regimes for a population of Internet users to execute a given task after receiving a message.

We argue that WWW, Scientific Collaboration and OSS reuse networks are not undirected, as assumed in some studies; they are in fact directed networks where outlinks and inlinks demonstrate different degree distributions. Faloutsos et al. [17], Krapivsky et al. [24], Tanimoto [25] and more also assumed that preferential attachment is the dominant link formation mechanism in directed networks, resulting in power-law degree distribution. At the same time, we claim that there are different (out) inlink formation mechanisms within directed networks which result in degree distribution distinctions. We propose two hypotheses to explain causal effect of link formation mechanism on degree distribution.

We prove the hypotheses both analytically and empirically. In the analytical approach, apart from using indegree-based preferential attachment mechanism to prove our claims, we apply other link formation mechanisms such as outdegree-based preferential attachment, fitness-based preferential attachment. In the empirical section, we first consider the sample network of open-source-software (OSS) projects reuse to identify the distinction between indegree and outdegree distribution, then analyze whether this distinction holds in the corpus of each of those OSS projects, and at different system decomposition levels of package–package and class–class dependencies.

2. Theoretical development and hypotheses

2.1. Inlink and outlink formation logic

As already mentioned, inlink and outlink do not lead to similar types of degree distribution. Here we give few examples to demonstrate the logic behind this distinction.