Physica A 474 (2017) 49–60

Contents lists available at ScienceDirect

Physica A

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

Fitness networks for real world systems via modified preferential attachment

Ke-ke Shang [a,b,*], Michael Small [b,c], Wei-sheng Yan [a]

[a] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, PR China

[b] School of Mathematics and Statistics, The University of Western Australia, Crawley, Western Australia, 6009, Australia

[c] Mineral Resources, CSIRO, Kensington, Western Australia, 6151, Australia

HIGHLIGHTS

- We first use a link prediction method as a metric for network fitness.
- Our methods can construct the fitness networks for real world under our metrics.
- Link sparsity and the size of a network are key factors for network reconstruction.
- The network that is limited by geographic location differs from other networks.
- We observe that human behaviours can reflect real friendships.

ARTICLE INFO

Article history:
Received 30 September 2016
Received in revised form 11 January 2017
Available online 18 January 2017

Keywords:
Complex network
BA model
Assortative
Degree distribution
Rich-club
Network sparsity

ABSTRACT

Complex networks are virtually ubiquitous, and the Barabási and Albert model (*BA* model) has became an acknowledged standard for the modelling of these systems. The so-called *BA* model is a kind of preferential attachment growth model based on the intuitive premise that popularity is attractive. However, preferential attachment alone is insufficient to describe the diversity of complex networks observed in the real world. In this paper we first use the accuracy of a link prediction method, as a metric for network fitness. The link prediction method predicts the occurrence of links consistent with preferential attachment, the performance of this link prediction scheme is then a natural measure of the "preferential-attachment-likeness" of a given network. We then propose several modification methods and modified *BA* models to construct networks which more accurately describe the fitness properties of real networks. We find that all features assortativity, degree distribution and rich-club formation can play significant roles for the network construction and eventual structure. Moreover, link sparsity and the size of a network are key factors for network reconstruction. In addition, we find that the structure of the network which is limited by geographic location (nodes are embedded in a Euclidean space and connectivity is correlated with distances) differs from other typical networks. In social networks, we observe that the high school contact network has similar structure as the friends network and so we speculate that the contact behaviours can reflect real friendships.

* Corresponding author at: School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, PR China.
E-mail address: keke.shang@uwa.edu.au (K.-k. Shang).

http://dx.doi.org/10.1016/j.physa.2017.01.066
0378-4371/

1. Introduction

The Barabási and Albert preferential attachment growth model (the *BA* model) [1] is the *de facto* standard model in the field of complex networks. The model is both intuitive and highly appealing. Popularity, as characterized by node degree, is a sufficient mechanism to lead to the emergence of scale-free degree distribution. As a consequence of the success of the BA model, it is often treated as the generative process for a wide range of real scale-free network systems. However, the *BA* model follows the simple principle that popularity is attractive [2] whereas the real world is often more complicated [2–4]. Similarity (or affinity) between nodes [2], minimum node degree [3] and the dominance of low degree nodes [4] all play significant supporting roles while constructing fitness networks.[1]

Emerging as the solution to a quite separate problem, link prediction can help us to analyse and predict network structure in real-world networks with characteristic structure [5–7]. By taking the *BA* model as the underlying generative pattern, we aim to adapt a newly described link prediction method (*PAI*) [6,8] to predict links based on the degree to which preferential attachment is followed and hence measure the level of preferential attachment in real networks. Hence, we propose using the accuracy of the *PAI* algorithm to measure the fitness between these network construction models and real networks. Obviously, the BA model cannot model the degree distribution for many real networks. Conversely, the configuration model [9,10] can achieve the same degree distribution as real networks, but it is difficult to maintain the connectivity of the networks and the algorithm readily generates self-loops when mimicking real networks. Hence, we also use the degree distribution and the connectivity as metrics of networks fitness.

In our previous studies [3,4], we observed that the assortativity of a *BA* network differs from that of *typical* real networks. Furthermore, previous studies already proposed the assortative preferential attachment to construct the scale-free network with a given assortativity value [11] or can control their model parameter to change the network assortativity [12]. A specific modified preferential attachment also has been proposed to describe the local assortativity distribution for the Internet Autonomous System level networks [13]. In this paper, based on Xulvi-Brunet and Sokolov's algorithm [14], we use corresponding randomized methods to modify the assortativity of an unbiased *BA* network to construct fitness networks—networks which better match real observed data.

However, we find that these assortativity modification methods are not successful in all cases. Moreover, since the actual degree distribution of *BA* models will usually differ from that of real networks, we further propose two modified methods to modify the degree distribution and the product of nodes degrees of the *BA* networks. Finally we note that in our previous study we found that the rich-club plays a significant role in driving the other properties of networks [15]. On the other side, our modified methods cannot work well for the large size and the sparse network. Hence, to construct the fitness network simpler and more efficient, we propose one additional modified *BA* model to construct networks with more rich-clubs and one additional modified *BA* model to construct networks with more randomly links. In addition, compare these two additional models, we can study the role of the rich-club. We then propose a further two modified models which more effectively construct fitness networks.

In this paper, we use 14 different real datasets to construct a range of archetypal "real" networks. We use the *BA* model to construct corresponding networks. Meanwhile, we compute the connectivity, the accuracy of *PAI* and the degree distribution as metrics of network fitness. Our various modified methods (which will be described below) and network generation models all maintain the connectivity of the original real networks. However, compared to real networks, we find that there is only one *BA* network which matches according to the second metric. Hence, we use our methods to modify the assortativity of the *BA* networks, after which, we find that there are now six modified *BA* networks coincide according to the second metric. Finally, we use our new methods to modify the degree distribution and the product of nodes degrees of *BA* networks, following this we find that there 12 networks are able to match the model with the second metric. Based on the rich-club phenomena and random networks, we propose two network models *RCBA* and *RBA* respectively. Moreover, we propose the corresponding modified models of *RCBA* (*MRCBA*) and *RBA* (*MRBA*). Finally, we can combine our modified methods and models to construct all fitness networks that fit the second metric. Our models also have degree distributions more similar to the original real networks. In addition, we find that the human behaviours can reflect underlying real ("true") friendships. The network which is limited by the geographic location has the special structure, and the rich-club, the sparsity and the size of network all play the key roles for the network reconstruction.

2. Data

Different kinds of networks have different structures for the reconstruction, to consider the diversity of networks, we employ large open and publically available data sets from 14 different undirected networks: from the social network to the engineering network, from the sparse network to the dense network, from the small size network to the large size network. (1) *Football* network [16]: a node represents one country, a link stands for a pair of nodes at least sharing one common football player which participated in the World Championship in Paris, 1998. (2) *Dolphin* network [17]: a node represents one dolphin which in a community living off Doubtful Sound, a link stands for there is associations between a pair of nodes. (3) *PolBooks* network [16]: a node represents one book about US politics which is sold by Amazon.com, a link stands for

[1] In our paper, the fitness network means that a network has the similar structure, degree distribution and connectivity with the original real network.