Physica A 507 (2018) 99–109

Contents lists available at ScienceDirect

Physica A

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

Analysis on large-scale rating systems based on the signed network

Ke Gu [a,b], Ying Fan [a,*], An Zeng [a,*], Jianlin Zhou [a], Zengru Di [a]

[a] School of Systems Science, Beijing Normal University, 100875 Beijing, China
[b] Department of Information Management, Economics and Management College, Beijing Institute of Petrochemical Technology, 102617 Beijing, China

HIGHLIGHTS

- Construct unweighted signed bipartite networks for user–object rating systems.
- Reveal the basic properties of real signed bipartite networks.
- Analyze the basic motif of signed bipartite networks: quadrangle.
- Introduce a novel projection method Signed Common Neighbors to get the projection to signed user-network.

ARTICLE INFO

Article history:
Received 16 January 2018
Received in revised form 17 April 2018

Keywords:
Online rating systems
Negative edges
Signed bipartite networks
Signed common neighbors
Projection to signed user-network

ABSTRACT

In many user–object online rating systems, it is obviously that the ratings usually show the users' attitudes: like or dislike the objects. Inevitably there is a need to introduce the sign into the rating systems. We first focus on how to construct signed bipartite networks on rating systems and reveal the basic properties of them. We also analyze the basic motif of signed bipartite networks: quadrangle. Then we introduce a novel projection method Signed Common Neighbors (SCN) to get the projection to signed user-network. The basic statistics of the projections show that SCN can well reflect the roles of negative edges.

1. Introduction

In the big data era, the rating of users on online products are becoming increasingly accessible. The large-scale rating systems have rich information. Our study focuses on such social systems. Many online rating systems actually have two-mode nature. The online rating system including the information of users, objects and the ratings can be represented by a weighted bipartite network, where the nodes are respectively users and objects, the edges represent the users rate the objects and the rating is the weight of edge. One way to discuss bipartite network is to analysis the origin bipartite networks which can keep the information to the maximum extent and there are some metrics on bipartite network [1]. We have also done some research on the bipartite network [2,3]. For example, we propose the notion of bipartite network clustering coefficient [3]. The other way to study bipartite network is to transform two-mode networks into one-mode projection networks.

In some online rating systems the user can directly give the object positive or negative rating, and the negative rating means that the user dislikes the object. But in many online rating systems, negative rating is not necessary for 'dislike'

* Corresponding authors.
E-mail addresses: yfan@bnu.edu.cn (Y. Fan), anzeng@bnu.edu.cn (A. Zeng).

https://doi.org/10.1016/j.physa.2018.05.048
0378-4371/

Table 1
Original information of the networks.

Network	*Object*	*m*	*n*	*E*	*Rating range*
Netflix	Movie	3 000	2779	197 248	1,2,3,4,5
Joke	Joke	50 691	140	1 725 047	−10–10

Note: The *Object* denotes the object user rates; *m*, *n* and *E* denote the number of users, objects and ratings; *Rating range* denotes the range of the ratings.

comment, the user cannot give the object negative rating. Most of the previous studies on online rating systems actually only focused on the ratings. In fact, the rating usually shows the user's attitudes: likes or dislikes the object. According to experience, in the online rating system that the user can only give positive rating, high rating means the user likes the object and low rating means the user dislikes the object. Based on this, we can introduce signed bipartite networks into online rating systems. The positive/negative edge of signed bipartite network means the user likes/dislikes the object. The signed bipartite network is a very important form of network structure. The study of signed networks was originally focused in the field of sociology, and in the 1940s, Heider explored the structural balance theory which is the basic theory of signed networks [4]. Cartwright and Harary introduced the concept of balanced signed graph which is based on the structural balance theory [5]. Most of the original research of signed network is applied in the so-called one-mode networks, especially in the signed social networks. Negative edges play an important role. Using the sign properties of edges to analysis real networks has practical applications [6–9].

The existence of negative edges improves the accuracy of the study but also challenges many concepts and methods of unsigned networks. Our research focuses on how to correctly locate the role and significance of the negative edge and reasonably handle the relationship between the positive and negative edges. Plenty of researches on signed networks are mainly about one-mode networks, and seldom focus on signed two-mode networks. Online rating systems actually have two-mode and signed nature. On the basis of combining the information of negative edge and bipartite structure, we construct the signed bipartite online rating network model, and make further analysis.

In this paper, we mainly concern about two online rating systems: Netflix and Joke. We also research the other four rating systems, but the results are similar to Netflix, so we only present the results of Netflix and Joke. We construct unweighted signed bipartite networks based on ratings for these user–object systems. Then we discuss basic properties presented by the real-world signed bipartite networks and compare the value of metrics with random networks. Unlike that triangle is the basic motif of one-mode graph, the basic motif of signed bipartite graph is quadrangle. We identify seven signed quadrangles based on perspective of users and discuss the motifs by comparing the statistical result of seven quadrangles on real-world networks with the results on two comparison models. In fact, the quadrangle is composed of different signed triples. And we find that the number of triples representing two users simultaneously hating the same object is the smallest. Based on this, we propose a new projection model and study the properties of the projection network. The innovative projection method Signed Common Neighbors (SCN) based on the similarity of users is proposed to generate signed user-network. The parameters of SCN are related to the ratio of different signed triples. There are some bipartite network projection methods and some simple signed bipartite network projection methods [10]. The novelty of SCN is that it is based on the statistical nature of the signed triples. The basic statistics of the projection signed user-networks show that SCN can well reflect the integrated use of positive and negative edges information and also reflect the different roles of the positive and negative edges.

2. Database description

To research the rating systems from the point of signed bipartite networks, we select two user–object rating data sets to analyze : Netflix[1] and Joke.[2] Table 1 shows the original information of the networks. We also research the other four rating systems: Movielens,[3] Douban,[4] RYM[5] and BookCrissing,[6] which the users rated movies, music, rock albums and books. The results of these four networks are similar to Netflix, so we only present the results of Netflix and Joke.

In Netflix dataset each user at least rates 20 movies. The users of Joke rate 150 jokes, but 10 jokes do not have ratings, so we remove these 10 jokes information. Joke is special, the users can give positive and negative points to the jokes directly, and 0 ratings given by users are deleted from the data set.

1 http://www.netflixprize.com/.
2 http://www.ieor.berkeley.edu/~goldberg/jester-data.
3 http://www.grouplens.org/.
4 http://music.douban.com/.
5 http://rateyourmusic.com/.
6 http://www.ieor.berkeley.edu/~goldberg/jester-data.