User Ratings Analysis in Social Networks Through a Hypernetwork Method

Article *in* Expert Systems with Applications · April 2015

DOI: 10.1016/j.eswa.2015.05.054

Expert Systems with Applications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

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

User ratings analysis in social networks through a hypernetwork method

Qi Suo [a,1], Shiwei Sun [b,*], Nick Hajli [c,2], Peter E.D. Love [d]

[a] School of Economics and Management, Qingdao University of Science and Technology, Qingdao 266061, China
[b] Raymond J. Harbert College of Business, Auburn University, Auburn, AL 36849, USA
[c] Newcastle University Business School, 5 Barrack Road, Newcastle upon Tyne NE1 4SE, UK
[d] School of Civil and Mechanical Engineering, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Social network
Hypernetwork
Topological characteristics
Collaborative recommendation

ABSTRACT

This study utilizes the critical properties of a complex social network to reveal its intrinsic characteristics and the laws governing the way information propagates across the network to identify the central, active users and opinion leaders. The hypernetwork method is applied to analyze user ratings in social networks (SNSs). After introducing the concept of a hypernetwork and its topological characteristics such as node degree, the strength of the node and node hyperdegree, collaborative recommendations in hypernetworks are formulated based on the topological characteristics. Finally, the new method developed is applied to analyze data from the *Douban* social network. In this hypernetwork, users are defined as hyperedges and the objects as nodes. Three hypernetworks focused on reviews of books, movies and music were constructed using the proposed method and found to share a similar law of trends. These topological characteristics are clearly an effective way to reflect the relationship between users and objects. This research will enable SNSs providers to offer better object resource management and a personalized service for users, as well as contributing to empirical analyses of other similar SNSs.

1. Introduction

With the advent of Web 2.0, social networks have become one of the most important channels for spreading information (Liang et al., 2014). A large number of social networking websites (SNSs) have emerged (e.g., *Facebook*, *Twitter* and *MovieLens*) and the numerous Internet content providers and discussion forums are now challenging the position of traditional media. This has developed a platform to facilitate interaction of individuals. Social networks have contributed to creating a more open public sphere by expanding the social distribution of comments and other information. The increasing importance of social networks has begun to attract a great deal of attention from scholars (e.g., Deng, Huang, & Xu, 2014; Rodder, Brenner, & Kulmann, 2014). With the help of social networks, users can not only share their experience, but also explore other users' collections to find interesting content (Hajli & Lin, 2014; Li et al., 2014; Liang et al., 2014). The information posted, such as ratings and comments can reflect users' behaviors and preferences.

The information created based on Web 2.0 social platforms and crowd-sourcing systems (Doan, Ramakrishnan, & Halevy, 2011) is commonly referred to as user-generated content (UGC). Users can easily make new acquaintances, collaborate with each other and form online communities with others that share similar interests. However, it is often difficult for a user to make informed choices given the huge numbers of books, movies, and web pages that are now available to them. As a consequence, helping people to efficiently extract the information that they truly need is a major challenge (Resnick & Varian, 1997).

Physicists have combined complex network theory (Barabási & Albert, 1999; Watts & Strogatz, 1998) and collaborative filtering (CF) (Herlocker, Konstan, Terveen, & Riedl, 2004; Schafer, Frankowski, Herlocker, & Sen, 2007) in an attempt to create better recommendation engines in social networks and the principles of physical dynamics, including mass diffusion (Zhou, Ren, Medo, & Zhang, 2007), and heat conduction (Zhang, Blattner, & Yu, 2007), have been applied in a CF algorithm. A popular approach has been to construct recommendation data via a user-object bipartite network where the nodes are divided into two sets, but only connections between two nodes in different sets are allowed (Araújo, Moreira, Furtado, Pequeno, & Andrade, 2014; Estrada & Rodriguez-Velazquez, 2005; Li & Chen, 2013; Lind, González, & Herrmann, 2005; Ramasco, Dorogovtsev, & Pastor-Satorras, 2004).

* Corresponding author. Tel.: +1 334 444 4000.
E-mail addresses: qustssqq@163.com (Q. Suo), szs0100@auburn.edu (S. Sun), Nick.Hajli@Newcastle.ac.uk (N. Hajli), p.love@curtin.edu.au (P.E.D. Love).
[1] Tel.: +86 13573829192.
[2] Tel.: +44 7951537481.

http://dx.doi.org/10.1016/j.eswa.2015.05.054
0957-4174/© 2015 Published by Elsevier Ltd.

Although all the efforts mentioned above have contributed to improving our understanding of social networks, in many cases, due to the highly diverse edge types and the complexity of the network structure, complex networks based on ordinary graphs can no longer provide a complete description of these real-world systems. Building on the foundation of hypergraph theory, hypernetworks provide a good approach for investigating the topological characteristics of social networks (Berge, 1973; Berge, 1989) and new applications of hypernetworks have begun to be proposed for topological properties and evolving models (e.g., Liu, Yang, & Hu, 2014; Liu, Li, Tang, Ma, & Tian, 2014; Yang & Liu, 2014). However, other than the properties discussed above, most of the important properties of hypernetworks have not yet been defined. Although definitions of topological properties such as node degree, node hyperdegree, and hyperedge degree have been given in previous studies, they have not taken the connection strength into consideration and thus lack critical information when attempting to depict real networks. It is therefore imperative to extend the basket of fundamental topology indicators to include factors such as the strength of the node, the strength of a hyperedge and hyperedge hyperdegree.

This study addresses this deficiency by extending the basic concepts and topological characteristics of complex networks to hypernetworks, opening up new possibilities for the topological analysis of complex systems represented by hypergraphs. These concepts and the associated calculation method are then applied to an online social network, *Douban*, to analyze empirical data as a case study. The findings reveal that the characteristics and laws governing the comments posted on SNSs can be identified and the central, active users and opinion leaders singled out. Collaborative filtering recommendation strategies can then be applied to make personalized recommendations.

The paper is organized as follows: Section 2 introduces the concept and topological characteristics of a hypernetwork; Section 3 applies a collaborative filtering algorithm into hypernetworks; Section 4 presents an application of this method using *Douban* datasets; and Section 5 discusses the results and suggests directions for future research.

2. Hypernetworks and their topological characteristics

2.1. The concept of a hypernetwork

Hypergraph theory is based on the concept that with hyperedge contains an arbitrary number of nodes rather than the two used in ordinary graphs (Berge, 1973; Berge, 1989). A hypernetwork can be described in terms of hypergraphs (Estrada & Rodríguez-Velázquez, 2006) and can effectively be used to represent the relative influences and interactions of a variety of nodes. For example, a chemical reaction can be viewed as a hyperedge where the nodes are chemicals. Similarly, in an ecological hypernetwork nodes represent species and hyperedges represent groups of species that compete for common prey. This type of competitive hypernetwork also reflects the state of the competition between species. Therefore, hypernetworks provide a powerful tool for accurately depicting real-life networks.

A number of scholars have discussed various aspects of hypernetworks. Represented by hypergraphs, the characteristics of concepts such as subgraph centrality and clustering for complex networks have been studied in three hypernetworks (e.g., Estrada & Rodríguez-Velázquez, 2006) and others have examined the theory of random hypergraphs and their applications (e.g., Ghoshal, Zlatić, Caldarelli, & Newman, 2009). The tripartite hypergraph model was extended by defining additional quantities and empirically measuring these quantities for two real-world folksonomies (Zlatić, Ghoshal, & Caldarelli, 2009), while a supernetwork model of internet public opinion has been used to examine the functions of indexes such as node superdegree, superedge-superedge distance, and superedge overlap (Ma & Liu, 2014). A framework for clustering and community detection in some systems using hypergraph representations has also been proposed (Michoel & Nachtergaele, 2012), as well as an algorithm based on a quality function for measuring the goodness of different partitions of a tripartite hypergraph into communities (Liu & Murata, 2011) and the chaotic synchronization of hypergraphs (Krawiecki, 2014). Another new concept related to hypernetworks, a hyperstructure, has been proposed and its efficiency defined (Criado, Romance, & Vela-Pérez, 2010).

Some studies have proposed evolving models to describe hypernetworks. For example, Zhang and Liu (2010) used an evolutionary hypergraph model to identify emerging statistical properties, after which they compared the model with a real-world data set; Wang, Rong, Deng, and Zhang (2010) also proposed an evolving model for uniform hypernetworks based on their growth and preferential attachment mechanisms. Two knowledge generation dynamic evolving models for scientific collaboration hypernetworks have been developed (Liu, Yang, et al., 2014; Liu, Li, et al., 2014) and it has been suggested that a local-world evolving hypernetwork model share scale-free properties (Yang & Liu, 2014).

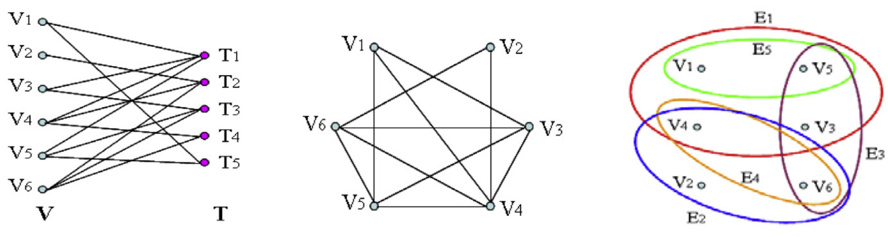
The mathematical definition of a hypergraph is given as follows. Let $V = \{v_1, v_2, \ldots, v_n\}$ be a finite set, and let $E_i = \{v_{i_1}, v_{i_2}, \cdots, v_{i_k}\}(v_{i_j} \in V, j = 1, 2, \ldots, k)$, $E^h = \{E_1, E_2, \ldots, E_m\}$ be a family of subsets of V. The pair $H = (V, E^h)$ is known as a hypergraph. The elements in V are the nodes, and $E_i\ (1, 2, \cdots, m)$ represents a set of non-empty subsets of V called a hyperedge. In a hypergraph, two nodes are considered to be adjacent if there is a hyperedge that contains both of these nodes. Two hyperedges are deemed to be adjacent if their intersection is not empty. If $|E_i| = \mathrm{u}(i = 1, 2, \ldots, m)$, $H = (V, E^h)$ is a u-uniform hypergraph, $|E_i| = 2, i = 1, 2, \ldots, m)$, $H = (V, E^h)$ degrades to a graph. Based on these definitions, a hypernetwork is a generalization of the hypergraph concept. Complex networks can be regarded as a special case of hypernetworks where each hyperedge contains only two nodes.

Consider the case of an online movie review system. This can be described by different kinds of graphs, including a bipartite, a projection and a hypergraph. Let V and T denote the users and the movies, respectively. As illustrated in Fig. 1(a), only connections between two nodes in different sets is allowed. Each node in the user-set is connected with the movies that user has reviewed in the objects-set. A bipartite graph is usually compressed by a projection graph, where two user nodes are connected when they have reviewed at least one common movie node, as shown in Fig. 1(b). Although one-mode projection graphs are always less informative than bipartite graphs, they are often convenient in that they directly show the relationships among a particular set of nodes. In Fig. 1(c), the movie can be described as a hyperedge. By analyzing the main topological characteristics in the hypergraph, the relationships between users and movies, among users, and among movies can be obtained. The hypergraph is capable of conveying a considerable more amount of information than either of the other two types of graphs; describing a hypernetwork using a hypergraph is a powerful approach for accurately depicting the activities in real-life networks.

2.2. Topological characteristics of hypernetworks

The seven main topological characteristics of a hypernetwork are discussed below. The characteristics are defined and calculations performed based on the following two criteria: (1) the property that each hyperedge in the hypernetwork may contain a

Please cite this article in press as: Suo, Q., et al. User ratings analysis in social networks through a hypernetwork method. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.05.054



(a) bipartite graph (b) projection graph (c) hypergraph

Fig. 1. Examples of a bipartite graph, a projection graph and a hypergraph.

number of arbitrary nodes; and (2) when each hyperedge contains only two nodes, the hypernetwork could be degraded to a complex network.

2.2.1. Node degree

Nodes are connected by hyperedges. The node degree of node i, D_i, represents the number of nodes connected to that node. In the example shown in Fig 1(c), $D_{v_1} = 3, D_{v_3} = 4$.

2.2.2. The Strength of the node

Node degree is based on the number adjacent to the node i, without considering the weight of the link among adjacent nodes. Here, W_{ij} denotes the number of hyperedges that encircle both node i and node j. The strength of node i, S_i, is defined as follows

$$S_i = \sum_{j \in N_i} W_{ij} \tag{1}$$

The neighborhood of node i, henceforth represented as N_i, corresponds to the set of nodes adjacent to i. In Fig. 1(c), since both v_3 and v_5 are encircled by hyperedges E_1 and E_3 and $W_{v_3 v_5} = 2$, $W_{v_3 v_1} = 1, W_{v_3 v_4} = 1, W_{v_3 v_6} = 1$, thus $S_{v_3} = 5$.

2.2.3. Node hyperdegree

The node hyperdegree of node i, D_{H_i}, refers to the number of connected hyperedges of that node (Wang et al., 2010; Zlatić et al., 2009). In Fig. 1(c), $D_{H_{v_1}} = 2, D_{H_{v_3}} = 2$, and $D_{H_{v_6}} = 3$.

2.2.4. Hyperedge degree

In hypernetworks, if two hyperedges contain the same node, this indicates that these two hyperedges are connected by a common node. Hyperedge degree, D_{E_i}, is defined as the number of other hyperedges with which a certain hyperedge is linked through its nodes (Wang et al., 2010). In Fig 1(c), since hyperedges E_1, E_3, E_5 all contain v_5, these three hyperedges E_1, E_3, E_5 are adjacent. Hyperedges E_1, E_2, E_4 all contain v_4, indicating that these three hyperedges E_1, E_2, E_4 are also adjacent. Thus, $D_{E_1} = 4, D_{E_5} = 2$.

2.2.5. The strength of a hyperedge

Hyperedge degree is based on the number adjacent to E_i, without considering the weight of the link among adjacent hyperedges. Here, $W_{E_iE_j}$ denotes the number of nodes contained in both hyperedge E_i and hyperedge E_j. We define S_{E_i} as the strength of hyperedge E_i as follows

$$S_{E_i} = \sum_{j \in N_{E_i}} W_{E_iE_j} \tag{2}$$

The neighborhood of hyperedge E_j, henceforth represented as N_{E_i}, corresponds to the set of hyperedges adjacent to E_i. In Fig. 1(c), since both E_1 and E_2 contain the node $v_4, W_{E_1E_2} = 1$. Similarly, $W_{E_1E_3} = 2, W_{E_1E_4} = 1, W_{E_1E_5} = 2$, and so $S_{E_1} = 6$.

2.3. Hyperedge Hyperdegree

The hyperedge hyperdegree of hyperedge $E_i, D_{H_{E_i}}$, refers to the number of nodes encircled by that hyperedge. In Fig. 1(c), for example, $D_{H_{E_1}} = 4, D_{H_{E_3}} = 3$.

2.4. Hyperedge–Hyperedge distance

The hyperedge–hyperedge distance $M_{E_iE_j}$ is defined as the length of the shortest path between two hyperedges E_i and E_j that are reachable via common nodes. In order to calculate the shortest distance between hyperedges, all hyperedges are converted to "hypernodes" (Ma & Liu, 2014). If two hyperedges contain at least one common node, an edge exists between these two "hypernodes", and the weight of the edge is determined by the number of common nodes. Here we only consider the number of common nodes between these two hyperedges; thus greater the value, the closer the distance between these two hyperedges. If two hyperedges are adjacent, then the distance between E_i and E_j, hence, $M_{E_iE_j}$ is defined as

$$M_{E_iE_j} = 1/W_{E_iE_j}, \quad j \in N_{E_i} \tag{3}$$

Based on the hypernode network of which they are both members, the distance for hyperedges that are not adjacent can also be obtained. The Floyd algorithm which is a basic algorithm to calculate the shortest path between any two points in a weighted graph can be used to calculate the shortest distance between two arbitrary hyperedges. Fig. 2 shows the result of transforming the hypergraph in Fig. 1(c) to a hypernode network; here, $M_{E_2E_5} = 1.5$.

3. Collaborative filters and personal recommendations in hypernetworks

This section describes how information on users' ratings can be transformed into a hypernetwork in which nodes are objects and

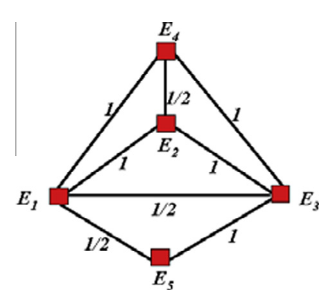


Fig. 2. The hypernode network.

hyperedges are users in such a way that all the objects evaluated by a single user are linked by the same hyperedge. The problem here is to identify the objects that are of interest to a particular user in a specific hypernetwork. Until now, collaborative filtering has been the method most widely used to perform this function, where the basic approach is to select those objects favored by other users who are similar to the target user.

3.1. Similarity of hyperedges

The similarity of hyperedges represents the similarity between two users based on the objects they have both evaluated. The Pearson correlation coefficient is defined as

$$pearson(E_i, E_j) = \frac{\sum_{t \in CR_{E_i} \cap CR_{E_j}} (r_{E_i t} - \bar{r}_{E_i})(r_{E_j t} - \bar{r}_{E_j})}{\sqrt{\sum_{t \in CR_{E_i} \cap CR_{E_j}} (r_{E_i t} - \bar{r}_{E_i})^2} \sqrt{\sum_{t \in CR_{E_i} \cap CR_{E_j}} (r_{E_j t} - \bar{r}_{E_j})^2}} \tag{4}$$

where CR_{E_i} denotes the node set contained in E_i, namely the set of objects that user i has evaluated; $r_{E_i t}$ denotes the rating awarded by user i for object t; and $\bar{r}_{E_i}$ denotes the average rating of user i for objects evaluated by both i and j. A known disadvantage of the Pearson correlation coefficient is that the similarity scores between two users may be very large when they have only a small number of common objects, which is referred to as the data sparsity problem.

Another way to deal with this problem is to utilize the Jaccard similarity coefficient (Jaccard, 1908), which is defined as

$$jaccard(E_i, E_j) = \frac{|CR_{E_i} \cap CR_{E_j}|}{|CR_{E_i} \cup CR_{E_j}|} \tag{5}$$

In general, the greater the number of common objects evaluated by two users, the more similarity there is likely to be between them. However, the Jaccard similarity coefficient only takes the number of common objects into account, without considering the difference between the ratings within the calculation. Note that the similarity is not only related to the number of common objects, but also to how the objects were rated. From this point of view, a combination of the Pearson and the Jaccard similarities is a better alternative (Xie, Ma, & Yang, 2013). Indeed, experiments have shown that the product of the Pearson and the Jaccard similarities provides the best similarity function for the MovieLens dataset (Candillier, Meyer, & Fessant, 2008), where, the hyperedge similarity function is obtained by the following expression

$$SuperSim(E_i, E_j) = pearson(E_i, E_j) \times jaccard(E_i, E_j) \tag{6}$$

3.2. Recommendation

For the target user i, calculating the hyperedge similarity enables us to filter the set of users whose evaluations are similar to the evaluations assigned by the target user. The first N users are within the neighborhood of user i. Looking at the ratings by neighborhood, we can calculate the predicted ratings for user i 's uncollected objects. Sorting the uncollected objects in descending order by the predicted ratings, the first M objects will be recommended. The predicted ratings ranking list is defined as

$$r_{E_i l} = \frac{\sum_{k \in N_{E_i}} Super\ Sim(E_i, E_k) r_{E_k l}}{\sum_{k \in N_{E_i}} Super\ Sim(E_i, E_k)} \tag{7}$$

4. Datasets and empirical analysis

In this section, we apply the proposed hypernetwork method to conduct a user ratings analysis using a dataset from a popular social network, *Douban*. Launched in 2005, *Douban* is a typical Chinese Social Networking Site (SNS) that allows registered users to create content related to films, books, music, and recent events and activities in Chinese cities. The three shared layers of films, books, and music are the core functions. *Douban* is open to both registered and unregistered users. In addition to the site serving as a social network website and record keeper, registered users can recommend potentially interesting books, movies and music; unregistered users can find ratings and reviews of books, movies and music on the site. *Douban* currently has over 100 million users and so a huge number of ratings and reviews of books, movies and music are posted on this site every day. We selected *Douban* as our target SNS for the following two reasons. First, unlike *Facebook*, which released its rating system in 2013, *Douban* has more historical data for research, which should make our results more reliable in analyzing uses' precise interests and preferences. Second, unlike *Movielens* and *Amazon*, *Douban* is a comment website that is widely used for finding intensive ratings and reviews of books, movies and music. Analyzing different types of rating datasets will enable us to

Table 1
Dataset distribution extracted from *Douban*.

Datasets	Number of users	Number of objects	Number of records	Time period
Books	102	6,591	9,399	From 2005-8-1 to 2011-11-5
Movies	182	11,457	49,117	From 2005-8-1 to 2011-11-5
Music	241	30,073	53,296	From 2005-8-1 to 2011-11-5

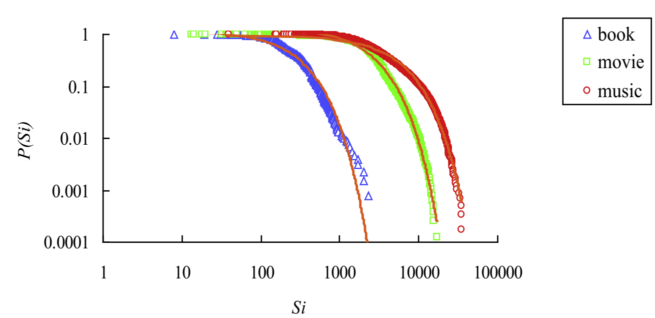


Fig. 3. Cumulative distribution function of the strengh of nodes for each hypernetwork (log–log plot).

Table 2
The strength of nodes in each of the three hypernetworks.

Datasets	Minimum $S_{i\min}$	Maximum $S_{i\max}$	Average $\bar{S}_i$
Books	2	2288	288.7
Movies	13	16,891	2680.3
Music	39	34,675	4555.8

Table 3
The results of the node hyperdegree analysis for the three hypernetworks.

Dataset	Minimum $D_{H_i\min}$	Maximum $D_{H_i\max}$	Average $\bar{D}_{H_i}$
Books	1	29	1.43
Movies	1	92	4.28
Music	1	53	1.77

Please cite this article in press as: Suo, Q., et al. User ratings analysis in social networks through a hypernetwork method. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.05.054

identify the specific characteristics of the various ratings among reviews of books, movies and music.

4.1. Datasets

We can now conduct an empirical study of the aforementioned characteristics using real world datasets collected from the online Chinese social network *Douban* (http://www.douban.com). This web site contains a multifaceted social network, including groups, locations, comments and ratings of content. Registered users can contribute to the system and share their consumption experiences by evaluating objects on a discrete ratings scale from one to five, where higher ratings mean "more preferred", and comment on them. They can also read how others have talked about objects and provide feedback.

Users can be connected on the platform based on individual users' interests in different kinds of objects. In this study of *Douban*, online communications remains the central theme. Our objective was to further explore the dimension of users' ratings and to this end we collected three ratings datasets over equal periods of time, namely book ratings, movie ratings and music ratings (Table 1), using the hypernetwork method to analyze these datasets. Taking the book ratings dataset as an example, each entry has three attributes: personid, bookid, and ratings. This dataset consists of 9399 records of 6591 books selected by 102 users. A node in this hypernetwork represents a book, and a hyperedge connects the sets of books evaluated by the same user. In total, there are 6591 nodes that are encircled by 102 hyperedges. Since a book may be evaluated by more than one user and the users may read more than one book, most hyperedges will inevitably overlap with others, thus creating the book hypernetwork. For the sake of comparison, the other two hypernetworks based on movie and music datasets are also obtained in the same way. The first step is to analyze the topological characteristics of these three hypernetworks.

4.2. The distribution of the strength of nodes

If two objects are evaluated by the same user, their corresponding nodes belong to the same hyperedge. With the aid of hyperedges, a node is connected to other nodes via their co-membership in hyperedges. Given that the use of the strength of the node is more appropriate than node degree for accurately depicting the real links among adjacent nodes, we will not discuss the node degree here. The results are shown in Table 2, which shows that the the strength of nodes for music is significantly larger than that of either books or movies ($\bar{S}_{i(book)} < \bar{S}_{i(movie)} < \bar{S}_{i(music)}$). There is a simple explanation for this; since it takes a much shorter time to listen to a song than to read a book or watch a movie, the total number of songs in the dataset is much bigger than either of the others. Thus, each node is far more likely to be connected to other nodes. According to the results in Table 2, we can obtain several nodes with the largest strength. In our study, we regard the nodes with largest strengths as the central nodes in the hypernetwork.

The cumulative distributions of the strength of nodes strengths are shown in Fig. 3, which shows that they each follow an exponential distribution:

$$\begin{aligned}
&\text{The function for books} \quad y = 1.1786e^{-0.0042x} \quad (R^2 = 0.9533)\\
&\text{The function for movies} \quad y = 1.386e^{-0.0005x} \quad (R^2 = 0.9876)\\
&\text{The function for music} \quad y = 0.9904e^{-0.0002x} \quad (R^2 = 0.9887)
\end{aligned} \tag{8}$$

As these are exponential distributions, the probability that two nodes will be connected to each other is random and uniform. Consequently, most of the nodes' connection numbers will be

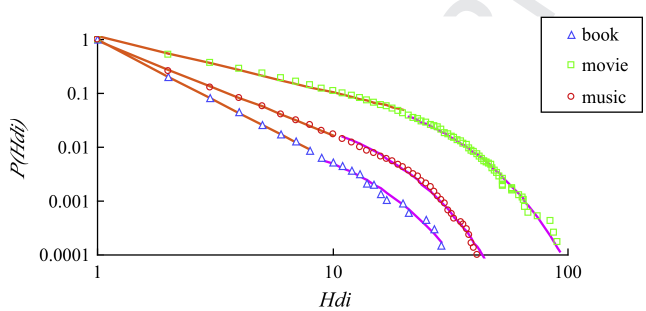


Fig. 4. Cumulative probability distribution function of the node hyperdegree for the three hypernetworks (log–log plot).

Table 4
The results of the analysis of the strength of hyperedges in the three hypernetworks.

Dataset	Minimum $S_{E_i\,min}$	Maximum $S_{E_i\,max}$	Average $\bar{S}_{E_i}$
Books	1	802	161.4
Movies	3	21,323	4095.3
Music	1	8449	894.7

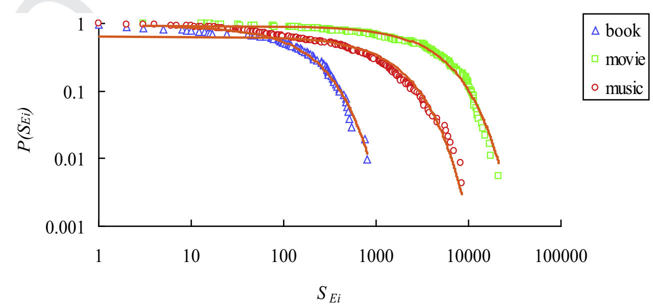


Fig. 5. Cumulative distribution function of the strength of hyperedges in each hypernetwork (log–log plot).

Table 5
The results of hyperedge hyperdegree.

Dataset	Minimum $D_{H_{E_i}\,min}$	Maximum $D_{H_{E_i}\,max}$	Average $\bar{D}_{H_{E_i}}$
Books	6	883	92.2
Movies	5	1997	269.9
Music	1	6959	221.1

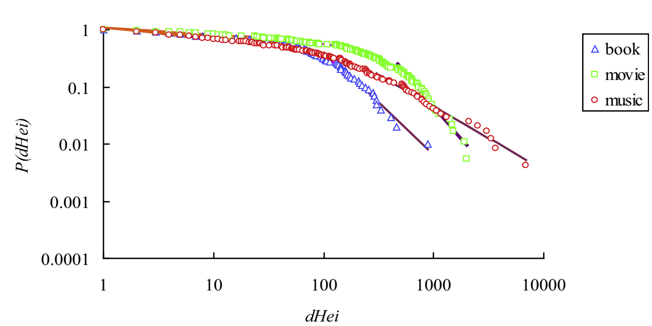


Fig. 6. Cumulative distribution functions of the hyperedge hyperdegrees for the three hypernetworks (log–log plot).

roughly the same, which means that values that are either much higher or much lower than the average will be rare. The degree distribution difference is thus not as obvious as in a power-law distribution. Consider the book dataset as an example. If a node is encircled by a small number of hyperedges, this indicates that the book has been read by a small number of users. This might be because the content of that book is not popular or most of the users are not interested in it. However, if a node appears in a large number of hyperedges or in the biggest hyperedge (big means that the hyperedge hyperdegree is big), the probability that it will connect to other nodes is greater. As a result, the nodes are assigned a larger strength.

4.3. The distribution of the node hyperdegree

A social network analysis that is based on a regular graph does not allow us to understand how many groups a particular node is participating in, but this information can be obtained easily from a hypernetwork. The node hyperdegree is used to denote the number of users that have evaluated each object. The results are shown in Table 3, which reveals that most of the objects are evaluated by only a few users, resulting in a relatively sparse dataset. According to the results in Table 3, we can obtain several central nodes with the largest hyperdegree. The more people post comments, the more attention the corresponding objects receive.

The cumulative distribution of the node hyperdegree is shown in Fig. 4, and the corresponding functions for the three hypernetworks are as follows:

$$
\begin{aligned}
&\text{Books}\ \ y=\begin{cases}0.9841x^{-2.2553}, & 0<x\leqslant 8 \quad (R^2=0.9997)\\ 0.0284\exp(-0.1756x), & 8<x<30 \quad (R^2=0.9789)\end{cases}\\
&\text{Movies}\ \ (y=\begin{cases}1.1563x^{-1.0493}, & 0<x\leqslant 20 \quad (R^2=0.9904)\\ 0.2026\exp(-0.081x), & 20<x<100 \quad (R^2=0.988)\end{cases}\\
&\text{Music}\ \ y=\begin{cases}0.9325x^{-1.7424}, & 0<x\leqslant 10 \quad (R^2=0.9992)\\ 0.089\exp(-0.1568x), & 10<x<60 \quad (R^2=0.9872)\end{cases}
\end{aligned}
\tag{9}
$$

These three curves can be well fitted by power-law distributions with an exponential cutoff. The downward bend in the trailing edge is usually attributed to resource limitations such as a restricted comment period or limited resource availability; the users may simply not be able to track down all the books, movies or music published in *Douban*. Once a node reaches a certain value, no additional connections can be made. Hence, the probability that users will evaluate more objects decays faster than would be the case for a power law and the distribution has an obvious cut-off rather than a long tail. Our analysis shows that the hyperdegree of most nodes is actually less than ten.

4.4. The distribution of the strengths of hyperedges

If two users have evaluated the same object, then the two hyperedges overlap. Common nodes connect a hyperedge to other hyperedges, and the hyperedge degree denotes the number of users that have evaluated the same objects as a particular user, without considering how many common objects they have evaluated. Given that the strength of a hyperedge has a more practical

Table 6a
The 5 users with the shortest distances in the books dataset.

Rank	Personid	Average shortest distance	Hyperedge hyperdegree	Rank for hyperedge hyperdegree
1	4403499	0.320848	207	13
2	1491149	0.323596	883	1
3	1092298	0.329899	410	2
4	2287877	0.331021	145	21
5	1161690	0.331072	306	5

Table 6b
The isolated users in the books dataset.

Rank	Personid	Hyperedge hyperdegree	Rank for hyperedge hyperdegree
1	1339175	3	84
2	2252935	6	92
3	2252955	2	93
4	4506010	3	94
5	50171557	1	102

Table 7
The 5 users with the shortest distances in the movies dataset.

Rank	Personid	Average shortest distance	Hyperedge hyperdegree	Rank for hyperedge hyperdegree
1	1491149	0.108526	1997	1
2	4403626	0.109083	1918	2
3	1181019	0.109245	1463	4
4	1898300	0.109759	856	13
5	44260269	0.109788	1122	6

Table 8a
The 5 users with the shortest distances in the music dataset.

Rank	Personid	The average shortest distance	Hyperedge hyperdegree	Rank for hyperedge hyperdegree
1	1439016	0.218946	2109	6
2	1215937	0.219091	6959	1
3	2253006	0.219998	2554	5
4	1092563	0.220201	3678	2
5	1491071	0.221262	1061	9

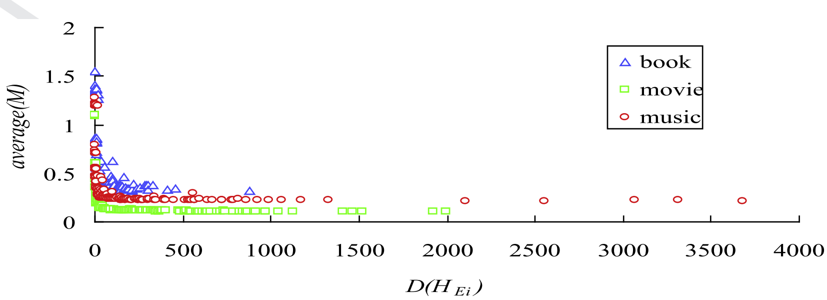


Fig. 7. The average distance between hyperedges.

significance, however, we will focus on the strength of hyperedges rather than its degree here. The results are shown in Table 4, which reveals that the strength of hyperedge for movies is significantly larger than that of either books or music. There is a simple explanation for this: since the number of users that evaluate each movie is much higher than that of either of the other two, the number of objects that each user has evaluated is also higher, so each hyperedge is more likely to be connected to other hyperedges. According to the results in Table 4, we can get several hyperedges with the largest hyperedge hyperdegree. In our research, we regard these hyperedges as the central hyperedges in the hypernetwork. The corresponding users are central users in the hypernetwork.

The cumulative distributions of the strength of hyperedges in each hypernetwork are shown in Fig. 5, which shows that they once again follow exponential distributions with the following functions:

$$\begin{array}{ll} \text{Books} & y = 0.9095e^{-0.0054x} \quad (R^2 = 0.9852) \\ \text{Movies} & y = 0.9217e^{-0.0002x} \quad (R^2 = 0.98) \\ \text{Music} & y = 0.6429e^{-0.0006x} \quad (R^2 = 0.9553) \end{array} \tag{10}$$

Given the nature of exponential distribution, the probability that two hyperedges will be connected to each other is random and uniform. Consequently, most of the hyperedges' connection numbers are roughly the same, which means that a value that is much higher or lower than the average is rare. A user with a large hyperedge strength may take part in a large number of ratings or evaluate objects with the largest hyperdegree. Thus, the probability that he/she is connected to other users is greater and the hyperedge will have a larger strength.

4.5. The distribution of hyperedge hyperdegree

We are also interested in seeking information on the number of nodes that a particular hyperedge contains. This can be obtained in a very straightforward manner from a hypernetwork. The hyperedge hyperdegree is used to denote the number of objects that each user has evaluated: the larger the value, the more objects the user has evaluated. The results are shown in Table 5. Notice that the hyperdegree for books is much smaller than either of the other two. This indicates that the number of books that each user reads and comments on is far smaller than the other two media. The users in movie datasets are most active and the number of users who evaluate individual movies is bigger than that for either of the other two media. According to the results shown in Table 5, we can get several hyperedges with the largest strength of hyperedge hyperdegree in each dataset. In our research, these hyperedges represent the most active users.

The cumulative distributions of the hyperedge hyperdegree are shown in Fig. 6; the functions are as follows:

$$\begin{array}{ll}
\text{Books} & y = \begin{cases} 1.0873x^{-0.1657}, & 0 < x \leqslant 50 \quad (R^2 = 0.9618) \\ 0.8264\exp(-0.0092x), & 50 < x \leqslant 150 \quad (R^2 = 0.9714) \\ 2180.6x^{-1.846}, & 150 < x < 1000 \quad (R^2 = 0.9747) \end{cases} \\
\text{Movies} & y = \begin{cases} 1.0848x^{-0.1244}, & 0 < x \leqslant 60 \quad (R^2 = 0.9587) \\ 0.7703\exp(-0.003x), & 60 < x \leqslant 450 \quad (R^2 = 0.9896) \\ 264308x^{-2.2538}, & 450 < x < 2000 \quad (R^2 = 0.9571) \end{cases} \\
\text{Music} & y = \begin{cases} 1.0822x^{-0.2013}, & 0 < x \leqslant 50 \quad (R^2 = 0.9801) \\ 0.5728\exp(-0.0047x), & 50 < x \leqslant 300 \quad (R^2 = 0.9898) \\ 96.471x^{-1.1092}, & 300 < x < 7000 \quad (R^2 = 0.9793) \end{cases}
\end{array} \tag{11}$$

As the figure shows, all three curves exhibit similar trends. For both small and large values of the hyperedge hyperdegree, they display power-law distributions, but for intermediate values they follow exponential distributions. This may be because the number of objects that each user has evaluated is roughly the same and is near the average, which results in an exponential distribution in the central portion of the curves. Conversely, since nodes with large or small hyperedge hyperdegrees are limited, this results in a power-law distribution at both ends of each curve.

Table 8b
The isolated users in the music dataset.

Rank	Personid	Hyperedge hyperdegree	Rank for hyperedge hyperdegree
1	1126486	1	222
2	1408060	1	223
3	2890304	1	224
4	11264920	1	225
5	36669121	1	226

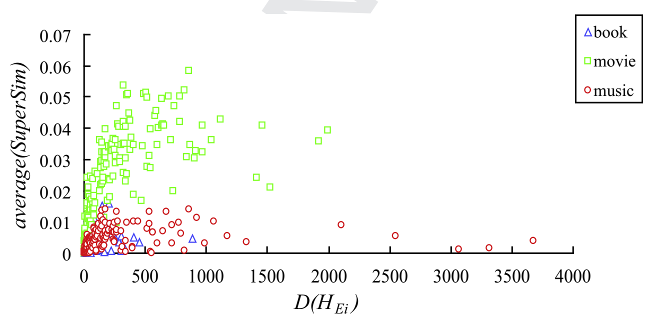


Fig. 8. The average similarity between hyperedges.

Table 9
The collaborative recommendations for the top 5 users with the shortest distances in the books dataset.

Rank	Personid	Top 5 users with the greatest similarity personid ($SuperSim(E_i, E_k)$)	The Objectid for the recommendation($r_{E_i l}$)
1	4403499	2287877 (0.12), 33371957 (0.10), 4506070 (0.09), 36048106 (0.09), 44260269 (0.08)	1023045(2.24), 1770782(2.06), 1367964(1.94), 1029791(1.88), 4138982(1.84)
2	1491149	1161690 (0.05), 2287877 (0.03), 1009907 (0.02), 1092298 (0.02), 1988740 (0.02)	4220020(2.75), 1358873(2.56), 1004821(2.44), 1007433(2.23), 1361249(2.20)
3	1092298	4403499 (0.06), 2287877 (0.04), 1040761 (0.03), 2019605 (0.03), 2392838 (0.03)	1017143(3.77), 1046265(2.67), 1047138(2.61), 1007305(2.61), 1090043(2.57)
4	2287877	4403499 (0.12), 1585841 (0.10), 2602505 (0.08), 1040761 (0.07), 2704486 (0.07)	1529893(2.58), 1914078(2.54), 1029159(2.42), 1023500(2.56), 1873231(2.56)
5	1161690	1491149 (0.05), 2287877 (0.05), 1040761 (0.04), 4403499 (0.03), 1092298 (0.03)	1090043(2.88), 1029791(2.60), 1008074(2.57), 1066462(2.33), 1023045(2.32)

Please cite this article in press as: Suo, Q., et al. User ratings analysis in social networks through a hypernetwork method. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.05.054

4.6. The results of hyperedge–hyperedge distance

The average distance between a particular hyperedge and other hyperedges is shown in Fig. 7. Here, $\overline{M}_{E_iE_j(movie)} < \overline{M}_{E_iE_j(music)} < \overline{M}_{E_iE_j(book)}$, which indicates that there is a closer interrelationship between users in the movie dataset than that in either of the other two. From 4.4, we know that the strength of hyperedges denotes the total number of associated users and $\overline{S}_{E_i(book)} < \overline{S}_{E_i(music)} < \overline{S}_{E_i(movie)}$. Since the strength of hyperedges in the movie dataset is the highest in any of the three datasets, the average distance in the movie dataset is the smallest.

The users with the shortest average distance and isolated users (there are no isolated users in the movies dataset) are listed in Tables 6–8. The users with the shortest average distance have correspondingly high hyperedge hyperdegrees; conversely, the hyperdegrees of the isolated users are usually small. This is mainly because the more nodes a hyperedge encircles, the more hyperedges that hyperedge connects to. As a general rule, a hyperedge with a large hyperdegree connects to a greater number of other hyperedges, which results in shorter distances, while a hyperedge with a small hyperdegree has few, if any, connections to others. This means that hyperedges with shorter distances have more influence than hyperedges with longer distances. Those users with the shortest distance can thus be seen as the opinion leaders in hypernetworks (Liu, Yang, et al., 2014; Liu, Li, et al., 2014; Ma & Liu, 2014). They are not only the most active users in a hypernetwork, but also play an important role in spreading public opinion.

Table 10
The collaborative recommendations for the top 5 users with the shortest distances in the movies dataset.

Rank	Personid	Top 5 users with the greatest similarity personid ($SuperSim(E_i, E_k)$)	The Objectid for the recommendation(r_{E_il})
1	1491149	1092318 (0.17), 1898300 (0.15), 1161569 (0.14), 1057620 (0.14), 22529950 (0.14)	2132495(4.21), 2213597(3.65), 1652587(3.65), 1820156(3.60), 3148748(3.27)
2	4403626	44260269 (0.17), 1181019 (0.17), 1898300 (0.17), 1057620 (0.15), 1161672 (0.14)	1305164(4.43), 1820156(4.28), 1292402(3.98), 1292950(3.81), 3550132(3.71)
3	1181019	44260269 (0.24), 1057620 (0.19), 1126627 (0.17), 1898300 (0.17), 4403626 (0.17)	1295644(4.64), 5344178(4.55), 1292000(4.09), 1299131(4.09), 3072124(3.89)
4	1898300	44260269 (0.21), 1057620 (0.19), 2287877 (0.18), 22529950 (0.18), 1161672 (0.18)	1291549(4.62), 1291999(4.43), 1300616(4.03), 3205624(3.68), 1295409(3.66)
5	44260269	1181019 (0.24), 2572220 (0.22), 1898300 (0.21), 1057620 (0.18), 22529950 (0.18)	2209573(5), 3793023(4.24), 1291561(4.14), 1858711(4.12), 1291548(3.92)

Table 11
The collaborative recommendations for the top 5 users with the shortest distances in the music dataset.

Rank	Personid	Top 5 users with the greatest similarity personid ($SuperSim(E_i, E_k)$)	The Objectid for the recommendation(r_{E_il})
1	1439016	1491071 (0.07), 2502133 (0.06), 1009892 (0.06), 1491428 (0.06), 1215937 (0.06)	1417475(2.94), 2153935(2.82), 2359621(2.61), 3566603(2.47), 2995812(2.45)
2	1215937	1439016 (0.06), 1009892 (0.04), 2502133 (0.03), 1491071 (0.02), 2253006 (0.02)	2347182(4.00), 1394767(2.91), 1394547(2.68), 4323489(2.53), 2072279(2.51)
3	2253006	1491071 (0.06), 1439016 (0.06), 1092563 (0.05), 2502133 (0.04), 22529950 (0.04)	2072279(3.55), 1397543(3.27), 1419566(3.24), 1394791(3.09), 1439133(3.03)
4	1092563	1009666 (0.05), 2253006 (0.05), 1439016 (0.05), 1561355 (0.05), 1491428 (0.04)	3041487(3.43), 1394547(3.26), 3590980(2.88), 1394718(2.70), 2131368(2.70)
5	149107(3.31)	2502133 (0.09), 1126518 (0.08), 1439016 (0.07), 2252993 (0.06), 3883291 (0.06)	1415369(2.46), 1394798(2.16), 2132581(2.16), 5365287(2.16), 1394571(2.13)

4.7. The results of collaborative filters and personal recommendations

For each hyperedge, the average hyperedge similarity between a particular hyperedge and other hyperedges is shown in Fig. 8. As the graph clearly demonstrates, the similarity of the movies dataset is significantly higher than that of either of the other two. In this respect, the precision of the collaborative recommendation will be higher in the movies dataset.

The previous section showed that opinion leaders are those with the shortest hyperedge–hyperedge distance in each dataset, and hyperedge similarity can also be used to identify the top 5 users whose interests are most similar to a particular user as they constitute the neighborhood of that user. We can thus predict a user's ratings for unevaluated objects based on his or her neighbors' rating data, so the objects getting the highest ratings from them are likely to be of the most interest to that user as well. The first 5 objects are thus recommended to the users, as shown in Tables 9–11 and these ratings for objects will vary from user to user as the ratings of objects recommended to users are consistent with each user's personalized features.

5. Conclusions and Implications for future research

In this paper, the hypernetwork method was applied to analyze user ratings in social networks. Our results lead to the following observations and contributions to the normative literature.

First, based on the concept of a hypernetwork and the definition of characteristics in complex networks, several new topological characteristics of hypernetworks are defined, namely *the strength of the node*, the *strength of a hyperedge*, and *hyperedge hyperdegree*. The collaborative recommendation approach in hypernetworks was also introduced and is confidently expected to provide a solid foundation for future research in this area.

Second, a new method for applying hypernetworks to analyze user ratings data is proposed, which will contribute to empirical analyses of other similar networks. Hypernetworks can be constructed where the users are defined as hyperedges and the objects as nodes. This method offers a very effective way of depicting the relationships between users and objects.

Third, we utilized empirical data to analyze the distribution network topology characteristic from different perspectives and portrayed the resource evaluation relationships among various characteristics of users, resources, and the connection strengths between them. This involved evaluating the characteristics and laws of the resources of social network users at a macro level.

Please cite this article in press as: Suo, Q., et al. User ratings analysis in social networks through a hypernetwork method. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.05.054

Our results indicate that this approach may be a good way for Internet service providers to make more effective use of their user resource management and develop better personalized recommendations. Future research directions that build on the findings reported herein include:

- These characteristics provide a starting point for developing a deeper understanding and in-depth analysis of social networks. The mechanisms that dominate the emergence of hypernetworks are still relatively unknown, so evolving models capable of effectively depicting the growth of hypernetworks will be a potential focus of future research.
- This study analyzed data from the *Douban* SNS, although we believe our methodology can be also applied to analyze data from other SNSs such as *Twitter*, *Facebook*, and *Movielens*. A comparative features analysis of the different social networks' resource comments would thus be interesting.
- As yet, we have only considered the user's rating data and ways to further develop accurate personalized recommendations remain challenging. We posit that a more advanced recommendation needs to integrate more user properties in order to make an in-depth analysis of user characteristics.
- The collaborative recommendations in this study were based on a very basic standard algorithm. Adopting an improved algorithm for future research could markedly improve the prediction accuracy.
- Notably, this study did not consider temporal aspects of users' comments on resources. Future studies could focus on the mechanisms of human dynamics at different review stages by including the factor of comment time. This may be a better way of dealing with users' constantly shifting interests and preferences.
- This research provides the impetus to test these assumptions. As such, a further study should to consider this stream.

References

Araújo, E. B., Moreira, A. A., Furtado, V., Pequeno, T. H., & Andrade, J. S. Jr, (2014). Collaboration networks from a large CV database: dynamics, topology and bonus impact. *PloS one, 9*(3), e90537.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Berge, C. (1973). *Graphs and hypergraphs* (Vol. 7). Amsterdam: North-Holland Publishing Company.

Berge, C. (1989). *Hypergraphs: Combinatorics of finite sets* (Vol 45). Amsterdam: North-Holland Publishing Company.

Candillier, L., Meyer, F., & Fessant, F. (2008). Designing specific weighted similarity measures to improve collaborative filtering systems. In *Advances in data mining. Medical applications, E-commerce, marketing and theoretical aspects* (pp. 242–255). Berlin, Heidelberg: Springer.

Criado, R., Romance, M., & Vela-Pérez, M. (2010). Hyperstructures, a new approach to complex systems. *International Journal of Bifurcation and Chaos, 20*(03), 877–883.

Deng, S. G., Huang, L. T., & Xu, G. D. (2014). Social network-based service recommendation with trust enhancement. *Expert Systems with Applications, 41*(18), 8075–8084.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowd sourcing systems on the world-wide web. *Communications of the ACM, 54*(4), 86–96.

Estrada, E., & Rodríguez-Velázquez, J. A. (2006). Subgraph centrality and clustering in complex hyper-networks. *Physica A: Statistical Mechanics and its Applications, 364*, 581–594.

Ghoshal, G., Zlatić, V., Caldarelli, G., & Newman, M. E. J. (2009). Random hypergraphs and their applications. *Physical Review E, 79*(6), 066118.

Hajli, N., & Lin, X. (2014). Exploring the security of information sharing on social networking sites: The role of perceived control of information. *Journal of Business Ethics*, 1–13.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5–53.

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles, 44*, 223–270.

Krawiecki, A. (2014). Chaotic synchronization on complex hypergraphs. *Chaos, Solitons & Fractals, 65*, 44–50.

Li, J., Peng, W., Li, T., Sun, T., Li, Q., & Xu, J. (2014). Social network user influence sense-making and dynamics prediction. *Expert Systems with Applications, 41*(11), 5115–5124.

Li, X., & Chen, H. (2013). Recommendation as link prediction in bipartite graphs: a graph kernel-based machine learning approach. *Decision Support Systems, 54*(2), 880–890.

Liang, B., Liu, Y. Q., Zhang, M., Ma, S. P., Ru, L. Y., & Zhang, K. (2014). Searching for people to follow in social networks. *Expert Systems with Applications, 41*(16), 7455–7465.

Lind, P. G., González, M. C., & Herrmann, H. J. (2005). Cycles and clustering in bipartite networks. *Physical Review E, 72*(5), 056127.

Liu, J. G., Yang, G. Y., & Hu, Z. L. (2014). A knowledge generation model via the hypernetwork. *PloS One, 9*(3), e89746.

Liu, X., & Murata, T. (2011). Detecting communities in *k*-partite *k*-uniform (hyper) networks. *Journal of Computer Science and Technology, 26*(5), 778–791.

Liu, Y., Li, Q., Tang, X., Ma, N., & Tian, R. (2014). Superedge prediction: What opinions will be mined based on an opinion supernetwork model? *Decision Support Systems., 64*, 118–129.

Ma, N., & Liu, Y. (2014). Superedge Rank algorithm and its application in identifying opinion leader of online public opinion supernetwork. *Expert Systems with Applications, 41*(4), 1357–1368.

Michoel, T., & Nachtergaele, B. (2012). Alignment and integration of complex networks by hypergraph-based spectral clustering. *Physical Review E, 86*(5), 056111.

Ramasco, J. J., Dorogovtsev, S. N., & Pastor-Satorras, R. (2004). Self-organization of collaboration networks. *Physical Review E, 70*(3), 036106.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM, 40*(3), 56–58.

Rodder, W., Brenner, D., & Kulmann, F. (2014). Entropy based evaluation of net structures-deployed in social network analysis. *Expert Systems with Applications, 41*(17), 7968–7979.

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In The. Adaptive (Ed.), *Web* (pp. 291–324). Berlin, Heidelberg: Springer.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Wang, J. W., Rong, L. L., Deng, Q. H., & Zhang, J. Y. (2010). Evolving hypernetwork model. *The European Physical Journal B-Condensed Matter and Complex Systems, 77*(4), 493–498.

Xie, H., Ma, J., & Yang, L. (2013). Similarity of user ratings play a positive information filter in social networks. *Journal of the Korean Physical Society, 62*(9), 1221–1226.

Yang, G. Y., & Liu, J. G. (2014). A local-world evolving hypernetwork model. *Chinese Physics B, 23*(1), 018901.

Zhang, Y. C., Blattner, M., & Yu, Y. K. (2007). Heat conduction process on community networks as a recommendation model. *Physical Review Letters, 99*(15), 154301.

Zhang, Z. K., & Liu, C. (2010). A hypergraph model of social tagging networks. *Journal of Statistical Mechanics: Theory and Experiment, 2010*(10), P10005.

Zhou, T., Ren, J., Medo, M., & Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. *Physical Review E, 76*(4), 046115.

Zlatić, V., Ghoshal, G., & Caldarelli, G. (2009). Hypergraph topological quantities for tagged social networks. *Physical Review E, 80*(3), 036118.

Please cite this article in press as: Suo, Q., et al. User ratings analysis in social networks through a hypernetwork method. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.05.054