

SCIENTIFIC REPORTS



OPEN

Uncovering the essential links in online commercial networks

Wei Zeng^{1,2,3}, Meiling Fang^{1,2,3}, Junming Shao^{2,3} & Mingsheng Shang⁴

Received: 22 April 2016

Accepted: 09 September 2016

Published: 29 September 2016

Recommender systems are designed to effectively support individuals' decision-making process on various web sites. It can be naturally represented by a user-object bipartite network, where a link indicates that a user has collected an object. Recently, research on the information backbone has attracted researchers' interests, which is a sub-network with fewer nodes and links but carrying most of the relevant information. With the backbone, a system can generate satisfactory recommendations while saving much computing resource. In this paper, we propose an enhanced topology-aware method to extract the information backbone in the bipartite network mainly based on the information of neighboring users and objects. Our backbone extraction method enables the recommender systems achieve more than 90% of the accuracy of the top-L recommendation, however, consuming only 20% links. The experimental results show that our method outperforms the alternative backbone extraction methods. Moreover, the structure of the information backbone is studied in detail. Finally, we highlight that the information backbone is one of the most important properties of the bipartite network, with which one can significantly improve the efficiency of the recommender system.

Recently, recommender systems have been widely studied and designed to assist individuals in finding their interested contents such as books, movies and music. So far, many recommendation algorithms have been proposed, such as collaborative filtering (CF)¹, matrix factorization², latent semantic indexing³, multi-dimensional scaling⁴ and so on. In addition, a recommender system can be represented by a user-object bipartite network, where a node denotes a user or an object and a link indicates that a user has collected an object. By relying on the network topology, some network propagation processes such as the mass diffusion^{5,6} and the heat conduction⁷, have been applied in the recommender system. Researchers also find that the hybridization of these two methods can solve the accuracy-diversity dilemma in the recommender system⁸. More recently, a semi-local diffusion method is proposed to solve the sparsity problem in the recommender system⁹.

One of the most important challenges in the research on recommender system is how to improve the efficiency of a recommendation algorithm. It is of great importance for a recommender system to generate instant object recommendations for individual users to keep their loyalty. In the recent, some network manipulation methods are proposed to address the problem. Zeng¹⁰ studied the relevance of users in a recommender system and found that there exists a small group of key users. With them, the recommender system can generate satisfactory recommendations. That is, not all the information in the user-object bipartite is necessary when recommending objects for users. Similar conclusion can be found in ref. 11, which reveals that the performance of recommendation algorithms can be improved by eliminating some redundant links in the network. The authors proposed time-aware and topology-aware link removal algorithms to extract the information backbone in a recommender system.

The backbone is one of the most important properties of the complex network^{12,13}. It is a sub-network with fewer links and nodes but preserving the topological properties or the function of the original network such as the degree distribution¹⁴, transportation ability¹⁵ and betweenness¹⁶. In the user-object bipartite network, an information backbone is defined as a sub-network with a group of users, a set of objects and their links, which carries most of the relevant information for object recommendations. Only relying on the information backbone, a recommender system's efficiency can be greatly improved.

Although some methods are proposed to uncover the information backbone, there are still some problems left. For instance, most of these methods are utilized in unipartite networks^{13,16–18}. Few of them can be used in bipartite

¹Trusted Computing and Automated Reasoning Lab, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China. ²Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China. ³Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, China. ⁴Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, P. R. China. Correspondence and requests for materials should be addressed to W.Z. (email: zwei504@uestc.edu.cn) or M.S. (email: msshang@cigit.ac.cn)

networks. Moreover, the information of neighboring users and objects is not considered for the topology-aware backbone and the evolution of the network is not taken into account for the time-aware backbone¹¹. In this work, we propose a topology-aware backbone extraction method which exploits the structure of the subgraph consisting of neighboring users and objects. After verification on three real-world bipartite networks, the obtained results show that our method performs better than the alternative backbone extraction techniques. Moreover, structure properties of the information backbone are analyzed in detail.

Results

A recommender system can be naturally denoted by a user-object bipartite network $G(U, O, E)$, where $U = \{u_1, u_2, \dots, u_n\}$, $O = \{o_1, o_2, \dots, o_m\}$ and $E = \{e_1, e_2, \dots, e_l\}$ are sets of users, objects and links, respectively. It can also be represented by an adjacency matrix $A = \{a_{i\alpha}\}$, where the element $a_{i\alpha} = 1$ if user i has collected object α , and 0 otherwise (We use Greek and Latin letters, respectively, for object-related and user-related indices). For a target user to whom we will recommend objects, each of her uncollected objects will be assigned a score by the recommendation algorithm and the top- L objects with the highest scores will be recommended. Different algorithms generate different object scores and thus different recommendation lists for users.

To the best of our knowledge, some methods are proposed to uncover the information backbone in a bipartite network. The first one is simply based on the popularity of links, with the underlying hypothesis that the relevance of a link can be reflected by its popularity (Popularity for short). The popularity of link $e_{i\alpha}$ is defined as $k_i k_\alpha$, where k_i and k_α are the degree of user i and object α , respectively. The information backbone thus consists of links with the largest popularity. The second one is to randomly select a set of links as the information backbone. This method is used as a benchmark for comparison (Random for short). In the third method, the frequency of rectangles that each link belongs to is utilized to measure the relevance of links. A rectangle is defined as a subgraph consisting of a pair of users and objects. Those links which appear most frequently are selected as the information backbone (Rectangle for short). As mentioned above, these methods didn't take into account the information of neighboring users and objects, which may make a recommendation algorithm inaccurate. In this paper, we propose a subgraph-based method to uncover the information backbone, which computes the frequency of user-oriented and object-oriented subgraphs that each link belongs to. For user i , we can get her K nearest neighbors by ranking the cosine similarity: $s_{ij} = |I_i \cap I_j| / \sqrt{k_i k_j}$, where I_i is the set of objects collected by user i . Those neighboring users and their collected objects make up a subgraph. Likewise, we can also obtain a set of objects which are the most similar to i 's collected objects. Each of i 's uncollected objects will be assigned a score: $P_{i\alpha} = \sum_{\beta \in I_i} s_{\alpha\beta}$, where $s_{\alpha\beta} = |U_\alpha \cap U_\beta| / \sqrt{k_\alpha k_\beta}$ is the similarity between object α and β . Those objects with highest scores and their links can also comprise a subgraph. We then count the frequency that a link belongs to these two kinds of subgraphs. By repeating the above process over all users, we can get the final frequency of each link and those links with the largest frequency will be chosen as the information backbone (Subgraph for short). A visual representation of the subgraph-based method is given in Fig. 1.

Once we get the information backbone, only the links in the backbone are used to run the recommendation algorithm. We then compare the performance of the recommendation algorithm relying on the backbone with those relying on the original network. Tests are executed between our information backbone extraction method and four recommendation algorithms (see Methods), namely mass diffusion (MD for short)⁵, mass diffusion and heat conduction hybrid method (Hybrid for short)⁸, User-based collaborative filtering (UCF for short)¹ and Item-based collaborative filtering (ICF for short)¹⁹.

We run these four recommendation algorithms on the original bipartite network and the information backbone, respectively. For those users who have no links in the backbone, we recommend the most popular objects as a compromise. In order to evaluate the performance of our methods, we utilize three data sets in this paper, i.e. Douban, Flixster and Flickr (see Table 1). Precision as well as Ranking score are computed to evaluate the performance of recommendation algorithms (see Methods and SI).

Generally speaking, the more data we used in a recommender system, the higher accuracy we can get. In this paper, our purpose is to find the most relevant links in the user-object bipartite network, with which a recommendation algorithm can achieve a satisfactory accuracy. The result is presented in Fig. 2, where r denotes the fraction of links in the information backbone. When $r = 1$, all the links will be used in the recommendation algorithms, equivalent to the traditional method. For the subgraph-based method, the number of neighboring users and objects is set to 10 (see the results with different neighbor size in SI).

We make use of *precision* index to measure the top- L accuracy of the algorithm. Naturally, the accuracy of the recommendation algorithm is lowered if less information is used ($r < 1$). However, the accuracy doesn't change too much if we use the subgraph-based backbone. Taking the UCF algorithm in the Douban dataset for example, 96.2%(0.075/0.078) of the accuracy can be achieved when we use only 20% links ($r = 0.2$). It outperforms other methods especially when the size of the information backbone is very small. The similar results can be found in Flickr and Flixster datasets. This is of great importance since the algorithmic efficiency of recommendation methods can be largely improved if we take into account fewer links. In addition to *precision*, we also choose the *ranking score* as the accuracy metric and the results are presented in Fig. 3. To save space, we only plot the result of the UCF recommendation algorithm. It can be seen that recommendation algorithms have the best performance in the backbone extracted by random method. This is because objects selected by the random method are more diverse than other methods and more objects will get resources distributed by users' selected objects⁹. From practical point of view, online retailers are generally more concerned with the accuracy of the top- L object recommendations. When only the links in the backbone are considered, some users will become isolated, with no adjacent links in the network. The ratio of those isolated users are 4%(480/11898), 3.9%(797/20641), 3.6%(360/10000) for Douban, Flickr and Flixster, respectively. For these users, we recommend the most popular objects.

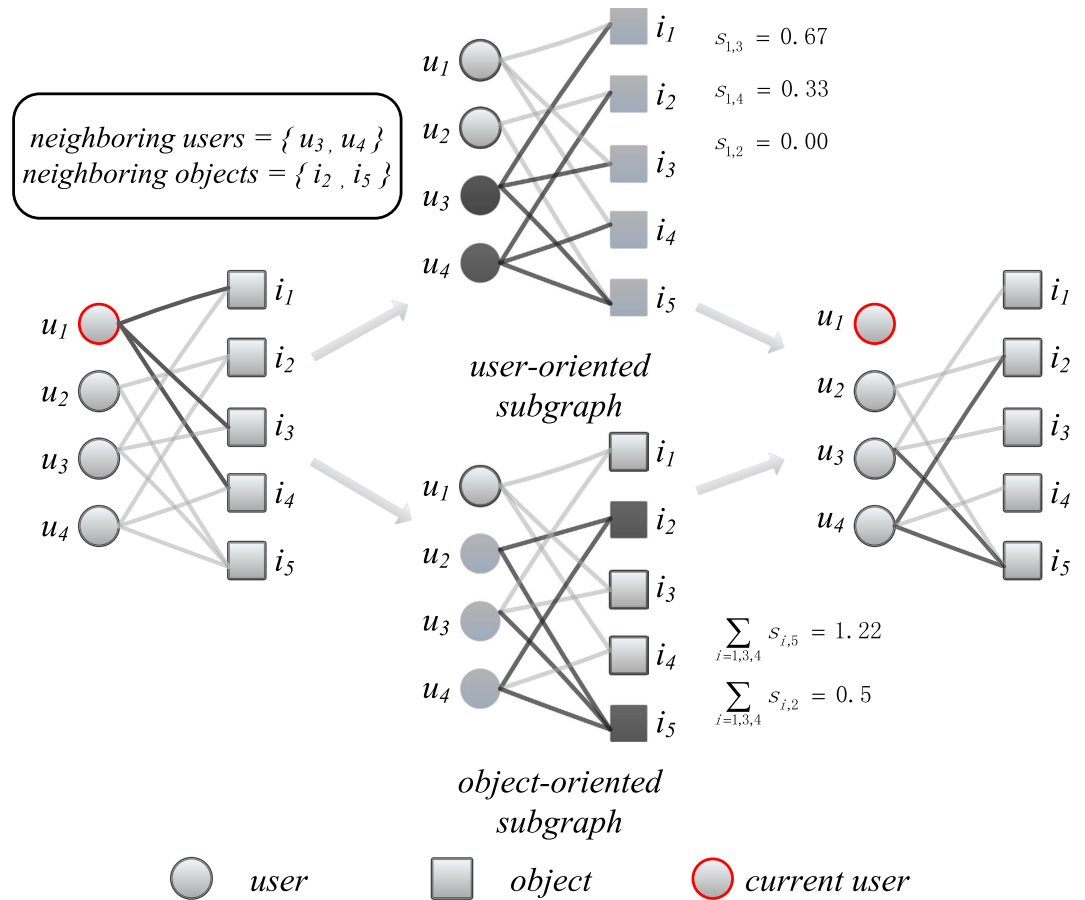


Figure 1. A visualization of the subgraph-based method. Given user u_1 , the user-oriented and object-oriented subgraph are constructed by relying on u_1 's neighboring users and objects.

Dataset	#Users, n	#Objects, m	#Links, l	Sparsity
Douban	11,898	222,815	2,037,736	7.69×10^{-4}
Flickr	20,641	58,768	1,795,730	1.48×10^{-3}
Flixster	10,000	40,758	2,106,568	5.17×10^{-3}

Table 1. The statistics of Douban, Flickr and Flixster datasets. The sparsity is defined as $\frac{l}{n \times m}$.

From the Fig. 2, it can be seen that the random method is sometimes better than the method based on computation (e.g. popularity-based method). In order to uncover the relevance of links in detail and to see their differences, we evenly divide all links into 10 groups C_1, C_2, \dots, C_{10} according to their weights computed by the information backbone extraction methods. If $i < j$, the weights of all links in C_i are larger than that in C_j . We also choose the above four recommendation algorithms to test the information backbone extraction methods and the results are presented in Fig. 4, where only the links in C_i are used to run the algorithm. Likewise, we make use of the *precision* index to test the accuracy of the algorithms. Firstly, for subgraph-based, popularity-based and rectangle-based methods, the accuracy of recommendation algorithms considering links in group C_i is better than C_j with $i < j$, which means all these methods have the ability to uncover the relevant links compared to the random method. Besides, when only the links in C_1 are considered, the subgraph-based method is the best one compared to other methods, which means the relevant links uncovered by subgraph-based method carry the most information of the recommender system.

Moreover, we investigate the structure of the information backbone. For simplicity, the relative size of the information backbone (r) is set as 0.2. We then compute the average degree ($\langle k_u^s \rangle$) of users and the average degree ($\langle k_o^s \rangle$) of objects which are selected by users in the information backbone. The result is presented in Table 2. Firstly, $\langle k_u^s \rangle$ in the rectangle-based information backbone is the largest, which means many large degree users are selected in the backbone. Secondly, the backbone obtained by random method has the smallest $\langle k_o^s \rangle$ since this method chooses diverse objects. For the graph-based backbone, its objects are less diverse than objects selected by random method, but its $\langle k_o^s \rangle$ is much smaller than $\langle k_o^s \rangle$ of the popularity-based and rectangle-based backbone.

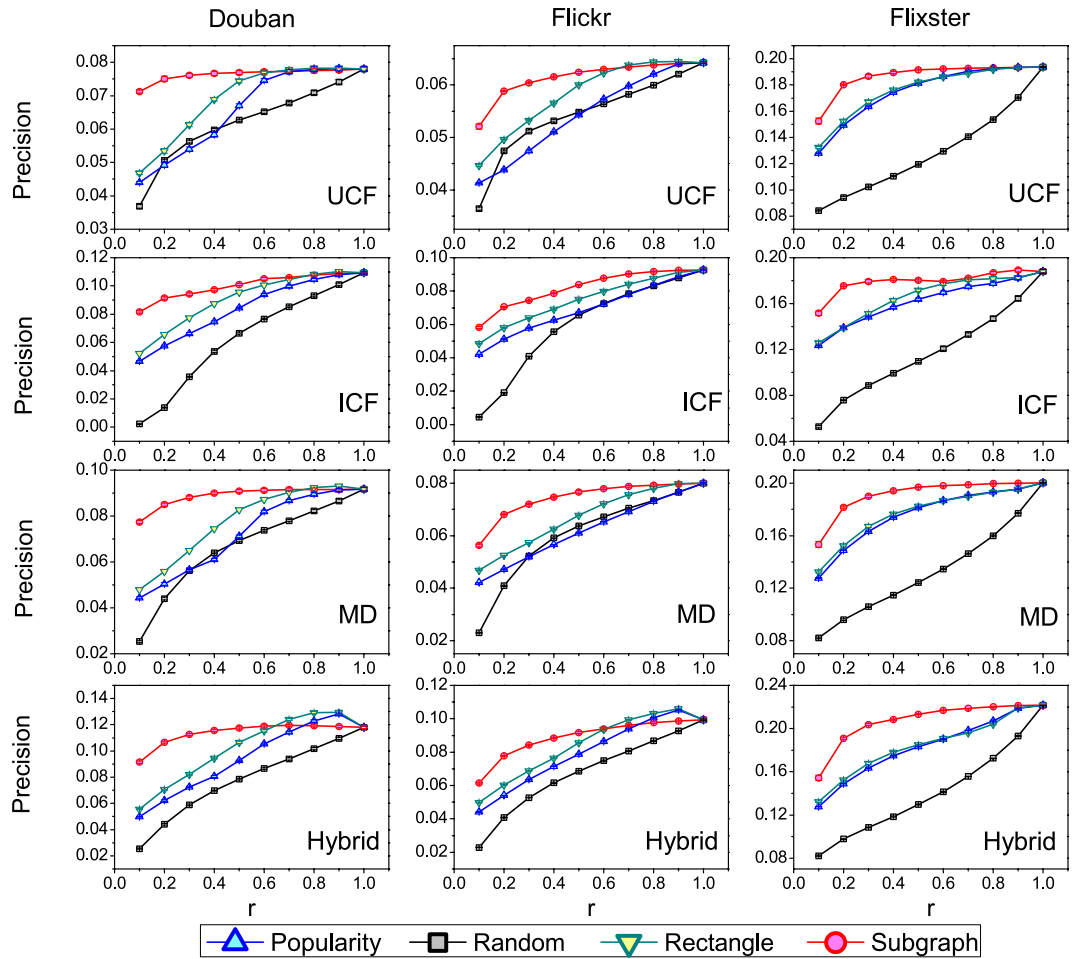


Figure 2. The recommendation accuracy contributed only by the backbone in recommender system. The recommendation length L is set to 20. For the subgraph-based method, the number of neighboring users and objects is set to 10. r is the ratio of the size of the backbone to the whole system. The error bars are obtained based on 5 independent instances of training and probe set.

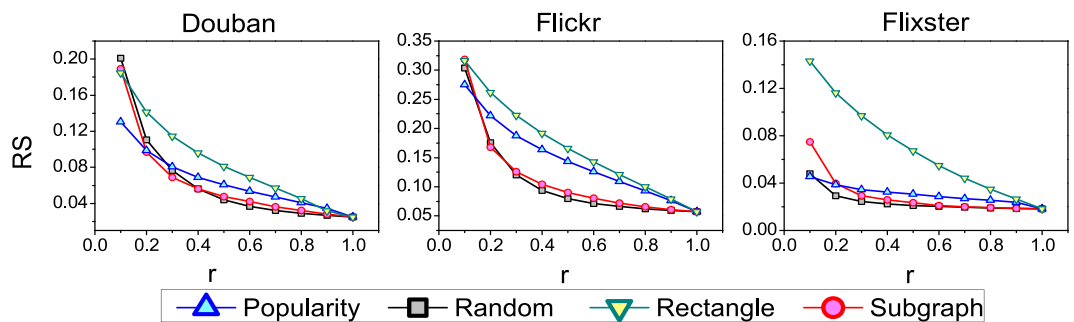


Figure 3. The recommendation accuracy (ranking score) contributed only by the backbone. For the subgraph-based method, the number of neighboring users and objects is set to 10. r is the ratio of the size of the backbone to the whole system.

In addition, we choose another three indices to uncover the structure of the information backbone: degree heterogeneity, clustering coefficient and diffusion coverage. The degree heterogeneity is defined as $\langle k_\alpha^2 \rangle / \langle k_\alpha \rangle^2$ where $\langle k_\alpha \rangle$ is the average degree of objects. The results are presented in Table 3, which reveals that the subgraph-based backbone has the largest degree heterogeneity, in other words, this method tends to select those high quality objects in the backbone²⁰.

For an object α , its clustering coefficient is defined as $c_{ij}(\alpha) = \frac{q_{\alpha ij}}{(k_i - \eta_{\alpha ij})(k_j - \eta_{\alpha ij}) + q_{\alpha ij}}$, where i and j are neighboring users of object α , $q_{\alpha ij}$ is the number of common neighbors between i and j (not counting α), $\eta_{\alpha ij} = 1 + q_{\alpha ij}$.

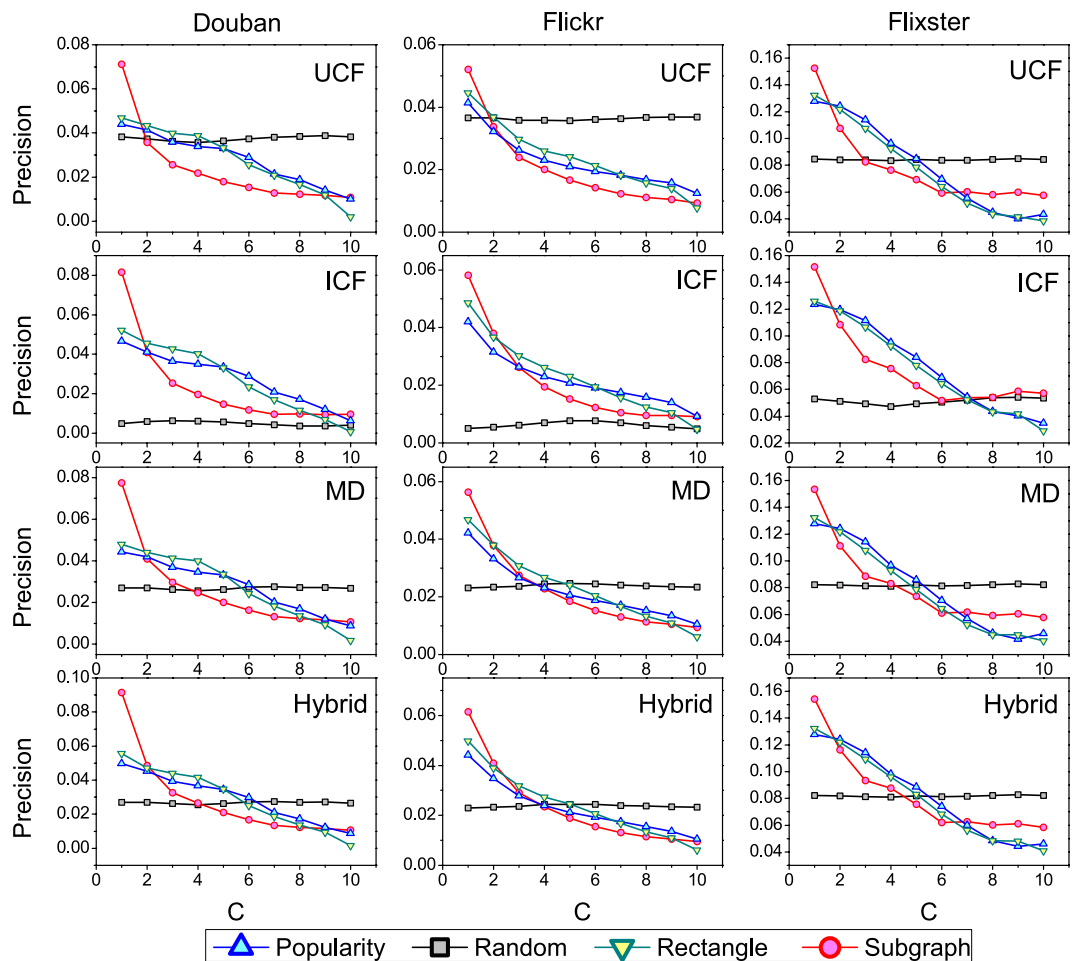


Figure 4. The performance of recommendation algorithms in each subset. For the subgraph-based method, the number of neighboring users and objects is set to 10. The recommendation length L is set to 20.

Methods	Douban		Flickr		Flixster	
	$\langle k_u^s \rangle$	$\langle k_o^s \rangle$	$\langle k_u^s \rangle$	$\langle k_o^s \rangle$	$\langle k_u^s \rangle$	$\langle k_o^s \rangle$
Random	150.33	15.57	71.01	37.44	176.64	84.89
Subgraph	145.50	42.48	72.11	84.67	175.39	318.71
Popularity	413.85	65.48	86.83	184.34	461.66	338.10
Rectangle	392.49	121.08	101.61	236.45	480.21	889.41

Table 2. The average degrees of users and the average degrees of objects which are selected by these users in the information backbone.

The clustering coefficient of α is defined as the average of $c_{ij}\alpha$ over all the users. The clustering coefficient of the network is the average clustering coefficient of all the objects²¹. Generally speaking, the backbone has a smaller clustering coefficient since a lot of links are removed. For the Douban and Flickr datasets, the subgraph-based backbone has the largest clustering coefficient, which means the objects in the backbone are highly overlapped. This could explain why the subgraph-based backbone has the best predictive accuracy²². For the Flixster dataset, the rectangle-based backbone has the largest clustering coefficient since the degree distribution of object is highly uneven (see Table 2).

The diffusion coverage is defined as the ratio of relevant objects to all the objects in the network⁹. Given a diffusion starting from user i , in the first step, it finds the relevant users who have collected the same objects with i . In the second step, those objects which are selected by the relevant users are found and marked as relevant objects. From the Table 3, the diffusion coverage of the backbone is much smaller than that of the original network, which means many irrelevant objects are filtered out. For the popularity-based and the rectangle-based backbone, their diffusion coverage is too small, which affects the diversity of a recommendation algorithm. Taking the UCF in the Douban dataset for example, its diversity decreases from 0.5 to 0.1 (see the results in SI).

	Original	Random	Subgraph	Popularity	Rectangle
Degree heterogeneity					
Douban	25.91	10.88	47.44	12.41	6.75
Flickr	14.58	9.69	49.64	16.55	9.78
Flixster	26.00	12.59	26.65	9.14	3.06
Cluster coefficient					
Douban	5.906e-05	2.046e-05	2.237e-05	1.368-e05	1.746e-05
Flickr	9.588e-04	3.107e-04	4.895e-04	1.045e-04	2.011e-04
Flixster	2.391e-04	6.961e-05	2.387e-04	2.304e-04	3.190e-04
Diffusion Coverage					
Douban	0.8470	0.1364	0.1074	0.0211	0.0097
Flickr	0.7723	0.1438	0.1530	0.0572	0.0291
Flixster	0.9911	0.3256	0.1128	0.0313	0.0077

Table 3. The characteristics of original network and the information backbone.

Discussion

In summary, this paper proposes an enhanced topology-aware method to uncover the information backbone in a user-object bipartite network. The basic assumption of our method is that the more frequently a link is used, the more relevant it is. We firstly find out the subgraph related to neighboring users and objects, and then count the frequency of a link appearing in these subgraphs. Compared to existing methods, our method utilized more information. By investigating the structure of the backbone, it is found that the backbone extracted by our method has a higher cluster coefficient and lower diffusion coverage. It means the objects in the backbone are highly overlapped and a lot of irrelevant objects are filtered out, which could explain why the subgraph-based backbone has the best predictive accuracy. This work may have wide applications in practice. For one thing, the efficiency of a recommendation algorithm can be significantly improved since a lot of links are removed. Although it takes time to get the information backbone, the backbone is quite relatively stable in real systems. Taking the Douban dataset for example, more than 97% links stay the same in two adjacent months. Therefore, it is enough to update the information backbone once a month, which will significantly reduce the computational cost. For another, it can be very helpful to study topological properties or functions of the original network according to the structure of the backbone.

There are still many open issues. On one hand, we concentrate on the topology-aware backbone extraction method without considering the time factor. A recommender system is actually an evolving system, therefore the information backbone should be changing dynamically. To better understand the backbone, one could study the evolution of the original network and the backbone, respectively. On the other hand, we only remove those irrelevant links in the recommender system. If both the irrelevant links and users are removed simultaneously, the efficiency of the recommendation algorithm can be further improved¹⁰.

Methods

Data description. Douban, launched on March 6, 2005, is a Chinese Web 2.0 website providing user reviews and recommendation services of movies, books, and music²³. The raw data contains user activities before Aug 2010 and we randomly sample 11,898 users who have collected at least twenty songs. Flixster is a social movie site allowing users to share movie ratings, discover new movies and meet others with similar movie favor²⁴. The raw data consists of 147,612 users and we randomly sample 10,000 users who have collected at least twenty movies. Flickr is an image hosting and video hosting website, web services suite and online community²⁵. In this paper, data of individuals' membership on that are concerned. Therefore, we recommend groups for users instead of objects^{26,27} and randomly sample 20,641 users who have joined at least twenty groups (see Table 1). As it is quite difficult to provide accurate recommendations for those users with small degree, we filter out the users whose degrees are no larger than 20 in advance.

Evaluating recommender systems. Each data is randomly divided into two parts: the training set (E^T) and the probe set (E^P). The training set contains 80% of the original data and the recommendation algorithm runs on it²⁴. The probe set contains the remaining 20% links and will be used to test the performance of the recommendation algorithm. In our simulation, we run five independently random divisions of E^T and E^P and average the results²⁸.

Precision. The precision of a user i is defined as $P_i(L) = d_i(L)/L$, where L is the length of the recommendation list and $d_i(L)$ is the number of hit links, namely user i 's probe set links contained in the top- L recommendation. The precision of the whole system is defined as $P(L) = \frac{1}{n} \sum_{i=1}^n P_i(L)$. A higher precision value indicates a better recommendation²⁹.

Ranking score. We also consider a global accuracy metric: ranking score(RS)²³. For a target user i , the recommendation algorithm will give her uncollected objects a rank and the ranking score is defined as $RS_{i\alpha} = p/N_p$, where p is the ranking position of object α in the probe set and N_i is the total number of uncollected

objects. The lower the ranking score, the better the recommendation. The mean value of RS is defined as $\langle RS \rangle = \frac{1}{|E^p|} \sum_{i\alpha \in E^p} RS_{i\alpha}$.

Mass diffuse. It works by assigning objects an initial level of “resource” denoted by the vector \vec{f} (where f_α is the resource possessed by object α), and then redistributing it via the transformation $\vec{f}^i = W\vec{f}$, where $W_{\alpha\beta} = \frac{1}{k_\beta} \sum_{j=1}^n \frac{a_{j\alpha} a_{j\beta}}{k_j}$ is a column-normalized $m \times m$ matrix. For a target user, the resulting recommendation list of uncollected objects is sorted according to \vec{f}^i in descending order and top- L objects with the most resources will be recommended.

Hybrid algorithm. When recommending objects for user i , the hybrid method works by assigning one unit of resource to each object which is collected by user i . The initial resources are denoted by the vector \vec{f} where f_α is the resource possessed by object α . Then they will be redistributed via the transformation $\vec{f}^i = W\vec{f}$, where $W_{\alpha\beta} = \frac{1}{k_\alpha^{1-\lambda} k_\beta^\lambda} \sum_{j=1}^n \frac{a_{j\alpha} a_{j\beta}}{k_j}$ is the redistribution matrix, with $k_\alpha = \sum_{l=1}^n a_{l\alpha}$ and $k_j = \sum_{\gamma=1}^m a_{j\gamma}$ denoting the degree of object α and user j , respectively. λ is a tunable parameter which adjusts the relative weight between the mass diffusion algorithm ($\lambda = 1$) and heat conduction algorithm ($\lambda = 0$).

User-based collaborative filtering. In the user-based collaborative filtering method, the basic assumption is that similar users usually have common interests and collect the same objects. Accordingly, the recommendation score of object α for the target user i is $p_{i\alpha} = \sum_{j=1}^n s_{ij} a_{j\alpha}$, where s_{ij} is the cosine similarity between user i and j .

Item-based collaborative filtering. It provides each individual user with objects which are similar to her selected objects. That is, for user i , the recommendation score of object α is $p_{i\alpha} = \sum_{\beta=1}^m a_{i\beta} s_{\alpha\beta}$ where $s_{\alpha\beta}$ is the cosine similarity between objects α and β . Objects will be sorted in descending order according to $p_{i\alpha}$ and the highly ranked objects will be recommended to i .

References

- Adomavicius, G. & Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data. Eng.* **17**, 734–749 (2005).
- Koren, Y., Bell, R. & Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **42**, 30–37 (2009).
- Hofmann, T. Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst.* **22**, 89–115 (2004).
- Zeng, W., Zeng, A., Liu, H., Shang, M. S. & Zhang, Y. C. Similarity from multi-dimensional scaling: Solving the accuracy and diversity dilemma in information filtering. *PLoS One* **9**, e111005 (2014).
- Zhou, T., Ren, J., Medo, M. & Zhang, Y. C. Bipartite network projection and personal recommendation. *Phys. Rev. E* **76**, 046115 (2007).
- Zhang, Y. C. *et al.* Recommendation model based on opinion diffusion. *Europhys. Lett.* **80**, 68003 (2007).
- Zhang, Y. C., Blattner, M. & Yu, Y. K. Heat conduction process on community networks as a recommendation model. *Phys. Rev. Lett.* **99**, 154301 (2007).
- Zhou, T. *et al.* Solving the apparent diversity/accuracy dilemma of recommender systems. *Proc. Natl. Acad. Sci. USA* **107**, 4511–4515 (2010).
- Zeng, W., Zeng, A., Shang, M. S. & Zhang, Y. C. Information filtering in sparse online systems: recommendation via semi-local diffusion. *PLoS One* **8**, e79354 (2013).
- Zeng, W., Zeng, A., Liu, H., Shang, M. S. & Zhou, T. Uncovering the information core in recommender systems. *Sci. Rep.* **4** (2014).
- Zhang, Q. M., Zeng, A. & Shang, M. S. Extracting the information backbone in online system. *PLoS One* **8**, e62624 (2013).
- Kim, D. H., Noh, J. D. & Jeong, H. Scale-free trees: The skeletons of complex networks. *Phys. Rev. E* **70**, 046126 (2004).
- Serrano, M. Á., Boguñá, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **106**, 6483–6488 (2009).
- Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M. & Apkarian, A. V. Scale-free brain functional networks. *Phys. Rev. Lett.* **94**, 018102 (2005).
- Wu, Z., Braunstein, L. A., Havlin, S. & Stanley, H. E. Transport in weighted networks: partition into superhighways and roads. *Phys. Rev. Lett.* **96**, 148702 (2006).
- Grady, D., Thiemann, C. & Brockmann, D. Robust classification of salient links in complex networks. *Nat. Commun.* **3**, 864 (2012).
- Liu, Y., Tang, M., Zhou, T. & Do, Y. Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. *Sci. Rep.* **5**, 9602 (2015).
- Liu, Y., Tang, M., Zhou, T. & Do, Y. Improving the accuracy of the k-shell method by removing redundant links: From a perspective of spreading dynamics. *Sci. Rep.* **5**, 13172 (2015).
- Linden, G., Smith, B. & York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput.* **7**, 76–80 (2003).
- Zhou, Y. B., Lei, T. & Zhou, T. A robust ranking algorithm to spamming. *Europhys. Lett.* **94**, 48002 (2011).
- Lind, P. G., González, M. C. & Herrmann, H. J. Cycles and clustering in bipartite networks. *Phys. Rev. E* **72**, 056127 (2005).
- Huang, Z., Zeng, D. D. & Chen, H. C. Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Manage. Sci.* **53**, 1146–1164 (2007).
- Huang, J. M., Cheng, X. Q., Shen, H. W., Zhou, T. & Jin, X. Exploring social influence via posterior effect of word-of-mouth recommendations. in Proceedings of the fifth ACM international conference on Web search and data mining: WSDM’12, Seattle, USA. New York: ACM Press (February, 2012).
- Jamali, M. & Ester, M. A matrix factorization technique with trust propagation for recommendation in social networks. In Proceedings of the fourth ACM conference on Recommender systems: RecSys’10, Barcelona, Spain. New York: ACM Press (March, 2010).
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P. & Bhattacharjee, B. Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement: IMC’07, San Diego, USA. New York: ACM Press (May, 2007).
- Zeng, W. & Chen, L. Heterogeneous data fusion via matrix factorization for augmenting item, group and friend recommendations. In Proceedings of the 28th Annual ACM Symposium on Applied Computing: SAC’13, Coimbra, Portugal. New York: ACM Press (May, 2013).
- Chen, L., Zeng, W. & Yuan, Q. A unified framework for recommending items, groups and friends in social media environment via mutual resource fusion. *Expert Syst. Appl.* **40**, 2889–2903 (2013).

28. Lü, L. Y. & Liu, W. P. Information filtering via preferential diffusion. *Phys. Rev. E* **83**, 066119 (2011).
29. Herlocker, J. L., Konstan, J. A., Terveen, L. G. & Riedl, J. T. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **22**, 5–53 (2004).

Acknowledgements

We thank An Zeng for helpful advice. This work was supported by the National Natural Science Foundation of China (61370150, 61403062, 61433014, 61502078), Scientific Research start-up Foundation (ZYGX2015KYQD073), the Sichuan Provincial Science and Technology Department (2015GZ0143) and Chongqing Research Program of Basic Research and Frontier Technology (cstc2015jcyjB0244).

Author Contributions

M.F., W.Z. and M.S. designed the research. M.F. performed the experiments. M.F. and W.Z. analysed the data. M.F., M.S., J.S. and W.Z. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zeng, W. *et al.* Uncovering the essential links in online commercial networks. *Sci. Rep.* **6**, 34292; doi: 10.1038/srep34292 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016