Long-Term Effects of Recommendation on the Evolution of Online Systems *

ZHAO Dan-Dan(赵丹丹)[1], ZENG An(曾安)[2**], SHANG Ming-Sheng(尚明生)[1**], GAO Jian(高见)[1]

[1]*School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054*

[2]*Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland*

We employ a bipartite network to describe an online commercial system. Instead of investigating accuracy and diversity in each recommendation, we focus on studying the influence of recommendation on the evolution of the online bipartite network. The analysis is based on two benchmark datasets and several well-known recommendation algorithms. The structure properties investigated include item degree heterogeneity, clustering coefficient and degree correlation. This work highlights the importance of studying the effects and performance of recommendation in long-term evolution.

The digital revolution has brought to us the so-called "information overload": there is too much information for online users to deal with. As a result, nowadays there is hardly an e-commerce website without some form of information filtering or recommendation service. Recommender systems seek to predict users' non-considered preference typically through collaborative or content-based filtering.[1−4] Researchers have developed lots of methods to improve the effectiveness of recommendation, such as matrix factorization,[5] restricted Boltzmann machines,[6] social tags[7] and the ensemble method.[8]

Online commercial systems can be well described by bipartite networks[9] where users and items are represented by nodes, and an edge means that a user has selected an item. In recent years, the network-based recommendation methods became the focus in the literature. For example, the mass diffusion[10] and heat conduction[11] are two personalized recommendation algorithms based on the diffusion process on bipartite networks. A hybrid algorithm of them is shown to effectively solve the accuracy-diversity dilemma.[12] The network manipulation method is also introduced to improve the recommendation performance.[13]

Even studied intensively, most previous works concentrate on evaluating the performance of single recommendation based on training-probe set division. It is not sure that a well-performed recommendation method in a single step can enjoy high performance in the long term. Some methods may make the online network evolve to an unhealthy state. For example, the whole market might be dominated by several super popular items thus users only have limited choices. On the other hand, items may have very even popularity, which makes the quality of the objectives indistinguishable from the degree. Therefore, it is important to study how different recommendation algorithms affect the structure of an online network in long-term evolution.

Fortunately, the tremendous wave of research on complex networks[14−17] in the past decade provides us with a powerful tool to uncover the structure properties and function of the bipartite network. Due to the wide existence of the bipartite network in reality, such as human sexual network[18] and collaboration network,[19] great effort has been made to study its empirical analysis,[20] node strength connectivity correlation,[21] projection into monopartite network[22] and topology metrics.[23] In this Letter, we investigate some topology measurements including the heterogeneity of degree distribution, network clustering and degree correlation. Some well-known recommendation algorithms are considered and the results show that they indeed have different effects on the network evolution.

To begin our analysis, an online commercial system is described by the user-item bipartite network. Specifically, we consider a system of M users and N items represented by a bipartite network with adjacency matrix A, where the element $a_{i\alpha} = 1$ if a user i has collected an item α, and $a_{i\alpha} = 0$ otherwise (throughout this paper we use Greek and Latin letters, respectively, for item- and user-related indices).

In this study, we use data with time information from two online systems: Movielens (http://www.grouplens.org/) and Netflix (http://www.netflixprize.com/). In both systems, users can rate an item (movie) from 1 to 5. A higher rating indicates that the user likes the item more. In order to obtain a binary network, we consider that an edge exists between the two nodes only if a user rates an item higher than 2. The descriptions of the two datasets are given in Table 1.

We will briefly introduce several well-known rec-

---

ommendation algorithms considered in this work: popularity-based recommendation (PR), user-based collaborative filtering (UCF),[1] item-based collaborative filtering (ICF),[1] mass diffusion (MD)[10] and heat conduction (HC).[11] Among these methods, PR is the simplest one which recommends items to each user based on item popularity. The recommendation list is composed of the bestsellers which have no edge with the target user before in descending order. The other methods are personalized and recommend items to each user based on his/her history records.

Table 1. Properties of the two datasets.

Dataset	M	N	Edge number	Sparsity
Movielens	803	4231	119476	3.52×10^{-2}
Netflix	2191	3316	160977	2.22×10^{-2}
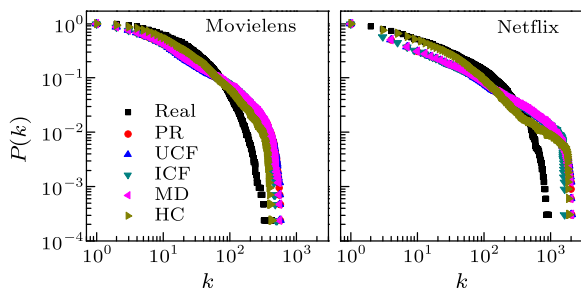
Fig. 1. (Ccolor online) The log-log plots of cumulative item degree distribution ($P(k)$) after evolution in a real and five artificial networks.

The collaborative filtering has become one of the main approaches for recommendation in real online systems and it provides recommendation based on user or item similarities. The UCF algorithm follows this process: First, the similarities between the target user and the rest of the users are calculated. Then the recommendation scores of uncollected items for user i is calculated by

$$P_{i\alpha} = \sum_{j=1}^{m} s(i,j)a_{j\alpha},$$

where $s(i,j)$ denotes the similarity between user i and j. Actually, the measure of similarity is subject to definition. Here we use the Salton similarity measure.[4] Denote Γ_i as the neighbor set of user i and k_i as the degree of i, and the Salton index can be expressed as

$$s(i,j) = \frac{|\Gamma_i \cap \Gamma_j|}{\sqrt{k_i k_j}}.$$

The resulting recommendation list of uncollected items for the target user i is then sorted according to $P_{i\alpha}$ in descending order.

Instead of considering users' similarities, ICF is based on the similarities between items. The recommendation scores can be written as

$$P_{i\alpha} = \sum_{\beta=1}^{n} s(\alpha,\beta)a_{i\beta}.$$

Like UCF, the similarities here between items are calculated based on the Salton index.
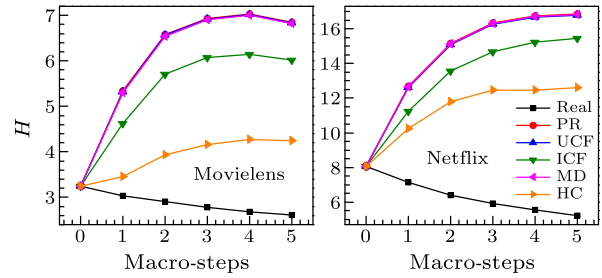
Fig. 2. (Color online) Evolution of item degree heterogeneity (H) when different recommendation methods are implemented in real systems.

The MD and HC algorithms work similarly by assigning one unit of resource to each item selected by the target user, but reallocating the resource through the user-item bipartite network in two different ways. The final resource that item α received from the target user i can be computed by

$$P_{i\alpha} = \sum_{\beta=1}^{n} w_{\alpha\beta}a_{i\beta},$$

where $w_{\alpha\beta}$ is an element of the resource redistribution matrix W. Finally, the recommendation list will be generated by sorting the resource of the items in descending order. The difference between MD and HC lies in the definition of W. In MD, each item distributes its resource equally to its neighboring users. Next, each user redistributes his/her resource equally back to the neighboring items. Mathematically, the redistribution matrix of MD is

$$w_{\alpha\beta} = \frac{1}{k_\beta} \sum_{l=1}^{m} \frac{a_{l\alpha}a_{l\beta}}{k_l}.$$

The HC method also has two steps. In the first step, users receive the average resource from the neighboring items. Then the resource goes back to items in the same way (i.e., each item obtains the average resource of its neighboring users). The resource redistribution matrix reads

$$w_{\alpha\beta} = \frac{1}{k_\alpha} \sum_{l=1}^{m} \frac{a_{l\alpha}a_{l\beta}}{k_l}.$$

In order to investigate how recommendation affects the structure properties of the user-item network, we design the following network evolution scenario. The edges of real data are sorted in order of time. The initial network consists of the first half of the edges. The subsequent edges are equally divided into five subsets according to time. The edges in each subset will be added to the network in one macro-step by the recommendation algorithms. In each macro-step, a recommendation list is generated for each user. The number of edges for each user to be added is determined

by the degree increase of the real data in the corresponding macro-step. For simplicity, we assume that users would accept recommendation by selecting the item with the highest recommendation score, resulting in a deterministic evolution of the network. We perform only one realization of network evolution for each recommendation method. Note that after each macro-step, a new recommendation list for each user will be generated based on the new user-item network. Therefore, the recommendation has a cumulative effect on the network evolution. User degree after each macro-step will be exactly the same as that in real data. However, item degree and connection between users and items will be different from the real case due to the effects of recommendation. As discussed above, five recommendation algorithms will be studied. We will compare the evolution of their resulting artificial networks to the real network (i.e. the case where the edges are added simply based on real data). We do not consider the growth of the system size here since introducing new users or items may involve the cold-start problem.

One of the basic topology properties of a complex network is the degree distribution $p(k)$. Figure 1 shows the cumulative item degree distribution $P(k)$ after five macro-steps of evolution. In both systems, the cumulative item degree distributions of real data follow the stretched exponential distributions.[20] Clearly, all the recommendation algorithms significantly change the degree distributions of the real systems. To quantify the effects of different recommendation on the item degree distribution over time, we calculate the item degree heterogeneity as $H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$.[24] The greater the item degree heterogeneity is, the more the uneven items' degrees are. Actually, the degree heterogeneity is a measure of the global diversity of the online systems. If online retail systems have low global diversity, only several of the most popular items survive and monopolize the markets. On the other hand, with a large global diversity, the popularity is distributed on a large space of items which is generally beneficial compared to a limited space spanned by only the popular items. Figure 2 shows the comparison of the resulting H index from different methods as a function of macro-steps. The curves of PR, UCF and MD overlap with each other and achieve the highest H, indicating that they all tend to recommend popular items. Interestingly, even though UCF and MD are claimed to be personalized, their effects on item degree distribution are more or less the same as the non-personalized PR method. On the other hand, the other two methods recommend items with smaller degree instead. The H from HC is even smaller than that from ICF.

Actually, items' recommendation scores from different recommendation algorithms have been analytically calculated based on the mean-field method.[25] The results show that items' recommendation scores

from MD and UCF are proportional to the degrees of the items. Items' recommendation scores from HC are independent of their degrees. For ICF, the solution of items' recommendation scores strongly depends on the definition of items' similarity measure. Therefore, we observe that the results of MD and UCF are very similar to the PR method. HC leads to relatively homogenous item degree distribution. The results of ICF are between those of PR and HC.

Though the recommendation scores from some algorithms are proportional to items' degrees, they cannot reproduce the features of real networks as the classic preferential attachment does. This is because the recommendation system cannot present each user with the whole recommendation list. Since only the top L ($L = 1$ in our case) items will be shown to the users, the items in the bottom of the list cannot receive any links. This mechanism makes several items (which frequently appear in most users' top L lists) attract most of the links, resulting in a more heterogeneous item degree distribution than the real data.
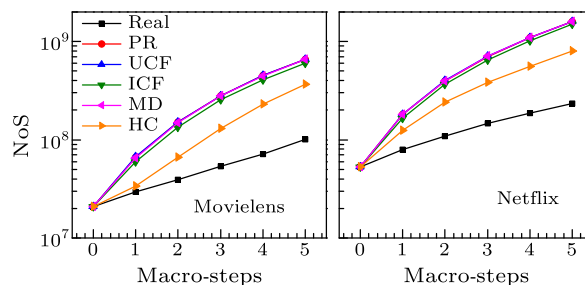

Fig. 3. (Color online) Evolution of number of squares (NoS) when different recommendation methods are implemented in real systems.
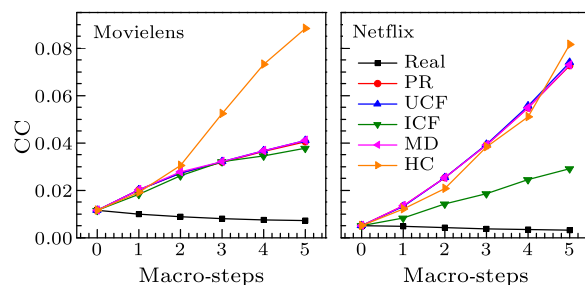
Fig. 4. (Color online) Evolution of clustering coefficient (CC) when different recommendation methods are implemented in real systems.

It is found that the personalized recommendation mainly relies on the square motifs of the bipartite networks.[22] Here we adopt both the number of squares (NoS) and clustering coefficient (CC) to monitor the network evolution. In a bipartite network, the clustering coefficient for a node i is calculated by dividing the number of squares (the smallest clique in the bipartite network) passing through i by the number of i's all possible squares.[23] The value of CC varies from 0 to 1 and a high value indicates a strong local con-

nection. Figure 3 shows the evolution of the NoS in different networks where NoS increases exponentially in the real system and all the recommendation algorithms result in a higher NoS than the real data. It is reasonable since almost all these methods are based on the square motif to realize personal recommendation. Among these algorithms, NoS of HC increases most slowly and other methods are having almost the same effects as PR. As for the clustering coefficient in Fig. 4, most of the methods result in a bigger CC than the real data, just like the case of NoS. It is interesting that even though NoS of HC is the lowest, its CC is the highest among all these artificial networks. The reason is that HC tends to recommend novel items, which significantly strengthens the local connectivity of small degree items. There are many such small degree items, so the final clustering coefficient for the whole system is largely increased.
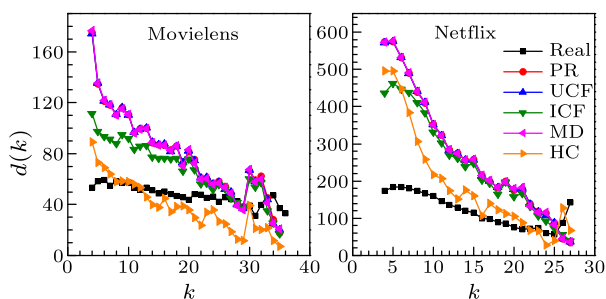
Fig. 5. (Color online) Degree of correlation between users and items after evolution in real and artificial networks. For a given item degree k, its corresponding $d(k)$ is obtained by averaging over all the users whose degrees are in the range of $[a(k^2 - k), a(k^2 + k)]$, where a is chosen as $\frac{1}{2}\log 5$ for a better illustration.

It has been pointed out that the negative degree correlation between users and their selected items is an important feature of an online commercial system.[20] We investigate whether such a feature is significantly influenced by different recommendation methods in the long term. For a given user i, we obtain the average degree of the items selected by him/her as d_i. To see a clear trend, we average d_i over all the users with the same degree k_i (denoted as $d(k)$). Figure 5 shows the relation between $d(k)$ and user degree k in different networks after evolution. Both real systems exhibit negative correlations. Generally, inactive users with low degree have not much experience in exploring new items so they are more likely to select popular items while active users are inclined to try novel items, thus both the networks are disassortative mixing. Almost all the recommendation methods enhance the negative correlation. Actually, this effect mainly comes from the small degree users. With a limited number of edges, they only provide the recommendation algorithms with limited information so recommender systems can only find a few relevant items for them and the large degree items are generally easier

to find (reached by diffusion). Consequently, d's of these small degree users increase after recommendation. Results of HC are the closest to the real case and the curves of PR, UCF, MD overlap just like the H and CC cases.

In summary, we have studied the long-term effects of different recommendation algorithms on the evolution of online commercial systems. We design a bipartite network evolution scenario by assuming that users always accept recommendation (to see the pure effects of recommendation on the topology changes). We study several structure indices including the degree heterogeneity, clustering coefficient and degree correlation. We find that the recommendation algorithms have reinforcing effects on these indices. Generally, the degree heterogeneity and clustering coefficient are increased in the long term. Moreover, the recommendation algorithms result in a stronger negative degree correlation between users and items. Though claimed to be personalized, the MD and UCF algorithms have almost the same effects on network topology as the non-personalized popularity-based method. Among the algorithms we considered, the HC is a special one. Since it is inclined to recommend novel items, its reinforcing effects are the weakest.

Actually, we examined the accuracy of each recommendation method after long-term network evolution. The results show that the recommendation accuracy decreases in the long term if all users entirely rely on the recommendation to select items. We remark that it is interesting to investigate the possible combinations to optimize the network evolution for long-term accuracy, which asks for research in the future.

References

[1] Ekstrand M D et al 2011 *Found. Trends Hum.-Comput. Interact.* **4** 81
[2] Linden G et al 2003 *IEEE Int. Comput.* **7** 76
[3] Adomavicius G and Tuzhilin A 2005 *IEEE Trans. Knowl. Data Eng.* **17** 734
[4] Lü L Y et al 2012 *Phys. Rep.* **519** 1
[5] Koren Y et al 2009 *Computer* **42** 30
[6] Salakhutdinov R et al 2007 *Proc. 24th ICML* p 791
[7] Shang M S and Zhang Z K 2009 *Chin. Phys. Lett.* **26** 118903
[8] Dror G et al 2012 *Proc. IEEE* **100** 2722
[9] Holme P et al 2003 *Phys. Rev.* E **68** 056107
[10] Zhou T et al 2007 *Phys. Rev.* E **76** 046115
[11] Zhang Y C et al 2007 *Phys. Rev. Lett.* **99** 154301
[12] Zhou T et al 2010 *Proc. Natl. Acad. Sci. U.S.A.* **107** 4511
[13] Zhang F and Zeng A 2012 *Europhys. Lett.* **100** 58005
[14] Strogatz S H 2001 *Nature* **410** 268
[15] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[16] Newman M E J 2003 *SIAM Rev.* **45** 167
[17] Boccaletti S et al 2006 *Phys. Rep.* **424** 175
[18] Liljeros F et al 2001 *Nature* **411** 907
[19] Zhang P P et al 2006 *Physica* A **360** 599
[20] Shang M S et al 2010 *Europhys. Lett.* **90** 48006
[21] Shi J J et al 2009 *Chin. Phys. Lett.* **26** 078902
[22] Huang Z et al 2007 *Manage. Sci.* **53** 1146
[23] Lind P G et al 2008 *Phys. Rev.* E **77** 056102
[24] Zhang Q M et al 2013 *PLoS ONE* **8** e62624
[25] Zeng A et al 2012 *Europhys. Lett.* **97** 18005