The code used in the experiments in this paper is available online at:
https://github.com/hongruz/Evolving-RSs/blob/master/EvolvingRSsExperimentsCode.zip
The data/datasets used in the experiments in this paper can be downloaded from the given urls in Reference [25-29].

Evolving Recommender Systems: Modeling, Analysis and Experiments

Hongru Zhu[1], Luoyi Fu[1], Xinbing Wang[1], Songwu Lu[2]
[1]Shanghai Jiao Tong University, China; [2]University of California, Los Angeles, USA
Email: [1]{auberginepp,yiluofu,xwang8}@sjtu.edu.cn ,[2]slu@cs.ucla.edu

ABSTRACT

In many realistic recommender systems (RSs), user ratings on items and social connections among users usually interact in the context of an embedded evolving process where new users and items constantly arrive over time. However how to mathematically model such evolving RSs, along with the corresponding quantitative characterizations, remains unexplored.

Motivated by this, we take the initiative to propose a novel evolving RS model, which, as validated by our empirical results, can well capture some basic features of RSs. Particularly, two types of results are presented in this paper. (i) Our model is primarily based on a weighted bipartite graph structure composed of users and items. With edge weights abstracted from user ratings on their purchased items, an evolving process is proposed to highlight the establishment of social connections for new users and the characterization of their purchases based on rating-driven preferential attachment. (ii) By analytical derivations, along with further empirical validation on eight different real RS datasets, we disclose several RS properties. Specifically, in addition to power-law degree distributions of both users and items, we find that the distribution of total rating scores given by each user or received by each item is bounded by power-law distributions determined by the rating scale in the RS. And interestingly, the beginning of the distribution of total rating scores, instead of being pure power-law, resembles the expected rating probability mass function in certain situations.

Our model discloses how users' rating behaviors, together with RS parameters like rating scales, can alter item markets by affecting item popularity, and can provide insights into the design and analysis of RSs for RS operators and developers.

1 INTRODUCTION

Recommender systems (RSs) [1] are traditionally designed to assist users to find relevant information and offer customized information access for specific domains. In an effective RS where users are provided with relevant and useful items, they are more willing to make purchases or revisit the website, which in turn increases the site traffic and revenue. Regarding such economic benefits, RSs tend to improve their performances and provide more accurate recommendations by trying out some new approaches, including the introduction of "social" features in their recent design. For instance, Amazon allows users to post their purchases on social media like Facebook or Twitter. Netflix shows users the films that their friends have recently watched. Yelp makes use of opinions from users' friends to personalize restaurant rankings. These features bring apparent benefits to websites revenue since friends often affect each other on preferences and purchases. Considering that friends probably share common interests and are likely to trust each other, it is almost certain that users tend to follow recommendations from their friends [2, 3]. Therefore it is essential to explore the structure of the underlying social ties, which contribute to provide more trusted recommendations, to improve RSs.

Regarding this, there have been some initial efforts directed toward performance improvement of RSs through exploitation of social characteristics. For instance, Su et al. [4] study friend recommendation in Twitter, and show that recommendations of popular users are more likely to be accepted than recommendations of "average" users. With both local and global metrics introduced, Wong et. al. [5] analyze the efficiency of social RSs, aiming at simultaneously maximizing individual's benefit and the efficiency of network in information dissemination. Other types of work include investigating the limit of RSs that rely heavily on popularity [6] and more precise recommendation from the perspective of user interests [7]. However, all those works fail to consider a key factor, i.e., the evolution of RSs, a common phenomenon in a flurry of RSs. For example, in purchasing websites such as Amazon, new items are constantly increasing along with new customers arriving and making purchases. The same phenomenon holds in Youtube, where new videos keep being uploaded and new users are registering.

In this paper we try a different approach to emphasize the evolution of RSs. We alternatively view RSs as evolving networks composed of groups of users and items and utilize the network structure to better capture "social" features in the meantime. Yet one difficulty to approach the problem from this perspective is to characterize the interactions both within the user group and between users and items. These two types of interactions corresponds to users' social connections and users' purchases of items in realistic RSs. With the constant arrival of new users and items these interactions will gradually alter the item market, and the user and item popularity, which can be presented by the total rating scores given or received, will also evolve in the rating-driven evolving process, normally resulting in a virtuous cycle:

- A newly arrived user (item), with a high probability, will show preference on popular item (user) whose total rating is high.
- Preference, with a high probability, will result in a new rating to popular item (user), adding up to its total rating and strengthening its popularity.
- As the RS keeps evolving, popular item (user) become more popular while outmoded ones gradually become neglected.

Under such circumstance, the concrete representation of popularity, namely the degree distribution and the distribution of total rating scores manifest their crucial roles in RSs and can be the key factor influencing newly added users and items. Therefore it is essential to a fundamental question in this context, and the one we intend to tackle in this paper is the following: What affects the user and item popularity in an evolving RS, and how? For instance, suppose a group of users rate a set of movies on a movie recommender system. If we change some RS parameters like setting the rating scale from a scale of one to five stars to a scale of one to ten stars, will the movie popularity greatly differs in these two settings?

To understand how popularity is affected in the evolving process, we start by introducing an rating-driven model to capture the evolving and "social" features of RSs. As Figure 1 illustrates, the evolving process in our model incorporates two symmetrical aspects, i.e., the arrival of new users and items. At each time step, a new user arrives, selects an existing user (someone who possibly shares common interests with him and has a high total rating score) as the "prototype" and establishes connection with him. Then the new user chooses among the items purchased by the "prototype" with a probability proportional to the rating that the "prototype" gives to each item. The new user will further rate those selected items according to his own judgement and give ratings from a discrete rating set $\{1, 2, \ldots, R\}$. A symmetrical process also occurs to newly added items.

Based on the proposed model, we derive some general results for our evolving RSs. First we prove that when an RS evolves in the above prescribed way, the degrees of users and items are power-law distributed, and the exponents relate to the ratio of the number of users and items as well as the least number of ratings each has given or received. Then we proceed to conclude that under any assumption of users' rating habits the distribution of total ratings given or received can be upper and lower bounded by power-laws, which, to a great extent, are affected by the rating scale in the RS. Additionally, we also look into the beginning of the distribution of total ratings where total rating scores are small and prove that it is far from being a pure power-law and resembles the expected rating probability mass function in certain situations.

Our main contributions are summarized as follows:

- **Modeling**: We develop an evolving RS model from the weighted bipartite graph structure which well captures basic features of evolving RSs. Our proposed evolving process highlights the establishment of social connections for new users and the characterization of their purchases based on rating-driven preferential attachment. We also exhibit the alternate influence of user and item popularity in the evolving process.
- **Analysis**: We analyze our evolving RS model and show that the degree distribution and the distribution of total rating scores are either power-law or bounded by power-laws determined by RS parameters like the rating scale. We also look into the beginning of the distribution of total rating scores, which resembles the expected rating probability mass function in certain situations.
- **Experiments**: We empirically validate our model through experiments on eight real RS datasets, from which we demonstrate that our model fits well to real data and well captures features of realistic RSs. Our validation further indicates that these findings could provide insights for RS operators and developers into the RS design and analysis from both user and item perspective.

The rest of the paper is organized as follows. In Section 2 we provide the relevant literature. We define our model in Section 3 and mathematically analyze its properties in Section 4. Theoretical results are further verified through experiments in Section 5. We conclude in Section 6. The proofs are either presented in line or available at Appendix.

2 RELATED WORKS

We note that there is no prior work, other than ours, that focuses on the analysis of evolving RSs. However, there is prior work on non-evolutional RS models and the modelling of evolving networks.

Prior studies in non-evolutional RS models have covered serval aspects on the improvement and design of RSs. Some are devoted to the RS performance analysis from algorithmic perspectives [8–10]. Others are directed toward the RS performance evaluation from network aspects [5, 11]. Most recently, several studies also exploit social features in RS design. Vahabi et al. [12] devise a novel RS exploiting the anticipated social-network information diffusion. Bressan et al. [6] introduce a RS model based on popularity and the power of users in influencing others. Zeng et al. [13] clarify the existence of core users carrying most of the information for recommendation in RSs. Liu et al. [7] propose a unified RS to integrate user interests and evolving sequential preference with temporal interval assessment.

There is also a flurry of prior work in evolving network sciences. A series of studies [14–17] have illustrated the structure of evolving social networks with the arrival and departure of users [18, 19] and the temporary dynamics of interest [20]. Some work features the evolving process in the network model. Chung et al. [21] introduce the assumption of either vertex-arrival or edge-arrival at each time step. Ghoshal et al. [22] propose a model that is able to clarify the role of individual elementary mechanisms. Lattanzi et al. [23] introduce the model of affiliation networks, where preferential attachment and edge copy are emphasized in the proposed evolving process.

3 MODEL OF EVOLVING RS

In this section we begin with the model structure and introduce new definitions in the context of our evolving RS model. Then we present our basic assumptions as well as the evolving algorithm. Finally we use a concrete situation to help understand our model intuitively and provide remarks.

3.1 Mathematical Modeling

Model Structure

We use a simple weighted bipartite graph structure $B(U, I)$ to present our model. Vertices in set U represent RS users and vertices in set I represent items in RSs. Intuitively, an edge between user vertex u and item vertex i indicates that user u makes a purchase and gives a rating on item i. The edge has weight $w_{(u,i)}$, which is the rating score given to item i by user u.

New Definition

We introduce a new definition to be used in our proposed evolving RS model $B(U, I)$.

Definition 3.1 (VERTEX WEIGHT). Given an arbitrary vertex v in $B(U, I)$, let $N(v)$ be the set of vertices connected to v. The vertex weight of v is the sum of edge weights on all edges connected to vertex v, namely,

$$W(v) = \sum_{t \in N(v)} w_{(v,t)}$$

Naturally, the vertex weight is a representation of the total rating score a user gives or an item receives in RSs.

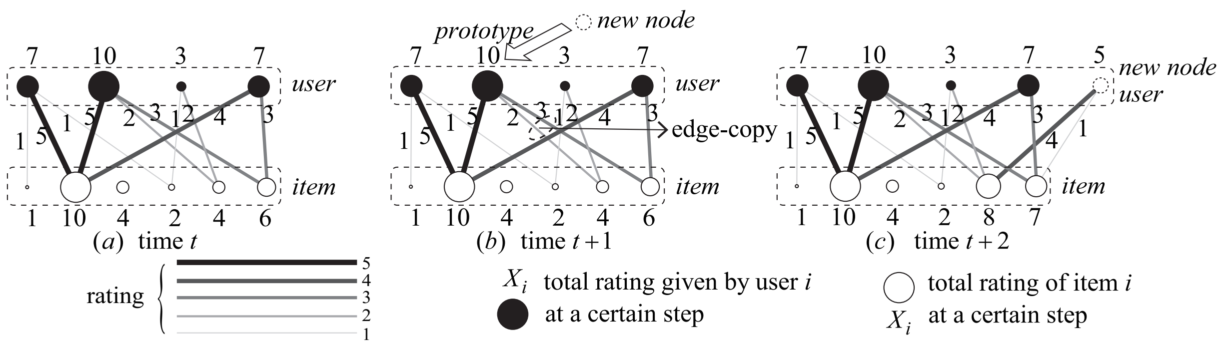


Figure 1: An illustration of how the recommender system evolves in terms of users, items, and ratings.

Basic Assumption

Real users have their own rating habits on different items. We introduce assumptions on how users rate items in our model and propose two frameworks. In the first framework ratings are affected by either the new user or the new item. In the second one ratings are affected by the joint influence of the user and the selected item.

Two frameworks have the followings in common:

(1) All rating scores are sampled from a mixture distribution.

(2) There are K basic user types and each real user u reflects a mixture of those K types with weight vector v_u when rating items.

(3) There are L basic item levels and each real item i reflects a mixture of those L levels with weight vector t_i when being rated.

(4) For all users, their weight vectors are distributed according to some parameterized distribution F (possibly Gaussian) with a parameter vector θ and the probability density function (pdf) is f_θ.

(5) For all items, their weight vectors are distributed according to some parameterized distribution G (possibly Gaussian) with a parameter vector γ and the pdf is g_γ.

(6) The global expectation of a rating score is Er.

In the first framework we assume:

(1) For any basic user type k there is a unique corresponding rating probability mass function (pmf) $h_k(r)$, $r = 1, 2, ..., R$. Symmetrically for any basic item level l, there is a unique corresponding rating pmf $h^l(r)$, $r = 1, 2, ..., R$.

(2) When a newly added user u gives his/her ratings on selected items, his/her rating pmf can be presented as:

$$H_u(r) = \sum_{k=1}^{K} v_u(k)h_k(r).$$

Symmetrically the rating pmf for newly added item i is

$$H^i(r) = \sum_{l=1}^{L} t_i(l)h^l(r).$$

Yet in the second framework we assume:

(1) Given any pairs of the basic user type k and the basic item level l, there is a unique corresponding rating probability mass function (pmf) $h_k^l(r)$, $r = 1, 2, ..., R$.

(2) When a user u gives his/her rating on the item i, his/her rating pmf can be presented as:

$$H_u^i(r) = \sum_{k=1}^{K}\sum_{l=1}^{L} v_u(k)t_i(l)h_k^l(r).$$

Evolving Algorithm

The evolving process in $B(U, I)$ is shown in Algorithm 1.

Algorithm 1: Rating-Driven Evolving Process in $B(U, I)$

Fix two integers $c_u, c_i > 0$, and let $\beta \in (0, 1)$.
Fix an integer R as the highest rating score.
Initialization at $t = 0$**:**
Weighted bipartite graph $B(U, I)$ is a simple graph with at least $c_u c_i$ edges, where each vertex in U has at least c_u edges and each vertex in I has at least c_i edges. Meanwhile, the edge weights are sampled from the assumed mixture distribution in the discrete rating set $\{1, 2, . . . , R\}$ with expectation Er.
At $t > 0$**:**
begin
(Evolution of U**)** With probability β:
begin
(Arrival) A new vertex u is added to U.
(Preferentially chosen Prototype) A vertex $u' \in U$ is chosen as $prototype$ with a probability proportional to its vertex weight, namely the sum of edge weights on all edges connected to it.
(Edge-copy) c_u edges are "$copied$" from u'; that is, c_u neighbors of u', denoted by $n_1, n_2, \ldots, n_{c_u}$ are chosen with a probability proportional to the weight of edges in between (without replacement). Edges $(u, n_1), (u, n_2), \ldots, (u, n_{c_u})$ are added to the graph with weights sampled from the assumed mixture distribution in the discrete rating set $\{1, 2, . . . , R\}$.
end
(Evolution of I**)** With probability $1 - \beta$, a new vertex i is added to I following a symmetrical process, adding c_i edges to i.
end

3.2 An Intuitive Understanding

Our model captures a large spectrum of real RSs. For example, let us consider a movie RS where user and movie groups correspond to set U and I in $B(U, I)$ respectively. When a new user registers, he tends to watch movies that receive high ratings from one of his friends or one popular RS user, who may have many reviews and a trusted taste for movies. That friend can be viewed as "*prototype*" and influence the new user on selection of movies to watch and write reviews on. Similarly when a new movie comes, before it advertises on potential viewers, it may seek in the existing movie group another popular movie similar in theme or cast and treat it as "*prototype*". The new movie is likely to build connections with those who have watched the "*prototype*" and given positive reviews. These two processes alternate and evolve as the RS grows.

Remarks. Our weighted bipartite graph model only produces weights in positive integers yet in real RS, some may allow for the existence of decimal rating scores like 1.5 stars and produces implicit ratings like the play counts for music or CDs. However, this doesn't hurt since we can always map those into explicit positive integer ratings according to preestablished rules. Meanwhile, one particular advantage of our weighted model is the introduction of vertex weights, which serve as concrete indicators of popularity besides vertex degrees. As in conventional unweighted models proposed in prior literature, the evolutionary forms shown by edges and vertices are mainly related to degree distributions. In the context of our model, the vertex weight distribution also affects the evolutionary forms and provides insights into how ratings affect purchases in the future and vice versa. Our model emphasizes the establishment of social connections between users and thus embeds social characteristics in the analysis of evolving RSs.

4 ANALYSIS OF DEGREE AND VERTEX WEIGHT DISTRIBUTIONS

We present here the theoretical results regarding degree and vertex weight distributions in $B(U, I)$. Theorem 4.2 shows power-law degree distributions. Theorems 4.3 and 4.4 show power-law bounds for vertex weight distributions. Proposition 4.7 and Theorem 4.8 analyze the beginning of vertex weight distributions.

4.1 Degree Distribution of Evolving RSs

We start with the degree distribution in our RS model $B(U, I)$ and introduce Lemma 4.1 before we state Theorem 4.2 regarding degree distributions.

Lemma 4.1 ([24]). *If a sequence a_t satisfies the recursive formula $a_{t+1} = (1 - b_t/t)a_t + c_t$ for $t \geq t_0$, where $\lim_{t\to\infty} b_t = b > 0$ and $\lim_{t\to\infty} c_t \geq c$ exists. Then $\lim_{t\to\infty} a_t/t$ exists and equals $c/(1 + b)$.*

With an approach similar to that in [23], We derive Theorem 4.2.

Theorem 4.2. *For the weighted bipartite graph $B(U, I)$ generated after n steps, when $n \to \infty$, the ensemble average of the degree sequence of vertices in U (resp. I) follows a power-law distribution with exponent $\alpha = -2 - \frac{c_u\beta}{c_i(1-\beta)}$ $\left(\alpha = -2 - \frac{c_i(1-\beta)}{c_u\beta}\right)$.*

Proof. Provided in Appendix. A. □

4.2 Vertex Weight Distribution of Evolving RSs

Degree distributions reflect popularity from the perspective of purchase times. Here we investigate another reflection of popularity, i.e., the vertex weight, namely the total ratings given or received. we give the upper and lower bounds of vertex weight distributions in Theorems 4.3 and 4.4 separately and provide illustrations on techniques used to prove these two theorems in Figure 2. We propose Theorem 4.8 regarding the beginning of vertex weight distributions.

4.2.1 *Upper Bound of Vertex Weight Distribution*.

Theorem 4.3. *For the weighted bipartite graph $B(U, I)$ generated after n steps, when $n \to \infty$, the ensemble average of the vertex weight sequence of vertices in U (resp. I) has an upper bound which follows a power-law distribution with exponent $\alpha = -\frac{1+C}{R}$, where C is a constant and equals $Er\left(1 + \frac{c_u\beta}{c_i(1-\beta)}\right)$ $\left(Er\left(1 + \frac{c_i(1-\beta)}{c_u\beta}\right)\right)$ for any vertex with vertex weight greater than $c_uR + 2R$ ($c_iR + 2R$).*

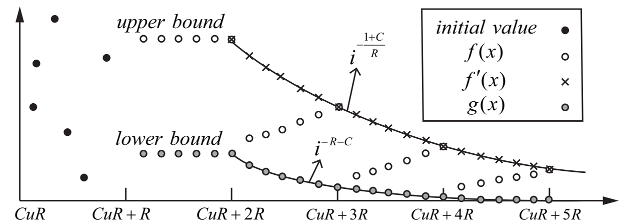


Figure 2: An illustration of proving upper and lower bounds of the vertex weight distribution.

Proof. The proof of Theorem 4.3 includes 3 sequential parts:
I. Derivation of the recurrence relation of vertex weight sequence
II. Derivation of the intermediate upper bound
III. Derivation of the power-law upper bound

PART I. ***Derivation of the Recurrence Relation***

Let V_t^k be the expected number of vertices in U with vertex weight k at time t. When $k > c_uR + R$ we have

$$\begin{aligned}V_t^k =&V_{t-1}^k - E[\#\text{ of vertices in } U \text{ with vertex weight } k\\ &\text{at time } t-1 \text{ and increase at time } t]\\ &+ E[\#\text{ of vertices in } U \text{ with vertex weight} < k\\ &\text{at time } t-1 \text{ and increase to } k \text{ at time } t].\end{aligned}$$

The vertex weight of vertices in U can increase if and only if a new vertex is added to I. Thus we have

$$\begin{aligned}V_t^k =&V_{t-1}^k - (1-\beta)c_i\frac{kV_{t-1}^k}{W_{t-1} + W_{B_0}}\\ &+ (1-\beta)\sum_{j=k-R}^{k-1} E[\#\text{ of vertices in } U \text{ with vertex weight}\\ &j \text{ at time } t-1 \text{ and increase to } k \text{ at time } t|\\ &\text{a new vertex is added to } I].\end{aligned}$$

Similar to the proof of Theorem 4.2 the probability that an edge is selected in a single selection is proportional to its weight. Thus,

$$\begin{aligned}&E[\#\text{ of vertices with vertex weight } j \text{ at time } t-1 \text{ that is chosen}\\ &\text{as end point}|\text{ a new vertex is added to } I] = \frac{jc_iV_{t-1}^j}{W_{t-1} + W_{B_0}}.\end{aligned}$$

Since we have

$$\begin{aligned}&E[\#\text{ of vertices in } U \text{ with vertex weight } j \text{ at time } t-1\\ &\quad\text{and increase to } k \text{ at time } t|\text{ a vertex is added to } I]\\ =&E[\#\text{ of vertices in } V_{t-1}^j \text{ that is chosen as end points}|\\ &\quad\text{a new vertex is added to } I]\\ &\times E[Pr[\text{the edge weight is assigned to be } k-j]],\end{aligned}$$

We call $H(r) = E[Pr[\text{the edge weight is assigned to be } r]]$ to be the expected rating pmf for new users. Using our assumptions in Section 3, we have the following results under the two different frameworks. In the first framework, we have:

$$\begin{aligned}H(r) &= E[Pr[\text{the edge weight is assigned to be } r]]\\ &= \int_\gamma g_\gamma H_\gamma(r)d\gamma\\ &= \int_\gamma g_\gamma \sum_{l=1}^{L} t_\gamma(l)h^l(r)d\gamma.\end{aligned}$$

Or in the second framework, we have:

$$\begin{aligned}H(r) &= E[Pr[\text{the edge weight is assigned to be } r]]\\ &= \int_\theta \int_\gamma f_\theta g_\gamma H_{\theta,\gamma}(r) d\gamma d\theta\\ &= \int_\theta \int_\gamma f_\theta g_\gamma \sum_{k=1}^{K} \sum_{l=1}^{L} v_\theta(k) t_\gamma(l) h_k^l(r) d\gamma d\theta.\end{aligned}$$

Since we know that

$$\sum_{r=1}^{R} h_k^l(r) = 1, \sum_{r=1}^{R} h^l(r) = 1,$$

we always have

$$\sum_{r=1}^{R} H(r) = 1.$$

we can derive

$$V_t^k = V_{t-1}^k \left(1 - \frac{(1-\beta)c_i k}{W_{t-1} + W_{B_0}}\right) + (1-\beta) \sum_{j=k-R}^{k-1} \frac{c_i j V_{t-1}^j H(k-j)}{W_{t-1} + W_{B_0}}.$$

Let $X_k = \lim_{t\to\infty} V_t^k / t$. Again, using Lemma 4.1, we get the following recurrence relation

$$X_k = \frac{(1-\beta)\frac{c_i}{Er(c_u\beta + c_i(1-\beta))}}{1 + (1-\beta)\frac{c_i k}{Er(c_u\beta + c_i(1-\beta))}} \sum_{j=k-R}^{k-1} H(k-j) j X_j,$$

namely,

$$X_k = \frac{1}{k+C} \sum_{j=k-R}^{k-1} H(k-j) j X_j.$$

PART II. ***Derivation of the Intermediate Upper Bound***

This high order homogeneous recurrence relation is not detachable as in the proof of Theorem 4.2. Without loss of generality, we assume an arbitrary set of initial values which are not all zeros, as shown in Figure 2 and suppose f is one intermediate upper bound. We have the following 2 cases:

When $c_u R + R + 1 \le k \le c_u R + 2R$, let

$$f(k) = \max_{cuR+R+1 \le j \le cuR+2R} X_j.$$

When $k > c_u R + 2R$, $H(r)$ sums to 1 and we have

$$\begin{aligned}X_k &= \frac{1}{k+C} (H(R)(k-R)X_{k-R} + \ldots + H(1)(k-1)X_{k-1})\\ &\le \frac{1}{k+C} \times \max_{k-R \le j \le k-1} f(j)\\ &\quad \times (H(R)(k-R) + H(R-1)(k-R+1) + \ldots + H(1)(k-1))\\ &\le \frac{k-1}{k+C} \max_{k-R \le j \le k-1} f(j),\end{aligned}$$

namely,

$$f(k) = \frac{k-1}{k+C} \max_{k-R \le j \le k-1} f(j). \tag{1}$$

Before we use induction to find the upper bound, we first deal with the initial case when k is in the range $[c_u R + 2R + 1, c_u R + 3R]$.

Recall that when $c_u R + R + 1 \le k \le c_u R + 2R$, $f(k)$ is fixed and equals $f(c_u R + 2R)$. Thus for $c_u R + 2R + 1 \le k \le c_u R + 3R$, since $\frac{k-1}{k+C} < 1$, using Equation (1) we have

$$\max_{c_u R+2R+1 \le j \le c_u R+3R} f(j) < f(c_u R + 2R),$$

which is demonstrated in Figure 2.

Moreover, since $\frac{k-1}{k+C}$ is strictly increasing, we also have

$$\max_{c_u R+2R+1 \le j \le c_u R+3R} f(j) = f(c_u R + 3R).$$

Ultimately, we get the following two equations for the initial case when $c_u R + 2R + 1 \le k \le c_u R + 3R$:

$$f(k) = \frac{k-1}{k+C} f(c_u R + 2R), \quad \max_{c_u R+2R+1 \le j \le c_u R+3R} f(j) = f(c_u R + 3R).$$

By induction, suppose

if $n = K$, when $c_u R + nR + 1 \le k \le c_u R + (n+1)R$, we have

$$f(k) = \frac{k-1}{k+C} f(c_u R + nR), \text{ and}$$

$$\max_{c_u R+nR+1 \le j \le c_u R+(n+1)R} f(j) = f(c_u R + (n+1)R).$$

Next we show the conditions above also hold when $n = K + 1$.

We know from Equation (1) that

$$f(k) = \frac{k-1}{k+C} \max_{k-R \le j \le k-1} f(j) < \max_{k-R \le j \le k-1} f(j).$$

Thus we can derive the following inequality

$$\max_{c_u R+(K+1)R+1 \le j \le c_u R+(K+2)R} f(j) < f(c_u R + (K+1)R).$$

Hence we have

$$\begin{aligned}f(k) &= \frac{k-1}{k+C} f(c_u R + (K+1)R)\\ &= \frac{k-1}{k+C} f(c_u R + nR)\\ &< f(c_u R + nR).\end{aligned}$$

Since $\frac{k-1}{k+C}$ is increasing as k increases, $f(k)$ is increasing in the given range for k and

$$\max_{c_u R+nR+1 \le j \le c_u R+(n+1)R} f(j) = f(c_u R + (n+1)R).$$

Now we can conclude that for any positive integer $n \ge 2$, when $c_u R + nR + 1 \le k \le c_u R + (n+1)R$, we have

$$f(k) = \frac{k-1}{k+C} f(c_u R + nR).$$

PART III. ***Derivation of the Power-law Upper Bound***

We can derive the following relation

$$f(k) = \frac{k-1}{k+C} f(k-R),$$
$$\text{for every } k = c_u R + 3R, c_u R + 4R, \ldots$$

Recursively, the above equation results in

$$\begin{aligned}f(k) &= \frac{k-1}{k+C} \cdot \frac{k-R-1}{k-R+C} \cdot \frac{k-2R-1}{k-2R+C} \cdot \ldots \cdot \frac{c_u R + 2R - 1}{c_u R + 2R + C}\\ &\quad \cdot f(c_u R + 2R)\\ &= \frac{\frac{k-1}{R}}{\frac{k+C}{R}} \cdot \frac{\frac{k-R-1}{R}}{\frac{k-R+C}{R}} \cdot \frac{\frac{k-2R-1}{R}}{\frac{k-2R+C}{R}} \cdot \ldots \cdot \frac{\frac{c_u R+2R-1}{R}}{\frac{c_u R+2R+C}{R}} f(c_u R + 2R)\\ &= \frac{\Gamma(\frac{k-1}{R}+1)}{\Gamma(\frac{k+C}{R}+1)} \frac{\Gamma\left(\frac{c_u R+2R+C}{R}+1\right)}{\Gamma\left(\frac{c_u R+2R-1}{R}+1\right)} f(c_u R + 2R)\\ &\sim \left(\frac{k}{R}\right)^{-\frac{1+C}{R}},\end{aligned}$$

for every $k = c_u R + 3R, c_u R + 4R, \ldots$.

As illustrated in Figure 2, we want to find a new upper bound f'. To do this, we first define the initial case

$$f'(c_u R + 2R) = f(c_u R + 2R),$$

and then instead of $k = c_u R + 3R, c_u R + 4R, \ldots$, suppose for any positive integer $k > c_u R + 2R$,

$$f'(k) = \frac{\Gamma(\frac{k-1}{R}+1)}{\Gamma(\frac{k+C}{R}+1)} \frac{\Gamma(\frac{c_u R+R+C}{R}+1)}{\Gamma(\frac{c_u R+R-1}{R}+1)} f'(c_u R + 2R).$$

$f(k)$ is increasing in interval $[c_u R + nR + 1, c_u R + (n+1)R]$ for any positive integer n while $f'(k)$ is always decreasing and $f(c_u R + nR) = f'(c_u R + nR)$. As shown in Figure 2, $f'(k)$ is also an upper bound of X_k and it follows a power-law distribution with exponent $\alpha = -\frac{1+C}{R}$.

Thus we have proved that the upper bound of the ensemble average of the vertex weight distribution in set U follows a power-law distribution with exponent $\alpha = -\frac{1+C}{R}$. Symmetrically, we can prove a similar result for the vertices in I. □

4.2.2 ***Lower Bound of Vertex Weight Distribution***.

Theorem 4.4. *For the weighted bipartite graph $B(U, I)$ generated after n steps, when $n \to \infty$, the ensemble average of the vertex weight sequence of vertices in U (resp. I) has a lower bound which follows a power-law distribution with exponent $\alpha = -R - C$, where C is a constant and equals $Er(1+\frac{c_u\beta}{c_i(1-\beta)}) \left(Er\left(1+\frac{c_i(1-\beta)}{c_u\beta}\right)\right)$ for any vertex weight greater than $c_u R + 2R$ ($c_i R + 2R$).*

Proof. Provided in Appendix. B. □

Theorem 4.2 is consistent with Theorems 4.3 and 4.4 when the degree distribution is viewed as a special case of the vertex weight distribution where all edge weights are assigned to 1 in the weighted graph.

When $R = 1$ it is easy to verify that the exponents of both upper and lower bound power-law distributions are $-1 - C$, which is exactly the same as the exponent of the degree distribution we derive in Theorem 4.2. And note that the following inequality always holds for positive integer R.

$$-R - C \le -\frac{1+C}{R}.$$

4.2.3 ***Beginning of Vertex Weight Distribution***.

In Sections 4.2.1 and 4.2.2, we show the vertex weight distribution is bounded by power-laws when the vertex weight k is greater than a certain value. Here we proceed to explore the beginning of the vertex weight distribution when k is relatively small. Defined as the above, V_t^k is the expected number of vertices in U with vertex weight k at time t. Again let $X_k = \lim_{t\to\infty} V_t^k/t$. We define a new random variable S to be the vertex weight of a newly added vertex in U with a pmf $s(k)$. Symmetrically we define another random variable S' to be the vertex weight of a newly added vertex in I with a pmf $s'(k)$. Similar to the proof of Theorems 4.3 and 4.4, we define $H(r)$ to be the expected rating pmf for new users and $H(r)'$ to be the expected rating pmf for new items.

We introduce the following 3 propositions before providing Theorem 4.8. The first two propositions give the recurrence relation and an upper bound of X_k. The third one describes the distributions of random variables S and S'. Due to space limitations, proofs of these 3 propositions are provided respectively in Appendix. C, D, E.

Proposition 4.5. *For the ensemble average of the vertex weight distribution in set U (resp. I), when $c_u < k < c_u R$ ($c_i < k < c_i R$), the recurrence relation of X_k is*

$$X_k = \frac{1}{k+C} \sum_{j=k-R}^{k-1} H(k-j) j X_j + \frac{C\beta s(k)}{k+C}$$

$$\left(X_k = \frac{1}{k+C} \sum_{j=k-R}^{k-1} H'(k-j) j X_j + \frac{C(1-\beta)s'(k)}{k+C}\right),$$

where $C = Er\left(1+\frac{c_u\beta}{c_i(1-\beta)}\right) \left(Er\left(1+\frac{c_i(1-\beta)}{c_u\beta}\right)\right)$.

Proposition 4.6. *For $c_u \le k \le c_u R$ (resp. in I, $c_i \le k \le c_i R$), X_k have an upper bound $l(k) \le 1$.*

Proposition 4.7. *When c_u (resp. c_i) is large enough, S (S') follows a unimodal probability distribution.*

Now we introduce Theorem 4.8 regarding the difference of vertex weight sequence and the difference of the nonhomogeneous term in the recurrence relation.

Theorem 4.8. *Let $p(k)$ be the nonhomogeneous term in recurrence relation in set U (resp. I),*

$$p(k) = \frac{C\beta s(k)}{k+C} \quad \left(p(k) = \frac{C(1-\beta)s'(k)}{k+C}\right)$$

. For $c_u < k < c_u R$ ($c_i < k < c_i R$), asymptotically

$$|\Delta X_k - \Delta p(k)| \le H_u + \frac{1}{c_u + C}$$

$$\left(|\Delta X_k - \Delta p(k)| \le H_i + \frac{1}{c_i + C}\right)$$

where

$$C = Er\left(1 + \frac{c_u\beta}{c_i(1-\beta)}\right)\left(Er\left(1+\frac{c_i(1-\beta)}{c_u\beta}\right)\right),$$

and

$$H_u = \max\{H(1), H(R)\} + \sum_{r=1}^{R-1} |H(r+1) - H(r)|,$$

$$H_i = \max\{H'(1), H'(R)\} + \sum_{r=1}^{R-1} |H'(r+1) - H'(r)|.$$

Proof. Provided in Appendix. F. □

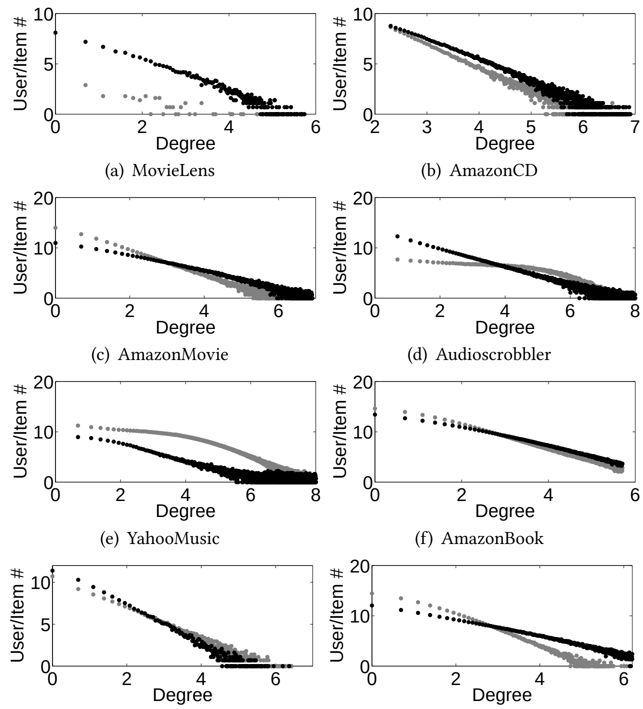
Here we discuss more about function $p()$ in set I, which is closely related to the beginning of the vertex weight distribution of items. For $p(k) = \frac{C(1-\beta)s'(k)}{k+C}$, it is a product of $s'(k)$ and another power function $q(k) = \frac{C(1-\beta)}{k+C}$ with exponent equal to -1. We know that for the derivative of a power function $q(k)$ with exponent -1 is

$$\frac{dq(k)}{dk} < 0, \frac{dq(k)}{dk} \to 0, k \to \infty.$$

When c_i is large enough, $s'(k)$ is a unimodal pmf. Since the absolute derivative of $q(k)$ is rather small and leads to steady $q(k)$, with a high probability, the product $p(k)$ increases before some value and then decreases toward 0 as k increases. By Proposition 4.7 and Theorem 4.8, we can roughly depict the vertex weight distribution

Table 1: Dataset statistics and fitting parameters

Dataset	Rating scale	Number of users	Number of items	Number of ratings	α (item degree distribution)	Parameter C
MovieLens	[1,5]	700	10,000	100,000	-2.397	4.191
AmazonMovie	[1,5]	1,884,911	198,797	4,607,047	-2.018	3.054
AmazonCD	[1,5]	1,460,632	470,130	3,749,004	-2.184	3.552
YahooMusic	[1,100]	1,948,882	98,211	11,557,943	-2.015	51.256
Audioscrobbler	[1,5]	146,946	1,493,930	24,296,858	-2.003	3.009
AmazonBook	[1,5]	5,518,811	2,078,816	22,507,155	-2.287	3.861
BookCrossing	[1,10]	278,858	271,379	1,149,781	-2.335	7.343
AmazonElectronics	[1,5]	3,431,122	464,673	7,824,482	-2.019	3.057



(a) MovieLens (b) AmazonCD (c) AmazonMovie (d) Audioscrobbler (e) YahooMusic (f) AmazonBook (g) BookCrossing (h) AmazonElectronics

Figure 3: Degree distribution of users (grey dot) and items (black dot) in logarithm

of items for $c_i \leq k \leq c_i R$ and it is very probable that $p(k)$ and X_k are both unimodal function and the vertex weight distribution will demonstrate a peak at the beginning of the sequence.

When c_i is small, for instance in the extreme case when c_i equals 1, pmf $s'(k)$ is the same as the expected rating pmf for new items $H'(r)$. Thus the shape of $p(k)$ and X_k will largely depend on the shape of $H'(r)$. $p(k)$ will fluctuate when $H'(r)$ has fluctuations since a power function tends to be steady and has small absolute derivatives when $k > 1$. Therefore the shape of $p(k)$ and X_k, or rather the vertex weight distribution will resemble $H'(r)$.

Symmetrically we have similar results for the analysis in set U.

5 EXPERIMENTS

In this section, we discuss several experiments on real RS datasets. We provide experiment settings in Section 5.1 and performance evaluations in Section 5.2.

Theorem 4.2 says that the degrees in the groups of users and items in RSs are power-law distributed. A first set of experiments shows the degree distributions from different datasets. We then try to fit them with our theoretical results and illustrate phenomena observed in real datasets. Then as we move on to validate Theorems 4.3, 4.4 and 4.8, we carry out experiments focusing on the vertex weight of items from those datasets. To answer the question we raise in Section 1, what affects the popularity and how, we present how different datasets with different rating scales have distinct vertex weight distributions. Finally, we focus on the begin of vertex weight distributions from particular datasets and seek reasonable illustrations.

5.1 Experiment Settings

As shown in Figure 2, we use 8 publicly available RS datasets: the MovieLens movie recommendation datasets [25], the Audioscrobbler music artist rating datasets [26], the Yahoo music artist rating datasets [27], the BookCrossing book recommendation datasets [28], and the Amazon product recommendation datasets for movies, CDs, electronics and books [29]. These datasets vary in size, types of products and ratio of user and item numbers, testing the applicability of our model in different situations.

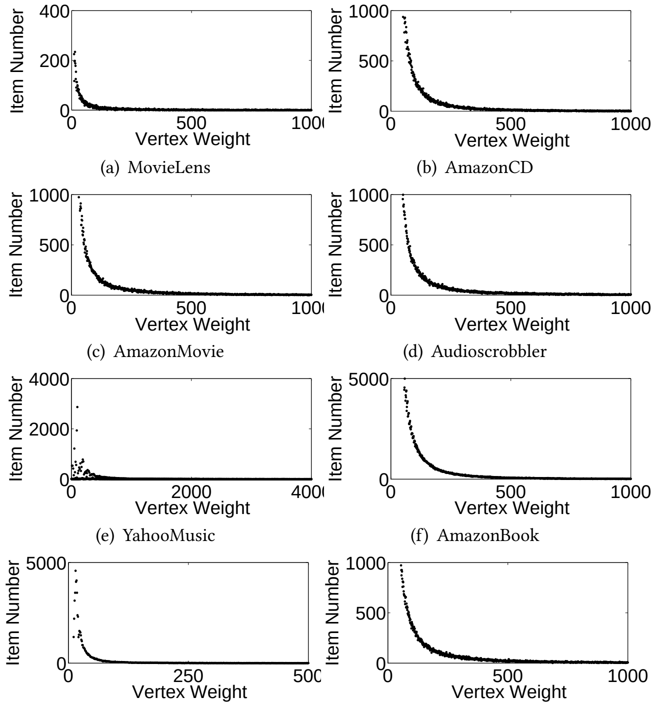
When we conduct experiments on those datasets that have implicit ratings, we set up a rule to generate corresponding integer rating scores. For instance, in Audioscrobbler dataset, we map users' play count for musical artist to explicit integer ratings in interval [1, 5]. Specifically we set the mapping rule as follows. If the user listens only once to an artist, then the rating is 1. When the play count is in range [2,4],[5,9] or [10,19], the rating is set to 2,3 or 4 separately. A rating of 5 will be given if and only if the user listens to the artist no less that 20 times. After converting implicit and decimal ratings to explicit integer ratings, we record the dataset statistics in Table 1.

5.2 Performance Evaluations

5.2.1 ***Degree Distributions of Items and Users.***

The first experiments show degree distributions from those datasets are power-laws. Figure 3 exhibits the degree distributions of users and items in 8 datasets in logarithm. These results fit to the symmetrical features of degree distributions of users and items in our bipartite graph model.

Clearly, item degree distributions have unambiguous fits of power-laws in some datasets while in others like the Yahoo dataset the distribution seems divergent with high degrees. This phenomenon might result from a bias in favor of popular items receiving much attention in the dataset. We fit item degree distributions in 8 datasets and avoid using divergent parts with high degrees. As shown in Table 1, the exponents α of item degree distributions are ciphered out and they appear to be lower than -2, which is



(a) MovieLens (b) AmazonCD (c) AmazonMovie (d) Audioscrobbler (e) YahooMusic (f) AmazonBook (g) BookCrossing (h) AmazonElectronics

Figure 4: Vertex weight distribution of items

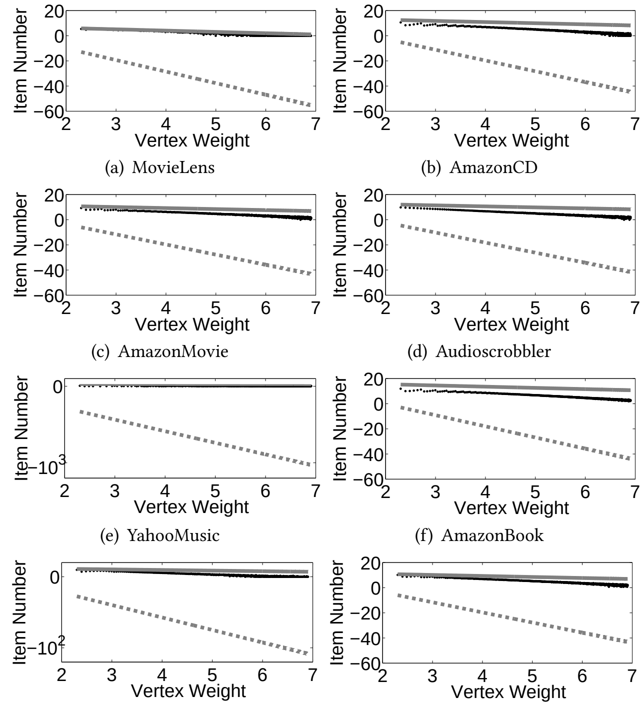
consistent with Theorem 4.2 that the exponents should be

$$\alpha = -2 - \frac{c_i(1-\beta)}{c_u\beta}. \tag{2}$$

Moreover, using Equation (3) we calculate parameter C to be used in deriving the exponents of upper and lower bounds of item vertex weight distributions in Theorems 4.3 and 4.4. and provide them in Table 1.

$$C = Er\left(1 + \frac{c_i(1-\beta)}{c_u\beta}\right). \tag{3}$$

Figure 3 also shows a noticeable common feature for 2 music artist recommendation datasets, the Audioscrobbler dataset and the Yahoo dataset. User degree distributions in these two datasets, i.e. the grey plots in Figure 3(d) and 3(e) seem to be piecewise. When degrees are small, the fitted power-laws have small absolute exponents while the absolute exponents are larger when degrees are large. One explanation consistent with our model is that users in these RSs can be roughly separated into 2 groups, senior users and junior users. Senior users have used the RS for long and given many ratings, and junior users, on the contrary, do not often use the RS or have recently arrived. In the context of music preference and recommendation, these two types of users have more distinct differences. Senior users might be more socially interactive and have particular taste for music while junior users are not. Senior users with large vertex weights are more probable to become "prototype" and interact with others like sharing their preference list. As time increases, behavioral patterns of some junior users are altered and they start to build up social influences. These features provide insights into how the evolving process with social interactions can influence user degree distributions.



(a) MovieLens (b) AmazonCD (c) AmazonMovie (d) Audioscrobbler (e) YahooMusic (f) AmazonBook (g) BookCrossing (h) AmazonElectronics

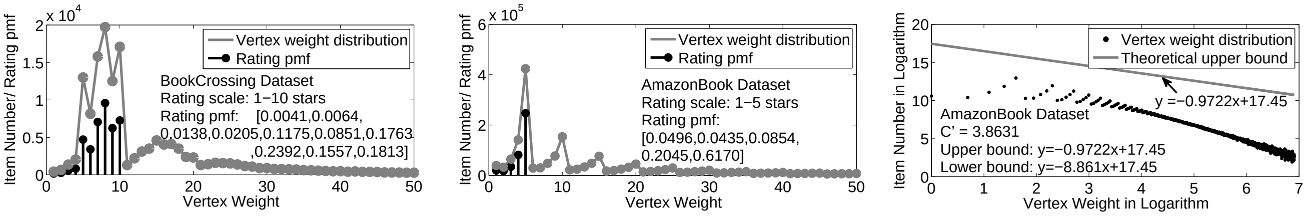
Figure 5: Upper (solid grey) and lower bound (dash grey) of vertex weight distribution (black) of items in logarithm

5.2.2 *Vertex Weight Distribution of Items.*

The second set of experiments are to validate our theoretical results regarding vertex weight distributions. In Figure 4, we plot item vertex weight distributions and in Figure 5 we plot them with their theoretical upper and lower bounds in logarithm. These bounds are calculated using parameter C in Table 1.

As shown in Figure 5, item vertex weight distributions are well bounded, implying that the distribution follows either a single power-law or a sum of power-laws with bounded exponents. We know from Theorems 4.3 and 4.4 that the theoretical bounds are affected by rating scales or rather the highest possible rating score R, which could account for the differences in the performance of vertex weight distributions in different datasets. Given that Item vertex weight distributions are closer to upper bounds, we can see how rating scales may influence item vertex weight distributions by affecting the exponents of their power-law upper bounds.

Subsequently we look into the beginning of item vertex weight distributions. No plot in Figure 4 exhibits a peak at the beginning of the sequence, indicating a small c_i for selected datasets according to Proposition 4.7 and Theorem 4.8. With small c_i, item vertex weight distributions tend to have similar fluctuations to those in expected rating pmf for items $H'(r)$. We show item vertex weight distributions in book RSs from BookCrossing and AmazonBook in Figure 6. Figure 6(a) and 6(b) show that the beginning of item vertex weight distributions resemble $H'(r)$, suggesting that average users' rating habits can affect the beginning of item vertex weight distributions. In an evolutional view, small c_i leads to a large number of items having small vertex weights. New items are added with



(a) BookCrossing (b) AmazonBook (c) AmazonBook in logarithm

Figure 6: Vertex weight distribution for book recommendation datasets.

small vertex weights and are not competitive with other popular items. Users choose items with probabilities proportional to edge weights, indicating that popular items with large vertex weights are more likely to be selected by users and become more popular.

Surprisingly we find that in Figure 6(b), 5 different power-law curves seem to coexist and the logarithm plot from the same dataset in Figure 6(c) also supports the idea. With a detailed examination on the dataset, we present the rating pmf for different scores in Figure 6 and find that those users prefer to give a rating score of 5. Recall the recurrence relation in proofs of Theorems 4.3 and 4.4 is

$$X_k = \frac{1}{k+C} \sum_{j=k-R}^{k-1} H'(k-j) j X_j.$$

If $H'(5)$ is much larger, the recurrence relation yields to

$$X_k \approx \frac{H'(R)(k-R)X_{k-R}}{k+C},$$

indicating that X_k is dependent mostly on X_{k-5} and this results in 5 different power-law curves starting from 5 distinct initial values.

6 CONCLUSION

In this paper, we propose a novel evolving RS model that highlights the establishment of social connections for new users and the characterization of their purchases based on rating-driven preferential attachment. The superiority of our model lies in three aspects: good capture of realistic RSs, mathematically tractability and novelty in the context of evolving RSs with social characteristics. Subsequently we analyze our model and prove some basic properties including power-law degree distributions and the vertex weight distributions bounded by power-laws. We also investigate the beginning of the item vertex weight distribution, which resembles the expected rating pmf for new items. Our model is finally validated through experiments on 8 real RS datasets, from which we verify our theoretical results and demonstrate that our evolving model can well capture realistic RSs.

REFERENCES

[1] P. Resnick and H. Varian. "Recommender Systems," in *Communications of the ACM*, Vol. 40, No. 3, pp. 56–58, Mar., 1997.

[2] D. Kempe, J. M. Kleinberg and É Tardos, "Maximizing the Spread of Influence through a Social Network," in *Proc. of KDD*, 2003, pp. 137–146.

[3] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, W. Wei, Y. Wang and Y. Yuan, "Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate," in *Prof. of SDM*, Mesa, Arizona, USA, April, 2011.

[4] J. Su, A. Sharma and S. Goel. "The Effect of Recommendations on Network Structure," in *Proc. of WWW. 2016.*

[5] F. M. F. Wong, Z. Liu and M. Chiang, "On the Efficiency of Social Recommender Networks," in *IEEE/ACM Trans. on Networking*, Vol. 24, No. 4, pp. 2512–2524, Aug., 2016.

[6] M. Bressan, S. Leucci, A. Panconesi, P. Raghavan and E. Terolli, "The limits of popularity-based recommendations, and the role of social ties," in *Prof. of ACM KDD*, San Francisco, CA, USA, Aug., 2016.

[7] Y. Liu, C. Liu, B. Liu, M. Qu and H. Xiong, "Unified point-of-interest recommendation with temporal interval assessment," in *Proc. of ACM KDD*, San Francisco, CA, USA, Aug., 2016.

[8] B. J. Mirza, B. J. Keller and N. Ramakrishnan, "Studying Recommendation Algorithms by Graph Analysis," in *J. of Intelligent Information Systems* (2003) Vol. 20, No. 2, pp. 131–160, 2003.

[9] S. Chojnacki and M. A. Klopotek. "Random Graphs for Performance Evaluation of Recommender Systems," in *Journal of Control and Cybernetics*, Vol. 40, No. 2, 2011,

[10] Zhao, J., Zhang, H. and Y. Lian, "Analysis and Design of Personalized Recommender System Based on Collaborative Filtering," in *J. Internet of Things*, Vol. 312 of the Series Communications in Computer and Information Science, pp. 473-480, 2012.

[11] D. Lamprecht, M. Strohmaier and D. Helic, "Improving Reachability and Navigability in Recommender Systems," in *arXiv preprint arXiv:1507.08120*, 2015.

[12] H. Vahabi, I. Koutsopoulos, F. Gullo and M. Halkidi, "Difrec: A social-diffusion-aware recommender system," in *Proc. of CIKM*, pp. 1481-1490, Melbourne, VIC, Australia, Oct., 2015.

[13] W. Zeng, A. Zeng, H. Liu, M. Shang and T. Zhou, "Uncovering the Information Core in Recommender Systems," in Scientific Reports, 4:6140, Aug., 2014.

[14] M. Atzmueller, A. Ernst, F. Krebs, C. Scholz and G. Stumme, "On the Evolution of Social Groups During Coffee Breaks," in *Proc. of ACM WWW*, pp. 631–636, 2014.

[15] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar and D. Song, "Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications using Google+," in *Proc. of ACM IMC*, pp. 131-144, 2012.

[16] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel and S. Geirhofer, "Network Densification: The Dominant Theme for Wireless Evolution into 5G," in *IEEE Communications Magazine*, Vol. 52, No. 2, pp. 82–89, 2014.

[17] Y. Wu, N. Pitipornvivat, J. Zhao, S. Yang, G. Huang and H. Qu, "egoSlider: Visual Analysis of Egocentric Network Evolution," in *IEEE Trans. on Visualization and Computer Graphics*, Vol. 22, No. 1, pp. 260– 269, 2015.

[18] S. Wu, A. D. Sarma, A. Fabrikant, S. Lattanzi and A. Tomkins, "Arrival and Departure Dynamics in Social Networks," in *Proc. of ACM WSDM*, pp. 233–242, 2013.

[19] T. Zhang, P. Cui, C. Faloutsos, W. Zhu and S. Yang, "Come-and-Go Patterns of Group Evolution: A Dynamic Model", in *Proc. of ACM KDD*, 2016.

[20] B. Yin, Y. Yang and W. Liu, "Exploring Social Activeness and Dynamic Interest in Community-based Recommender System," in *Proc. of ACM WWW*, pp. 771–776, 2014.

[21] F. Chung and L. Lu, "Complex Graphs and Networks," in *Providence: American Mathematical Society*, Vol. 107, 2006.

[22] G. Ghoshal, L. Chi and A. Barab'asi, "Uncovering the Role of Elementary Processes in Network Evolution," in *Scientific Reports*, 2013.

[23] S. Lattanzi and D. Sivakumar, "Affiliation Networks," in *Proc. ACM STOC*, pp. 427-434, 2009.

[24] W. Aiello, F. Chung, L. Lu, "Random Evolution in Massive Graphs", in *42nd IEEE symposium on Foundations of Computer Science*, FOCS 2001.

[25] MoiveLens Dataset: http://grouplens.org/datasets/movielens/

[26] Audioscrobbler Music Artist Dataset: http://www-etud.iro.umontreal.ca/~bergstrj/audioscrobbler_data.html

[27] Yahoo Music Artist Rating Dataset: http://webscope.sandbox.yahoo.com/

[28] Book Crossing Dataset: http://www2.informatik.uni-freiburg.de/ cziegler/BX/

[29] Amazon Product Recommendation Dataset: http://jmcauley.ucsd.edu/data/amazon/links.html

APPENDIX

The proofs in this section appear in the order in which the original theorems or propositions are organized.

A. Proof of Theorem 4.2

Let E_t^k be the random variable that denotes the expected number of vertices in U of degree k at time t. We look into the cases for $k = c_u$ and $k > c_i$ respectively.

First we analyze the case when $k = c_u$. We have that

$$\begin{aligned} E_t^{c_u} =& E_{t-1}^{c_u} + Pr[\text{a new vertex is added to } U] \\ & - E[\#\text{ of vertices in } U \text{ with degree } c_u \\ & \quad \text{at time } t-1 \text{ whose degrees increase}]. \end{aligned}$$

In the evolving process, the degree of a vertex in U increases if and only if a vertex is added to I. Thus we have

$$\begin{aligned} E_t^{c_u} =& E_{t-1}^{c_u} + Pr[\text{a new vertex is added to } U] \\ & - (1-\beta) \\ & \quad E[\#\text{ of vertices in } U \text{ with degree } c_u \text{ at time } t-1 \\ & \quad \text{whose degrees increase} | \text{a vertex is added to } I] \\ =& E_{t-1}^{c_u} + \beta \\ & - (1-\beta) \sum_{k=1}^{c_i} Pr[\text{a vertex in } E_{t-1}^{c_u} \text{ is chosen} \\ & \quad \text{as endpoint for the } k\text{-th edge}], \end{aligned}$$

where the second equation comes from the linearity of expectation.

Since the "*prototype*" is chosen with a probability proportional to its vertex weight, and edges to "copy" are selected with probabilities proportional to edge weights, the probability of one edge to be chosen in a single selection is proportional to its edge weight. We have

$$\begin{aligned} & Pr[\text{a vertex in } E_{t-1}^{c_u} \text{ is chosen as endpoint for the } k\text{-th edge}] \\ & = \frac{E_{t-1}^{c_u} c_u Er}{W_{t-1} + W_{B_0}}, \end{aligned}$$

where Er is the global expectation of edge weight, namely the global expectation of a single rating. $W_{t-1} + W_{B_0}$ is the total edge weight in the bipartite graph at time $t-1$ and W_{B_0} is the total edge weight in the initial graph.

Thus we can derive

$$\begin{aligned} E_t^{c_u} =& E_{t-1}^{c_u} + \beta - (1-\beta) c_i \frac{E_{t-1}^{c_u} c_u Er}{W_{t-1} + W_{B_0}} + o(1) \\ =& E_{t-1}^{c_u} \left(1 - \frac{(1-\beta) c_u c_i Er}{(t-1)(c_u\beta + c_i(1-\beta))Er \pm o(t) + W_{B_0}}\right) \\ & + o(1) + \beta \\ =& E_{t-1}^{c_u} \left(1 - \frac{(1-\beta) c_u c_i Er}{(t-1)(c_u\beta + c_i(1-\beta))Er \pm o(t)}\right) \\ & + o(1) + \beta \\ =& E_{t-1}^{c_u} \left(1 - \frac{(1-\beta) c_u c_i Er}{(t-1)(c_u\beta + c_i(1-\beta))Er}(1 \pm o(1))\right) \\ & + o(1) + \beta. \end{aligned}$$

Using Lemma 4.1, we get the new equation

$$\begin{aligned} \lim_{t\to\infty} \frac{E_t^{c_u}}{t} &= \frac{\beta}{1 + (1-\beta)\frac{c_u c_i}{c_u\beta + c_i(1-\beta)}} \\ &= \frac{\beta(c_u\beta + c_i(1-\beta))}{c_u\beta + c_i(1-\beta) + (1-\beta)c_u c_i}. \end{aligned}$$

Next, we turn to analyze the general case when $k > c_i$. We know

$$\begin{aligned} E_t^k =& E_{t-1}^k - E[\#\text{ of vertices in U with degree } k \text{ at time } t-1 \\ & \quad \text{and then increase their degrees}] \\ & + E[\#\text{ of vertices in } U \text{ with degree} < k \text{ at time } t-1 \\ & \quad \text{and then increase degree to } k]. \end{aligned}$$

Notice that in the weighted bipartite graph there are no multiple edges and with an analysis similar to the case of $k = c_u$ we have

$$\begin{aligned} E_t^k =& E_{t-1}^k \left(1 - \frac{(1-\beta)c_i k Er}{(c_u\beta + (1-\beta)c_i)(t-1)Er}(1 \pm o(1))\right) \\ & + \frac{(1-\beta)c_i(k-1)Er}{(c_u\beta + (1-\beta)c_i)(t-1)Er}(1 + o(1))E_{t-1}^{k-1} + o(1). \end{aligned}$$

Let us define $Y_k = \lim_{t\to\infty} E_t^k / t$. Using Lemma 4.1, we transform the equation above into

$$\begin{aligned} Y_k &= \frac{(1-\beta)\frac{c_i(k-1)Y_{k-1}}{c_u\beta + c_i(1-\beta)}}{1 + (1-\beta)\frac{c_i k}{c_u\beta + c_i(1-\beta)}} \\ &= \frac{(k-1)}{k + 1 + \frac{c_u\beta}{c_i(1-\beta)}} Y_{k-1} \\ &= Y_{c_i} \prod_{j=c_i+1}^{k} \frac{j-1}{j + 1 + \frac{c_u\beta}{c_i(1-\beta)}} \\ &= Y_{c_i} \frac{\Gamma(k)}{\Gamma(k + 2 + \frac{c_u\beta}{c_i(1-\beta)})} \frac{\Gamma(c_i + 2 + \frac{c_u\beta}{c_i(1-\beta)})}{\Gamma(c_i)} \\ &\sim k^{-2-\frac{c_u\beta}{c_i(1-\beta)}}. \end{aligned}$$

Thus when time step t approaches infinity, the ensemble average of the degree of vertices in set U follows a power-law distribution with exponents $\alpha = -2 - \frac{c_u\beta}{c_i(1-\beta)}$. Using a symmetrical manner, we can also prove the similar results in set I.

B. Proof of Theorem 4.4

Similar to the proof of Theorem 4.3, the proof of Theorem 4.4 also includes 2 sequential parts:

I. Derivation of the recurrence relation of vertex weight sequence

II. Derivation of the power-law lower bound

PART I is identical to that in the proof of Theorem 4.3 and we start with PART II.

PART II. ***Derivation of the Power-law Lower Bound***

We assign an arbitrary set of initial values which are not all zero and suppose one lower bound of X_k is g. We have the following two cases:

When $c_u R + R + 1 \leq k \leq c_u R + 2R$, let

$$g(k) = \min_{c_u R+R+1 \leq j \leq c_u R+2R} X_j.$$

When $k > c_u R + 2R$,

$$\begin{aligned} X_k &= \frac{1}{k+C}(H(R)(k-R)X_{k-R} + \ldots + H(1)(k-1)X_{k-1}) \\ &\geq \frac{1}{k+C}(H(R)(k-R) + H(R-1)(k-R+1) \\ &\qquad + \ldots + H(1)(k-1)) \times \min_{k-R \leq j \leq k-1} g(j) \\ &\geq \frac{k-R}{k+C} \min_{k-R \leq j \leq k-1} g(j), \end{aligned}$$

namely,

$$g(k) = \frac{k-R}{k+C} \min_{k-R \leq j \leq k-1} g(j).$$

Since $\frac{k-R}{k+C} < 1$, we have $g(k) < \min_{k-R \leq j \leq k-1} g(j)$, which means the smallest $g(k)$ within an interval of length R is always the rightmost one and this results in

$$g(k) = \frac{k-R}{k+C} g(k-1).$$

From the above equation, we can derive

$$\begin{aligned} g(k) &= \frac{k-R}{k+C} \cdot \frac{k-1-R}{k-1+C} \cdot \ldots \cdot \frac{c_u R + 2R - R}{c_u R + 2R + C} g(c_u R + 2R) \\ &= \frac{\Gamma(k-R+1)}{\Gamma(k+C+1)} \frac{\Gamma(c_u R + 2R + C + 1)}{\Gamma(c_u R + 2R - R + 1)} g(c_u R + 2R) \\ &\sim k^{-R-C}. \end{aligned}$$

Therefore the lower bound of the X_k follows a power-law with exponent $\alpha = -R - C$, which is also illustrated in Figure 2. We can get the similar result for vertex weight distribution in set I in a symmetric manner.

C. Proof of Proposition 4.5

When $k < c_u$, $Y_t^k = 0$ since by definition there is no vertex in U with vertex weight less than c_u.

When $c_u < k < c_u R$, a newly added vertex in U could have vertex weight k. Thus,

$$\begin{aligned} V_t^k = & V_{t-1}^k - E[\text{\# of vertices in U with vertex weight} = k \\ & \qquad \text{at time } t-1 \text{ and increase at time } t] \\ & + Pr[\text{a new vertex with vertex weight } k \text{ is added to } U] \\ & + E[\text{\# of vertices in } U \text{ with vertex weight} < k \\ & \qquad \text{at time } t-1 \text{ and increase to } k \text{ at time } t]. \end{aligned}$$

So we have

$$\begin{aligned} V_t^k = & V_{t-1}^k \left(1 - (1-\beta)\frac{c_i k}{W_{t-1} + W_{B_0}}\right) + \beta s(k) \\ & + (1-\beta) \sum_{j=k-R}^{k-1} \frac{c_i j V_{t-1}^j}{W_{t-1} + W_{B_0}} H(k-j). \end{aligned}$$

Let $X_k = \lim_{t \to \infty} V_t^k / t$. Again, we use Lemma 4.1 and obtain

$$X_k = \frac{(1-\beta)\frac{c_i}{Er(c_u\beta + c_i(1-\beta))} \sum_{j=k-R}^{k-1} h(k-j) j X_j + \beta s(k)}{1 + (1-\beta)\frac{c_i k}{Er(c_u\beta + c_i(1-\beta))}},$$

namely,

$$X_k = \frac{1}{k+C} \sum_{j=k-R}^{k-1} H(k-j) j X_j + \frac{C\beta s(k)}{k+C},$$

where $C = Er\left(1 + \frac{c_u \beta}{c_i(1-\beta)}\right)$.

We can prove a symmetrical result for set I.

D. Proof of Proposition 4.6

In set U, since $Y_t^k = 0$ for $k < c_u$. The initial value of $X_k = 0$ when $k < c_u$.

We define $m(k) = \text{argmax}_{k-R \leq j \leq k-1} j X_j$ and we can get

$$\begin{aligned} X_k &= \frac{1}{k+C} \sum_{j=k-R}^{k-1} H(k-j) j X_j + \frac{C\beta s(k)}{k+C} \\ &\leq \frac{m(k) X_{m(k)}}{k+C} + \beta s(k) \\ &\leq l(m(k)) + \beta s(k). \end{aligned}$$

Since $X_k = 0$ when $k < c_u$, we have

$$\begin{aligned} l(k) &= l(m(k)) + \beta s(k) \\ &= l(m(m(k))) + \beta s(m(k)) + \beta s(k) \\ &= l(m(m(m(k)))) + \beta s(m(m(k))) + \beta s(m(k)) + \beta s(k) \\ &= \ldots \leq \beta \sum_{i=1}^{k} s(i). \end{aligned}$$

Given that $s(k)$ is the pdf of variable S and sums to 1, $l(k) \leq 1$ for $c_u \leq k \leq c_u R$. We can prove the case also holds in set I in a symmetrical manner.

E. Proof of Proposition 4.7

This is a direct result of the central limit theorem (CLT). Suppose random variable Z_i represents the edge weight for the i-th edge for a newly added vertex in U. Since $S = \sum_{i=1}^{c_u} Z_i$ and Z_i is i.i.d for different i, we define

$$S_{c_u} = \frac{\sum_{i=1}^{c_u} Z_i}{c_u}.$$

When c_u approaches ∞, by Lindeberg-Levy CLT, we have

$$\sqrt{c_u}(S_{c_u} - E[S_{c_u}]) \to N(0, \sigma^2),$$

which means S_{c_u} follows unimodal probability distribution when $c_u \to \infty$. Therefore S also follows unimodal probability distribution when c_u increases to ∞. Also it is obvious that when c_u is small enough, namely $c_u = 1$, the pmf $s(k)$ is the same as the rating pmf for new users $H(r)$. Symmetrically, we can easily prove the case for c_i and random variable S'.

F. Proof of Theorem 4.8

We prove the case for set U. From Propositions 4.5 and 4.6, we can derive the following result.

$$
\begin{aligned}
&\Delta X_k \\
=&\frac{1}{k+C}\sum_{j=k-R}^{k-1} H(k-j)jX_j + \frac{C\beta s(k)}{k+C} \\
&- \frac{1}{k+C-1}\sum_{j=k-R-1}^{k-2} H(k-j)jX_j - \frac{C\beta s(k-1)}{k+C-1} \\
=&\frac{H(1)(k-1)X_{k-1}}{k+C} + \left(\frac{H(2)(k-2)}{k+C} - \frac{H(1)(k-2)}{k+C-1}\right)X_{k-2} \\
&+ \ldots - \frac{H(R)(k-1-R)X_{k-1-R}}{k+C-1} + \Delta\left(\frac{C\beta s(k)}{k+C}\right).
\end{aligned}
$$

Further we have,

$$
\begin{aligned}
&\left|\Delta X_k - \Delta\left(\frac{C\beta s(k)}{k+C}\right)\right| \\
=&\left|\frac{H(1)(k-1)X_{k-1}}{k+C} + \left(\frac{H(2)(k-2)}{k+C} - \frac{H(1)(k-2)}{k+C-1}\right)X_{k-2}\right. \\
&\left. + \ldots - \frac{H(R)(k-1-R)X_{k-1-R}}{k+C-1}\right| \\
\leq&\left|\left(\frac{H(2)(k-2)}{k+C} - \frac{H(1)(k-2)}{k+C-1}\right)X_{k-2}\right| \\
&+ \left|\left(\frac{H(3)(k-3)}{k+C} - \frac{H(2)(k-3)}{k+C-1}\right)X_{k-3}\right| + \ldots \\
&+ \left|\frac{H(1)(k-1)X_{k-1}}{k+C} - \frac{H(R)(k-1-R)X_{k-1-R}}{k+C-1}\right| \\
\leq&\left|\frac{(H(2)-H(1))(k+C)(k-1) - H(2)(k-1)}{(k+C)(k+C-1)}\right| \\
&+ \left|\frac{(H(3)-H(2))(k+C)(k-2) - H(3)(k-2)}{(k+C)(k+C-1)}\right| + \ldots \\
&+ \max\{H(1), H(R)\} \\
\leq&\left|H(2) - H(1) - \frac{H(2)}{k+C}\right| + \left|H(3) - H(2) - \frac{H(3)}{k+C}\right| + \ldots \\
&+ \max\{H(1), H(R)\} \\
\leq&|H(2) - H(1)| + \left|\frac{H(2)}{c_u+C}\right| + |H(3) - H(2)| + \left|\frac{H(3)}{c_u+C}\right| \\
&+ \ldots + \max\{H(1), H(R)\} \\
\leq&H' + \frac{1}{c_u+C}.
\end{aligned}
$$

Symmetrically we can get similar results for the case in set I.