Evolving Bipartite Model Reveals the Bounded Weights in Social Networks: A Case Study in Recommendation Networks

Lingkun Kong, Xudong Wu, Hongru Zhu, Luoyi Fu, Xinbing Wang
Shanghai Jiao Tong University, China

ABSTRACT

Many realistic social networks can be presented by evolving bipartite graphs [1], in which dynamically added elements are divided into two entities and connected by links through these two entities, such as followees and followers in Twitter networks, authors and scientific topics in scholarly networks, users and items in recommendation networks, and etc. However, when adding weights of edges, i.e., connections between elements, how to mathematically model such evolving social networks, along with quantitative characterizations, remains unexplored.

Motivated by this, we develop a novel evolving bipartite model (EBM), which, based on empirically validated power-law distribution within multiple realistic social networks, discloses that the distribution of total weights given or received by each element in networks is determined by the weighting scale and bounded by certain ceilings and floors. Based on these theoretical results, once power-law is discovered in bipartite social networks, the overall weights of vertices can be predicted by EBM. To illustrate, in recommending networks, the evaluation of items, i.e., total rating scores can be determined in a small scope; in scholarly networks, the total numbers of publications under specific topics can be anticipated with a certain range; in Twitter social networks, the influence of users can be roughly measured. Therefore, we perform extensive experiments on 6 realistic recommendation networks with varying weighting, i.e., rating scales to further evaluate the performance of EBM by case studying, and experimental results demonstrate that knowing weighting scale, both the upper bound and the lower bound of total weights of vertices in social networks can be properly predicted by EBM.

1 INTRODUCTION

Social networks are theoretical constructs useful in the social sciences to study relationships between individuals, groups, organizations, or even entire societies [2]. Many of naturally arising social networks can be either directly modeled as bipartite graphs, such as affiliation networks [3, 4], or be obtained by the projection of a bipartite graph such as co-membership networks [5]. For instance, as many online stores allow users to post their evaluation of items, the social networks, namely recommendation networks [6] naturally emerge. Due to the affiliation among users and items, these recommendation networks can be viewed as bipartite graphs where users and items are two group of vertices, and the behaviors of users rating items generate edges between two group of vertices. Also, the human contact networks collected from social media can be analyzed as bipartite graphs. Using data from Twitter as an example [7], in which followees and followers represent two group of vertices, and their contacts denote edges in bipartite graphs.

Regarding this, there have been some initial efforts directed towards analysis of evolving social networks through bipartite graph models [8, 4, 9], and several works extend the bipartite graph model to tripartite and even K-entities graph models [10, 11]. However, few of them combine the weight sof edges in the bipartite-graph modeled social networks into their models. For instance, in recommendation networks, when users being linked to items by rating behaviors, the links are attached with rating scores. In other works, in the bipartite graphs of recommendation networks, edges connecting two group of vertices have weights.

To bridge this gap, in this paper we propose a novel evolving bipartite model (EBM) to incorporate the weights of edges into the whole framework. This model is founded on observed power-law distribution in bipartite social networks, and by both theoretical and experimental validation, EBM reveals that the total weights of vertices, e.g., the total rating scores items receive in recommendation networks, are bounded by certain upper and lower limits. To begin with, employing 10 realistic datasets which are consisted of recommendation networks, scholarly networks and Twitter social networks, we observed power-law distribution of vertices' degree and total weights in bipartite social networks. Based on these observations, we propose EBM to model social networks which is inspired by the intuition of *influence of the elderly*, i.e., the new arrived vertices in bipartite graph being impacted by weight distributions of the existing vertices. Further, with the purpose of explicitly clarifying mathematical notations in EBM, we characterize EBM by notations in the context of recommendation networks and analyze it theoretically. By theoretical analyzing results and experimental validations in 6 realistic recommendation networks, we reveal the bounded rating scores of items, which indicates that the total weights of vertices in social networks that can be modeled as bipartite graphs with power-law distribution are also bounded. Here, we summarize contributions of EBM by three aspects:

Weights Incorporated: The first contribution of EBM is to originally incorporate the weights of edges in bipartite graphs into the model construction of social networks. Translating rating scores given by users or received by items in recommendation networks into weights of edges, EBM properly considers the weights as one important component of the model. To illustrate, when a new user/item arrives at networks, EBM suppose the weights distribution of the new arrival user/item will be exposed to the impact from weights distributions of existing users/items. Further, EBM also portrays the bound of items' total rating scores in recommendation networks, which could be leveraged to predict bounds of total weights of vertices in other social networks.

Mathematical Justified: The next contribution of EBM is to offer detailed mathematical proof to consolidate the reasonability of our model. Based on the constructing methods of random arrival, preferential attachment, edge copying and weight distribution imitating, EBM successfully reproduces the observed power-law distributed degree and weights in bipartite social networks. Also, by theoretical analysis, EBM explains the reason of discontinuity in the distribution sequence of items' weights.

Table 1: Dataset statistics and fitting parameters

Dataset	Weighting scale	# Users/Authors/Followees	# Items/Topics/Followers	Weights	α	Parameter C
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
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
ArtificialIntelligence	[1,61]	3,277,927	1,174,863	11,807,837	—	—
Algorithms	[1,103]	1,687,270	1,260,825	7,519,219	—	—
ProgrammingLanguage	[1,65]	3,957,282	1,692	11,473,305	—	—
Twitter	[1,1102]	40,103,281	35,689,148	1,468,365,182	—	—

Experiment Validated: EBM's third contribution relies on extensive experimental validations in the context of recommending networks. In this paper, we launch exhaustive investigations on 6 realistic recommendation networks with varying weighting, i.e., rating scales to further evaluate the performance of EBM, and experimental results demonstrate that knowing weighting scale, both the upper bound and the lower bound of total weights of vertices can be properly predicted by EBM.

The paper is organized as follows. In Section 2, we discuss relevant literatures. In Section 3, We introduce our datasets and present our observation results. We illustrate our design of EBM and characterize it mathematically in Section 4. Then, we theoretically analyze EBM in Section 5. Section 6 is our experiments and we conclude in Section 7.

2 RELATED WORKS

There is a flurry of prior work in evolving network sciences. A series of studies [12, 13, 14, 15] have illustrated the structure of evolving social networks with the arrival and departure of users [16, 17] and the temporary dynamics of interest [18]. Some work features the evolving process in the network model. Chung et al. [19] introduce the assumption of either vertex-arrival or edge-arrival at each time step. Ghoshal et al. [20] propose a model that is able to clarify the role of individual elementary mechanisms. Lattanzi et al. [4] introduce the model of affiliation networks, where preferential attachment and edge copy are emphasized in the proposed evolving process. Liu et al. [11] develop an evolving K-Graph model which studies the hybrid interactions among multiple network entities. Liu et al. [9] propose an algorithm named Crosslayer 2-hop Path (C2P) which leverages bipartite evolving network structure to study problems in recommendation systems.

3 OBSERVATION

In this section, we launch our observation based on multiple real publicly available social network datasets. We first introduce datasets we use in Section 3.1 and then illustrate our discovery in Section 3.2, based on which EBM is founded.

3.1 Introduction of Datasets

In this paper, we observe 10 real publicly available social network datasets, including 6 datasets about recommendation networks, 3 datasets about scholarly networks and 1 dataset about the Twitter

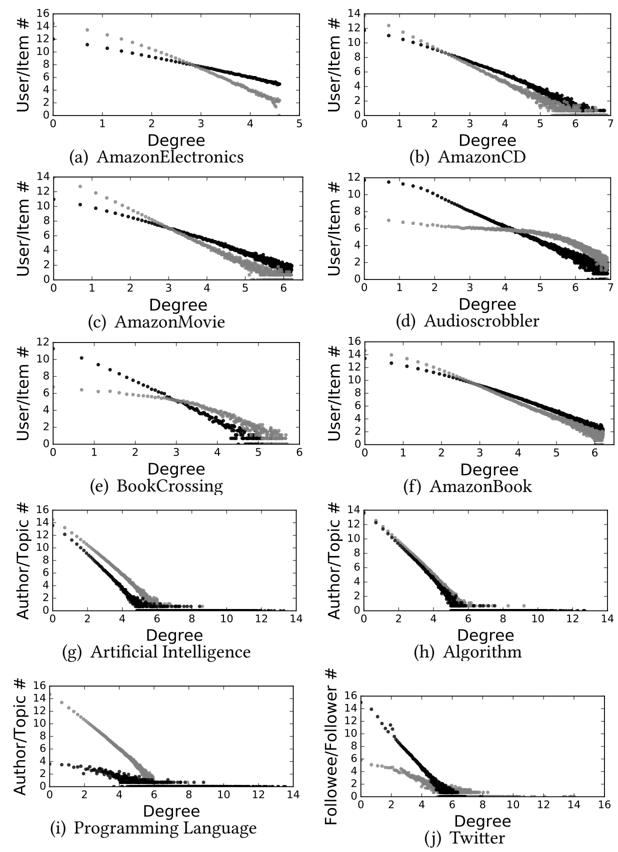


Figure 1: Degree, i.e., number of connections, distribution of users/ authors/ followees (grey dot) and items/ topics/ followers (black dot) in logarithm

social network. To illustrate, 6 recommendation datasets are comprised of the Audioscrobbler music artist rating dataset[1] [21], the BookCrossing book recommendation dataset[2] [22], and the Amazon product recommendation datasets for movies, CDs, electronics and books[3] [23]. The users and items in recommendation datasets are connected by users rating items, where weights are users' rating scores. 3 scholarly datasets separately contain scholarly data

[1] http://www-etud.iro.umontreal.ca/~bergstrj/audioscrobbler_data.html
[2] http://www2.informatik.uni-freiburg.de/~cziegler/BX/
[3] http://jmcauley.ucsd.edu/data/amazon/links.html

in scientific fields of artificial intelligence, algorithm and programming language from *Microsoft Academic Graph*[4] [24]. In each fields, authors and topics are linked by authors publishing papers under topics, and weights are presented by numbers of published papers. The dataset of Twitter social network[5] [7] divides Twitter users to be followees and followers, in which nodes are connected by followers following followees. In this dataset, we regard the weights of connections by influence of this connection. For instance, for a followee, a follower who himself has 100 followers has greater influence than a follower who only has 1 follower.

These datasets vary in size and type, thereby guaranteeing the generality of our observing results as well as the applicability of our model which will be presented in Section 5.

Remarks: When we conduct experiments on those datasets that have implicit weights, we set up a rule to generate corresponding integer weights. For instance, in Audioscrobbler dataset, we map users' play count for musical artist to explicit integer weights in interval $[1, 5]$. Specifically, we set the mapping rule as follows. If the user listens only once to an artist, then the weight is 1. When the play count is in range [2,4],[5,9] or [10,19], the weight is set to 2,3 or 4 separately. A weight 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. Also, Table 1 gives value of exponent α in power-law distribution and parameter C, which we will discuss in Section 6.

3.2 Power-distributed Degree and Weight

We start our observation with calculating the number of connections for elements in datasets. Using this value as degree, we analyze the distribution of connections between elements in 10 datasets in logarithm. As shown in Figure 1, there exhibits a power-law distribution of degree in different kinds of social networks. Actually, in later discussion, our model leverages this power-law distribution as an important theoretical premise.

Then, we count the total weight for elements in bipartite social networks. Similarly, we analyze the distribution of total weights given or received by elements in datasets in logarithm. As Figure 2 presents, there also exhibits a power-law distribution of total weights, i.e., few elements generate or receive high total weights in social networks, and the number of elements are power-law related with their corresponding total weights. This property also provides insights for us to build EBM.

Interestingly, by observing Figure 1 and 2, we find that the beginning of the distribution sequence shows great discontinuity. This phenomenon actually is well reproduced in EBM both in theory and practice and properly discussed in Section 6. Here, in advance, we say that the beginning of distribution sequences in bipartite social networks, instead of being pure power-law, resembles the expected rating probability mass function in certain situations.

4 MODEL OF EBM

Based on above observations, we design a novel model to capture these properties, which is named as evolving bipartite model (EBM).

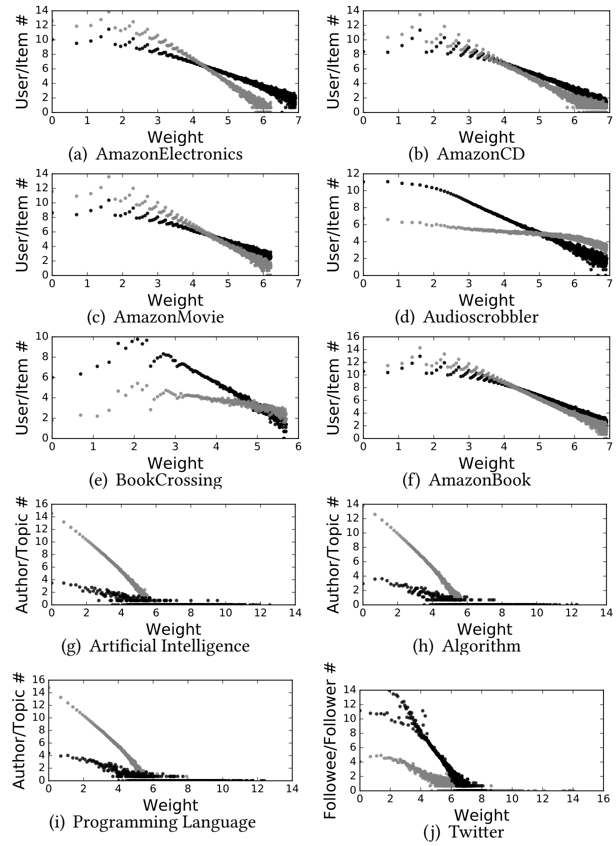


Figure 2: Total weight distribution of users/ authors/ followees (grey dot) and items/ topics/ followers (black dot) in logarithm

In this section, we begin with airing the intuition behind of this model. Then, we mathematically characterize EBM by introducing the model structure, new definitions, two basic assumptions and the evolving algorithm.

4.1 Intuition of EBM

Viewing social networks as evolving bipartite graphs composed of two sets of vertices, we assume that new vertices appear in social networks exposed to influence of the elderly, i.e., existing vertices.

Here we firstly use recommendation networks as an instance, in which users and items are two sets of vertices in bipartite graph, and users' rating scores are weights of edges in the graph. In the perspective of users, a new user is likely to learn from rating scores given by an influential elderly user, who is thus selected as a **prototype** and impacts the new user on launching rating scores, i.e., weights. Similarly, in the perspective of items, a new item which is yearn for a success, is likely to imitate an outstanding item, i.e., the item with high total rating score. Therefore, since this outstanding elderly item is chosen to be prototype, the rating score of the new item is, to some extent, allied with the score of its prototype. When using scholarly networks as an example, now authors and topics are vertices in the graph, and the numbers of publication are weights of edges. In authors' aspect, a new author might learn from elderly authors who have large amount

[4] http://acemap.sjtu.edu.cn/acenap/index.php/datasets.html
[5] http://an.kaist.ac.kr/traces/WWW2010.html

of publications; In topics' aspect, a rising topic is likely to be an branch new of an existed topic in which massive papers have been published. As for twitter networks, followees and followers represent two sets of vertices, and the weights now are influence of users in twitter networks. Naturally, with the purpose of pursuing headlines, new followers will choose to follow influential followees, and new followees, learning from the elderly that what might appeal influential followers, are likely to post similar tweets to draw attention from those influential followers.

By these examples, being impacted by the elderly seems to be an universality in bipartite social networks, and based on this intuition, with the constant arrival of new vertices and the weights of edges, a virtuous cycle will be normally resulted:

- A newly arrived vertex, with a high probability, will show preference on *popular* vertices, where popular indicates the specific element has high total weights in networks.
- Preference, with a high probability, will result in new weights to popular vertices, adding up to their total weights and strengthening their popularity.
- As the bipartite graph evolves, popular vertices become more popular while outmoded ones gradually become neglected, which contributes to final power-law distribution in the network.

Based on these evolving processes, we build the EBM. In rest parts of this paper, the EBM will be illustrated in detail and examined both by theoretical and experimental analysis, which, in reverse, can demonstrate that our intuitive assumption does reflect evolving mechanism in social networks.

4.2 Mathematical Characterization

In order to simplify mathematical symbols and endow our model with more concrete meaning, we characterize EBM under the specific case of social networks, i.e., users rating items. Here users and items can be extended to vertices in other bipartite social networks which yielded to power-law distribution, and rating scores can be extended to weights of edges.

Model Structure: We use a simple weighted bipartite graph structure $B(U, I)$ to present EBM. Vertices in set U represent users and vertices in set I represent items in recommendation networks. 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.

Definition of 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 recommendation networks.

Basic Assumptions: 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: As Figure 3 illustrates, the evolving process in EBM incorporates two symmetrical aspects, i.e., the arrival of new users and items. Here we use the arrival of a new user as an instance. 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 ratings 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. Algorithm 1 shows the mathematically characterized evolving process in $B(U, I)$.

Remarks: EBM only produces weights in positive integers. Yet in real recommendation networks or other social networks, some may allow the existence of decimal rating scores like 1.5 stars or allow decimal value of weights. However, this doesn't hurt since we can always map decimal or implicit ratings into explicit positive integer ratings according to preestablished rules in subsection 3.1.

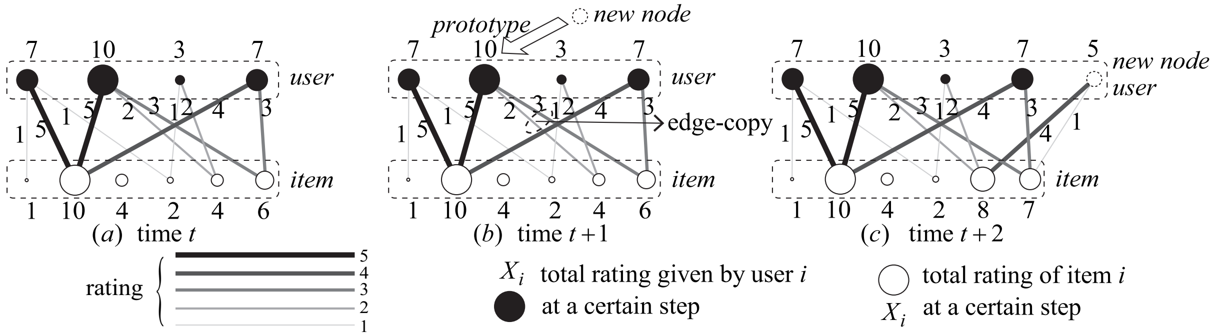


Figure 3: An illustration of how the recommender system evolves in terms of users, items, and ratings.

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 $\{1, 2, \ldots, R\}$ 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 $\{1, 2, \ldots, R\}$.
  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

---

5 THEORETICAL ANALYSIS OF EBM

We present here the theoretical analysis regarding degree and vertex weight distributions in $B(U, I)$. Theorem 5.2 shows power-law degree distributions. Theorems 5.3 and 5.4 show power-law bounds for vertex weight distributions. Proposition 5.7 and Theorem 5.8 analyze the beginning of vertex weight distributions.

5.1 Analysis of Degree Distribution

We start with the degree distribution in EBM and introduce Lemma 5.1 before we state Theorem 5.2 regarding degree distributions.

Lemma 5.1 ([25]). *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 approaches similar to [4], we derive Theorem 5.2.

Theorem 5.2. *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. □

5.2 Analysis of Vertex Weight Distribution

The vertex weight represents the total ratings given. Here by theoretical analysis of EBM, we give **the upper and lower bounds** of vertex weight distributions in Theorems 5.3 and 5.4 separately and provide illustrations on techniques used to prove these two theorems in Figure 4. We propose Theorem 5.8 regarding the beginning of vertex weight distributions.

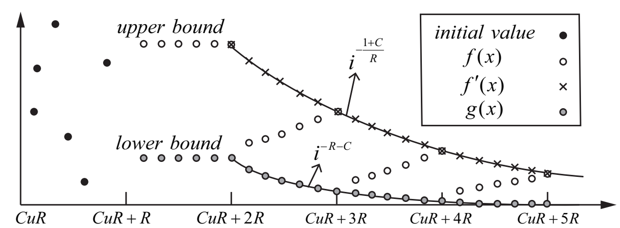


Figure 4: An illustration of proving upper and lower bounds of the vertex weight distribution.

5.2.1 *Upper Bound of Vertex Weight Distribution.*

Theorem 5.3. *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 $c_uR + 2R$ ($c_iR + 2R$).*

Proof. The proof of Theorem 5.3 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 $k > c_u R + R$ 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 in Theorem 5.2, the probability that an edge is selected in a single selection is proportional to its weight. Thus,

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

Since we have

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

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 5.1, we get the following recurrence relation

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

namely,

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

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

This high order homogeneous recurrence relation is not detachable as in the proof of Theorem 5.2. 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

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

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 $c_u R + 2R + 1 \le k \le c_u R + 3R$:

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

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

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

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

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

Recursively, the above equation results in

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

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

As illustrated in Figure 4, 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 4, $f'(k)$ 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. □

5.2.2 *Lower Bound of Vertex Weight Distribution*.

Theorem 5.4. *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 5.2 is consistent with Theorems 5.3 and 5.4 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 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 5.2. And note that the following inequality always holds for positive integer R.

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

5.2.3 *Beginning of Vertex Weight Distribution*.

In Sections 5.2.1 and 5.2.2, 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 5.3 and 5.4, 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 5.8. The first two give the recurrence relation and an upper bound of X_k. The third describes the distributions of random variables S and S'. Proofs of these 3 propositions are provided respectively by Appendix. C, D, E.

Proposition 5.5. *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 5.6. *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 5.7. *When c_u (resp. c_i) is large enough, S (S') follows a unimodal probability distribution.*

Now we introduce Theorem 5.8 regarding the difference of vertex weight sequence and the difference of the nonhomogeneous term in the recurrence relation.

Theorem 5.8. *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

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

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

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

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

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 5.7 and Theorem 5.8, we can roughly depict the vertex weight distribution 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.

6 EXPERIMENTS

In this section, we discuss experiments launched on 6 real recommendation networks presented in Table 1 to validate the reproducibility of EBM. As shown by observation results in Figure 1 that the degrees in the groups of users and items in recommendation networks are power-law distributed, Theorem 5.2 provides the equation for calculating the value of exponent in power-law distribution. A first set of experiments explore the value of exponent in real-world power-law distributions. Then, as we move on to validate Theorems 5.3, 5.4 and 5.8, we carry out experiments focusing

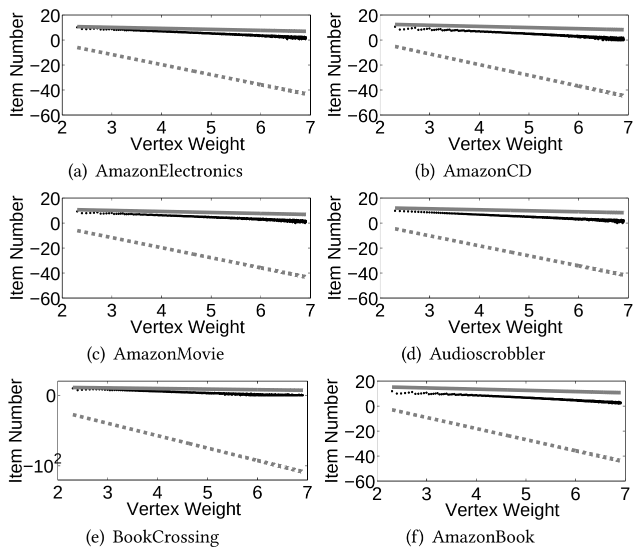


Figure 5: Upper (solid grey) and lower bound (dash grey) of vertex weight distribution (black) of items in logarithm

on the vertex weight of items from those datasets, and validate the bounded total rating scores in recommendation networks.

Remarks: Owing to the symmetrical characteristic between users and items, experimental results are highly identical on both sides of users and items. Here we only present experimental results from items' side, since the total rating score reflects the fame gained by items such as movies, books, and electronic products, etc., which draws more public attention.

6.1 Degree Distributions

Clearly, by observation, Figure 1 evidence the power-law distributed degree of users and items in 6 datasets in logarithm. We fit item degree distributions in 6 datasets and calculate the value of exponent in power-law distribution. 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 consistent with Theorem 5.2 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 5.3 and 5.4. and also present them in Table 1.

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

6.2 Vertex Weight Distributions

In this subsection, we validate our theoretical results regarding vertex weight distributions. In Figure 2, power-law distributions of item vertex weight, i.e., items' total rating scores have been observed. Here, we further plot them with their theoretical upper and lower bounds in logarithm in Figure 5, where 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 5.3 and 5.4 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 2 exhibits a consecutive curve at the beginning of the sequence, indicating a small c_i for selected datasets according to Proposition 5.7 and Theorem 5.8. 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 recommendation networks 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 small vertex weights and are not competitive with other existing 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 have higher weights in the future.

Moreover, 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 5.3 and 5.4 is

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

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.

7 CONCLUSION

In this paper, using 10 datasets about real social networks, we start with observing several interesting distributions in social networks. Based on observations, we propose a novel evolving bipartite model called EBM that highlights the establishment of social connections for new vertices and the characterization of their behaviors based on weighting-driven preferential attachment. The superiority of our model lies in three aspects: good capture of realistic social networks, mathematically tractability and novelty in predicting the bounds of final weights of connections. In the case study of recommendation networks, we also investigate the beginning of the item vertex weight distribution, which resembles the expected rating pmf for new items.

REFERENCES

[1] Aaron Pope, Daniel Tauritz, and Alexander Kent. Evolving bipartite authentication graph partitions. *IEEE Transactions on Dependable and Secure Computing*, 2017.

[2] John Scott. *Social network analysis*. Sage, 2017.

[3] Elena Zheleva, Hossam Sharara, and Lise Getoor. Co-evolution of social and affiliation networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1007–1016. ACM, 2009.

[4] Silvio Lattanzi and D Sivakumar. Affiliation networks. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 427–434. ACM, 2009.

[5] Mark EJ Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences*, 101(suppl 1):5200–5205, 2004.

[6] Bryan Perozzi, Michael Schueppert, Jack Saalweachter, and Mayur Thakur. When recommendation goes wrong: Anomalous link discovery in recommendation networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 569–578. ACM, 2016.

[7] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.

[8] Nesserine Benchettara, Rushed Kanawati, and Celine Rouveirol. Supervised machine learning applied to link prediction in bipartite social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*, pages 326–330. IEEE, 2010.

[9] Luoyi Fu Xinbing Wang Jiaqi Liu, Qi Lian. Who to connect to? joint recommendations in cross-layer social networks. In *to appear in IEEE INFOCOM 2018*.

[10] Jiaqi Liu Jinghao Zhao Luoyi Fu Xinbing Wang Lingkun Kong, Fengyu Deng. Msm: A multi-entity scholarly model for systematic understanding of evolving scholarly networks. ACM, 2017.

[11] Jiaqi Liu, Yuhang Yao, Xinzhe Fu, Luoyi Fu, Xiao-Yang Liu, and Xinbing Wang. Evolving k-graph: Modeling hybrid interactions in networks. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, page 29. ACM, 2017.

[12] Martin Atzmueller, Andreas Ernst, Friedrich Krebs, Christoph Scholz, and Gerd Stumme. On the evolution of social groups during coffee breaks. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 631–636. ACM, 2014.

[13] Neil Zhenqiang Gong, Wenchang Xu, Ling Huang, Prateek Mittal, Emil Stefanov, Vyas Sekar, and Dawn Song. Evolution of social-attribute networks: measurements, modeling, and implications using google+. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 131–144. ACM, 2012.

[14] Naga Bhushan, Junyi Li, Durga Malladi, Rob Gilmore, Dean Brenner, Aleksandar Damnjanovic, Ravi Sukhavasi, Chirag Patel, and Stefan Geirhofer. Network densification: the dominant theme for wireless evolution into 5g. *IEEE Communications Magazine*, 52(2):82–89, 2014.

[15] Yanhong Wu, Naveen Pitipornvivat, Jian Zhao, Sixiao Yang, Guowei Huang, and Huamin Qu. egoslider: Visual analysis of egocentric network evolution. *IEEE transactions on visualization and computer graphics*, 22(1):260–269, 2016.

[16] Shaomei Wu, Atish Das Sarma, Alex Fabrikant, Silvio Lattanzi, and Andrew Tomkins. Arrival and departure dynamics in social networks. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 233–242. ACM, 2013.

[17] Tianyang Zhang, Peng Cui, Christos Faloutsos, Yunfei Lu, Hao Ye, Wenwu Zhu, and Shiqiang Yang. Come-and-go patterns of group evolution: A dynamic model. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1355–1364. ACM, 2016.

[18] Bin Yin, Yujiu Yang, and Wenhuang Liu. Exploring social activeness and dynamic interest in community-based recommender system. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 771–776. ACM, 2014.

[19] Fan RK Chung and Linyuan Lu. *Complex graphs and networks*. Number 107. American Mathematical Soc., 2006.

[20] Gourab Ghoshal, Liping Chi, and Albert-László Barabási. Uncovering the role of elementary processes in network evolution. *Scientific reports*, 3, 2013.

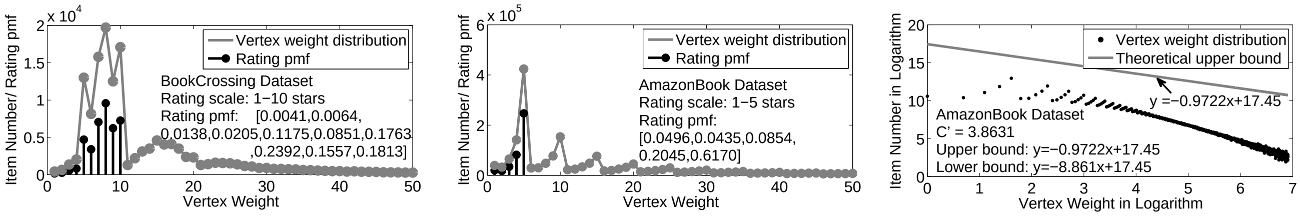
[21] Òscar Celma Herrada et al. *Music recommendation and discovery in the long tail*. Universitat Pompeu Fabra, 2009.

[22] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32. ACM, 2005.

[23] Julian McAuley and Alex Yang. Addressing complex and subjective product-related queries with customer reviews. In *Proceedings of the 25th International Conference on World Wide Web*, pages 625–635. International World Wide Web Conferences Steering Committee, 2016.

[24] Luoyi Fu Xinbing Wang. Acenap: A general-purpose network analysis dataset. 2017.

[25] William Aiello, Fan Chung, and Linyuan Lu. Random evolution in massive graphs. In *Handbook of massive data sets*, pages 97–122. Springer, 2002.



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

Figure 6: Vertex weight distribution for book recommendation datasets.

APPENDIX

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

A. Proof of Theorem 5.2

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

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

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

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

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 5.1, 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

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

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 5.1, 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 5.4

Similar to the proof of Theorem 5.3, the proof of Theorem 5.4 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 5.3 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 $c_uR + R + 1 \le k \le c_uR + 2R$, let

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

When $k > c_uR + 2R$,

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

namely,

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

Since $\frac{k-R}{k+C} < 1$, we have $g(k) < \min\limits_{k-R \le j \le k-1} g(j)$, 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

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

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 5.5

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 $c_u < k < c_uR$, 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

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

Let $X_k = \lim_{t\to\infty} V_t^k / t$. Again, we use Lemma 5.1 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 5.6

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 $m(k) = \operatorname{argmax}_{k-R \le j \le k-1} jX_j$ and we can get

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

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

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

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

E. Proof of Proposition 5.7

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 5.8

We prove the case for set U. From Propositions 5.5 and 5.6, 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,

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

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