Volume 6, Issue 8, pp. 72-74, 2022.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

Matrix Representation of Graph Probability Diffusion Model Based on Heterogeneous Social Network

Yang Wang
School of Computer and Information, Anqing Normal University, Anqing 246133, China

***Abstract*— *As a current research focus and hotspot, the graph probability diffusion model has been widely used in recent years because of its solid physical meaning and excellent recommendation performance. However, how to unify and represent the diffusion process of the model from the view of matrix theory? This issue is of great significance for unifying the algorithm representation and facilitating the model calculation. To address this challenge, this paper discusses the graph probability diffusion model from the perspective of matrix representation. Specifically, we first elaborate the purpose and purport of the graph probability diffusion model. Then we describe the matrix representation of the diffusion and recommendation process of this model. Finally, we give a concrete application example to validate the feasibility of the proposed method.*

***Keywords*— *Graph probability diffusion; Matrix representation; Heterogeneous Social Network; Resource allocation.*

I. Introduction

The graph probability diffusion model is subject to the law of conservation of mass. It traverses through heterogeneous social networks by the random walk strategy, which is essentially a mass diffusion algorithm. Generally, this model makes use of "resource" [1] to express users' preferences, and uses the mass diffusion algorithm to propagate the intended user' interest through heterogeneous social networks, so as to obtain the preference degree of the intended user for each item. Compared to traditional collaborative filtering, the graph probability diffusion model has better recommendation accuracy and low storage and computation cost. Moreover, it does not depend on the feature vector of items, which overcomes the difficulty of feature extraction in content-based recommendation [2-4]. In this paper, the resource allocation and recommendation process of the graph probability diffusion model based on heterogeneous social networks are introduced. Then the matrix representation of the graph probability diffusion model is described. Ultimately, an example analysis of the proposed method is given.

II. Graph Probability Diffusion Model

A. *Construction of heterogeneous social networks*

Considering a heterogeneous social network [5-6] $G(N, E)$, the node set $N = \{v \mid v \in U \vee v \in I\}$ consists of two disjoint sets U and I, where U is the users set, and I is the items set. Given $U=\{U_1, U_2, ..., U_m\}$, $I=\{I_1, I_2, ..., I_n\}$, the edge set E is denoted as $\{e_{i\alpha} \mid e_{i\alpha} = (U_i, I_\alpha, a_{\alpha i}) \wedge U_i \in U \wedge I_\alpha \in I\}$, where $e_{i\alpha}$ is the edge between U_i and I_α. If G is an unweighted heterogeneous social network, the weight of edge $e_{i\alpha}$ can be described as:

$$\begin{cases} a_{i\alpha} = 1 \text{ if user } i \text{ has selected item } \alpha\ , \\ a_{i\alpha} = 0 \text{ otherwise.} \end{cases} \quad (1)$$

Table 1 shows sample data of user-item interaction behavior. Based on the interaction, each user has a sequence of selected items.

TABLE 1. An example of user-item behavior data

User	Selected items
U_1	I_1, I_2
U_2	I_2, I_3
U_3	I_1, I_3, I_4

In the graph probability diffusion model, the user-item interaction behavior in Table 1 is first represented by the standard binary group （U_i, I_α） shown in Table 2, where each row in the table means that user U_i has selected item I_α.

TABLE 2. Normalized user-item behavior data set

User	Selected item
U_1	I_1
U_1	I_2
U_2	I_2
U_2	I_3
U_3	I_1
U_3	I_3
U_3	I_4

According to the data in Table 2, we can construct the user-item bipartite heterogeneous social network, as shown in Figure 1. The nodes in the graph are divided into two disjoint node sets. One of them is the user node set $\{U_1, U_2, U_3\}$ and the other is the item node set $\{I_1, I_2, I_3, I_4\}$. The nodes connecting each edge belong to the user node set and the item node set, respectively. After the user-item interaction behavior data is constructed as a heterogeneous social network, the graph probability diffusion model can be used for personalized recommendations.

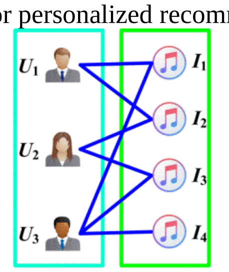

Fig. 1. the user-item bipartite graph constructed from sample data in Table 2

B. *Resource allocation and recommendation of graph probability diffusion model*

International Journal of Scientific Engineering and Science
Volume 6, Issue 8, pp. 72-74, 2022. ISSN (Online): 2456-7361

The resource allocation process of the graph probability diffusion model based on heterogeneous social networks can be decomposed into three steps:

(1) First, given the intended user U_i, the graph probability diffusion model provides an initial preference value (initial resource) reflecting U_i's preference for each item. The initial resource value of the item selected by U_i is 1, otherwise it is 0. As a result, each item receives a certain amount of initial resource, labeled $f(I)$.

(2) Then, U_i's preference is diffused in the heterogeneous social network. Each item allocates its resources equally to all adjacent users. In consequence, U_i's resource received from its adjacent items is:

$$f'(U_i) = \sum_{\beta=1}^{n} \frac{a_{\beta i} f(I_\beta)}{K(I_\beta)}, \tag{2}$$

where $K(I_\beta) = \sum_{i=1}^{m} a_{\beta i}$ is the degree of item I_β.

(3) Finally, each user returns the obtained resources evenly to the selected items. After the two-step graph probability diffusion, the ultimate obtained resource of item I_α is:

$$\begin{aligned} f''(I_\alpha) &= \sum_{i=1}^{m} \frac{a_{i\alpha} f(U_i)}{K(U_i)} \\ &= \sum_{i=1}^{m} \frac{a_{i\alpha}}{K(U_i)} \sum_{\beta=1}^{n} \frac{a_{\beta i} f(I_\beta)}{K(I_\beta)} \\ &= \sum_{\beta=1}^{n} \frac{1}{K(I_\beta)} \sum_{i=1}^{m} \frac{a_{\alpha i} a_{\beta i}}{K(U_i)} f(I_\beta), \end{aligned} \tag{3}$$

where $K(U_i) = \sum_{\beta=1}^{n} a_{\beta i}$ is the degree of U_i.

Formula (3) can be rewritten as follows:

$$f''(I_\alpha) = \sum_{\beta=1}^{n} \omega_{\alpha\beta} f(I_\beta), \tag{4}$$

where

$$\omega_{\alpha\beta} = \frac{1}{K(I_\beta)} \sum_{i=1}^{m} \frac{a_{\alpha i} a_{\beta i}}{K(U_i)}. \tag{5}$$

$\omega_{\alpha\beta}$ represents the resource contribution of B I_β to I_α in graph probability diffusion.

Given the intended user, after a graph probability diffusion process, each item acquires a certain amount of ultimate resource, which denotes the recommendation score assigned by the graph probability diffusion model. The higher the obtained score of an item, the more interest the intended user has in it [4]. All items that are not selected by the intended user U_i are sorted in descending order according to the ultimate resource amount $f''(I_\alpha)$. Ultimately, the top N items in the queue are recommended to the intended user U_i.

III. Matrix Representation of the Graph Probability Diffusion Model

In Section 2, the resource allocation and recommendation of the graph probability diffusion model is analyzed step by step from the perspective of physical process, which is complicated in form and tedious in calculation. To unify the form of the model and simplify the calculation of the corresponding algorithm, we discuss and analyze the matrix representation of the model from the perspective of matrix theory.

(1) Suppose **F** is the initial resource vector of the item.

(2) The transfer matrix for items to allocate their resources to adjacent users is:

$$\mathbf{M} = \left(\frac{a_{\beta i}}{K(I_\beta)}\right)^{\mathrm{T}}. \tag{6}$$

Matrix **M** is normalized by column vectors, where $a_{\beta i} / K(I_\beta)$ represents the proportion of resources allocated by item I_β to user U_i in the graph probability diffusion. Each column in matrix **M** represents the proportion distribution of resources diffused by the corresponding item to adjacent users.

(3) The transfer matrix for users to allocate their resources to adjacent items is:

$$\mathbf{N} = \left(\frac{a_{i\alpha}}{K(U_i)}\right)^{\mathrm{T}}. \tag{7}$$

Matrix **N** is also column vector normalization. $a_{i\alpha} / K(U_i)$ denotes the proportion of resources allocated to item I_α by user U_i in graph probability diffusion. Each column in matrix **N** represents the proportional distribution of resources that the corresponding user diffuses to adjacent items.

(4) The transition matrix of the graph probability diffusion process of item → user → item is:

$$\mathbf{W} = \mathbf{NM}, \tag{8}$$

where the element in matrix **W** is $\omega_{\alpha\beta}$ which has been expounded in formula (4).

(5) After a round (two steps) of graph probability diffusion, the ultimate resource vector of items is:

$$\mathbf{F}'' = \mathbf{WF}, \tag{9}$$

where **W** is the transition matrix of the graph probability diffusion process, and **F** is the initial resource vector of items.

After several iterations of the graph probability diffusion process, the distribution of the resource reaches a steady state. At this point, the items resource vector is:

$$\mathbf{F}^* = \mathbf{WF}^*. \tag{10}$$

IV. An Example Demonstration of the Proposed Matrix Representation Method

In order to explain our proposed matrix representation method of the graph probability diffusion model more clearly, taking the user-item bipartite heterogeneous social network [7] shown in Figure 1 as an example, we give a concrete application example to testify the feasibility of the proposed matrix representation method.

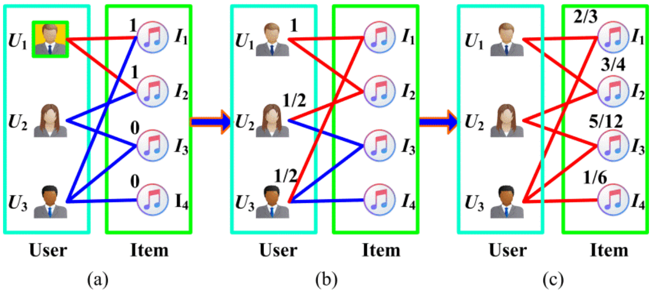

Figure 2. The process of graph probability diffusion

As shown in Figure 2(a), assuming U_1 is the intended user, a unit of resources is initially allocated to the items I_1 and I_2 that U_1 has selected. As a consequence, the initial resource vector of items is:

$$\vec{\mathbf{f}} = (1,1,0,0)^{\mathrm{T}} \tag{11}$$

Then, I_1 and I_2 evenly distribute the obtained resources to adjacent users, as shown in Figure 2(b). Thus, the resource vector of the intermediate stage of users is obtained as follows:

$$\vec{\mathbf{f}}' = (1,1/2,1/2)^{\mathrm{T}} \tag{12}$$

Then, the resources of U_1, U_2, and U_3 are redistributed back to item nodes. As shown in Figure 2(c), the ultimate resource vector of items is:

$$\vec{\mathbf{f}}'' = (2/3,3/4,5/12,1/6)^{\mathrm{T}} \tag{13}$$

The above graph probability diffusion process can simply computed with our proposed matrix representation method.

According to Figure 2, the transfer matrix of item nodes allocating resources to adjacent users is:

$$\mathbf{M} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 1 \end{pmatrix} \tag{14}$$

The transfer matrix for users to allocate resources to adjacent item nodes is:

$$\mathbf{N} = \begin{pmatrix} 1/2 & 0 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/3 \\ 0 & 0 & 1/3 \end{pmatrix} \tag{15}$$

Therefore, the transition matrix of the graph probability diffusion process is:

$$\mathbf{W} = \mathbf{NM} = \begin{pmatrix} 5/12 & 1/4 & 1/6 & 1/3 \\ 1/4 & 1/2 & 1/4 & 0 \\ 1/6 & 1/4 & 5/12 & 1/3 \\ 1/6 & 0 & 1/6 & 1/3 \end{pmatrix} \tag{16}$$

The initial resource vector of the item node in the example is: $\vec{\mathbf{f}} = (1,1,0,0)^{\mathrm{T}}$

After a round of graph probability diffusion of item → user → item, the ultimate resource vector of items is:

$$\vec{\mathbf{f}}'' = \mathbf{W}\vec{\mathbf{f}} = \begin{pmatrix} 5/12 & 1/4 & 1/6 & 1/3 \\ 1/4 & 1/2 & 1/4 & 0 \\ 1/6 & 1/4 & 5/12 & 1/3 \\ 1/6 & 0 & 1/6 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 2/3 \\ 3/4 \\ 5/12 \\ 1/6 \end{pmatrix} \tag{17}$$

Ultimately, I_3 is recommended to the intended user U_1 because its ultimate resource score (5/12) is higher than the ultimate resource score of I_4 (1/6).

V. Conclusion and Discussion

In order to unify the representation and facilitate the calculation of the graph probability diffusion model, we propose a matrix representation method of the model from the view of matrix theory. Firstly, the diffusion and recommendation process of the model are described in detail, and then, the graph probability diffusion model is theoretically summarized and unified from the perspective of matrix representation. Finally, a concrete application example is given, which shows the efficiency and feasibility of this method. Similar to the matrix representation of the graph probability diffusion model, all graph diffusion algorithms, including the graph activation diffusion, the graph heat diffusion, and the random walk model can also be represented and calculated by the matrix theory. Importantly, this method also provides a solid theoretical basis for the future research on dynamic incremental graph diffusion models.

References

[1] Zhou T, Kuscsik Z, Liu J G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of the Sciences of the United States of America, 2010, 107(10): 4511-4515. https://doi.org/10.1073/pnas.1000488107
[2] Lü L, Medo M, Yeung C H, et al. Recommender systems. Physics Reports, 2012, 519(1): 1-49. https://doi.org/10.1016/j.physrep.2012.02.006
[3] Yu F, Zeng A, Gillard S, et al. Network-based recommendation algorithms: A review [J]. Physica A: Statistical Mechanics and its Applications, 2016, 452: 192-208. https://doi.org/10.1016/j.physa.2016.02.021
[4] Campana M G, Delmastro F. Recommender Systems for Online and Mobile Social Networks: A survey [J]. Online Social Networks and Media, 2017, 3-4: 75-97. https://doi.org/10.1016/j.osnem.2017.10.005
[5] Zare H, Nikooie M A, Moradi P. Enhanced recommender system using predictive network approach. Physica A: Statistical Mechanics and its Applications, 2019, 520: 322-337. https://doi.org/10.1016/j.physa.2019.01.053
[6] Alhijawi B, Obeid N, Awajan A, et al. New hybrid semantic-based collaborative filtering recommender systems. International Journal of Information Technology, 2022, 14: 1-7(*Early Access*). https://doi.org/10.1007/s41870-022-01011-x
[7] Wang Y, Han L, Qian G, et al. Personalized recommendation via multi-dimensional meta-paths temporal graph probabilistic spreading. Information Processing & Management, 2022, 59(1): 102787. https://doi.org/10.1016/j.ipm.2021.102787

This research was funded by the Natural Science Foundation of the Colleges and Universities in Anhui Province of China under Grant No. KJ2021A0640. *Yang Wang, Ph. D., associate professor, master tutor. His research fields include intelligent information processing, machine learning, and recommendation system, etc. E-mail: wangy@aqnu.edu.cn.*