Helper Recommendation with seniority control in Online Health Community

Junruo Gao* Chen Ling† Carl Yang‡ Liang Zhao §

Abstract

Online health communities (OHCs) provide an essential platform for patients with similar health conditions to share experiences and offer moral support. However, many time-sensitive questions from patients often remain unanswered due to the multitude of threads and the random nature of patient visits in OHCs. Traditional recommendation systems solely based on similarity for recommendations cannot be directly applied in OHCs. They tend to overlook the influence of patients' dynamically changing features (e.g., health stages), affecting their ability to provide meaningful responses to questions. To address this, we propose a novel recommender system scenario designed for OHCs, which differs from traditional recommender systems in several ways. Firstly, it's challenging to model the social support factors that form helper-seeker links in OHCs. Secondly, the impact of patients' historical activities is complex to quantify. Lastly, ensuring recommended helpers have the requisite expertise is crucial. To overcome these challenges, we develop a Monotonically regularIzed diseNTangled Variational Autoencoders (MINT) model. This model formulates interactions between seekers and helpers as a dynamic graph, using encoded historical activities as node features. We also introduce a graph-based disentangle VAE to capture patient features and a monotonic regularizer to ensure the logical pairing of seekers and helpers. Our extensive experiments show the effectiveness of our approach.

1 Introduction

Online health communities (OHCs) have emerged as a growing platform for patients and their families to acquire knowledge about illnesses, facilitate the exchange of information and experiences, and connect with individuals who have undergone similar health challenges [17]. Unlike general online communities, OHCs can take diverse forms, including blogs, forums, and social media platforms [22], enabling patients to engage in various forms of asynchronous social communication. Within OHCs, patients have access to various discus-

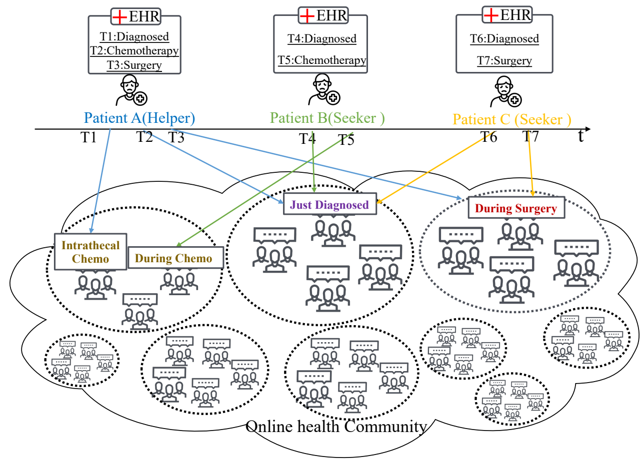

Figure 1: The dynamic activities of patients in OHCs. Patient A went through more health stages and visited more threads that would help Patient B and Patient C answer their questions on corresponding threads.

sion threads (e.g., *During Surgery* and *During Chemo* in Fig. 1) where they can seek advice on illness-related questions or help others by sharing their own experiences and emotions. However, due to the asynchronous nature of OHCs and the randomness of patient visits, there is often a long waiting time for many questions to be answered. To enhance communication efficiency and effectiveness within OHCs, it is crucial to develop a recommender system that can connect patients seeking urgent health-related advice (i.e., solution seekers) with experienced patients or healthcare providers (i.e., problem helpers).

Most existing recommendation algorithms [10, 20] have made success in modeling collaborative filtering relationships between users and items. However, recommending appropriate problem helpers to solution seekers requires capturing the affinity between users and their health stages, as well as their interactions with visited threads. These aspects go beyond what classical user-item algorithms can effectively capture. Another research avenue [5,12,18] is primarily focused on recommending friends within social networks. By leveraging existing social relationships, these approaches frame the recommendation problem as a link prediction task. For instance, researchers [21] propose to model the users' historical interests and build implicit relationships by

*gaojr1@chinatelecom.cn, Department of Intelligent Edge Cloud, Chinatelecom Cloud Technology Co., Ltd

†chen.ling@emory.edu, Emory University

‡j.carlyang@emory.edu, Emory University

§liang.zhao@emory.edu, Emory University

computing the similarity for friendship recommendations. However, the helper recommendation task in OHCs cannot be adequately addressed as a simple link prediction problem, as a more intricate logical relationship exists between solution seekers and problem helpers. Specifically, recommended helpers in OHCs must possess specific experience with the relevant illness to effectively respond to the queries posed by seekers. Consequently, the helper recommendation task cannot be satisfactorily solved solely by the conventional link prediction problem. A more intricate logical relationship exists, where the suitability of a helper to address a seeker's questions depends on their particular experience. For instance, as depicted in Fig. 1, even though both patient B and patient C have previously visited the thread *Just Diagnosed* before and may appear similar, patient B can not adequately address the questions posed by patient C in the thread *To do list for Surgery*.

Given the distinctive recommendation scenario in OHCs, the straightforward adaptation of existing recommendation algorithms to pair "problem helpers" with "solution seekers" faces several significant obstacles. *1). Difficulty in modeling the implicit interaction between problem helpers and solution seekers due to heterogeneous factors.* Unlike traditional user-item recommender systems, the interactions between problem helpers and solution seekers in OHCs are influenced by diverse and heterogeneous factors, including historical activities and experienced health stages. These factors intricately shape the nature of patient interactions, making the modeling of social support within OHCs a complex task. *2). Difficulty in distinguishing the influence of historical activities in OHCs to comprehensively characterize patients.* Patients tend to have a combination of static and evolving features, encompassing inherent knowledge about their illness as well as dynamically acquired knowledge through their interactions with various threads over time. Properly discerning the impact of historical activities is crucial for comprehensive patient characterization, as it necessitates considering both time-invariant and time-varying features. However, disentangling these features during the modeling process poses a significant challenge. *3). Difficulties in ensuring the competence of predicted helpers in addressing seekers' questions.* In OHCs, relying solely on similarity measures between patients is inadequate for accurate seeker-helper recommendations, as there exist logical seniority orders [23] as Fig. 1 shows. The existing recommendation algorithms are currently unable to incorporate such logical seniority in a straightforward manner. Additionally, employing a simple filter based on primary patients is suboptimal for an end-to-end system, as different patients possess varying levels of expertise, making it time-consuming to filter for each seeker before making recommendations.

To cope with these challenges, we develop a novel approach called **M**onotonically regular**I**zed dise**NT**angled VAE (**MINT**) for the recommendation of problem helpers to solution seekers in OHCs. MINT is specifically designed to overcome the identified obstacles by incorporating three key components. Specifically, we formulate the interactions between seekers and helpers as a dynamic graph to deal with the first challenge and consider the modeled historical information as patient node features. To tackle the second challenge, we propose graph-based Disentangled VAE to learn time-varying and time-invariant features. Lastly, to solve the third challenge, we design a monotonic regularization to enforce the correspondence between time-varying features and the senior level and leverage a senior constraint to guarantee the senior logic between seekers and predicted helpers.

- **A novel framework to model the social support between patients based on a dynamic graph.** This graph-based representation allows us to capture the complex and heterogeneous nature of social support in OHCs, with the modeled historical information serving as essential patient node features.
- **A graph-based Disentangled VAE to characterize the time-invariant and time-varying features.** This disentanglement facilitates a comprehensive understanding of patients' features, capturing the dynamic evolution of their knowledge while preserving the stable aspects of their expertise.
- **A monotonic regularizer and a seniority level constraint to guarantee the logic between seekers and predicted helpers.** The monotonic regularizer enables learned time-varying features to monotonically correlate with user's seniority. The seniority constraint guarantees the recommended helpers possess the sufficient expertise to assist seekers.
- **Extensive experiments to validate the effectiveness of MINT.** The results show the superiority of our proposed model over the comparison methods by 17% and 9% on the two real-world datasets.

2 Related work

Recommender System based graph. Graph Neural Networks (GNNs) offer a promising solution by effectively leveraging the underlying graph structure of the user-item interactions, which have been widely employed in recommender systems. GNNs can be applied not only in static recommendation scenarios [10,24] but also in sequential scenarios to capture the temporal dynamics of user preferences and item interactions. By representing users and items as nodes in a graph and the sequential interactions as edges in the graph, GNN layers can learn the temporal dependencies and evolution of the nodes [3, 6, 28, 29]. This enables the model to effectively capture the evolution of user preferences and transitions between items. Furthermore, in order to learn different factors that influence user-item interactions, disentangled graph methods are proposed. They combine the principles of collaborative filtering, which utilizes user-item interactions, with disentanglement learning, which aims to disentangle underlying factors of variations in the data [25]. Zheng et al. propose CLSR to improve recommendations by understanding both short-term and long-term user interests [31]. However, only considering static user profiles is inadequate for directly modeling the dynamic accumulation of experiences in their anti-cancer process. Therefore, it is not suitable for recommendations in OHCs.

Instead of recommending specific products or items, friend recommendations, also known as social recommendations is another important aspect in the realm of online social media, as highlighted by [5, 12, 18]. The intricate information embedded in social networks can be harnessed to enhance the learning process by considering the social effects among users. This type of recommendation can enhance social interactions, foster community engagement, and expand users' social circles. Although the method proposed in [21] using graph-based method models user interests or features to compute user similarity for matching purposes, they are not applicable to seeker-helper recommendation scenarios as they neglect to incorporate logical seniority orders among patients.

Online Health Communities: Online health communities (OHCs) can naturally be formed as a heterogeneous network of user, patient, and discussion threads [13, 14], which have offered platforms for acquiring health-related information. There are various OHCs related to different diseases [1, 2, 16], and a substantial number of patients seeking healthcare assistance. To enhance the service provided to patients, several approaches have been developed to mine and analyze patients' activities within OHCs. [4] focus on capturing patients' sentiments expressed in their discussions to identify reliable medical knowledge. In order to further aid healthcare organizations in supporting patients, Gao et al. [7, 8] propose to infer patients' health stages based on their online activities. In [23], helpers in OHCs tend to possess more specialized expertise compared to seekers and propose explicitly assessing their level of expertise. Despite its significance, few methods have addressed the problem of recommendation within OHCs. Recently, health-related recom-

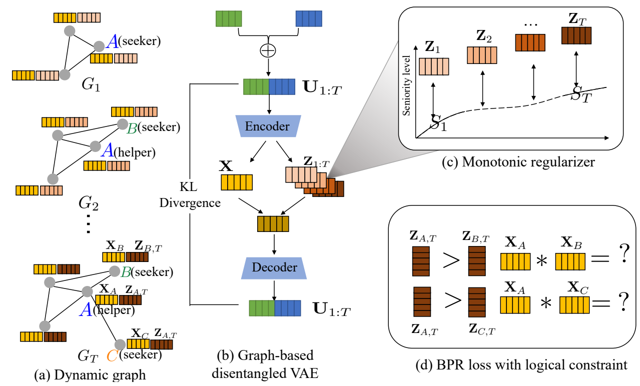

Figure 2: Overall Framework. (a) represents the dynamic graph that captures the interactions between patients. Nodes denote the patients; edge denotes the helper answering the questions on particular threads at some time. (b) shows the modeling process of the historical activities of patients to obtain time-invariant and time-varying features. (c) depicts the monotonic regularizer and seniority level constraint to guarantee the logical seniority orders. (d) shows the recommendation under logical constraint

mender systems have widely emerged to assist patients. A recent approach [27] models patients' health histories, interests, and needs to recommend suitable threads to patients. Aiming to suggest doctors to patients who are best suited to provide relevant and suitable answers to their inquiries, Liu et al. [15] develope a health recommender system by leveraging doctors' profiles and analyzing their past conversations. However, they fail to consider the social support network among patients, which is often broader in scope, more abundant, adaptable, and accessible than the support offered by healthcare professionals. It helps individuals share their experiences anytime and anywhere.

3 Proposed Method

3.1 Problem Formulation To model the unique communication dynamics among patients in OHCs, we introduce a dynamic social support graph denoted as $G = \{G_1, \cdots, G_T\}$. Each snapshot graph $G_t = (U_t, E_t)$ captures the interactions between patients up to time t, as illustrated in Fig. 2(a). Here, U_t represents the set of nodes (i.e., patients), E_t denotes the edges formed when helpers respond to questions posed by seekers on specific threads at time t, and $E_t \subseteq U_t \times U_t$. S denotes the set of seeker-helper pairs. The interactions between patients are significantly influenced by heterogeneous historical activities within OHCs, such as the threads they have visited and their experienced health stages. Upon gaining experience, she would acquire the ability to assist patients B and C in resolving their issues. To capture the influence of historical activities on patient interactions, we incorporate them as node features $\mathbf{u}_t$ at each time point t. Specifically, we denote the ordered thread sequences that patients have visited as $v_{1:T} = (v_1, \cdots, v_t, \cdots, v_T)$, and the health stage sequences that patients have gone through as $h_{1:T} = (h_1, \cdots, h_t, \cdots, h_T)$. For instance, as depicted in Fig. 1, patient A visited the thread *Diagnosed*, when she was in the health stage *Just Diagnosed*. And when she progressed to the health stage *Chemotherapy*, she developed an interest in the thread *Intrathecal chemo*. The embeddings of the thread sequences and health stage sequences can be represented as $\mathbf{v}_{1:T} = (\mathbf{v}_1, \cdots, \mathbf{v}_t, \cdots, \mathbf{v}_T)$ and $\mathbf{h}_{1:T} = (\mathbf{h}_1, \cdots, \mathbf{h}_t, \cdots, \mathbf{h}_T)$, respectively. Here, $\mathbf{v}_t \in \mathbb{R}^{1\times D_v}$ denotes the embedding of the thread visited by patients at time t, and $\mathbf{h}_t \in \mathbb{R}^{1\times D_h}$ represents the embedding of the health stage experienced by patients at time t. The dimensions of the thread and health stage embeddings are denoted as D_v and D_h, correspondingly. Thus, at each time t, the embeddings of patient nodes can be computed as $\mathbf{u}_t = concat[\mathbf{v}_t, \mathbf{h}_t]$, where $\mathbf{u}_t \in \mathbb{R}^{m\times(D_v+D_h)}$. The patient node embedding sequence is denoted as $\mathbf{u}_{1:T} = (\mathbf{u}_1, \cdots, \mathbf{u}_t, \cdots, \mathbf{u}_T)$, which records the ordered historical activities of patients. In addition, we assign a quantitative value to represent the seniority of the patient at time t, which means the patient's knowledge and experience of the disease, expressed as $s_t \in \mathbb{R}$ for the seniority levels of seekers and $o_t \in \mathbb{R}$ for the seniority levels of helpers at time t. This value can be calculated based on factors such as the number of threads visited, the number of health stages experienced, and the duration of their presence on OHCs. Similarly, we can obtain a sequence of seniority level $s_{1:T} = (s_1, \ldots, s_t, \ldots, s_T)$ and $o_{1:T} = (o_1, \ldots, o_t, \ldots, o_T)$ for different time.

Given the visited thread sequence $v_{1:T}$, health stage sequence $h_{1:T}$, senior level sequence $s_{1:T}$, and the social support dynamic graph G, the objective is to recommend suitable helpers for addressing seekers' questions. However, this problem poses significant challenges that make it highly complex to tackle: 1). Difficulty in modeling the interactions between problem helpers and solution seekers due to heterogeneous reasons. 2). Difficulty in differentiating the time-invariant and time-varying influence of historical information $v_{1:t}$, $h_{1:T}$, and $s_{1:T}$. 3). Difficulty in guaranteeing the logical seniority orders between predicted helpers and seekers.

3.2 The Objective of seeker-helper modeling

To tackle the first challenge, we introduce a framework based on graph G that models patient relationships. Node features U record patients' historical activities, while edges E represent their participation in shared threads as either seekers or helpers, as depicted in Fig. 2(a). For the second challenge, we employ a graph-based Disentangled VAE to capture both time-invariant and time-varying patient features by using historical thread sequences $v_{1:t}$ and health stage sequences $h_{1:T}$, as shown in Fig. 2(b). To address the third challenge, we introduce a monotonic regularizer to ensure that the patient's time-varying feature $\mathbf{z}_{1:T}$ aligns with their seniority level $s_{1:T}$. A seniority level constraint is applied to these time-varying features to characterize the roles of helper and seeker, detailed in Fig. 2(c-d).

To effectively train our comprehensive model, we aim to maximize the evidence lower bound (ELBO) loss of the graph-based Disentangled VAE, denoted as $\mathcal{L}_{dis}$ (refer to Eq.(3.2)), which enables the learning of both time-varying and time-invariant features. In order to capture the stability of time-varying features over short time periods, we incorporate a smoothness constraint, denoted as $\mathcal{L}_{smo}$ (refer to Eq.(3.3)), which minimizes the L_2 distances between embeddings of adjacent time steps. In addition to minimizing the Bayesian Personalized Ranking (BPR-loss) $\mathcal{L}_{bpr}$ (detailed in Eq.(3.4)), which is computed based on the time-invariant features of patients, we minimize a monotonic regularization $\mathcal{L}_{reg}$ (detailed in Eq.(3.5)) to enforce the time-varying features to have monotonic relation with seniority level. Moreover, in order to guarantee that the recommended helpers can have the ability to solve the seekers' problems, a seniority-level constraint is introduced. The overall objective function is shown as follows:

$$\min_{\Theta} \mathcal{L} = \alpha\mathcal{L}_{dis} + \gamma\mathcal{L}_{smo} + \lambda\mathcal{L}_{bpr} + \beta\mathcal{L}_{reg}, \quad \text{s.t.}, s_t < o_t, \quad t \in [1:T] \tag{3.1}$$

where α, γ, λ, β are non-negative trade-off weights for each corresponding components. These weights determine the relative contribution of each component to the overall optimization process. Θ is the set of all parameters that need to be optimized in the objective function Eq. (3.1).

3.3 Graph-based disentangled variational autoencoder

The graph-based disentangled variational autoencoder (VAE) model [30] is designed to characterize patients based on their visited threads and experienced health stages, as well as the social support graph. In this model, we assume that the sequence of patient node embeddings, denoted as $\mathbf{u}_{1:T} = (\mathbf{u}_1, \cdots, \mathbf{u}_t, \cdots, \mathbf{u}_T)$, is generated from a latent variable $\mathbf{x}$ that can be factorized into two disentangled variables: the time-invariant features $\mathbf{x}$ (representing patients' intrinsic knowledge) and the time-varying features $\mathbf{z}_{1:T}$ (representing patients' state after visiting threads at different time points). The prior distribution of the time-invariant features $\mathbf{x}$ is defined as a standard Gaussian distribution: $\mathbf{x} \sim \mathcal{N}(0,1)$. The time-varying features $\mathbf{z}_{1:T}$ follow a sequential prior, where $\mathbf{z}_t|\mathbf{z}_{<t} \sim \mathcal{N}(\mu_t, \mathrm{diag}(\sigma_t^2))$. Here, $[\mu_t, \mathrm{diag}(\sigma_t^2)] = f_\theta(\mathbf{z}_{<t})$, and μ_t and $\mathrm{diag}(\sigma_t^2)$ are the parameters of the prior distribution conditioned on all previous time-varying features $\mathbf{z}_{<t}$. The model parameter θ can be parameterized as a recurrent network, such as LSTM or GRU, where the hidden state is updated over time. We define the generating distribution of patient features conditioned on the time-invariant features $\mathbf{x}$ and time-varying features $\mathbf{z}_t$ as follows:

$$p(\mathbf{u}_{1:T}, \mathbf{z}_{1:T}, \mathbf{x}, G_{1:T}) = p(\mathbf{x}) \prod_{t=1}^{T} p(\mathbf{u}_t \mid \mathbf{x}, \mathbf{z}_t, G_t)\, p(\mathbf{z}_t|\mathbf{z}_{<t})$$

$$\mathbf{u}_t|\mathbf{z}_t, \mathbf{x}, G_t \sim \mathcal{N}(\mu_{u,t}, \mathrm{diag}(\sigma_{u,t}^2))$$

where $[\mu_{u,t}, \mathrm{diag}(\sigma_{u,t}^2)] = g_\phi(\mathbf{z}_t, \mathbf{x}|G_t)$. During the inference process, the graph-based Variational Autoencoder employs variational inference to approximate the posterior distributions. Specifically, we approximate the posterior distributions as follows:

$$\mathbf{x}|G_T \sim \mathcal{N}\left(\mu_x, \mathrm{diag}\left(\sigma_x^2\right)\right), \quad \mathbf{z}_t \sim \mathcal{N}\left(\mu_t, \mathrm{diag}\left(\sigma_t^2\right)\right)$$

$$[\mu_x, \mathrm{diag}(\sigma_x^2)] = h_\delta(\mathbf{u}_{1:T}), [\mu_{u,t}, \mathrm{diag}(\sigma_{u,t}^2)] = h_\eta(\mathbf{u}_{\le t}).$$

δ and η correspond to the recurrent encoder functions h_δ and h_η for computing the approximated posterior distributions, respectively. Specifically, the time-invariant features are conditioned on the entire sequence and encoded using h_δ, whereas the time-varying features are inferred using h_η and conditioned only on previous time points. Thus, our inference model can be factorized as:

$$q(\mathbf{z}_{1:T}, \mathbf{x} \mid \mathbf{u}_{1:T}, G_{1:T}) = q(\mathbf{x} \mid \mathbf{u}_{1:T}, G_{1:T}) \prod_{t=1}^{T} q(\mathbf{z}_t \mid \mathbf{u}_{\le t})$$

The loss of graph-based disentangled VAE in (3.1) is a timestep-wise negative variational lower bound:

(3.2)

$$\mathcal{L}_{dis} = \mathbb{E}_{q(\mathbf{z}_{1:T}, \mathbf{x}|\mathbf{u}_{1:T}, G_{1:T})}\left[-\sum_{t=1}^{T} \log p(\mathbf{u}_t \mid \mathbf{z}_t, \mathbf{x}, G_t)\right]$$
$$+ \mathrm{KL}\left(q(\mathbf{x} \mid \mathbf{u}_{1:T}, G_{1:T}) \| p(\mathbf{x} \mid G_{1:T})\right)$$
$$+ \sum_{t=1}^{T} \mathrm{KL}\left(q(\mathbf{z}_t \mid \mathbf{u}_{\le t}) \| p(\mathbf{z}_t \| p(\mathbf{z}_t \mid \mathbf{z}_{<t})\right)$$

Additionally, the smoothing constraint in Equation (3.1) can be written as follows:

$$\mathcal{L}_{smo} = \|\mathbf{z}_t - \mathbf{z}_{t-1}\|_2^2 \quad (3.3)$$

We capture the interactions between patients in terms of question-answering dynamics within the social support graph. The adjacency matrix of the social support graph is defined as follows:

$$\mathcal{A}_T = \begin{pmatrix} 0 & \mathcal{R}_T \\ \mathcal{R}_T^{\mathrm{T}} & 0 \end{pmatrix}$$

where the interaction matrix $\mathcal{R}_T \in \mathbb{R}^{m \times m}$ represents the connections between patients within the social support graph. Each entry of $\mathcal{R}$ is assigned a value of 1 if there is an interaction between the corresponding pair of patients, otherwise, it is 0. To initialize the patient nodes on the graph, we use the obtained time-invariant features $\mathbf{x}$ as the embeddings for the 0-th layer, denoted as $\mathbf{e}^{(0)} = \mathbf{x}$. Subsequently, the patient embeddings at the l-th layer can be expressed as: $\mathbf{e}^{(l)} = \left(\mathcal{D}^{-\frac{1}{2}} \mathcal{A} \mathcal{D}^{-\frac{1}{2}}\right) \mathbf{e}^{(l-1)}$, where $\mathcal{D}^{-\frac{1}{2}} \mathcal{A} \mathcal{D}^{-\frac{1}{2}}$ is used to compute the symmetrically normalized matrix. $\mathcal{D}$ represents a $2m \times 2m$ degree matrix. Each entry $\mathcal{D}_{ii}$ in $\mathcal{D}$ corresponds to the number of nonzero entries in the i-th row vector of the adjacency matrix $\mathcal{A}$. As a result, the final time-invariant features of patients can be represented as: $\mathbf{e} = \frac{1}{L} \sum_{l=0}^{l=L} \mathbf{e}^{(l)}$.

3.4 Recommending under logical seniority orders

The learned patient embeddings $\mathbf{e}$ capture both the intrinsic time-invariant information and collaborative signals. To differentiate between seekers and helpers, we utilize $\mathbf{e}_p \in \mathbb{R}^{a \times D_p}$ and $\mathbf{e}_q \in \mathbb{R}^{b \times D_q}$, where $a \le m$, $b \le m$, D_p, and D_q represent the dimensionality of the time-invariant features. Moreover, the roles of seekers and helpers are determined by their activities, specifically whether they answer or ask questions on the threads within OHCs. The BPR-loss in Eq. (3.1) can be expressed as:

$$\mathcal{L}_{bpr} = -\sum_{(a,+,-) \in S} \ln \sigma\left(\hat{r}_{a,+} - \hat{r}_{a,-}\right) \quad (3.4)$$

where $\hat{r}_{a,+} = \mathbf{e}_p^{(a)} \mathbf{e}_q^{(+)^T}$ calculates the similarity score between a seeker and a positive sampled helper, $\hat{r}_{a,-} = \mathbf{e}_p^{(a)} \mathbf{e}_q^{(-)^T}$ computes the similarity between seekers and negatively sampled patients. However, while the BPR-loss can learn the similarities between seekers and helpers, it does not guarantee that the recommended helpers possess sufficient experience to assist the seekers. In order to address this, we introduce a monotonic regularizer to enforce a monotonic correlation between the time-varying features $\mathbf{z}_t$ and the seniority level. To derive the regularizer, we refer to the standard definition of monotonicity [19]: if $f(x) \ge f(y)$, then $x \ge y$, where x and y represent any latent variables. However, directly computing such a non-linear relationship is challenging. Instead, we penalize the following term in the objective as an equivalent formulation: $\max(0, -(f(x) - f(y)) \cdot (x - y))$. Take seekers as an example, in our specific context, the monotonic correlation is implemented through the regularizer term $\mathcal{L}_{reg}$, which is defined as:

$$\mathcal{L}_{reg} = \sum_{i}^{n} \mathrm{ReLU}[(s_t - s_{t-1}) * ([\mathbf{z}_{t-1}]_i - [\mathbf{z}_t]_i)] \quad (3.5)$$

where n represents the embedding size of time-varying features and i denotes each entry of the embeddings, ReLU [9] is the activation function. For helpers, s can be replaced by o. The regularizer, which ensures the monotonic correlation, can be incorporated into the feature learning loss given in Eq. (3.2). This addition strengthens the training objective by encouraging the desired monotonic behavior in the time-varying features. Building upon the monotonically regularized time-varying features, we further introduce a seniority-level constraint to enforce the logical ordering between the recommended helpers and seekers. That is, at the same time t, each dimension of the helpers must be larger than each dimension of the corresponding seekers. This constraint between the pairs seeker p and the helper q is formulated as: $[\mathbf{z}_{p,t}]_i < [\mathbf{z}_{q,t}]_i$, where $[\mathbf{z}_{p,t}]_i$, $[\mathbf{z}_{q,t}]_i$ denote the i-th entry of the time-varying features for seeker and helper, respectively. A computable form of this constraint can be written as:

$$\mathcal{L}_{cons} = \sum_{\{p,q\}\in S} \sum_{i} ([\mathbf{z}_{p,t}]_i - [\mathbf{z}_{q,t}]_i) \tag{3.6}$$

We put $\mathcal{L}_{cons}$ into $\mathcal{L}_{reg}$, which is used to train the model together. By enforcing the time-varying features to monotonically increase with the seniority level, this constraint ensures that the recommended helpers possess the necessary capabilities to address the problems of the seekers.

4 Experiments

Our proposed method, MINT, is evaluated on two real-world OHC datasets along with four state-of-the-art comparison methods. Furthermore, we analyze the trade-off between efficiency and accuracy, providing insights into the balance achieved by our approach. Sensitivity analysis and ablation studies are also performed to examine the impact of hyperparameters and different components of the proposed method. Lastly, a case study was conducted, highlighting the necessity of the recommendation for seekers in OHCs.

4.1 Dataset and Data preparation The Breast Cancer Community is one of the largest OHCs, serving as a valuable resource for patients to gain disease knowledge, seek and offer social support, and connect with others in similar health stages [7]. Patient interactions within the community occur in various threads consisting of questions and answers. The dataset, covering 2014 to 2018, captures patients' evolving health stages as they engage with different threads over time. In totoal, this community contains $3,948$ patients, 719 seekers, $3,827$ helpers, and $16,360$ interactions.

The Bladder Cancer Community, another prominent OHC, focuses on bladder cancer information exchange. Its dataset, spanning 2006 to 2021, parallels the Breast Cancer Community dataset in terms of informational features. In both cases, we excluded threads with fewer than ten patient visits and omitted non-helpful interactions. Patients' thread visits and interactions were organized chronologically. In totoal, this community contains 296 patients, 189 seekers, 243 helpers, and $9,867$ interactions.

We also filtered out irrelevant replies and identified helper and seeker roles based on patients' activities, such as asking and answering questions and offering both professional and emotional support. Importantly, patients' roles are not static. Initially, early-stage cancer patients are more likely to be seekers, but over time or upon recovery, they often transition to helpers. The datasets were divided into training, validation, and testing sets at an 80%, 10%, and 10% ratio, respectively.

4.2 Experiment Setup

4.2.1 Comparison methods. To demonstrate the effectiveness of our proposed MINT method, we conducted comparisons with state-of-the-art methods from three distinct categories. 1) *Traditional method:* BPR-MF [20] is a well-established recommendation algorithm known for its strong performance across various recommendation tasks. 2) *Sequential methods:* SASRec [11] has shown promising performance in sequential recommendation tasks, particularly in scenarios where temporal dependencies and complex patterns are crucial. 3) *Graph-based methods:* NGCF [24] and LightGCN [10] leverage the bipartite graph structure of user-item interactions to learn user and item embeddings, facilitating effective recommendations. DGCF [25] seeks to overcome limitations in traditional methods by explicitly modeling and disentangling different aspects of user preferences and item features, but ignores the temporal information. GraFRank [21] stands as the pioneering work in exploring the utilization of GNNs for modeling social user-user interactions and recommendations. It effectively captures expressive user representations through the integration of multiple feature modalities and user-user interactions.

4.2.2 Implementation Details. In the MINT model, we used two LSTMs to encode time-varying features and one LSTM coupled with a single-layer perceptron for time-invariant features. The decoder included two-layer perceptrons with non-linear transformations. Patient embeddings were created by concatenating visited thread and health stage embeddings into a 16-dimensional space. The dimensions of both time-varying features $\mathbf{z}_t$ and time-invariant features $\mathbf{x}$ were set to 8. A graph layer of 3 was found to yield the best results. For training, we used a batch size of 256

Table 1: Experimental results on Breast Cancer Community Dataset and Bladder Cancer Community Dataset with respect to four evaluation metrics. "N" is the abbreviation of NDCG (Normalized Discounted Cumulative Gain), and "H" is the abbreviation of HIT.

Method	Breast Cancer Community Dataset							Bladder Cancer Community Dataset						
	N@3	H@3	NG@5	H@5	N@10	H@10	MRR	N@3	H@3	N@5	H@5	N@10	H@10	MRR
MF	0.0151	0.0204	0.0177	0.0267	0.0220	0.0402	0.0215	0.1672	0.2110	0.1972	0.2836	0.2345	0.3991	0.1998
SASRec	0.0144	0.0149	0.0162	0.0194	0.0223	0.0388	0.0241	0.2156	0.2604	0.2304	0.2959	0.2400	0.3254	0.2217
NGCF	0.0096	0.0124	0.0113	0.0165	0.0145	0.0264	0.0144	0.1809	0.1298	0.3212	0.1862	0.5274	0.2542	0.1820
LightGCN	0.0139	0.0177	0.0167	0.0245	0.0225	0.0425	0.0220	0.2556	0.3249	0.2941	0.4200	0.3295	0.5272	0.2781
DGCF	0.0132	0.0122	0.0152	0.0138	0.0201	0.0366	0.0149	0.3511	0.2960	0.4383	0.4240	0.5652	0.6329	0.4418
GraFRank	0.0187	0.0248	0.0216	0.0321	0.0267	0.0458	0.0258	0.4618	**0.5522**	0.4380	0.5881	0.4851	0.6135	0.4467
Ours	**0.0244**	**0.0281**	**0.0270**	**0.0345**	**0.0327**	**0.0526**	**0.0278**	**0.4861**	0.5396	**0.5120**	**0.6029**	**0.5505**	**0.7194**	**0.5071**

and employed the Adam optimizer with learning rate: 0.001. The hyperparameter λ for BPR-loss was set at 1, and the smoothing coefficient γ was set at 0.1. The sensitivity of parameters α and β is discussed in Sec. 4.5. In comparison, baseline models like BPR-MF, NGCF, DGCF, LightGCN, and GrafRank were primarily focused on collaborative filtering and did not account for patients' historical activities. To maintain a fair comparison, we integrated LSTM to capture this historical information for initial feature representations in these models. In SASRec, the maximum sequence length was set to 10, roughly equivalent to the average number of patient interactions. The model also used 16-dimensional embeddings, a batch size of 256, and the Adam optimizer with a learning rate of 0.001. For graph-based models, the layer parameter was set to 3, in line with recommendations in their respective papers. DGCF's disentanglement factors were set to 4.

4.2.3 Evaluation Metrics.

We adopt Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain (NDCG@K), and HIT@K as main evaluation metrics [26]. Specifically, the MRR measures the rank of the ground truth entity relative to all other entities. NDCG@K evaluates the relevance of recommended entities based on their positions in the ranking list, considering the discounted cumulative gain. Lastly, HIT@K indicates the proportion of correct recommendations within top K positions.

4.3 Results

The comprehensive performance evaluation of both the compared methods and the proposed method, MINT, is presented in Table 1 for both datasets. During the validation phase, MINT achieved the highest performance on the Breast Cancer Community dataset when the parameters α and β were set to 0.01 and 0.001, respectively. For the Bladder Cancer Community dataset, the best performance was obtained with parameter values of $\alpha = 0.001$ and $\beta = 0.01$. Statistical analysis reveals that the performance improvement achieved by the proposed method over each comparison method is statistically significant at a significance level of 5%, except for the prediction results on the online Bladder Cancer Community dataset, where HIT@3 does not show significant improvement.

Table 2: Ablation Study on the two OHC datasets.

Method	Breast Cancer Community			
	NDCG@5	HIT@5	NDCG@10	HIT@10
w/ V	0.0192	0.0274	0.0248	0.0449
w/o S	0.0196	0.0275	0.0253	0.0429
Ours	0.0270	0.0345	0.0327	0.0526
Method	**Bladder Cancer Community**			
	NDCG@5	HIT@5	NDCG@10	HIT@10
w/ V	0.4971	0.5753	0.5301	0.6735
w/o S	0.5012	0.5793	0.5331	0.6747
Ours	0.5120	0.6029	0.5505	0.7194

MINT excels across various evaluation metrics, outpacing other methods. Specifically, it shows relative gains of 17.33% and 9.52% across all metrics for the Breast Cancer and Bladder Cancer Community datasets, respectively. NGCF, a graph-based approach, fares the worst, likely due to the negative impact of non-linear activation functions and feature transformation matrices on prediction performance. BPR-MF underperforms compared to LightGCN, highlighting the benefit of incorporating high-order information. SASRec outdoes BPR-MF by considering time-serial features but falls short of MINT because it lacks a comprehensive graph neural network and has a limited focus on temporal factors. DGCF and GrafRank perform well by modeling features from various angles, yet they don't surpass MINT. GrafRank, despite its attention mechanisms for capturing diverse content, ranks highest among comparison algorithms but still falls behind MINT. Its limitations are evident in the complex task of seeker-helper recommendations in OHCs and in its inability to effectively use historical information to inform patient interactions.

4.4 Ablation Study

We further conduct the ablation study to investigate the importance of each component of MINT. We present two variants of the proposed MINT: (1) For the first ablated model, instead of

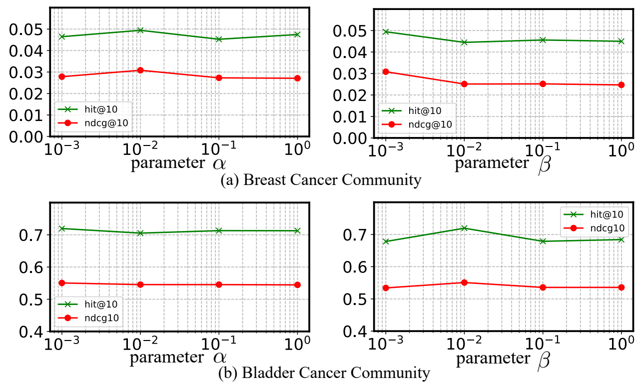

Figure 3: The sensitivity of parameter α and β for MINT. When tuning α, β was fixed to 0.001 for Breast Cancer OHC and 0.01 for Bladder Cancer OHC. When tuning α, β was fixed to 0.01 for Breast Cancer OHC and 0.001 for Bladder Cancer OHC.

modeling the time-invariant and time-varying features of patients, we only learn the time-varying features with VAE and apply monotonic regularization, noted as w/ V. (2) For the second ablated model, we do not use the monotonic regularizer and seniority level constraint to enforce the learning process of time-varying features, noted as w/o S. Table 2 shows performance metrics such as NDCG@5, HIT@5, NDCG@10, and HIT@10 for three models in both Breast Cancer and Bladder Cancer Communities. The results indicate that removing any components from our proposed MINT model leads to performance degradation. Using only the VAE without incorporating time-invariant and time-varying features fails to capture patients' intrinsic knowledge, which is crucial for calculating similarity. While the w/o S model performs comparably on the Bladder Cancer Community dataset, as seen in Table 1, it lacks a monotonic regularizer and seniority-level constraint. This limitation makes it difficult to ensure that the recommended patient is actually equipped to answer the question, leading to consistent underperformance compared to the default settings.

4.5 Impact of Hyper-parameters In our parameter analysis, we focus on the sensitivity of two key coefficients, α and β, and their impact on MINT's performance. The performance metrics used are NDCG@10 and HIT@10, and the results are presented in Fig. 3. For the Breast Cancer Community dataset, initial settings of $\beta = 0.001$ and $\alpha = 0.01$ show consistent results across varying parameter values, confirming the model's robustness. Similarly, the Bladder Cancer Community dataset reveals optimal performance when α and β are individually set to 0.001 and 0.01.The ablation study in Table 2 underscores the importance of disentangled

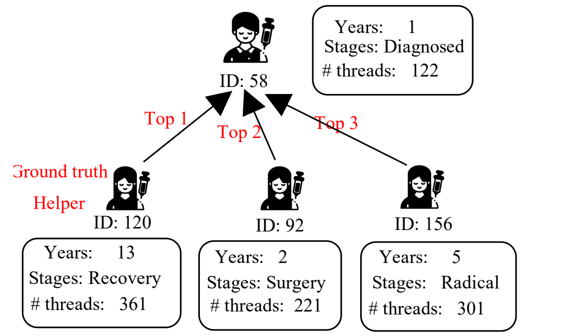

Figure 4: 'Years', 'Stages', and '#thread' respectively represent how long the patient has been on OHC, the health stage during questioning or answering, and the number of threads visited.

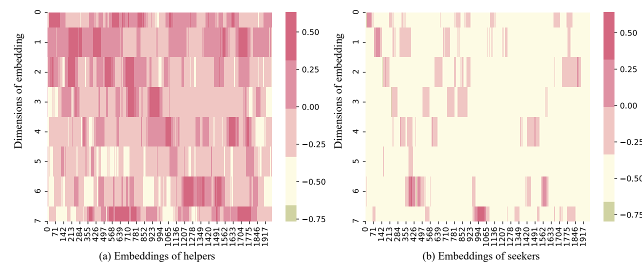

Figure 5: The visualization of the seeker and helper representations under logical seniority orders.

feature learning and seniority-level constraints in the MINT model. While BPR-Loss is a foundational component, these additional features ensure that the model doesn't solely focus on similarity measures, but also considers the hierarchical relationships between seekers and helpers. This is crucial for real-world applications. Overall, these added components significantly enhance the performance and relevance of the MINT approach.

4.6 Case Study MINT not only excels in performance but also ensures that recommended helpers are capable of assisting seekers. Fig. 4 presents a case study featuring a patient (ID: 58) who joined the OHC a year ago and is in *Diagnosed* health stage. MINT successfully identifies the best-suited helper (patient ID: 120) who has recovered from the illness with rich anti-cancer experience. Thus, he is the most suitable helper with the highest seniority. In Fig. 5, visualizations of seeker and helper representations confirm that MINT maintains the seniority order between recommended helpers and seekers, demonstrating the effectiveness in matching appropriate helpers with seekers.

5 Conclusion

To address the critical need for reliable healthcare information in OHCs, we design a seeker-helper recommender system specifically tailored for OHCs. Unlike traditional user-item recommendation methods, our framework captures the complex social dynamics among patients and utilizes the diverse data available in OHCs. The proposed model introduces a Graph-based Disentangled VAE that handles both static and evolving patient information. We also include a monotonic regularizer to ensure the stable evolution of time-varying features. A seniority-level constraint is integrated to maintain a logical expertise hierarchy among users. By incorporating these elements, our recommender system effectively tackles OHC-specific challenges and identifies experienced helpers to answer seekers' questions.

References

[1] Bladder, *Bladder*. Available at `https://bladdercancersupport.org`, 2005.

[2] Breast, *Breast*. Available at `https://community.breastcancer.org/`, 1991.

[3] J. Chang, C. Gao, Y. Zheng, Y. Hui, Y. Niu, Y. Song, D. Jin, and Y. Li, *Sequential recommendation with graph neural networks*, in SIGIR, 2021, pp. 378–387.

[4] S. Chen, X. Guo, T. Wu, and X. Ju, *Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis*, Inf. Process. Manag., (2020), p. 102253.

[5] S. Cheng, B. Zhang, G. Zou, M. Huang, and Z. Zhang, *Friend recommendation in social networks based on multi-source information fusion*, ICMLC, (2019), pp. 1003–1024.

[6] Z. Fan, Z. Liu, J. Zhang, Y. Xiong, L. Zheng, and P. S. Yu, *Continuous-time sequential recommendation with temporal graph collaborative transformer*, in CIKM, 2021, pp. 433–442.

[7] Y. Gao, T. Chowdhury, L. Wu, and L. Zhao, *Modeling health stage development of patients with dynamic attributed graphs in online health communities*, TKDE, (2022).

[8] Y. Gao, L. Wu, H. Homayoun, and L. Zhao, *Dyngraph2seq: Dynamic-graph-to-sequence interpretable learning for health stage prediction in online health forums*, in ICDM, 2019, pp. 1042–1047.

[9] X. Glorot, A. Bordes, and Y. Bengio, *Deep sparse rectifier neural networks*, in AISTATS, 2011.

[10] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, *Lightgcn: Simplifying and powering graph convolution network for recommendation*, in SIGIR, 2020, pp. 639–648.

[11] W.-C. Kang and J. McAuley, *Self-attentive sequential recommendation*, in ICDM, 2018, pp. 197–206.

[12] S. Li, X. Song, H. Lu, L. Zeng, M. Shi, and F. Liu, *Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm*, Expert Systems with Applications, 139 (2020), p. 112839.

[13] C. Ling, C. Yang, and L. Zhao, *Deep generation of heterogeneous networks*, in ICDM, 2021, pp. 379–388.

[14] ———, *Motif-guided heterogeneous graph deep generation*, KAIS, 65 (2023), pp. 3099–3124.

[15] X. Lu, Y. Zhang, J. Li, and S. Zong, *Doctor recommendation in online health forums via expertise learning*, arXiv preprint arXiv:2203.02932, (2022).

[16] MedHelp, *Medhelp*. Available at `http://www.medhelp.org`, 1982.

[17] P. K. Mo and N. S. Coulson, *Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online hiv/aids support group*, Cyberpsychology & behavior, (2008), pp. 371–374.

[18] H. Ning, S. Dhelim, and N. Aung, *Personet: Friend recommendation system based on big-five personality traits and hybrid filtering*, TCSS, (2019), pp. 394–402.

[19] M. Pemberton and N. Rau, *Mathematics for economists: an introductory textbook*, 2007.

[20] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, *Bpr: Bayesian personalized ranking from implicit feedback*, in UAI, 2009.

[21] A. Sankar, Y. Liu, J. Yu, and N. Shah, *Graph neural networks for friend ranking in large-scale social platforms*, in WebConf, 2021, pp. 2535–2546.

[22] M. Van der Eijk, M. J. Faber, J. W. Aarts, J. A. Kremer, M. Munneke, B. R. Bloem, et al., *Using online health communities to deliver patient-centered care to people with chronic conditions*, JMIR, (2013).

[23] V. Vydiswaran and M. Reddy, *Identifying peer experts in online health forums*, BMC medical informatics and decision making, 19 (2019), pp. 41–49.

[24] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, *Neural graph collaborative filtering*, in SIGIR, 2019, pp. 165–174.

[25] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua, *Disentangled graph collaborative filtering*, in SIGIR, 2020, pp. 1001–1010.

[26] W. Wu, L. He, and J. Yang, *Evaluating recommender systems*, in ICDIM, 2012, pp. 56–61.

[27] C. C. Yang and L. Jiang, *Enriching user experience in online health communities through thread recommendations and heterogeneous information network mining*, TCSS, 5 (2018), pp. 1049–1060.

[28] L. Zhang, L. Zhao, S. Qin, D. Pfoser, and C. Ling, *Tg-gan: Continuous-time temporal graph deep generative models with time-validity constraints*, in The WebConf., 2021, pp. 2104–2116.

[29] M. Zhang, S. Wu, X. Yu, Q. Liu, and L. Wang, *Dynamic graph neural networks for sequential recommendation*, TKDE, (2022).

[30] Y. Zhang, S. Gu, Y. Gao, B. Pan, X. Yang, and L. Zhao, *Magi: Multi-annotated explanation-guided learning*, in ICCV, 2023, pp. 1977–1987.

[31] Y. Zheng, C. Gao, J. Chang, Y. Niu, Y. Song, D. Jin, and Y. Li, *Disentangling long and short-term interests for recommendation*, in WebConf., 2022, pp. 2256–2267.