DynGraph2Seq: Dynamic-Graph-to-Sequence Interpretable Learning for Health Stage Prediction in Online Health Forums

Yuyang Gao
George Mason University
ygao13@gmu.edu

Lingfei Wu
IBM Research
wuli@us.ibm.com

Houman Homayoun
George Mason University
hhomayou@gmu.edu

Liang Zhao
George Mason University
lzhao9@gmu.edu

Abstract—Online health communities such as the online breast cancer forum enable patients (i.e., users) to interact and help each other within various subforums, which are subsections of the main forum devoted to specific health topics. The changing nature of the users' activities in different subforums can be strong indicators of their health status changes. This additional information could allow health-care organizations to respond promptly and provide additional help for the patient. However, modeling complex transitions of an individual user's activities among different subforums over time and learning how these correspond to his/her health stage are extremely challenging. In this paper, we first formulate the transition of user activities as a dynamic graph with multi-attributed nodes, then formalize the health stage inference task as a dynamic graph-to-sequence learning problem, and hence propose a novel dynamic graph-to-sequence neural networks architecture (DynGraph2Seq) to address all the challenges. Our proposed DynGraph2Seq model consists of a novel dynamic graph encoder and an interpretable sequence decoder that learn the mapping between a sequence of time-evolving user activity graphs and a sequence of target health stages. We go on to propose dynamic graph hierarchical attention mechanisms to facilitate the necessary multi-level interpretability. A comprehensive experimental analysis of its use for a health stage prediction task demonstrates both the effectiveness and the interpretability of the proposed models.

Index Terms—deep learning, dynamic graph, sequence prediction, health stage prediction.

I. INTRODUCTION

Online healthcare forums and communities [1]–[3] such as the Breast Cancer Community have greatly changed the way patients seek health-related information and have become an important part of patients' lives. The communications and interactions between patients in online forums can provide valuable information about a patient's emotional well-being and behaviors related to the management of their health that conventional clinical data collected from hospital information systems and electronic health records (EHR) is unable to capture. The synergies between the information on patients' online communication and health status make possible a unique and wide range of research topics on health informatics [4]–[6] that rely on both patients' interactions in online forums as well as their health stage records.

However, the health stage information in the online health community has some unique challenges and characteristics. First, though some patients share their disease history, as

Fig. 1. An example of patient signature that contains cancer diagnosis and treatment history.

shown in Figure 1, such information is not provided or is simply missing for many others. For instance, over 36% active users that registered within recent 2 years have not yet shared their disease history in the Breast Cancer Community. Second, different subforums under specific topics are often correlated to specific disease stages. For example, in the online breast cancer forum, the patients who are active in the "Chemotherapy - Before, During, and After" subforum typically look for information related to their Chemotherapy treatment. Third, as the patients' health conditions progress over time, they often move from one set of subforums to others that are more related to their new health stages. Therefore, for each patient, these transitions among subforums can lead to an inter-connected subforum activity network that evolves over time, which could be highly entangled with the progress of patient's health status, as shown in Figure 2.

The ability to accurately infer users' missing health stage information is crucial, as this could enable health care organizations to better support patients by pinpointing the most valuable information for each at their particular health stage [7]. To infer the missing user health stage information, the correspondence between the users' forum activities and their health stage history needs to be accurately identified and modeled. Naturally, the networked and time-evolving forum activity data can be formulated as a dynamic sequence of user activity transition graphs that change over time. In addition, the target user health stage history can be formulated as a sequence that needs to be inferred. Thus, without loss of generality, a new generic task is presented here where the goal is to learn the mapping from a sequence of graph-structured data to a target sequence. In this paper, we limit our scope to the domain of online health forums and focus on health stage sequence prediction based on online health forums data.

However, capturing the high-level mapping between the evolution of the user activity networks and the changes in
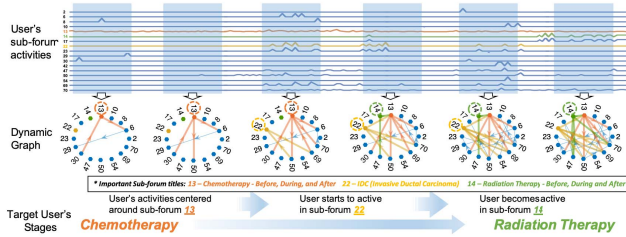
IEEE
computer
society

Fig. 2. An example of user forum activities and the corresponding health stage evolution. In the first two time windows, the user is mainly active in Subforum #13 while going through chemotherapy treatment. In the third time window, the user starts to be active in Subforum #22 at about the time when she undergoes IDC treatment. Finally, in the last three time windows, the user becomes active in Subforum #14 when she enters the "Radiation Therapy" health treatment stage.

the corresponding user's health stage can be very difficult due to the following challenges: **1) Difficulty in modeling the forum data, which is dynamic, networked, and multi-attributed.** A user's activities in the various subforums can change dynamically over time and these activity transitions naturally bridge different subforums. **2) Difficulty in learning the association between a sequence of user activity networks and the corresponding sequence of health stages**. The sequence of user activity networks contains complicated graph-structured information that dynamically evolves over time. Developing end-to-end learning between such dynamic complex data and a specific sequence is highly difficult. **3) Lack of interpretability of the health stage sequence inference process.** The sequence of user activity networks has a two-level hierarchical structure, namely node (i.e., subforum) to network level, and network to health stage level. It is thus a major objective to incorporate this hierarchical structural information into the development of an interpretable health stage inference process.

In this paper, we formally define the generic learning problem of health stage sequence inference using online forum data and propose the first framework to address the aforementioned challenges effectively. The contribution of this paper is four-fold: 1) we define the health stage inference problem in online health forums and formulate the user activities as transition graphs that are capable of modeling user dynamic transitions between subforums and their complex relationships; 2) we propose a novel deep neural encoder-decoder framework for learning the mapping between complex dynamic graph sequence inputs and the target output sequence; 3) we propose a new dynamic graph hierarchical attention mechanism that captures both the time-level and node-level attention, thus providing model transparency throughout the whole inference process; 4) experiments on online health forum dataset demonstrate that our proposed models outperform conventional sequence inference methods. In addition, our qualitative analyses and case studies provide interpretable insights into the learning results of the proposed model and its variations.

II. RELATED WORK

Our model draws inspiration from the research fields of online health community analysis, dynamic graph learning,

attention mechanisms, and neural encoder-decoder models.

A. Online Health Communities Analysis

A number of studies have focused on the analysis and utilization of online health communities data. Popular social media is good for aggregate level pattern mining tasks [8], [9]. However, their power is limited for discovering individual-level health stages and health network patterns due to the privacy issues involved and data scarcity. There have been several analyses of breast cancer forum data [4], [5] and, more recently, machine learning models have been used for longitudinal analysis [6] and some binary classification tasks [7]. However, we are the first to propose a general framework that can achieve health stage sequence inference using online forum data.

B. Dynamic Graph Representation Learning

As an emerging topic in the graph representation learning domain, dynamic graph learning has attracted a great deal of attention from researchers in recent years [10]–[12]. However, these graph embedding techniques typically focus on learning representations of the graphs, such as node embedding, but in many real-world applications the aim is to learn some high-level knowledge from the graph data, such as graph classification tasks [13], [14] and graph to sequence tasks [15], [16]. An end-to-end learning model is thus needed to learn the mapping between the whole sequence of graph data and the target output sequence, instead of merely focusing on learning node representations.

C. Attention Mechanism

The attention mechanism was first proposed by [17] and has been widely used for machine translation tasks [18], [19]. The attention mechanism has also been introduced in the graph representation learning domain [20], [21]. However, there is little to no work that focuses specifically on studying the unique hierarchical structure that is naturally present in dynamic graphs.

D. Neural Encoder-Decoder Models

The neural encoder-decoder models [17], [22] have been widely extended to model the mapping of general object inputs to their corresponding sequences [23], [24]. Recent advances in graph deep learning and graph convolutional networks have enabled various graph deep learning models to handle challenges in the domains of graph generation [25]–[27] and graph-to-sequence learning [28]. However, there have been no reports of work that explores dynamic graph to sequence learning, where the natural sequential order contained in a dynamic graph and its sequences might be advantageous for neural encoder-decoder models.

III. PROBLEM FORMULATION

A. User Forum Activities as a Dynamic Graph

An activity transition network is formulated naturally as follows. User activities are first partitioned into a series of

time windows. We then begin by formulating a node for each subforum, with a transition from one forum to the other deemed to occur if the most active forum (based on visiting time or number of postings) switches from the former to the latter, creating a directed 'edge' between them. Each node (i.e., subforum) also records the user activity in the forum to build the activity transition network. Naturally, such time-ordered activity transition networks can be formally defined as dynamic graphs, also known as temporal networks in the network science literature [29], that capture the complex dynamic characteristics and time-evolving features of graphs, as defined in the following.

Definition 1. (dynamic graph). *A dynamic graph $\mathcal{G} = \{G_1, G_2, \cdots, G_T\}$ is an ordered sequence of $t = 1, \cdots, T$ separate graphs on the same set of $|V| = N$ nodes, with each snapshot graph $G_t(V, E_t)$ characterized by a weighted adjacency matrix $A_t \in \mathbb{R}^{N \times N}$ and a set of node features $F_t \in \mathbb{R}^{N \times D}$ for a given time window, where D represents the total number of node features.*

We can now formulate the activity transition networks as a dynamic graph, illustrated in Figure 2. Here, the dynamic graph contains a sequence of snapshot graphs $G_1, G_2, \cdots, G_6$ that characterize user activities in the online forum for a given time period, where G_t represents the snapshot graph $G_t(V, E_t)$ for simplicity. Each node $v \in V$ represents a subforum devoted to a specific topic and the edges E_t capture the user's movement between different subforums at a given time window shown as blue boxes. Each node v contains a set of features $F_{t,v}$ that represents the topics covered by the specific subforum. By formulating user activities as dynamic graphs, the mapping between the evolution of the user activity and the changes of user's health stages will be preserved.

B. Learning Sequence from Dynamic Graph

As we can see from Figure 2, there is a clear mapping between the evolution of the user activity dynamic graph and changes in the corresponding user's health stage. Motivated by this observation, we can formulate such problems as a general dynamic graph to sequence problem as follows:

Given a dynamic graph $\mathcal{G} = \{G_1, G_2, \cdots, G_T\}$ as input data, the goal is to predict the target sequence $S = \{s_1, s_2, \cdots, s_M\}$, where $s_m \in \mathbb{V}$ is the mth token of the output sequence in vocabulary $\mathbb{V}$; and T and M are the input graph sequence length and output sequence length, respectively. Formally, this problem is equivalent to learning a translation mapping from input dynamic graph $\mathcal{G}$ to a sequence S as $\{G_1, G_2, \cdots, G_T\} \rightarrow \{s_1, s_2, \cdots, s_M\}$.

The translation mapping problem between some source objects and target sequences has been widely studied, including both graph-to-sequence [28] and sequence-to-sequence [22], [30] formulations. However, dynamic-graph-to-sequence translation is more complex and poses several unique challenges, namely 1) Difficulty in comprehensively modeling the dynamic multi-attributed network-structured data, as both complex relationships and dynamic evolving characteristics need to
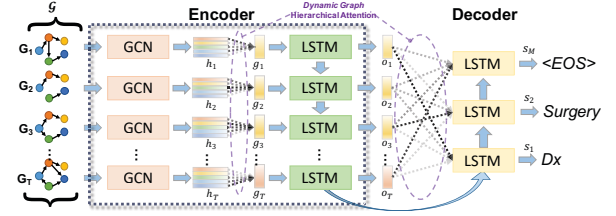

Fig. 3. The proposed end-to-end dynamic graph-to-sequence learning (Dyn-Graph2Seq) framework. It includes a novel dynamic graph encoder and a sequence decoder with dynamic graph hierarchical attention.

be captured; 2) The temporal dependency of snapshot graphs in the dynamic graph need to be modeled and constrained by the learning model; and 3) The learned translation mapping is often obscure and hard to explain or verify. This is because the original low-level representation (i.e. the node level at a specific time) is aggregated into the high-level representation (i.e. the dynamic graph as a whole), making it much more difficult to backtrack and explain the correspondence.

IV. DYNAMIC GRAPH-TO-SEQUENCE MODEL

A. Dynamic Graph Encoder

The base model of our graph convolutional network for each snapshot graph is inspired by graph2seq [28], which was originally proposed for addressing static graph-to-sequence learning problems. The Graph2Seq model employs an inductive node embedding algorithm that generates bi-directional node embeddings by aggregating information from a node local forward and backward neighborhood within k hops for a static graph. We extend this idea for dynamic graphs by applying such graph convolution on each snapshot graph within dynamic graph inputs. Specifically, suppose the total number of hops is k, then the hidden representation of n-th node in the snapshot graph G_t after applying the first graph convolutional layer will be computed as follows:

$$h_{t,n}^{\vdash} = mean(\{\sigma(W_t^{\vdash(1)} F_{t,u} + b_t^{\vdash(1)}), u \in \mathcal{N}_{\vdash}(v)\}) \quad (1)$$

$$h_{t,n}^{\dashv} = mean(\{\sigma(W_t^{\dashv(1)} F_{t,u} + b_t^{\dashv(1)}), u \in \mathcal{N}_{\dashv}(v)\}) \quad (2)$$

$$h_{t,n}^{(1)} = concat[h_{t,n}^{\vdash}, h_{t,n}^{\dashv}] \quad (3)$$

where $\mathcal{N}_{\vdash}(v)$ represents the set of forward neighbor nodes of node v, whereas $\mathcal{N}_{\dashv}(v)$ represents the set of backward neighbor nodes; $W_t^{\dashv(1)}, b_t^{\dashv(1)}$ and $W_t^{\vdash(1)}, b_t^{\vdash(1)}$ are learnable parameters for the first convolution layer. $F_{t,u}$ is the feature vector of node u in a snapshot graph at time step t; $\sigma(\cdot)$ represents the activation function of the network (e.g. ReLU); the $mean(\cdot)$ function takes the element-wise mean of the set of vectors in the equation; and $concat[vec1, vec2]$ concatenates the two row vectors into a single row vector.

Likewise, for hop k, the hidden representation of the n-th node in the snapshot graph G_t can be computed via the hidden representations computed from layer $k - 1$. Finally, after applying k layers of convolutions, the final hidden representation of the n-th node in the snapshot graph G_t will be output as $h_{t,n} = h_{t,n}^{(k)}$.

In order to capture the high-level representation of graphs for end-to-end graph learning, aggregating node level embeddings to graph level embedding that conveys the entire graph
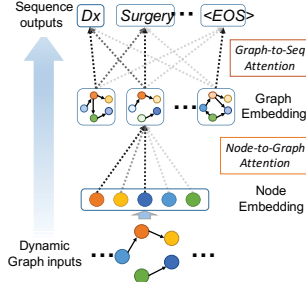
1044

Fig. 4. The proposed dynamic graph hierarchical attention mechanism: node-to-graph and graph-to-sequence attention

information is essential. To achieve this, we adopt the max pooling operation proposed by [13], [28] as the base aggregation function, which feeds the node embeddings $h_{t,n}$ to a fully-connected layer and then applies the max pooling method element-wise for each snapshot graph G_t to yield a sequence of graph-level representations g_t. To model the graph dynamic changes and long-term dependencies throughout the M steps, we utilize Long Short Term Memory (LSTM) networks [31] as a graph embedding sequence encoder to learn the entire dynamic graph-level embedding.

B. Sequence Decoder with Dynamic Graph Hierarchical Attention

Once the dynamic graph encoder takes the sequence of snapshot graphs G_t and aggregates node embeddings to generate a sequence of graph-level embeddings that capture the entire dynamic graph's global characteristics, the LSTM layer will output the final hidden-state of encoder C_T to summarize all the graph-level embeddings. Then, in the sequence decoding phase, we utilize a conventional sequence decoder [32] and set the initial cell state of the decoder as C_T in order to decode the target sequence S.

However, there are two issues with this simple sequence decoder: 1) the effectiveness of the sequence decoder depends on the length of the dynamic graph sequence; and 2) the predicted user's health stage sequence need to be interpretable based on the dynamic graph sequence at both the time-level and node-level . To handle the above questions pertaining to model interpretability, we propose a novel dynamic graph hierarchical attention mechanism that includes **node-to-graph** and **graph-to-sequence** attention that is capable of enhancing the interpretability for node embedding aggregation and capture the hierarchical structure of user online forum activities over time more effectively.

1) Node-to-Graph Attention:
Once the node embeddings of a graph have been computed, an average or max pooling operation [13], [28] is typically employed as the base aggregation function to obtain the graph-level embedding for the current graph. Although this works well in their individual settings, it does not work properly in our case since not all node embeddings contribute equally to the representation of the graph. For example, although a patient may view multiple subforums within a given time period, only a few important subforums will be correlated with the specific health stage of the patient. Therefore, it

is vital to identify these important nodes (subforums) that contribute most to representing the embedding of the current graph. Inspired by [33], we adopt the feed-forward attention to aggregate the node embeddings and formulate the graph-level embeddings. Figure 4 shows an example of how the node-to-graph attention is computed for a snapshot graph G_t. For a given snapshot graph at step t, the *node-to-graph attention* is given as follows:

$$e_{t,n} = a(h_{t,n}) \quad \alpha_{t,n} = \frac{\exp\left(e_{t,n}\right)}{\sum_{k=1}^{N} \exp\left(e_{t,k}\right)} \quad g_t = \sum_{n=1}^{N} \alpha_{t,n} h_{t,n}$$

where the function $a(\cdot)$ is a learnable function that depends on the node embeddings $h_{t,n}$; and g_t denotes the aggregated graph-level embedding for a snapshot graph at step t. In this formulation, the attention weights $\alpha_{t,n}$ explicitly model the importance of each node n when constructing the graph-level representation of g_t. Clearly, we can utilize the attention weight information for each node to pinpoint which nodes (subforums) are highly related to the current health stage. We will discuss the interpretability of our node-to-graph attention in detail in the experimental Section.

2) Graph-to-Sequence Attention:
Once the graph-level embedding g_t has been obtained for each snapshot graph G_t, the whole sequence of graph embeddings $g = \{g_i\}_{i=1}^{T}$ is fed into the sequence decoder, which generates the global hidden embedding c that characterizes the entire sequence of dynamic graph information. Following the conventional encoder-decoder setup, c is set as the initial hidden state for the sequence decoder from which to generate the target sequence of the health stages.

Although the hidden vector c theoretically contains all the information needed for generating the target sequence, the encoder's hidden representation o_t also contains valuable information about the snapshot graph information at that time step during the sequence encoding. To reward such snapshot graphs, we use the attention mechanism and introduce graph-to-sequence level attention to measure the importance of each snapshot graph with the target sequence. Specifically, as shown in Figure 4, the graph-to-sequence attention takes the sequence of hidden states for each graph $o = \{o_1, \cdots, o_T\}$ in the dynamic graph sequence as additional inputs to the decoder. This forces the decoder to consider both the current hidden state and the attention alignments between each word generated and for the whole sequence o.

V. Experiments

We evaluated the performance of our proposed model utilizing a real-world online health forum, namely the *breast cancer community*. All the experiments were conducted on a 64-bit machine with Intel(R) Xeon(R) W-2155 CPU 3.30GHz processor, 32GB memory and an NVIDIA TITAN Xp GPU.

A. Experimental Settings

Online Breast Cancer Community Dataset: The Breast Cancer Community [1] is one of the largest online forums designed for patients to share information related to breast cancer. The forum data collected for this study covers an 8

TABLE I

PERFORMANCE EVALUATION FOR HEALTH STAGE PREDICTION. THE SCORES WERE OBTAINED FROM 20 INDIVIDUAL RUNS AND PRESENTED IN A MEAN ± STANDARD DEVIATION (SD) FORMAT.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE
NMT(seq2seq) (w/o att)	55.5±2.38	38.4±0.91	27.1±0.90	19.2±0.87	71.6±1.04
NMT(seq2seq) (w/ att)	57.8±1.86	40.4±1.21	29.0±1.28	20.1±1.06	72.9±0.86
Graph2Seq (w/o att)	57.5±1.72	41.5±0.94	29.8±0.72	20.3±0.85	75.8±1.20
Graph2Seq (w/ att)	58.2±2.19	41.1±1.38	30.1±0.83	21.0±0.51	76.2±0.96
DynGraph2Seq (w/o att)	60.9±1.53	43.7±1.00	31.5±0.63	22.1±0.48	79.3±0.80
DynGraph2Seq (w/ att)	62.3±1.46	44.7±1.29	32.0±0.94	22.5±1.13	80.8±0.36

year period from the beginning of 2010 to the end of 2017. To create user subforum activity transition graph sequences, we defined user activities as being when they posted new topics or replied to existing topics and the time window was set as one month. After removing common words and stop words, we extracted the 100 top frequency keywords from the forum content to construct the feature vectors for the subforums. We randomly selected 70% of users who provided their health stage history for training, another 10% for validation, and the remaining 20% for testing. The predicted health stage sequences were validated against the real health stage history extracted from the users' signatures. The vocabulary of the health stages consists of {'Dx'[1], 'Chemotherapy', 'Targeted', 'Hormonal', 'Radiation', 'Surgery'}.

1) Evaluation Metrics:
We used BLEU scores [34] and ROUGE-1 score [35] as evaluation metrics for determining the closeness of the model predicted health stage history and the ground truth.

2) Comparison Methods:
NMT(seq2seq) The Neural Machine Translation model [32] is a widely used state-of-the-art sequence-to-sequence model for machine transition tasks. Since the NMT model can only handle simple sequence inputs, we simplified the input data by concatenating the transition sequences of user activity for each month together in time order. The subforum features are omitted in such formulations.

Graph2seq The Graph2seq model [28] is a general-purpose encoder-decoder model for static graph to sequence learning. Since the model cannot handle dynamic graphs, we simplified the input by aggregating all the edges that appeared in the dynamic graph together into a single static graph.

3) Hyper-parameter Settings:
We used the Adam optimizer [36] with a learning rate of 0.001 and a batch size of 50 for model training; greedy search was used for sequence decoders. Hyper-parameters were searched based on the highest scores achieved on the validation set.

B. Performance

Table I shows the model performance of the baseline and proposed models. The scores were obtained from 20 individual runs and presented in a mean ± standard deviation (SD) format. In general, our proposed DynGraph2Seq framework significantly outperformed both the Seq2Seq and Graph2Seq

---

[1] Short for Oncotype DX test, an initial diagnosis that analyzes how a cancer is likely to behave and respond to treatment.
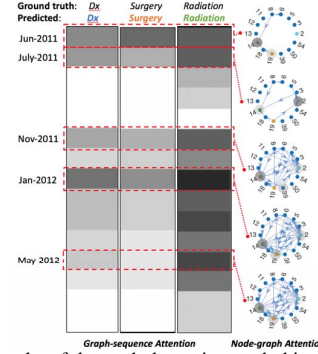

Fig. 5. An example of learned dynamic graph hierarchical attention by DynGraph2Seq. The darker the color, the greater the attention being paid.

baselines for the various model settings and evaluation metrics. The basic DynGraph2Seq framework with the proposed dynamic graph hierarchical attention achieved the best score on all the metrics, outperforming the baseline models by 7% - 17% on the BLEU scores and 6% - 13% on the ROUGE scores. The baseline Graph2Seq model also achieved good scores, but was not as competitive as our proposed model. This was largely because Graph2Seq model failed to capture the dynamic characteristics of user activity with only static graph inputs. The Seq2Seq model performed badly due to its inability to model the complex relationships between the subforums with simple sequence inputs.

C. Interpretablity Analysis

Figure 5 shows an example of the learned dynamic graph hierarchical attention by DynGraph2Seq. The left part of the figure shows the graph-to-sequence attention learned by the model, where each column is a grayscale heatmap representing the amount of attention being paid to each snapshot graph when the model predicted a specific health stage. The darker the color, the greater the attention being paid. We can see much attention was paid to the graphs around the months being labeled in the figure. The graphs for each labeled months are shown on the right. Interestingly, the graphs in the first two months attracted more attention from the model because those were the months when the patient first became active in the breast cancer online forum. The last two labeled snapshot graphs relate approximately to the time when the user engaged in extensive activities in a wide variety of subforums.

To understand why these particular snapshot graphs were important , we went one step deeper by examining the node-to-graph level attention. The red spots on the nodes shown on

the right side of Figure 5 represent the amount of attention being paid to each node (i.e. subforum). Again the darker the red spot, the greater the attention being paid. Now the attention becomes even more interesting and interpretable. For example, when constructing the representation of the May-2012 snapshot graph, Subforum #14 was assigned the most attention. The title of it is actually "Radiation Therapy - Before, During and After", which is strongly correlated to the health stage 'Radiation'. Likewise, we further discovered that Subforum #2, entitled "Not Diagnosed but Worried", has a strong correlation with 'Dx' and Subforum #19, entitled "DCIS (Ductal Carcinoma In Situ)", is a strong indicator for 'Surgery'. These observed correspondences confirm that the proposed dynamic graph hierarchical attention mechanism greatly enhances the interpretability of the model.

VI. CONCLUSION

In this paper, we formulated the task of health stage inference using online health forum data as a dynamic graph-to-sequence learning problem and propose a novel Dyn-Graph2Seq architecture that can handle this new type of learning problem effectively. Our DynGraph2Seq model consists of a novel dynamic graph encoder and an interpretable sequence decoder to learn the mapping between a sequence of time-evolving user activity graphs and a sequence of target health stages. In addition, we developed a dynamic graph hierarchical attention to facilitate the multi-level interpretability. Our comprehensive experiments and analyses for health stage prediction demonstrate both the effectiveness and the interpretability of the proposed models.

ACKNOWLEDGEMENT

REFERENCES

[1] "Breast cancer community," https://community.breastcancer.org/.

[2] "American cancer society," http://www.cancer.org.

[3] "ehealth forum," http://ehealthforum.com.

[4] N. Elhadad, S. Zhang, P. Driscoll, and S. Brody, "Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions," in *AMIA Annual Symposium Proceedings*, vol. 2014. American Medical Informatics Association, 2014, p. 516.

[5] S. Zhang, E. Bantum, J. Owen, and N. Elhadad, "Does sustained participation in an online health community affect sentiment?" in *AMIA Annual Symposium Proceedings*, vol. 2014, 2014, p. 1970.

[6] S. Zhang, E. Grave, E. Sklar, and N. Elhadad, "Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks," *Journal of biomedical informatics*, vol. 69, pp. 1–9, 2017.

[7] M. Jha and N. Elhadad, "Cancer stage prediction based on patient online discourse," in *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, 2010, pp. 64–71.

[8] L. Zhao, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting," in *KDD 2016*. ACM, 2016, pp. 2085–2094.

[9] J. Wang and L. Zhao, "Multi-instance domain adaptation for vaccine adverse event detection," in *WWW 2018*, 2018, pp. 97–106.

[10] R. Trivedi, M. Farajtbar, P. Biswal, and H. Zha, "Representation learning over dynamic graphs," *arXiv preprint arXiv:1803.04051*, 2018.

[11] L.-k. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, "Dynamic network embedding by modeling triadic closure process." in *AAAI*, 2018.

[12] P. Goyal, S. R. Chhetri, and A. Canedo, "dyngraph2vec: Capturing network dynamics using dynamic graph representation learning," *arXiv preprint arXiv:1809.02657*, 2018.

[13] J. Wu, S. Pan, X. Zhu, C. Zhang, and S. Y. Philip, "Multiple structure-view learning for graph classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 3236–3251, 2018.

[14] L. Wu, I. E.-H. Yen, Z. Zhang, K. Xu, L. Zhao, X. Peng, Y. Xia, and C. Aggarwal, "Scalable global alignment graph kernel using random features: From node embedding to graph embedding," in *KDD 2019*, 2019, pp. 1418–1428.

[15] K. Xu, L. Wu, Z. Wang, Y. Feng, and V. Sheinin, "Sql-to-text generation with graph-to-sequence model," *arXiv preprint arXiv:1809.05255*, 2018.

[16] K. Xu, L. Wu, Z. Wang, M. Yu, L. Chen, and V. Sheinin, "Exploiting rich syntactic information for semantic parsing with graph-to-sequence model," *arXiv preprint arXiv:1808.07624*, 2018.

[17] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[18] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[19] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *NAACL*, 2016, pp. 1480–1489.

[20] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, vol. 1, no. 2, 2017.

[21] J. Yang, J. Lu, S. Lee, D. Batra, and D. Parikh, "Graph r-cnn for scene graph generation," *arXiv preprint arXiv:1808.00191*, vol. 2, 2018.

[22] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[23] A. Eriguchi, K. Hashimoto, and Y. Tsuruoka, "Tree-to-sequence attentional neural machine translation," *arXiv preprint arXiv:1603.06075*, 2016.

[24] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.

[25] X. Guo, L. Wu, and L. Zhao, "Deep graph translation," *arXiv preprint arXiv:1805.09980*, 2018.

[26] M. Simonovsky and N. Komodakis, "Graphvae: Towards generation of small graphs using variational autoencoders," *arXiv preprint arXiv:1802.03480*, 2018.

[27] Y. Li, O. Vinyals, C. Dyer, R. Pascanu, and P. Battaglia, "Learning deep generative models of graphs," *arXiv preprint arXiv:1803.03324*, 2018.

[28] K. Xu, L. Wu, Z. Wang, and V. Sheinin, "Graph2seq: Graph to sequence learning with attention-based neural networks," *arXiv preprint arXiv:1804.00823*, 2018.

[29] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, and A.-L. Barabási, "The fundamental advantages of temporal networks," *Science*, vol. 358, no. 6366, pp. 1042–1046, 2017.

[30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *NIPS 2014*, 2014, pp. 3104–3112.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[32] M. Luong, E. Brevdo, and R. Zhao, "Neural machine translation (seq2seq) tutorial," *https://github.com/tensorflow/nmt*, 2017.

[33] C. Raffel and D. P. Ellis, "Feed-forward networks with attention can solve some long-term memory problems," *arXiv preprint arXiv:1512.08756*, 2015.

[34] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *ACL 2002*, 2002, pp. 311–318.

[35] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, 2004.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.