Explainable Session-based Recommendation with Meta-path Guided Instances and Self-Attention Mechanism

Jiayin Zheng
1811523@mail.nankai.edu.cn
College of Computer Science,
Nankai University
Tianjin, China

Juanyun Mai
juanyunmai@mail.nankai.edu.cn
College of Computer Science,
Nankai University
Tianjin, China

Yanlong Wen†
wenyl@nankai.edu.cn
College of Computer Science,
Nankai University
Tianjin, China

ABSTRACT

Session-based recommendation (SR) gains increasing popularity because it helps greatly maintain users' privacy. Aside from its efficacy, explainability is also critical for developing a successful SR model, since it can improve the persuasiveness of the results, the users' satisfaction, and the debugging efficiency. However, the majority of current SR models are unexplainable and even those that claim to be interpretable cannot provide clear and convincing explanations of users' intentions and how they influence the models' decisions. To solve this problem, in this research, we propose a meta-path guided model which uses path instances to capture item dependencies, explicitly reveal the underlying motives, and illustrate the entire reasoning process. To begin with, our model explores meta-path guided instances and leverages the multi-head self-attention mechanism to disclose the hidden motivations beneath these path instances. To comprehensively model the user interest and interest shifting, we search paths in both adjacent and non-adjacent items. Then, we update item representations by incorporating the user-item interactions and meta-path-based context sequentially. Compared with recent strong baselines, our method is competent to the SOTA performance on three datasets and meanwhile provides sound and clear explanations.

CCS CONCEPTS

• **Information systems** → *Data mining*.

KEYWORDS

session-based recommendation, explainable recommendation, meta-path

ACM Reference Format:

Jiayin Zheng, Juanyun Mai, and Yanlong Wen† . 2022. Explainable Session-based Recommendation vwith Meta-path Guided Instances and Self-Attention Mechanism. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3477495.3531895

† Corresponding author.

SIGIR '22, July 11–15, 2022, Madrid, Spain

ACM ISBN 978-1-4503-8732-3/22/07...$15.00
https://doi.org/10.1145/3477495.3531895

1 INTRODUCTION

Very recently users value increasingly their privacy and prefer not to log in when utilizing online platforms, which highlights the importance of anonymous session-based recommendation (SR for short) [18] [7]. Unlike traditional recommendations, SR predicts the next item based only on the interactions within a single session, without using long-term histories or user profiles [16]. Given the anonymity and short duration of the sessions, good SR models must successfully capture item relationships and mine more data [18] to compensate for the lack of user knowledge and long-range preferences. Some existing SR models use neural networks to analyze item dependencies [7] [16] and some use global session graphs [15] or item attributes [10] to enrich the information.

However, most of the existing SR models [7] [16] [10] [15], are not explainable, which weakens the persuasiveness and satisfaction of their recommendations [17]. Even though some SR methods use attention weights to reflect item importance or mine patterns such as sequential co-occurrence to explain their findings [3], their explanations cannot explicitly uncover the underlying motivations. For example, the model in [3] captures sequential patterns $i_1 \longrightarrow i_2$ but cannot explain why i_2 is often clicked after clicking i_1. Furthermore, data-mining-based explainable methods perform poorly in comparison to deep learning-based models, and their explanations are insufficient as well [3]. All these emphasize the necessity of an explainable SR model that can provide accurate predictions and clear explanations of what users' latent interests are, how the model captures them and uses them for reasoning.

To this end, we propose a path-based explainable SR model that uses pathways over knowledge graphs (KGs) as assisting information and explanations. Unlike other explainable recommendations, path-based methods don't require user profiles or long-range interactions, making it ideal for anonymous short-session recommendations. In addition, item attributes in KGs can help solve data sparsity issues and improve recommendation accuracy [10]. Furthermore, the path-based methods have better interpretability than embedding techniques such as TransE [1] and TransR [8] since we can leverage path instances to model users' motives, reason over these paths and use this reasoning process for explanations.

Despite the wealth of information gained from the KG, a path-based explainable SR approach still faces several challenges. The first one is the vast number of possible paths, which leads to a significant rise in processing time and cost. To address this problem, we employ the concept of meta-path [13] and explore meta-path schemas [2] rather than all candidate pathways. Another one is that a user's motivation can be quite complicated and comprise multiple underlying components, with interest shifts in between [18].

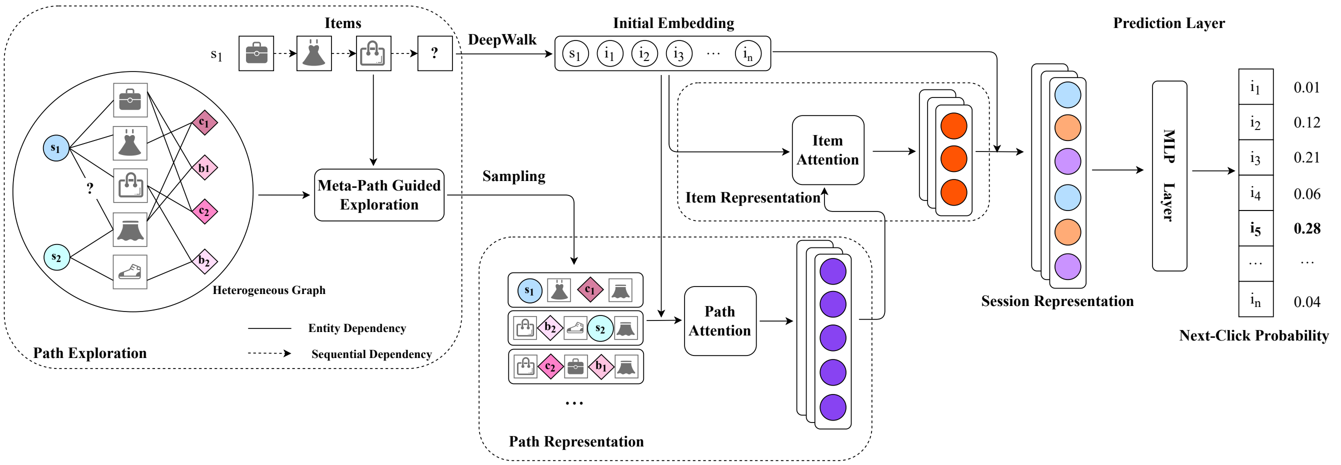

Figure 1: The overall framework of our model. It first utilizes the session histories and KG to build a heterogeneous graph, searches meta-path guided path instances over this graph and does sampling based on node similarities. Then, the model computes the path representations and uses them to update the item representations. Finally, it concatenates all item embeddings within this session to serve as the session context and produces the final probabilities.

Taking s_1 in Figure 1 as an example, the motivations of the forth click can be multi-fold: 1) The dress and skirt are of the same category. 2) The previous items and skirt share common features which match the user's preferences. 3) A user who shares a similar taste previously clicked the skirt. To represent these latent interests and their shifts (which are often overlooked by previous path-based models [2]), we explore meta-path guided paths among both adjacent and non-adjacent items, which is a highlight of our work and proven to be beneficial by the experimental results.

In summary, our main contributions are as follows:

- To the best of our knowledge, we are the first to build a meta-path guided explainable SR model which leverages the knowledge graph to explore item dependencies and the meta-path instances for explanations.
- By reasoning over meta-path-based instances between both adjacent and non-adjacent items, our algorithm can explicitly reveal multiple user latent interests, capture their shifts and assess their importance to the final results.
- Experiments on three datasets showcase our model's superiority and explainability over the SOTA models.

2 PRELIMINARIES

Definition 1. ***Heterogeneous graph.*** *In a heterogeneous graph, the nodes N and edges E are in different types, that is, |N|>1, |E|>1.*

For instance, in the product recommendation graph, the nodes could be of various types, including the items, sessions, categories and brands; and the edges could represent different relations.

Definition 2. ***Meta-path.*** *A meta-path [13] is a relation sequence connecting object pairs in a heterogeneous graph [6] and the specific path under the meta-path is called a path instance.*

For example, the meta-path *SICI* is denoted by $session \xrightarrow{includes} item \xrightarrow{belongs\ to} category \xrightarrow{has} item$, representing that the session includes the first item and that the two items are of the same category. The path $s_1 \longrightarrow i_1 \longrightarrow c_2 \longrightarrow i_2$ is a path instance under this meta-path.

3 METHODOLOGY

3.1 Path Exploration and Representation

Meta-path guided Exploration. Meta-paths [13] are particularly suited to extract different-aspect features since they can depict relational compositions between multiple types of entities in a heterogeneous graph [2]. In our scenario, the meta-paths are of two kinds: session-item meta-paths such as *SICI* and item-item meta-paths such as *ISIUI*. We first build a heterogeneous graph based on the session histories and item attributes, as shown in Figure 1, and then search meta-path guided path instances over this graph.

It's worth noting that if we simply seek item-item paths between successive items, we'll miss out on some underlying connectivities such as the one between the dress and skirt in s_1 shown in Figure 1, limiting the algorithm's ability to describe crossing latent interests. For instance, obviously, there are interest shifts between the dress, handbag and skirt. A model, if only considers paths between consecutive items, will fail to capture the user's preference for the category c_1 which is only revealed by the clicks on the handbag and skirt. Hence, we explore paths between i_k and its neighbouring itemset S = $\{i_{k+1}$ ..., $,i_{k+\epsilon}\}$ where ϵ is a hyperparameter to control the scope of modeling item-item paths. A bigger ϵ will increase the number of path instances, but a too-small ϵ will result in insufficient interest modeling.

Initialize Session and Item Embedding. The initial embedding of sessions and items is critical since it lays the foundation for path and item representation. The initialization procedure is required not only to efficiently model entities co-occurrence but also to capture global high-order relationships. Considering that DeepWalk [12], based on Word2Vec, is an efficient method to embed the nodes in a graph and optimize their co-occurrence probability, here we use it to initialize the embeddings, taking the random truncated walks in a session-item bipartite as input sentences.

Path Representation with Self-Attention. We need to learn the representations of different pathways after acquiring the meta-path guided instances and initial embeddings. Because there are many path instances under each meta-path schema, we have to first sample them. For example, if we search paths between i_1 and i_2 for the meta-path *IBICI*, there could be several intermediary items that are of the same brand with i_1 and also of the same category with i_2. Hence, we have to select from these candidate intermediary items. We assume that if the intermediary item is more similar to the start and end items, this path instance is more likely to represent more hidden user interests. Taking s_1 in Figure 1 as an example, if we explore *IBICI* paths between the first black handbag and its adjacent dress, we can find there could be many instances following this meta-path schema which represent the user's interest in the brand and the category. However, if the intermediary item of one path shares more common features with the black handbag and the dress such as are in the same color or both tweed, this path instance could also reveal the user's preference for a specific color or the tweed style. As a consequence, we filter out the nodes that are less similar to the previous nodes one by one and use the remaining pathways as the final sampling results.

We use the Word2Vec [12] technique to embed these path instances after sampling, treating path instances as sentences and nodes as tokens. The path instances are then reweighted using a multi-head self-attention technique. Self-attention is an attention mechanism that uses the relationships between distinct positions in a single sequence to compute a representation of that sequence [14], which can be used to analyze the effects of different nodes along the path. We also add a multi-head mechanism to the self-attention module to capture multiple latent interests hidden behind a single path. The calculations below demonstrate how our multi-head self-attention module works.

$$Attention(Q_\phi, K_\phi, V_\phi) = Softmax(\frac{Q_\phi K_\phi^T}{\sqrt{d_k}})V_\phi \tag{1}$$

$$MultiHead(Q_\phi, K_\phi, V_\phi) = Concat(head_1, ..., head_m)W^O \tag{2}$$

where $head_i$ is $Attention(W_i^Q Q_\phi, W_i^K K_\phi, W_i^V V_\phi)x$. Query Q, key K and value V are self-attention variables associated with path ϕ, and W is the weights. d_k is the dimensionality (here $d_k = 100$) and m is the number of items in this session. Concat(.) is the concatenation operation.

3.2 Update the Item Representation

Intuitively, three aspects determine the representation of a specific item: 1) its basic features, 2) the items in front of it, and 3) the user's latent interest behind his/her click on this item. Since we've already used DeepWalk and the path instances (meta-path-based context) to model the items' basic properties and hidden motives, the only thing left to do now is to integrate them and generate the final item representation. Considering the mutual effects between the path instances and the interactions, here we use the representations of prior items (a combination of all preceding information) and path instances (between neighboring items) together to update the item representation, allowing it to capture both long-term and short-term user interests. For example, for the third item of the session s_1 in Figure 2, since there is only one path instance connecting its prior items and itself, we use the representations of the black bag and the path in between to update the representation of the white bag. Additionally, because attention weights can reflect the importance of multiple path instances, an item-attention module is used here. Notably, to capture the temporal dependency between items, we update the item representations sequentially. The following is how our representation module works.

Table 1: Dataset Detailed Information.

Datasets	Session	Item	Attributes
Musical Instruments	11600	11457	1185 brands and 429 categories
Automotive	24000	15184	2544 brands and 567 categories
Diginetica	30000	23315	735 categories, 10 price log values and 4323 name tokens

$$h_{k,m} = ReLU(W_{k-m}h_{k-m} + W_{\phi_{k-m\to k}} h_{\phi_{k-m\to m}} + b_k) \odot h_{k-m} \tag{3}$$

$$h_k = mean - pooling(\{h_{k,m}\}_{m=1}^{\epsilon}) \tag{4}$$

where $h_{k,m}$ means the representation gained from the path instances between i_k and i_{k-m} and h_{k-m} is the item i_{k-m}'s latent representation. $\phi_{k-m\to i}$ is the instance from the $(k-m)^{th}$ item to the k^{th} item. W_i and b_i represent the weights and bias of the i-th variable.

3.3 Prediction Layer

To represent the entire session, we concatenate session embedding and item representations into a vector. It's worth noting that the item representations incorporate not only their own characteristics but also the latent interests beneath the path instances. As a result, we model item features, dependencies, and side information such as item attributes and high-order relationships all at once by constructing this vector.

$$h_{s,i} = Concat(h_s, h_1, ..., h_n) \tag{5}$$

Here $h_{s,i}$ denotes the explicit mutual vector of the session, item, and implicit effect of session-item and item-item path instances.

We use a Multilayer Perceptron (MLP) layer to generate the final next-click probability for all items [6] and the likelihood is calculated as follows. The representation of the current session and the embeddings of all items are the inputs to this layer, and the outputs are the likelihood of the items being the next-click items.

$$r_{s,i} = MLP(h_{s,i}) \tag{6}$$

The loss function is defined as the cross-entropy of the predictions:

$$\mathcal{L}_S = -\sum_{s\in\mathcal{S}}\sum_{i\in I}\{r_{si}\log(\hat{r}_{si}) + (1-r_{si})\log(1-\hat{r}_{si})\} \tag{7}$$

where y denotes the one-hot encoding vector of the ground truth.

4 EXPERIMENTS

4.1 Experimental Settings

4.1.1 Datasets. We evaluate our algorithm on three datasets, including the music instruments, automotive dataset from Amazon

Table 2: Comparison results with baselines. The best performances are in bold.

	Amazon Musical Instruments		Amazon Automotives		Diginetica	
	Hit@20	MRR@20	Hit@20	MRR@20	Hit@20	MRR@20
GCE-GNN	22.575	6.887	31.105	12.212	44.232	13.044
MKM-SR	37.818	11.382	41.224	13.892	7.970	1.643
STAMP	51.310	14.072	51.990	16.722	43.574	12.910
SR-GNN	54.391	13.267	52.050	14.575	46.821	**15.322**
Ours (ε=1)	56.620	14.695	53.767	**17.554**	46.230	10.857
Ours (ε=2)	**71.724**	**23.557**	**61.900**	17.139	**47.032**	13.521

platforms [11] and the Diginetica dataset from CIKM Cup 2016[1]. More details are shown in Table 1. These datasets especially the Diginetica datset are commonly used to evaluate SR models.

Following previous works [2], we first select users who have more than twelve interactions, order their interactions by timestamp and retain the latest twelve ones. Then we choose the first four items as bridge items, the next seven items as training items and the rest as the test items.

4.1.2 Baseline models. To demonstrate the superiority of our method, we compare it with the following SOTA models.

- **GCE-GNN** [15]: It learns two levels of item embeddings from session graph and global graph and aggregates them with a soft attention mechanism.
- **MKM-SR** [10]: It incorporates user micro-behaviors and item knowledge into multi-task learning. Similar to our work, it leverages item knowledge to serve as side information but cannot use these item attributes to explain its results.
- **STAMP** [9]: This model considers both long-term preference and current interest and uses a well designed priority neural network to model the last-clicks.
- **SR-GNN** [16]: It builds session graphs and use graph neural network to capture the complex transition of items.

4.1.3 Parameter settings. For our model, we use the Adam optimizer, setting the initial learning rate to 0.001 and the weight decay to 0.0005. To ensure a fair comparison, the batch size is set to 50 and the embedding size is set to 100 for all models.

4.2 Impact of Considering Non-adjacent Items

One contribution of our work is that we explore item-item paths in both adjacent and non-adjacent items to comprehensively capture user latent interests and their shifts. To validate its effectiveness, we conduct comparison experiments on the Amazon musical instrument dataset, as shown in Table 3. If ϵ equals 1, it means that the item-item paths are only between consecutive items. We can see that a larger ϵ leads to the Hit@20 performance improvement (26.68% and 27.60% respectively for ϵ equals 2 and ϵ equals 3) but also an increase in the number of path instances (1.91 times and 2.73 times respectively) and processing time (1.88 times and 2.71 times respectively). Therefore, to balance the computing cost and recommendation performance, we set ϵ to 2 for the model comparison in Table 2.

[1] http://cikm2016.cs.iupui.edu/cikm-cup

Table 3: Ablation study results on Amazon Musical Instrument dataset

ϵ	#item pairs	path processing time (s/1000 sessions)	Hit@20	MRR@20
1	15867	1002.83	56.620	14.695
2	30282	1887.24	71.724	23.557
3	43246	2715.90	72.247	25.588

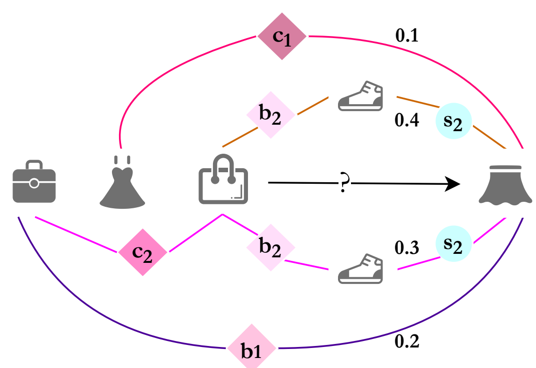

Figure 2: Case Study

4.3 Comparison Results

We utilize the metrics Top 20 Hit Ratio (Hit@20) and Mean Reciprocal Rank (MRR@20) to evaluate the performance of baselines and our model. The results are shown in Table 3.

Basically, our model outperforms all baselines on Hit@20 and the majority of baselines on MRR@20 on all three datasets. These results demonstrate that our explainable SR model can deliver sound interpretations without sacrificing recommendation accuracy. Particularly, compared with the MKM-SR model which also uses item knowledge, our model achieves higher accuracy, justifying the effectiveness of our meta-path guided instances in capturing user interests. Additionally, the advantage of exploring paths among non-adjacent items is obvious and stable across all three datasets.

Our model's MRR@20 is slightly lower than the SR-GNN on the Diginetica dataset and there might be two factors: 1) The Diginetica dataset has more sessions, allowing SR-GNN to collect more global information. 2) It's difficult to differentiate items merely based on the current item attributes since so many products share the same price log values or name tokens such as some stop words, which makes these attributes less helpful for modeling user interest.

4.4 Case Study

In this section, with illustrations on how to produce an explanation for a case randomly picked, we demonstrate our model's interpretability. Compared with other explainable SR models, our model, by leveraging path instances and attention mechanism, can explicitly reveal what the users' underlying motives are, how they shift and influence the results.

Here we still use the session s_1 in Figure 1 as an example and explain the process of predicting the probability of the skirt to be the next-click item. As depicted in Figure 2, there are four meta-path guided path instances between the first three items and the skirt, colored by four colors, representing four possible latent interests for clicking or purchasing the skirt. For example, the second path instance means that there is a pair of shoes shares the same brand with the white handbag and also belongs to the same session with the skirt to be predicted. To be more specific, if the user prefers items in b_2 and s_2, the skirt is more likely to be clicked.

Furthermore, the importance of these path instances is indicated by their attention weights. More precisely, the second path instance has the highest attention weight, implying that the most important reason for our model's judgment is that the model "guesses" that the user prefers items from the brand b_2 and that shares the similar taste with the user in s_2. Despite the fact that the black shirt and the skirt belong to the same category c_1, our model concludes that, in comparison to his preference for the brand b_2 and his similarity to the user in s_2, the user has little preference for the category c_1.

5 CONCLUSIONS

With the mind-blowing explainable SR model, our study not only makes accurate predictions but also well explains its reasoning process. Our framework provides two outstanding scientific merits: First, our model uses path instances to depict hidden interests and infer users' preferences, revealing new ways to generate explicit and sound explanations for SR without relying on long-term interactions or user information. Second, our model's strong performance and explainability show that meta-path guided instances can successfully expose users' intentions and that the reasoning process over these paths can accurately mimic users' decision-making process. For future work, we will incorporate item knowledge especially the meta-path guided method into existing models [4] [5] so that we can create better semantic interpretations of the latent space representations or clearer reasons for the models' decisions.

ACKNOWLEDGMENTS

This research is supported by Chinese Scientific and Technical Innovation Project 2030 (No. 2018AAA0102100), National Natural Science Foundation of China (No. 62077031). We thank the reviewers for their constructive comments.

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[2] Hongxu Chen, Yicong Li, Xiangguo Sun, Guandong Xu, and Hongzhi Yin. 2021. Temporal meta-path guided explainable recommendation. In *Proceedings of the 14th ACM international conference on web search and data mining*. 1056–1064.

[3] Jiayi Chen, Wen Wu, Wenxin Hu, Wei Zheng, and Liang He. 2021. SSR: Explainable Session-based Recommendation. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[4] Junsu Cho, SeongKu Kang, Dongmin Hyun, and Hwanjo Yu. 2021. Unsupervised Proxy Selection for Session-based Recommender Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 327–336.

[5] Zeyu Cui, Yinjiang Cai, Shu Wu, Xibo Ma, and Liang Wang. 2021. Motif-aware Sequential Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1738–1742.

[6] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1531–1540.

[7] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1419–1428.

[8] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*.

[9] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: short-term attention/memory priority model for session-based recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1831–1839.

[10] Wenjing Meng, Deqing Yang, and Yanghua Xiao. 2020. Incorporating user micro-behaviors and item knowledge into multi-task learning for session-based recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1091–1100.

[11] Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 188–197.

[12] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[13] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment* 4, 11 (2011), 992–1003.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[15] Ziyang Wang, Wei Wei, Gao Cong, Xiao-Li Li, Xian-Ling Mao, and Minghui Qiu. 2020. Global context enhanced graph neural networks for session-based recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 169–178.

[16] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 346–353.

[17] Yongfeng Zhang and Xu Chen. 2018. Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192* (2018).

[18] Huachi Zhou, Qiaoyu Tan, Xiao Huang, Kaixiong Zhou, and Xiaoling Wang. 2021. Temporal Augmented Graph Neural Networks for Session-Based Recommendations. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1798–1802.