HashtagMeta: Fake News Mitigation for Hashtag Recommendation

Jialiang SHI†, Takahiro KOMAMIZU†, and Ichiro IDE†

† Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464–8603, Japan
E-mail: †shij@cs.is.i.nagoya-u.ac.jp, ††taka-coma@acm.org, †††ide@i.nagoya-u.ac.jp

Abstract Hashtag recommendation systems facilitate user engagement in discussions on social media, but they also foster the propagation of fake news by inadvertently promoting misleading hashtags. To address this issue, we propose HashtagMeta, a novel Graph Neural Network (GNN) based framework designed to mitigate the spread of fake news via hashtags. HashtagMeta is based on the Heterogeneous Information Network (HeteIN) consisting of Tweets and hashtags, where the Tweet-hashtag relationship represents the containment of hashtags in each Tweet. To recommend relevant hashtags when a user attempts to post a Tweet, it estimates the relevant hashtags to a Tweet based on the connectivity in the network, which is realized via a GNN model. Concurrently, HashtagMeta dealing with mitigating fake news consists of two main steps: (1) Unsupervised step; Fake news Tweets are removed from the network, node similarities are calculated by PathSim which can model the semantic relationships in Tweets, and hashtag-hashtag edges based on the similarities are added to enhance the connectivity of the network to deal with the data sparsity issue, and (2) Supervised step; Heterogeneous Graph ATtention (HGAT) network is used for learning hashtag embeddings and minimize the propagation likelihood of fake news. Experimental evaluation demonstrates that the proposed effectively mitigates misleading hashtags while retaining relevance between hashtags and Tweets, providing a robust solution to misinformation and establishing a foundation for future graph-based research on fake news mitigation.

Key words Recommendation system, social media, fake news mitigation

1 Introduction

Hashtag recommendation systems enable users to quickly and conveniently join discussions while writing Tweets [1]. However, the growing influence of social media and the prominent role of hashtags have heightened the risk of fake news propagation, with these systems often contributing to the issue [2] [13]. Users frequently adopt trending hashtags to participate in discussions but may unknowingly encounter ones filled with misinformation (e.g., Tweets flagged as fake information by fact-checking Websites). For instance, as shown in Fig. 1, hashtags like `#wakeupamerica` and `#MAGA`, both tied to Donald Trump's election, reveal varying proportions of true and fake Tweets despite their similar themes. This problem is exacerbated when users rely on recommended hashtags without realizing that even similar hashtags can carry misleading contents. To create a safer online environment, developing fake news-aware hashtag recommendation models is essential to curb the spread of misinformation [4] [5].

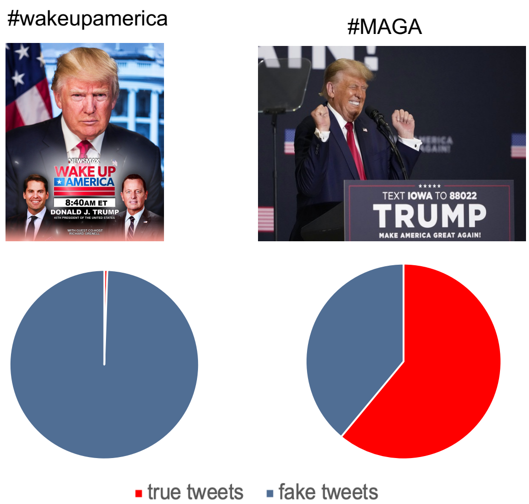


Figure 1: Results with similar topic hashtags for `#wakeupamerica` and `#MAGA`. The pie charts show the true/fake tweet ratios, where `#MAGA` has a higher ratio of true Tweets compared to `#wakeupamerica`. (Images from (注1) and (注2))

To address this challenge, we propose a Graph Neural Network (GNN)-based framework, HashtagMeta, designed to reduce the propagation of hashtags in fake news Tweets. The basic data structure for hashtag recommendation is a Heterogeneous Information Network (HeteIN) composed of two types of nodes (Tweet and hashtag), relationships among nodes, and Tweets nodes are associated with true/fake annotations; a Tweet is annotated with a fake la-

(注1)：https://www.instagram.com/newsmax/p/DAHDYO1sDCd/ (Accessed: February 12th, 2025)
(注2)：https://tucson.com/news/local/subscriber/donald-trump-arizona-election-2024-lawsuit-courts-challenger-14th-amendment (Accessed: February 12th, 2025)

bel if a fact-checking organization recognizes it as misinformation. HashtagMeta comprises two main steps: unsupervised step and supervised step. In the unsupervised step, it removes Tweet nodes associated with fake labels from the network. This process aims to focus more on the contextual semantics of true Tweets on the Tweet-hashtag network (purified network). Next, to deal with the sparsity issue of the original network that is not associated with relationships among hashtags, HashtagMeta adds hashtag-hashtag relationships based on the network-based node similarities. To realize this, HashtagMeta incorporates PathSim [18] to capture the semantic relationships among hashtags on the purified network. Finally, HashtagMeta reconstructs the network by adding hashtag-hashtag edges to the original network based on the k most similar hashtag nodes for each hashtag node, strengthening the semantic relationships within the Tweet-hashtag network. In the supervised step, a Heterogeneous Graph ATtention (HGAT) network [19] is applied to the reconstructed network to learn node embeddings and compute the hashtag recommendation loss. This enables the algorithm to learn which hashtags are more likely associated with fake news and prioritize reducing their propagation likelihood.

In summary, our contributions are as follows:

- **HashtagMeta — A novel graph-based approach to hashtag recommendation**: To the best of our knowledge, this is the first attempt to study the mitigation of spread of fake news in hashtag recommendation systems. We propose a GNN-based approach to effectively model the propagation characteristics of hashtags, addressing the challenge of mitigation of spread of misleading hashtags.
- **Mitigation of misleading hashtags while preserving semantic information**: Experimental evaluation reveals that HashtagMeta not only mitigates the spread of misleading hashtags but also preserves the core semantic information within the whole dataset. By leveraging the semantic relationships among hashtags, the proposed method enhances the robustness of hashtag recommendation.

2 Related Research

This work is closely related to the studies on fake information mitigation recommendation.

2.1 Recommendation for Fake Information Mitigation

Recommender systems help Tweets filter through large amounts of redundant information to retrieve desired contents [1]. However, in recent years, despite their ability to recommend accurate items, recommender systems have also been observed to suggest low-quality contents. On social media, where fake news is widespread, these systems may recommend popular but low-veracity items, which in turn can contribute to the spread of misinformation [14] [24]. This issue underscores the need for solutions within recommender systems to address the problem of misinformation on social media [4] [5].

To address this issue, researchers have proposed various strategies. Some methods [8] [12] aim to nudge Tweets toward higher-quality content without directly challenging their beliefs, thereby increasing engagement with trustworthy sources. A notable approach is Rec4Mit [21], a veracity-aware and event-driven recommendation model that delivers accurate news tailored to Tweets' preferences while mitigating fake news. Other studies focus on utilizing information from URLs to recommend fact-checked content to specific groups, such as professional fact-checkers [20]. Recently, Sallami et al. [16] have demonstrated that incorporating metrics to assess Tweets' trustworthiness can effectively reduce the spread of false information while also addressing the limitations of the traditional collaborative filtering paradigm, which tends to amplify the spread of fake news.

Despite these efforts, most existing approaches focus solely on whether misinformation itself is being recommended, overlooking its presence in other easily spreadable forms. For instance, misinformation often appears within hashtags [24], which are widely used on social media to discuss and propagate content. To address this gap, the proposed method adopts a novel perspective that considers not only the dissemination of veracity information but also the realistic scenario of misinformation embedded within hashtags. This approach broadens the scope of combating misinformation in recommender systems [3].

2.2 GNN-based Recommendation System

Recommender systems can be divided into three stages: shallow models, neural models, and GNN-based models. Early models used collaborative filtering to calculate interaction similarity, followed by collaborative filtering model-based methods like matrix factorization [10] and factorization machines [15], which approached recommendation as a representation learning task. However, these methods struggled with complex user behaviors and data input. Neural network-based models, such as neural collaborative filtering [7] and deep factorization machine [6], were proposed to improve prediction capacity.

The proposed method uses HGAT network [19] which is a kind of GNN to perform embedding learning for hashtags and Tweets.

3 HashtagMeta: Proposed Framework

We model hashtag recommendation as a link prediction task for a Heterogeneous Information Network (HeteIN) consisting of Tweets and hashtags. Link prediction is a task that estimates a missing link between nodes in a HeteIN, hashtag recommendation is modeled as predicting links between a Tweet and hashtags based on the present information in the HeteIN. A HeteIN for hashtag-related information is denoted as $G = (V, E, T, \delta)$, where T denotes a set of node types, $V = \bigcup_{t \in T} V$, such that $\bigcap_{t \in T} V_t = \emptyset$, represents a set of nodes where V_t corresponds with each node type $t \in T$, $E \subseteq V \times V$ is a set of relationships between nodes, and $\delta : V \to L$ is a node labeling function where L is the set of labels. In particular, here,

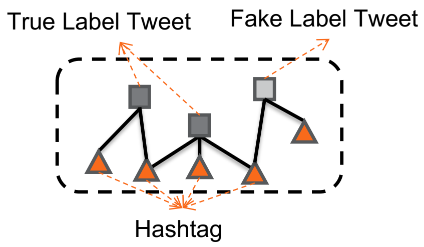


Figure 2: Visual example of a Heterogeneous Information Network (HeteIN) of hashtag-related information. It consists of two types of nodes, Hashtag (triangle) and Tweet (square), and relations of Hashtag-Tweet, Tweets are labeled either True (Black) or Fake (Gray).

$T = \{\mathcal{H}, \mathcal{T}\}$ ($\mathcal{H}$ and $\mathcal{T}$ represent hashtag and Tweet , respectively), $L = \{\text{True}, \text{Fake}\}$, δ is only applied to Tweets, and the relationships in G appear only between Tweet and hashtag, i.e., $E \subseteq V_{\mathcal{T}} \times V_{\mathcal{H}}$, that represents containment of a hashtag in a Tweet. Figure 2 illustrates an example of the HeteIN of hashtag data.

To address the recommendation task, we prepare HashtagMeta, a HeteIN-based link prediction framework with the feature combination of metapath-based similar hashtag nodes. As illustrated in Fig. 3, HashtagMeta consists of (1) Unsupervised step to calculate the PathSim under true hashtags and reconstruct a network, and (2) Supervised step that feeds the reconstructed network to GNN to compute the hashtag recommendation loss. Details of each step are explained below.

3.1 Unsupervised Step: Discover Semantics using Metapath

To better reflect the semantics under different nodes, we propose to calculate PathSim [18], which is a Metapath-based similarity measure to incorporate the semantics under Metapath specified to true/fake context. As shown in the unsupervised step in Fig. 3, firstly, we remove fake Tweets and their connected relations from the original network to understand the context of purified information. Secondly, we compute the top-k (e.g., $k = 5$) similar nodes according to PathSim. Specifically, a Metapath is a path defined in the graph of the network schema [18]. Given a schema network $G_S = (T, E_S)$ where $E_S \subseteq T \times T$, a Metapath P is defined as consecutive edges in G_S, which can be represented in a more interpretable way as: $t_0 \to t_1 \to \ldots \to t_m$, where $t_i \in T$ and m is the length of P. Given a symmetric Metapath P whose source and destination nodes are $t \in T$, the PathSim-based similarity $o(x, y)$ between two same type of instances x and y of t is as follows:

$$o(x, y) = \frac{2 \cdot \left|\{p_{x \sim y} : p_{x \sim y} \in P\}\right|}{\left|\{p_{x \sim x} : p_{x \sim x} \in P\}\right| + \left|\{p_{y \sim y} : p_{y \sim y} \in P\}\right|}, \tag{1}$$

where $p_{x \sim y}$ is a path instance of P in the data graph G of G_S between x and y. Similarly, $p_{x \sim x}$ (resp. $p_{y \sim y}$) is that between x (resp. y).

In particular, we select a Metapath $\mathcal{H} \to \mathcal{T} \to \mathcal{H}$ that represents the relationship between hashtags through Tweets, meaning that hashtags sharing common Tweets are related to each other. As shown in Fig. 4, we then compute $o(A, B)$. In this example, there is one path from node A to node A, two paths from node B to node B, and one path from node A to node B. Therefore, we calculate the similarity between nodes A and B as $o(A, B) = \frac{2 \cdot m''}{m''' + m'} = \frac{2 \cdot 1}{2+1} = 0.67$.

Based on the similarity scores, we reconstruct a HeteIN G' (called reconstructed graph) which is a super-graph of G. First, all the elements in G are copies to G', i.e., $G' = (V', E', T, \delta)$ such that $V' = V$ and $E' = E$. For each hashtag node $v \in V'_{\mathcal{H}} \subseteq V'$, top-$m$ similar nodes $W_{\mathcal{H}} \subseteq V'_{\mathcal{H}}$ are extracted. Then, edges between v and each node w in $W_{\mathcal{H}}$ are added to G' such that $E' \leftarrow E' \cup \{v\} \times V'_{\mathcal{H}}$.

3.2 Supervised Step

Based on the reconstructed graph G', we use a GNN as a framework to evaluate whether the hashtag can be linked to a Tweet. We use the Hierarchical Graph ATtention (HGAT) [22] network as a GNN, that aggregates messages in a hierarchical manner, and adopt to the link prediction task to compare the preference scores between hashtag-Tweet relation.

3.2.1 HGAT: Heterogeneous Graph ATtention Network

HGAT consists of three parts, as shown in Fig. 5: (1) Type-specific projection, (2) Adaptive node-level attention, and (3) Adaptive relation-level attention.

Firstly, the method projects different types of embeddings, hashtag and Tweet, into the same space as:

$$\mathbf{v}_i^{(0)} = \phi(v_i, t) \text{ s.t. } v_i \in V_t, \tag{2}$$

where $\phi : V \times T \to \mathbb{R}^d$ is a projection function to a corresponding node v of the given type $t \in T$ in the d_1-dimensional embedding space.

Secondly, the method integrates neighboring nodes of the same type by the adaptive node-level attention [19] in the d_2-dimensional space. We apply a shared weight matrix $\mathbf{W}_0 \in \mathbb{R}^{d_1 \times d_2}$ to transform the input features $\mathbf{v}_i$ to $\mathbf{z}_i$ as:

$$\mathbf{z}_i = \mathbf{W}_0 \cdot \mathbf{v}_i^{(0)}. \tag{3}$$

Then, we calculate the attention score α_{ij} with a shared weight matrix $\mathbf{W}_1 \in \mathbb{R}^{2d_2 \times d_3}$ between i and j, where j is the index of a neighbor node v_j of node v_i that connects directly to node $v_i \in V$ through a specific relation. We normalize this by the softmax function as:

$$\alpha_{ij}^m = \text{softmax}\left(\mathbf{W}_1 \cdot (\mathbf{z}_i \parallel \mathbf{z}_j)\right), \tag{4}$$

where $\parallel$ represents the concatenation operation.

We use a weight matrix $\mathbf{W}_3^m \in \mathbb{R}^{d_3 \times d_4}$ and α_{ij} as coefficients, to linearly combine neighboring nodes features with multi-head attention for node i in the relation r as:

$$\mathbf{v}_i^{(1)} = \Big\|_{m=1}^{M} \text{ReLU}\left(\sum_{j \in N(i)} \alpha_{ij}^m \mathbf{z}_j + \mathbf{W}_3^m (\mathbf{v}_i^{(0)} \odot \mathbf{v}_j^{(0)})\right), \tag{5}$$

where M is the number of multi-heads, $N(i)$ is the set of neighbor nodes of v_i, and $\odot$ represents the Hadamard product, ReLU is the

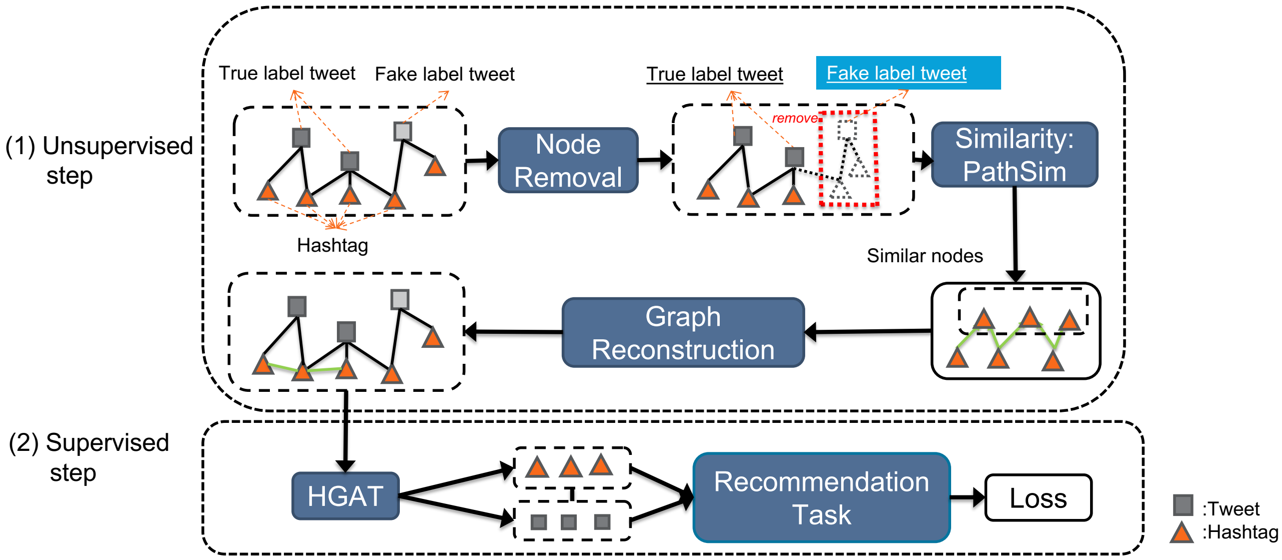


Figure 3: Overview of HashtagMeta composed of unsupervised and supervised steps. In the unsupervised step, firstly, fake Tweets are removed from the HeteIN to focus exclusively on the context of true Tweets. Secondly, PathSim similarities are calculated to capture and understand the underlying semantics within the true Tweets. Thirdly, the network is further enhanced by adding hashtag-hashtag edges (green). These edges are established based on the most similar (i.e., top-k) hashtags for each hashtag, strengthening semantic connections within the true Tweet network and leveraging the PathSim similarity. Finally, in the supervised step, the reconstructed graph data is fed into the Heterogeneous Graph ATtention (HGAT) network [19], to generate output node embeddings and compute the hashtag recommendation loss.

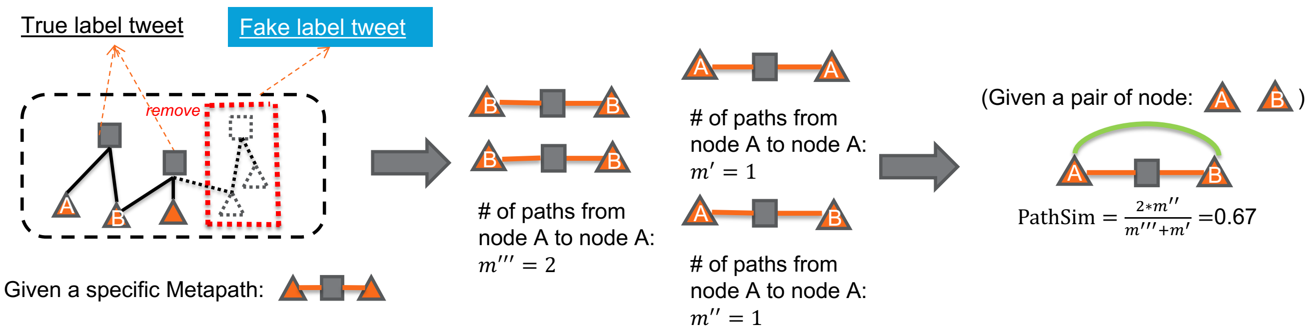


Figure 4: Example of calculating the PathSim score between nodes A and B under a specific metapath.

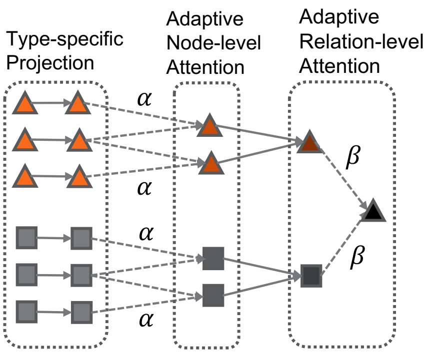


Figure 5: Overview of the Heterogeneous Graph ATtention(HGAT) [19] network.

activation function, Rectified Linear Unit.

Third, the method introduces the adaptive relation-level attention to learn the importance of each relation and fuse all relation-specific node embeddings. In particular, we apply a shared non-linear weight matrix $\mathbf{W}_r \in \mathbb{R}^{d_4 \times d_4}$ to transform the relation-specific node and use the trainable vector $\mathbf{q}$ to calculate the similarities and average it for all node embeddings of a specific relation to obtain the importance score $w_{i,r}$ for node i as:

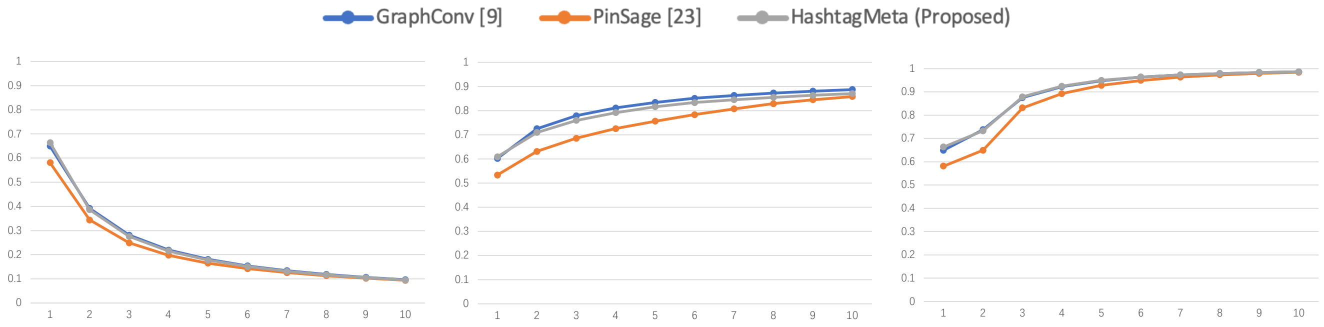
$$w_{i,r} = \frac{1}{|V_r|} \sum_{i \in V_r} \tanh\left(\mathbf{W}_r \cdot \mathbf{v}_i^{(l)} + \mathbf{b}\right) \mathbf{q}^{\top}, \tag{6}$$

where V_r denotes the set of nodes in a specific relation r and $\mathbf{b}$ the bias vector.

We then normalize $w_{i,r}$ to obtain the final relation-level attention weight $\beta_{i,r}$ and and fuse the relation-specific node embeddings $\mathbf{v}_i^{(l)}$ with it to obtain the final node embedding $\mathbf{v}_i^L$ as:

$$\beta_{i,r} = \frac{\exp(w_{i,r})}{\sum_{r \in R} \exp(w_{i,r})}, \tag{7}$$

$$\mathbf{v}_i^L = \sum_{r=1}^{R} \beta_{i,r} \mathbf{v}_i^{(l)}, \tag{8}$$



(a) Precision@k (b) Recall@k (c) NDCG@k

Figure 6: Comparison of GNN architectures in the HashtagMeta framework: HashtagMeta is based on HGAT [22], GraphConv [9] is based on GCN, and PinSage [23] is based on GAT.

where R represents the number of the different relations between hashtag and Tweet.

3.3 Objective Function for Recommendation

In this framework, the objective is to predict whether a connection exists between a hashtag and a Tweet. Therefore, we adopt the task of link prediction, which can compare the preference scores between nodes connected by a Tweet-hashtag relation in the original graph against the scores between a random pair of Tweet and hashtag nodes. For example, given a connected Tweet-hashtag pair, we expect the score between Tweet and hashtag would be higher than that between Tweet and a random hashtag node. Correspondingly, we use an inner product predictor $s(v_t, v_h) = \mathbf{v}_t^L \cdot \mathbf{v}_h^L$ to calculate the score between the embeddings of Tweet and hashtag and infer the cross-entropy recommendation loss L as:

$$L = \sum_{v_t \in V_t} \sum_{v_h \in N_h(v_t)} \max\left(0, 1 - s(v_t, v_h) + s(v_t, v_h')\right), \tag{9}$$

where $v_t \in V_t$ is a Tweet node, $v_h \in V_h$ is a hashtag node that is a neighbor of v_t in the Tweet-hashtag relation, and $v_h' \in V_h$ is a randomly selected hashtag node in G', respectively.

4 Recommendation Experiments

4.1 Settings

a) Dataset

In this evaluation, a Tweet-hashtag network constructed by FakeNewsNet [17], which is a comprehensive dataset with Tweets, naturally including the hashtag-Tweets relationship. It also includes true/fake tags on Tweets that are flagged by fact-checking Websites. In this experiment, for testing, we randomly split the Tweet-hashtag relationships in the Tweet-hashtag network into training, validation, and test sets with a ratio of 8:1:1. The nodes and edges in the validation and test sets are removed from the Tweet-hashtag network and the remaining graph is used as the training data.

b) Evaluation Metric

To ensure that our results are explainable in terms of both general metrics and hashtag veracity evaluation, we utilize Precision and Recall as general recommendation metrics and redefine the Normalized Discounted Cumulative Gain (NDCG) as:

$$\mathrm{NDCG} = \frac{1}{Q} \sum_{j=1}^{|Q|} Z \sum_{m=1}^{k} \frac{2^{R(j,m)\cdot\theta} - 1}{\log_2(1+m)}, \tag{10}$$

where Q is a set of test source nodes, Z is the normalization factor that is the inverse of the ideal DCG value for each test source node j, $R(j, m)$ is the relevance judgment, and θ represents the veracity ratio of true Tweets for a hashtag that a higher value indicates a stronger association with true Tweets.

As NDCG is chosen here, because it is designed to handle non-binary notions of relevance, making it easy to extend the concept of relevance to represent the ratio of true Tweets in this fake news mitigation recommendation. This approach was applied to evaluate hashtag veracity in top-k rankings, with k ranging from 1 to 10.

c) Implementation

For HashtagMeta, we set the learning rate to 0.005, batch size to 412, the training epochs to 50, the maximum number of similar nodes to be retrieved by PathSim [18] to 10. We randomly sample 100 negative hashtags for each Tweet in the test phase.

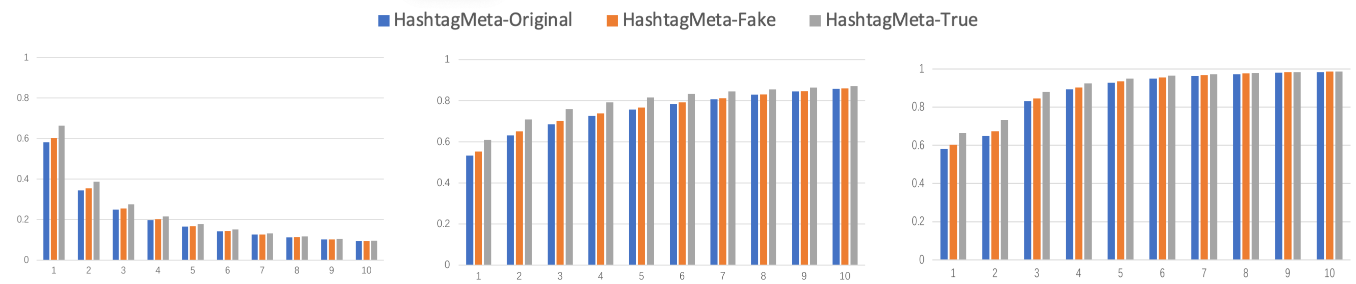
d) Baseline Methods

We compare the proposed HashtagMeta with two baselines, GraphConv [9] and PinSage [23], by replacing the HGAT [22] component of HashtagMeta with them:

- **GraphConv**: Graph Convolutional Network (GCN)-based approach that aggregates neighbor information equally for smoother node features.
- **PinSage**: Graph Attention Network (GAT)-based approach that combines random walks and graph convolutions to generate embeddings of nodes (i.e., items) that incorporate both graph structure as well as node feature information.

4.2 Performance Comparison: Choice of GNN Model

We present the performance of all models in Fig. 6. In terms of Precision (Fig. 6a) and Recall (Fig. 6b) that evaluate recommendation performance without considering fake Tweets, we found that HashtagMeta achieved the highest Precision, while GraphConv was the best and HashtagMeta was comparable to GraphConv for Re-



(a) Precision@k (b) Recall@k (c) NDCG@k

Figure 7: Comparison of Edge Addition: As a baseline, HashtagMeta-Original represents HashtagMeta without edge addition, HashtagMeta-Fake represents HashtagMeta that adds edges based on fake Tweets instead of true trees in HashtagMeta to observe the effect of adding edges based on fake tweet contexts.

call.

For NDCG (Fig. 6c), which takes the ratio of true Tweets into account (θ in Eq. 10), HashtagMeta outperformed all models. Therefore, HashtagMeta proved to be the most effective model overall, particularly excelling in NDCG, which highlights its ability to recommend a higher proportion of real information. Although the performance gap between HashtagMeta and GraphConv was small, it indicated that the proposed framework that learns semantics from the purified graph excluding fake Tweets enhanced the performance of recommending hashtags related to true Tweets.

4.3 Performance Comparison: Edge Addition

We evaluate the effect of addition of hashtag-hashtag relationships. To compare, the following models are constructed:

1. HashtagMeta-Original: This method skips the unsupervised step, and proceeds to the supervised step based on the original network.
2. HashtagMeta-Fake: This method adds hashtag-hashtag edges based on fake news Tweets. This is realized to swap the removal of Tweets in the unsupervised step of HashtagMeta into removal of true Tweets.
3. HashtagMeta-True: This method is exactly the proposed method introduced in Section 3.

As shown in Fig. 7, we found that HashtagMeta-True performed the best. HashtagMeta-Original performed worse than the other two methods, indicating that densifying the network had a positive effect. In comparison between HashtagMeta-True and -Fake methods, HashtagMeta-True performed better, indicating that the number of true Tweets was larger than that of fake Tweets, therefore, the added hashtag-hashtag relationships were more contextualized. Also, the fact that HashtagMeta-True performed better than HashtagMeta-Fake in NDCG (Fig. 7c) indicates that HashtagMeta-True could recommend hashtags avoiding recommending hashtags more associated with fake Tweets.

5 Conclusion

We proposed HashtagMeta, a framework that integrates an unsupervised node similarity calculation method with a supervised GNN approach to effectively learn neighbor relationship information from true Tweets for hashtag recommendation. This framework not only could recommend a suitable hashtag for a Tweet, but also could provide truthful hashtags. Specifically, the framework employed Metapaths in the unsupervised step to select specific true news Tweets, calculate node similarity using PathSim [18] and reconnect nodes. In the supervised step, it utilize HGAT [22] to predict Hashtag-Tweet relationships through link prediction.

While various other Metapath-based similarity computation methods exist [11], the emergence of novel data types on social media and the Web presents new opportunities to develop advanced fake news mitigation strategies and foster a healthier ecosystem [13]. Looking forward, we believe that effectively integrating multiple heterogeneous information sources will enable us to capture the diversity of data and recommend more satisfying and contextually relevant hashtags for Tweets.

Acknowledgment

This work was partly supported by JSPS KAKENHI JP21H0355.

References

[1] Areej Alsini, Du Q Huynh, and Amitava Datta. Hashtag recommendation methods for Twitter and Sina Weibo: a review. *Future Internet*, 13(5):129–148, May 2021.

[2] Canyu Chen, Haoran Wang, Matthew Shapiro, Yunyu Xiao, Fei Wang, and Kai Shu. Combating health misinformation in social media: Characterization, detection, intervention, and open issues. *Computing Research Repository, arXiv Preprints,* arXiv:1301.3781, January 2022.

[3] Mehdi Elahi, Dietmar Jannach, Lars Skjærven, Erik Knudsen, Helle Sjøvaag, Kristian Tolonen, Øyvind Holmstad, Igor Pipkin, Eivind Throndsen, Agnes Stenbom, et al. Towards responsible media recommendation. *AI and Ethics*, 2:103–114, May 2022.

[4] Miriam Fernandez and Alejandro Bellogín. Recommender systems and misinformation: The problem or the solution? In *Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems co-located with 14th ACM Conference on Recommender Systems*, page 40–50, March 2020.

[5] Miriam Fernandez, Alejandro Bellogín, and Iván Cantador. Analysing the effect of recommendation algorithms on the spread of misinformation. In *Proceedings of the 16th ACM Web Science Conference*, pages 159–169, December 2024.

[6] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xi-

uqiang He. DeepFM: A factorization-machine based neural network for CTR prediction. *Computing Research Repository, arXiv Preprints,* arXiv:1703.04247, November 2017.

[7] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pages 173–182, June 2017.

[8] Benjamin D Horne, Mauricio Gruppi, and Sibel Adali. Trustworthy misinformation mitigation with soft information nudging. In *Proceedings of the IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications*, pages 245–254, November 2019.

[9] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *Computing Research Repository, arXiv Preprints,* arXiv:1609.02907, January 2016.

[10] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, June 2009.

[11] Jiawei Liu, Chuan Shi, Cheng Yang, Zhiyuan Lu, and Philip S Yu. A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources. *AI Open*, 3:40–57, June 2022.

[12] Kuan-Chieh Lo, Shih-Chieh Dai, Aiping Xiong, Jing Jiang, and Lun-Wei Ku. All the wiser: Fake news intervention using user reading preferences. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 1069–1072, August 2021.

[13] Yvonne Oshevwe Okoro, Oluwatoyin Ayo-Farai, Chinedu Paschal Maduka, Chiamaka Chinaemelum Okongwu, and Olamide Tolulope Sodamade. A review of health misinformation on digital platforms: Challenges and countermeasures. *International Journal of Applied Research in Social Sciences*, 6(1):23–36, March 2024.

[14] Adri Priadana and Sylvert Prian Tahalea. Hashtag activism and message frames: social network analysis of Instagram during the COVID-19 pandemic outbreak in Indonesia. *Journal of Physics: Conference Series*, 1836(1):012031, 2021.

[15] Steffen Rendle. Factorization machines. In *Proceddings of the 2010 IEEE International Conference on Data Mining*, pages 995–1000, July 2010.

[16] Dorsaf Sallami, Rim Ben Salem, and Esma Aïmeur. Trust-based recommender system for fake news mitigation. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, pages 104–109, December 2023.

[17] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3):171–188, June 2020.

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

[19] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *Computing Research Repository, arXiv Preprints,* arXiv:1710.10903, 1050:4, February 2018.

[20] Nguyen Vo and Kyumin Lee. The rise of guardians: Fact-checking URL recommendation to combat fake news. In *Proceddings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 275–284, April 2018.

[21] Shoujin Wang, Xiaofei Xu, Xiuzhen Zhang, Yan Wang, and Wenzhuo Song. Veracity-aware and event-driven personalized news recommendation for fake news mitigation. In *Proceedings of the ACM Web Conference 2022*, pages 3673–3684, July 2022.

[22] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *Proceedings of the 19th International Conference on World Wide Web*, pages 2022–2032, August 2019.

[23] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for Web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 974–983, December 2018.

[24] Diana Zulli. Evaluating hashtag activism: Examining the theoretical challenges and opportunities of #blacklivesmatter. *Participations*, 17(1):197–215, January 2020.