UNIVERSITY OF CALIFORNIA

Los Angeles

Modeling and Discovering Authentic and Effective Influencers on Social Media via Graph Neural Network Learning

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Computer Science

by

Seungbae Kim

2021

ABSTRACT OF THE DISSERTATION

Modeling and Discovering Authentic and Effective Influencers on Social Media via Graph Neural Network Learning

by

Seungbae Kim
Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2021
Professor Wei Wang, Co-Chair
Professor George Varghese, Co-Chair

Influencer marketing, a word-of-mouth marketing strategy that leverages prominent individuals on social media, has been gaining great traction in recent years. As the number of influencers has explosively increased with the rapid growth of the influencer marketing industry, several issues in evaluating and discovering valuable influencers have been raised, including the influencer fraud problem. These issues arise from utilizing solely quantitative metrics such as the number of followers to identify influencers on social media. Although these quantitative indicators can measure the influence of social media users to some extent, it is challenging to evaluate the quality of influencers with such metrics.

In this dissertation, we propose several graph learning frameworks that incorporate various input sources from social media including text, images, and graphs, to evaluate influencers in the first place. To that end, we build and analyze social networks of influencers based on the posting, interacting, and advertising behaviors of influencers to find valuable insights by understanding the social relations of influencers. The frameworks presented in this dissertation take the influencer social networks with decent features from multi-modal

inputs and then assess qualities of influencers, including transparency, loyalty, authenticity, and efficiency of influencers. As a result, the proposed methodologies not only address pressing issues in the influencer marketing industry but also advance multi-modal graph learning-based applications.

The dissertation of Seungbae Kim is approved.

Kai-Wei Chang

Yizhou Sun

George Varghese, Committee Co-Chair

Wei Wang, Committee Co-Chair

University of California, Los Angeles

2021

To my wife,

for her endless support, care, and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

In memory of Prof. Mario Gerla, my former advisor who has always supported my research in a new field.

This work would not have been possible without the support and encouragement of my advisor, committee members, and family. First of all, I would like to express my sincere gratitude to my advisor, Prof. Wei Wang, who warmly welcomed me as a member of SCAI lab and gave me an invaluable opportunity to continue my research at UCLA. She has always fully supported the ideas and directions of my research and has given me generous assistance to help me grow as an independent and initiative researcher. I would also like to thank Prof. George Varghese, who is my teaching mentor. He has provided me with many opportunities to become a great teacher and has always assisted me to get through every difficult moment I had during my Ph.D. In addition, I would like to thank Prof. Yizhou Sun and Prof. Kai-Wei Chang for their insightful feedback on my research and kind answers to my questions.

I also want to express my infinite gratitude to my family. I would like to thank my parents, HyunJoun and HyeJeong Kim, for their endless love, faith, and support. I also want to thank my parents-in-law, MoonHyung Kim and JeongSook Choi, for their unconditional support and encouragement. My lovely children, Sooah and Noah, have continually provided the motivation to finish my degree. Most importantly, I would like to express my gratitude to my wife, Min Jung. She has been extremely supportive of me every step of the way, from the moment I prepared to start my Ph.D. to the moment I finished the degree. Without her tremendous understanding and encouragement, it would be impossible for me to complete my study.

VITA

2004 - 2009	B.S., Computer Science and Engineering, Chung-Ang University, Seoul, South Korea
2009 - 2011	M.S., Computer Science and Engineering, Seoul National University, Seoul, South Korea
2011 - 2014	Research Engineer, Institute for Information Technology Convergence, KAIST, Daejeon, South Korea
2014 - 2021	Graduate Student Researcher, Computer Science Department, UCLA

PUBLICATIONS

Seungbae Kim, Xiusi Chen, Jyun-Yu Jiang, Jinyoung Han, and Wei Wang. Evaluating audience loyalty and authenticity in influencer marketing via multi-task multi-relational learning. In *Proceedings of the 15th International AAAI Conference on Web and Social Media* (ICWSM 2021).

Seungbae Kim, Jyun-Yu Jiang, and Wei Wang. Discovering Undisclosed Paid Partnership on Social Media via Aspect-Attentive Sponsored Post Learning. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (WSDM 2021).

Seungbae Kim, Jyun-Yu Jiang, Jinyoung Han, and Wei Wang. Influencerrank: Discovering effective influencers via graph convolutional attentive recurrent neural networks. Under review.

Seungbae Kim, and Jinyoung Han. Detecting Engagement Bots on Social Influencer Marketing. In *Proceedings of the 12th International Conference on Social Informatics* (SocInfo 2020).

Seungbae Kim, Jyun-Yu Jiang, Masaki Nakada, Jinyoung Han, and Wei Wang. Multimodal Post Attentive Profiling for Influencer Marketing. In *Proceedings of the 31st World Wide Web Conference* (WWW 2020).

Xiao Yang, **Seungbae Kim**, and Yizhou Sun. How Do Influencers Mention Brands in Social Media? Sponsorship Prediction of Instagram Posts. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (ASONAM 2019).

Sara Melvin, Jonathan Lin, **Seungbae Kim**, and Mario Gerla. Hitchhiker: A Wireless Routing Protocol in a Delay Tolerant Network Using Density-Based Clustering. In *Proceedings of the IEEE 88th Vehicular Technology Conference* (VTC-Fall 2018).

Seunghyun Yoo, **Seungbae Kim**, Joshua Joy, and Mario Gerla. Promoting Cooperative Strategies on Proof-of-Work Blockchain. In *Proceedings of the International Joint Conference on Neural Networks* (IJCNN 2018).

Seungbae Kim, Jinyoung Han, Seunghyun Yoo, and Mario Gerla. How Are Social Influencers Connected in Instagram?. In *Proceedings of the 9th International Conference on Social Informatics* (SocInfo 2017).

Seungbae Kim, and Mario Gerla. Socio-Geo: Social Network Routing Protocol in Delay Tolerant Networks. In *Proceedings of the International Conference on Computing, Networking and Communications* (ICNC 2017).

CHAPTER 1

Introduction

Influencer marketing is a word-of-mouth marketing strategy that utilizes popular social media users to reach potential customers, promote brand awareness, and advertise products and services. Since companies and marketers can easily target niche markets through the social networks of influencers, the influencer marketing industry has been rapidly growing in recent years. To identify and discover influencers, companies and marketers heavily depend on utilizing quantitative indicators including the number of likes and follower counts. However, such metrics can only quantify the size of user influence on social media but disregard the quality of posts, social relationship with brands and audiences. This leads to pressing issues in influencer marketing such as influencer fraud problems and transparency in advertising posts.

In this dissertation, we aim to evaluate the quality of influencers by developing graph learning frameworks that incorporate multiple data sources from social media, thereby addressing the issues in influencer marketing. Specifically, we pursue a two-step approach: analyzing social networks of influencers and developing learning frameworks. In the first step, we construct multiple social networks of influencers, including the influencer network, the brand mentioning network, and the audience engagement network. By analyzing these social networks, we understand the distinct characteristics of influencers in different levels of qualities. Second, we propose graph learning frameworks that take the constructed networks to assess influencers from various perspectives, e.g., transparency, efficiency, and audience loyalty and authenticity.

1.1 Scope of the Research

- **Building large-scale datasets.** Although there have been a large number of research issues to be addressed in influencer marketing, many researchers have had to rely on small-scale data due to the absence of a large-scale influencer dataset. To foster other researchers to conduct in-depth studies in the influencer marketing domain, we construct the influencer datasets by collecting influencer profiles and their posts from Instagram which is the most popular social media for influencer marketing. We propose a multi-modal multi-task classification model that can automatically classify influencers and their posts into certain categories such as fashion, beauty, travel, and food, by learning unique features from images and texts and applying post-level attention.

- **Analysis on social networks of influencers.** We build various social networks of influencers to understand their behaviors from different perspectives. From the influencer network, we seek distinct social behaviors of influencers who have different interests. We also analyze the brand mentioning networks to understand how influencers mention brand names in their social media posts and different posting behaviors in sponsored posts from non-sponsored posts. Finally, from the audience engagement network, we aim to discover unique engagement patterns of social bots that generate fake likes and comments.

- **Multi-modal graph learning methods for influencer evaluation.** We propose several influencer evaluation metrics that can assess the quality of influencers. We then develop graph learning frameworks that exploit multiple input sources to discover high-quality influencers. By learning the social relationships of influencers and their behaviors, the proposed methods can address the issues in influencer marketing such as discovering hidden sponsorship in advertising posts, detecting engagement bots, and predicting engagement rates.

1.2 Contributions

- This dissertation is the first research to propose graph learning frameworks that incorporate multiple input sources including text and images, for evaluating influencers on social media. As the proposed frameworks utilize the data from any type of social media and excel other baseline methods, social media platforms and influencer marketing companies can adopt the frameworks to discover highly qualified influencers.

- We propose new influencer evaluation metrics including transparency, efficiency, and audience loyalty and authenticity. The transparency measures how transparent influencers are about their sponsorship disclosures, and the efficiency represents the expected return on investment, taking into account volatility over time. The loyalty is an indicator of how consistently the audience engages with influencers, and the authenticity is a value that shows the intimacy between influencers and bots.

- We foster interdisciplinary research between computer science and marketing by releasing two large-scale datasets that contain Instagram influencers and their millions of posts. We expect a variety of research activities in the field of marketing as the datasets contain influencer profiles, images, brand information, and metadata with detailed information about the posts.

1.3 Overview

The rest of the dissertation is organized as follows: Chapter 2 summarizes the relevant works for influencer marketing and graph neural networks. We introduce the three social networks of influencers and present analysis results in Chapters 3, 4, and 5. Chapter 6 describes a multi-modal influencer classifier to construct the large-scale datasets. Chapters 7, 8, and 9 present the graph learning frameworks and their experimental results for influencer evaluation. Chapter 10 concludes this dissertation with a summary of our work.

CHAPTER 2

Related Work

2.1 Influencers on Social Media

The explosive growth of the influencer marketing industry has led researchers to study influencers in various fields such as in marketing [44, 50, 105, 26], health [37], and travel [124]. De *et al.* [44] find that the number of followers is positively correlated to the popularity of influencers whereas the popularity correlates with the opinion leadership only in the limited cases. Casalo *et al.* [26] also reveal that the originality and uniqueness of the posts are the key factors to be an influencer. It has been also studied that influencers' attractiveness and authenticity have a positive impact on the perception of audiences which consequently makes them be more likable users in the social network [105]. These results suggest that understanding influencers' posting behavior and characteristics of their posts help find influencers. In addition, Evans *et al.* [50] study the effects of disclosing sponsorship on the advertising social media posts, and find that sponsorship disclosures increase advertising recognition and negatively impact the attitudes and behavioral intention. Furthermore, some works analyze influencers' brand mentioning behaviors [176] and emoji usage [58], and the results demonstrate that each influencer has unique properties.

2.2 Influence Prediction in Social Networks

To find influencers in social networks, most existing studies rely on social media features to measure the influence. For example, the number of followers, posts, reposts, and mentions are well-known metrics to measure the influence of a user [10, 156]. Based on the measures, the regression tree model has used [10] and rank results have aggregated to rank influencers [156], respectively. Some works use the information from the network structure with the PageRank algorithm [126]. Silva *et al.* [152] build the user-content graph that shows the content creation and propagation relationships between users and contents. Then they use random walks over the content diffusion network. Romero *et al.* [141] propose the passivity of nodes to measure how likely the information is propagated in the social networks, and then apply the PageRank to rank the users. Liu *et al.* [103] consider the time domain over the user trust network in the proposed framework to classify influencers into one of three categories, emerging influencers, holding influencers, and vanishing influencers. In addition to social network features, some studies propose to use machine learning with statistical features. Li *et al.* [101] extract network-based, content-based, and user activeness-based statistical features, e.g., the number of followers, and length of posts, to predict the influence of users. Segev *et al.* [148] use simple statistics of posts and users, e.g., the number of likes, comments, followers, and posts, to measure the user influence using a regression model. Some previous works, on the other hand, exploit graphical information. Zhang *et al.* [181] exploit the social influence locality to predict retweet behaviors. Qiu *et al.* [134] utilize mini-batches of sub-graphs and apply the attention mechanism to predict the influence of users on social networks. Chen *et al.* [32] propose recurrent convolutional networks to consider temporal effect on information cascade prediction. However, most previous works fail to consider temporal dynamics in the social relationships and characteristics of users.

2.3 Brand Loyalty in Social Media Marketing

Audience loyalty is an indicator that shows how consistently a social media user makes engagements to a particular brand or influential celebrity (i.e., influencer) to express their interests or make interactions [48]. Marketers seek loyal audiences for their marketing campaigns since loyal audiences have more trust, positive engagement, and repurchase than other users [16]. Audience loyalty can be established, enhanced, and maintained by having persistent interactions and offering enjoyable social media contents [83, 123]. In social media, audience loyalty is often measured based on the engagements suggested from many previous studies which showed a positive relationship between audience loyalty and engagements [70, 162, 47, 19]. The loyalty of each audience can be measured based on their engagements, hence the audience loyalty of a brand or influencer can be expressed as a retention rate, which indicates how many audiences make returning engagement over time [135, 3].

2.4 Fake Followers and Fake Engagements

The authenticity of social media users is a measure of whether the social media account is real or fake. Since the numbers of followers or engagements are often considered as popularity [44], social media users can manipulate their popularity by purchasing fake followers or fake engagements (a.k.a. link farming) [41, 149, 15]. To understand and address this problem, inauthentic user (i.e., bot) detection has been broadly studied [54, 125]. Many studies found distinct engaging or following behaviors of bots that are different from authentic users. Sen *et al.* [149] exploit engaging frequency and topical information, Kudugunta *et al.* [95] suggest encoding contextual information from user profiles with RNN, and Chavoshi *et al.* [28] focus on synchronized behavior of inauthentic accounts to detect bots on social media. Recently, Yang *et al.* [175] propose a generalized and scalable bot detection framework optimized with various validation sets.

2.5 Sponsorship Disclosure in Influencer Marketing

As influencer marketing has become a popular advertising method in recent years [105, 92, 10], several previous studies show the effect of disclosing sponsorship. Evans *et al.* [50] find that sponsorship disclosure helps audiences recognize paid partnerships but lowers purchase intention. Stubb *et al.* [155] find that impartiality disclosure, e.g., adding “This is not sponsored post”, helps generate high influencer credibility. Moreover, Evans *et al.* [49] investigate the effects of sponsorship text disclosure and sponsor pre-roll video advertising on YouTube. They find that the sponsor pre-roll advertising help audiences to understand sponsorship transparency. Yang *et al.* [176] reveal that distinct characteristics of sponsored posts, e.g., less number of usertags, longer caption than non-sponsored posts, that help exclusive promotion in advertising posts. Wojdynski *et al.* [171] present a metric to measure sponsorship transparency based on consumers' perceptions. However, this study only uses a very small number of sample posts and automated method has not been proposed.

2.6 Social Networks of Influencers

As influencer marketing receives a tremendous attention from marketers due to its effectiveness, many researchers have studied various aspects of influencer social networks. Rios *et al.* [138] presented a method to identify influencers in the social network that is built on user activities. They used semantic analysis to filter out futile links to find influencers. Arenas *et al.* [6] also proposed to use social networks of consumer reviews to find influencers. They found that identified influencers have a larger scope on various categories with high centrality than normal users. Kim *et al.* [90] built a social network of influencers based on their followers and followees to understand social relationships among influencers. They found that influencers with similar topics make clusters in the network and have more common followers than other influencers with no common interests. Yang *et al.* [176] presented the brand mentioning network that connects influencers and brands based on their brand mentioning tags. They proposed a neural network model that can predict sponsorship of social media post published by influencers by learning graphical features from the brand mentioning network.

2.7 Social Bot Detection

Bot detection in social media have been extensively investigated since bots can adversely affect authentic users and companies [54]. Some previous studies focused on various sets of features to represent bots on social media. Schuchard *et al.* [147] analyzed the properties of user conversation networks such as centrality, node degrees, and communities to detect bots. Varol *et al.* [163] presented a large number of high-level user representations including friendship, network, sentiment, temporal, and language features, and evaluated the proposed features using well-known learning algorithms. Wang *et al.* [167] presented a clustering model to group users with similar network structures. They used a sequence of events to build a user network. Kudugunta *et al.* [95] proposed a neural network based bot detection model that learns sequential embeddings of contextualized tweets by using a long short-term memory (LSTM).

2.8 User Profiling on Social Media

Many scholars have studied the user classification (or user profiling) on online social networks (OSNs) as social media has been grwoing explosively. Some researchers proposed a machine learning-based user classifier on Twitter for three tasks, detecting users' political affiliation, ethnicity, and business affiliation [129, 130]. They employed the Gradient Boosted Decision Trees framework [57] as a classification model and learned four feature sets which are profile features, tweeting behavior features, linguistic content features, and social network features. The authors showed that the linguistic features which encapsulated users' lexical usage information show a robust performance. Hung *et al.* [75] proposed a tag-based user profiling method for social media recommendation. The method finds a set of tags from users' profiles and posts and calculates weights for the tags to build the user profile. You *et al.* [178] focused on visual content to classify users' interests and proposed a classifier that learns from 748 Pinterest users' photo albums. To profile a user, the authors analyzed the users' individual images and then aggregated the image analysis results to obtain the user's interest distribution. These studies used various information for classifying users in OSNs, but most of the work paid little attention to use both textual and visual features in classifying special individuals, influencers, on social media. Farnadi *et al.* [51] presented a deep neural network that takes multi-modal features that represent user characteristics. They extracted high-level textual information using LIWC [131], facial information from profile images using Oxford Face API [24], and a set of neighbors from a page like relationship network using Node2Vec [65]. They introduced the power-set combination approach to aggregate multi-modal information. However, the proposed features may not be applicable for identifying the main interest of a given social media user because most features represent high-level user characteristics.

2.9 Graph Neural Networks (GNN)

In recent years, Graph Neural Networks (GNN) [144] have been studied to apply the concept of neural networks to the data with underlying graph structures. Among the various versions of GNNs, Graph Convolutional Networks (GCN) [94] have gained massive attention due to its effective convolutional filters that are able to capture both graph structures and neighboring node features. However, it omitted the edge type and node type contained in typical heterogeneous graphs which could be also very informative for producing the node/edge representations. To incorporate multiple relational information on top of GCNs, Wang *et al.* [168] propose the signed heterogeneous information network embedding (SHINE) that separates networks by link types to generate embeddings for each relation and then combines all embeddings at the end. Since SHINE treats multi-relational information as multiple homogeneous networks, it fails to learn interactions between different types of relations. Graph Transformer Networks (GTNs) [179] introduces meta-path to learn interactions among multiple relations but the performance of the framework might be depending on the quality of the generated meta-paths, which means bad meta-paths can easily propagate errors, harnessing the overall performance. Schlichtkrull *et al.* [145], on the other hand, propose Relational GCNs (R-GCNs) which utilizes weight parameter sharing between different relation types instead of considering multiple homogeneous networks. They apply weight matrix decomposition to optimize a large number of parameters. In this way, they force interactions between different relations, tackling the drawbacks of SHINE to some extent. Wang *et al.* [169] propose Heterogeneous graph Attention Network (HAN) which applies the attention mechanism at the node-level and semantic-level to learn the importance of nodes and meta-paths, respectively. While HAN is able to capture the knowledge from heterogeneous information networks, it can also be dependent on the meta-paths, which suffer from the same shortcoming of GTNs.

2.10 Graph Convolutional Recurrent Networks

Graph convolutional networks (GCNs) [94] are the neural network architecture for graph-structured data. GCNs deploy spectral convolutional structures with localized first-order approximations so that the knowledge of both node features and graph structures can be leveraged. The robustness and the effectiveness of GCNs have already brought on successes in many fields, such as recommender systems [177], computer vision [128], and popularity prediction [165]. However, while real-world data that can be modeled as graphs dynamically changes over time, temporal information cannot be easily captured from GCNs. To learn the temporal dynamics of structural graphs, previous studies suggest to combine GCN and recurrent neural networks (RNNs). Seo *et al.* [150] propose the models that (i) stack up graphs to make RNN inputs and (ii) consider convolutions in RNNs, which can learn a sequence of structural information. They find that each model outperforms the other models depending on applications such as video prediction and natural language modeling. Pareja *et al.* [127] propose another approach to capture graph dynamics. Instead of using a sequence of graph embedding as inputs of RNN, they first use RNN to acquire the knowledge of network parameter dynamics. This approach can benefit in a case where a node dynamically appears and disappears.

2.11 Multi-Task Learning

Multi-Task Learning (MTL) enables us to model multiple related tasks by sharing representations [143]. By jointly learning multiple related tasks, the knowledge acquired from one task can be applied to other tasks hence improving the performance of all tasks. Moreover, MTL also helps generalize the model by leveraging information from related tasks as an inductive bias [25]. Thanks to such advantages, MTL has been used in many kinds of applications of machine learning such as computer vision [182], natural language processing [38], speech recognition [91], and user profiling [92]. Our proposed framework has multiple tasks that jointly learn representations from the multi-relational GCNs.

2.12 Multimodal Deep Learning

Multimodal deep learning [121] is the method that can effectively generate the joint representations of different modalities by leveraging multiple sources of inputs. Since different modalities usually carry different information, combining various types of information can improve performance in learning tasks. The effectiveness of multimodal learning has led other researchers to exploit multi modalities, such as speaker identification [137], social media user profiling [92], visual question answering [106], sentiment prediction [30], depression detection [151], and speech recognition [74]. Despite the excellent performance of multimodal learning and the popularity of influencer marketing, no previous studies have yet adopted multimodal learning to evaluate influencers on social media.

CHAPTER 3

Analysis on Influencer Relationship in Social Networks

3.1 Background

Influencer marketing [61] that utilizes popular and influential users in online social media, dubbed as 'social influencers', recently has gained a great attention as a new marketing strategy [11, 68, 96]. Brands such as NIKE or Starbucks seek to advertise their products to potential customers through such social influencers who may be known as experts and thus can influence their followers [62, 66]. Social influencers are often regarded as special individuals who can create valuable content and/or have high reputations in specific fields [27]. A recent McKinsey article reported that social influencer marketing plays a great role in attracting consumers to buy products [20]. Also, it has been reported that brands spent 121 billion U.S. dollars on influencer marketing in 2015 [153].

This in turn has led many researchers and firms to study various aspects of social influencer marketing. Bhatt *et al.* found that a future purchase can be affected by the adoption of the product by his/her friends, and this influence remains mostly local to first-adopters and their immediate friends [17]. Leskovec *et al.* showed that purchasing from recommendation follows a power-law distribution, and most person-to-person recommendations do not spread beyond the initial purchase of a product [99]. Goodman *et al.* proposed a social media valuation algorithm that evaluates the defined index values for bloggers to determine whether the bloggers are influencers or not [64]. However, little effort has yet been paid to how social influencers interact each other. Understanding how social influencers manage their social relationships with other influencers and how they share common potential customers can provide a valuable insight for brands who would like to effectively hire a set of influencers and maximize the advertising effect.

This paper presents the first attempt to analyze social relationships and interactions among influencers in Instagram. By conducting a measurement study on Instagram, we collected and analyzed 218 social influencers, who are followed by 8.9 M users, and registered in *Popular Pays*, a popular influencer marketing platform that connects brands and influencers [67].

Using the collected dataset, we analyze (i) how social influencers establish relationships with others, in comparison to general users in Instagram, and (ii) how social influencers interact one another. We find that social influencers tend to (i) have a large number of followers who are mostly their fans and potential customers of brands, (ii) follow other influencers and make reciprocal social relationships with them, and (iii) share common followers with other influencers. By exploring social relationships and interactions among social influencers, we reveal that influencers who have higher node degrees and more bidirectional edges tend to have more common followers with other influencers. We also show that influencers with similar interests or same occupations tend to follow each other, have more interactions, and have more common followers.

3.2 Dataset

We collected the information of 218 social influencers, who had participated in marketing campaigns in the *Popular Pays* Instagram page. *Popular Pays* is a popular influencer marketing company that connects brands and influencers in Instagram. *Popular Pays* selectively posts advertising content created by their registered influencers. By crawling the *Popular Pays* page, we could obtain their 218 registered social influencers.

To examine the relationships among social influencers in *Popular Pays*, we first fetched the lists of followers and followees of the influencers. We collected 8.9 M followers and 167 K followees of the 218 influencers. We then downloaded the html files of Instagram pages of all the collected followers and followees to retrieve their profiles, e.g., their numbers of followers, followees, or posts. To analyze interaction among social influencers, we also collected all the comments on content posted by the influencers. We obtained 13 M comments from the 325 K posts in the influencers' Instagram pages. We also fetched the user profile pages of the influencers to classify them into four major occupations: (i) photographers, (ii) bloggers, (iii) designers, and (iv) others.

For the purpose of comparison, we randomly collected 948 Instagram users who have less than 10 K followers. These 'general' users have 378 K followers, 780 K followees, and 63 K comments in their 158 K posted content.

3.3 Influencers vs. General Users

We first investigate how social influencers maintain social relationships with other users via 'following'. Table 3.1 compares the influencers and the general users. Overall, as shown in Table 3.1, the social influencers show significantly different characteristics in terms of their connectivities or activities, in comparison with the general users. For example, influencers tend to create more number of content, and they are likely to be followed by more number of users than general Instagram users; influencers have 40 K followers and post 1,490 content while general users have 400 followers and 117 postings on average. While influencers are not likely to follow others, general users tend to have a large number of followees since they often use Instagram to see images of celebrities or popular brands.

We also examine how different influencers share common followers and followees. Two influencers, who share many common followers or followees, may have a strong tie in a network. As shown in Table 3.1, the influencers have 1.1 M common followers, which accounts for 16.55% of total number of unique followers, while only 1.81% of followers are common among general users. This implies that followers of influencers tend to also follow other influencers as well. We also find that Influencers have more common followees than general users; 20.16% of total followees are common across the influencers. Note that general users tend to have a relatively high common followees because they often follow specific famous

Table 3.1: A comparison between the influencers and general users.

	Influencers	***General***
Number of users	218	948
Avg. followers	40,322	400
Avg. followees	765	823
Avg. posts	1,490	117
Total followers	8,790,208	378,069
Total followees	166,837	780,435
Common followers	1,104,999 (16.55%)	6,689 (1.81%)
Common followees	21,890 (20.16%)	55,084 (9.63%)
Distinct followers	6,676,252	369,255
Distinct followees	108,580	572,137
Reciprocal users	63,687	135,189
Avg. followers of followers	1,040	1,704
Avg. followers of followees	82,735	107,474
Avg. followers of reciprocal users	17,739	1,756
Avg. # comments with parasocial users	0.19	0.01
Avg. # comments with reciprocal users	7.60	0.23

celebrities or global brands who may have more than 1 M followers.

We next focus on how influencers have reciprocal relationships with others. People tend to have reciprocal relationships when they know each other or share similar interests. Thus, reciprocal users may have closer relationships and more interactions than parasocial users. The 218 influencers have reciprocal relationships with 64 K users, which indicates that 38.17% of influencers' followees are reciprocal users. However, only 0.72% of followers are also followed by the influencers. Considering the fact that 99% of followers of influencers are parasocial, most users who follow influencers can be regarded as their fans. Interestingly, general users tend to be more reciprocal with their followers than their followees, which

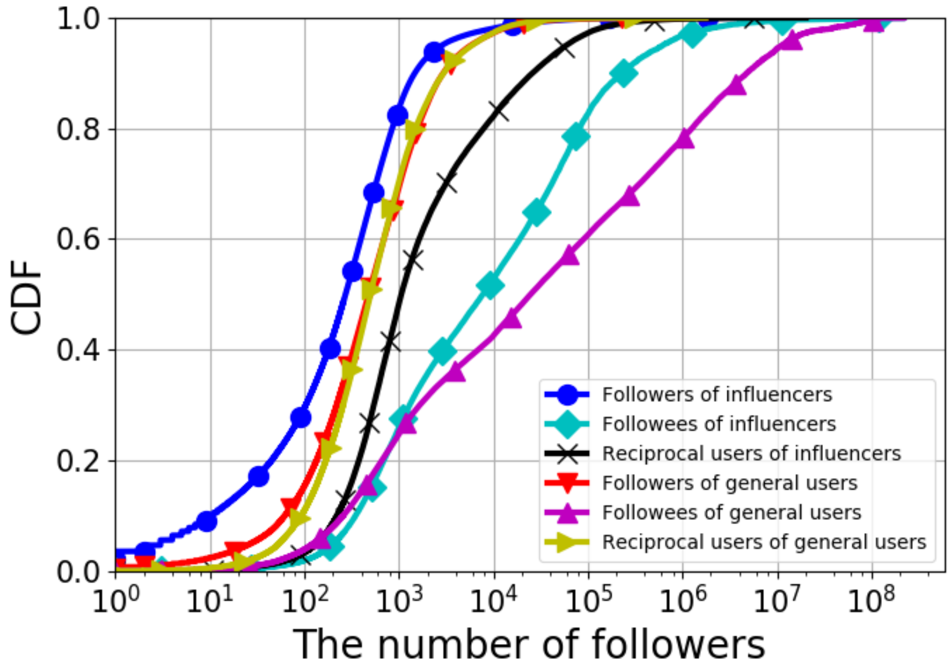


Figure 3.1: Number of followers of influencers' or general users' followers, followees, and reciprocal (or bi-directional) users.

implies that general users are likely to have reciprocal relationships with their friends and families, whereas general users may not be followed by their followees as they are often celebrities or popular brands.

We next analyze how influencers frequently interact with other users by investigating the number of comments on content posted by the influencers. As shown in Table 3.1, influencers tend to have more interaction with their followers and followees than general users. We also find that influencers have more interactions with reciprocal friends, who tend to have a large number of followers, than parasocial users.

To investigate how followers, followees, or reciprocal (or bi-directional) friends of influencers have many numbers of followers, we plot the distributions of numbers of followers of influencers' followers, followees, and reciprocal users, respectively, in Figure 3.1. We also plot the numbers of followers of general users' followers, followees, and reciprocal users, respectively, in Figure 3.1. We find that followers of both influencers and general users have

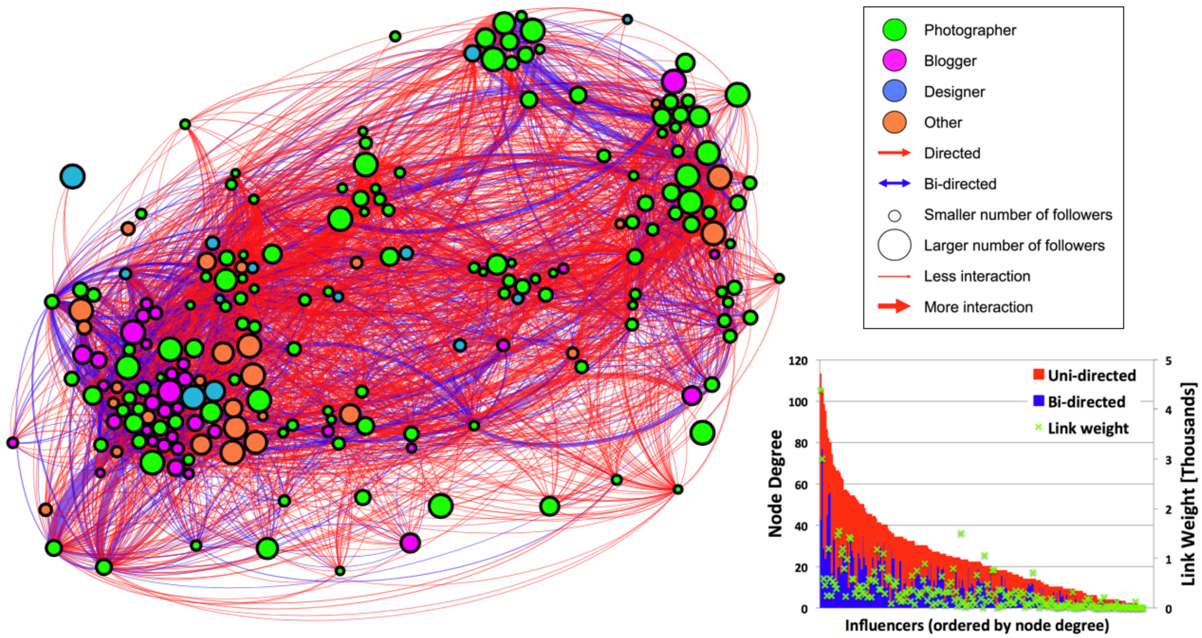


Figure 3.2: An illustration of the influencer network. Influencers who participate in the marketing campaigns form a mesh-style social network. Among 218 influencers, 212 of them are connected among each other.

much smaller numbers of followers than followees of influencers and general users. This result confirms that followers are mostly fans of influencers, and both influencers and general users tend to follow famous people. Interestingly, the numbers of followers of reciprocal users for influencers and general users are substantially different. The distributions for general users' followers and reciprocal users are almost identical, whereas influencers' reciprocal users have a large number of followers. This reveals that influencers tend to have mutual relationships with other influencers.

3.4 Connectivities among Social Influencers

In this section, we provide an in-depth analysis on how influencers are connected to each other. It has been reported that new links are attached to nodes in proportion to the its popularity [13], and users tend to connect to other users who are close in a network [115].

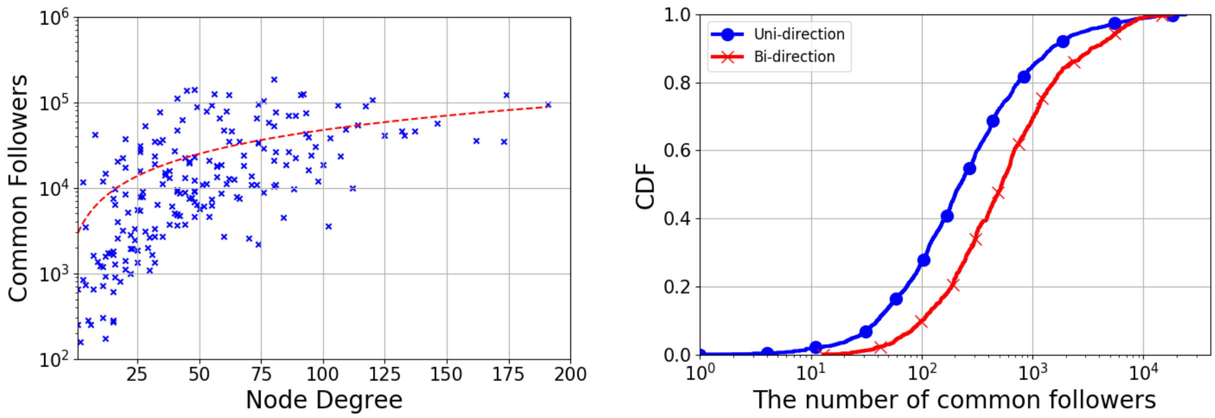
More specifically, we seek to answer the following question – How do two connected influencers share common followers? Answering such a question can provide important insight into homophily or synergy of connected influencers from a social influencer marketing perspective.

To investigate how social influencers are connected one another, we first define the notion of the *influencer network* where a node indicates an influencer and a directed edge represents 'following'. A weight on an edge from A to B represents the number of comments A writes on B's content. If A and B follows each other, the edge between A and B is bidirectional, and the weight on the bidirectional edge is the sum of the number of comments left in other's content.

Figure 3.2 illustrates the influencer network. Different node colors represent the influencer's occupations (e.g., photographer, designer, etc.). A larger node indicates the more number of followers the influencer has. The red edge between two influencer represents that one influencer follows the other influencer while they have a blue edge if they follow each other. The width of edges indicates the level of interaction, i.e., the number of comments. As shown in Figure 3.2, influencers are highly connected to each other. When we look at the distributions of node degrees and link weights in Figure 3.2, we find that some influencers have social relationships with more than 100 influencers, and nodes with higher degree tend to have larger link weight values. On average, an influencer follows 25 influencers and has written 322 comments.

To investigate how an influencer with different connectivities in the network share followers with other influencers, we plot the number of common followers shared by an influencer (with different degree) with others in Figure 3.3(a). Each blue marker in Figure 3.3(a) represents the number of total common followers that an influencer has. As shown in Figure 3.3(a), if influencers have more connections to the other influencers, they tend to have more common followers.

We next examine how two influencers with parasocial (uni-directional) or reciprocal (bi-directional) relationships share common followers in Figure 3.3(b). As shown in Fig-



(a) Node degree vs. his/her # of common followers with others

(b) Edge direction vs. # of common followers on the edge

Figure 3.3: Influencers' connectivities are associated with their common followers with other influencers.

ure 3.3(b), two influencers who reciprocally follow each other have more common followers than those with parasocial relationships. This may be because the bi-directional relationship between two influencers may lead to expose postings to people who follow one influencer, which results in following the other influencer as well.

We finally investigate whether influencers with same occupations share more common followers than the ones with different occupations in Table 3.2. As shown in Table 3.2, among 218 influencers, 139 of them, which accounts for 64%, are photographers. The remaining influencers are composed of 37 bloggers, 13 designers, and 29 other occupations such as actors or sports stars. Except the photographers, influencers with same occupations have more common followers than influencers with different occupations. Bloggers have 2,477 common followers on average between two bloggers while they have 1,126 common followers with photographers. Also, 72% of total edges among bloggers are bidirectional ones. This implies that influencers who have similar interests or same occupations tend to share common followers, who may have similar interests. Note that nodes with same colors (i.e., same occupations) tend to locate closely in the given network as shown in Figure 3.2. Photographers show somewhat disparate patterns compared to others since their interests or topics

Table 3.2: Influencers who have same occupation tend to share many common followers.

		Edges To			
Edges From	Values	**Photographer(P)**	**Blogger(B)**	**Designer(D)**	**Other(O)**
P	#Edges(Bi%) Avg. weight #Common	2,838 (42.6%) 12.9 863.6	348 (64.9%) 12.9 1,217.1	193 (38.9%) 7.5 479.2	394 (40.6%) 14.7 1,113.7
B	#Edges(Bi%) Avg. weight #Common	446 (50.7%) 11.1 1,125.5	161 (72.0%) 18.7 2,476.6	47 (23.4%) 3.7 678.4	90 (33.3%) 4.5 435.1
D	#Edges(Bi%) Avg. weight #Common	134 (56.0%) 12.5 643.7	16 (68.8%) 19.9 2,036.6	8 (50.0%) 30.0 6,842.6	23 (43.5%) 17.2 1,339.1
O	#Edges(Bi%) Avg. weight #Common	363 (44.1%) 14.0 1,147.7	60 (50.0%) 8.3 582.3	20 (50.0%) 7.8 1,248.5	76 (47.4%) 13.5 1,437.4

could be diverse. For example, fashion photographers might have closer relationships with designers than food photographers.

In summary, influencers with similar interests or same occupations tend to follow and interact each other, and hence those influencers with same occupations are likely to have strong ties and form communities in the influencer network. The mutual ties between two influencers eventually lead to a greater number of common followers, which in turn results in their co-evolution in social influencer marketing.

CHAPTER 4

Analysis on Brand Mentioning Behaviors of Influencers in Social Networks

4.1 Background

Nowadays, social media has become an essential marketing channel. Companies have started using social media to advertise products, increase brand awareness, and have better communication with consumers. However, unlike traditional media, social media allows all users to generate their own contents. Moreover, since it is well known that people trust friends' recommendation more than brands' advertisement [33], influencers can have more impact on the audience for marketing than companies. As a consequence, companies utilize brand mentioning by influencers to increase brand awareness and purchase intention of potential customers [102]. Therefore, it is important for companies to understand the brand mentioning behaviors of influencers thereby leveraging the advantages of social media as a marketing channel.

Influencer marketing is one of the social media marketing strategies that utilize the brand mentioning effect [44]. Since influencers are users with a notable number of followers [90], companies can effectively advertise their products to the influencers' followers if influencers recommend the products with brand mentions. There are two types of brand mentions, the sponsored brand mentioning (a.k.a., paid media) and the non-sponsored brand mentioning (a.k.a., earned media) [154]. If influencers get paid from companies by mentioning the brands then it is the sponsored mentioning. On the other hand, the non-sponsored brand mentioning is what influencers mention brands' names without having any sponsorship.

Many researchers have studied on various social media marketing strategies including influencer marketing or brand mentioning. Stephen et al. [154] examined the effectiveness of earned media and found that social earned media greatly affects sales compared to the traditional earned media. It has also been studied that popular influencers are more likable [44] and they are strongly tied and have common followers [90]. However, no study yet has focused on the brand mentioning behavior of influencers and compared characteristics of the sponsored brand mentioning and the non-sponsored brand mentioning.

In this work, we study brand mentioning practice on Instagram which is known as the most popular social media platform for influencer marketing [119]. We collect brand mentioning posts from Instagram and label them as either the sponsored or the non-sponsored by examining the sponsorship indicator within the corresponding posts. We then build the ***Brand Mentioning Network*** which is an information network of influencers and brands who are connected by brand mentioning posts created by the influencers. By analyzing the brand mentioning network, we reveal influencers' brand mentioning behavior and characteristics of the sponsored and the non-sponsored posts, so that we provide valuable insights for influencers and marketers. Furthermore, we propose a neural network based model that can classify the sponsorship of social media posts. Our model learns network structures of the brand mentioning network as well as other node properties including the number of followers and the number of hashtags. We find that the network embedding plays an important role to detect sponsored posts. The experimental results show that Our model outperforms baseline methods with 80% accuracy of classifying sponsorship of given posts.

4.2 Dataset

We collected influencers and brands data from Instagram. We first collected brand sponsored posts, where influencers explicitly mention names of brands, to find influencers and brands on Instagram. We only kept the influencers who are followed by more than 10,000 people, because that is the generally required number of followers to be considered as an influencer on Instagram. We then downloaded Instagram posts from the influencers and found all brand mentioning posts created by the influencers. Note that we define the user tagging of a brand name in a post as brand mentioning in this study as user tagging is the widely used function to mention a specific user on Instagram. We examined all usertags in the influencers' posts to find brand mentioning. We finally obtained 804,397 brand mentioning posts that are created by 18,523 influencers and contain 26,910 mentioned brands.

According to the Federal Trade Commission (FTC)'s endorsement guides [40], influencers are required to disclose brands' name and sponsorship information when they post a paid advertisement on social media. Based on the FTC's guideline, we were able to further classify the brand mentioning posts into sponsored and non-sponsored posts by checking sponsorship indicators. On Instagram, influencers can provide the sponsorship information either by marking a post as a sponsored post[1] or using special hashtags that indicate sponsorship (e.g., *#ad*, *#sponsored*, *#paidAD*). Therefore, we classified a brand mentioning post as a sponsored post if the post contains one or more sponsorship indicators. We found 139,386 sponsored posts and 665,011 non-sponsored posts in our dataset. Note that we also fetched Instagram profile pages of the influencers and brands to acquire the number of followers.

4.3 Brand Mentioning Network

To study influencers' brand mentioning behavior and understand different characteristics between sponsored posts and non-sponsored posts, we define the notion of the ***Brand Mentioning Network*** where a node is either an influencer, a brand, or a post. Influencer nodes have edges to post nodes which represent a posting relation between an influencer and a post, and post nodes have edges to brand nodes which account for a mentioning relation. Figure 4.1 shows an example of the brand mentioning network with an influencer node, two post nodes, and two brand nodes. In the example, the Influencer A mentions the brand B in the post 1 and both brand B and C in the post 2. Using our dataset, we build the brand mentioning network which consists of 848,814 nodes and 2,000,023 edges.

[1] `https://help.instagram.com/116947042301556`

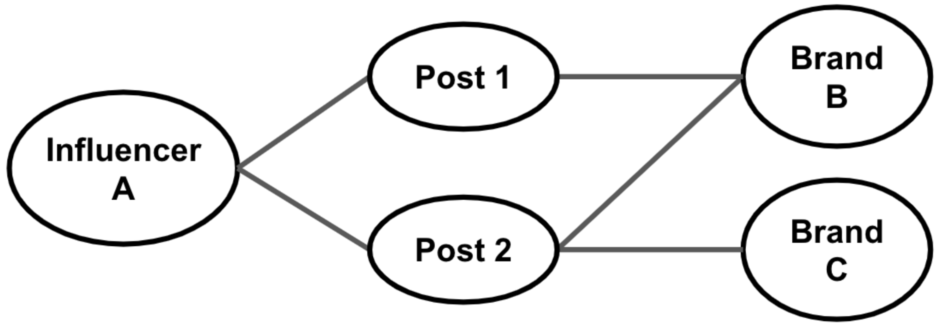
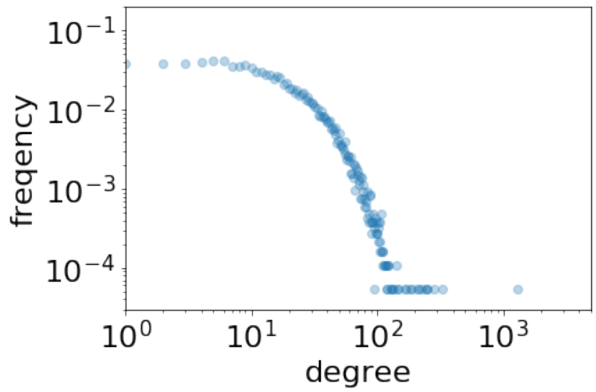


Figure 4.1: Example of the brand mentioning network.

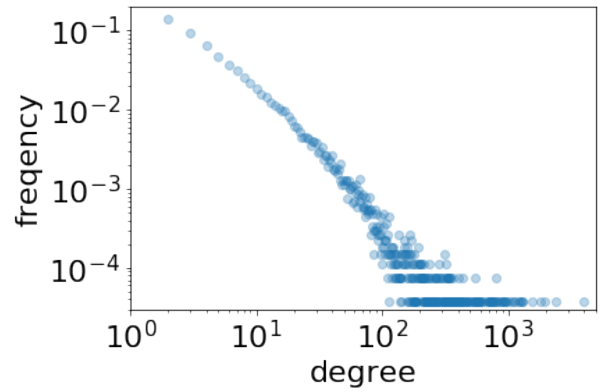
4.3.1 Brand Mentioning Network Analysis

We first analyze the distributions (dist.) of brand mentions in the brand-influencer network where post is not included, and each edge represents an unique mentioning relation between an influencer and a brand. Figure 4.2(a) shows the degree dist. of influencer nodes that reveals how many brands each influencer mentions; and Figure 4.2(b) shows that of brand nodes describing how many influencers that mention the brand. Interestingly, these two dist. are different: most influencers have medium low degree while brands' degree follows the power-law dist. In addition, 34% and 92% of influencers has the degree less than 10 and 50 respectively, and 0.5% of them has degree more than 100. This reveals that most influencers tend to mention a limited number of brands in posts whereas only a few of them mention lots of brands. On the other hand, the numbers are 73%, 94% and 2% for brands, implying that there are more popular brands (e.g., Starbucks, Zara) mentioned by a large number of influencers while most of the brands such as small restaurants or retailers have been mentioned by a small number of influencers.

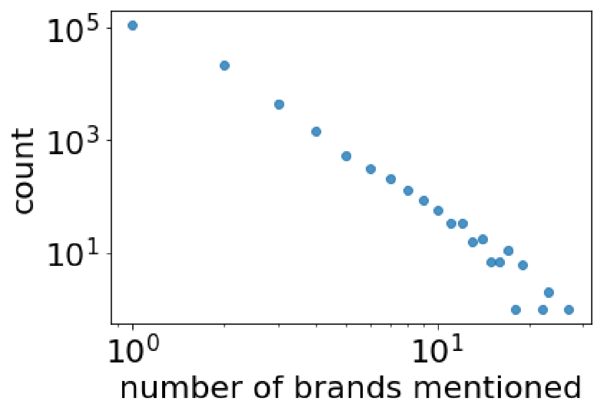
We also examine the number of brands mentioned in the sponsored posts and the non-sponsored posts as shown in Figure 4.2(c) and Figure 4.2(d). We notice that most posts only mention a few brands regardless of sponsorship of the post. Note that 79.3% of the sponsored posts and 73.5% of the non-sponsored posts mention only one brand. We also observe that the non-sponsored posts tend to have more brands mentioned compared to the sponsored posts; 5.3% of the non-sponsored posts mention more than three brands while only 1.0% of



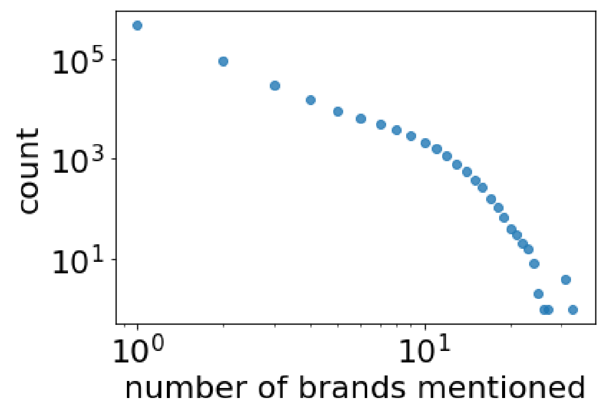
(a) Influencers



(b) Brands



(c) Sponsored post



(d) Non-sponsored post

Figure 4.2: Distributions of brand mentioning in the network

the sponsored posts do. That is because the sponsors want to have the exclusive mentioning of their brand names by the influencers, in the paid advertisement.

We next investigate relationships between the brands and the influencers who are connected through the post nodes in the Brand Mentioning Network. Figure 4.3 shows all brand mentions in our dataset where each dot indicates a mentioning relation between an influencer and a brand, and its color indicates the number of occurrence. We find that popular influencers (e.g., millions of followers) are likely to mention famous brands while micro-influencers (e.g., less than 100K followers) tend to mention any brands regardless of brand popularity. This tendency can be possibly explained as the influencers often get paid from the brands by mentioning brand names in their posts, and most small brands would not be able to

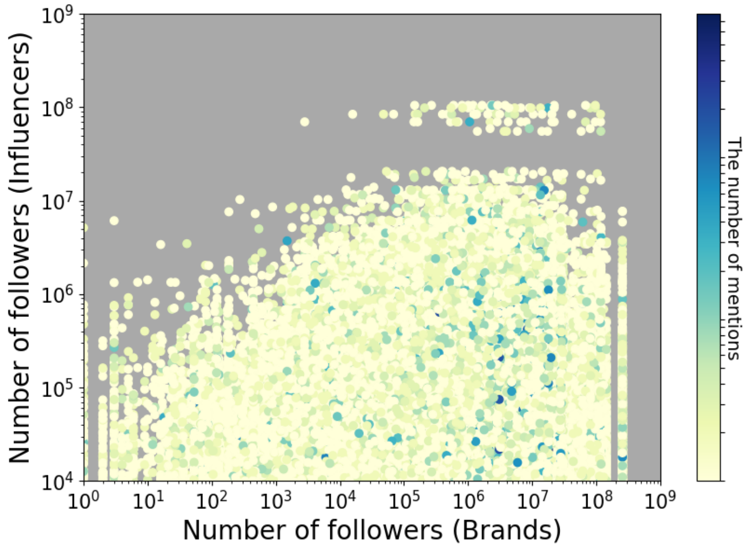


Figure 4.3: Popular influencers who have millions of followers tend to mention more popular brands than other brands with small number of followers. Micro-influencers tend to mention brands regardless of brand popularity.

hire famous celebrities due to the limitation of marketing budget. We also observe that the influencers tend to mention the same brand frequently if brands are popular; As shown in Figure 4.3, more popular brands have darker dots than the less popular brands.

4.3.2 Sponsored Posts vs. Non-sponsored Posts

In this subsection, we further compare the network-level and the post-level properties of the sponsored posts and the non-sponsored posts to understand the effect of sponsorship on brand mentioning behavior.

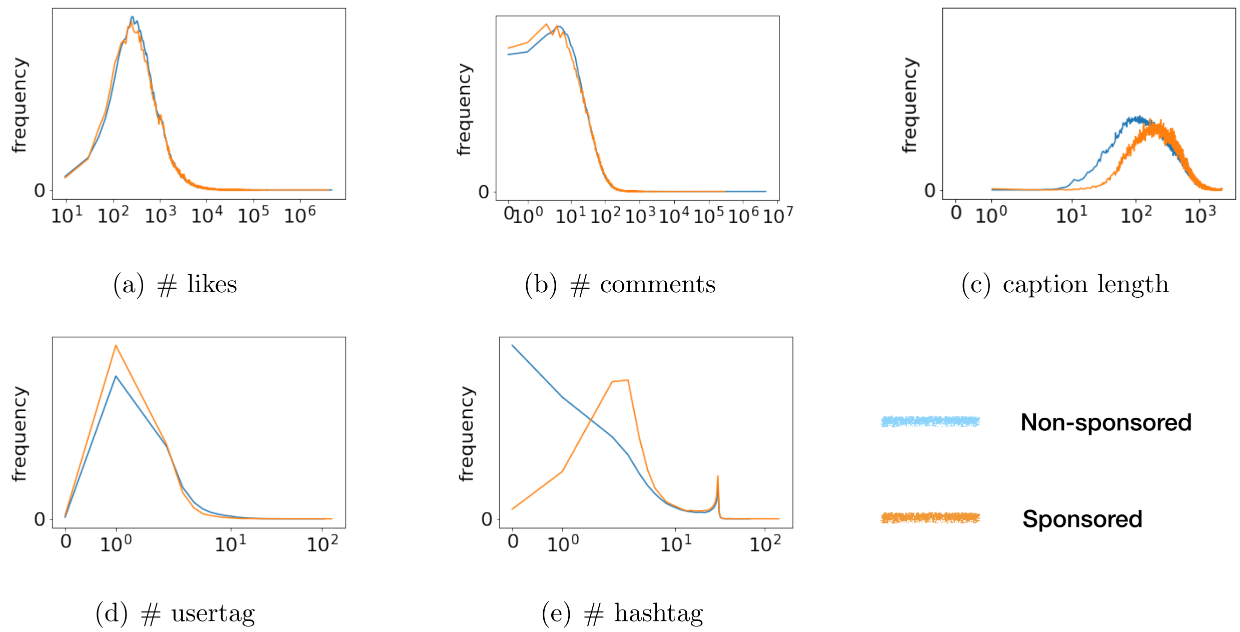
We first test the Spearman coefficient scores to measure the monotonic correlation between properties in Brand Mentioning Network as shown in Table 4.1. Here the degree of brands indicates the number of posts that mention them; and that of influencer indicates the number of mentioning posts they make. We have two observations as follows: (i) We find that the sponsored and non-sponsored degrees of brands are positively correlated with the number of followers of the corresponding brands. This reveals that popular brands tend to

Table 4.1: Correlations between the brand mentioning network properties (*:$p<0.01$, **:$p<0.001$)

Correlation	Spearman coef.
Brand's sponsored degree & # of Brand's followers	0.29**
Brand's non-sponsored degree & # of Brand's followers	0.46**
Influencer's sponsored degree & # of Influencer's followers	0.09**
Influencer's non-sponsored degree & # of Influencer's followers	0.10**

be mentioned and sponsor influencers more than brands that have less number of followers. Although both coefficients show a positive correlation, the correlation of the non-sponsored posts is stronger than that of the sponsored posts. This is because brands with a small number of followers are less likely to be mentioned without sponsorship due to their low brand awareness. (ii) We also find that the number of brand mentions by influencers is not correlated with the number of followers of the corresponding influencers. This implies that there is no significant difference between the brand mentioning behavior of macro influencers and that of micro influencers.

We next analyze statistics of post attributes of the sponsored and the non-sponsored posts. Figure 4.4 shows the histograms of the five post attributes: the number of likes, the number of comments, caption length, the number of usertags, and the number of hashtags. Interestingly, the sponsored and the non-sponsored posts have nearly identical user reactions, in terms of the number of likes and comments, as shown in Figure 4.4(a) and Figure 4.4(b), respectively. This suggests that disclosing sponsorship information in the paid advertisement does not affect the number of likes or comments compared to the non-sponsored posts. We find, however, that there are differences between the sponsored and the non-sponsored posts when influencers create the posts. As shown in Figure 4.4(c), influencers write longer captions in the sponsored posts than the non-sponsored posts since they probably want to share detailed experience on the advertised products and recommend the products to their



(a) # likes (b) # comments (c) caption length

(d) # usertag (e) # hashtag

Figure 4.4: Histogram of post attributes

audience. We also find that influencers use only one usertag more in the sponsored posts to disclose the sponsor name exclusively, and use more hashtags in the sponsored posts than in the non-sponsored posts to expose names of brands' marketing campaigns or sponsorship information as shown in Figure 4.4(d) and Figure 4.4(e). Note that there is a small peak in Figure 4.4(e) since those posts use the maximum number of hashtags which is set as 30 on Instagram.

4.4 Sponsorship Classification

In this section, we propose a neural network model that can automatically classify the sponsorship of posts by learning a network structure of the brand mentioning network as well as node features in the network.

Our proposed model utilizes three sets of features as follows; (i) ***network feature*** which is a network embedding of the brand mentioning network generated by using LINE [159] and its dimension is 128, (ii) ***post feature*** which contains the five post attributes which are the number of likes, comments, usertags, and hashtags, and caption length, and (iii) ***follower***

Table 4.2: Performance of sponsorship classification with different features. P, F, and N indicate the post features, the follower features, and the network features, respectively.

Classifier	Features	Precision	Recall	F1 Score	Accuracy
Proposed model	P	0.633	0.840	0.722	0.677
	P, F	0.673	0.783	0.723	0.699
	N	0.712	0.799	0.746	0.731
	P, F, N	0.795	0.819	**0.806**	**0.803**
Random forest	P	0.638	0.853	0.730	0.684
	P, F	0.715	0.846	0.775	0.754
	N	0.656	0.800	0.721	0.690
	P, F, N	0.737	0.835	0.783	0.769

feature which describes the number of followers of influencers and brands. Note that we normalize all values in the features with the maximum value of the corresponding feature accordingly.

Our model takes the three sets of features of an input post as an input layer, and puts the input into a single hidden layer which dimension size is set to 256. We then add a softmax function to predict the sponsorship of the given input post.

In the brand mentioning dataset, we have 139,386 sponsored posts and 665,011 non-sponsored posts. For training the model, we use the down-sampling to have the same number of sponsored and non-sponsored posts. We then split the dataset into training set (80%) and testing set (20%). We set a batch size as 128 and the number of epochs as 50 with the early stopping. We evaluate the performance of our model with three different baselines that use different sets of features; (i) only post feature, (ii) post feature & follower feature, and (iii) only network feature. We measure the accuracy, precision, recall, and F1 score 10 times by randomly splitting the labeled posts, and average the values. For evaluation purpose, we also use a random forest classifier as the baseline.

Table 4.2 reports the evaluation results of classifiers with different sets of input features and models. We find that the network features significantly improve the model performance which suggests that brand mentioning behavior learned from the network plays an important role in sponsorship detection. Note that our proposed model achieves 80% accuracy while the accuracy of the model is below 70% without the network features. Our model also outperforms to the baseline model which is the random forest classifier that utilizes the same set of features.

CHAPTER 5

Analysis on Influencer Fraud Behaviors in Influencer and Audience Network

5.1 Background

Influencer marketing has become an effective word-of-mouth method that increases brand awareness among numerous potential customers [44, 92]. The effectiveness of the influencer marketing can be suffered from "influencer fraud" that can damage marketing campaigns of companies [108, 88]. That is, influencers may manipulate their influence by obtaining more engagements on their posts using engagement bots that automatically generate likes or comments on social media. Such influencer fraud needs to be seriously considered by marketers since likes or comment counts are used directly as a measure of the success of influencer marketing campaigns [73]. If an influencer hired by a company manipulates his/her influence using engagement bots, the company obtains an advertising effect that is less than what it paid to hire the influencer. In addition, users may lose their trust in influencers, adversely affecting the influencer marketing industry.

Since bots often adversely affect users, companies, and marketers around social media [54], many researchers have investigated the characteristics of bots on social media, such as conversational patterns and network properties [147], user profile and content sentiment [163], which can shed light on differences between bots and authentic users. Also, there have been attempts to develop bot detection models by learning graph similarity [167] or a sequence of contextualized information [95]. However, little attention has been paid to developing bot detection models for influencer marketing, where such a model is useful to accurately assess the influence of each influencer.

To understand the behavior characteristics of bots in social media marketing, we first conduct an empirical study on social media bots by comparing them with authentic audiences of influencers. To this end, we build an influencer engagement network that is a social network of influencers and their audience connected through social engagements, e.g., via liking posts or commenting. In our study, the constructed influencer engagement network is composed of 14,221 influencers, 9,290,895 users, and 65,848,717 engagements.

By analyzing the influencer engagement network, we identify a group of potential engagement bots who generate tons of engagements to a large number of influencers but have zero followers. We perform an in-depth analysis on the distinct characteristics of the engagement bots and reveal that the engagement bots have lower local clustering coefficients in the influencer engagement network than normal users since the bots tend to engage in the posts of random influencers while normal users usually follow influencers with similar interests. Our analysis further reveals that the identified engagement bots tend to write short comments, which are similar to each other, because they are likely to use a set of pre-populated comments to automatically write comments.

Based on lessons learned, we propose a deep learning model that can detect the engagement bots from influencers' audiences. Our model learns contextualized information from comments, engagement behavior, and structural information of the network by taking multimodal inputs including text features, behavior features, and graph features, respectively. The experimental results show that the proposed model outperforms well-known baseline methods by achieving about 80% accuracy. The results also reveal that all three input feature sets are useful to represent social media users while the behavior features play the most important role in detecting engagement bots among the three input feature sets.

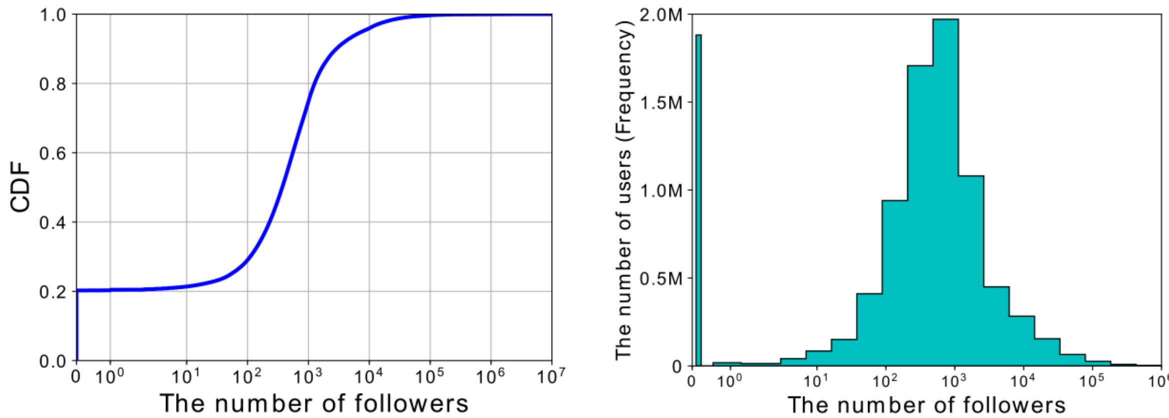
5.2 Influencer Engagement Network

5.2.1 Definition of Influencer Engagement Network

We define the notion of the influencer engagement network as a directed weighted graph $G = (V, E, W)$. In the given network, a node $v \in V$ represents a user who is an influencer or an audience. Two nodes v_i and v_j in the network are connected if node v_i engages in the activities by node v_j. That is, a directed edge $e_{ij} \in E$ from v_i to v_j exists when the user v_i engages in a post published by the influencer v_j through liking or commenting. Note that an influencer can also engage in another influencer's activities, therefore, edges can exist among influencer nodes. Each edge e_{ij} in the network has a weight value $w_{ij} \in W$ which represents the amount of engagements between two nodes v_i and v_j. We calculate the total number of likes and comments that are generated by a node v_i to a node v_j to derive a weight value w_{ij}.

5.2.2 Engagement Data

To build the influencer engagement network, we use the Instagram Influencer Dataset [92]. The dataset contains 33,935 influencers and their 10 M Instagram posts published from 2010 to 2018, where 87% of the posts had been published in 2017 and 2018. In this study, we only use the posts that were published after January 1st in 2017 from the original dataset. After filtering out influencers who have less than 100 posts during the given time period, we use the 6,244,555 Instagram posts published by 14,221 influencers who are followed by at least 10,000 users. We then identify a list of users who have engaged in the posts through liking or writing comments. More specifically, we collect 65,848,717 engagements (21,374,920 likes and 44,473,797 comments) generated by 9,290,895 unique users from the Instagram posts. Note that each post in Instagram provides a partial list of associated users, and we exclude all self-engagements in our study. In addition to the engagement data, we further collect the number of followers for each user by crawling his/her Instagram profile page to examine the characteristics of the users.



(a) CDF

(b) PDF

Figure 5.1: Distributions of the number of followers of the engaged users.

5.2.3 Network Analysis

Based on the constructed influencer engagement network, we analyze (i) the number of followers of users who engage in this network, (ii) the amount of engagements (i.e., weights in the network), and (iii) the connectivity of the network, i.e., how a user or an influencer is connected with other users and/or influencers, to understand engaging behavior of users in the influencer engagement network. We further seek to identify a set of suspicious audience in terms of engaging behavior, who can be considered as engagement bots.

5.2.3.1 Number of Followers of the Engaged Users

We first investigate the number of followers of the influencers' audiences since influencers' posts can be further propagated to the followers of the audience. In other words, influencers can have more influence if their audiences have a large number of followers. Therefore, understanding the size of the social networks of the audience is important in influencer marketing. Figures 5.1(a) and 5.1(b) show the distributions of the number of followers of the engaged users by the cumulative distribution function and probability density function, respectively. As shown in Figure 5.1(a), we first find that around 20% of the engaged users do not have followers. We denote this user group having no follower as the zero-follower

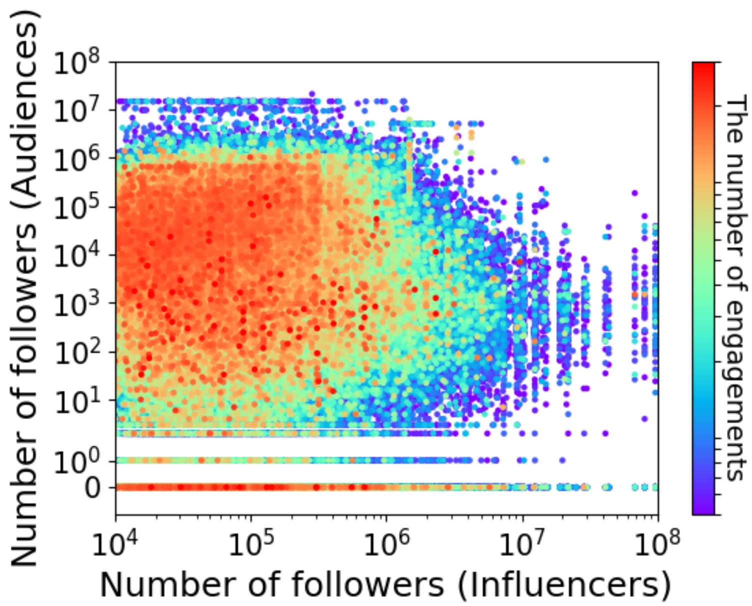
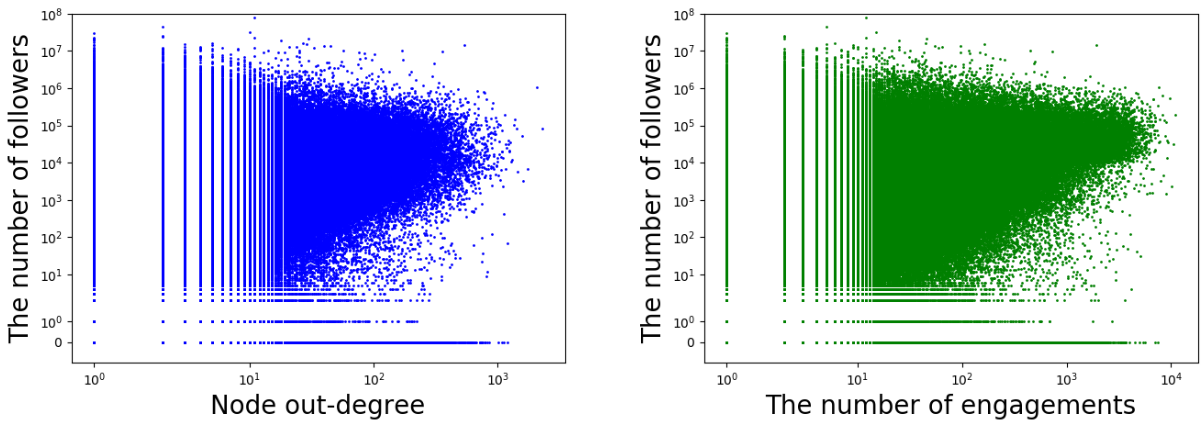


Figure 5.2: Amount of engagements (i.e., weights in the given network) for all pairs of influencers and users.

audience. The users in the zero-follower audience are active in engaging in influencers' activities but inactive in having social relationships with other users. We particularly analyze their engaging behavior in Section 5.2.4 since some of the users in the zero-follower audience are suspicious as 'bot accounts' [54]. The remaining audience excluding the zero-follower audience follows a normal distribution whose median value is 557 as shown in Figure 5.1(b). Note that 46% of the engaged users have between 100 and 1,000 followers, and only 4% of the engaged users have more than 10,000 followers who can be also considered as influencers.

5.2.3.2 Amount of Engagements

We next examine the amount of engagements (i.e., weights in the network) for all unique pairs of influencers and users. The weight value refers to the total number of engagements, thus, high weights indicate consistent and loyal engaging relationships between audience and influencers. Figure 5.2 shows the weights in different colors where red represents higher weight values while purple represents lower weight values. Each dot in Figure 5.2 indicates engagements between a user and an influencer. We find that micro-influencers who have less



(a) Number of engaged influencers (b) Number of total engagements

Figure 5.3: Engagement behavior of influencers' audience.

than 100,000 followers, in general, tend to have more active engagements with their audience than macro-influencers who are followed by millions of users. This suggests that audience of micro-influencers show more consistent and loyal engaging behavior than the audience of macro-influencers. We also find that influencers actively interact with other influencers through engagements. On one hand, most of the active interactions between influencers can be found from influencers less than a million followers. On the other hand, influencers with more than a million followers tend to receive engagements from normal users. This can be due to that influencers with common interests tend to have interactions [90] whereas macro-influencers, who are more likely the celebrities, are likely to focus on their own activities. The result implies that advertising through micro-influencers can be more effective than hiring a few macro-influencers since advertising posts published by micro-influencers are further propagated to the social networks of engaged influencers. Additionally, we find some users in the zero-follower audience show a substantially large amount of engagements. This suspicious users with abnormal engagement behavior might be the social bots [54].

5.2.3.3 The Degree of Engagements

To understand how a user engages in multiple influencers, we analyze the node degree of users and their total number of engagements in Figures 5.3(a) and 5.3(b), respectively. Since a user node has directional edges from itself to engaged influencer nodes, its out-degree represents the total number of influencers engaged by the corresponding user. As shown in Figure 5.3(a), we find that users tend to have higher out-degrees if they have more followers. In other words, influencers are densely connected with other influencers in the engagement network, whereas normal users usually engage in activities of a relatively smaller number of influencers thereby having lower out-degrees. The same tendency can be found in Figure 5.3(b). Despite only 4% of the total audiences have more than 10,000 followers as shown in Figure 5.1, they account for 78% of engagement relationships with more than 100 engagements.

Figure 5.3 shows that there are a set of users with high out-degrees and a large number of engagements, but having no followers, unlike influencers. More specifically, this group of users has different engagement behavior compared to other users as they are not socially connected with other users but contribute a huge number of engagements to numerous influencers. This unique engagement behavior can be generated by social media bots [54]. To conduct an in-depth study on these suspicious user accounts, we set a threshold on the node out-degree as 300, and assume that the zero-follower audience whose node out-degree is higher than 300 can be considered as bots. In our dataset, 206 users in the network are classified as (potential) bots.

5.2.4 Fake Engagement by Bots

Understanding the characteristics of the potential engagement bots can be useful in identifying influencers who are involved in fake engagements. To this end, we analyze distinct attributes of the potential bots in terms of their (i) local clustering coefficient and (ii) commenting behavior.

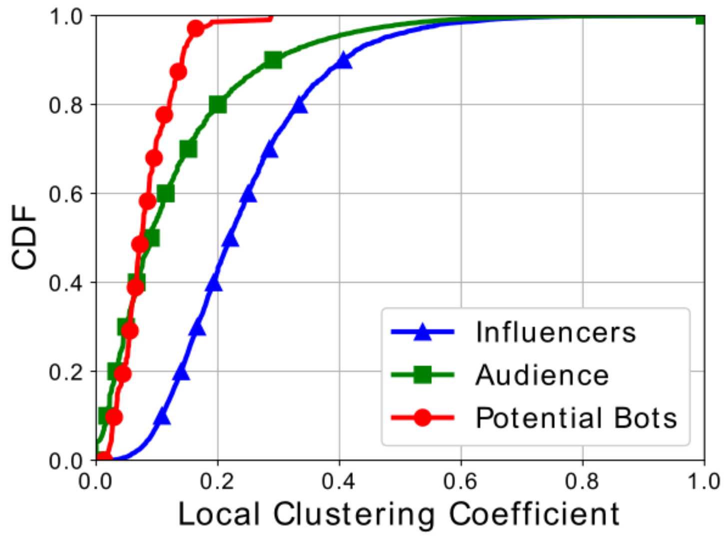


Figure 5.4: The potential bots have relatively lower local clustering coefficient values than the normal audience of the influencers while the influencers tend to have higher clustering coefficients than the normal users.

5.2.4.1 Local Clustering Coefficient

It is a widely observed property of social networks that a node's friends are often themselves friends in the network. This property is also observed in the influencer engagement network where a user engages in the influencers who themselves have interactions through engagements. To examine how normal audiences and potential bots show different engagement behavior, we calculate the local clustering coefficient of each node in the engagement network. Note that the local clustering coefficient of a node represents the ratio of friendship between the node's neighbors to all possible pairs of neighbors. Suppose user i engages in a set of influencers $N(i) = \{j | (i, j) \in E\}$ where E is the edge set in the engagement graph G. Then the local clustering coefficient of user i can be calculated as follows:

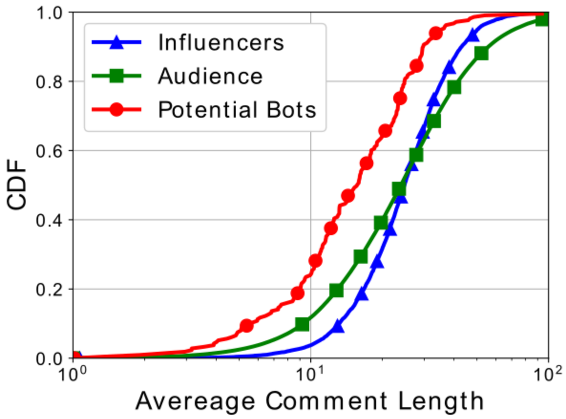
$$F(i) = \frac{|\{e_{jk} \in E | j, k \in N(i)\}|}{\{|N(i)| \times (|N(i)| - 1)\} \div 2}$$

Figure 5.4 shows the local clustering coefficients of the influencers, the normal audiences, and the potential bots, respectively. We first find that the potential bots tend to have lower clustering coefficients than normal users. Note that most of the potential bots have the clustering coefficients lower than 0.2. This is because normal users tend to follow or engage in a group of influencers with similar interests, hence having active social interactions among the influencers. On the other hand, the potential bots are likely to engage in random influencers thereby having little interactions between their neighbor nodes. We also find that the influencers have significantly higher clustering coefficients than other users. This confirms that influencers form clusters in the engagement network by liking or writing comments to posts published by other influencers. As reported in [90], influencers actively tend to interact with other influencers to get attention from followers of the other influencers and ultimately increase their followers. The active engagement behavior of the influencers also helps their posts to be exposed to a set of users who may like or comment on the same posts that the influencers already posted.

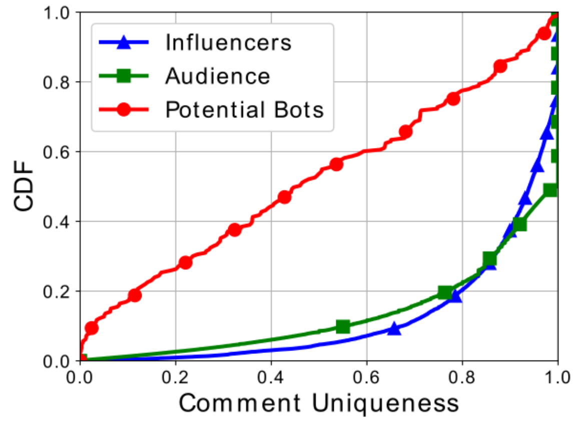
5.2.4.2 Commenting Behavior

To analyze distinct commenting behavior by potential bots, we next investigate all the comments on the influencers' posts. Note that we have 44,473,797 comments in our dataset and we discard influencers' self comments on their own posts in this analysis. We first analyze the length of the comments as shown in Figure 5.5(a). We find that potential bots write short comments on the influencers' posts than other normal users. The average length of comments written by the influencers, the normal audience, and the potential bots are 27.0, 29.7, and 17.5, respectively. This implies that the potential bots write comments with simple expressions such as "`Nice picture!`" whereas normal users tend to have more information or their feelings on their comments, for example, "`Like your dress! Where did you buy it?`" thereby having longer comments than the bots. We also find that the influencers have smaller variance on the average comment length than the normal audience, although they have similar average values. This suggests that the influencers are not likely to use simple emojis or a single word on their comments, which are often used by the normal audience, to interact with other influencers.

We next investigate the uniqueness of the comments written by users and bots. Since



(a) Comment length



(b) Comment uniqueness

Figure 5.5: The potential bots tend to write short comments and have very low comment uniqueness since they use a set of pre-populated comments to automatically make engagements to other social media users.

social media bots usually use a set of pre-populated comments, the exact same comments by a single bot can be attached to the influencers' posts. The normal users, on the other hand, write comments by themselves, therefore comments are likely to be different from each other. Suppose user i writes a set of comments $C(i) = \{c_n\}_{n=1}^{|C^i|}$, where $|C^i|$ indicates the number of comments written by user i. We then define the comment uniqueness of the user $H(i)$, the rate of the number of unique comments to the total number of comments, as follows:

$$H(i) = 1 - \frac{|\{c_x \in C^i | c_x = c_y, c_y \in C^i\}|}{|C^i|}$$

For example, a user has the comment uniqueness as 1 if all of the comments written by the user are different from each other. If a user uses the same set of comments repeatedly then the user has the comment uniqueness as 0. Figure 5.5(b) shows the distributions of the comment uniqueness values for the influencer, the normal audience, and the potential bots, respectively. We find that the potential bots have significantly lower comment uniqueness than the normal users. Note that the average comment uniqueness of the influencers, the normal audience, and the potential bots are 0.88, 0.87, and 0.47, respectively. This result confirms that the potential bots engage in influencers by using the same set of comments

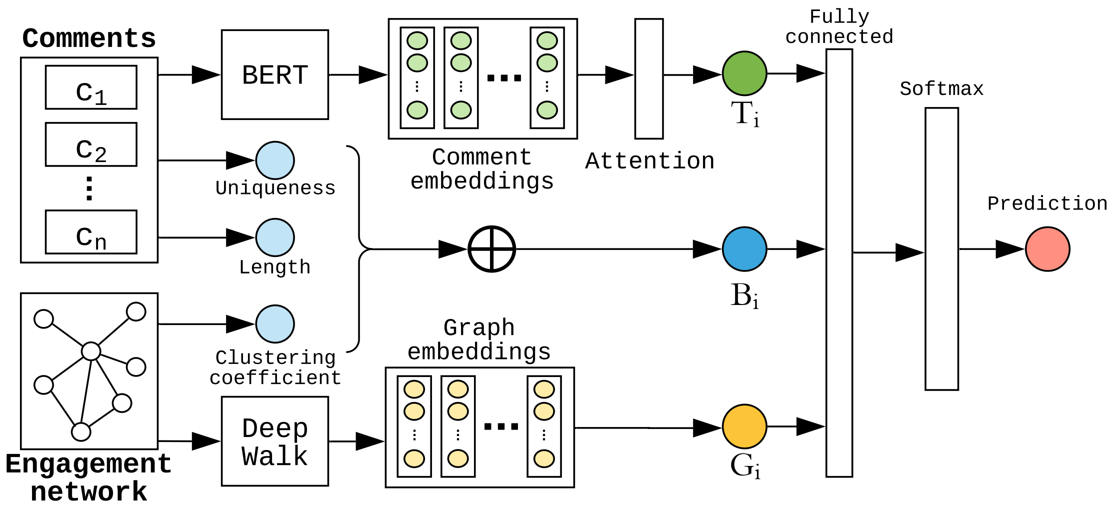


Figure 5.6: The overall architecture of the proposed model for detecting engagement bots.

rather than creating a new comment for each post.

5.3 Engagement Bot Detection

In this section, we propose a deep learning model that can detect engagement bots from a numerous audience who engages in the posts of influencers.

5.3.1 Model Description

5.3.1.1 Model Architecture

Figure 5.6 illustrates the overall architecture of the proposed model for detecting potential engagement bots. The aim of the model is to detect engagement bots from audience data by learning their social relationships and engagement behaviors. The proposed model utilizes three sets of input features including text features, behavior features, and graph features.

To model contextualized information from comments written by the audience, we apply the pre-trained neural language model, BERT [46]. The BERT takes all comments of a user i, $C(i) = \{c_1, c_2, \cdots, c_n\}$, and generates the comment embedding whose dimension is 768.

Note that we set the sequence length of an input comment as 100 since most comments have their lengths shorter than 100. We then apply the attention mechanism [8] to give more weights on more important comments to represent each user. The attention layer projects the comment embeddings into a hidden space and estimates the importance score of each comment. We finally obtain the text features, T_i, by using a weighted combination of comment embeddings and corresponding scores. We set the dimension of the text features as 128.

The proposed model also exploits the behavior features, B_i, to capture distinct engagement behavior of bots from the normal audience. The behavior features contain the three engagement properties, the comment uniqueness, the average comment length, and the local clustering coefficient, analyzed in Section 5.2.4. We compute the property values of each user by taking user comments and the structural information of the engagement network, then concatenate the values to generate the behavior features.

In addition to the text and behavior features, we also encode the influencer engagement network to learn social relationships based on the user engagements. To generate a graph representation, G_i, we use DeepWalk [132] that employs short random walks from nodes in the network. Note that we set the number of walks, the length of the random walk, and the dimension of the graph embedding as 10, 20, and 64, respectively.

We concatenate the three sets (text, behavior, and graph) of each user representation, and add a fully connected layer with the Rectified Linear Unit (ReLU) as the activation function. The dimension of the hidden layer is set to 128. Finally, we utilize the cross-entropy for the loss function to detect the engagement bots.

5.3.1.2 Training Procedure

Since the number of users who are considered as potential bots is highly imbalanced to the total number of audience in our data, we use both oversampling and undersampling techniques to have balanced training data. More specifically, we first apply SMOTE [29] to the 206 bot accounts to make 10,000 samples. We then undersample normal users to have

Table 5.1: Performance comparison with well-known baseline methods. Features T, G, and B represent the text features, the graph features, and the behavior features, respectively.

Method	Features	Precision	Recall	F-1 score	Accuracy
SVC	*All*	0.783	0.720	0.750	0.755
RandomForest	*All*	0.792	0.760	0.776	0.776
Proposed Model	T,G	0.714	0.652	0.682	0.689
	B,G	0.750	0.652	0.698	0.711
	T,B	0.762	0.696	0.727	0.733
	All	**0.826**	**0.760**	**0.792**	**0.796**

the same number of samples with the bot accounts. After making the balanced dataset, we split our data into training and testing sets with 8:2 ratio. We set the number of epochs and the learning rate as 100 and 0.0001, respectively. Note that we do not use follower counts of the users in the engagement network, which are directly implicated for identifying bot accounts, to prevent information leak in the training procedure.

5.3.2 Experiment Results

Table 5.1 shows the performances of the proposed model and two well-known baseline methods including the Support Vector Classifier (SVC) and the Random forest. To evaluate the performance, we use the precision, recall, F-1 score, and accuracy as evaluation metrics. As shown in Table 5.1, we find that our model outperforms the baseline models that use the same sets of features with the proposed model. This suggests that the proposed neural network is more effective in capturing useful features for detecting possible engagement bots than others.

In addition to the performance comparison with the baseline methods, we also conduct experiments with different sets of input features to analyze the importance of the different input features. We measure the performance of the model by excluding one set of features from the text, behavior, and graph features, respectively. The results show that the performances of the proposed model without one set of features are lower than the baseline methods. This implies that all three sets of features efficiently represent the distinct characteristics of the engagement bots from normal users. We find that the behavior features are more effective for detecting the engagement bots than the text and graph features. Note that the performance losses are observed in the F-1 score of the behavior, text, and graph features as 13.9%, 11.9%, and 8.1%, respectively. This reveals that the comment uniqueness, the average comments length, and the clustering coefficient values in the behavior features are important predictors in capturing the engagement behavior of bots. Moreover, the text features, which are the attentive comment embeddings, also can exhibit distinctive commenting patterns of the engagement bots. The graph features, on the other hand, show relatively less performance loss than the other features. That is probably because a few normal users tend to have high out-degrees thereby having similar graph representations with the engagement bots.

CHAPTER 6

Multimodal Post Attentive Profiling for Influencer Marketing

6.1 Introduction

Influencer marketing [61], which utilizes special individuals in social media, has gained great attention from brands. Brands expect to promote brand awareness and advertise products to social networks of influencers [105, 82, 53, 161, 66], who have 'influence' over a large number of followers [10, 27], since customers are often more likely to trust influencers' recommendations than brands' advertisements [33, 122, 161]. It has been reported that the global influencer market value was estimated to be 2 billion U.S. dollars as of 2017 and will increase to 10 billion U.S. dollars by 2020 [42]. The growing interest in influencer marketing has led many social media users to participate in marketing campaigns and create advertising content [76]. For example, Instagram influencers created 9.7 million brand-sponsored posts in 2016, and the volume will reach to 32.3 million posts in 2019 [56].

Due to its popularity, brands tend to increase their budgets for influencer marketing [98], and researchers have started studying various aspects of influencer marketing such as roles of influencers in social media [44, 50, 90, 157, 105, 82]. However, big challenges remain for both brands and researchers in the rapidly growing market. So far, most brands have relied on influencer marketing agencies to hire influencers, but such agencies usually have a limited number of influencers who registered in their services, which may limit the chance to find more proper influencers, i.e., influencers who work with other agencies or did not register to any marketing agency are not considered [113]. Also, previous studies on influencer marketing mostly relied on small datasets that are acquired through surveys of influencers [44, 50, 105] or finding a few influencers on Instagram [82, 90, 157] due to lack of available influencer data.

We, therefore, believe developing an influencer profiling model that can classify influencers with specific interests can provide valuable information for brands in their influencer hiring process or marketing strategies. Furthermore, constructing a large-scale and informative influencer dataset using such a model can foster researchers to conduct in-depth research, for example, analyzing the influencers-brands relationship, targeted audiences' responses, and influencer marketing effectiveness, as well as building efficient recommendation systems. Consequently, such research will enable brands to understand how to initialize influencer marketing campaigns, manage relationships with influencers, and effectively promote advertisements through influencers. Also, an influencer profiling model enables brands to easily find influencers for their marketing campaigns, without hiring them from marketing agencies.

To shed light on the above issue, we propose a multimodal convolutional neural network model that uses text and image information to classify influencers into specific interests such as beauty or travel. More specifically, our model takes the text and image features from social media posts published by an influencer to generate an influencer embedding by using the attention mechanism. The attention helps find posts that are more relevant to the influencer's topic, thereby obtaining a better influencer representation than existing user profiling methods [129, 130, 75, 178, 51]. In addition to the influencer classification task, our model also classifies all posts into certain post categories (e.g., fashion, food, or interior). In our work, we use Instagram as our research context since it is the most popular social media website for influencer marketing [118]. We define a (special) individual as an influencer if he/she has more than 1,000 followers, which is the minimum number of followers required in influencer marketing in general [114]. Note that our model can be generalized and applied to any social media (e.g., Facebook, Twitter, Pinterest), where users post images with texts.

6.2 Multimodal Post Attentive Influencer Profiler

In this section, we first state the problem to describe the objectives of our proposed model. we then present the proposed framework to classify both influencers and their posts by taking text and image features from posts.

6.2.1 Problem Statement

Here we formally define the goal of this paper. Given the social media posts of an influencer $P = \{p_1, \cdots, p_n\}$, we aim to classify the influencer into the corresponding category of $c \in \mathcal{C}$, where $\mathcal{C}$ is the classification space of categories. Moreover, each social media post p_i multimodally consists of a raw image p_i^I and a piece of texts p_i^T. More specifically, a raw image can be considered as a tensor of pixel values while texts of a post are an ordered lists of tokens.

6.2.2 Framework Overview

Figure 6.1 shows the overall schema of the proposed model, *Influencer Profiler*. The multimodal post encoder first encodes a set of posts of a given influencer to generate the post representations. The post encoder takes text and image from a post and obtains text features and image features by using pre-trained models, BERT [46] and Inception-v3 [158], respectively. The post encoder concatenates the text and image features to make a post representation. Next, the post attention layer takes post representations as an input and calculates a score for each post. The distribution of post scores is then used to output the influencer representation. Finally, our model predicts the category of a given influencer based on the influencer representation. Our proposed model is also capable of predicting the category of a social media post by learning the post representations.

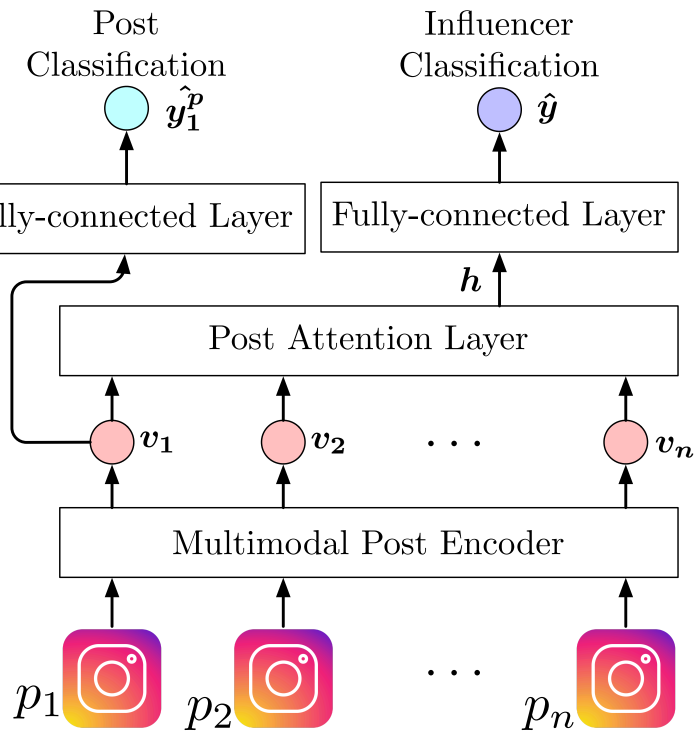


Figure 6.1: The overall schema of *Influencer Profiler*.

6.2.3 Multimodal Post Encoder

To leverage the multimodal knowledge in posts, we propose the multimodal post encoder to derive a continuous representation for each post of the influencer using both text and image information. Figure 6.2 illustrates the architecture of the multimodal post encoder. Texts are encoded by a pre-trained BERT model [46] as text features while the image features of a post are derived by a pre-trained Inception-v3 [158].

Image Features. To generate post image features of a post p_i, we use the pre-trained Inception-v3 [158] model. We apply the transfer learning technique using the pre-trained model instead of training the model from scratch because the number of manually labeled posts is relatively small. In the transfer learning, the parameters on frozen layers are fixed and never updated. We fine-tune only the top 2 layers because those are what directly influence on the determination of the category, and we want to keep the same trained low-level feature detectors in the hidden layers, which was trained with much larger and more robustly spread

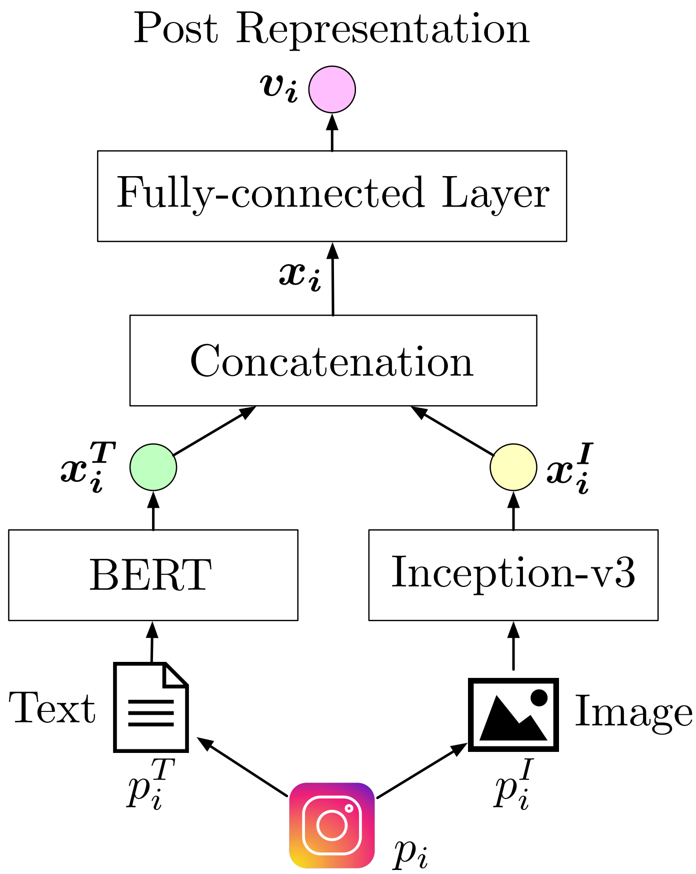


Figure 6.2: The architecture of the multimodal post encoder.

dataset (ImageNet [45]). ImageNet contains 1 M images in 1,000 classes, which covers quite a variety of categories. In the image feature network, there are 49 hidden layers. We add a global spatial average pooling layer on top of the original output layer of Inception-v3, then add a fully-connected layer with the rectified linear unit (ReLU) activation function after the pooling layer to generate an image feature vector, $\boldsymbol{x_i^I}$, which has 1,024 dimensions.

Text Features. We exploit the pre-trained text model to derive text features. We use the BERT [46] model, which has 12 layers with 110 M parameters because it can capture contextualized information from text by applying the bidirectional transformers in the training procedure. We set the maximum sequence length as 128 and use only the last layer to obtain the text features. We then select the output of `[CLS]` token which is inserted at the beginning of an input sentence. Finally, output text feature vector $\boldsymbol{x_i^T}$ for a post p_i has 768 dimensions.

After deriving the image and text features, the ultimate feature vector can be derived by concatenating the features of two resources:

$$\boldsymbol{x_i} = [\boldsymbol{x_i^I}; \boldsymbol{x_i^T}].$$

Finally, the continuous representation $\boldsymbol{v_i}$ of the post p_i can be derived as:

$$\boldsymbol{v_i} = \mathrm{ReLU}\left(\mathcal{F}\left(\boldsymbol{x_I}\right)\right),$$

where $\mathcal{F}\left(\cdot\right)$ is a fully-connected layer; $\mathrm{ReLU}(\cdot)$ is the activation function.

6.2.4 Post Attentive Influencer Encoder

The influencer encoder generates a continuous influencer representation by taking a set of post feature vectors $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_n}\}$, We use the attention mechanism [8] to obtain the influencer embedding since all posts are not equally important to represent the category of the given influencer. For example, a small number of beauty posts published by a food influencer may not be very important to decide the influencer class. Instead, food posts should be considered as more important posts for classifying the category of food influencer. Therefore, the attention mechanism can be applied to weigh higher scores on more important posts.

For each post p_i of the influencer, a fully-connected layer with an activation function is first applied to project the post features into a hidden space as:

$$\boldsymbol{r_i} = \tanh\left(\mathcal{F}_a\left(\cdot\right)\right),$$

where $\mathcal{F}_a(\cdot)$ is a fully-connected later; $\tanh(\cdot)$ is the activation function. A trainable context vector $\boldsymbol{r^c}$ is then exploited to estimate the importance α_i of each post p_i with a softmax function as:

$$\alpha_i = \frac{\exp(\langle \boldsymbol{r_i}, \boldsymbol{r^c} \rangle)}{\sum_j \exp(\langle \boldsymbol{r_j}, \boldsymbol{r^c} \rangle)},$$

where $\langle \boldsymbol{r_i}, \boldsymbol{r_j} \rangle$ denotes the inner-product of $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$. Finally, the influencer representation $\boldsymbol{h}$ can be constructed by a weighted combination of post features as:

$$\boldsymbol{h} = \sum_i \alpha_i \cdot \boldsymbol{v_i}.$$

Influencer Classification. Based on the influencer representation $\boldsymbol{h}$, the logits of influencer classification can be computed as:

$$\hat{\boldsymbol{y}} = \mathcal{F}_I(\mathrm{ReLU}(\mathcal{F}_h(\boldsymbol{h})),$$

where a fully-connected layer $\mathcal{F}_h(\cdot)$ and the activation function $\mathrm{ReLU}(\cdot)$ perform a non-linear transformation while another fully-connected layer $\mathcal{F}_I(\cdot)$ infers the ultimate logits of categories for influencer classification.

6.2.5 Auxiliary Post Classification

In addition to influencer classification as the main task, we propose to improve the model by considering an auxiliary task. More specifically, our proposed model further classifies each post of the influencer into a post category $c_p \in \mathcal{C}_p$. Note that the classification space of auxiliary post classification $\mathcal{C}_p$ is not necessary to be identical to the space of influencer classification $\mathcal{C}$. If the post features are effective enough for post classification, the knowledge in the auxiliary task can be also leveraged to the main task of influencer classification.

By taking the representation $\boldsymbol{v_i}$ derived from the post encoder for each post p_i, the model computes the logits of post classification as follows:

$$\hat{\boldsymbol{y_i^P}} = \mathcal{F}_p(\mathrm{ReLU}(\mathcal{F}_s(\boldsymbol{v_i})),$$

where we conduct a non-linear projection with a fully-connected layer $\mathcal{F}_s(\cdot)$ and an activation function $\mathrm{ReLU}(\cdot)$; another fully-connected layer $\mathcal{F}_p(\cdot)$ generates the logits for post classification.

6.2.6 Multi-task Learning

In this work, we learn the ***Influencer Profiler*** with multi-task learning for both tasks of influencer classification and post classification.

For influencer classification, we treat the task as a multi-class classification problem and utilize the cross-entropy [63] as the loss function. More precisely, the loss function of influecner classification as the main objective can be computed as:

$$\text{loss}_{\text{main}} = \sum_{c \in \mathcal{C}} P(c \mid \boldsymbol{y}) \log P(c \mid \hat{\boldsymbol{y}}),$$

where $P(c \mid \boldsymbol{y})$ is the ground truth class distribution; $P(c \mid \boldsymbol{haty})$ is the estimated probability for the influencer category c by the logits $\hat{\boldsymbol{y}}$ and a softmax function.

As an auxiliary task, post classification can be also treated as a multi-class classification. Hence, the loss function of post classification can be written as:

$$\text{loss}_{\text{aux}} = \sum_{p_i \in P} \sum_{c \in \mathcal{C}_p} P\left(c \mid \boldsymbol{y_i^p}\right) \log P\left(c \mid \hat{\boldsymbol{y}}_{\boldsymbol{i}}^{\boldsymbol{p}}\right),$$

where $P\left(c \mid \boldsymbol{y_i^p}\right)$ is the ground truth class distribution for the post p_i; $P\left(c \mid \hat{\boldsymbol{y}}_{\boldsymbol{i}}^{\boldsymbol{p}}\right)$ can be also estimated by a softmax function.

Finally, the ultimate objective for multi-task learning can be a combination of two loss function as:

$$\text{loss} = \text{loss}_{\text{main}} + \text{loss}_{\text{aux}}.$$

6.3 Data Collection and Annotation

In this section, we first describe our dataset used in this paper. We collected Instagram posts that contain images, texts, and meta-data (e.g., tags, numbers of likes and comments), and user profiles (e.g., biography, number of posts, numbers of followers and followees). We then illustrate how we define categories of influencers and categories of posts.

6.3.1 Dataset Collection

To collect the post and user data of Instagram influencers, we first need to find user names of influencers. Since there is no function to find influencers on Instagram, we employed the following data collection process that includes the three steps: (i) collecting user names of potential influencers by searching posts that contain specific hashtag(s) which are widely used by Instagram influencers, (ii) filtering out users with less than 1,000 followers, and (iii) downloading posts including images and texts from the collected influencers. Note that we followed the social media data collection ethics by collecting only public posts and anonymizing the dataset [160]. To this end, we removed personal information from users' profiles and converted Instagram user names to random integer numbers using a hashmap. We used a third-party Instagram API called *InstaLooter* [97] to collect the post and user data from Instagram.

As the first step of our data collection, we searched posts with hashtag(s) to collect user names of potential influencers. According to the FTC's Endorsement Guides [40], influencers are required to disclose brand information by explicitly mentioning 'paid advertisement' and their relationships with brands if they advertise brands' products. Thus, on Instagram, most influencers are likely to use particular marketing related hashtags, for example, #*ad* or #*sponsored*, to indicate that the posts are paid advertisements, or are likely to use usertags to disclose the brand names. Therefore, we collected Instagram posts that contain the hashtag #*ad* which is the most commonly used hashtag for influencer marketing on Instagram. Note that we use the hashtag #*ad* to find influencers with diverse interests because this hashtag is widely used regardless of types of influencers. The other hashtags (e.g., #*OOTD*, #*EatingForTheInsta*) also can be used in finding potential influencers in certain categories. As of October 2019, there are 10,372,177 Instagram posts that contain #*ad*. We periodically queried the hashtag #*ad* in every 10 minutes, and downloaded meta-data of newly updated posts that contain Instagram user names. Note that a search result page displays all the public posts with the queried hashtag, which are displayed in the chronological order by the

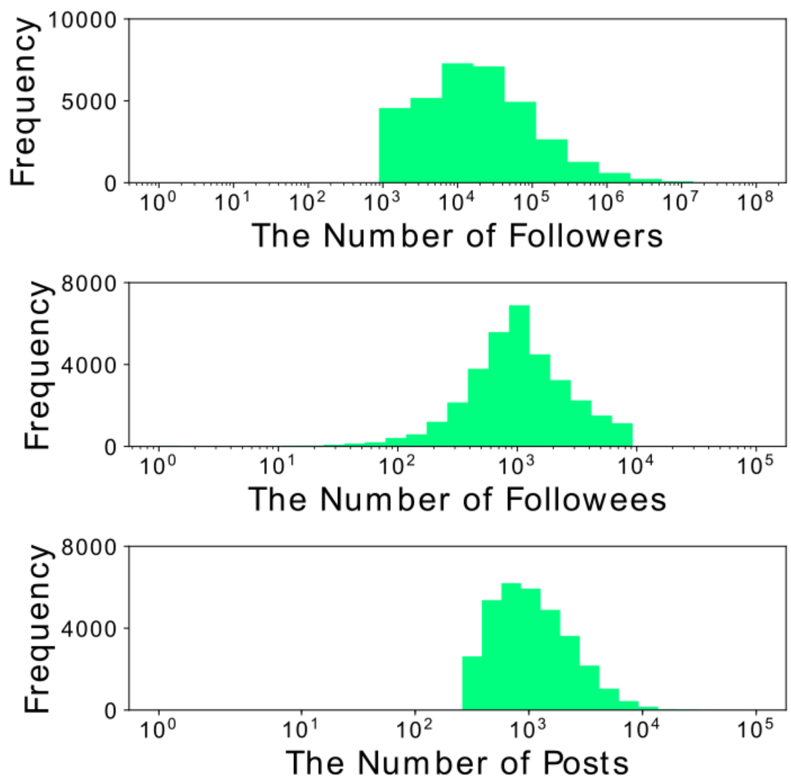


Figure 6.3: Distributions of numbers of followers, followees, and posts of Instagram users in the collected influencer dataset.

posting date[1]. We had collected the data for 92 days from October 1st, 2018 to January 1st, 2019, which contains 828,045 posts by 107,656 unique Instagram users who may be potentially influencers.

Since any user can use the marketing-related hashtags in their posts even if they do not work with brands, the obtained data may have a possibility of containing the mixture of posts by influencers and non-influencers. Hence, we filtered out those non-influencers who have less than 1,000 followers in the second step of our data collection. By downloading an HTML page of each user which includes his/her biography and number of followers, we could examine the number of followers of each user, and narrowed down to 81,928 users. We further excluded users who had less than 300 posts. Finally, our dataset includes 33,935

[1] `https://help.instagram.com/355932664593846?helpref=faq_content`

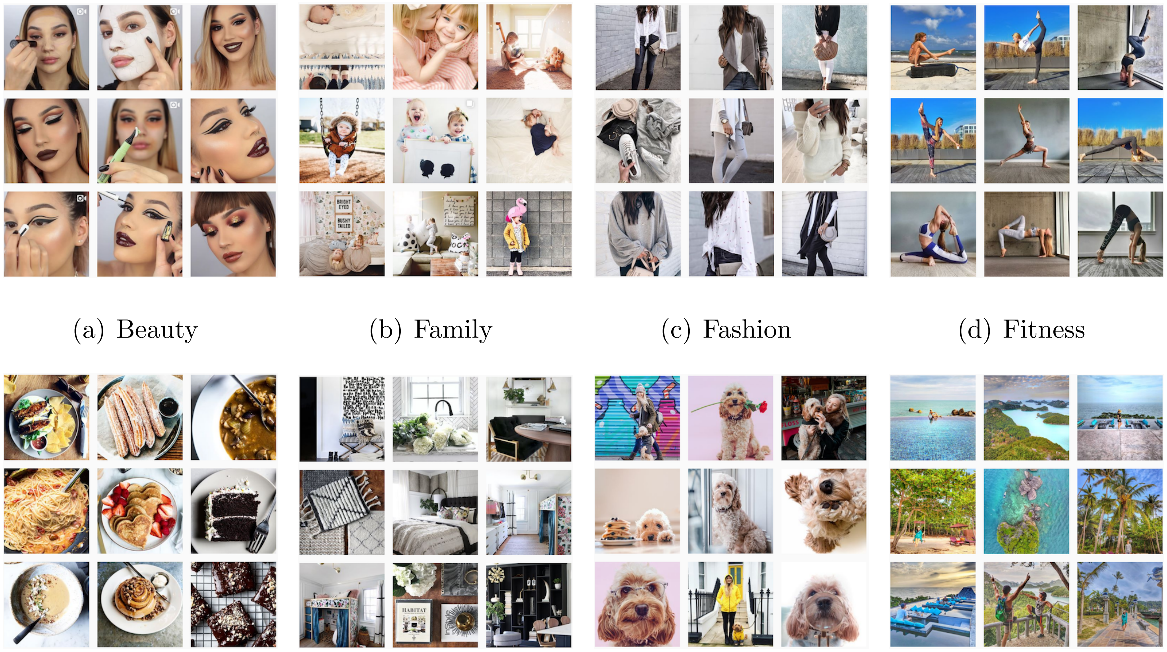
influencers. Figure 6.3 shows the distributions of numbers of followers, followees, and posts of the obtained influencers. Notice that the portions of influencers who have followers less than 10,000, 25,000, and 50,000 are 37.38%, 60.22%, and 73.77%, respectively, and they are often dubbed as *micro-influencers* [114]. This result confirms that micro-influencers are actively participating in influencer marketing on Instagram, and brands are likely to work with them to reach targeted audiences.

In the last step of our data collection, we collected the post information of identified influencers. We downloaded 300 recent post information of each influencer, which includes image files and their meta-data information. We only used the latest 300 posts to reflect influencers' recent interests; an influencer could have changed his/her interest (e.g., fashion, travel) a long time ago, which can possibly lead a misclassification of the current influencer's category. Also, 300 posts for each influencer are sufficient to accurately classify the influencer category. The meta-data of a post contains a caption, hashtags, usertags, time-stamp of posting, number of likes, and associated comments. We finally obtained 10,180,500 posts from 33,935 influencers.

6.3.2 Category Identification

Influencer Categories. The number of followers of an influencer is one of the important factors to evaluate the influencer [10]. It is widely used to categorize the types of influencers, e.g., micro-influencers or macro-influencers [87]. In addition, topics (or categories) that influencers are interested in are also a crucial factor in evaluating influencers for a marketing purpose; influencers are expected to be experts in a specific topic that brands are interested in. Therefore, brands seek to find influencers who are specialized in specific fields to reach the target audiences effectively.

In this work, we define major categories of influencers based on the topic modeling. As influencers tend to introduce themselves by describing their interests in biography, we apply the Latent Dirichlet Allocation (LDA) topic model [166] to the biographies of the 33,935 influencers in the dataset. We first found ten topics, each of which is composed



(a) Beauty (b) Family (c) Fashion (d) Fitness

(e) Food (f) Interior (g) Pet (h) Travel

Figure 6.4: Example images of Instagram posts of influencers in the eight influencer categories.

of ten words, and then manually selected the most proper words that can represent the topics of influencers' interests. In other words, we removed the words that cannot represent influencers' interests such as 'influencer', 'blogger', 'loving', and etc. We finally identify eight major categories of influencers, *beauty*, *family*, *fashion*, *fitness*, *food*, *interior*, *pet*, and *travel*, as shown in Figure 6.4. To verify whether our identified eight categories are applicable to actual influencer marketing on Instagram, we investigated over 30 influencer marketing companies, e.g., *HYPR*[2], *DeepSocial*[3], to examine how they identify influencer categories.

Post Categories. While we categorize topics of influencers into the eight categories, we define 10 categories of posts, which are the same categories of the eight influencer categories and two additional post categories including *product* and *other*. We add the *product* post category, which contains only products (e.g., cosmetics, fashion accessories, electronics) in

[2] `https://hyprbrands.com/`

[3] `https://deep.social/`

the photo because many influencers often use product photos for advertising purposes. Posts that are not classified into any category, e.g., music, sports, arts, or humor posts, are labeled as *other* category. Note that categories are not limited to the classes proposed in this paper, and can be easily added as needed.

6.3.3 Category Labeling

Influencer Labeling. We manually label influencers by examining their biographies and posts. We make two sets of labeled influencers for (i) training and validation and (ii) testing. In the first set, we label 200 influencers in each one of the eight influencer categories, which makes 1,600 labeled influencers in the first set. We then split the set into 8:2 for training and fine-tuning the model. While the first set is balanced to avoid biased results, we randomly select and label 1,142 influencers in the second set for testing purposes.

Post Labeling. To train data for the post classifier, we randomly select posts of the influencers from our dataset, and then manually label them into one of the 10 categories. Note that we do not take account of the posts that do not contain any text in the caption. Then, we obtain more than 1,000 posts for each post category and select 1,000 posts, which contain images as well as their captions, in each post category. Finally, we have 10,000 labeled posts.

6.4 Experiments

6.4.1 Experimental Setting

We use TensorFlow [1] to implement our proposed model. We set the dimension size of the output of the post encoder as 256. For training the model, we set learning-rate as 10^{-3} and dropout probability as 0.5. The batch sizes for the post encoder and the influencer encoder are set as 128 and 16. We apply the heavy data augmentation to all images in our dataset by spacial rotation, spacial shifting, flipping, zooming, and channel shifting as a data preprocessing.

6.4.2 Baseline Methods

To evaluate the performance of our model, we consider two categories of baseline methods: (i) *Feature* and (ii) *Model*.

6.4.2.1 Feature baselines

In this category, we have four user profiling baseline methods that exploit different sets of input features from our model, to understand the importance of proposed text and image embedding to represent an influencer. (i) *Tag* method [75] uses tag information to classify social media users. We extract hashtags, usertags, and URLs in captions and find high ranked tags to implement this baseline. Here, we set the number of top tags as 50 for each influencer category. (ii) *Text&Social* method [129, 130] has four feature sets: social media profile, posting behavior, linguistic content, and social network features. To generate each feature, we (i) extract category-specific keywords in users' biographies for the social media profile feature, (ii) analyze posting interval and the number of total posts of users for the posting behavior feature, (iii) find prototypical words from users by estimating conditional probability of a word in a given for the linguistic content feature, and (iv) find the followers and following counts for the social network feature. (iii) *User* method [51] uses textual characteristics of captions, image features from profile pictures, and neighbor information based on like relationship. We use the first two features as the following user information is not available in our dataset. (iv) *Image* method [178] classifies users' interest by using CNN based image classifier. We exclude the text input layer and use the only image input layer to implement this baseline method.

6.4.2.2 Model baselines

In this category, baseline methods take the same input features as our proposed model but use different machine learning approaches. Therefore, we can evaluate the novelty of our proposed model. We use the following four well-known learning methods: (i) Gaussian Naive

Bayes, (ii) K-Nearest Neighbors, (iii) SVC, and (iv) Random Forest. Note that we aggregate all post features by averaging values since the input layer has multiple post features.

6.4.3 Experimental Results

6.4.3.1 Influencer Classification

We first evaluate the performance of the influencer classification task. Table 6.1 shows overall classification accuracy and F1 score of each influencer class. In the Feature baseline methods, *Tag* method shows 71.1% accuracy and the lowest F1 scores in most influencer categories since tags from posts are insufficient to represent the category of influencers. *Text&Social* and *User* methods improve classification performance by taking features from captions, biographies, and profile images. However, these baseline methods have poor F1 scores in specific influencer categories (e.g., fitness and pet). On the other hand, *Image* method has 90.02% accuracy and higher F1 scores than other baseline methods in the same category. This suggests that image information is more informative than other social media information for representing influencers' interests. The baseline methods in the Model category show better classification performance than Feature baseline methods. Note that all four baselines have around 94% accuracy. This implies that our proposed input features play a significant role in classifying influencer categories. Finally, Our proposed model, *Influencer Profiler*, outperforms all the baselines by achieving 98.32% accuracy. This is because our model uses the distribution of attention scores to focus on more important posts that can represent influencers. The model baselines, instead, use average values of posts which can lead a classifier to make a wrong decision by considering all posts equally.

The number of posts. We next examine how many posts are sufficient to accurately classify influencer categories. The computation cost will be remarkably reduced if we can classify influencers correctly with only a small number of posts. Figure 6.5 shows the influencer classification accuracy of the proposed model and baseline methods with a different number of input posts. The result reveals that the proposed model performs well with a

Table 6.1: Model performance: influencer classification results with our proposed model and the eight baseline methods. The proposed model outperforms all the baseline methods across all influencer categories.

Model Type	Model	Accuracy	F-1 Scores on Different Influencer Categories							
			Beauty	Family	Fashion	Fitness	Food	Interior	Pet	Travel
Feature Baselines	*Tag*[75]	71.10%	0.606	0.656	0.735	0.739	0.877	0.540	0.780	0.716
	Text&Social[129, 130]	81.35%	0.736	0.747	0.855	0.745	0.889	0.722	0.830	0.816
	User[51]	84.06%	0.905	0.836	0.871	0.663	0.949	0.854	0.690	0.849
	Image[178]	90.02%	0.835	0.915	0.911	0.772	0.934	0.898	0.870	0.952
Model Baselines	*GaussianNB*	93.26%	0.992	0.986	0.960	0.774	0.913	0.944	0.725	0.945
	KNeighbors	93.87%	0.905	0.902	0.961	0.857	0.968	0.946	0.936	0.942
	SVC	94.13%	0.855	0.882	0.970	0.878	0.988	0.989	0.940	0.934
	RandomForest	94.65%	0.956	0.983	0.967	0.807	0.949	0.989	0.845	0.934
Proposed Model	***Influencer Profiler***	98.32%	0.994	0.988	0.989	0.898	0.991	0.986	0.990	0.977

small number of posts while performances of the baseline methods significantly drop. Note that the accuracy of the Influencer Profiler with 20 input posts is 92.9% and that of the baseline methods are ranged from 65% (Random Forest) to 80% (SVC). The robustness of the model comes from the attention method which helps find more relevant posts to represent influencers even with a small number of posts. Therefore, our model can greatly benefit by reducing the time to collect and process a large amount of data.

Components importance. To understand the benefits of each component in the proposed model, we compare the performance of the full model with a model with no attention mechanism and a model with no fine-tuning. Note that we use mean values to aggregate post features in the model with no attention. We then use the following three input sets, only text, only image, and both text and image, to analyze the effect of input modalities on the models with different components. We first observe a large performance difference depending on the input sets as shown in Figure 6.6. The accuracy of the full model with only text, only image, and both text and image, are 89.04%, 95.47%, and 98.32%, respectively. This result suggests that image features are more useful than text features in identifying influ-

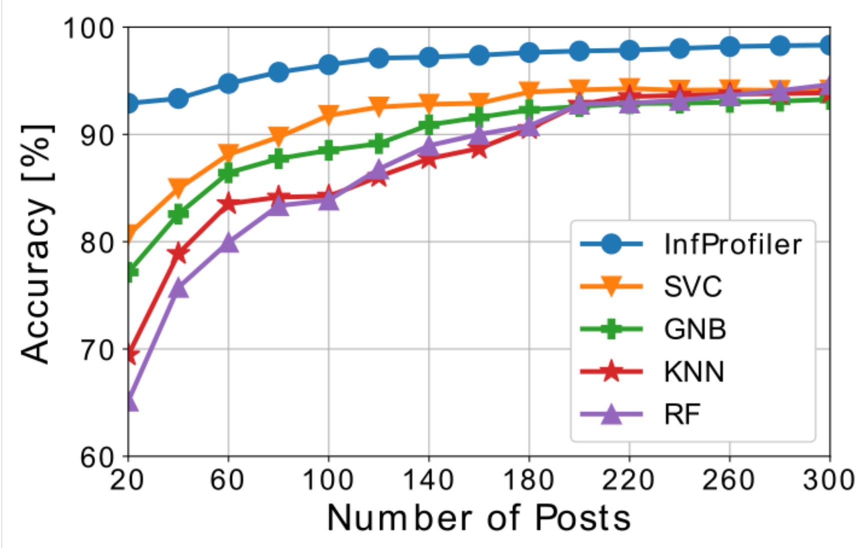


Figure 6.5: Influencer classification accuracy scores on different number of posts. Influencer Profiler shows more robust performance than baseline methods.

encer categories. This is probably because the style of writing captions for each influencer differs more than the typical style of images in certain categories. We further improve the performance by taking both text and image which implies that contextualized information helps provide unique features with image information. Figure 6.6 also shows the performance gain of using attention and fine-tuning. The accuracy of the model with no fine-tuning with multi-modality is 97.41% which demonstrates that the fine-tuning the hidden layer gives a moderate gain. Attention mechanism, however, helps the model to achieve very high classification scores by weighting higher scores on more important posts which can represent the influencer category. Note that the model with no attention has 95.18% accuracy, which is a large loss from the full model (98.32% accuracy).

6.4.3.2 Post Classification

To evaluate the performance of the post classification, we conduct experiments on the 10,000 labeled posts which is divided into training and testing with an 8:2 ratio. Table 6.2 shows the post classification results when the model uses (i) the text features only, (i) the image

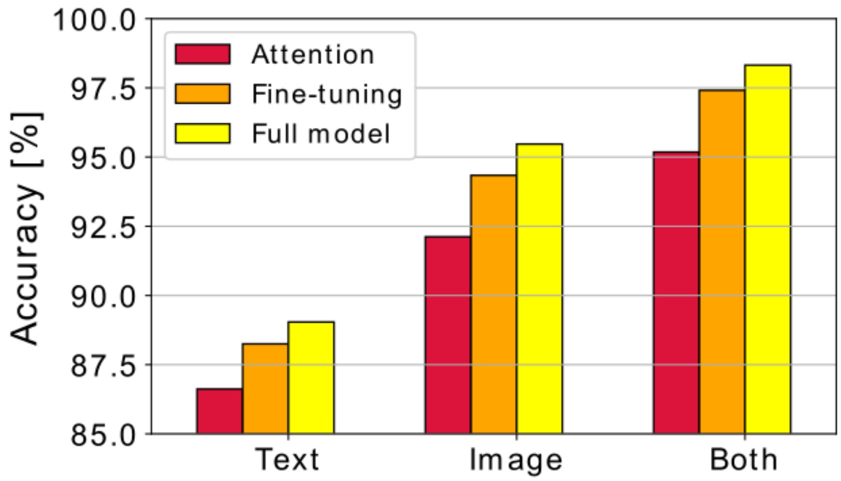


Figure 6.6: The effects of different modalities of input data, fine-tuning, and attention on influencer classification performance. Multi-modal inputs and attention mechanisms help the model gain high classification performance.

Table 6.2: Post classification results of the proposed model and baseline methods across the ten post categories.

Model	Input	Accuracy	F-1 Scores on Different Post Categories									
			Beauty	Family	Fashion	Fitness	Food	Interior	Other	Pet	Product	Travel
GaussianNB	Text	40.70%	0.495	0.387	0.327	0.497	0.491	0.551	0.308	0.315	0.441	0.297
	Image	84.65%	0.956	0.726	0.850	0.847	0.937	0.938	0.819	0.709	0.804	0.837
	Text & Image	87.00%	0.904	0.731	0.850	0.887	0.948	0.930	0.844	0.871	0.858	0.874
KNeighbors	Text	38.85%	0.462	0.255	0.335	0.414	0.477	0.521	0.176	0.374	0.491	0.364
	Image	81.75%	0.946	0.690	0.826	0.812	0.931	0.935	0.793	0.623	0.773	0.794
	Text & Image	88.70%	0.923	0.706	0.839	0.886	0.954	0.939	0.840	0.922	0.891	0.927
SVC	Text	36.20%	0.507	0.325	0.308	0.392	0.296	0.548	0.273	0.161	0.458	0.281
	Image	67.00%	0.857	0.514	0.795	0.662	0.842	0.884	0.825	0.429	0.382	0.720
	Text & Image	71.50%	0.870	0.568	0.819	0.687	0.835	0.884	0.845	0.479	0.509	0.814
RandomForest	Text	31.80%	0.452	0.248	0.268	0.311	0.376	0.436	0.212	0.259	0.367	0.247
	Image	78.30%	0.866	0.624	0.770	0.772	0.859	0.902	0.736	0.757	0.714	0.811
	Text & Image	76.25%	0.909	0.624	0.738	0.791	0.830	0.871	0.701	0.770	0.624	0.724
Influencer Profiler	Text	60.90%	0.696	0.522	0.515	0.714	0.703	0.718	0.487	0.556	0.556	0.635
	Image	90.75%	0.970	0.752	0.874	0.928	0.955	0.958	0.884	0.897	0.905	0.931
	Text & Image	96.20%	0.982	0.891	0.932	0.975	0.990	0.985	0.955	0.968	0.955	0.979

features only, and (i) both the text and image features. We first find that our model significantly outperforms other baseline methods. Our model achieves 96.2% accuracy while K-Nearest Neighbors shows the highest accuracy, which is 88.7%, among the baseline methods. This demonstrates that our model can effectively capture unique characteristics for each post category. We also observe that the image features help the model to achieve higher classification performance than the text features. The accuracy of using only the text features and the image features of our proposed model are 60.9% and 90.75%, respectively. We observe that the post classification with the image features outperforms since influencers often write very short or random captions that are not related to the topic of the posts. The results also reveal that we can accurately classify posts with both image and text features together. This implies that contextualized information in captions assists the post classifier to identify the class of corresponding posts accurately.

6.4.4 Influencer Analysis

6.4.4.1 Influencer Categorization

We now apply our model to our entire influencer dataset. That is, we classify 33,935 influencers into the eight influencer categories. Table 6.3 shows the number of influencers in each category and corresponding percentages. We find that fashion is the most dominant influencer category in our dataset which accounts for 44.16% of total influencers. This is nearly three times the number of travel influencers which is the second dominant influencer category. Travel, family, and food categories are also considered as major types of influencers while fitness, interior, and pet categories have less than 5%. Note that the percentages of the number of influencers in Table 6.3 are applied to our Instagram dataset, and other social media platforms may have different percentages for the influencer categories.

Table 6.3: The number of classified influencers and percentages for each influencer category. Fashion is the most dominant influencer category in our dataset.

Influencer Category	Influencer counts	Percentage
Fashion	14,986	44.16%
Travel	5,113	15.07%
Family	4,928	14.52%
Food	3,973	11.71%
Beauty	2,035	6.00%
Fitness	1,210	3.56%
Interior	1,163	3.43%
Pet	527	1.55%
Total	33,935	100.00%

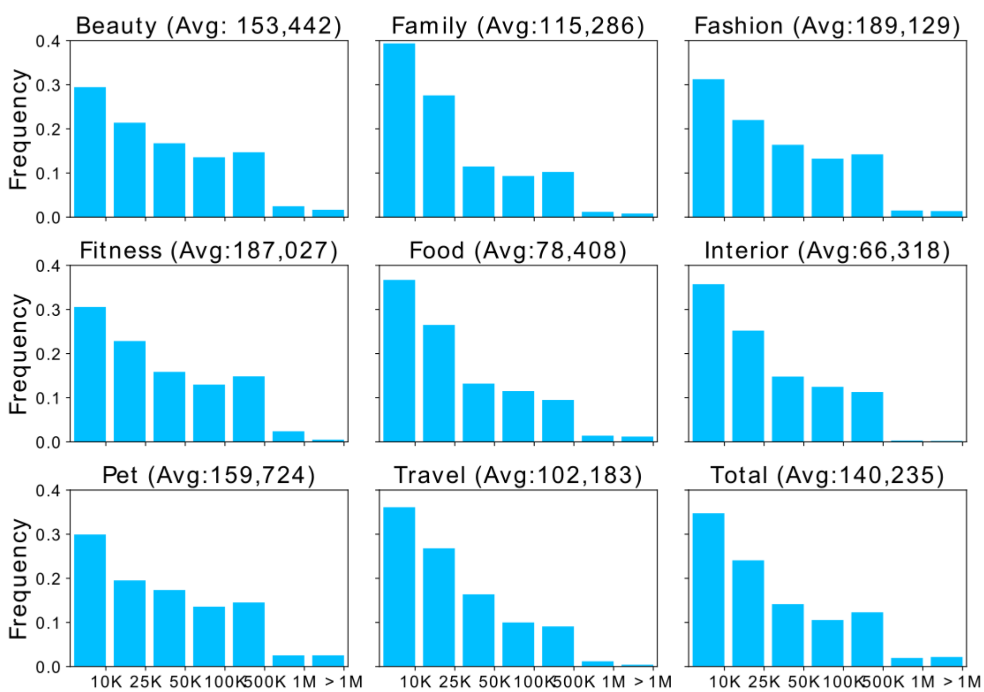


Figure 6.7: Distributions of portions of the number of followers in each category.

6.4.4.2 Size of Potential Customers

Given the influencer categories and influencers' follower counts, we examine the distributions of portions of the number of followers in each influencer category as shown in Figure 6.7. We categorize the influencers into seven groups by the number of followers, which are less than 10 K, 25 K, 50 K, 100 K, 500 K, 1 M, and greater or equal to 1 M followers, to understand the portion of 'micro' and 'macro' influencers. In general, influencers who have less than 10 K and 25 K followers are the most dominant in all categories in terms of the number of influencers, which we already observed in Figure 6.3. However, the portion of each group varies by category, for example, there are more micro-influencers in family, food, and travel categories than other categories, while beauty, fashion, and fitness categories tend to have more famous influencers who have more than 500 K followers. We also find that the average numbers of followers of food and interior influencers are relatively smaller than those of influencers in other categories since there is small number of famous macro-influencers (e.g., $> 1\ M$ followers) who are interested in food or interior. We believe this result can give insight to brands who are participating in influencer marketing or are planning to hire influencers.

6.4.4.3 Posting Behavior

Understanding how influencers (with different interests) upload posts and how their followers engage in the posts is important to brands who want to hire influencers for their marketing campaigns. Given the classified influencers and their posts, we examine the posting behavior of influencers with different interests. Table 6.4 shows the portions of post categories[4] that influencers with different interests have uploaded. As shown in Table 6.4, the influencers mainly upload posts related to their interests, but rarely upload posts related to other post categories. We find that pet, food, and interior influencers have relatively higher posting rates, which are 0.79, 0.68, and 0.6, respectively, for their own major post categories than other influencers (e.g., fitness and family influencers only have 0.39 and 0.42 posting rates

[4]'be','fm','fa','fi','fo','in','pr','pe','ot', and 'tr' indicate 10 post categories, *categories*, *beauty*, *family*, *fashion*, *fitness*, *food*, *interior*, *product*, *pet*, *other*, and *travel*, respectively.

Table 6.4: Posting behavior of influencers.

Post category	Influencer category							
	Beauty	Family	Fashion	Fitness	Food	Interior	Pet	Travel
Be	**0.62**	0.02	0.05	0.04	0.02	0.01	0.01	0.02
Fm	0.03	**0.55**	0.04	0.05	0.03	0.06	0.03	0.04
Fa	0.08	0.11	**0.58**	0.15	0.03	0.03	0.02	0.09
Fi	0.01	0.03	0.03	**0.52**	0.02	0.01	0.01	0.02
Fo	0.02	0.06	0.02	0.02	**0.71**	0.05	0.02	0.05
In	0.01	0.05	0.02	0.01	0.02	**0.72**	0.01	0.03
Pr	0.14	0.05	0.05	0.08	0.06	0.04	0.04	0.06
Pe	0.01	0.01	0.01	0.01	0.01	0.01	**0.83**	0.01
Ot	0.06	0.04	0.05	0.07	0.03	0.01	0.01	0.02
Tr	0.02	0.08	0.15	0.05	0.07	0.06	0.02	**0.66**

in beauty and family post categories, respectively). We also observe that beauty influencers have the highest posting rates among all the influencer categories since they tend to actively upload cosmetics and skincare product images to share product information with their audience.

We next investigate how Instagram users engage in posts uploaded by influencers with different interests. Figure 6.8 shows the distributions of the engagement rate, which is calculated by dividing the number of likes of each post by the number of followers of the corresponding influencer, for different post categories. We find two observations from this analysis. First, the engagement rates vary depending on the post category. In general, *beauty*, *family*, *fashion*, *fitness*, and *pet* posts attract more attention than *food*, *interior*, *product*, and *travel* post categories. This reveals that photos with a human face(s) receive more likes than photos with the only product(s) or landscape, which is consistent with the result of the previous study [9]. Second, *food* and *interior* influencers show relatively low engagement rates than influencers in other categories. Note that the average engagement rate of all the influencers is 0.035 while those of both *food* and *interior* influencers are 0.024 and 0.023, respectively. Also, their engagement rates are mostly similar across the post categories. In

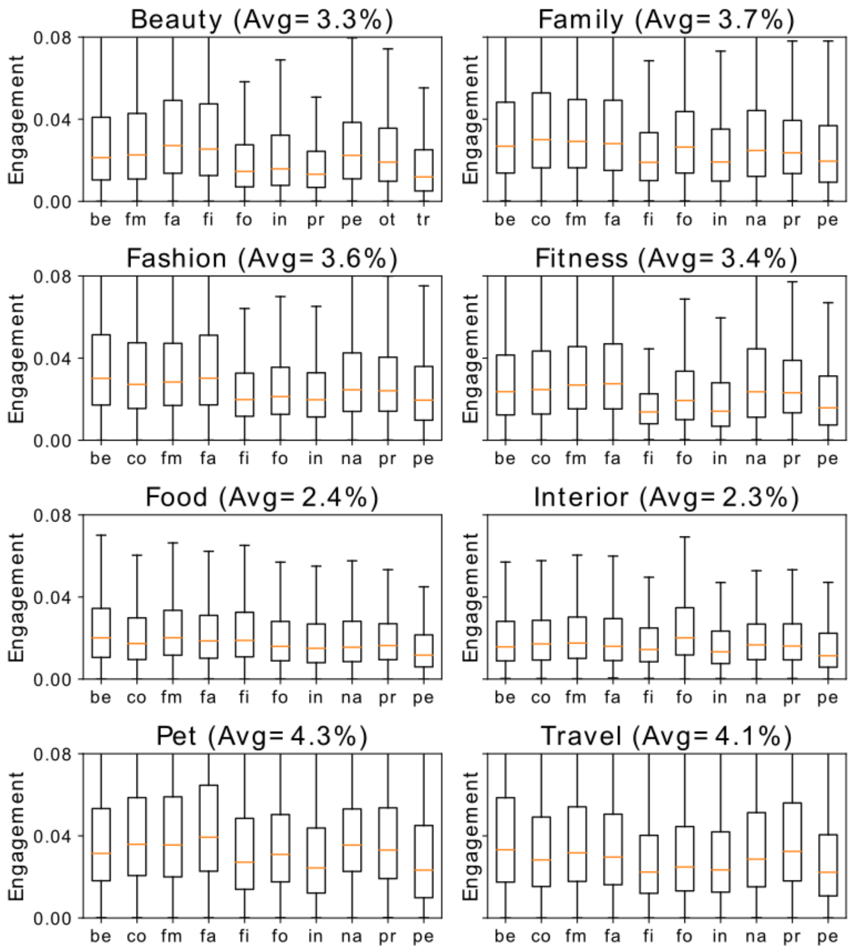


Figure 6.8: Distributions of engagement rates by post category. Photos with human face(s) receive more likes than faceless photos such as food, interior, and products.

other words, audiences who follow influencers who primarily post faceless photos, such as food and interior photos, tend to show limited and constant interest to the content uploaded by such influencers. These findings can give insight to brands and influencers, who are involved in influencer marketing, on how to select influencers in influencer marketing and/or how to create advertising content.

CHAPTER 7

Discovering Undisclosed Paid Partnership on Social Media via Aspect-Attentive Sponsored Post Learning

7.1 Background

Influencer marketing has been gaining significant attention from marketers as an essential advertising method recently [109]. As the rapid growth of the influencer marketing industry results in numerous paid advertisements in social media, the transparency issue of advertising posts has been raised. According to the regulations from the Federal Trade Commission (FTC) [40], the Advertising Standards Authority (ASA) [7], and the Organisation for Economic Co-operation and Development (OECD) [55], influencers are required to conspicuously disclose sponsorship when they publish paid advertisements. That is, mentioning brand names and the relationship between a mentioned brand and an influencer in paid advertisements, thereby having transparency in advertising posts. However, a noticeable number of influencers fail to disclose paid partnerships with brands in their advertising posts, either because they are not aware of the regulations [2] or because they are concerned about lowering the effectiveness of the advertisement [50]. Surprisingly, the recent survey [2] reveals that only 52% of influencers and 60% of marketers have a good understanding of the regulation. This implies that the lack of legal knowledge and education for social media users can lead to social issues amid the rapid growth of social media. Figure 7.1 shows an example of a paid media where the influencer advertises the product of the brand in the absence of mentioning sponsorship.

The sponsored posts without disclosing the sponsorship may cause the following problem.

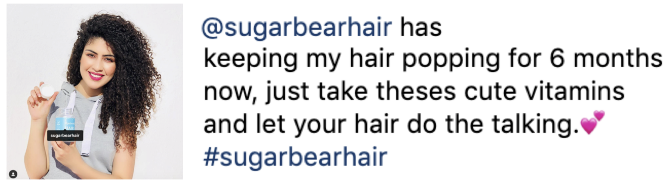


Figure 7.1: An example of paid media that fails to disclose sponsorship. Despite the influencer advertises a product and mentions a certain brand name, no sponsorship is disclosed.

Audiences will be increasingly skeptical toward the influencers' posts, and hence influencers lose the trust. Influencer marketing can only become effective when people think that influencers are trusted sources of information [105]. Furthermore, the lack of transparency in the advertising posts can negatively impact brand image [50]. For the steady growth of the influencer marketing industry with proper advertising practice, the FTC has monitored and warned a few famous celebrities in social media who violated the endorsement regulations[1]. However, it is impractical to monitor the millions of influencers on social media.

Although influencer marketing has gained noticeable attention recently, only a limited number of studies focused on sponsorship disclosure in influencer marketing. Some previous works examine the effect of the presence of sponsorship in social media [50, 155, 176] and suggest that disclosing sponsorship helps audiences to identify the post as an advertisement but lowers purchase intention. Moreover, some researchers [171] attempt to measure the sponsorship transparency of paid advertisement by considering audiences' perceptions. While the previous works discuss the importance of detecting undisclosed sponsorship of social media posts, no study has proposed a method yet. Additionally, the previous works solely rely on a small number of survey results, thereby lacking evaluation with a large dataset.

In this paper, we propose a learning-to-rank based model, *Sponsored Post Detector (SPoD)*, that can detect sponsorship of social media posts. Our model incorporates three different aspects (i.e., modalities) on social media including graph, text, and image to represent the social media posts. We first employ the Graph Convolutional Networks (GCNs) [94] to lever-

[1] `https://www.ftc.gov/news-events/press-releases/2017/04/ftc-staff-reminds-influencers-brands-clearly-disclose`

age the characteristics of posts and the social relationship among influencers and brands. To adopt GCNs, we construct a heterogeneous network that connects influencers, posts, and brands. Besides the graph features, we also generate image and text features of each post to further describe the characteristics of the posts. Particularly, we use the pre-trained Inception-V3 model to obtain the image object features which have 1,000 categories [158] and utilize BERT [46] to create contextualized features of social media post captions. Moreover, we apply attention [164] over the three sets of features to estimate the importance of each aspect of social media posts, thereby utilizing more important aspects to detect hidden sponsorship. In addition to the attentive post features, we conduct a manifold regularization method to optimize the model performance. More specifically, we propose to exploit posting time and mentioned brands from social media posts for temporal regularization. For example, we place more weight on posts created at similar times and mentioning the same brand, that is, posts that likely belong to the same marketing campaign. With the proposed temporal regularization, our model takes the attentive post features as input to rank given social media posts by their sponsorship scores.

7.2 Problem Statement

In this section, we formally define the goal of this paper. Suppose we have a set of posts $P = \{p_n\}_{n=1}^{|P|}$ published by a set of users $U = \{u_m\}_{m=1}^{|U|}$. Each post can be represented as $p_n = (t_n, a_n, b_n, l_n)$ where t_n, a_n, b_n, and l_n denote text, image(s), mentioned brand(s), and posting time of the post, respectively. Note that a post mentions at least one brand where the brand is in the set of brands $B = \{b_k\}_{k=1}^{|B|}$. Given text and images of a post, we extract text features and image features, $\boldsymbol{X^T} \in \mathbb{R}^{n \times f}$ and $\boldsymbol{X^I} \in \mathbb{R}^{m \times g}$, respectively, where f and g are the numbers of features. Moreover, we have graph features, $\boldsymbol{X^G} \in \mathbb{R}^{n \times h}$, where h is the number of features. The graph features can be generated from a heterogeneous network $\mathcal{G} = (E, V)$ where the vertices $V = (U, P, B)$ are composed of users, posts, and brands, respectively. Given a set of posts P, we aim to rank the posts by learning the distinctive features of sponsored posts so that sponsored posts with the absence of sponsorship disclosure can be discovered.

7.3 Methodology

In this section, we present the proposed framework to detect paid partnerships by discovering sponsored social media posts.

7.3.1 Framework Overview

Here we briefly give an overview of our proposed framework, SPoD, as shown in Figure 7.2. To leverage multimodal inputs of social media posts, we utilize three encoders, including graph encoder, text encoder, and image encoder. The graph encoder takes the heterogeneous network that includes users, posts, and brands as an input. Moreover, each node in the graph has a set of features as contextual representations to indicate the entity characteristics. Based on the heterogeneous structures of different entities and their features, graph convolutional networks (GCNs) are applied to derive appropriate node representations. In addition to GCN-encoded features, the text and image of each post are encoded by a contextualized text encoder and an image encoder, respectively. We then apply attention [164] over the three sets of features to estimate their importance and generate the post representations. The sponsorship scores of candidate posts are then computed based on all of the corresponding features and ranked for discovering sponsored posts. Finally, we optimize the sponsorship scores by conducting temporal regularization based on posting time and mentioned brands from the input post set.

7.3.2 Aspect-Attentive Heterogeneous Post Encoder

To acquire decent representations of posts, we utilize three encoders including the graph, text, and image encoders to capture knowledge from different aspects, and apply aspect-attention over the three sets of features as shown in Figure 7.2.

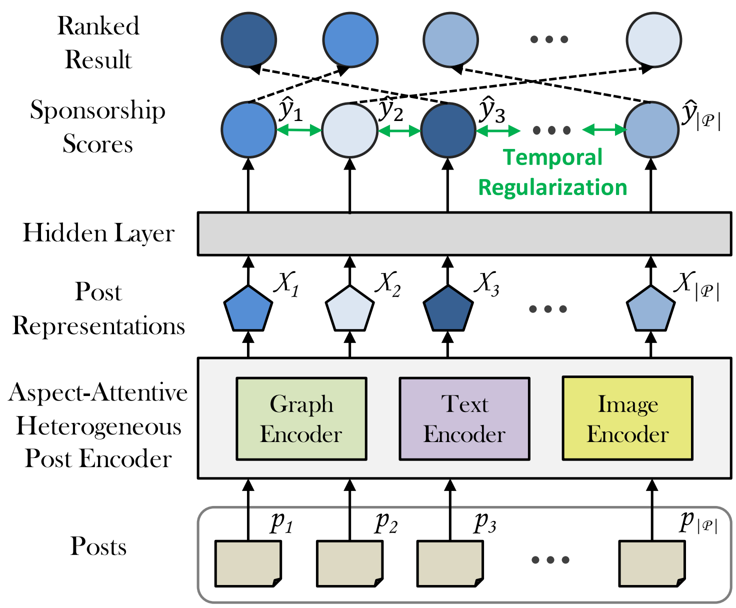


Figure 7.2: The overall framework of the SPoD. The aspect-attentive heterogeneous post encoder generates post representations. The estimated sponsorship scores are optimized by conducting temporal regularization for detecting sponsored posts.

7.3.2.1 Graph Encoder

To model posts with the graphical structure, we first construct a heterogeneous network and then apply graph convolutional networks (GCNs) [94] to derive GCN-encoded features for each candidate post.

Heterogeneous Network Construction. To construct a heterogeneous network, we consider three different entities including posts, users (i.e., influencers), and brands mentioned in posts. Note that the constructed heterogeneous network in our framework can be flexibly expanded with any additional relevant entities. The edges in the heterogeneous network indicate the interactions between entities behind nodes. The node of each post is linked to the node of the author user. If a brand is mentioned in a post, the post node has an edge to the brand node. Note that since more than one brand can be mentioned in a post, a post node can have multiple edges to brand nodes. More specifically, the edges of the network are represented by a sparse matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, where $A_{ij} = 1$ if the i-th and j-th nodes are

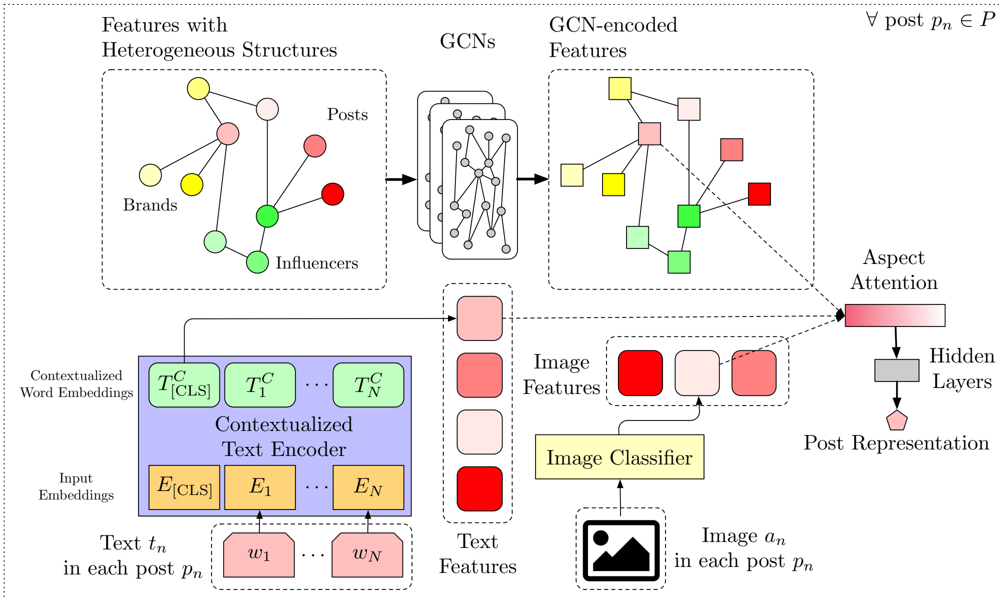


Figure 7.3: The illustration of the post encoder using heterogeneous information.

connected; otherwise, $A_{ij} = 0$.

Node Features. Each node in the network has a set of features while all of the features can be represented as

$$\boldsymbol{Z} = [\boldsymbol{Z^P}; \boldsymbol{Z^U}; \boldsymbol{Z^B}] \in \mathbb{R}^{N \times d},$$

where $\boldsymbol{Z^P}$, $\boldsymbol{Z^U}$, and $\boldsymbol{Z^B}$ are the features of nodes for posts, influencers, and mentioned brands, respectively; N and d are the number of all nodes in the network and the number of features for each node, respectively. The detailed features in this paper are defined in Section 7.3.2.5.

Graph Convolutional Networks To leverage the knowledge of structural information, we propose to apply GCNs to encode node representations with both node features and network structures. First, the adjacency matrix $\boldsymbol{A}$ is transformed into a normalized adjacency matrix $\hat{\boldsymbol{A}}$ as follows:

$$\hat{\boldsymbol{A}} = \boldsymbol{D}^{-\frac{1}{2}} A \boldsymbol{D}^{-\frac{1}{2}},$$

where $\boldsymbol{D}$ is the diagonal matrix of node degrees. GCNs can then be operated based on the normalized adjacency matrix $\hat{\boldsymbol{A}}$ and the feature matrix $\boldsymbol{Z}$.

To model complicated network structures, we consider multi-layer GCNs by propagating information through different layers. Formally, the outputs of the i-th layer in GCNs, $H^{(i)} \in \mathbb{R}^{N \times k}$, can be computed as follows:

$$\boldsymbol{H^{(i)}} = \sigma\left(\boldsymbol{\hat{A}H^{(i-1)}W^{(i-1)}}\right),$$

where k is the number of hidden dimensions in GCNs; $\boldsymbol{H^{(i-1)}}$ is the outputs of the previous layer; $\boldsymbol{W^{(i-1)}}$ is a matrix of layer-specific trainable weights; $\sigma(\cdot)$ is a nonlinear activation function. Note that $\boldsymbol{H^{(0)}} = \boldsymbol{Z}$ as the base case. Finally, the GCN-encoded representations $\boldsymbol{X}$ can be computed by concatenating the outputs of different layers as follows: $\boldsymbol{X^G} = \left[\boldsymbol{H^{(1)}}, \boldsymbol{H^{(2)}}, \ldots, \boldsymbol{H^{(L)}}\right]$, where L is the number of layers in GCNs.

7.3.2.2 Text Encoder

To model contextualized knowledge from text in posts, we encode the text t_n in a given post p_n. Instead of learning from scratch, we apply the pre-trained state-of-the-art neural language model, BERT [46]. Note that any potential language model can be applied to the text encoder. Given the text in a post t_n, a length j sequence of words $c = [c_1, c_2, \cdots, c_j]$, BERT adds an initial token `[CLS]` to alleviate the positional bias in input embedding $c' = [\texttt{[CLS]}, c'_1, c'_2, \cdots, c'_{j'}]$. The input word sequences of n posts, $\mathbb{Q}^{(j'+1)\times n}$, then goes to the transformer $\mathbf{F}_T$ with o layers. We generate the output of the o-th transformer $\mathbf{F}_T(\mathbf{F}_T^{o-1}) = D^o = [d^o_{[CLS]}, d^o_1, d^o_2, \cdots, d^o_{j'}]$ as the representation of each word in the word sequences. Finally, we only adopt the $d^o_{[CLS]}$ to form the text representation $\boldsymbol{X^T}$.

7.3.2.3 Image Encoder

In addition to the graph and text, we use images attached in the posts since images are known as one of the most important factors in effectively advertising products in social media marketing [111]. For example, the influencer in Figure 7.1 holds the product in the image for advertising purposes. We apply the pre-trained Inception-V3 model trained with one-million images in 1,000 object categories [158] to avoid training images from scratch. Particularly, the image encoder takes input images from a post and generates a feature vector with g dimensions where each dimension represents the probability that the image contains the corresponding object. That is a list of s_n images from a post p_n, $a_n = [a_{n1}, a_{n2}, \cdots, a_{ns}]$. We use the maximum value for each dimension while aggregating the feature vectors as follows: $\mathbf{F}_I(a_n) = \text{max-pool}(\{a_{ni} \mid 1 \leq i \leq s_n\})$, where the function max-pool$(\cdot)$ remains the maximum value for each dimension over the feature vectors of s_n images. Finally, the image representation $\boldsymbol{X^I}$ contains the image object vectors from n posts.

7.3.2.4 Aspect-Attention

To estimate the importance of features from different aspects, including the heterogeneous graph, texts, and images, we apply the attention mechanism over the three sets of features. We first apply a fully-connected hidden layer to each set of features to have the same dimension of the features as:

$$\boldsymbol{V^G} = \mathcal{F}_G\left(\boldsymbol{X^G}\right), \boldsymbol{V^T} = \mathcal{F}_T\left(\boldsymbol{X^T}\right), \boldsymbol{V^I} = \mathcal{F}_I\left(\boldsymbol{X^I}\right),$$

where $\mathcal{F}_G(\cdot)$, $\mathcal{F}_T(\cdot)$, and $\mathcal{F}_I(\cdot)$ are fully-connected layers for the graph, text, and image features, respectively. The sets of features from different modalities can be represented as:

$$\boldsymbol{V} = [\boldsymbol{V^G}, \boldsymbol{V^T}, \boldsymbol{V^I}].$$

The importance α_i of feature $\boldsymbol{V_i}$ can be estimated as:

$$\alpha_i = \frac{\exp(\boldsymbol{r_i} \cdot \boldsymbol{r^c})}{\sum_j \exp(\boldsymbol{r_j} \cdot \boldsymbol{r^c})},$$

where $\boldsymbol{r_i} = \tanh\left(\mathcal{F}\left(\boldsymbol{V_i}\right)\right)$ is the hidden representation of the feature $\boldsymbol{V_i}$; $\mathcal{F}(\cdot)$ is a fully-connected layer; $tanh(\cdot)$ is the activation function; $\boldsymbol{r^c}$ is the context vector for importance estimation. The estimated score α_i is multiplied with corresponding features $\boldsymbol{V_i}$ to get weighted values, and the representation of the post can be derived by taking all the weighted values of the graph, text, and image features as follows:

$$\boldsymbol{X} = \sum_i \alpha_i \cdot \boldsymbol{V_i}.$$

Table 7.1: Node features that represent characteristics of each type of node in GCNs.

Category	Feature	Description
Node	Node Type	Node type in the heterogeneous network.
Influencer	Keywords	The normalized frequency of keywords.
	Followers	Number of followers.
	Followees	Number of followees.
	Posts	Number of published posts.
	Influncer Category	Major interest of the influencer.
Posts	Likes	Number of likes in a post.
	Comments	Number of comments in a post.
	Hashtags	Number of hashtags(#) in a post.
	Usertags	Number of usertags(@) in a post.
	Caption Length	Length of text in a post.
	Images	Number of images in a post.
	Posting Day	The day a post was published.
Brand	Followers	Number of followers.
	Followees	Number of followees.
	Posts	Number of published posts.
	Brand Category	Business type of the brand.

7.3.2.5 Node Features in GCNs

To represent characteristics of nodes in the network, we incorporate four types of node features, including node type, influencer, post, and brand as shown in Table 7.1. Note that our proposed framework is not limited to these features, therefore, any potential feature can be additionally applied to the model.

- **Node type features.** To indicate one of the three types of nodes, including post, influencer, and brand, we apply the one-hot coded node type feature in this category.

- **Influencer features.** We exploit the normalized frequency of keywords to capture textual patterns of influencers. We select the frequently used keywords based on their Chi-square values. Note that we use the top 100 keywords in this study. We also use the number of followers, followees, and published posts features that represent the reputation of influencers [90]. Moreover, we exploit the major interest of the influencers such as Food and Interior [92].

- **Post features.** To represent post characteristics, we exploit the features which are widely used in any social media. We first obtain the numbers of likes and comments in a post which can represent the popularity of the given post. Since it is well known that people tend to avoid advertising [35, 85], such post popularity features can help provide a distinguishable representation of sponsored posts. We also employ the numbers of hashtags and usertags that are particularly used for disclosing names of brands, products, or marketing campaigns in paid advertisements [176]. Additionally, we use the number of images in a post while most social media accept multiple images in a post, and the day a post was published (e.g., Sunday, Monday) since publishing time affects the popularity of advertising posts in social media [117].

- **Brand features.** To characterize the brand nodes, we have the business type of the brands[2]. Additionally, we use the number of followers, followees, and published posts to measure brand awareness [176].

7.3.3 Sponsorship Estimation and Ranking

To estimate the sponsorship score of a post, all of the GCN-encoded features, text features, and image features can be useful because many aspects of the post are considered.

For a post i, all of the features are concatenated as the ultimate representation $\boldsymbol{X_i}$. The predicted sponsorship score $\hat{y}_i$ of the post i can then be generated by a linear unit with a fully-connected hidden layer as follows:

$$\hat{y}_i = \mathcal{F}_p\left(\sigma\left(\mathcal{F}_h\left(\boldsymbol{X_i}\right)\right)\right),$$

where $\mathcal{F}_h(\cdot)$ and $\mathcal{F}_p(\cdot)$ are two fully-connected hidden layers; $\sigma(\cdot)$ is a nonlinear activation

[2] `https://business.instagram.com/`

function. Therefore, the candidate posts can be ranked by the predicted sponsorship scores.

7.3.4 List-wise Learning to Rank

Since our goal is to rank posts by their likelihood scores to be sponsored posts, it is intuitive to apply learning to rank approaches to deal with the problem. More specifically, in this paper, we modify the ListMLE [172], which is list-wise learning to rank approach that can benefit overall ranking performance.

Suppose that $\boldsymbol{X}$ is the set of features for posts to be ranked; Y is the output space of permutations of the posts; P_{XY} is an unknown but fixed joint probability distribution of X and Y. If a ranking function can be represented by $\hat{\boldsymbol{y}} : X \rightarrow Y$, the expected loss $R(\hat{\boldsymbol{y}})$ to be optimized can be derived as follows:

$$R(\hat{\boldsymbol{y}}) = \int_{X \times Y} L(\hat{\boldsymbol{y}}(\boldsymbol{X_i}), \boldsymbol{y}) \partial P(\boldsymbol{X_i}, \boldsymbol{y}),$$

where $\boldsymbol{y} \in \boldsymbol{Y}$ is a permutation; $\boldsymbol{X_i} \in \boldsymbol{X}$; $L(\hat{\boldsymbol{y}}(\boldsymbol{X_i}), \boldsymbol{y})$ is the 0-1 loss between the ranked result $\hat{\boldsymbol{y}}(\boldsymbol{X_i})$ and the position in the permutation $\boldsymbol{y}$ such that

$$L(\hat{\boldsymbol{y}}(\boldsymbol{X_i}), \boldsymbol{y}) = \begin{cases} 1 & , \text{ if } \hat{\boldsymbol{y}}(\boldsymbol{X_i}) \neq \boldsymbol{y} \\ 0 & , \text{ if } \hat{\boldsymbol{y}}(\boldsymbol{X_i}) = \boldsymbol{y} \end{cases}$$

To make the training process more efficient, candidate lists with n labeled posts are sampled from the whole training space in each iteration. Given independently and identically distributed samples in a candidate list $S = \{(X_i, \boldsymbol{y_i})\}_{i=1}^{n} \sim P_{XY}$, we minimize the empirical loss R_S as follows:

$$R_S(\hat{\boldsymbol{y}}) = \frac{1}{n} \sum_{i=1}^{n} L(\hat{\boldsymbol{y}}(\boldsymbol{X_i}), \boldsymbol{y_i}),$$

where $\boldsymbol{y_i}$ is the ground truth permutation.

7.3.5 Temporal Regularization

Timing is important for publishing posts in influencer marketing [117], so the redundancy between different posts with similar published times can be leveraged to improve the performance for sponsorship estimation. In addition, brands usually hire a number of influencers for a marketing campaign at a time. For example, the posts that are published at a similar time and mention the same brand name are more likely to be sponsored posts. In this paper, therefore, we conduct manifold regularization [81] by using the redundancy between the posts and their published times. Formally, the regularization loss $Q(\hat{\boldsymbol{y}})$ can be defined as:

$$Q(\hat{\boldsymbol{y}}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(|\hat{y}_i - \hat{y}_j| \times \frac{w_b(i,j)}{\max(|l_i - l_j|, 1)}),$$

$$w_b(i,j) = \begin{cases} 1 & , \text{ if } b_i = b_j \\ 10^{-1} & , \text{ if } b_i \neq b_j \end{cases},$$

where $\hat{y}_i$ and $\hat{y}_j$ are the estimated sponsorship scores of the posts p_i and p_j mentioning the brands b_i and b_j at time l_i and l_j; $w_b(i,j)$ indicates the brand-based regularization weight. Note that the posting time difference, $|l_i - l_j|$, is measured in days. Finally, the ultimate objective for discovering sponsorship L can be a combination of two loss functions as;

$$L(\hat{\boldsymbol{y}}) = R_S(\hat{\boldsymbol{y}}) + w_l \cdot Q(\hat{\boldsymbol{y}}),$$

where w_l is the weight for manifold regularization.

7.4 Experiments

7.4.1 Experimental Dataset

7.4.1.1 Dataset Construction

To evaluate our proposed model, our dataset samples influencer posts from Instagram, which is the most popular social media platform for influencer marketing [109]. Note that we implement the data collection method in [92] and comply with the Instagram policy[3]. To find posts that mention brand names, we first collect a set of brands on Instagram by searching branded content, i.e., sponsored posts. Note that Instagram provides the branded content tool for influencers to disclose sponsorship by showing a partnered brand name on

[3] https://help.instagram.com/325135857663734

the top of a post[4]. From the searched sponsored posts, we obtain 26,910 brand names. Next, we find brand mentioning posts that contain at least one brand name by searching user tags in the corresponding caption. To reduce noises in the dataset, we filter out a post if it is published by a user with less than 1,000 followers which is a generally required number of followers to be considered as an influencer[5]. Finally, we collect 1,601,074 brand mentioning posts that are published from 2013 to 2019 by 38,113 influencers. Note that the number of posts is exponentially grown over time; the average follower count of the influencers is 127,279.

7.4.1.2 Sponsorship Labeling

Since our goal is to find sponsored posts that do not disclose paid partnerships, we first classify the posts in the dataset into two classes, including "Sponsored" and "Unknown". We label the posts as 'Sponsored' if the posts explicitly disclose sponsored relationships by using certain keywords. More specifically, a given post is labeled as 'Sponsored' if the post either uses the branded content tool from Instagram or has one of the following hashtags, `#ad`, `#sponsored`, and `#paidAD`, which are widely used hashtags for sponsorship disclosure in influencer marketing [50, 176]. The remaining posts that are not identified as sponsored posts are labeled as 'Unknown'. That is, the 'Unknown' posts may contain non-sponsored and sponsored posts with no sponsorship disclosure. We label all posts in the dataset and finally have 221,710 'Sponsored' posts and 1,379,364 'Unknown' posts which account for 13.8% and 86.2%, respectively. After labeling the posts, we remove all of the sponsorship-related keywords and hashtags from the posts to prevent information leakage in the experiments. Therefore, our model can detect sponsored posts without relying on such keywords. To evaluate the performance of detecting sponsored posts from the 'Unknown' posts, we further manually investigate the unknown posts. The details of manual labeling procedure are

[4] `https://business.instagram.com/a/brandedcontentexpansion`

[5] `https://www.digitalmarketing.org/blog/how-many-followers-do-you-need-to-be-an-influencer`

described in Section 7.4.4.2.

7.4.1.3 Heterogeneous Network

We build the heterogeneous network by using the 1,601,074 posts which include 2,273,578 brand mentions. As a result, the network has 38,113 influencer nodes, 26,910 brand nodes, and 1,601,074 post nodes with 3,874,652 edges.

7.4.2 Experimental Settings

To measure the performance of the proposed SPoD, we treat the task as a one-class ranking problem and assign a relevance score for each post so that posts with higher scores are more likely to be sponsored posts. That is, the posts labeled as 'Sponsored' have relevance 1 while the 'Unknown' posts have the relevance 0. As SPoD ranks the candidate posts by their sponsorship scores, the relevances are used to evaluate the rank quality. More specifically, we use mean average precision (MAP), mean reciprocal rank (MRR), and average precision (AP) as our evaluation metrics.

We use TensorFlow [1] to implement our model. We split the dataset into three partitions for training, validation, and testing with a ratio of 7:1:2 by randomly selecting the posts. Therefore, the ratios of sponsored posts and unknown posts on three partitions are the same. Additionally, we ensure that the same influencers are not included across the training, validation, and testing sets to avoid information leakage from learning relationships between influencers and brands (e.g., a certain influencer repetitively advertises a certain brand). We tune the parameters with the validation set and set a single hidden layer with 128 hidden nodes. The learning rate and the dropout probability are set as 10^{-3} and 0.5, respectively. We set the regularization weight, w_l, as 10^{-4}.

7.4.3 Comparative Baseline Methods

We compare the performance of the proposed model with the baseline methods in three different categories, including *Ranking*, *Graph*, and *Text*.

Ranking Baselines. As our model applies the learning-to-rank approach, we apply the identical feature sets of the proposed model for the ranking baselines, therefore, we can evaluate the model capability for the ranking task of the proposed model. We deploy three ranking baseline methods as follows: *ListNet* (LN) [23] is a list-wise learning-to-rank algorithm that exploits gradient descent on neural networks to optimize a list wise loss function. *MART* [57] is a pair-wise learning-to-rank algorithm that uses gradient boosted decision trees for prediction tasks. *LambdaMART* (LM) [21] directly optimize rank cost functions by using gradient boosted regression trees based on *MART*.

Graph Embedding Baselines. The baseline methods in this category only exploit the graphical structure without other information. We deploy the LINE [159] and the GCN [94] as two graph baseline methods. The baselines use the heterogeneous network as input features and disregard the text and image features.

Text Modeling Baselines. In addition to the ranking and the graph baselines, we also have two text baseline methods since influencers usually reveal paid partnerships using text. As the baselines, we deploy two state-of-the-art language models, Embeddings from Language Models (ELMo) [133] and Bidirectional Encoder Representations from Transformers (BERT) [46].

7.4.4 Experimental Results

In this section, we evaluate the performance of our proposed SPoD compared to the baseline methods with the following two steps: (i) We first examine the sponsored post ranking performance without taking into account sponsored posts in the unknown posts. (ii) We then investigate highly ranked unknown posts to evaluate the performance of detecting hidden sponsored posts in the unknown posts.

Table 7.2: Performance comparison with the baseline methods. SPoD significantly outperforms all types of baseline methods. The temporal regularization and aspect-attentive components improve the ranking performance.

Method	MAP	MRR	AP@k			
			10	100	1000	10000
LN [23]	0.250	0.500	0.714	0.643	0.487	0.380
LM [21]	0.269	1.000	0.867	0.451	0.461	0.395
MART [57]	0.290	1.000	0.507	0.398	0.432	0.421
LINE [159]	0.317	1.000	0.894	0.701	0.587	0.473
GCN [94]	0.370	1.000	0.935	0.744	0.709	0.566
ELMo [133]	0.352	1.000	0.926	0.751	0.714	0.608
BERT [46]	0.376	1.000	0.947	0.788	0.755	0.653
SPoD w/o regularization	0.558	1.000	1.000	0.967	0.956	0.902
SPoD w/o aspect-attention	0.573	1.000	1.000	0.973	0.960	0.913
SPoD	**0.592**	**1.000**	**1.000**	**0.994**	**0.984**	**0.941**

7.4.4.1 Sponsored Posts Ranking Performance

Table 7.2 shows the performance of the proposed SPoD and the baseline methods. We find that ranking baselines show poor ranking performance compared to the graph and the text baselines. Despite the graph and the text baselines only exploit the network features and the contextualized text features, respectively, these baselines outperform the ranking baselines which use all the proposed features. This is because the text features can be easily over-fitted over the complicated structures in ranking baseline methods. In other words, the ranking baselines fail to leverage the contextualized features. We also find that, in the graph and the text baselines, GCN and BERT have better rank quality than LINE and ELMo, respectively, which are adopted to our proposed post encoder.

Finally, our proposed model, SPoD, significantly outperforms all of the other baseline methods. More specifically, SPoD obtains a 57.45% improvement in mean average precision over BERT. Unlike ranking baselines, SPoD separates text and image features from GCN

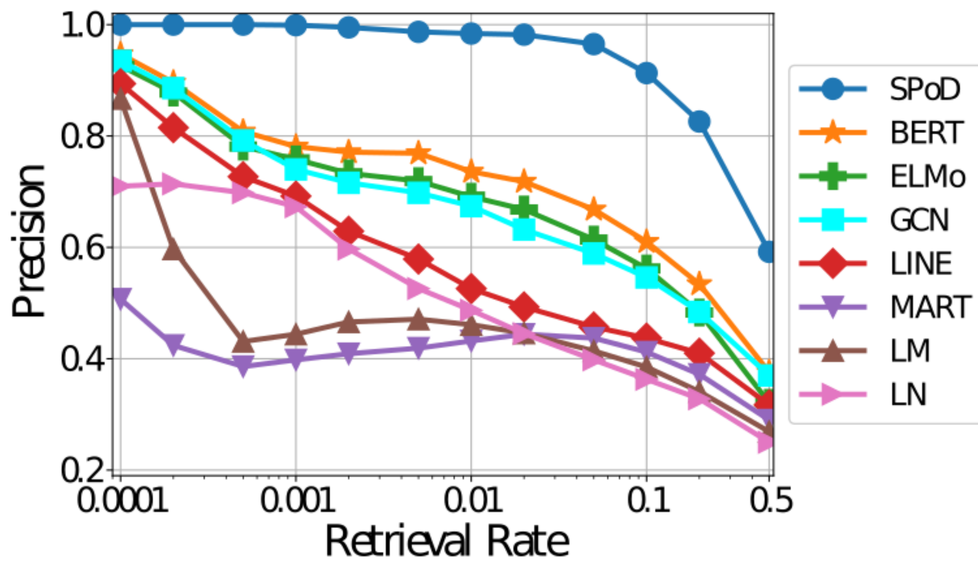


Figure 7.4: Precision at retrieval rates of the proposed model and the baseline methods. SPoD shows efficient and robust ranking performance compared to the baseline methods.

features and then applies attention over the features, thus effectively utilizing more important post features for ranking. Furthermore, SPoD shows a more robust ranking performance than other baselines. Note that average precision at 10,000 of SPoD is 0.941 while the performance other baseline methods tend to decrease if the ranked list size increases. Figure 7.4 shows precision at retrieval rates of SPoD and the other baseline methods. While the precision of the baseline methods significantly drops as the retrieval rates increase, SPoD shows highly robust performance. That is, SPoD has high ranking accuracy even while finding a large number of sponsored posts. Furthermore, we perform an ablation study by removing the temporal regularization and the aspect-attention. As shown in Table 7.2, SPoD loses a 5.74% and 3.21% rank performances in the measure of MAP without using the proposed regularization and the aspect-attention, respectively. This suggests that the proposed regularization that exploits the redundancy between the posts and their published times plays an important role in discovering sponsorship of social media posts. This also reveals that the aspect-attention effectively generates post representations by finding more important features.

Table 7.3: Precision results on the top-ranked unknown posts. SPoD outperforms other baseline methods in detecting undisclosed sponsorship of the unknown posts.

Precision@k	10	50	100	150	200
GCN [94]	0.600	0.540	0.420	0.380	0.355
BERT [46]	0.800	0.820	0.750	0.673	0.580
SPoD w/o regularization	**1.000**	0.960	0.910	0.873	0.810
SPoD w/o aspect-attention	**1.000**	**1.000**	0.950	0.920	0.860
SPoD	**1.000**	**1.000**	**0.990**	**0.967**	**0.920**

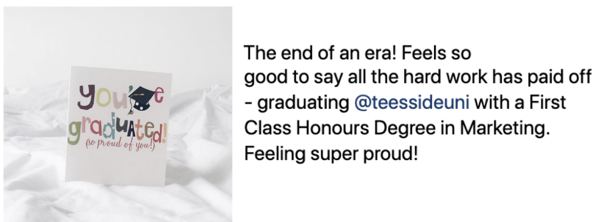
7.4.4.2 Detecting Unlabeled Sponsored Posts

Since the goal of SPoD is to detect sponsored posts which do not clearly disclose sponsored relationship with brands, we evaluate the performance of detecting such sponsored posts by investigating the ranking results from Section 7.4.4.1. To this end, we first extract a set of highly ranked unknown posts and then examine their images and captions to manually label the posts. To ensure the quality of our labeling procedure, two authors of this paper have carefully read and understood the FTC's endorsement regulation, then investigated the top 200 unknown posts from each ranking result of SPoD, SPoD without regularization, BERT, and GCN. More specifically, we decide an unknown post as a sponsored post when an influencer exclusively promotes a certain product or service by expressing appreciation for sponsorship indirectly in text and holding the product to show brands in images. Cohen's kappa coefficient of our labels is 0.784 which suggests that our labeling result is highly reliable [112]. Note that there are only 9 disagreements out of the 200 posts. We consider an unknown post as a sponsored post only if both labelers agree.

Table 7.3 shows the precision of detecting sponsored posts from the unknown posts. The result demonstrates that SPoD is very effective in discovering the sponsored posts with the absence of sponsorship disclosure compared to the baseline methods. Note that SPoD gains 58.6% and 159.2% improvements in precision scores at 200 over BERT and GCN, respectively by achieving 0.920 precision score. This implies that the proposed post representations are

(a) Sponsored post 1

(b) Sponsored post 2



(c) Non-sponsored post 1

(d) Non-sponsored post 2

Figure 7.5: Examples of successfully detected sponsored posts with absence of sponsorship disclosure, and highly ranked non-sponsored posts.

remarkably useful to identify sponsorship of social media posts even if the paid partnership is not explicitly disclosed. On the other hand, using only graph structural or contextualized information may fail to detect such posts. We also find that SPoD loses 11.96% and 6.52% performance in the precision at 200 without using the temporal regularization and the aspect-attention, respectively. This suggests that proposed temporal regularization significantly improves the performance of detecting undisclosed sponsorship by detecting advertising posts in the same marketing campaigns. This also reveals that the aspect-attention is useful to obtain more important knowledge from different aspects of social media posts.

Figure 7.5 showcases the example posts that are in the top 200 of the rank result of SPoD. The posts in Figure 7.5(a) and 7.5(b) are the sponsored posts with absence of sponsorship disclosure. In the sponsored posts, the influencers tend to describe details of the products by sharing their experience and recommending the products in the captions. In addition to the text, we observe the evidence of sponsorship in the image as they hold the products

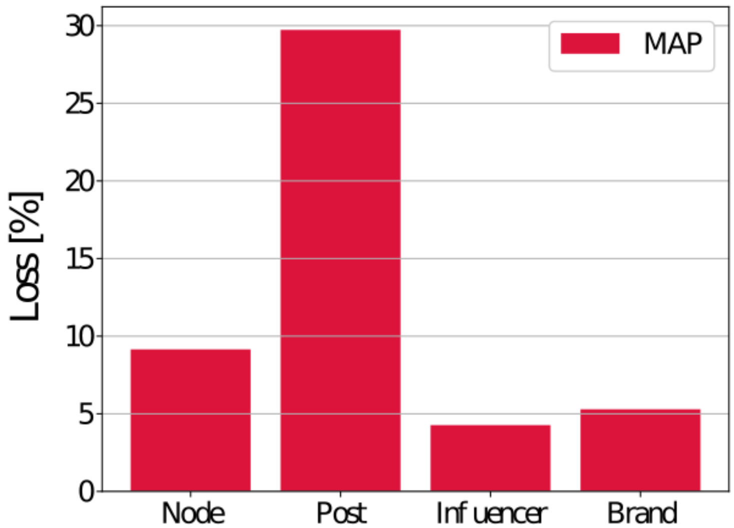


Figure 7.6: Analysis on the node features in GCNs.

for advertising. However, the sign of sponsorship disclosure is not found from the text and hashtags. On the other hand, Figure 7.5(c) and 7.5(d) show the non-sponsored posts that are in high-rank positions. These posts may have been highly ranked due to an object in the images (e.g., products), contextual similarity, and social relation between influencers and brands.

7.4.5 Analysis and Discussions

In this section, we first study the effectiveness of the proposed model with different feature sets to understand the impact of features to discover sponsored posts. We then evaluate the performance of SPoD with the sets of posts that have different caption lengths.

7.4.5.1 Feature Importance

To understand the importance of each feature set in the proposed SPoD for detecting sponsored posts, we evaluate the rank performance over different feature sets. Figure 7.6 shows the performance loss of MAP scores of the models trained with the features excluding one particular node feature against the full model as the leave-one-out analysis. We find that

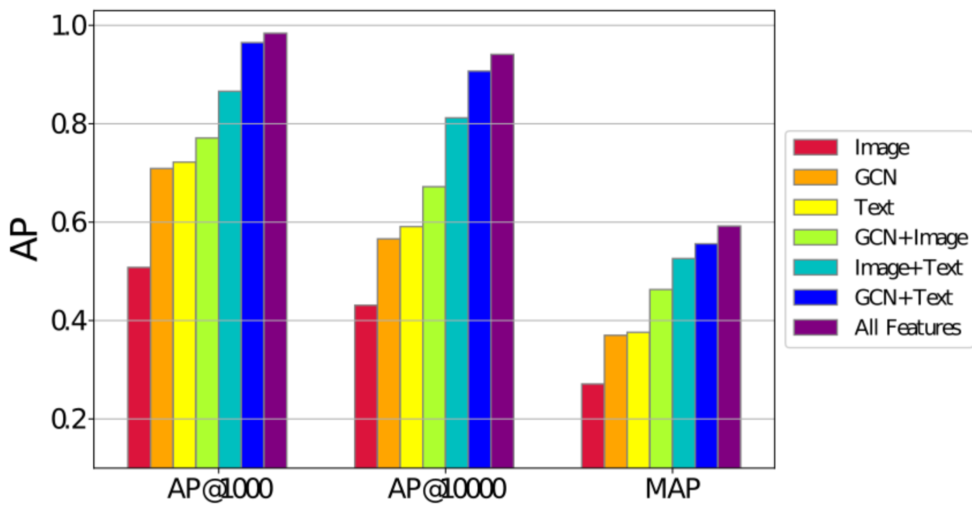


Figure 7.7: Analysis on the Image, Text, and GCN features.

the post features have a larger loss value than other node features since sponsored posts have distinct characteristics from non-sponsored posts [176]. Figure 7.6 also shows that the node type features have a large loss value. This suggests that social relations between influencers and brands that are indirectly learned from node types provide valuable information to detect sponsorship of posts.

Figure 7.7 shows the average precision scores of the proposed SPoD over (i) only image features, (ii) only GCN features, (iii) only text features, (iv) GCN and image features, (v) image and text features, (vi) GCN and text features, and (vii) All features. The result reveals that the text features significantly improve the performance while the image features contribute to the slight improvement. Note that SPoD loses 27.86%, 12.55%, and 6.47% performance in MAP when it excludes text, GCN, and image features, respectively. This suggests that the contextualized information from captions is very useful for discovering sponsored posts. Due to the nature of paid advertisements, influencers try to recommend the products, convey detailed information of products, and to make a good impression on the brand in the text [171]. This consequently makes the paid advertisements to have distinct contextualized features from non-sponsored posts. The image features, on the other hand, have fewer benefits in discovering sponsored posts compared to the other features. Unlike

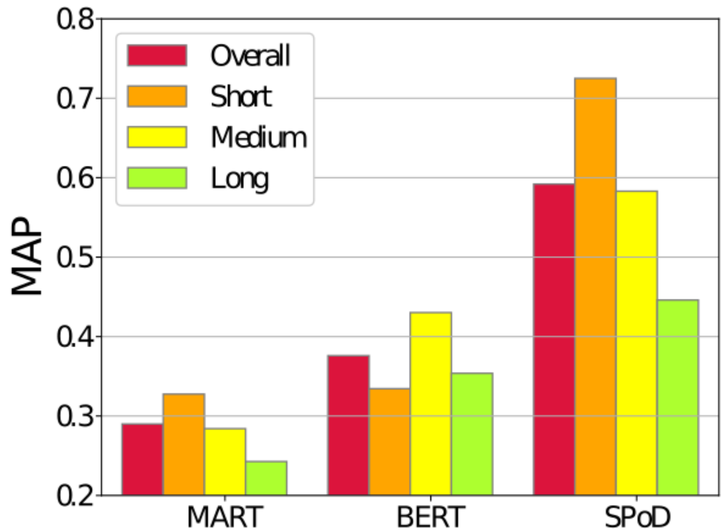


Figure 7.8: Analysis on the caption length of posts.

the text features, which naturally show similar characteristics due to the commonality of language, images can be generated in various ways by different users, thereby making it difficult to rank the posts.

7.4.5.2 Caption Length of Posts

Since we find that textual information plays an important role in determining whether given social media posts are sponsored, we investigate the performance of the proposed model over different sets of posts with various caption lengths. We split the test dataset into three subsets, including short, medium, and long posts, which have less than 250 (34.5%) and 500 (31.9%), and more than 500 (33.6%) characters in a caption, respectively. Note that we use the same training set for the post sets with different caption lengths thus our model can be used for posts with various caption lengths to detect sponsorship. Figure 7.8 shows the performance of SPoD and two baseline methods over the post sets in different caption lengths. By comparing the performance of MART and BERT, we observe that BERT obtains noticeable improvement in medium and long length posts. This reveals that the contextualized information greatly help detect paid partnership in the posts. However, because of the same reason, BERT fails to improve the performance on the short posts due to insufficient contextualized information from captions. As shown in Figure 7.8, the proposed SPoD outperforms the baseline methods over all the post sets in various caption lengths. Particularly, SPoD remarkably improves the performance in the short post set. This implies that SPoD effectively leverages the knowledge of both graph structures and the various features to detect the sponsorship of social media posts. For example, a sponsored post with a very short caption can be detected by our model by understanding the social relations and characteristics of adjacent influencers and brands.

CHAPTER 8

Evaluating Audience Loyalty and Authenticity in Influencer Marketing via Multi-task Multi-relational Learning

8.1 Background

Influencer marketing, one of the popular social media marketing strategies, utilizes influential social media users as marketing channels to reach a large number of target audiences [44]. Many brands have been increasingly sponsoring influencers in recent years to advertise their products or services [176, 93] since audiences tend to have more interactions and trust in influencers than brands [33, 105]. The cost of influencer marketing paid by brands is usually determined by quantitative metrics such as the follower count of the influencers or the average number of likes they receive [34, 41, 73]. However, the number of followers or likes can be easily manipulated by influencers through buying fake followers and engagements [5, 89]. Brands can be suffered from such influencer fraud behavior, e.g., wasting of marketing costs and losing trust from their audiences, which adversely affect the influencer marketing industry. To tackle this problem, the quality of the audience of influencers should be evaluated instead of just using simple metrics like the number of followers or the average number of likes, which can be easily manipulated.

To understand and evaluate the audience of social media users, researchers have studied different metrics to assess the quality of the audience, especially in the marketing domain [136, 173, 110]. The quality of the audience can be evaluated from two main perspectives, *loyalty* and *authenticity*, which represent the level of interest toward brands and the

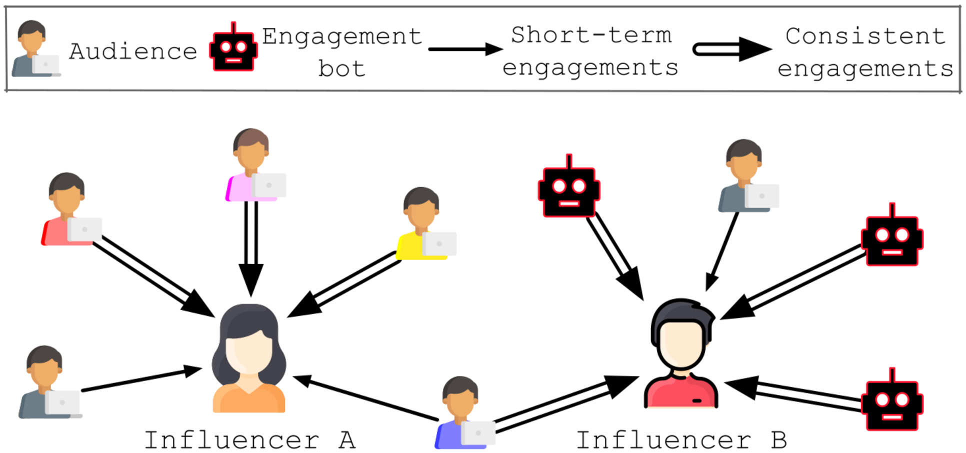


Figure 8.1: Illustration of engagement relationships for two influencers and their audiences. Influencer A has many loyal audiences who consistently make engagements, whereas Influencer B is connected to inauthentic audiences (engagement bots) who generate fake engagements.

genuineness of the audience's engagements, respectively [100]. Figure 8.1 compares two influencers with different quality of audiences. Influencer A can be considered as an influencer with high-loyalty audiences since audiences show consistent engagement behavior. On the other hand, although influencer B receives consistent engagements, a large portion of B's audiences are fake ones. This demonstrates that simply considering either one of loyalty or authenticity limits accurate audience quality evaluation. Therefore, both of them should be considered to evaluate the quality of audiences for given influencers.

Brand loyalty has been widely studied since the degree of loyalty is closely related to the success of marketing campaigns. Loyal customers tend to show positive engagements and solid trust in certain brands while other relatively disloyal customers are likely to make short-term engagements [16]. Some previous studies focus on correlations between brand loyalty and other factors such as online advertisements [12], frequency of interactions [120], and characteristics of published contents [48]. These previous studies mostly focus on understanding brand loyalty in social media marketing. However, little attention has been paid to use brand loyalty for evaluating audience quality, especially in influencer marketing.

Also, most previous works rely on surveys for evaluation, thus fail to propose learning-based models to automatically predict the loyalty of the audience in social media.

The authenticity, another audience evaluation metric, has been widely investigated for detecting bots in social media. Bots in social media have been causing various problems including popularity manipulation via fake followers and engagements [54, 43, 89, 5, 149, 15]. Many previous studies analyze distinct behaviors of bots such as link farming [149, 28, 15] and propose methods to detect fake engagements or fake followers [175, 95]. However, to our knowledge, no work has yet suggested evaluating influencers based on their audience authenticity.

To address the above limitations towards evaluating the quality of audiences of influencers, we propose a computational audience evaluation framework based on Audience Loyalty and Authenticity in Influencer Marketing ($ALAIM$), which can predict multiple audience quality scores together. To integrate both loyalty and authenticity into our model, we formulate our audience evaluation problem as a multi-task ranking problem. Note that the proposed framework is naturally optimized in an end-to-end manner, eliminating the need for conducting human evaluation such as surveys. The proposed framework first takes audience engagements (e.g., likes and comments) as input, and then learns the engagement behavior of audiences and their social relationships with other users to predict the loyalty and authenticity of audiences.

More specifically, our model consists of the following three components: (i) the contextualized engagement encoder, (ii) multi-relational GCNs (Graph Convolutional Networks), and (iii) multi-task decoder. In the contextualized engagement encoder, we use contextualized knowledge from user comments to generate user embeddings that represent distinct user commenting behavior. We next construct multi-relational engagement networks based on the different types of engagements and take the contextualized user embeddings as node features to learn different engaging relationships among users. Lastly, the multi-task decoder estimates the loyalty and authenticity scores of users with the outputs from the multi-relational GCNs.

8.2 Problem Statement

Our goal is to rank influencers based on their audiences' loyalty and authenticity by learning the engagement representations in the influencer-audience social network. In this section, we formally define the influencer-audience heterogeneous information network, and two evaluation metrics, audience loyalty, and authenticity.

Definition 1 ***Influencer-Audience Heterogeneous Information Network*** *is a social network of social media influencers and their audiences, who are connected based on engagement behaviors, including likes and comments. Given the two types of vertices corresponding to influencers ($\mathcal{V}_I$) and audiences ($\mathcal{V}_A$), the influencer-audience network $\mathcal{G} = \{\{\mathcal{V}_I, \mathcal{V}_A\}, \{\mathcal{E}_L, \mathcal{E}_C\}\}$ can be defined where $\mathcal{E}_L$ and $\mathcal{E}_C$ represent liking and commenting as engagement behaviors, respectively.*

To evaluate the quality of audiences, we define two metrics: *Audience Retention Rate* and *Influencer Fraud Score*, which measure the loyalty and authenticity of audiences, respectively.

Definition 2 ***Audience Retention Rate (ARR)*** *is the ratio of the number of audiences who consistently make engagements to the total number of audiences who have at least one engagement. Since the level of audience engagement may vary depending on the influencer's activity, we take into account the temporal variation of audience engagement by using multiple time frames. Given an influencer $i \in \mathbf{I}$ and a set of audience users $j, \forall j \in \mathbf{A}$, the audience retention rate can be calculated as:*

$$ARR_i = \frac{1}{|t|} \times \sum_t \frac{|e_{ij}^{t+1} \cap e_{ij}^t|}{|e_{ij}^t|},$$

where t is a time period and $e_{ij} \in \{\mathcal{E}_L \cup \mathcal{E}_C\}$.

An influencer has a high audience retention rate when his/her audience users consistently make engagements.

Definition 3 ***Influencer Fraud Score (IFS)*** *measures the intimacy between the influencers and engagement bots. For example, an influencer with a high IFS is more likely to be a fraudulent influencer who may have purchased fake engagements generated by social bots. Denote the sets of influencers, audiences, and bots are* $\mathbf{I}$, $\mathbf{A}$, *and* $\mathbf{B}$, *respectively. Given an influencer* $i \in \mathbf{I}$, *audiences* $j(\forall j \in \mathbf{A})$ *and pre-defined social bots* $b(\forall b \in \mathbf{B})$, *the IFS can be calculated as follows:*

$$IFS_i = \frac{f(v_i, v_b)}{f(v_i, v_j)} \times |e_{ib}|$$

where $f(v_1, v_2)$ *is the total number of engagements from* v_2 *to* v_1.

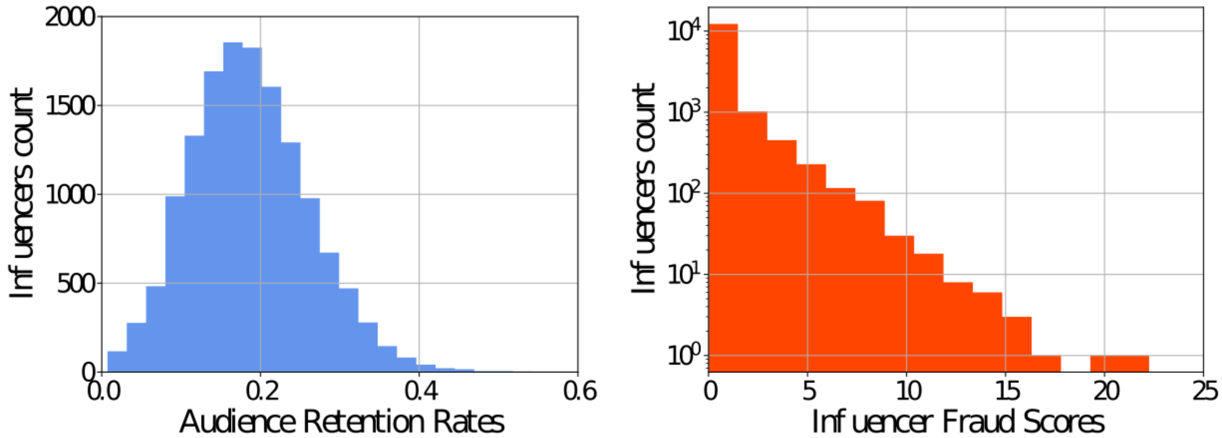
Based on the above two metrics, a set of influencers can be ranked in two different ways that essentially show the loyalty of audiences and their authenticity, by learning embeddings from the given influencer-audience heterogeneous information network.

8.3 Dataset

In this section, we describe the influencer-audience dataset and analyze audience engagement behaviors based on our proposed metrics.

8.3.1 Audience Data Collection

To evaluate influencer's audiences, we use the Instagram influencer dataset [92] which contains 33,935 influencers and their 10,180,500 Instagram posts. On average, we have 300 posts per influencer. The influencers in the dataset are classified into eight different categories including beauty, family, fashion, fitness, food, interior, pet, and travel. Each post in the dataset has a list of audiences who have engaged by liking or writing comments, thus we can collect influencer audience information. We notice that the posts in the dataset have been published between November 2010 and January 2018, but 87% of the posts were published after January 2017. That is, the number of posts exhibits a power-law distribution over influencers, which is likely to be attributed to the different posting habits between influencers. For accurate audience analysis, we only include posts published after January 2017, exclud-



(a) ARR

(b) IFS

Figure 8.2: Distributions of ARR and IFS of the influencers. ARR has a normal distribution, but IFS has an exponential distribution. The IFS distribution suggests that a small portion of influencers is related to engagement bots while most influencers do not purchase fake engagements.

ing 13% of the posts that were too outdated. We also exclude influencers with less than 10,000 followers and less than 100 posts since these influencers are considered as inactive influencers. Finally, we collect 9,290,895 unique audiences who had generated 21,374,920 likes and 44,473,797 comments to 6,244,555 Instagram posts published by 14,221 influencers.

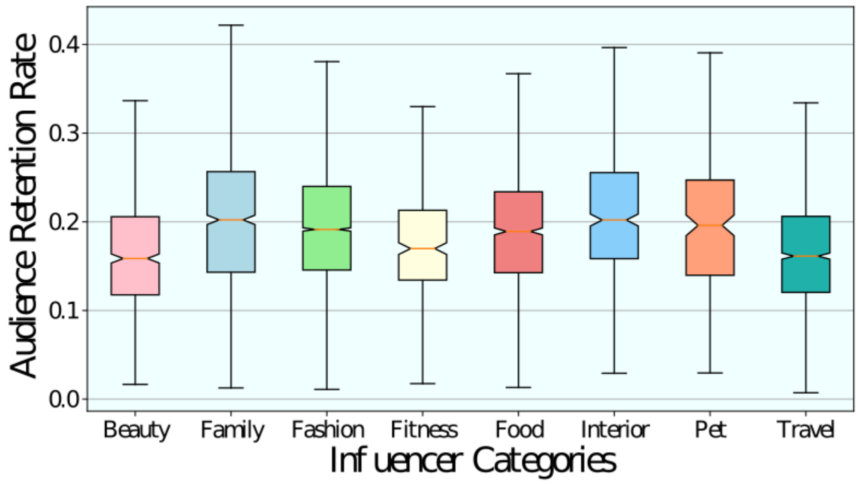
8.3.2 Analysis on Audience Evaluation

8.3.2.1 ARR Distribution

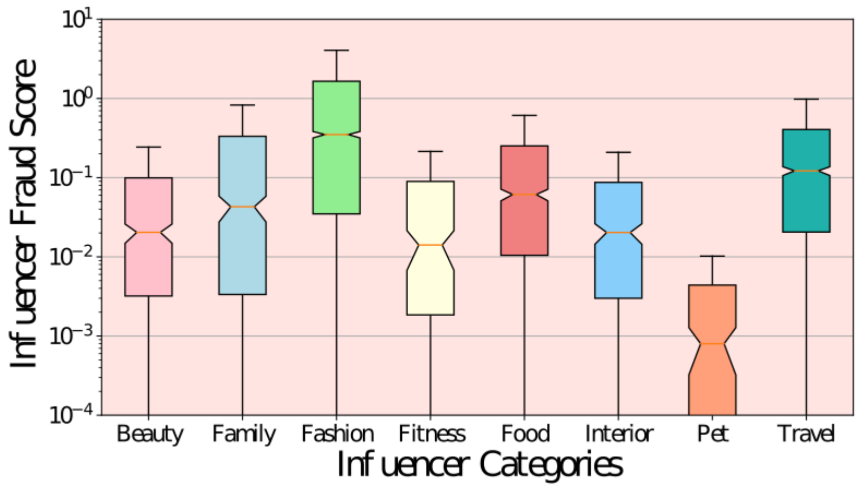
To calculate the ARR of every influencer, we first split the dataset into thirteen timeframes, each of which represents each month starting from January 2017 to January 2018. Based on the ARR definition, we first calculate the ratios of loyal audiences between two months (e.g., Jan.-Feb., Feb.-Mar.), and then compute the average of 12 ratio values to obtain the ARR for each influencer. Figure 8.2(a) shows the ARR distribution of the influencers that has a normal distribution. The standard deviation, average, and median ARR values of the influencers are 0.067, 0.187, and 0.182, respectively. This suggests that most influencers have similar ARR values, but some influencers have ARR values that are significantly higher or lower than the average value; the maximum and minimum ARR values are 0.545 and 0.007, respectively. Note that the average standard deviation of ARR over the timeframes is 0.036. This demonstrates that our proposed ARR well reflects the temporal variation of audience engagement.

8.3.2.2 IFS Distribution

To compute the IFS of the influencers, we find potential engagement bot accounts from the influencer-audience dataset based on the definition of engagement bots used in the previous studies [149, 4, 89]. According to these studies, inauthentic users, who are considered as potential bots, show different behaviors from authentic users; inauthentic users tend to have zero or a few numbers of followers and posts, generate lots of engagements, and have high similarity in their written comments. Based on these characteristics, we identify 1,822 bots that have zero followers and posts but have generated more than 1,000 engagements in our dataset. To verify whether the bots we found in our dataset are inauthentic users, we fetch Instagram pages of the 1,822 bot accounts and 5,000 randomly selected authentic user accounts as of January 2021. We find that 96.8% of the bot accounts have been deleted from Instagram while only 7.5% of the authentic user accounts are removed. This suggests that the identified bot accounts are highly likely to be inauthentic users since Instagram removes inauthentic activities and accounts [140, 77]. The IFS distribution of the influencers is shown in Figure 8.2(b). Unlike the ARR distribution, the scores show exponential distribution. This represents that most influencers have very low IFS while a few influencers are heavily connected to inauthentic audiences. Note that the median value is 0.077, and 81% of the total influencers have IFS less than 1.0 whereas only 2.7% of influencers have IFS greater than 5.0.



(a) Audience Retention Rate



(b) Influencer Fraud Score

Figure 8.3: Distributions of ARR and IFS of the influencers across their categories. Audience loyalty and authenticity values are varied over different types of influencers. Beauty and travel influencers tend to have lower audience loyalty than influencers in other categories, and many fashion influencers are connected to a large set of inauthentic audiences.

8.3.2.3 Audience Analysis on Influencer Categories

We further investigate the loyalty and authenticity of audiences across different influencer categories. Figure 8.3 shows the ARR and IFS distributions over the eight categories. We find that the family and interior influencers tend to have more loyal audiences than the beauty and travel influencers as shown in Figure 8.3(a). Note that median ARR values for family, interior, beauty, and travel are 0.202, 0.202, 0.158, and 0.161, respectively. We also observe that IFS values have a larger variance by category than ARR. In the dataset, fashion influencers tend to have more connections with inauthentic audiences than influencers in other categories. On the other hand, pet influencers are not likely to be related to the engagement bots in our dataset. Since the characteristics of audiences are different across the influencer categories, we comprehensively evaluate the proposed framework with different types of influencers in the experiment section.

8.3.2.4 Correlations with Audience Evaluation Metrics

We next examine the Pearson correlation between two evaluation metrics, ARR and IFS, to find mutuality in the two ranking lists. The correlation coefficient between ARR and IFS is 0.177 as shown in Table 8.1 which indicates a slightly positive correlation. Since bots consistently generate engagements, the influencers who have high IFS values might have high ARR values. However, high ARR values do not necessarily mean the influencers are connected to engagement bots, thereby having a weak correlation. In addition to the correlation between the two evaluation metrics, we also perform correlation studies between the evaluation metrics and the degree of influencer nodes in the influencer-audience network to check potential information leakage during training the proposed model. As shown in Table 8.1, there are no correlations between the influencer node degree and the audience evaluation metrics. This confirms that the popularity of influencers, which can be represented by the node degree in the network, is not related to the loyalty or authenticity of audiences.

8.4 Methodology

In this section, we present our proposed model. The overview of the proposed framework is illustrated in Figure 8.4. The framework consists of three components, engagement encoder,

Table 8.1: Pearson correlation between evaluation metrics (ARR, IFS) and the degrees of the influencer nodes (Degree)

	coefficient	p-value
ARR & IFS	0.177	< 0.001
ARR & Degree	-0.071	< 0.001
IFS & Degree	-0.027	< 0.005

multi-relational GCNs, and multi-task decoder.

8.4.1 Contextualized Engagement Encoder

To learn distinct engagement behaviors of users, we exploit user comments that may contain unique contextual characteristics of the users. The contextualized engagement encoder first merges all comments written by each user to make a sequence of concatenated comments as follows:

$$\mathbf{C}_u = \|c_0, c_1, \cdots, c_{|C_u|}\| \ (\forall user\ u \in \{\mathbf{I} \cup \mathbf{A}\}).$$

We utilize the concatenated user comments instead of learning contextualized knowledge from each comment separately since most user comments in social media are very short thus not having sufficient textual information.

To capture unique contextualized engagement features from a long sequence of user comments, we adopt the pre-trained Longformer [14], which is capable of processing long documents. As the concatenated comments generally have long sequences, it is difficult to apply the transformer-based models that utilize full self-attention due to its quadratic scaling. Longformer, however, addresses this issue by introducing global attention on special tokens thereby reducing the number of attention procedures and saving memory usage. In our framework, global attention is applied over user comments thus the relations across the comments can be induced. We generate contextualized embeddings as follows:

$$\mathbf{X} = Longformer(\mathbf{C}) \in \mathbb{R}^{n \times d},$$

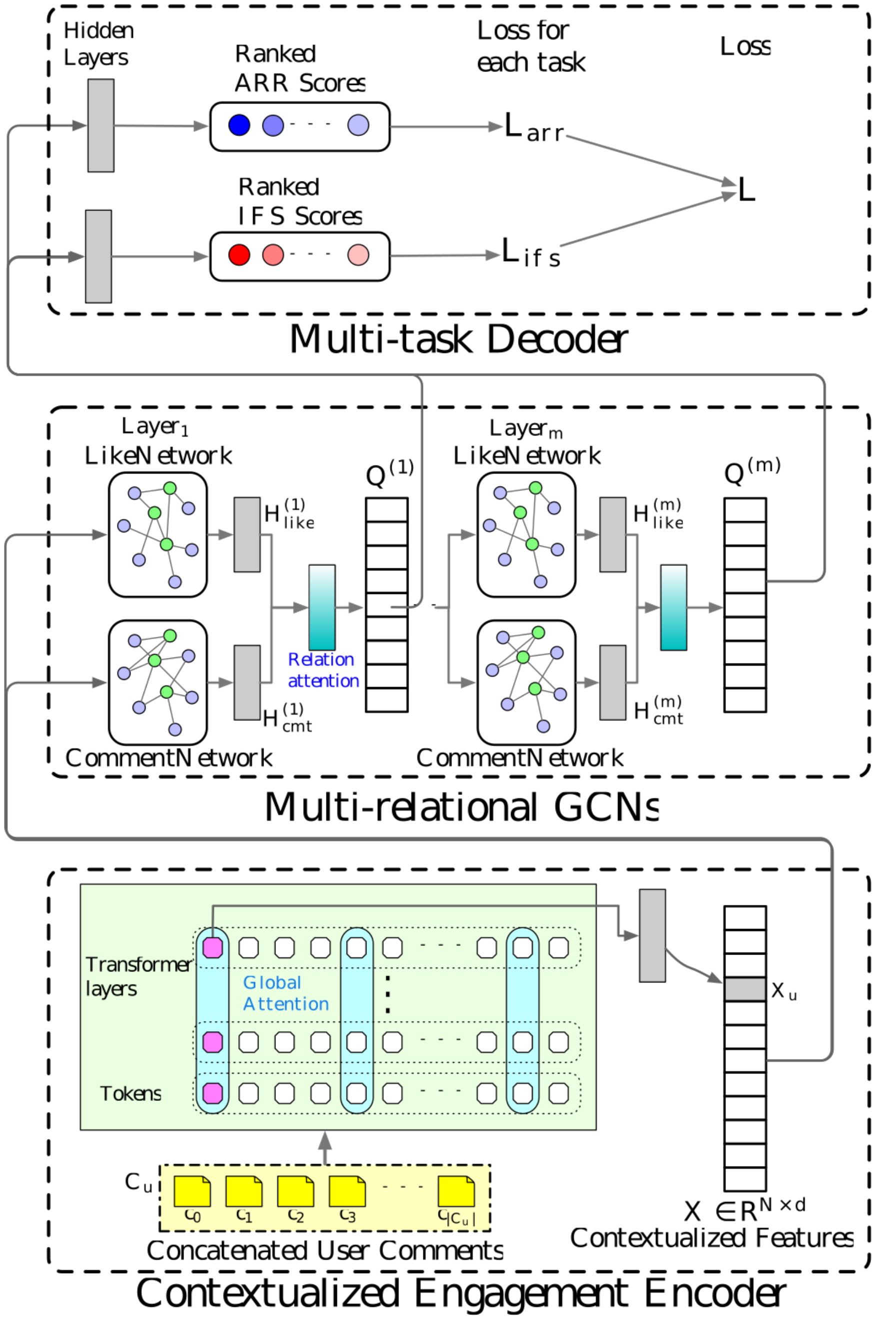


Figure 8.4: The overall framework of the proposed ALAIM. The framework consists of three components including the contextualized engagement encoder, the multi-relational GCNs, and the multi-task decoder.

where n is the number of users and d is the dimension size of a contextualized embedding.

8.4.2 Multi-Relational GCNs

To model social relations between influencers and their audiences based on engagements, we apply graph convolutional networks (GCNs) [94] to the influencer-audience heterogeneous information networks. Since the network has two relations, $\mathcal{R} = \{r_l, r_c\}$, where r_l and r_c represent like and comment relations, respectively, we use the multi-relational GCNs to learn interactions between the two types of engagements. For each relation $r \in \mathcal{R}$, we obtain a normalized adjacency matrix $\hat{\boldsymbol{A}}_{\boldsymbol{r}} = \boldsymbol{D}_r^{-\frac{1}{2}} A_r \boldsymbol{D}_r^{-\frac{1}{2}}$, where A_r is the adjacency matrix and D_r is the diagonal node degree matrix of a relation r. The output of the $i+1$-th layer in GCNs of relation r is then calculated as follows:

$$\boldsymbol{H}_{\boldsymbol{r}}^{(\boldsymbol{i+1})} = \sigma\left(\hat{\boldsymbol{A}}_{\boldsymbol{r}} \boldsymbol{Q}^{(\boldsymbol{i})} \boldsymbol{W}_{\boldsymbol{r}}^{(\boldsymbol{i})}\right),$$

where σ is ReLU activation function; $\boldsymbol{W}_{\boldsymbol{r}}^{(\boldsymbol{i})}$ is the weight parameters of relation r at the previous layer; $\boldsymbol{Q}^{(\boldsymbol{i})}$ is the outputs of the i-th layer. Note that the initial node features $\boldsymbol{Q}^{(\boldsymbol{0})} = \mathbf{X}$ for the both like and comment GCNs. We then apply attention over the outputs of the GCNs with different relations to acquire the output of the multi-relational layer as follows:

$$\boldsymbol{Q}^{(\boldsymbol{i})} = \sum_r \alpha_r \cdot \boldsymbol{H}_{\boldsymbol{r}}^{(\boldsymbol{i})},$$

where α_r is the estimated importance weight of relation r. This can be computed by using a softmax function as follows:

$$\alpha_r = \frac{\exp(\tanh\left(\mathcal{F}\left(\boldsymbol{H}_{\boldsymbol{r}}\right)\right))}{\sum_i^{|\mathcal{R}|} \exp(\tanh\left(\mathcal{F}\left(\boldsymbol{H}_{\boldsymbol{i}}\right)\right))},$$

where $\mathcal{F}()$ is a fully-connected layer and tanh() is the activation function.

Finally, the output of Multi-relational GCNs can be obtained as follows:

$$\mathbf{Q} = [\boldsymbol{Q}^{(\boldsymbol{1})}, \boldsymbol{Q}^{(\boldsymbol{2})}, \cdots, \boldsymbol{Q}^{(\boldsymbol{m})}],$$

where m is the number of layers in the multi-relational GCNs.

8.4.3 Multi-task Decoder

To conduct multiple tasks by learning the influencer-audience embeddings, the proposed multi-task decoder predicts corresponding scores for each task. In the framework, we have two ranking tasks to evaluate the influencers based on the audience retention rate (ARR) and the influencer fraud score (IFS). Note that another advantage of our framework is that it is easily extendable for any potential task, which utilizes influencer-audience embeddings, can be added as an additional task in the decoder.

We first estimates audience retention rates $\hat{y}_{arr}$ and influencer fraud scores $\hat{y}_{ifs}$ as follows:

$$\hat{y}_{arr} = \mathcal{F}_a\left(\sigma\left(\mathcal{F}_b\left(\mathbf{Q}\right)\right)\right),$$

$$\hat{y}_{ifs} = \mathcal{F}_c\left(\sigma\left(\mathcal{F}_d\left(\mathbf{Q}\right)\right)\right),$$

where σ is the ReLU activation function; $\mathcal{F}_a$, $\mathcal{F}_b$ and $\mathcal{F}_c$, $\mathcal{F}_d$ are fully-connected layers to predict ARR and IFS, respectively.

To properly rank the influencers based on the predicted values, we propose to use a list-wise learning-to-rank approach [172]. Denote that y_{arr} and y_{ifs} are ground truth for ARR and IFS, respectively; m is the list size. The losses for ARR and IFS then can be computed as:

$$\mathcal{L}_{arr}(\hat{\boldsymbol{y}}_{arr}) = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{l}(\hat{\boldsymbol{y}}_{arr\_i}(\boldsymbol{Q_i}), \boldsymbol{y_{arr\_i}}),$$
$$\mathcal{L}_{ifs}(\hat{\boldsymbol{y}}_{ifs}) = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{l}(\hat{\boldsymbol{y}}_{ifs\_i}(\boldsymbol{Q_i}), \boldsymbol{y_{ifs\_i}}),$$

where $\boldsymbol{l}(\hat{\boldsymbol{y}}_i(\boldsymbol{Q_i}), \boldsymbol{y_i})$ 0-1 loss that returns 0 when the ranked result equals to the ground truth and 1 otherwise.

Finally, the ultimate objective for the multi-task learning by summing up the losses as follows:

$$\mathcal{L} = \mathcal{L}_{arr} + \mathcal{L}_{ifs}.$$

Note that we use a combination of two loss functions since the amount of sampled data points for both tasks is the same thereby having equal proportions.

8.5 Experiments

8.5.1 Experimental Setting

8.5.1.1 Implementation Details

We split the dataset into train, validation, and test sets with a 7:1:2 ratio and use the same sets for all models. To prevent potential information leakage from learning the relationship among influencers and audiences, we ensure that the same influencers are not included across the three sets. We train the model by using posts from January to October in 2017, fine-tune the model using posts published in November 2017, and test with posts published in December 2017 and January 2018 to prevent temporal leakage. After fine-tuning the model with the validation set, we set the parameters as follows: the number of layers in the multi-relational GCNs as 2, the number of GCN features as 128, the batch size as 256, the list size for list-wise learning as 5, and the learning rate as 10^{-3}.

8.5.1.2 Baseline Methods

To compare the performance of the proposed ALAIM with other models, we consider the following three baseline methods. Convolutional neural networks (CNN) are considered as the first baseline method to understand the benefits of using the GCN-based approach. We also evaluate two open-sourced GCN-based methods, GCN [94] and R-GCN [145], to demonstrate the novelty of the proposed model. Note that we generate contextualized features per influencer by merging all comments on posts published by the corresponding influencer for the CNN baseline. For the GCN model, we combine the like and the comment networks into a single engagement network to make one adjacency matrix since the model only considers a single relational network. R-GCN learns interactions between different relation types by sharing the weight parameter. We consider all of the baseline methods as a single-task learning framework therefore we train the models separately for each task. For node features in the GCN-based baseline methods, we use one-hot encoded node type information that

Table 8.2: The number of influencers in different relevance levels over the two audience evaluation metrics, ARR, and IFS.

Relevance	Criteria	Number of Influencers
Audience Retention Rate (ARR)		
4	$ARR \geq 0.25$	2,744 (19.30%)
3	$0.25 > ARR \geq 0.20$	3,000 (21.10%)
2	$0.20 > ARR \geq 0.15$	3,790 (26.65%)
1	$0.15 > ARR \geq 0.10$	2,994 (21.05%)
0	$0.10 > ARR$	1,693 (11.90%)
Influencer Fraud Score (IFS)		
2	$IFS \geq 1.0$	2,671 (18.78%)
1	$1.0 > IFS \geq 0.1$	3,982 (28.00%)
0	$0.1 > IFS$	7,568 (53.22%)

indicates whether a node is an influencer or an audience. Moreover, we extend the GCN-based baseline methods by adding our proposed contextualized engagement encoder, named GCN+ and R-GCN+, thereby having the same node features as ALAIM. In addition to the three baseline methods, we have ALAIM-single which is a single-task learning model of the proposed framework to study the performance gain from the joint learning of multiple tasks.

8.5.1.3 Relevance Assignments

We use the normalized discounted cumulative gain (NDCG) [79] to measure the ranking performances of the models. To assign graded relevance values to the influencers, we divide the influencers into five and three different levels based on their ARR and IFS, respectively. Table 8.2 shows the number of influencers in different relevance levels and their criteria. Note that the three relevance levels for IFS can be denoted as groups of influencers who have

Table 8.3: Ranking performance measured by NDCG score. The proposed ALAIM outperforms other baseline methods in both audience evaluation tasks.

	NDCG@K					
	10	100	200	300	500	1000
Audience Retention Rate (ARR)						
CNN	0.372	0.428	0.411	0.437	0.403	0.479
GCN	0.627	0.622	0.613	0.601	0.592	0.616
GCN+	0.647	0.640	0.656	0.631	0.637	0.719
R-GCN	0.660	0.674	0.648	0.635	0.621	0.644
R-GCN+	0.682	0.726	0.702	0.687	0.657	0.727
ALAIM-single	0.841	0.784	0.759	0.738	0.740	0.766
ALAIM	**0.895**	**0.811**	**0.767**	**0.750**	**0.766**	**0.773**
Influencer Fraud Score (IFS)						
CNN	0.337	0.351	0.38	0.396	0.411	0.413
GCN	0.554	0.597	0.575	0.584	0.591	0.592
GCN+	0.650	0.648	0.653	0.639	0.666	0.704
R-GCN	0.673	0.622	0.612	0.640	0.638	0.620
R-GCN+	0.681	0.654	0.658	0.651	0.669	0.705
ALAIM-single	0.803	0.719	0.690	0.701	0.741	0.760
ALAIM	**0.820**	**0.735**	**0.705**	**0.711**	**0.744**	**0.764**

high, moderate, and low risks to be connected to bots. Based on the relevance values, we aim to rank influencers with high relevance scores in the first position.

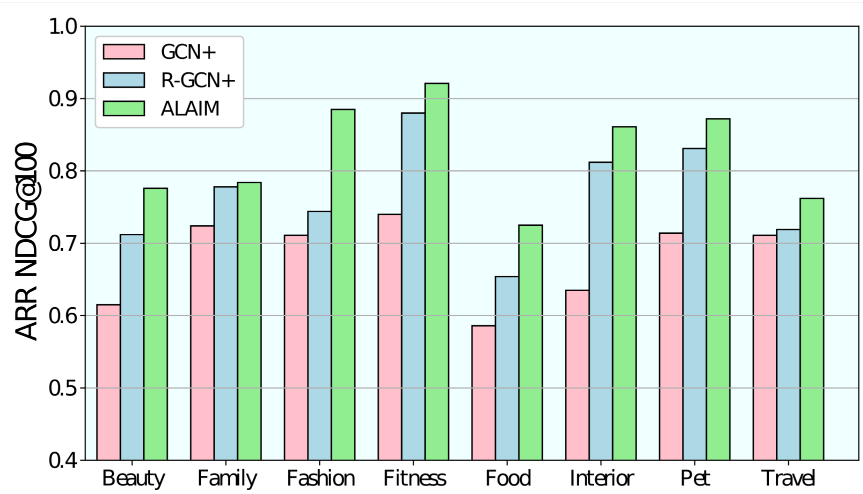
8.5.2 Experimental Results

8.5.2.1 Ranking Performance Evaluation

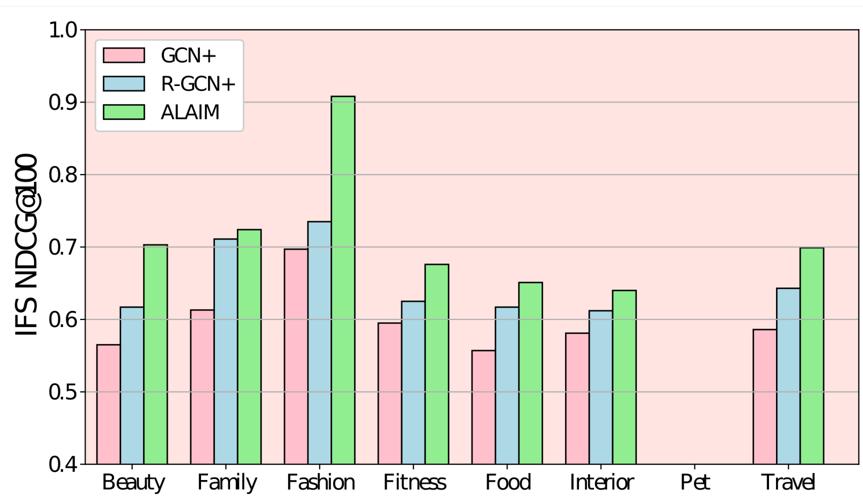
We first compare the ranking performance of ALAIM with other baseline methods. Table 8.3 shows NDCG scores of the models in the two audience evaluation tasks. As shown in Table 8.3, CNN shows poor ranking performance compared to the other GCN-based methods in both loyalty and authenticity evaluation ranking tasks. That is because the social relationships with audiences and their engaging behaviors are ignored in the CNN baseline method. This highlights the benefits of applying the GCN-based approach that enables the model to learn the features of neighbor nodes. Among the GCN-based models, GCN has relatively lower ranking performances than others. Since GCN does not adopt multi-relational types, potential knowledge from interactions between different types of relations can not be learned. R-GCN, on the other hand, shows better rank quality than GCN by exploiting the multi-relational networks to capture interactions between the like and comment networks. Note that R-GCN has 0.660 and 0.673 at NDCG@10 for ARR and IFS tasks, respectively, compared to that of GCN are 0.627 and 0.554. We also find that GCN+ and R-GCN+ which use the proposed contextualized node features have higher NDCG scores than GCN and R-GCN. This suggests that the contextualized features over comments are beneficial for both audience evaluation tasks. Our proposed model with a single task, ALAIM-single, outperforms other baseline methods at any position of the ranking. This demonstrates that node-level attention in the proposed multi-relational GCNs helps estimate the importance of nodes in different relations thereby inducing decent audience embeddings. Finally, ALAIM, which incorporating multi-task learning, outperforms all baseline methods. Note that there are no significant differences between ALAIM and ALAIM-single at NDCG@1000, but performance improvement can be found in the higher-ranking position. This implies that knowledge is shared from the related task benefits to evaluate audiences.

8.5.2.2 Ranking Performance across Influencer Categories

We next investigate the ranking performance over different influencer categories since the interests or engagement behaviors of audiences may be varied upon the types of influencers' expertise. We randomly select influencers in each category from the test dataset to assure the influencers for testing are not in the training set. Note that all the pet influencers in the testing set are assigned to the relevance level 0 in IFS, thus IFS ranking results for the



(a) ARR ranking results



(b) IFS ranking results

Figure 8.5: Ranking performance across the influencer categories. ALAIM shows robust ranking results for loyalty and authenticity evaluations.

pet influencers are always zero. Figure 8.5 shows the ranking results of ALAIM and baseline methods measured by NDCG@100 for ARR and IFS across the eight influencer categories. We find that ranking performance varies over the categories but the proposed ALAIM shows robust ranking results across all influencer categories and outperforms other baselines. This suggests that our proposed framework can be applied to evaluate specific types of influencers.

8.6 Analysis and Discussions

In this section, we conduct analytical studies to discuss (i) the audience evaluation metrics, (ii) the contextualized engagement embedding, and (iii) node features.

8.6.1 Analysis on Evaluation Metrics

In influencer marketing, the number of followers and likes, and the engagement rate, which is the ratio of the average number of likes to the number of followers, have been widely used to find effective influencers [22, 170]. However, those metrics may fail to measure the quality of influencers' audiences who are potential customers of marketing campaigns. Therefore, in this study, we propose two metrics, ARR and IFS, to evaluate the loyalty and authenticity of audiences, respectively. To understand the utility of the proposed ALAIM and the importance of ARR and IFS, we carry out a case study by investigating highly ranked influencers.

Table 8.4 shows the fifteen example influencers, who are ranked by the proposed metrics and engagement rate, with their number of followers and the average number of likes. We first find that influencers A to E, who are selected based on engagement rate, have remarkably high engagement rates; they have engagement rates higher than 7%. Note that influencers with engagement rates of 2-3% and 4-6% can be considered as good and excellent, respectively [170]. However, these influencers have average ARR values, and some of them (e.g., influencers D & E) have very high IFS values. This suggests that if only the engagement rate is considered, influencers using fake engagements or influencers with less loyal audiences may be included. In addition, the influencers obtained this way are usually micro-influencers with relatively few followers compared to other influencers; influencers from A to E have less than 100,000 followers. That is because the engagement rate is generally inversely proportional to the number of followers. On the other hand, the proposed ALAIM successfully finds a set of influencers who have many loyal audiences. For example, influencers from F to J not only have high ARR values but also have good engagement rates and wide ranges of the number

Table 8.4: The example influencers ranked by different audience engagement metrics including the engagement rate (EngRate), ARR, and IFS. Although the engagement rate is a widely used indicator in influencer marketing, considering only the engagement rate may ignore loyal audiences or include engagement bots.

Influencer	Followers	Avg. likes	EngRate	ARR	IFS
Engagement Rate (EngRate)					
A	16,530	2,018	**12.21%**	0.171	0.006
B	12,507	1,186	**9.48%**	0.187	1.330
C	41,060	3,640	**8.87%**	0.101	0.000
D	16,320	1,273	**7.80%**	0.175	2.275
E	17,921	1,357	**7.57%**	0.175	7.398
Audience Retention Rate (ARR)					
F	1,970,911	47,036	2.39%	**0.420**	0.000
G	784,270	27,563	3.51%	**0.389**	0.000
H	17,025	567	3.33%	**0.369**	0.004
I	555,461	29,798	5.36%	**0.360**	0.000
J	21,759	1,230	5.65%	**0.340**	0.082
Influencer Fraud Score (IFS)					
K	13,131	1,363	10.38%	0.209	**13.961**
L	125,024	2,930	2.34%	0.330	**9.316**
M	35,146	1,976	5.62%	0.232	**7.548**
N	36,551	1,396	3.82%	0.274	**6.647**
O	494,239	8,441	1.71%	0.305	**4.801**

of followers. We can also utilize ALAIM to filter out influencers with fraudulent behavior. Influencers from K to O have good engagement rates and ARR values, but they should not be recommended to marketers due to high IFS values.

Table 8.5: Ranking results of the proposed ALAIM with different contextualized features. The model with Longformer outperforms the model with BERT since it efficiently learns engagement behavior from the long sequence of very short user comments.

	NDCG@K					
	10	100	200	300	500	1000
Audience Retention Rate (ARR)						
ALAIM-NoContext	0.773	0.672	0.662	0.629	0.657	0.704
ALAIM-Bert	0.845	0.771	0.746	0.733	0.741	0.752
ALAIM	**0.895**	**0.811**	**0.767**	**0.750**	**0.766**	**0.773**
Influencer Fraud Score (IFS)						
ALAIM-NoContext	0.720	0.652	0.619	0.629	0.668	0.681
ALAIM-Bert	0.767	0.719	0.680	0.688	0.735	0.759
ALAIM	**0.820**	**0.735**	**0.705**	**0.711**	**0.744**	**0.764**

In summary, the proposed ALAIM is useful in finding influencers with good performance indicators. Unlike the number of likes or the engagement rates, which are accessible from most social media platforms, the ARR and IFS cannot be simply inferred as they require audience engagement information. Therefore, we believe that the proposed framework shows great utility in general influencer marketing since it only takes the engagement network and the contextualized engagement embeddings as input but not utilizes other information such as the degree of the engagements and bot labels.

8.6.2 Analysis on Contextualized Engagement Embedding

The proposed framework uses Longformer [14] to capture the contextualized information from the audiences' comments. To understand the importance of the contextualized engagement embeddings, we conduct experiments with (i) original ALAIM, (ii) ALAIM-Bert, and (iii) ALAIM without contextualized embeddings. We first employ BERT [46] to generate the engagement embeddings. Since the main drawback of BERT is that the computational cost of the attention calculations grows quadratically with the length of an input sequence, we generate a BERT feature for each comment without concatenating all comments into one. We then combine the generated comment BERT features to make the BERT-based engagement embedding. We also deploy ALAIM without the contextualized engagement encoder, named ALAIM-NoContext, for comparison purposes. For this model, we use one-hot encoded node type information as the node features.

Table 8.5 shows the ranking results of the proposed framework with three different contextualized embeddings. We find that ALAIM-NoContext has the lowest ranking performance since it only relies on the knowledge learned from multi-relational networks. This demonstrates that contextual information over comments is very useful to capture distinct characteristics of each audience. For example, an embedding of a loyal audience who tends to write comments with positive sentiments must be different from an embedding of another user who usually uses simple words or emojis to write comments. The results also present that ALAIM-Bert has lower ranking performances than ALAIM with Longformer. This implies that BERT sometimes fails to learn contextualized information from very short comments which contain only a couple of words thereby the combined BERT features can not represent the engaging behavior of an audience well. Besides the ranking performance, we confirm that Longformer significantly reduces the computational time and memory consumption compared to BERT. In our experimental setting, Longformer is about 10 times faster and saves about 10 times of memory than BERT.

8.6.3 Analysis on Node Features

We propose to use only the contextualized engagement embeddings as node features to make the framework general; our framework requires minimal engagement information as input. However, any potential information, given from social media platforms, that represents the characteristics of a node can be added as a node feature as our proposed framework takes the GCN-based approach. In this analysis study, we use the influencer category information as additional node features since we found that both ARR and IFS are varied across different

Table 8.6: NDCG@100 scores of the proposed framework with different node features across the influencer categories. The proposed framework can have performance gain by adding node features on specific tasks.

	NDCG@100							
	Beauty	Family	Fashion	Fitness	Food	Interior	Pet	Travel
Audience Retention Rate (ARR)								
ALAIM	0.776	**0.784**	0.885	**0.921**	0.725	0.861	0.872	0.762
ALAIM-category	**0.782**	0.783	**0.890**	**0.921**	**0.739**	**0.868**	**0.874**	**0.766**
Influencer Fraud Score (IFS)								
ALAIM	**0.703**	0.724	**0.908**	0.676	**0.651**	0.640	0.000	0.699
ALAIM-category	0.698	**0.728**	**0.908**	**0.701**	0.643	**0.672**	0.000	**0.708**

influencer categories.

Table 8.6 shows NDCG@100 scores of the proposed framework with and without the influencer category information as node feature. Note that We use the one-hot coded category node features for ALAIM-category. We find that the extra node features improve the ranking performance on some categories of influencers in certain tasks. For example, ALAIM-category shows better ranking performances in (i) beauty, fashion, and food categories on the ARR task, and (ii) family, fitness, and travel categories on the IFS task, than ALAIM with no category information as a node feature. On the other hand, the category node features do not improve the performance in some other categories. This reveals that utilizing additional node features can be beneficial for specific tasks. Therefore, marketers or social media platforms may add unique features to the proposed framework for their objectives.

CHAPTER 9

InfluencerRank: Discovering Effective Influencers via Graph Convolutional Attentive Recurrent Neural Networks

9.1 Introduction

Influencers are known as individuals who influence a magnificent number of people on social media. This, in turn, has attracted great attention to marketers since influencers and their huge fan bases can be considered as marketing channels and audiences, respectively [44, 50]. More recently, companies have started hiring influencers to advertise products for targeted audiences and expand brand awareness. It has been reported that the value of the influencer marketing industry can reach $15 billion by 2022 [146].

Due to the rapid growth of social media and influencer marketing, discovering effective influencers on social media has become increasingly important [139, 84]. For measuring user influence on social media, well-known metrics, such as the numbers of followers, retweets, and mentions, have been widely applied [10, 148]. In addition, information propagation [141, 152, 86], social connections [101], network centrality [31], and multi-relational network [107, 88] have been used to identify influencers on social media. Among the various measures, the *effectiveness of influence* [103], often measured by the engagement rate [44, 105, 39], has been considered as crucial in identifying effective influencers especially in the marketing domain. The engagement rate can be calculated as the ratio of the average number of likes to the number of followers, which essentially shows how much audiences engage with the corresponding influencer.

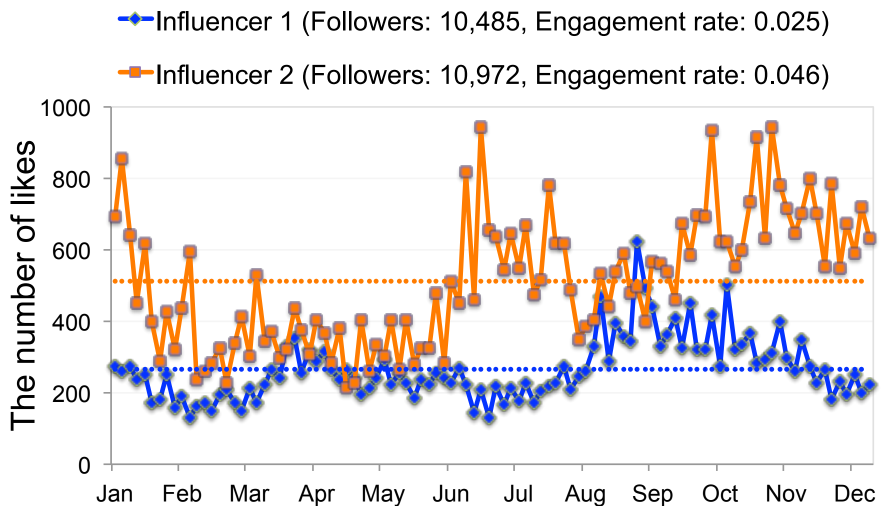


Figure 9.1: The number of likes on posts published by two influencers across the time. Although two influencers have similar numbers of followers, the average numbers of likes (i.e., dotted lines) are significantly different. Additionally, the number of likes dynamically changes over time.

To discover the effective influencers (i.e., influencers with high engagement rates), previous work used posting behaviors of influencers or characteristics of their posts. For example, some researchers utilized the social networks among influencers [141, 103, 52]; analyzed post contents to derive statistical features in identifying influencers [101]. However, none of these studies jointly and comprehensively modeled posting behaviors, post characteristics, and social networking behaviors, which may result in a biased or partial representation of the effectiveness of influencer. For example, as shown in Figure 9.1, two influencers have similar numbers of followers hence they may be considered as having similar effectiveness, but their actual engagement rates are shown to be significantly different. Although some methods applied the PageRank algorithm on influencer-content graphs [152] and independently derived the features of influencers and posts [103], the PageRank algorithm can be biased to a certain type of nodes [18] when the relations between influencers and posts are ignored by independent features.

To address this issue, we propose to use a heterogeneous network to model the effectiveness of influencers with their posting behavior, social networking behavior, and post characteristics together. In addition, considering historical behavioral patterns can be further beneficial to discover effective influencers since posting behavior of an influencer can change dynamically over time. For instance, as shown in Figure 9.1, although an influencer does not receive many likes in the most recent time period, he/she may receive many number of likes in the future if he/she was used to get great attention in the past. Moreover, analyzing time-varying behavior patterns can provide more evidence on the robustness of an influencer. For example, the unstable performance (or effectiveness) of an influencer over time may not be desirable even if he/she satisfies the performance in the most recent time period. Hence, taking such time-varying behavior patterns into account for discovering effective influencers is essential. However, most of the prior studies only focused on the most recent information without considering the historical patterns of influencers.

In this paper, we propose *InfluencerRank*, a learning framework, that discovers effective influencers in social media by learning historical behavioral patterns of influencers. For comprehensively representing the effectiveness of an influencer, we build a heterogeneous information network that consists of influencers, hashtags, user tags, and image objects used by influencers for each historical time period [174, 180]. To learn the complex posting behaviors, social networking, and post characteristics of each influencer, we apply graph convolutional networks (GCNs) [94] with well-designed influencer features, thereby deriving the influencer representation at a certain period. Based on the influencer representations over different historical time periods, the attentive recurrent neural network is proposed to learn the sequential and temporal behaviors to derive an ultimate representation. Finally, a learning-to-rank framework ranks a list of influencers to discover the ones who are more effective than others. The utility is expected to significantly increase given a recent decision of Instagram, one of the most popular influencer marketing platform [119], that hides the number of likes on each post [78, 104] to help mental health issues of social media users [142]. Unlike prior work, the number of likes is not used in discovering influencers in *InfluencerRank*, hence our proposed model can be particularly used by brands with relatively small business sizes, who may be suffering from the heavy expense of discovering effective influencers among millions of candidates [76] in a situation where the number of likes is hidden from other users.

9.2 Problem Statement

In this section, we formally define the effectiveness metric of an influencer and then formulate the problem of discovering effective influencers.

Definition 4 ***Engagement rate*** *is a widely-used metric in influencer marketing that shows how much audiences actively engage with an influencer [44, 105, 39]. Given an influencer* u*, the engagement rate of the influencer at time* t *is calculated as follows:*

$$E_u^t = \frac{l_u^t}{f_u}$$

where f_u *is the number of followers who follow the influencer* u *and* l_u^t *is the average number of likes on content posted by the influencer* u *at timestamp* t*.*

Based on the definition of influencer effectiveness, we introduce the influencer ranking problem. Let U be the set of influencers. For each timestamp t, we suppose that an influencer u has published a set of posts P_u^t. Given the set of influencers U and their posts published until time k, $\{P_u^t \mid 1 \leq t \leq k\}$, the goal of this work is to discover influencers with high engagement rates at time k by ranking all influencers $u \in U$ so that $E_{u_i}^k$ is greater than $E_{u_j}^k$ if the influencer u_i is ranked higher than the influencer u_j.

9.3 Influencer Ranking Model Framework

In this section, we propose InfluencerRank that learns the temporal dynamics of the engagement rates of influencers to automatically discover highly effective influencers. Figure 9.2 shows the overall framework of the proposed InfluencerRank. The framework takes a series of influencer social networks as input, where each network is composed of influencers and different entities, including but not limited to image objects, hashtags, and other users in social media. The graph convolutional networks (GCNs) are then applied to the input social

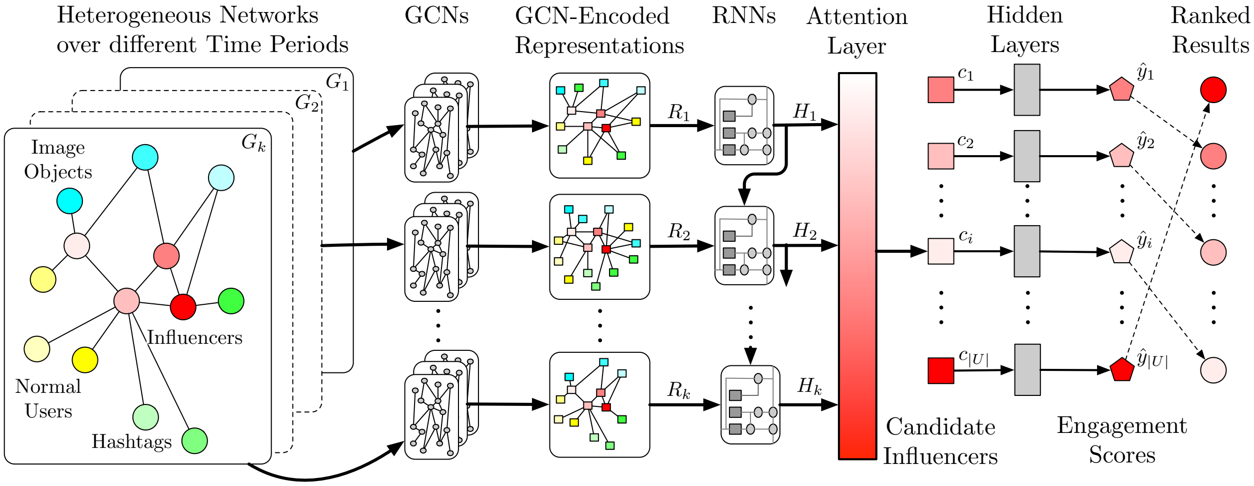


Figure 9.2: The overall framework of the proposed *InfluencerRank*.

networks to derive appropriate node representations that capture social relationships and posting characteristics of influencers at a certain time. The GCN-encoded representations across different times are then fed into a recurrent neural network to learn from the sequence of the node representations. The attention mechanism is then applied to the whole sequence of representations to finally derive the effectiveness scores of candidate influencers and rank them for discovering effective influencers. In sum, Table 9.1 further summarizes the major notations in this paper.

9.3.1 Heterogeneous Information Networks

To represent the dynamics of the engagement rates on a sequence of time, we build k heterogeneous networks $\mathbb{G} = \{G_1, G_2, \cdots, G_k\}$ based on the influencers and other relevant entities. Hence, G_t can further characterize the relationships of influencers and their posting behaviors at time t.

9.3.1.1 Heterogeneous Nodes and Embedded Features

We build a heterogeneous network G_t for time t with four different types of nodes, including influencers, hashtags, image objects, and other users in social media. Given an influencer u,

Table 9.1: Summary of notations and their descriptions.

Notation	Description
E_u^t	the engagement rate of an influencer u at time t.
l_u^t	the average number of engagements on contents posted by the influencer u at time t.
f_u^t	the number of followers for an influencer u at time t.
U	the set of influencers.
P_u^t	the posts published by the influencer u at time t.
$G_t = (\boldsymbol{X_t}, \boldsymbol{A_t})$	the heterogeneous network for time t with the node features X_t and the adjacency matrix A_t.
$\boldsymbol{\hat{A}_t}$	Normalized adjacency matrix transformed from $\boldsymbol{A_t}$.
d	the number of dimensions for embedded node features.
$\boldsymbol{D}$	the diagonal degree matrix of $\boldsymbol{A_t}$.
r	the number of hidden dimensions in GCNs.
$\boldsymbol{F^{(i)}}$	the outputs of the i-th GCN layer.
$\boldsymbol{W^{(i)}}$	the weight matrix between $\boldsymbol{F^{(i)}}$ and $\boldsymbol{F^{(i+1)}}$.
$\boldsymbol{R_t}$	the GCN-encoded representation for time t.
$\boldsymbol{H_t}$	the hidden states in the RNN for time t.
$\boldsymbol{S}$	the list of hidden states in the RNN over time.
τ_t	the importance weight for $\boldsymbol{H_t}$.
$\mathcal{F}_a(\cdot)$	the fully-connected layer for deriving τ_t.
α_t	the normalized importance weight for $\boldsymbol{H_t}$.
$\boldsymbol{c_u}$	the final representation of the influencer u.
$\boldsymbol{\hat{y}_u}$	the predicted engagement score for the influencer u.
$\mathcal{F}_b(\cdot), \mathcal{F}_c(\cdot)$	the fully-connected layers for inferring $\boldsymbol{\hat{y}_u}$.

we extract all of the hashtags $\{h_i\}_{i=1}^a \in H$ and mentioned users (i.e., user tags) $\{v_j\}_{j=1}^b \in V$ from posts P_u^t, where a and b indicate the number of extracted hashtags and mentioned users, respectively. Note that the mentioned users can be either influencers, brands, or other normal users. In addition, the categories of objects shown in the posted images $\{o_k\}_{k=1}^{c} \in O$ are also considered as nodes. Since each type of node has unique features, we denote the node features of influencers, mentioned users, hashtags, and object categories in images as $\boldsymbol{X_t^U}$, $\boldsymbol{X_t^V}$, $\boldsymbol{X_t^H}$, and $\boldsymbol{X_t^O}$, respectively. We then represent embedded features of each node as $\boldsymbol{X_t} = [\boldsymbol{X_t^U}; \boldsymbol{X_t^V}; \boldsymbol{X_t^H}; \boldsymbol{X_t^O}] \in \mathbb{R}^{N \times d}$, where N is the total number of all four types of nodes and d is the number of embedded node features. Note that we describe the details of node features utilized for the experiments in Section 9.3.5.

9.3.1.2 Edge Construction and Adjacency Matrix

The edges in the heterogeneous network indicate the interactions between entities behind nodes. For example, if an influencer mentioned the hashtag `#makeup` and posted an image of cosmetic products, the influencer node will be connected to the node of the `#makeup` hashtag and the node of the cosmetic image object. Given a timestamp t, we make a sparse adjacency matrix $\boldsymbol{A_t} \in \mathbb{R}^{N \times N}$, where $A_{ij}^t = 1$ indicates a connection between the i-th and j-th nodes.

Finally, a set of k heterogeneous networks $\mathbb{G}$ with the sets of node features and adjacency matrices can be constructed as follows:

$$\mathbb{G} = \{G_1, G_2, \cdots, G_k\},$$

where $G_t = \boldsymbol{(X_t, A_t)}$ indicates both the node embedded features $\boldsymbol{X_t}$ and the heterogeneous network structure A_t at time t.

9.3.2 Graph Convolutional Networks

For the heterogeneous network G_t of each time t, our proposed InfluencerRank applies Graph Convolutional Networks (GCNs) [94] to generate node representations over time. GCNs first generate a normalized adjacency matrix $\boldsymbol{\hat{A}_t}$ by transforming the adjacency matrix $\boldsymbol{A_t}$ with the diagonal degree matrix D as $\boldsymbol{\hat{A}_t} = \boldsymbol{D}^{-\frac{1}{2}} A_t \boldsymbol{D}^{-\frac{1}{2}}$. GCNs then stack multiple GCN layers where each layer takes outputs of the previous layer and performs nonlinear transformation to propagate information through different layers. The i-th layer in GCNs then outputs

$\boldsymbol{F^{(i)}} \in \mathbb{R}^{N\times r}$ as follows:

$$\boldsymbol{F^{(i)}} = \sigma\left(\boldsymbol{\hat{A}_t F^{(i-1)} W^{(i-1)}}\right),$$

where r is the number of hidden dimensions in GCNs, $\boldsymbol{F^{(i-1)}}$ is the outputs of the previous layer, $\boldsymbol{W^{(i-1)}}$ is a matrix of trainable weights, and $\sigma(\cdot)$ is a nonlinear activation function. We use $\boldsymbol{X_t}$ for $\boldsymbol{F^{(0)}}$ as the input of the first GCN layer. The final output of the GCNs $\boldsymbol{R_t}$ at time t can be represented as follows:

$$\boldsymbol{R_t} = \left[\boldsymbol{F^{(1)}}, \boldsymbol{F^{(2)}}, \ldots, \boldsymbol{F^{(e)}}\right],$$

where e is the number of layers in GCNs.

Finally, we can obtain a sequence of GCN-encoded node representations, $[\boldsymbol{R_1}, \ldots, \boldsymbol{R_k}]$, to implicitly represent the knowledge about influencers over time.

9.3.3 Attentive Recurrent Neural Networks

9.3.3.1 Learning Graph Dynamics

Based on the sequence of GCN-encoded node representations, $[\boldsymbol{R_1}, \ldots, \boldsymbol{R_k}]$, InfluencerRank applies Recurrent Neural Networks (RNNs) to the model framework. More specifically, we employ Gated Recurrent Units (GRUs) [36], which use update gate and reset gate inside the unit to carry information flow over many time periods, to capture long-term temporal dependencies from the heterogeneous networks. Note that we decide to use GRU instead of Long Short-Term Memory (LSTM) [72] in conducting performance comparison in Section 9.5.4. Each GRU takes hidden states from the previous unit and the GCN representations as input and then outputs hidden states of the current time. More formally, the hidden states at time t, H_t is computed as follows:

$$H_t = (1 - z_t)H_{t-1} + z_t\tilde{H}_t,$$

where z_t is an update gate at time t and $\tilde{H}_t$ is the candidate state at time t. The candidate state is updated as follows:

$$\tilde{H}_t = tanh(W \cdot [r_t \odot H_{t-1}, R_t]),$$

where r_t is an reset gate at time t, $\odot$ is an element-wise multiplication, and R_t is the GCN representations at time t. Finally, InfluencerRank obtains the whole states of GRUs as follows:

$$\boldsymbol{S} = [\boldsymbol{H_1}, \boldsymbol{H_2}, \ldots, \boldsymbol{H_k}] .$$

9.3.3.2 Attention over Time

To acquire the final influencer representations, InfluencerRank applies the attention mechanism [8] to the whole state embeddings derived from GRUs $\boldsymbol{S}$. The attention mechanism allows InfluencerRank to learn the dynamics of the engagement rates by taking into account only a certain time period.

For each timestep t, InfluencerRank estimates the importance weight of the corresponding state embedding by applying a projection as:

$$\tau_t = \tanh\left(\mathcal{F}_a\left(\boldsymbol{H_t}\right)\right),$$

where $\mathcal{F}_a(\cdot)$ is a fully-connected layer; $tanh(\cdot)$ is the activation function. We then compute the weights of each timestep by using a softmax function as:

$$\alpha_t = \frac{\exp(\tau_t)}{\sum_{i=1}^{k} \exp(\tau_i)}.$$

Finally, InfluencerRank derives the ultimate representation of candidate influencers by using the weighted sum as follows:

$$\boldsymbol{c} = \sum_{i=1}^{k} \alpha_i \cdot \boldsymbol{H_i}.$$

9.3.4 Engagement Score Estimation

For an influencer u, InfluencerRank takes the corresponding ultimate representation $\boldsymbol{c_u}$ as the input and then predicts an engagement score $\hat{y_u}$ that is proportional to the engagement rate E_u^k as follows:

$$\hat{\boldsymbol{y_u}} = \mathcal{F}_c\left(\mathrm{ReLU}\left(\mathcal{F}_b\left(\boldsymbol{c_u}\right)\right)\right),$$

where a non-linear transformation is carried out in a fully-connected layer $\mathcal{F}_c(\cdot)$ with the ReLU activation function and the engagement rate is estimated in another fully-connected layer $\mathcal{F}_b(\cdot)$.

9.3.4.1 List-wise Ranking and Optimization

InfluencerRank treats the task as a ranking problem and optimizes the ranking performance with a list-wise learning-to-rank framework [172]. Suppose Z is the set of features for influencers to be ranked; Y is the space of all possible rankings. During training, we sample m labeled influencers from the whole training space as an i.i.d. candidate ranked list $S = \{(Z_i, \boldsymbol{y_i})\}_{i=1}^{m} \sim P_{ZY}$, where P_{ZY} is the unknown target joint probability distribution of Z and Y. Therefore, the corresponding loss $\mathcal{L}_S$ can be considered as:

$$\mathcal{L}_S(\boldsymbol{\hat{y}}) = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{l}(\boldsymbol{\hat{y}}(\boldsymbol{Z_i}), \boldsymbol{y_i}),$$

where $\boldsymbol{l}(\boldsymbol{\hat{y}}(\boldsymbol{Z_i}), \boldsymbol{y})$ is the 0-1 loss between $\boldsymbol{\hat{y}}(\boldsymbol{Z_i})$ and the rank in $\boldsymbol{y}$; $\boldsymbol{y_i}$ denotes the ground-truth ranking.

9.3.5 Node Features

In this subsection, we describe node features in the heterogeneous network. To understand the relationship between the engagement rate of an influencer and the characteristics of the corresponding influencer, we introduce six types of node features, including node type, profile, image, text, posting, and reaction features.Note that most of the features are only applicable for influencer nodes while the remaining nodes (e.g., hashtags, image objects) hold zeros for the inapplicable features. For the feature engineering, we deploy the average, median, minimum, and maximum values for the features that need to be aggregated with statistics.

- **Node type features.** The one-hot coded feature that indicates one of the four node types, including influencers, other users, hashtags, and image objects.

Table 9.2: Six categories of node features that represent node characteristics.

Category	Feature	Description
Node	Node Type	Node type in the heterogeneous network.
Profile	Followers and Followees	Numbers of followers and followees.
	Posts	Number of published posts.
	Influncer Category	Major interest of the influencer.
Image [59]	Brightness	Perception of luminance of posted images.
	Colorfulness [69]	Chromatic level of posted images.
	Color Temperature	Degree of warmness of posted images.
Text [71]	Hashtags	Number of hashtags(#) in a post.
	Usertags	Number of usertags(@) in a post.
	Emojis	Number of emojis in a post.
	Length	Length of text in a post.
	Post Sentiment [60]	Sentiment scores of text in a post.
Posting	Category Rate	Ratio of the posts in a certain category.
	Advertising Rate	Ratio of the advertising posts.
	Feedback Rate	Ratio of the responded posts.
	Posting Interval	Time interval between consecutive posts.
Reaction	Comment Sentiment [60]	Sentiment scores of post comments.

- **Profile features.** For each influencer node, we exploit the numbers of followers, followees, and posts which are the most commonly used metrics to measure user influence in social networks [139]. Additionally, we consider a category of influencers from eight influencer categories defined in the previous study [92].

- **Image features.** The previous study [59] showed that the characteristics of images on social media posts affect its popularity. In addition to the image objects which are considered as nodes in the heterogeneous network, we add the attributes of visual perception of the images to understand how influencers create images. We compute the brightness, colorfulness [69], and color temperature of the posted images based on their RGB values.

- **Text features.** To understand how textual usage of influencers affects the engagement rate, we retrieve various text features. More specifically, we use the numbers of hashtags, user tags, and emojis that are widely used functions on social media, and the length of captions which can represent how much detailed information is in the caption [71]. Moreover, we also calculate the sentiment scores of captions to learn how positive or negative emotions are carried through the captions by using VADER [60].

- **Posting features.** The features in this category can provide information about how influencers use social media from various aspects. We first exploit the portion of the number of posts in one of the ten post categories [92] to the total number of posts to understand the posting behavior of influencers. In addition to the post category rate feature, we also examine the portion of the number of advertising posts to the total posts published by an influencer; posting too many paid advertisements can show negative impacts on the popularity [50, 176]. We also consider the feedback rate and posting interval, which are the measures of the interaction with their followers and activeness, respectively. The feedback rate is calculated as the ratio of the number of posts that contain the influencers' responses to the user comments to the number of total posts. The posting interval is the average time gap between posts that are in chronological order.

- **Reaction features.** We use the user comments to generate the user reaction feature. Specifically, we compute the sentiment scores of comments that are written by audiences of the influencers' posts. Note that we do not consider the number of likes and comments as node features since it can directly imply the engagement rates of influencers.

9.4 Experiments

9.4.1 Experimental Dataset

9.4.1.1 Dataset Construction

To evaluate the proposed *InfluencerRank*, we use the Instagram influencer dataset [92]. The dataset includes profiles of influencers, and their posts, including both images and all metadata. We only keep the posts that were published in the range of January 1st, 2017 and December 31st, 2017, to build temporal influencer networks. As a result, the dataset consists of 18,397 influencers and 2,952,075 posts. For the experiments, we split the dataset into the training dataset, which contains posts from January to November, and the testing dataset that contains posts published in December.

9.4.1.2 Heterogeneous Network Construction

To build the temporal heterogeneous networks, we first divide the whole dataset into 12 subsets by one-month intervals. Note that we conduct experiments to analyze ranking performances across different temporal window sizes in Section 9.5.1, thereby having the proper time intervals. We then extract all hashtags and user tags from the post captions and detect objects from the images. As a consequence, 1,151,082 unique hashtag nodes, 532,468 other user nodes, and 1,000 image object nodes are found across the networks and connected to the corresponding influencer nodes. To further reduce noises in the dataset, we remove every auxiliary node (i.e., hashtags, other users, and image objects) with only a single edge while edges with normalized frequencies less than 0.01 are also discarded. After the pruning process, 18,397 influencers, 20,744 other users, 67,695 hashtags, and 996 image objects are in the networks (i.e., 107,832 nodes), and a total of 15,090,225 edges remain across the networks.

Table 9.3: Statistics of influencers in the dataset across different relevance levels and criteria for the engagement rates.

Relevance	Engagement rate $E(\cdot)$	Number of Influencers
5	$E(\cdot) \geq 0.10$	1,274 (6.92%)
4	$0.10 > E(\cdot) \geq 0.07$	1,678 (9.12%)
3	$0.07 > E(\cdot) \geq 0.05$	2,321 (12.62%)
2	$0.05 > E(\cdot) \geq 0.03$	4,509 (24.51%)
1	$0.03 > E(\cdot) \geq 0.01$	6,882 (37.41%)
0	$0.01 > E(\cdot)$	1,734 (9.42%)

9.4.2 Experimental Settings

9.4.2.1 Evaluation Metrics.

Based on the definition in Section 9.2, we first compute the engagement rates for all influencers in across all timesteps as the ground truths. Note that the average engagement rate is 0.038 and the median engagement rate is 0.029. We utilize two metrics to evaluate the performance of ranking influencers.

- *Normalized Discounted Cumulative Gain* (NDCG) [80]: First, we divide all of the influencers into six groups with different thresholds on the engagement rates and relevance levels from 0 to 5. Table 9.3 further shows the statistics of influencers in the dataset across different relevance levels and criteria for the engagement rates. We then treat the relevance levels as ground truths to evaluate the ranking performance with the metric of NDCG.

- *Rank-Biased Precision* (RBP) [116]: To avoid losing valuable information while converting the engagement rates to the six relevance levels, we directly use the engagement rates with the metric of RBP. We set the probability p as 0.95 to measure rank quality.

9.4.2.2 Implementation Details

For the hyperparameter tuning, we use a validation set which contains posts published by the 18,397 influencers in January 2018. Since our model is optimized with the validation set, we can avoid potential information leakage from the testing set. After tuning the model, we set the numbers of dimensions of the graph embeddings and GCN features as 128, and the number of GCN layers as 2. Each batch contains 1,024 lists of influencers, and each list includes 10 randomly selected influencers for list-wise learning. The learning rate and the dropout probability are set as 0.001 and 0.5, respectively.

9.4.2.3 Baseline Methods

We compare the performance of *InfluencerRank* with nine baseline methods in three different categories, including *User*, *Ranking*, and *Graph*.

User Baselines. The baseline methods in this category exploit information on social media to measure the popularity of users with certain features. Since user popularity is often to be considered as an important factor in influencer hiring [44, 26, 105], we develop three methods from previous studies as the baseline approaches in this category as follows:

- *User Popularity* (UP) [10] estimates the popularity of users by applying ordinary linear regression (OLS) with a set of hand-crafted user-seed features, the number of followers, the number of friends, the number of posts.

- *Post Popularity* (PP) [111] predicts the popularity of users based on their social media posts. More specifically, an L2 regularized loss support vector machine (SVM) tackles the task with several post–related features, including brands, user mentions, sentiment scores, and image aesthetics and concepts.

- *User Activity* (UA) [101] applies features in three aspects, including network-based features, post-related features, and activity-based features. For the network features, UA exploits the PageRank scores [126] while the post-related features include the statistics of user mentions. The activity-based features describe the user activities like the comments in the posts of other users.

Ranking Baselines. We deploy two ranking baseline methods that exploit the same features described in Section 9.3.5 without knowledge of graphical structure.

- *ListNet* (LN) [23] is a list-wise learning-to-rank algorithm, using neural networks with gradient descent to model appropriate permutation probabilities to simultaneously optimize the overall ranking performance.

- *LambdaMART* (LM) [21] adopts the technique of the lambda function to optimize the specific evaluation metrics. Based on MART, LamdaMART can directly optimize the NDCG scores as a list-wise learning-to-rank algorithm.

Graph Baselines. The baseline methods in this category implement graph neural networks (GNNs) based learning models. Note that the applied features of *InfluencerRank* and graph baselines are identical so that we can fairly evaluate the model novelty and capability for the ranking task.

- *Graph Convolutional Recurrent Network* (*GCRN*) [150] extracts features by stacking graph CNNs and then applies LSTM to learn sequence of structures.

- *DeepInf* [134] generates a node representation by utilizing convolutional neural and attention networks. A fixed-size sub-network of each node is extracted by using random walk and node features are considered to incorporate both network structure and user-specific features.

- *Recurrent Cascades Convolutional Networks* (*CasCN*) [32] predicts the size of information cascades in networks. CasCN samples sub-cascade graphs to learn the local structures by using CGNs, and then applies RNNs to learn the dynamics of the cascade graphs.

Table 9.4: *RBP*, *NDCG@K* scores, and training time of the proposed InfluencerRank and the nine baseline methods.

Method	*RBP*	*NDCG@K*					Time
		1	10	50	100	200	(sec)
UP [10]	0.025	0.800	0.436	0.413	0.406	0.368	347
PP [111]	0.028	1.000	0.519	0.465	0.442	0.425	**295**
UA [101]	0.024	0.800	0.518	0.494	0.438	0.436	330
LN [23]	0.026	1.000	0.610	0.511	0.465	0.441	481
LM [21]	0.031	1.000	0.648	0.546	0.493	0.477	563
GCRN [150]	0.028	1.000	0.629	0.557	0.513	0.467	612
DeepInf [134]	0.031	1.000	0.697	0.567	0.549	0.512	525
CasCN [32]	0.033	1.000	0.751	0.645	0.572	0.543	1109
EGCN [127]	0.038	1.000	0.812	0.679	0.616	0.577	1483
InfluencerRank	**0.043**	**1.000**	**0.864**	**0.720**	**0.661**	**0.614**	648

- *Evolving Graph Convolutional Networks* (*EGCN*) [127] generates node representations in evoloving networks. EGCN exploits RNNs to learn graph dynamics and uses it as the parameters of the GCNs thereby capturing evolving sequence.

9.4.3 Experimental Results

Table 9.4 shows RBP, NDCG scores, and training time of InfluencerRank and the nine baseline methods for discovering influencers with high engagement rates. All of the three methods in the user baselines, which exploit the social media features, obtain low ranking results. This suggests that only considering social media features is insufficient to discover effective influencers. The ranking baseline methods, on the other hand, show better ranking performance compared to the user baseline methods since they use our proposed features. It demonstrates that our proposed features are very useful to capture the characteristics of influencers, thereby discovering effective influencers. Next, most of the graph baseline methods outperform the user baselines and ranking baselines. More specifically, among the graph baseline methods, GCRN [150] shows limited ranking performance improvement since it only resorts to temporal-spatial structures of graphs without taking into account the node features. DeepInf [134] demonstrates better ranking performance than GCRN by exploiting the graph convolutional networks that take advantage of the network structures of different entities while the features in different aspects provide sufficient knowledge to describe both influencers and other entities in the graph. Both CasCN [32] and EGCN [127] further improve performance by applying recurrent neural networks to adjacency matrices of temporal graphs. This suggests that the learning dynamics of graph structures with node features over time is beneficial in discovering effective influencers.

Finally, our proposed approach, InfluencerRank, outperforms all of the baseline methods. This is because our model derives informative influencer representations over time by using the graph convolutional networks and the attentive neural network, and effectively learns the dynamics of influencer characteristics and engagement rates. The results also show that InfluencerRank is able to learn the latent influencer representations in a reasonable amount of training time compared to the other baseline methods. CasCN and EGCN, on the other hand, have significantly longer training time than InfluencerRank. This is probably because the proposed framework successfully learns the importance of hidden states in the RNNs by applying attention while other baseline methods combine RNNs with GCNs without taking the importance of each temporal graph into account.

9.5 Analysis and Discussions

In this section, we conduct six analyses to understand the importance of (i) the temporal window size, (ii) the temporal information, (iii) the model components, (iv) the type of RNNs, (v) the heterogeneous networks, and (vi) the input features. We then evaluate the performance of InfluencerRank on various sets of influencers which are grouped by the size of audiences.

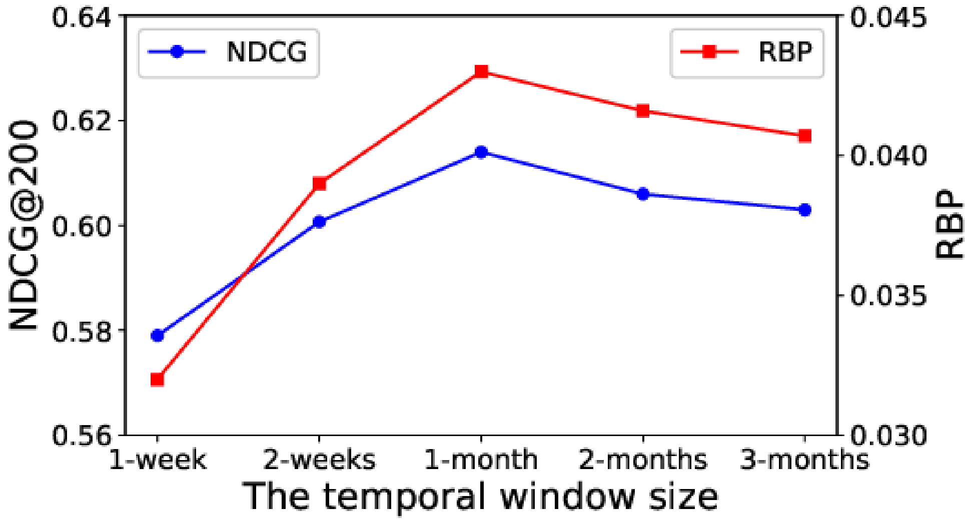


Figure 9.3: Ranking performance over different temporal window sizes. InfluencerRank trained with the 1-month window size shows the best performance.

9.5.1 Analysis on Temporal Window Size

We first investigate the effect of different temporal window sizes for heterogeneous network construction. To that end, we split the training dataset which has posts in 11 months period into sub-datasets by five different temporal window sizes including 1 week, 2 weeks, 1 month, 2 months, and 3 months. Note that we use the same testing dataset across the five different window sizes for consistent performance comparison. The RBP and NDCG@200 scores of the InfluencerRank over the different temporal window sizes are shown in Figure 9.3. We find that the model trained with the networks divided by 1-month intervals shows the best ranking performance whereas the model trained with the 1-week temporal window has the lowest ranking scores. InfluencerRank loses 5.7% performance on NDCG when the model is trained with the 1-week window size compared to the model trained with the 1-month window size. This suggests that the heterogeneous networks of the models, which are trained with temporal window size shorter than 1-month, have insufficient information to learn the dynamics of engagement rates. We also observe that the ranking performance gradually decreases while we use the longer temporal window size. This implies that the model trained with a large window size fails to take into account the variance of engagements by using the

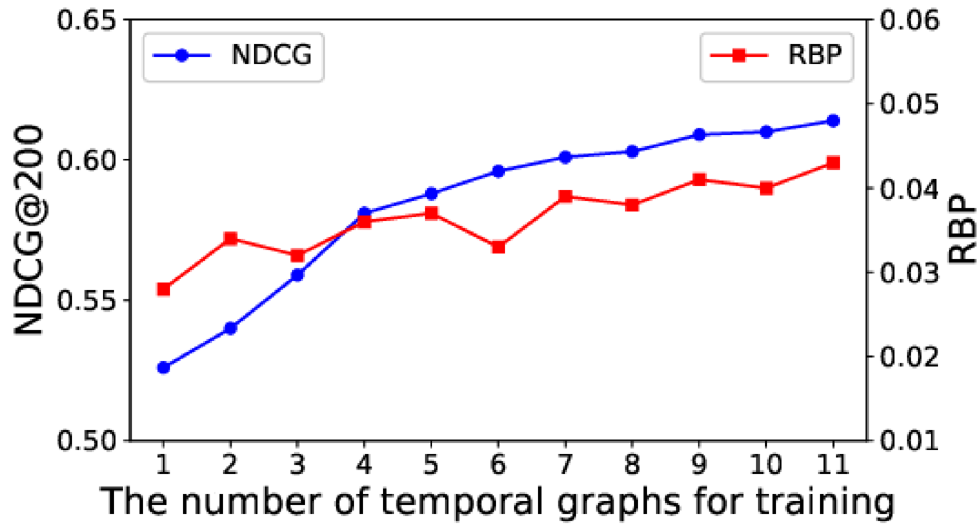


Figure 9.4: Ranking performance on different lengths of timestamps. InfluencerRank achieves higher ranking scores with longer history.

average like counts of all posts in each sub-dataset. The analysis results also demonstrate that learning the temporal dynamics of the engagement rates is very important to find effective influencers.

9.5.2 Analysis on Temporal Information

We next evaluate the ranking performance of the proposed model by using the different number of temporal input networks for training the model. Figure 9.4 shows RBP and NDCG@200 scores of the InfluencerRank over the number of temporal graphs. Note that the model uses the most recent temporal graphs. For example, a model trained with two temporal graphs learns two networks in October and November in our dataset. We observe that the performance significantly drops when InfluencerRank obtains insufficient historical information. InfluencerRank loses 15% performance on NDCG if the model uses only one graph compared to the model that considers all temporal graphs. The result confirms that only considering the most recent network degrades the performance since the engagement rates of influencers vary over time. We also find that as the number of temporal networks increases, the model has gradually less performance gain.

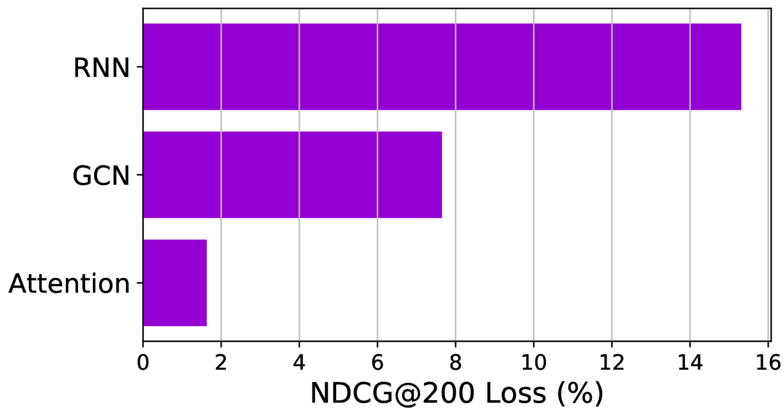


Figure 9.5: Performance losses after removing each of the components. The RNNs are the most important component to discover effective influencers.

9.5.3 Analysis on Model Components

The proposed model consists of three major components, including the graph convolutional networks, the recurrent neural networks, and the attention network. We conduct an ablation study by excluding each component from the model framework to understand the importance of model components on discovering influencers with high engagement rates. Figure 9.5 shows the performance losses of NDCG@200 scores over the three model components. We find that the model which excludes the RNN component has significant performance loss compared to the full model. This suggests that disregarding to learn sequential temporal information leads to performance degradation since engagement rates of an influencer change over time. The model that discards the GCN component also shows large performance loss. This is because the model fails to learn structural information with embedded node features. This demonstrates that learning social relationships of influencers with other users, tags, and image objects plays an important role in discovering effective influencers. We observe that the attention component has relatively less impact on the performance than other model components whilst it still enhances the model by considering the importance of temporal graph embeddings.

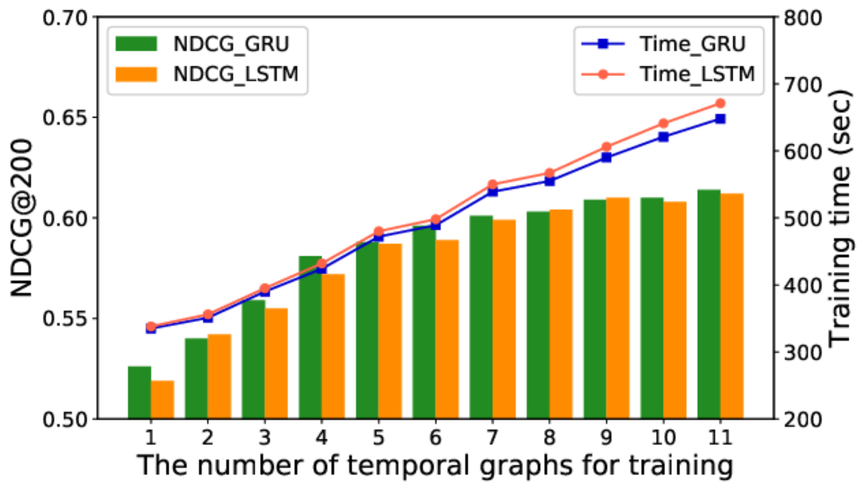


Figure 9.6: NDCG scores and training times of InfluencerRank trained with GRU and LSTM on different number of temporal graphs. InfluencerRank with GRU tends to have better ranking performance and shorter training time than the model with LSTM.

9.5.4 Analysis on Recurrent Neural Networks

In the proposed InfluencerRank framework, we employ gated recurrent units (GRUs) [36] for the recurrent neural networks. However, the GRU can be replaced with a long short-term memory (LSTM) [72]. To make a design decision which recurrent architecture to employ, we train InfluencerRank with GRU and LSTM. Figure 9.6 shows NDCG@200 scores and training times of InflurncerRank with two RNN architectures on different number of temporal graphs. We observe that no significant difference in the NDCG scores of models using GRU and LSTM, but the model with GRU tends to have slightly higher scores. The results also show that InfluencerRank with GRU has shorter training times than LSTM across the different number of temporal graphs. More specifically, the time difference gradually increases as the number of temporal graphs for training increases. Note that GRU is 1% faster than LSTM when the model only takes one temporal graph and 3.4% faster when the model uses 11 graphs for training. GRU shows better performance than LSTM in our task and that is probably because GRU has simpler network than LSTM and also benefits from the short input sequence length.

Table 9.5: Rank evaluation on the different network structures. Note that $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{H}$, and $\boldsymbol{O}$ represent influencer, other mentioned user, hashtag, and image object nodes in the network, respectively.

Node Type	*RBP*	*NDCG@K*				
		1	10	50	100	200
All nodes	0.043	1.000	0.864	0.720	0.661	0.614
$\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{H}$	0.039	1.000	0.848	0.704	0.648	0.589
$\boldsymbol{U}$, $\boldsymbol{H}$, $\boldsymbol{O}$	0.040	1.000	0.863	0.716	0.654	0.602
$\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{O}$	0.037	1.000	0.861	0.708	0.657	0.597

9.5.5 Analysis on Heterogeneous Networks

We study the importance of the proposed heterogeneous network to find effective influencers. To understand the importance of individual auxiliary node type, we train InfluencerRank with the network without the type of auxiliary node. Table 9.5 shows the RBP and the NDGC scores of InfluencerRank with different types of networks. The results show that the model trained with all types of nodes achieve higher RBP and NDCG scores than the other models that exclude a type of auxiliary nodes. This confirms that the graphical structure in InfluencerRank helps improve performance in finding effective influencers. We also observe that NDCG scores of the model without the image object nodes are lower than that of the model excluding hashtags and other user nodes. Note that excluding the image object nodes drops the performance of NDCG@200 by 4.1%, whereas excluding hashtag and other user nodes only drops the score by 2.8% and 1.9%. This is probably because each image object node can densely connect a large number of similar influencers together as it has a greater number of edges than a hashtag node and a user node.

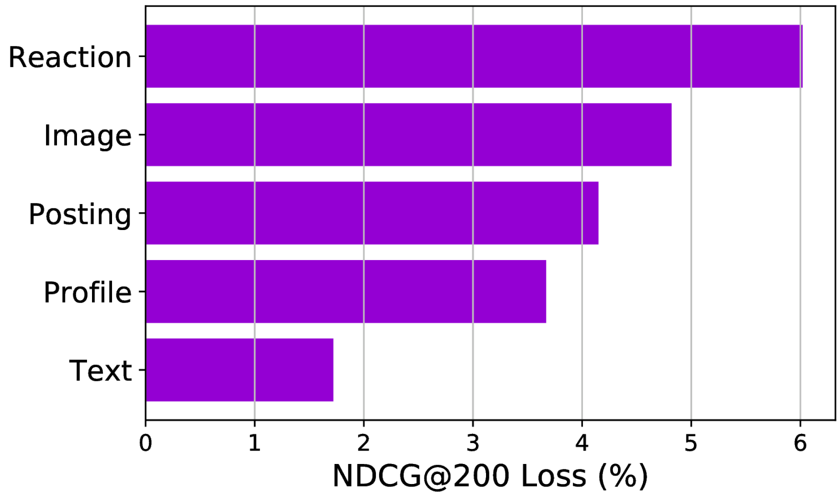


Figure 9.7: Performance losses of NDCG@200 over node feature categories. The image features which represent visual perception and the reaction features which include sentiment scores of user comments have more impact on the effective influencer discovery than other types of features.

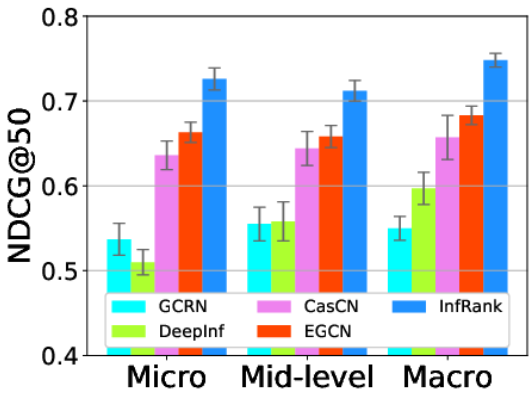
9.5.6 Analysis on Node Features

The benefit of using GCNs comes from considering network structure information with node features. To understand the importance of node features, we first evaluate the performance of the model that excludes all node categories. The model without the whole node features significantly drops the ranking quality; the loss of NDGC@200 of the model without node features is 21.99%.

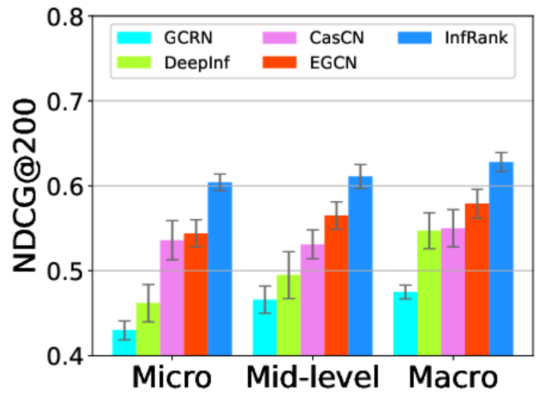
We then investigate the performance of InfluencerRank with variant sets of node features to study the importance of each category of the node features defined in Section 9.3.5. Figure 9.7 shows the performance loss of NDCG scores of the models trained with the node features excluding one particular node category against the full model as the leave-one-out analysis. The results reveal that the reaction feature category, which contains the sentiment scores of user comments on the posts, is more important than other categories to identify effective influencers. This indicates that the audience may show distinct reactions to influencers with high engagement rates. The image category, which includes the visual perception of images (e.g., brightness, colorfulness), also has higher loss values than other node feature categories. This suggests that influencers with high engagement rates may have different visual characteristics from other influencers. We further examine that the Pearson correlation scores between the effectiveness and the image features, including brightness, colorfulness, and color temperature of the influencers' images, are -0.16, -0.10, and 0.08, respectively. To understand the correlation between the image features and the effectiveness, we examine the brightness, colorfulness, and color temperature of the influencers' images. The Pearson correlation coefficient values are -0.16, -0.10, and 0.08 for the brightness, colorfulness, and color temperature, respectively. This reveals that effective influencers tend to post less bright, less colorful, and warmer images than other influencers. We also find that effective influencers have lower standard deviation values for all image features than other influencers. This suggests that the similarity of the visual characteristics of posted images can affect the effectiveness of influencers. For example, constantly posting images that have similar characteristics may help to become an effective influencer, whereas posting images that have different attributes from previously posted images may fail to attract audiences' attention. On the other hand, the text category, including the number of hashtags, user tags, emojis in a caption, and the sentiment scores of the caption, have the least impact to discover effective influencers. Although the statistical features to represent textual characteristics of influencers' posts have less impact than other features, InfluencerRank can improve the ranking performance by taking hashtags and user tags into account to the network structure.

9.5.7 Influencer Follower Size

In the influencer marketing industry, influencers are often divided into subgroups by the number of followers since it directly refers to the size of potential customers and hiring cost [44]. For example, companies with a sufficient marketing budget can hire influencers who are followed by millions of people while small retailers may collaborate with influencers with a small number of followers. Therefore, we evaluate the performance of InfluencerRank over groups of influencers with different sizes of followers. Although there are no standard



(a) NDCG@50



(b) NDCG@200

Figure 9.8: Performance evaluation on influencers with different sizes of followers. *InfluencerRank* shows consistently good performance regardless of the audience size of influencers.

criteria to classify influencers based on the number of followers, we utilize the following thresholds which are the generally accepted numbers to divide the influencers into three groups[1]. Influencers who are followed by less than 20,000 followers are classified as the *Micro influencers*. The *Mid-level influencers* have followers between 20,000 and 100,000, and *Macro influencers* have more than 100,000 followers. In our dataset, around 30% of influencers are the micro-influencers, 45% of them are the mid-level influencers, and the remaining 25% influencers are the macro-influencers. To evaluate the performance under the same conditions, we randomly select multiple sets of 1,000 influencers from each category and run the experiment 10 times.

Figure 9.8 shows the average NDCG scores of *InfluencerRank* and four baseline methods, including GCRN [150], DeepInf [134], CasCN [32], and EGCN [127] over the micro, mid-level, and macro-influencers. The results show that the proposed model has robust performance to discover effective influencers in the groups of all ranges of followers compared to the baseline methods. More specifically, DeepInf [134] fails to discover effective micro-influencers. This is probably because DeepInf disregards the temporal information which is critical to find micro-influencers who have relatively large variance on their features and engagement rates

[1] `http://www.mattr.co/pros-cons-micro-macro-mid-level-influencers/`

over time compared to macro-influencers who are matured. On the other hand, our proposed model can accurately find highly effective micro-influencers since their unique features are captured by sequential learning of temporal information.

CHAPTER 10

Conclusion

In this dissertation, we propose new influencer evaluation metrics and develop multi-modal graph learning frameworks to model and discover high-quality influencers on social media. More specifically, we first collect influencer data from Instagram and construct two large-scale influencer datasets. We then build three types of influencer social networks, including the influencer network, the brand mentioning network, and the audience engagement network, and conduct analytical studies to understand distinct features of high-quality influencers. Finally, we develop deep learning frameworks that take multi-modal inputs and the influencer social networks to learn unique representations of influencers. The contributions of this dissertation in modeling and discovering influencers with multi-modal data inputs can be further summarized as follows:

- We release two datasets to foster research in both computer science and marketing fields. To help categorize influencers, we develop InfluencerProfiler that applies post-level attention to automatically label influencers and their posts into one of eight topics. The first dataset, the Instagram Influencer Dataset (Chapter 6), contains over 10 million posts published by 33,935 influencers who have been classified into their topics. The second dataset, the Influencer and Brand Dataset, contains 1.6 million influencer posts that mention one of 26,910 brands on Instagram. The dataset can be specifically utilized to understand the advertising behaviors of influencers.
- We find unique behaviors of influencers by conducting in-depth analysis studies on the influencer network (Chapter 3), the brand mentioning network (Chapter 4), and the audience engagement network (Chapter 5). Our findings reveal that influencers with

similar interests have strong reciprocal relations and share many common followers, influencers put more effort to make sponsored advertising posts than non-sponsored posts, and social bots have distinct engaging behaviors and social relations with influencers.

- We investigate hidden sponsorship in advertising posts and present a learning framework that takes texts, images, and graphs to detect undisclosed paid partnerships with brands. SPoD (Chapter 7) applies aspect attention over multi-modal inputs and optimize with temporal regularization to generate unique representations of social media posts. The experimental results suggest that SPoD can effectively detect hidden sponsorships on social media and both aspect attention and temporal regularization improve the performance.

- We present an audience evaluation framework, ALAIM (Chapter 8), to consider the loyalty and authenticity of audiences in assessing influencers. ALAIM takes multi relations as input networks to capture the knowledge from interactions between relations. The results demonstrate that our proposed method outperforms conventional methods by finding influencers followed by high loyal audiences, and influencers related to engagement bots. ALAIM can be utilized in any social media platform to evaluate various aspects of audiences.

- We develop InfluencerRank (Chapter 9) that learns posting behaviors and social relations of influencers over time to predict engagement rates in the future. InfluencerRank applies attentive recurrent graph neural networks to understand the dynamics of user engagements depending on the posting characteristics. The extensive experiments and analysis indicate that attention over graphs across time significantly enhances the prediction performance and outperforms state-of-the-art approaches in discovering effective influencers.

REFERENCES

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] Activate. Exploring the brand and influencer relationship in influencer marketing. *State of Influencer Marketing Study*, 2018.

[3] R. Ahmad and F. Buttle. Customer retention: a potentially potent marketing management strategy. *Journal of strategic marketing*, 9(1):29–45, 2001.

[4] F. C. Akyon and M. E. Kalfaoglu. Instagram fake and automated account detection. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–7. IEEE, 2019.

[5] A. Anand, S. Dutta, and P. Mukherjee. Influencer marketing with fake followers. *IIM Bangalore Research Paper*, (580), 2019.

[6] F. J. Arenas-Marquez, M. R. Martínez-Torres, and S. Toral. Electronic word-of-mouth communities from the perspective of social network analysis. *Technology Analysis & Strategic Management*, 26(8):927–942, 2014.

[7] A. S. Authority. The labelling of influencer advertising. `https://www.asa.org.uk/uploads/assets/uploaded/e3158f76-ccf2-4e6e-8f51a710b3237c43.pdf`, 2019.

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

[9] S. Bakhshi, D. A. Shamma, and E. Gilbert. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 965–974, 2014.

[10] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74, 2011.

[11] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528, 2012.

[12] B. K. Balakrishnan, M. I. Dahnil, and W. J. Yi. The impact of social media marketing medium toward purchase intention and brand loyalty among generation y. *Procedia-Social and Behavioral Sciences*, 148:177–185, 2014.

[13] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[14] I. Beltagy, M. E. Peters, and A. Cohan. Longformer: The long-document transformer. *arXiv:2004.05150*, 2020.

[15] M. C. Benigni, K. Joseph, and K. M. Carley. Bot-ivistm: Assessing information manipulation in social media using network analytics. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, pages 19–42. Springer, 2019.

[16] M. Bergel and C. Brock. Visitors' loyalty and price perceptions: the role of customer engagement. *The Service Industries Journal*, 39(7-8):575–589, 2019.

[17] R. Bhatt, V. Chaoji, and R. Parekh. Predicting product adoption in large-scale social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1039–1048, 2010.

[18] C. Brezinski and M. Redivo-Zaglia. The pagerank vector: properties, computation, approximation, and acceleration. *SIAM Journal on Matrix Analysis and Applications*, 28(2):551–575, 2006.

[19] R. J. Brodie, A. Ilic, B. Juric, and L. Hollebeek. Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of business research*, 66(1):105–114, 2013.

[20] J. Bughin. Gettig a sharper picture of social media's influence. *McKinsey Quartely, July*, 2015.

[21] C. J. Burges, R. Ragno, and Q. V. Le. Learning to rank with nonsmooth cost functions. In *Advances in neural information processing systems (NIPS)*, pages 193–200, 2007.

[22] C. Campbell and J. R. Farrell. More than meets the eye: The functional components underlying influencer marketing. *Business Horizons*, 2020.

[23] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning (ICML)*, pages 129–136. ACM, 2007.

[24] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *2010 IEEE Computer society conference on computer vision and pattern recognition*, pages 2707–2714. IEEE, 2010.

[25] R. Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.

[26] L. V. Casaló, C. Flavián, and S. Ibáñez-Sánchez. Influencers on instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*, 2018.

[27] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, 2010.

[28] N. Chavoshi, H. Hamooni, and A. Mueen. Debot: Twitter bot detection via warped correlation. In *Icdm*, pages 817–822, 2016.

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[30] F. Chen, R. Ji, J. Su, D. Cao, and Y. Gao. Predicting microblog sentiments via weakly supervised multimodal deep learning. *IEEE Transactions on Multimedia*, 20(4):997–1007, 2017.

[31] W. Chen and S.-H. Teng. Interplay between social influence and network centrality: a comparative study on shapley centrality and single-node-influence centrality. In *Proceedings of the 26th international conference on world wide web*, pages 967–976, 2017.

[32] X. Chen, F. Zhou, K. Zhang, G. Trajcevski, T. Zhong, and F. Zhang. Information diffusion prediction via recurrent cascades convolution. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 770–781. IEEE, 2019.

[33] H. J. Cheong and M. A. Morrison. Consumers' reliance on product information and recommendations found in ugc. *Journal of Interactive Advertising*, 8(2):38–49, 2008.

[34] C. C. Childers, L. L. Lemon, and M. G. Hoy. # sponsored# ad: Agency perspective on influencer marketing campaigns. *Journal of Current Issues & Research in Advertising*, 40(3):258–274, 2019.

[35] C.-H. Cho. Why do people avoid advertising on the internet? *Journal of advertising*, 33(4):89–97, 2004.

[36] 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.

[37] A. E. Coates, C. A. Hardman, J. C. Halford, P. Christiansen, and E. J. Boyland. Social media influencer marketing and children's food intake: a randomized trial. *Pediatrics*, 143(4), 2019.

[38] R. Collobert and J. Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167, 2008.

[39] W. Comcowich. Macro-engagement of micro-influencers yields marketing dividends. *GleanInfo*, 2018.

[40] F. T. Commission. The ftc's endorsement guides: What people are asking, 2017.

[41] N. Confessore, G. J. Dance, R. Harris, and M. Hansen. The follower factory. *The New York Times*, 27, 2018.

[42] G. Contestabile. Influencer marketing in 2018: Becoming an efficient marketplace. *Adweek*, 2018.

[43] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, 80:56–71, 2015.

[44] M. De Veirman, V. Cauberghe, and L. Hudders. Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, 36(5):798–828, 2017.

[45] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[46] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[47] U. M. Dholakia and E. Durham. One café chain's facebook experiment. *Harvard Business Review*, 88(3):26, 2010.

[48] İ. E. Erdoğmuş and M. Cicek. The impact of social media marketing on brand loyalty. *Procedia-Social and Behavioral Sciences*, 58:1353–1360, 2012.

[49] N. J. Evans, M. G. Hoy, and C. C. Childers. Parenting "youtube natives": The impact of pre-roll advertising and text disclosures on parental responses to sponsored child influencer videos. *Journal of Advertising*, 47(4):326–346, 2018.

[50] N. J. Evans, J. Phua, J. Lim, and H. Jun. Disclosing instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. *Journal of interactive advertising*, 17(2):138–149, 2017.

[51] G. Farnadi, J. Tang, M. De Cock, and M.-F. Moens. User profiling through deep multimodal fusion. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 171–179. ACM, 2018.

[52] S. Feng, G. Cong, A. Khan, X. Li, Y. Liu, and Y. M. Chee. Inf2vec: Latent representation model for social influence embedding. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 941–952. IEEE, 2018.

[53] R. Ferguson. Word of mouth and viral marketing: taking the temperature of the hottest trends in marketing. *Journal of consumer marketing*, 25(3), 2008.

[54] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.

[55] T. O. for Economic Co-operation and Development. Good practice guide on online advertising. `http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CP(2018)16/FINAL&docLanguage=En`, 2019.

[56] Forrester. Number of brand sponsored influencer posts on instagram from 2016 to 2019. `https://www.statista.com/statistics/693775`, 2017.

[57] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[58] J. Ge and U. Gretzel. Emoji rhetoric: a social media influencer perspective. *Journal of Marketing Management*, 34(15-16):1272–1295, 2018.

[59] F. Gelli, T. Uricchio, M. Bertini, A. Del Bimbo, and S.-F. Chang. Image popularity prediction in social media using sentiment and context features. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 907–910, 2015.

[60] C. Gilbert and E. Hutto. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, volume 81, page 82, 2014.

[61] P. Gillin and G. A. Moore. *The new influencers: A marketer's guide to the new social media*. Linden Publishing, 2009.

[62] J. Goldenberg, S. Han, D. R. Lehmann, and J. W. Hong. The role of hubs in the adoption process. *Journal of Marketing*, 73(2), 2009.

[63] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[64] M. B. Goodman, N. Booth, and J. A. Matic. Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal*, 16(3), 2011.

[65] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.

[66] S. Guo, M. Wang, and J. Leskovec. The role of social networks in online shopping: information passing, price of trust, and consumer choice. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 157–166, 2011.

[67] A. Ha. Instagram marketing startup popular pays raises $2 million. *TechCrunch*, 2015.

[68] J. Han, D. Choi, B.-G. Chun, T. Kwon, H.-c. Kim, and Y. Choi. Collecting, organizing, and sharing pins in pinterest: interest-driven or social-driven? *ACM SIGMETRICS Performance Evaluation Review*, 42(1):15–27, 2014.

[69] D. Hasler and S. E. Suesstrunk. Measuring colorfulness in natural images. In *Human vision and electronic imaging VIII*, volume 5007, pages 87–95. International Society for Optics and Photonics, 2003.

[70] K. Hawkins and P. Vel. Attitudinal loyalty, behavioural loyalty and social media: An introspection. *The Marketing Review*, 13(2):125–141, 2013.

[71] J. Hessel, L. Lee, and D. Mimno. Cats and captions vs. creators and the clock: Comparing multimodal content to context in predicting relative popularity. In *Proceedings of the 26th International Conference on World Wide Web*, pages 927–936, 2017.

[72] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[73] D. L. Hoffman and M. Fodor. Can you measure the roi of your social media marketing? *MIT Sloan Management Review*, 52(1):41, 2010.

[74] D. Hu, X. Li, et al. Temporal multimodal learning in audiovisual speech recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3574–3582, 2016.

[75] C.-C. Hung, Y.-C. Huang, J. Y.-j. Hsu, and D. K.-C. Wu. Tag-based user profiling for social media recommendation. In *Workshop on Intelligent Techniques for Web Personalization & Recommender Systems at AAAI*, volume 8, pages 49–55, 2008.

[76] Instagram. 2m monthly advertisers on instagram. *Instagram blog*, 2017.

[77] Instagram. Reducing inauthentic activity on instagram. `https://about.instagram.com/blog/announcements/reducing-inauthentic-activity-on-instagram`, 2018. Accessed: 2018-11-19.

[78] Instagram. Giving people more control on instagram and facebook. *Instagram blog*, 2021.

[79] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *SIGIR*. ACM, 2000.

[80] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *ACM SIGIR Forum*, volume 51, pages 243–250. ACM New York, NY, USA, 2017.

[81] J.-Y. Jiang, P.-J. Cheng, and W. Wang. Open source repository recommendation in social coding. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017.

[82] S. V. Jin, A. Muqaddam, and E. Ryu. Instafamous and social media influencer marketing. *Marketing Intelligence & Planning*, 2019.

[83] S. Jun and J. Yi. What makes followers loyal? the role of influencer interactivity in building influencer brand equity. *Journal of Product & Brand Management*, 2020.

[84] J. Kang, D. Choi, E. Park, and J. Han. Understanding influential comments in online conversations. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 197–200, 2018.

[85] L. Kelly, G. Kerr, and J. Drennan. Avoidance of advertising in social networking sites: The teenage perspective. *Journal of interactive advertising*, 10(2):16–27, 2010.

[86] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[87] S. Khamis, L. Ang, and R. Welling. Self-branding,'micro-celebrity'and the rise of social media influencers. *Celebrity studies*, 8(2):191–208, 2017.

[88] S. Kim, X. Chen, J.-Y. Jiang, J. Han, and W. Wang. Evaluating audience loyalty and authenticity in influencer marketing via multi-task multi-relational learning. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 278–289, 2021.

[89] S. Kim and J. Han. Detecting engagement bots on social influencer marketing. In *International Conference on Social Informatics*, pages 124–136. Springer, 2020.

[90] S. Kim, J. Han, S. Yoo, and M. Gerla. How are social influencers connected in instagram? In *International Conference on Social Informatics*, pages 257–264. Springer, 2017.

[91] S. Kim, T. Hori, and S. Watanabe. Joint ctc-attention based end-to-end speech recognition using multi-task learning. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4835–4839. IEEE, 2017.

[92] S. Kim, J.-Y. Jiang, M. Nakada, J. Han, and W. Wang. Multimodal post attentive profiling for influencer marketing. In *Proceedings of The Web Conference 2020*, pages 2878–2884, 2020.

[93] S. Kim, J.-Y. Jiang, and W. Wang. Discovering undisclosed paid partnership on social media via aspect-attentive sponsored post learning. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 319–327, 2021.

[94] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[95] S. Kudugunta and E. Ferrara. Deep neural networks for bot detection. *Information Sciences*, 467:312–322, 2018.

[96] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600, 2010.

[97] M. Larralde. Instalooter. `https://github.com/althonos/InstaLooter`, 2016.

[98] J. Lawley. Brands to increase spend on influencer marketing despite fraud concerns. *MarketingWeek*, 2018.

[99] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5–es, 2007.

[100] D. Lewis and D. Bridger. *Soul of the New Consumer: Authenticity-What We Buy and Why in the New Economy*. Nicholas Brealey, 2011.

[101] Y.-M. Li, C.-Y. Lai, and C.-W. Chen. Discovering influencers for marketing in the blogosphere. *Information Sciences*, 181(23):5143–5157, 2011.

[102] A. Lipsman, G. Mudd, M. Rich, and S. Bruich. The power of "like": How brands reach (and influence) fans through social-media marketing. *Journal of Advertising research*, 52(1):40–52, 2012.

[103] S. Liu, C. Jiang, Z. Lin, Y. Ding, R. Duan, and Z. Xu. Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach. *Information sciences*, 306:34–52, 2015.

[104] T. Loren. Instagram is hiding likes: Here's everything you need to know. https://later.com/blog/hidden-likes-instagram/, 2019.

[105] C. Lou and S. Yuan. Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, pages 1–16, 2019.

[106] P. Lu, L. Ji, W. Zhang, N. Duan, M. Zhou, and J. Wang. R-vqa: learning visual relation facts with semantic attention for visual question answering. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1880–1889. ACM, 2018.

[107] N. Ma, Y. Liu, and Y. Chi. Influencer discovery algorithm in a multi-relational network. *Physica A: Statistical Mechanics and its Applications*, 510:415–425, 2018.

[108] J. Marciano. The real economic losses from influencer fraud. *Marketing Technology insights*, 2019.

[109] MarketingHub. The state of influencer marketing 2019 : Benchmark report. *Benchmark Report*, 2019.

[110] F. J. Martínez-López, R. Anaya-Sánchez, M. Fernández Giordano, and D. Lopez-Lopez. Behind influencer marketing: key marketing decisions and their effects on followers' responses. *Journal of Marketing Management*, 36(7-8):579–607, 2020.

[111] M. Mazloom, R. Rietveld, S. Rudinac, M. Worring, and W. Van Dolen. Multimodal popularity prediction of brand-related social media posts. In *Proceedings of the 24th ACM international conference on Multimedia (MM)*, pages 197–201. ACM, 2016.

[112] M. L. McHugh. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282, 2012.

[113] Mediakix. 4 major influencer marketing platform risks marketers must know. *Mediakix*, 2018.

[114] Mention. Instagram reports 2018. `https://mention.com/en/reports/instagram/`, 2018.

[115] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *Proceedings of the first workshop on Online social networks*, pages 25–30, 2008.

[116] A. Moffat and J. Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Transactions on Information Systems (TOIS)*, 27(1):1–27, 2008.

[117] S. Moro, P. Rita, and B. Vala. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research*, 69(9), 2016.

[118] K. Morrison. Instagram dominates influencer marketing (report). *Adweek*, 2017.

[119] A. Nanji. The most important social network for influencers. *MarketingProfs*, 2017.

[120] S. Neti. Social media and its role in marketing. *International journal of enterprise computing and business systems*, 1(2):1–15, 2011.

[121] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. Multimodal deep learning. In *International conference on machine learning (ICML)*, 2011.

[122] A. Nielsen. Nielsen: Global consumers' trust in 'earned' advertising grows in importance. *Business Wire*, 2012.

[123] T. M. Nisar and C. Whitehead. Brand interactions and social media: Enhancing user loyalty through social networking sites. *Computers in Human Behavior*, 62:743–753, 2016.

[124] Y. X. Ong and N. Ito. "i want to go there too!" evaluating social media influencer marketing effectiveness: a case study of hokkaido's dmo. In *Information and communication technologies in tourism 2019*, pages 132–144. Springer, 2019.

[125] M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel. Detection of bots in social media: A systematic review. *Information Processing & Management*, 57(4):102250, 2020.

[126] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[127] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. B. Schardl, and C. E. Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *AAAI*, pages 5363–5370, 2020.

[128] Z. Peng, L. Wu, J. Ren, R. Zhang, and P. Luo. Cuimage: A neverending learning platform on a convolutional knowledge graph of billion web images. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1787–1796. IEEE, 2018.

[129] M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks afficionados: user classification in twitter. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 430–438, 2011.

[130] M. Pennacchiotti and A.-M. Popescu. A machine learning approach to twitter user classification. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.

[131] J. W. Pennebaker and L. A. King. Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296, 1999.

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

[133] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contextualized word representations. In *Proc. of NAACL*, 2018.

[134] J. Qiu, J. Tang, H. Ma, Y. Dong, K. Wang, and J. Tang. Deepinf: Social influence prediction with deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2110–2119, 2018.

[135] F. F. Reichheld. Loyalty and the renaissance of marketing. *Marketing management*, 2(4):10, 1994.

[136] H. Reinikainen, J. Munnukka, D. Maity, and V. Luoma-aho. 'you really are a great big sister'–parasocial relationships, credibility, and the moderating role of audience comments in influencer marketing. *Journal of Marketing Management*, 36(3-4):279–298, 2020.

[137] J. Ren, Y. Hu, Y.-W. Tai, C. Wang, L. Xu, W. Sun, and Q. Yan. Look, listen and learn—a multimodal lstm for speaker identification. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[138] S. A. Rios, F. Aguilera, J. D. Nuñez-Gonzalez, and M. Graña. Semantically enhanced network analysis for influencer identification in online social networks. *Neurocomputing*, 326:71–81, 2019.

[139] F. Riquelme and P. González-Cantergiani. Measuring user influence on twitter: A survey. *Information processing & management*, 52(5):949–975, 2016.

[140] S. Rodriguez. Instagram could delete up to 10 million accounts as it cracks down on spam. `https://www.ibtimes.com/instagram-could-delete-10-million-accounts-it-cracks-down-spam-1749914`, 2014. Accessed: 2014-12-11.

[141] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 18–33. Springer, 2011.

[142] Royal Society for Public Health. The impact of cyberbullying on social media on children and young people's mental health. www.rsph.org.uk/about-us/news/instagram-ranked-worst-for-young-people-s-mental-health.html, 2017.

[143] S. Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[144] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.

[145] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607. Springer, 2018.

[146] A. Schomer. Influencer marketing: State of the social media influencer market in 2020. https://www.businessinsider.com/influencer-marketing-report, 2019.

[147] R. Schuchard, A. Crooks, A. Stefanidis, and A. Croitoru. Bots in nets: empirical comparative analysis of bot evidence in social networks. In *International Conference on Complex Networks and their Applications*, pages 424–436. Springer, 2018.

[148] N. Segev, N. Avigdor, and E. Avigdor. Measuring influence on instagram: a network-oblivious approach. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1009–1012, 2018.

[149] I. Sen, A. Aggarwal, S. Mian, S. Singh, P. Kumaraguru, and A. Datta. Worth its weight in likes: Towards detecting fake likes on instagram. In *Proceedings of the 10th ACM Conference on Web Science*, pages 205–209, 2018.

[150] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *International Conference on Neural Information Processing*, pages 362–373. Springer, 2018.

[151] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844, 2017.

[152] A. Silva, S. Guimarães, W. Meira Jr, and M. Zaki. Profilerank: finding relevant content and influential users based on information diffusion. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, pages 1–9, 2013.

[153] Statista. Brand activation marketing spending in the united states in 2015, by type. `http://www.statista.com/statistics/650998/brand-activation-marketing-spending-usa/`, 2015.

[154] A. T. Stephen and J. Galak. The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of marketing research*, 49(5):624–639, 2012.

[155] C. Stubb and J. Colliander. "this is not sponsored content"–the effects of impartiality disclosure and e-commerce landing pages on consumer responses to social media influencer posts. *Computers in Human Behavior*, 2019.

[156] K. Subbian and P. Melville. Supervised rank aggregation for predicting influencers in twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 661–665. IEEE, 2011.

[157] P. Suciati, M. Maulidiyanti, and A. Lusia. Cultivation effect of tourism tv program and influencer's instagram account on the intention of traveling. In *Proceedings of the International Conference on Social Sciences (ICSS)*, volume 1, 2018.

[158] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[159] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077. International World Wide Web Conferences Steering Committee, 2015.

[160] L. Townsend and C. Wallace. Social media research: A guide to ethics. *University of Aberdeen*, 1:16, 2016.

[161] E. Uzunoğlu and S. M. Kip. Brand communication through digital influencers: Leveraging blogger engagement. *International Journal of Information Management*, 34(5):592–602, 2014.

[162] M. van Asperen, P. de Rooij, and C. Dijkmans. Engagement-based loyalty: The effects of social media engagement on customer loyalty in the travel industry. *International Journal of Hospitality & Tourism Administration*, 19(1):78–94, 2018.

[163] O. Varol, E. Ferrara, C. Davis, F. Menczer, and A. Flammini. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 2017.

[164] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[165] R. Vijayan and G. Mohler. Forecasting retweet count during elections using graph convolution neural networks. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 256–262. IEEE, 2018.

[166] H. M. Wallach. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*, pages 977–984, 2006.

[167] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao. You are how you click: Clickstream analysis for sybil detection. In *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, pages 241–256, 2013.

[168] H. Wang, F. Zhang, M. Hou, X. Xie, M. Guo, and Q. Liu. Shine: Signed heterogeneous information network embedding for sentiment link prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 592–600, 2018.

[169] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, pages 2022–2032, 2019.

[170] J. Warren. The ultimate guide to instagram influencer marketing. `https://later.com/blog/instagram-influencer-marketing/`, 2020. Accessed: 2020-11-08.

[171] B. W. Wojdynski, N. J. Evans, and M. G. Hoy. Measuring sponsorship transparency in the age of native advertising. *Journal of Consumer Affairs*, 52(1):115–137, 2018.

[172] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li. Listwise approach to learning to rank: theory and algorithm. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*. ACM, 2008.

[173] M. Xiao, R. Wang, and S. Chan-Olmsted. Factors affecting youtube influencer marketing credibility: a heuristic-systematic model. *Journal of media business studies*, 15(3):188–213, 2018.

[174] C. Yang, Y. Xiao, Y. Zhang, Y. Sun, and J. Han. Heterogeneous network representation learning: A unified framework with survey and benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[175] K.-C. Yang, O. Varol, P.-M. Hui, and F. Menczer. Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1096–1103, 2020.

[176] X. Yang, S. Kim, and Y. Sun. How do influencers mention brands in social media? sponsorship prediction of instagram posts. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 101–104, 2019.

[177] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. 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. ACM, 2018.

[178] Q. You, S. Bhatia, and J. Luo. A picture tells a thousand words—about you! user interest profiling from user generated visual content. *Signal Processing*, 124:45–53, 2016.

[179] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim. Graph transformer networks. In *Advances in Neural Information Processing Systems*, pages 11983–11993, 2019.

[180] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 793–803, 2019.

[181] J. Zhang, J. Tang, J. Li, Y. Liu, and C. Xing. Who influenced you? predicting retweet via social influence locality. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(3):1–26, 2015.

[182] Z. Zhang, P. Luo, C. C. Loy, and X. Tang. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*, pages 94–108. Springer, 2014.