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

Seungbae Kim[1], Xiusi Chen[1], Jyun-Yu Jiang[1], Jinyoung Han[2] and Wei Wang[1]
[1] Department of Computer Science, University of California, Los Angeles
[2] Department of Applied Artificial Intelligence, Sungkyunkwan University
sbkim@cs.ucla.edu, xchen@cs.ucla.edu, jyunyu@cs.ucla.edu, jinyounghan@skku.edu, weiwang@cs.ucla.edu

Abstract

Since influencer marketing has become an essential marketing method, influencer fraud behavior such as buying fake followers and engagements to manipulate the popularity is under the spotlight. To address this issue, we propose a multi-task audience evaluation model that can assess both the loyalty and authenticity of influencers' audiences. More specifically, the proposed model takes engagement information of an influencer's audience, including likes and comments on social media posts, and predicts (i) the retention rate of the audience of the influencer and (ii) how the influencer is associated with fake audiences (or engagement bots). To learn the social interaction between influencers and their audiences, we build multi-relational networks based on the diverse engagement behavior such as commenting. Our model further utilizes the contextualized information captured in user comments to learn distinct engagement behavior of genuine and fake users. Based on the predicted loyalty and authenticity scores, we rank influencers to find those who are followed by loyal and authentic audiences. By using a large-scale Instagram influencer-audience dataset which contains 14,221 influencers, 9,290,895 audiences, and 65,848,717 engagements, we evaluate ranking performance, and show that the proposed framework outperforms other baseline methods.

Introduction

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 (De Veirman, Cauberghe, and Hudders 2017). Many brands have been increasingly sponsoring influencers in recent years to advertise their products or services (Yang, Kim, and Sun 2019; Kim, Jiang, and Wang 2021) since audiences tend to have more interactions and trust in influencers than brands (Cheong and Morrison 2008; Lou and Yuan 2019). 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 (Childers, Lemon, and Hoy 2019; Confessore et al. 2018; Hoffman and Fodor 2010). However, the number of followers or likes can be easily manipulated by influencers through buying fake followers and engagements (Anand, Dutta, and Mukherjee 2019; Kim and Han 2020). 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.

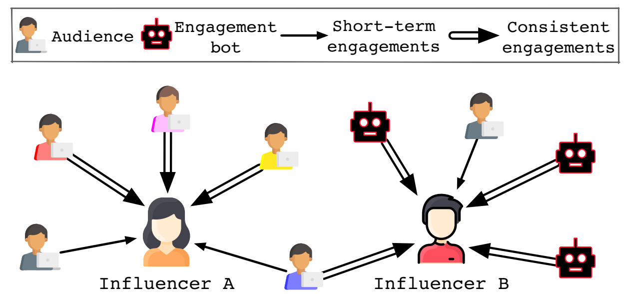


Figure 1: Illustration of engagement relationships for two influencers and their audiences. Influencer A has many loyal audiences who consistently make engagements, whereas Influencer B is connected to inauthentic audiences (engagement bots) who generate fake engagements.

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 (Reinikainen et al. 2020; Xiao, Wang, and Chan-Olmsted 2018; Martínez-López et al. 2020). 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 genuineness of the audience's engagements, respectively (Lewis and Bridger 2011). Figure 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 (Bergel and Brock 2019). Some previous studies focus on correlations between brand loyalty and other factors such as online advertisements (Balakrishnan, Dahnil, and Yi 2014), frequency of interactions (Neti 2011), and characteristics of published contents (Erdoğmuş and Cicek 2012). 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 (Ferrara et al. 2016; Cresci et al. 2015; Kim and Han 2020; Anand, Dutta, and Mukherjee 2019; Sen et al. 2018; Benigni, Joseph, and Carley 2019). Many previous studies analyze distinct behaviors of bots such as link farming (Sen et al. 2018; Chavoshi, Hamooni, and Mueen 2016; Benigni, Joseph, and Carley 2019) and propose methods to detect fake engagements or fake followers (Yang et al. 2020; Kudugunta and Ferrara 2018). 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.

We summarize our contributions as follows:

- To the best of our knowledge, this is the first learning-based framework for evaluating audiences in influencer marketing. We believe the proposed framework, ALAIM, is useful for brands and marketers who want to find influencers who are followed by loyal and authentic audiences.
- We propose a novel multi-relational framework that utilizes contextualized information. ALAIM uses user comments to form a long document thus captures contextualized knowledge that cannot be easily learned from each comment due to insufficient information. Additionally, the attention-based multi-relational GCNs can estimate the importance of each relation to model the engagement relationships between an influencer and their audiences.
- Since ALAIM uses multi-task learning, the output user representations can be generalized and extensible. That is, any potential audience evaluation metrics, e.g., activeness of audience, can be easily added in ALAIM as auxiliary tasks. In this way, ALAIM can be indisputably adopted to the applications that seek to use or evaluate the audience quality of social media users.

Related Work

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 (Erdoğmuş and Cicek 2012). Marketers seek loyal audiences for their marketing campaigns since loyal audiences have more trust, positive engagement, and repurchase than other users (Bergel and Brock 2019). Audience loyalty can be established, enhanced, and maintained by having persistent interactions and offering enjoyable social media contents (Jun and Yi 2020; Nisar and Whitehead 2016). 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 (Hawkins and Vel 2013; van Asperen, de Rooij, and Dijkmans 2018; Dholakia and Durham 2010; Brodie et al. 2013). 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 (Reichheld 1994; Ahmad and Buttle 2001). Therefore, in this work, we use the audience retention rate to measure the level of audience loyalty of a given social media user.

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 (De Veirman, Cauberghe, and Hudders 2017), social media users can manipulate their popularity by purchasing fake followers or fake engagements (a.k.a. link farming) (Confessore et al. 2018; Sen et al. 2018;

Benigni, Joseph, and Carley 2019). To understand and address this problem, inauthentic user (i.e., bot) detection has been broadly studied (Ferrara et al. 2016; Orabi et al. 2020). Many studies found distinct engaging or following behaviors of bots that are different from authentic users. Sen et al. (2018) exploit engaging frequency and topical information, Kudugunta and Ferrara (2018) suggest encoding contextual information from user profiles with RNN, and Chavoshi, Hamooni, and Mueen (2016) focus on synchronized behavior of inauthentic accounts to detect bots on social media. Recently, Yang et al. (2020) propose a generalized and scalable bot detection framework optimized with various validation sets. Although the previous studies propose decent bot detection methods, most of them focus on the single task, bot detection, and fail to incorporate other evaluation metrics.

Graph Neural Networks (GNN)

In recent years, Graph Neural Networks (GNN) (Scarselli et al. 2008) 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) (Kipf and Welling 2016) 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. (2018) 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) (Yun et al. 2019) 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. (2018), 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. (2019) 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.

Multi-Task Learning

Multi-Task Learning (MTL) enables us to model multiple related tasks by sharing representations (Ruder 2017). 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 (Caruana 1997). Thanks to such advantages, MTL has been used in many kinds of applications of machine learning such as computer vision (Zhang et al. 2014), natural language processing (Collobert and Weston 2008), speech recognition (Kim, Hori, and Watanabe 2017), and user profiling (Kim et al. 2020). Our proposed framework has multiple tasks that jointly learn representations from the multi-relational GCNs.

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 predefined 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.

Dataset

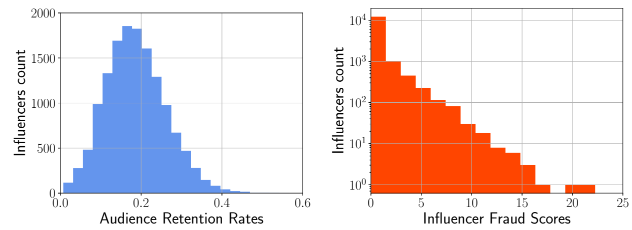
In this section, we describe the influencer-audience dataset and analyze audience engagement behaviors based on our proposed metrics.

Audience Data Collection

To evaluate influencer's audiences, we use the Instagram influencer dataset (Kim et al. 2020) 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, excluding 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.

Analysis on Audience Evaluation

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 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.

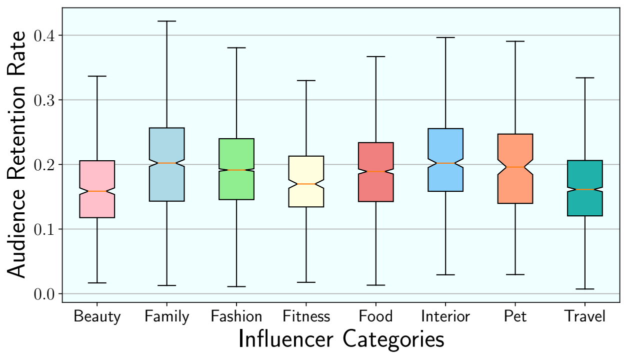


(a) ARR (b) IFS

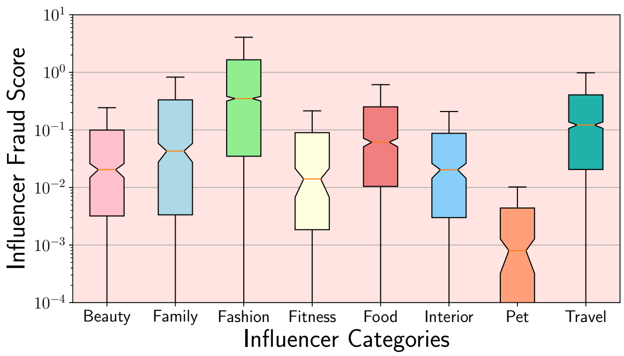
Figure 2: Distributions of ARR and IFS of the influencers. ARR has a normal distribution, but IFS has an exponential distribution. The IFS distribution suggests that a small portion of influencers is related to engagement bots while most influencers do not purchase fake engagements.

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 (Sen et al. 2018; Akyon and Kalfaoglu 2019; Kim and Han 2020). 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 (Rodriguez 2014; Instagram 2018). The IFS distribution of the influencers is shown in Figure 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.

Audience Analysis on Influencer Categories We further investigate the loyalty and authenticity of audiences across different influencer categories. Figure 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 5(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 influ-



(a) Audience Retention Rate



(b) Influencer Fraud Score

Figure 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.

encers 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.

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 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

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

Table 1: Pearson correlation between evaluation metrics (ARR, IFS) and the degrees of the influencer nodes (Degree)

the proposed model. As shown in Table 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.

Methodology

In this section, we present our proposed model. The overview of the proposed framework is illustrated in Figure 4. The framework consists of three components, engagement encoder, multi-relational GCNs, and multi-task decoder.

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. Note that we conduct an experiment to compare the performance between the contextualized information from concatenated comments and that from separated comments in Section Experiments.

To capture unique contextualized engagement features from a long sequence of user comments, we adopt the pre-trained Longformer (Beltagy, Peters, and Cohan 2020), 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:

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

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

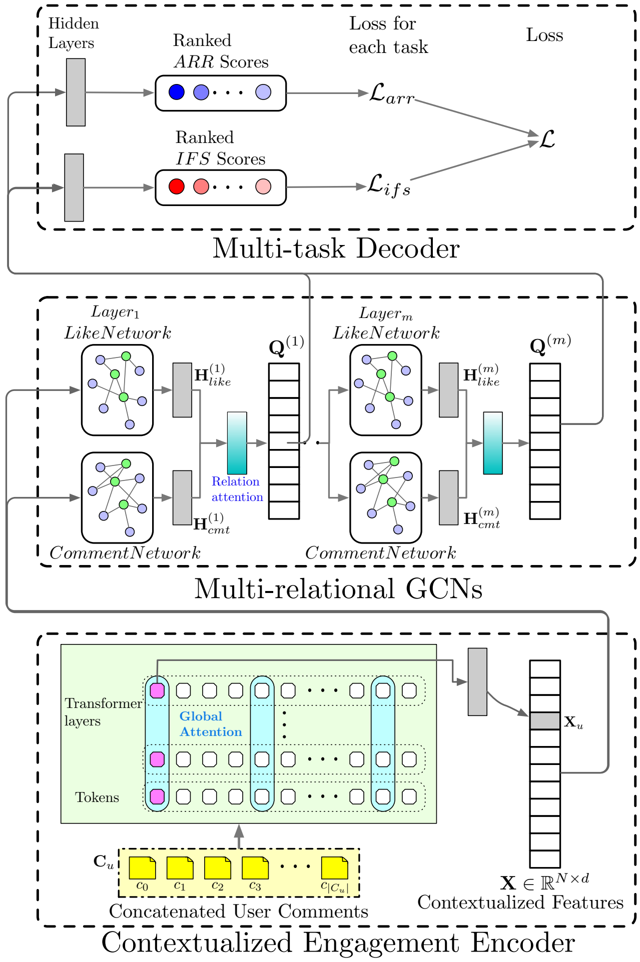


Figure 4: The overall framework of the proposed ALAIM. The framework consists of three components including the contextualized engagement encoder, the multi-relational GCNs, and the multi-task decoder.

Multi-Relational GCNs

To model social relations between influencers and their audiences based on engagements, we apply graph convolutional networks (GCNs) (Kipf and Welling 2016) 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:

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

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:

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

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.

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 (Xia et al. 2008). 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:

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

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

where $\boldsymbol{l}(\hat{\boldsymbol{y}}_i(\boldsymbol{Q}_{\boldsymbol{i}}), \boldsymbol{y}_{\boldsymbol{i}})$ 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.

Experiments

Experimental Setting

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}.

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 (Kipf and Welling 2016) and R-GCN (Schlichtkrull et al. 2018), 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. As we discuss in Section Related Work, 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 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.

Relevance Assignments We use the normalized discounted cumulative gain (NDCG) (Järvelin and Kekäläinen 2000) 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 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 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.

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%)

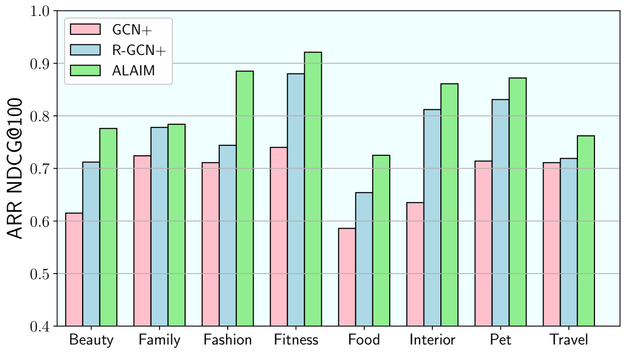
Table 2: The number of influencers in different relevance levels over the two audience evaluation metrics, ARR, and IFS.

	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**

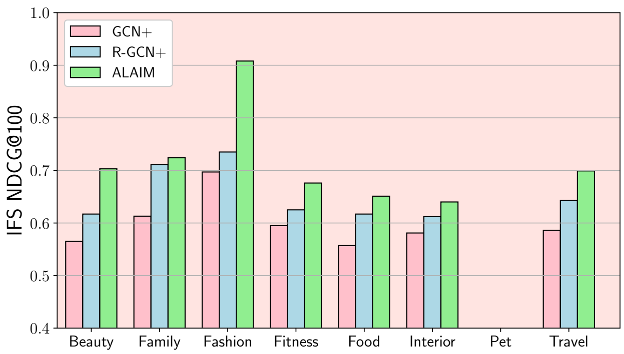
Table 3: Ranking performance measured by NDCG score. The proposed ALAIM outperforms other baseline methods in both audience evaluation tasks.

Experimental Results

Ranking Performance Evaluation We first compare the ranking performance of ALAIM with other baseline methods. Table 3 shows NDCG scores of the models in the two audience evaluation tasks. As shown in Table 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,



(a) ARR ranking results



(b) IFS ranking results

Figure 5: Ranking performance across the influencer categories. ALAIM shows robust ranking results for loyalty and authenticity evaluations.

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.

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 pet influencers are always zero. Figure 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.

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.

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 (Campbell and Farrell 2020; Warren 2020). 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 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 (Warren 2020). 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 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,

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**

Table 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.

but they should not be recommended to marketers due to high IFS values.

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.

Analysis on Contextualized Engagement Embedding

The proposed framework uses Longformer (Beltagy, Peters, and Cohan 2020) 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 (Devlin et al. 2018) 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.

	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**

Table 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.

Table 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.

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 influencer categories.

Table 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

	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**

Table 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.

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.

Conclusions

In this paper, we propose a multi-task multi-relational framework that can evaluate the audience quality of social media influencers. Our proposed framework, ALAIM, applies the long document transformer (Longformer) to the concatenated comments to generate contextualized embeddings that represent user commenting behaviors. The framework embeds the contextualized features to the multi-relational networks, which are constructed based on different types of engagements, to learn engagement relationships between influencers and audiences. The captured engaging behaviors of audiences are then used to estimate loyalty and authenticity scores of influencers. The experimental results show that the proposed ALAIM outperforms other baseline methods by acquiring knowledge from interactions between the multi-relational networks as well as the contextualized information. Moreover, ALAIM further enhances the model performance by jointly learning multiple related tasks. Since our proposed framework uses multi-task learning, the output can be generalized to represent social media users thus any potential audience evaluation metrics can be added as additional tasks. Also, any type of engagement can be easily adopted into the framework. Therefore, we believe our proposed framework can be utilized not only by influencer marketing stakeholders but also by the applications that require audience evaluations.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments. This work was supported by NSF DGE-1829071 and the framework of international cooperation program managed by the National Research Foundation of Korea (NRF-2020K2A9A2A11103842). Corresponding author is Jinyoung Han.

References

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

Akyon, F. C.; and Kalfaoglu, M. E. 2019. Instagram Fake and Automated Account Detection. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–7. IEEE.

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

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

Beltagy, I.; Peters, M. E.; and Cohan, A. 2020. Longformer: The Long-Document Transformer. *arXiv:2004.05150* .

Benigni, M. C.; Joseph, K.; and Carley, K. M. 2019. Botivistm: Assessing information manipulation in social media using network analytics. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, 19–42. Springer.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hoffman, D. L.; and Fodor, M. 2010. Can you measure the ROI of your social media marketing? *MIT Sloan management review* 52(1): 41.

Instagram. 2018. Reducing Inauthentic Activity on Instagram. https://about.instagram.com/blog/announcements/reducing-inauthentic-activity-on-instagram. Accessed: 2018-11-19.

Järvelin, K.; and Kekäläinen, J. 2000. IR evaluation methods for retrieving highly relevant documents. In *SIGIR*. ACM.

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

Kim, S.; and Han, J. 2020. Detecting Engagement Bots on Social Influencer Marketing. In *International Conference on Social Informatics*, 124–136. Springer.

Kim, S.; Hori, T.; and Watanabe, S. 2017. 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)*, 4835–4839. IEEE.

Kim, S.; Jiang, J.-Y.; Nakada, M.; Han, J.; and Wang, W. 2020. Multimodal Post Attentive Profiling for Influencer Marketing. In *Proceedings of The Web Conference 2020*, 2878–2884.

Kim, S.; Jiang, J.-Y.; and Wang, W. 2021. 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*, 319–327.

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

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

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

Lou, C.; and Yuan, S. 2019. Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising* 19(1): 58–73.

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

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

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

Orabi, M.; Mouheb, D.; Al Aghbari, Z.; and Kamel, I. 2020. Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management* 57(4): 102250.

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

Reinikainen, H.; Munnukka, J.; Maity, D.; and Luoma-aho, V. 2020. '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.

Rodriguez, S. 2014. 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. Accessed: 2014-12-11.

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

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

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

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

van Asperen, M.; de Rooij, P.; and Dijkmans, C. 2018. 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.

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

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

Warren, J. 2020. The Ultimate Guide to Instagram Influencer Marketing. https://later.com/blog/instagram-influencer-marketing/. Accessed: 2020-11-08.

Xia, F.; Liu, T.-Y.; Wang, J.; Zhang, W.; and Li, H. 2008. Listwise approach to learning to rank: theory and algorithm. In *ICML*, 1192–1199. ACM.

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

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

Yang, X.; Kim, S.; and Sun, Y. 2019. 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*, 101–104.

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

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