Interpreting Twitter User Geolocation

Ting Zhong[1], **Tianliang Wang**[1], **Fan Zhou**[1†], **Goce Trajcevski**[2], **Kunpeng Zhang**[3], **Yi Yang**[4]

[1]University of Electronic Science and Technology of China, China
[2]Iowa State University, Ames, IA, USA
[3]University of Maryland, College Park, USA
[4]The Hong Kong University of Science and Technology, Hong Kong
[†]Corresponding author: fan.zhou@uestc.edu.cn

Abstract

Identifying user geolocation in online social networks is an essential task in many location-based applications. Existing methods rely on the similarity of text and network structure, however, they suffer from a lack of interpretability on the corresponding results, which is crucial for understanding model behavior. In this work, we adopt influence functions to interpret the behavior of GNN-based models by identifying the importance of training users when predicting the locations of the testing users. This methodology helps with providing meaningful explanations on prediction results. Furthermore, it also initiates an attempt to uncover the so-called "black-box" GNN-based models by investigating the effect of individual nodes.

1 Introduction

Identifying geographic locations of users in online social networks (OSN) has become a key Internet service for many downstream applications, including location-based targeted advertising, emergency location identification, political election campaign, local event/place recommendation, natural disaster response, and remediation, etc. (Zheng et al., 2018). As such, the problem of *user geolocation* (UG) has received a great deal of research attention in the past decade (Han et al., 2012; Do et al., 2017; Miura et al., 2017; Rahimi et al., 2018; Bakerman et al., 2018).

Earlier efforts (Amitay et al., 2004; Wing and Baldridge, 2011; Han et al., 2012; Roller et al., 2012; Ahmed et al., 2013; Han et al., 2014; Chong and Lim, 2017) mainly focused on extracting indicative information from user-posted contents. These approaches rely on informative words that can link users to their specific locations via various natural language processing techniques such as topic modeling and other statistical models. More recently, researchers aimed at incorporating multi-aspect information, especially the user mention/interaction network to boost the performance of geolocation identification (Rahimi et al., 2015; Do et al., 2017; Rahimi et al., 2017, 2018; Hamouni et al., 2019). For example, (Rahimi et al., 2018; Wu et al., 2019) employ Graph Convolutional Networks (GCNs) (Kipf and Welling, 2017) or simplified GCN (Wu et al., 2019) to learn network structures for user geolocation. In addition, graph representation-based methods (Tang et al., 2015; Grover and Leskovec, 2016; Kipf and Welling, 2017; Hamilton et al., 2017; Qiu et al., 2018) have also been widely used for user geolocation (Do et al., 2017; Miura et al., 2017; Rahimi et al., 2018; Hamouni et al., 2019; Huang and Carley, 2019).

However, the existing methods lack model transparency and fail to provide meaningful explanations regarding the model behavior and prediction results, which prevents them from safety-critical applications. For example, when locating an emergency for a specific region, it would be more meaningful to explain why such prediction is made, rather than simply presenting numerical ranking values.

To address such limitations, we propose a general framework to explain the behavior of user geolocation models and the prediction results, by utilizing the influence function (Hampel et al., 2011; Koh and Liang, 2017) to quantify the impact of in-network users on the predicted outcome. The main assumption is that the prediction results from a trained geolocation model are typically affected by the knowledge learned from training data (i.e., all in-network users and their associated attributes). We demonstrate that the user geolocation (especially) for the network-based methods, is largely dominated by the geographical locations of the 1-hop neighboring nodes. This finding, on the one hand, enables demystification of the model behavior and quantitative measuring of the influence of individual users (both 1-hop and high-order prox-

imity nodes); and on the other hand, provides interpretation on the predicted locations for downstream applications/decision makers – which, in turn, points out a direction to further improve the model.

2 Preliminaries

The **User geolocation** problem is defined as predicting the user's "home" location (Zheng et al., 2018). Since each location is described by a *(longitude, latitude)* pair, the task can be converted into a typical classification problem where the labels/categories are region ids obtained through some space-partitioning methods, such as k-d trees, that divides the surface of the earth into closed and non-overlapping regions. Each user is associated with one (and only one) region-label that she/he belongs to. We use one-hot encoding to denote each label $\mathbf{y} \in \mathbb{R}^{1 \times c}$, where c is the number of regions. The user geolocation task thus becomes one of identifying the geographical locations, given the user tweet contents $\mathbf{X}$ and the mention network $\mathcal{G}$.

Graph Neural Networks (GNNs) (Bruna et al., 2014; Defferrard et al., 2016; Kipf and Welling, 2017; Hamilton et al., 2017; Velickovic et al., 2018; Ying et al., 2018; Xu et al., 2019; You et al., 2019) are powerful tools of representation learning for graph data, which has received increasing attention over the past several years (Zhou et al., 2018; Wu et al., 2020). The main idea is that, given a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with attributes $\mathbf{x}_v \in \mathbf{X}$ for every node v ($v \in \mathcal{V}$), a general GNN is to learn a vector representation of node v by:

$$\mathbf{x}_v^{(k)} = f_\mathrm{M}^{\theta_2}\left(\mathbf{x}_v^{(k-1)}, f_\mathrm{A}^{\theta_1}\left(\left\{\mathbf{x}_u^{(k-1)} \big| u \in \mathcal{N}(v)\right\}\right)\right),$$

where θ_1 and θ_2 are trainable parameters, and $\mathcal{N}(v)$ indicates the neighboring nodes of node v. Specifically, $f_\mathrm{A}^{\theta_1}$ aggregates information from neighbors, $f_\mathrm{M}^{\theta_2}$ merges the node representations from step $k-1$ with the computed neighborhood information. Both $f_\mathrm{A}^{\theta_1}$ and $f_\mathrm{M}^{\theta_2}$ can be arbitrary differentiable, permutation-invariant functions (e.g., neural networks) (Rahimi et al., 2018; Wu et al., 2019).

3 Interpreting Geolocation Results

Previous works (Rahimi et al., 2015; Do et al., 2017; Hamouni et al., 2019) typically leverage graph embedding methods (Grover and Leskovec, 2016; Kipf and Welling, 2017; Hamilton et al.,

2017; Wu et al., 2019) for user representation learning. In spite of the superior performance, GNNs (including those used for user geolocation) (Rahimi et al., 2018; Wu et al., 2019) model the process of learning and prediction in a "black-box" manner and, consequently, are limited in terms of interpretability. Therefore, it is important to understand how the model learns the data and why a particular prediction is made. To demystify the geolocation models, we intend to uncover and explain the training process and prediction outcomes.

There has been a growing interest to explain the model behavior and the predicted outcomes in the area of neural networks. In particular, an influence function (Cook and Weisberg, 1980; Hampel et al., 2011) is used to estimate the effect of individual training sample and interpret the results made by a specific model. While providing an elegant post-training interpretation, computationally-prohibitive cost of repeatedly retraining the model is avoided, spurring wide uses in a range of applications, including image classification (Koh and Liang, 2017), group effect (Koh et al., 2019), recommendation (Cheng et al., 2019), etc. However, whether influence functions can be applied to GNN-based models remained unclear. We take a step towards bridging this gap by tracing the geolocation results from the GNN-based models back to the important nodes in the mention network.

Specifically, let $z_1, \ldots, z_m$ denote the m training samples (nodes in the mention network $\mathcal{G}$), where each $z_i = (\mathbf{x}_i, \mathbf{y}_i)$ is an attribute vector of i-th node and its label. Removing one sample z from the training set can change the model optimal parameters from θ^* to θ_{-z}^*, where θ^* and θ_{-z}^* is the set of optimal parameters with and without the sample z, respectively. And θ_{-z}^* can be obtained by: $\theta_{-z}^* \stackrel{\text{def}}{=} \arg\min_{\theta \in \Theta} \sum_{z_i \neq z} \ell(z_i, \theta)$, where $\ell(z_i, \theta)$ is the loss of sample z_i. To estimate the influence of every removed training sample z and avoid retraining the model, Koh et al. (Koh and Liang, 2017) use an influence function strategy to efficiently approximate this behavior. The basic idea is to compute the change of optimal parameters if z was upweighted by some small ϵ, which gives the new parameters:

$$\theta_{\epsilon,z}^* \stackrel{\text{def}}{=} \arg\min_{\theta \in \Theta} \frac{1}{m} \sum_{i=1}^{m} \ell(z_i, \theta) + \epsilon \ell(z, \theta), \quad (1)$$

where the influence of upweighting z on the param-

eters θ^* is given by:

$$\mathcal{I}_{\text{up},\theta^*}(z) \stackrel{\text{def}}{=} \left.\frac{\partial \theta^*_{\epsilon,z}}{\partial \epsilon}\right|_{\epsilon=0} = -\mathbf{H}^{-1}_{\theta^*} \bigtriangledown_\theta \ell(z, \theta^*), \quad (2)$$

$\mathbf{H}_{\theta^*} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} \bigtriangledown^2_\theta \ell(z_i, \theta^*)$ is the Hessian matrix, and Eq. (2) shows that removing z is the same as upweighting it by $\epsilon = -\frac{1}{m}$. Thus, we can linearly approximate the parameter change of removing z as $\theta^*_{-z} - \theta^* \approx -\frac{1}{m}\mathcal{I}_{\text{up},\theta^*}(z)$ without retraining the model. The influence of upweighting a training node z on the loss for a testing node z_{test} can then be calculated according to the chain rule:

$$\mathcal{I}_{\text{up,loss}}(z, z_{\text{test}}) \stackrel{\text{def}}{=} \left.\frac{\partial \ell(z_{\text{test}}, \theta^*_{\epsilon,z})}{\partial \epsilon}\right|_{\epsilon=0}$$
$$= \bigtriangledown_\theta \ell(z_{\text{test}}, \theta^*)^\mathsf{T} \left.\frac{\partial \theta^*_{\epsilon,z}}{\partial \epsilon}\right|_{\epsilon=0}$$
$$= - \bigtriangledown_\theta \ell(z_{\text{test}}, \theta^*)^\mathsf{T} \mathbf{H}^{-1}_{\theta^*} \bigtriangledown_\theta \ell(z, \theta^*). \quad (3)$$

To speed up the computation, we use implicit Hessian-vector products (HVPs) to approximate $w_{\text{test}} \stackrel{\text{def}}{=} \mathbf{H}^{-1}_{\theta^*} \bigtriangledown_\theta \ell(z_{\text{test}}, \theta^*)$, and Eq. (3) can be rewritten as $\mathcal{I}_{\text{up,loss}}(z, z_{\text{test}}) = -w_{\text{test}} \bigtriangledown_\theta \ell(z, \theta^*)$. Since the Hessian $\mathbf{H}_{\theta^*}$ is positive semi-definite by assumption, we have:

$$w_{\text{test}} \equiv \arg\min_\beta \left\{ \frac{1}{2}\beta^\mathsf{T} \mathbf{H}_{\theta^*} \beta - \bigtriangledown_\theta \ell(z_{\text{test}}, \theta^*)^\mathsf{T} \beta \right\},$$

where the exact solution β can be obtained with conjugate gradients requiring only the evaluation of $\mathbf{H}_{\theta^*}\beta$ instead of explicitly computing $\mathbf{H}^{-1}_{\theta^*}$. We refer the reader to (Koh and Liang, 2017) for more detailed explanations on this topic, where the application of influence functions in computer vision is investigated. Note that the above method does not depend on a specific GNN implementation, i.e., it can be easily adapted to any GNN models.

4 Experiments

We now discuss in detail our experimental findings.

4.1 Datasets and Models

We investigate the interpretability of two user geolocation models: (1) SGC (Wu et al., 2019), which is a simplified GCN that removes the non-linearity in each layer of GCN and has achieved state-of-the-art performance for user geolocation. (2) A simple MLP model which combines the embedding of user-posted content and the node embedding for UG prediction. Here, we use doc2vec for user content embedding. For SGC, we use the published

code[1], where the tweet content is fed into a GNN as the node attribute. For MLP, we use node2vec to generate node representations. Note that we follow the evaluation metrics used by previous works.

We used two real-world Twitter datasets for evaluation: **GeoText** (Eisenstein et al., 2010) and **Twitter-US** (Roller et al., 2012). Their descriptive statistics are shown in Table 1.

Dataset	# Train	# Val	# Test	# Label
GeoText	5,685	1,895	1,895	129
Twitter-US	429,200	10,000	10,000	256

Table 1: Descriptive statistics of datasets.

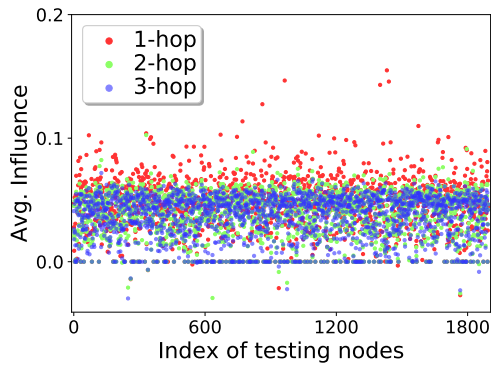| Method	GeoText			Twitter-US		
	Mean	*Median*	*Acc@161*	*Mean*	*Median*	*Acc@161*
GCN	546	45	60.1	485	71	62.3
SGC	531	40	61.1	479	70	62.5
MLP	555	46	60.2	545	86	59.6

Table 2: Performance comparison.

4.2 Interpretable Results

Before explaining the results, we compare the performance of SGC, MLP and GCN (Rahimi et al., 2018)[2] models as shown in Table 2. First of all, all three models can achieve similar performance results. While SGC performs slightly better, its superiority is not obvious compared to other GNN-based models and even the simple MLP model. This result implies that the improvement of UG results is limited even with the most advanced GNN-based models, at least on the two widely used benchmark datasets. Next, we turn to explain the results made by the SGC and MLP models. We omit other GCN-based models due to the limited space.
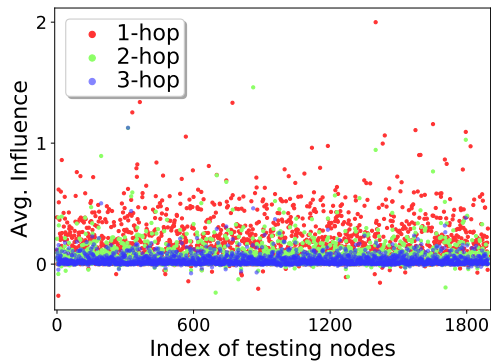
Influence of n-hop neighbors. For each test node, we *average* the influence value of the n-hop ($n = 1, 2, 3$) training neighbors in the mention network. As Figure 1 illustrates, 1-hop nodes usually have more positive influence on the test sample, while in contrast, the influence of 2-hop and 3-hop neighbors is relatively smaller. This is an intuitive interpretation of the geolocation results, but it verifies the effect of network-view modeling in user geolocation (Rahimi et al., 2015; Do et al., 2017;

---

[1]https://github.com/Tiiiger/SGC
[2]We used the implementation of GCN for geolocation provided by https://github.com/afshinrahimi/geographconv

(a) SGC

(b) MLP

Figure 1: The influence of n-hop neighbors.

(a) Overall vs. Positive vs. Negative.
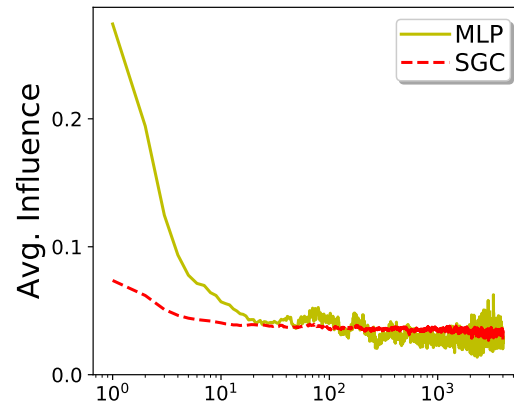
(b) Impact of distance.

Figure 2: (a) Overall performance vs. using only positive influence nodes vs. using only negative influence nodes of GeoText; (b) Average influence varies with distance (KM) between training and testing samples of GeoText.

Hamouni et al., 2019), i.e., the immediate neighbors can largely contribute to the user geolocation.

By comparing the results between SGC and MLP, we have the following observations. First, SGC is more sensitive to training samples, e.g., the influence value of SGC (y-axis) is far smaller than MLP. This is caused by the difference of fundamental training paradigms between SGC and MLP, i.e., SGC is a GNN-based model which considers the tweet content as attributes, but MLP embeds the tweet content and nodes independently. Therefore, a small change (removing a node) in SGC may incur significant influence on testing results. This result also implies that GNN-based models are more vulnerable to adversarial attacks, which is problematic for all existing GNN-based models as observed in recent works (Zügner and Günnemann, 2019a,b).

Second, for some nodes (indexed by x-axis) the average influence of their n-hop neighbors is negative, i.e., those data points that value are below 0. An interesting phenomena is that there are more such nodes for MLP as compared to SGC. We hypothesize that these nodes are the main reason of

the inferior performance of MLP. To verify this assumption, we scrutinize the prediction results of these nodes. As shown in Figure 2(a), where the accuracy of predicting these nodes in MLP is significantly lower than in SGC, which confirms our hypothesis and gives promising explanations on the classification error, i.e., the low classification accuracy might be caused by those nodes whose n-hop influence are negative (cf. Figure 1).

Influence of geographic distance. We also quantify the influence of geographic distance between training samples and testing samples. As Figure 2(b) shows, the close training data (e.g., less than 10KM) have higher impacts on locating users. However, the influence significantly drops with distance greater than 10KM. In another word, geographically far nodes (beyond a threshold) may
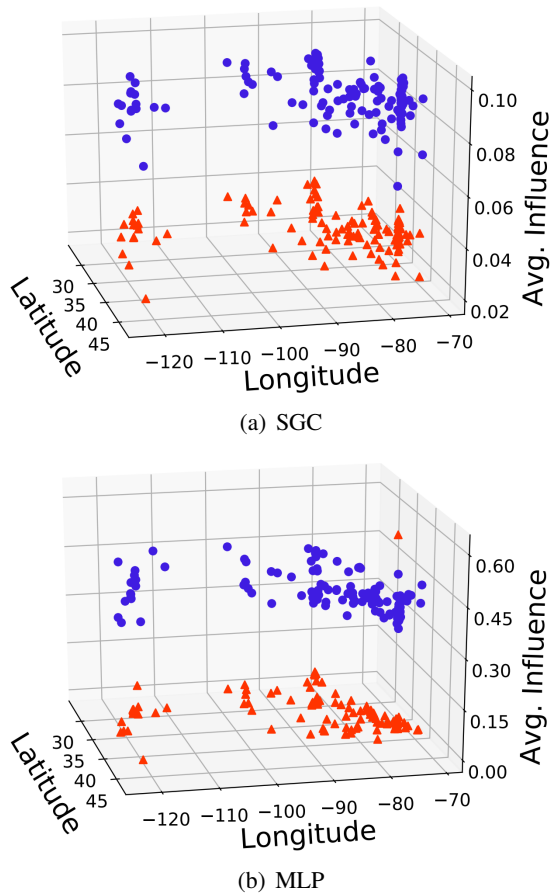
(a) SGC

(b) MLP

Figure 3: Visualization of the influence on all regions of GeoText. Upper cluster: the impact of *in-region* samples; Bottom cluster: the impact of *out-region* samples.

have less impact on geolocating a user.

Influence of regions. It is of interest to investigate the effect of regions on the user geolocation. The most important assumption in user geolocation is that geographically similar users should be topologically proximal. To quantify this assumption, for each region, we treat the training samples within this region as *in-region* nodes, and those out of it as *out-region* nodes. By calculating the mean influence of *in-region* and *out-region* samples for each test user, we can measure the influence of different regions. As Figure 3 shows, *in-region* samples (positive samples) have more significantly positive influence than *out-region* samples (negative samples). This result not only consolidates the motivation of most multi-view user geolocation models, but also suggests an important direction for improving the geolocation performance, i.e., paying more attention to the *in-region* samples.

5 Conclusion

In this work, we presented a framework for explaining the GNN-based models by extending the influence function to estimate the effect of samples in graph data. The experiments conducted on a specific task – user geolocation – provided intuitive explanations and enabled quantification of the influence of individual training samples. Some interesting observations include the effects of regions and the sensitivity of GNN-based models, which open potentials for further improvements that we plan to address in our future work.

Acknowledgement

References

Amr Ahmed, Liangjie Hong, and Alexander J Smola. 2013. Hierarchical geographical modeling of user locations from social media posts. In *Proceedings of the 22th International World Wide Web Conference*, pages 25–36.

Einat Amitay, Nadav Har'El, Ron Sivan, and Aya Soffer. 2004. Web-a-where: geotagging web content. In *The International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 273–280. ACM.

Jordan Bakerman, Karl Pazdernik, Alyson Wilson, Geoffrey Fairchild, and Rian Bahran. 2018. Twitter geolocation: A hybrid approach. *ACM Transactions on Knowledge Discovery from Data*, 12(3):34.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral networks and locally connected networks on graphs. In *Proceedings of the International Conference on Learning Representations*.

Weiyu Cheng, Yanyan Shen, Linpeng Huang, and Yanmin Zhu. 2019. Incorporating interpretability into latent factor models via fast influence analysis. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 885–893.

Wen-Haw Chong and Ee-Peng Lim. 2017. Tweet geolocation: Leveraging location, user and peer signals. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1279–1288. ACM.

R Dennis Cook and Sanford Weisberg. 1980. Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, 22(4):495–508.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852.

Tien Huu Do, Duc Minh Nguyen, Evaggelia Tsiligianni, Bruno Cornelis, and Nikos Deligiannis. 2017. Multiview deep learning for predicting twitter users' location. *Computing Research Repository*, arXiv:1712.08091. Version 1.

Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. 2010. A latent variable model for geographic lexical variation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1277–1287.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864.

Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034.

Parham Hamouni, Taraneh Khazaei, and Ehsan Amjadian. 2019. Tf-mf: Improving multiview representation for twitter user geolocation prediction. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. 2011. *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons.

Bo Han, Paul Cook, and Timothy Baldwin. 2012. Geolocation prediction in social media data by finding location indicative words. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 1045–1062.

Bo Han, Paul Cook, and Timothy Baldwin. 2014. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500.

Binxuan Huang and Kathleen M Carley. 2019. A hierarchical location prediction neural network for twitter user geolocation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*.

Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*.

Pang Wei Koh and Percy Liang. 2017. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1885–1894.

Pang Wei W Koh, Kai-Siang Ang, Hubert Teo, and Percy S Liang. 2019. On the accuracy of influence functions for measuring group effects. In *Advances in Neural Information Processing Systems*, pages 5255–5265.

Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi, and Tomoko Ohkuma. 2017. Unifying text, metadata, and user network representations with a neural network for geolocation prediction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1260–1272.

Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. 2018. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 459–467. ACM.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2015. Twitter user geolocation using a unified text and network prediction model. In *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics*, pages 630–636.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2017. A neural model for user geolocation and lexical dialectology. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 209–216.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2018. Semi-supervised user geolocation via graph convolutional networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2009–2019.

Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge. 2012. Supervised text-based geolocation using language models on an adaptive grid. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 1500–1510.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *Proceedings of the International Conference on Learning Representations*.

Benjamin P Wing and Jason Baldridge. 2011. Simple supervised document geolocation with geodesic grids. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 955–964.

Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. 2019.

Simplifying graph convolutional networks. In *Proceedings of the International Conference on Machine Learning*, pages 6861–6871.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How powerful are graph neural networks? In *Proceedings of the International Conference on Learning Representations*.

Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, William L. Hamilton, and Jure Leskovec. 2018. Hierarchical graph representation learning with differentiable pooling. In *Advances in Neural Information Processing Systems*, pages 4805–4815.

Jiaxuan You, Rex Ying, and Jure Leskovec. 2019. Position-aware graph neural networks. In *Proceedings of the International Conference on Machine Learning*, pages 7134–7143.

Xin Zheng, Jialong Han, and Aixin Sun. 2018. A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1652–1671.

Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, and Maosong Sun. 2018. Graph neural networks: A review of methods and applications. *Computing Research Repository*, abs/1812.08434. Version 4.

Daniel Zügner and Stephan Günnemann. 2019a. Adversarial attacks on graph neural networks via meta learning. In *Proceedings of the International Conference on Learning Representations*.

Daniel Zügner and Stephan Günnemann. 2019b. Certifiable robustness and robust training for graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 246–256.