Understanding Structural Hole Spanners in Location-Based Social Networks: A Data-Driven Study

Xiaoxin He, Yang Chen
School of Computer Science, Fudan University, China
{xxhe17,chenyang}@fudan.edu.cn

Abstract

Location-based social networks (LBSNs) have been popular around the world. Some recent studies have focused on using online/offline social interactions among individuals to explain social phenomena, strongly demonstrating that data collected by LBSNs can be utilized to analyze the behavior of users. However, how structural hole spanners (SHS) behave in a LBSN requires more investigation. In this paper, we crawl the entire social network and all published tips of Foursquare, a leading LBSN app with more than 60 million users, using a distributed approach. Based on the crawled massive user data, we discuss the behavior characteristics of SHS in demographic, spatiotemporal and linguistic aspects. We further develop a classification model to accurately identify SHS and ordinary users based on their behavioral data. Our model achieved a high classification performance, with an F1-score of 0.821 and an AUC value of 0.879.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Keywords: behavioral difference; location-based services; online social networks; urban computing; structural hole theory

ACM Reference Format:
Xiaoxin He, Yang Chen. 2021. Understanding Structural Hole Spanners in Location-Based Social Networks: A Data-Driven Study. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3460418.3480398

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3480398

1 Introduction

With the advance of mobile computing technologies and social networking services, location-based social networks (LBSNs), such as Foursquare [9] [33], Yelp [11] [36] and Dianping [8] [13], have blended in well in daily lives and enabled exploration of the intersection between physical world and digital world. Considering the ubiquity of LBSNs and the abundant data they are continuously collecting, extracting meaningful information about human behavior through large scale user-generated content (UGC) data has attracted significant attention from both academia and industry.

LBSNs bridge the gap between the physical world and the online social networking services. On one hand, LBSNs add a new dimension to an existing social network and facilitate the sharing of people's location-embedded experiences. On the other hand, LBSNs connect users by the interdependency derived from their location-tagged media content. The interdependency includes not only that two users co-occur in the same or similar physical location [34] but also the knowledge, e.g., common interests, behavior, and activities, inferred from an individual's location history and location-tagged data [39] [26].

Therefore, LBSN has been considered as a vital tool for extracting knowledge about human behavior from their interaction with the physical world. For example, some existing studies focused on user similarity/link prediction, experts/influencers detection and community discovery.

The structural hole theory [4] [20], which clarifies how people benefit from the positions they are occupying in a social network and from their social connections, has been applied to various scenarios of social networks [12] [40]. However, the relationship between the role of SHS and their online/offline activities requires more study. Our contributions are two-fold: 1) We provide a comprehensive analysis of the behavior of SHS in a representative LBSN. We find out how SHS differ from the ordinary users in demographic, spatiotemporal and linguistic aspects. 2) We propose a predication model to detect SHS in LBSNs, which achieves a high classification performance with an F1-score of 0.821 and an AUC value of 0.879. To our knowledge, very few existing studies in LBSNs have discussed human behavior from a perspective of structural hole theory or predicted SHS using UGC in LBSNs.

The rest of this paper is structured as follows. We review the related work in Section 2. We introduce the background of the Foursquare network, and the dataset we used for our study in Section 3. We provide a comprehensive analysis of the behavior of SHS in Section 4. We discuss our prediction model in Section 5 and conclude the paper in Section 6.

2 Related Work

Social influence. Social influence is generally related to social networks [1] [5] [37] and user interactions [17] [21] [29]. The former includes but is not limited to the number of followers and PageRank while the latter focuses on the number of likes, comments, forwards, etc, depicting a user's social influence in different ways. Besides individuals, there are also studies about the social impact of groups [18].

Structural hole theory. Sociology covers well-established ideas about how positions in social networks benefit the people who occupy them. For example, the structural hole theory demonstrates that users acting as intermediaries or bridges between different communities possess advantages, since they control the key information diffusion paths [20]. It points out the connections to different communities in a social network increase the social capital in competitive areas, whereas closed networks with homogeneous and repetitive information do not. According to this theory, when groups have no direct connections between each other, as if there exits a hole in the social fabric, which is called a structural hole. An individual who fills it is called a structural hole spanner. Structural hole theory has been applied to other theories of social networks [12] [40], as well as various practical scenarios, such as business [38], software development [2] and information diffusion [22].

Location-based social networks. Combining the features of social networking and geographic information sharing, LBSNs have long been popular in recent years. [32] [31] [19] focused on using online social interactions between individuals to explain social phenomena. These studies strongly indicate that the LBSN data can be used to analyze the behavior of user groups. However, there has been very few discussion about the social behavior of SHS in the context of LBSN. Our research aims to bridge the gap between cyberspace and physical places.

3 Dataset

Foursquare [13] [24] [28] has been a leading site for the combination of location-based services (LBS) and mobile social networking. Unlike traditional online social networks (OSNs) such as Facebook [7], all activities on Foursquare are location-specific with two key functions of conducting check-ins and leaving tips.

Since most mainstream OSNs apply per-IP address rate limits, it becomes time and resource consuming to crawl an entire large-scale OSN. As an alternative, most existing work [16] [27] [28] only selects a sampled sub-graph for research which may lead to an incomplete and even biased conclusion. To avoid such situation, we crawled the data of all 62.6 million Foursquare users accelerated by launching 40 crawlers. Each crawler was deployed on a virtual instance of the Microsoft Azure platform, with a unique IP address. For each user's data, we use the official API to conduct the crawling. Each Foursquare user has a unique numeric ID, and the IDs are assigned successively. Therefore, we can register a new Foursquare account to get the maximum ID, denoted by *max_uid*. We divide the entire Foursquare ID space [1,*max_uid*] evenly into 40 chunks, and each crawler is responsible for one chunk of IDs.

From August 1 to September 10 in 2015, we crawled all 62.6 million Foursquare users' profile pages and friend lists. In addition, we crawled all tips and venues published by the users. Noted that out of respect for Foursquare users' privacy, we only capture publicly visible information.

Based on the crawled friend lists of all Foursquare users, we modeled the entire Foursquare network using an undirected graph $G = (V, E)$. V is the collection of all Foursquare users and E is the collection of social relationships between users. Each node in V represents a user, and each edge in E represents a social connection. The degree of a node in G represents the number of friends of the corresponding user. Thus, we constructed a social graph G with 66,884,764 nodes and 1,546,131,581 edges.

4 Analysis

In this section, we start with introducing the network constraint metric in Section 4.1, then we discuss behavioral differences between SHS and ordinary users in the perspective of descriptive information (Section 4.2), spatial (Section 4.3), temporal (Section 4.4), and language characteristics (Section 4.5).

4.1 Labeling Structural Hole Spanners

Aiming to understand the behavioral preference of SHS in an OSN, we need a yardstick for determining them. In this study, the constraint metric was adopted to distinguish SHS from the ordinary users. Constraint was proposed by Burt [4] to measure the network closure which describes the degree to which a node in the network is directly or indirectly connected with other nodes. The more connected with other contacts, the stronger constraint a node will have. [30] [35] also use it to reflect the constraints between relationships. The calculation of constraint is as follows.

- Calculate the time and energy P_{ij} that i spends on j proportional of all of his time and energy:

$$P_{ij} = \frac{a[i,j] + a[j,i]}{sum(a[i,k] + a[k,i])} \quad k \neq i, \forall k \in N \tag{1}$$

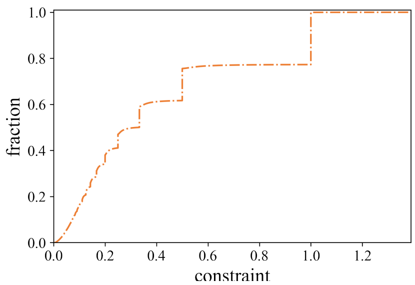


Figure 1. Distribution of the constraint.

- Calculate the constraint degree C_{ij} between node i and node j:

$$C_{ij} = (P_{ij} + \sum_{q} P_{iq}P_{qj})^2 \quad q \neq i, q \neq j \tag{2}$$

- Calculate the constraint metric C_{ij} of node i:

$$C_i = \sum_{j} C_{ij} \tag{3}$$

Due to the large scale of nodes and edges of Foursquare's social graph, we randomly select 100 thousand nodes and calculate their constraints. And we do not consider isolated nodes in our study. The distribution of the constraints is presented in Figure 1 . We sorted users by their constraints ascending. The top x% are labeled as SHS and the rest are labeled as ordinary users. We define the rest of users as ordinary users. It is worth mentioning that we also observe that the different thresholds of x will not affect the overall analysis results. For the sake of brevity, we use $x = 10$ by default unless specified.

4.2 Descriptive information

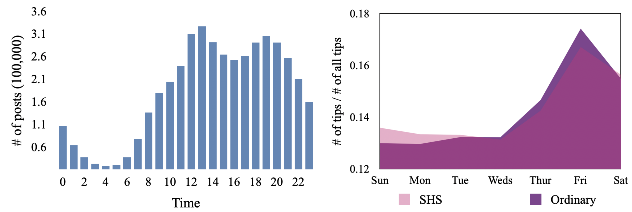
In this subsection, we discuss how SHS differ in gender, the number and length of tips and the tips with photos.

Gender. The number of male users is significantly higher than that of female users, with a ratio of 1.92. As for the user group of SHS, such difference is even more striking with a ratio of 2.7, while the ordinary users have a ratio of 1.91.

The number of tips. Foursquare users publish 3.33 tips on average while the average number of the tips published by SHS is 5 times more than that of ordinary users.

The length of tips. For the whole group of users, each tip is comprised of 22 words on average. But for the group of SHS, they tend to post longer tips, approximately 1.7 times longer than the group of ordinary users.

Tips with photos. Ordinary users seldom post tips with photos and only 5 out of 1,000 posts are with photos. However, for SHS, 3 out of 100 posts are with photos, which is five times higher compared with the ordinary users.



(a) Posting in a day. **(b)** Posting in a week.

Figure 2. Posting of a day and a week.

4.3 Spatial Characteristics

To analysis the spatial characteristic of SHS, we discuss in terms of the venue category, venue country and travel distance in this subsection.

Venue category. All activities on Foursquare are location-specific. Posts, the primary form of UGC on Foursquare, are tagged with according venues which can be divided into several categories, such as food, entertainment, and nightlife spot. To measure a user's diversity of physical mobility, we calculate the entropy of venue category. Specifically, we consider a user's posting as a discrete random event and the entropy of venue category can be used to characterise the uncertainty of user mobility. The larger the entropy is, the more random and less predicable a user's movement is.

The average entropy of venue category is 0.23. When the top 1% and top 10% users are labeled as SHS, the entropy of venue category is 4.44 times and 4.41 times of that of ordinary users, respectively. When the proportion of SHS arises, such trend always holds. Therefore, we believe that compared with ordinary users, SHS are more likely to explore different categories of places, instead of repeatedly visiting the same category of places.

Venue country. In addition to the entropy of venue category, a similar pattern persists at the national level. The entropy of venue country for the whole dataset is 0.03, while for the group of SHS is 0.10. Therefore, from the aspect of venue country, we again validate that SHS own more diversity in trajectory.

Travel distance. Travel distance refers to the mean distance between the venues where a user' two adjacent posts were published. For the entire dataset, the travel distance was 94.37 km. But for SHS, they tend to travel for a longer distance with average travel distance of 224 km, 242 km and 250 km respectively when top 1%, 5% and 10% are labeled as SHS, consistently higher than the 93 km, 86 km and 77 km of ordinary users.

4.4 Temporal Characteristics

For temporal factors, we consider the daily/weekly variation in posting and the post interval.

Within a day. We use timestamps to observe how users' posting varies within a day. We first convert the timestamp

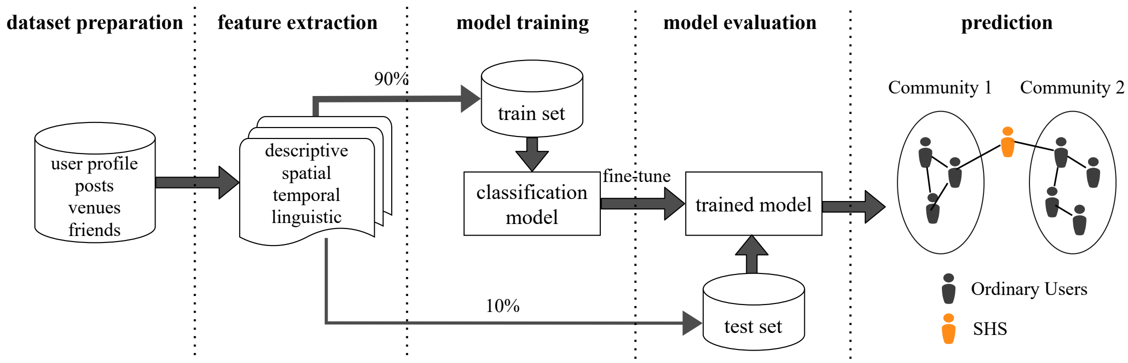


Figure 3. Design of the prediction model.

to local time to get the pattern of users' posts over 24 hours of a day. Then we divide 24 time slots and count the number of posts in each of them, as shown in Figure 2a. Again, we use entropy to measure the randomness of the posting time. Compared with ordinary users, SHS show more uncertainty about the time they post, with an entropy 3 times higher than others. Unlike posting at a regular time, their posting is an act more casual and flexible.

Within a week. As shown in Figure 2b, Friday is most popular, followed by Saturdays and Thursdays. Our result suggests that the peak of SHS at Friday (and Thursday) is less striking and the engagement on other days is slightly higher than ordinary users. That is to say, the participation of SHS in Foursquare is less influenced by the coming of weekend.

Post interval. Post interval is the mean time interval between each of the two adjacent posts. Considering there are users have never posted or only posted once, we fill their post interval with the maximum time interval we calculated. A shorter post interval is observed in the group of SHS in comparison of the group of ordinary users.

4.5 Language Characteristics

Since Foursquare is a world-wide LBSN, the corpus is made up of 45 different languages, with English, Indonesian and Spanish taking up the top 3. In order to facilitate the research but with representative, we selected English language as a subset from the data set for further study.

Vocabulary. People habitually convey their ideas through a set of own-style dictionaries. We first did natural language processing on all English posts, such as removing punctuation, converting to lowercase, and word segmentation. We then calculated the number and the entropy of words used by a user. The ordinary users have unique words of 45 and a entropy of 0.87 while SHS have 108 and 2.9, respectively.

Sentiment. To gain a deeper understanding of the semantics, especially the sentiment, we use Linguistic Inquiry and Word Count (LIWC) analyzer to categorise words according to their emotions into positive, negative, worried, angry and sad. Then we calculate the entropy of emotional words. As we label more users (1%,5%, and 10%) with low constraint as SHS, SHS consistently have a much greater entropy of emotion words in the aforementioned five types of sentiment.

Upper word. People occasionally use words in upper case with purposes. In Foursquare, every one of five posts contains a word in upper case and the frequency of upper-case word of SHS is slightly above the average with a ratio of 1.19.

5 Prediction Model

Nowadays, many popular OSNs, such as Facebook, allow a user to hide his or her friend list. And the inaccessible to the entire social graph information hinders the third-party application providers from determining whether a user is a SHS. Therefore, as an alternative, we proposed a supervised learning-based prediction model to identify SHS in LBSNs using their online/offline data. We introduce our system design in Section 5.1, experiment in Section 5.2 and feature importance study in Section 5.3.

5.1 System Design

Figure 3 presents the design of the prediction model. We roughly divided our system into five components: dataset preparation, feature extraction, model training, model evaluation and prediction. As described in Section 3, our data records user profiles, tips, venues, and friends. After cleaning and analysing the data, we extract features from users' location-centered data. In the part of model training, 90% of data is used as the training and validation set and the remaining 10% is used as the test set, feeding into various traditional machine learning algorithms. We evaluate the performance of different models and discuss the importance of different sets of features.

5.2 Experiment

We first randomly and respectively select 6,246 samples from positive and negative samples. After that we extract 19 features in accordance to the analysis in Section 4 and divide

Table 1. Experiment results of different models.

Model	Parameters	Precision	Recall	F1-Score	AUC
Random Forest	class_weight=balanced, max_depth=5, min_samples_split=7	0.777	0.765	0.770	0.766
XGBoost	booster=gbtree, gamma=0.1, max_depth=8, lambda=2, eta=0.001, subsample=0.7, min_child_weight=3	0.78	0.868	0.821	0.879
LightBoost	max_depth=5, learning_rate=0.05, num_leaves=500	0.781	0.748	0.763	0.770
CatBoost	eval_metric=AUC, depth=6, l2_leaf_reg=1, learning_rate=0.1	0.769	0.761	0.765	0.766
NaiveBayes	default	0.828	0.647	0.725	0.757
SVM	gamma=auto	0.770	0.807	0.787	0.783

Table 2. SHAP-base feature importance study.

rank	feature	Shapley values
1	time entropy (day)	180.722
2	post count	74.477
3	word entropy	69.299
4	post length	34.952
5	post interval	26.638
6	time entropy (week)	26.024
7	venue category entropy	21.936
8	travel distance	18.73
9	negative word entropy	13.892
10	uppercase word entropy	12.796

Table 3. Evaluation on different feature subsets.

Category	Precision	Recall	F1-Score	AUC
all	0.78	0.868	0.821	0.879
descriptive	0.750	0.816	0.779	0.775
spatial	0.761	0.774	0.766	0.767
temporal	0.779	0.776	0.776	0.784
language	0.779	0.768	0.772	0.776

them into four categories: descriptive information, spatial, temporal, and language features. Then we use several popular supervised machine learning models, including Random Forest [3], XGBoost [6], LightGBM [15], CatBoost [25], NaiveBayes [14], and Support Vector Machine(SVM) [10]. For simplicity and convenience, we do not use deep learning. For each model, we apply 10-fold cross-validation for the evaluation and use grid search to find the best parameters (Table 1).

There is no significant difference in the performance among models (Table 1). XGBoost has the highest score of AUC (0.879), closely followed by SVM (0.783). These decent performances indicate that the features extracted are competent in distinguishing SHS from a group of people.

5.3 Feature importance

Among these selected features, we use SHAP [23], a game theoretic approach to explain the output of machine learning models, to gain better understanding of the importance of features from different categories. SHAP connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. Features with large Shapley values are considered to be important. We summarise the top ten most important features in Table 2, two from the category of descriptive information, two from spatial features, three from temporal and three from language features. In other words, all categories of features contribute to the model's prediction.

We validate our conclusion by using each category independently for training. As we can see in Table 3, using only one category can achieve a comparable prediction performance to the original model.

6 Summary

To our knowledge, very few existing studies in LBSNs have discussed human behavior from a perspective of structural hole theory or predicted SHS using UGC in LBSNs. In this work, we regard SHS as a novel entry point to excavate valuable knowledge from massive data from LBSNs. We find out how SHS differ from others in descriptive information, spatiotemporal and linguistic characteristics. On top of that, we build a supervised machine learning-based prediction model to accurately distinguish between SHS and ordinary users based on their behavioral data. Our model achieved a high classification performance and also maintained a good performance when using only one category of features.

As future work, we will further improve the predication model to detect SHS with large scale of data. Moreover, we will consider applying the proposed model in other real-world LBSNs, which paves the way to develop novel physical and digital knowledge for urban computing.

Acknowledgment

This work is sponsored by National Natural Science Foundation of China (No. 62072115, No. 71731004, No. 61602122).

References

[1] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Everyone's an influencer: quantifying influence on Twitter. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011*. 65–74.

[2] Tanmay Bhowmik, Nan Niu, Prachi Singhania, and Wentao Wang. 2015. On the Role of Structural Holes in Requirements Identification: An Exploratory Study on Open-Source Software Development. *ACM Trans. Manag. Inf. Syst.* 6, 3 (2015), 10:1–10:30.

[3] Leo Breiman. 2001. Random Forests. *Mach. Learn.* 45, 1 (2001), 5–32.

[4] Ronald S Burt. 1992. *Structural holes*. Harvard university press.

[5] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and P. Krishna Gummadi. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010*.

[6] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785–794.

[7] Yang Chen, Jiyao Hu, Hao Zhao, Yu Xiao, and Pan Hui. 2018. Measurement and Analysis of the Swarm Social Network With Tens of Millions of Nodes. *IEEE Access* 6 (2018), 4547–4559.

[8] Qingyuan Gong, Yang Chen, Xinlei He, Zhou Zhuang, Tianyi Wang, Hong Huang, Xin Wang, and Xiaoming Fu. 2018. DeepScan: Exploiting Deep Learning for Malicious Account Detection in Location-Based Social Networks. *IEEE Commun. Mag.* 56, 11 (2018), 21–27.

[9] Qingyuan Gong, Yang Chen, Jiyao Hu, Qiang Cao, Pan Hui, and Xin Wang. 2018. Understanding Cross-Site Linking in Online Social Networks. *ACM Trans. Web* 12, 4 (2018), 25:1–25:29.

[10] Marti A. Hearst. 1998. Trends & Controversies: Support Vector Machines. *IEEE Intell. Syst.* 13, 4 (1998), 18–28.

[11] Amy Hicks, Stephen Comp, Jeannie Horovitz, Madeline Hovarter, Maya Miki, and Jennifer L. Bevan. 2012. Why people use Yelp.com: An exploration of uses and gratifications. *Comput. Hum. Behav.* 28, 6 (2012), 2274–2279.

[12] Hong Huang, Jie Tang, Lu Liu, Jarder Luo, and Xiaoming Fu. 2015. Triadic Closure Pattern Analysis and Prediction in Social Networks. *IEEE Trans. Knowl. Data Eng.* 27, 12 (2015), 3374–3389.

[13] Yue Huang, Yang Chen, Qian Zhou, Jin Zhao, and Xin Wang. 2016. Where are we visiting? Measurement and analysis of venues in Dianping. In *2016 IEEE International Conference on Communications, ICC 2016*. 1–6.

[14] George H. John and Pat Langley. 1995. Estimating Continuous Distributions in Bayesian Classifiers. In *UAI '95: Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*. 338–345.

[15] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, NeurIPS 2017*. 3146–3154.

[16] Xiangnan Kong, Jiawei Zhang, and Philip S. Yu. 2013. Inferring anchor links across multiple heterogeneous social networks. In *22nd ACM International Conference on Information and Knowledge Management, CIKM 2013*. 179–188.

[17] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue B. Moon. 2010. What is Twitter, a social network or a news media?. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010*. 591–600.

[18] Jianxin Li, Xinjue Wang, Ke Deng, Xiaochun Yang, Timos Sellis, and Jeffrey Xu Yu. 2017. Most Influential Community Search over Large Social Networks. In *33rd IEEE International Conference on Data Engineering, ICDE 2017*. 871–882.

[19] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2005. Geographic routing in social networks. *Proc. Natl. Acad. Sci. USA* 102, 33 (2005), 11623–11628.

[20] Zihang Lin, Yuwei Zhang, Qingyuan Gong, Yang Chen, Atte Oksanen, and Aaron Yi Ding. 2021. Structural Hole Theory in Social Network Analysis: A Review. *IEEE Transactions on Computational Social Systems* (2021), 1–16. https://doi.org/10.1109/TCSS.2021.3070321

[21] Qi Liu, Biao Xiang, Nicholas Jing Yuan, Enhong Chen, Hui Xiong, Yi Zheng, and Yu Yang. 2017. An Influence Propagation View of PageRank. *ACM Trans. Knowl. Discov. Data* 11, 3 (2017), 30:1–30:30.

[22] Tiancheng Lou and Jie Tang. 2013. Mining structural hole spanners through information diffusion in social networks. In *22nd International World Wide Web Conference, WWW 2013*. 825–836.

[23] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, NeurIPS 2017*. 4765–4774.

[24] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011. An Empirical Study of Geographic User Activity Patterns in Foursquare. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, ICWSM 2011*.

[25] Liudmila Ostroumova Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*. 6639–6649.

[26] Maayan Roth, Assaf Ben-David, David Deutscher, Guy Flysher, Ilan Horn, Ari Leichtberg, Naty Leiser, Yossi Matias, and Ron Merom. 2010. Suggesting friends using the implicit social graph. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 233–242.

[27] Salvatore Scellato, Cecilia Mascolo, Mirco Musolesi, and Vito Latora. 2010. Distance Matters: Geo-social Metrics for Online Social Networks. In *3rd Workshop on Online Social Networks, WOSN 2010*.

[28] Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo. 2011. Socio-Spatial Properties of Online Location-Based Social Networks. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, ICWSM 2011*.

[29] Jie Tang, Tiancheng Lou, and Jon M. Kleinberg. 2012. Inferring social ties across heterogenous networks. In *Proceedings of the Fifth International Conference on Web Search and Web Data Mining, WSDM 2012*. 743–752.

[30] Jie Tang, Tiancheng Lou, Jon M. Kleinberg, and Sen Wu. 2016. Transfer Learning to Infer Social Ties across Heterogeneous Networks. *ACM Trans. Inf. Syst.* 34, 2 (2016), 7:1–7:43.

[31] Anne LJ Ter Wal and Ron A Boschma. 2009. Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science* 43, 3 (2009), 739–756.

[32] Giorgio Topa. 2001. Social interactions, local spillovers and unemployment. *The Review of Economic Studies* 68, 2 (2001), 261–295.

[33] Gang Wang, Sarita Yardi Schoenebeck, Haitao Zheng, and Ben Y. Zhao. 2016. "Will Check-in for Badges": Understanding Bias and Misbehavior on Location-Based Social Networks. In *Proceedings of the Tenth International Conference on Web and Social Media, ICWSM 2016*. 417–426.

[34] Huandong Wang, Yong Li, Yang Chen, and Depeng Jin. 2019. Co-Location Social Networks: Linking the Physical World and Cyberspace. *IEEE Trans. Mob. Comput.* 18, 5 (2019), 1028–1041.

[35] Zhiyuan Wang, Yun Zhou, Jie Tang, and Jarder Luo. 2016. The Prediction of Venture Capital Co-Investment Based on Structural Balance Theory. *IEEE Trans. Knowl. Data Eng.* 28, 2 (2016), 537–550.

[36] Chenxi Yang, Yang Chen, Qingyuan Gong, Xinlei He, Yu Xiao, Yuhuan Huang, and Xiaoming Fu. 2018. Understanding the behavioral differences between american and german users: A data-driven study. *Big Data Min. Anal.* 1, 4 (2018), 284–296.

[37] Shaozhi Ye and Shyhtsun Felix Wu. 2013. Measuring message propagation and social influence on Twitter.com. *Int. J. Commun. Networks Distributed Syst.* 11, 1 (2013), 59–76.

[38] Yuyang Ye, Hengshu Zhu, Tong Xu, Fuzhen Zhuang, Runlong Yu, and Hui Xiong. 2019. Identifying High Potential Talent: A Neural Network Based Dynamic Social Profiling Approach. In *2019 IEEE International Conference on Data Mining, ICDM 2019*. 718–727.

[39] Yu Zheng. 2011. Location-Based Social Networks: Users. In *Computing with Spatial Trajectories*. Springer, 243–276.

[40] Thomas Zimmermann and Nachiappan Nagappan. 2008. Predicting defects using network analysis on dependency graphs. In *30th International Conference on Software Engineering, ICSE 2008*. 531–540.