# Structural Hole Theory in Social Network Analysis: A Review

Zihang Lin, Yuwei Zhang, Qingyuan Gong, Yang Chen, *Senior Member, IEEE*, Atte Oksanen, and Aaron Yi Ding, *Member, IEEE*

***Abstract*—Social networks now connect billions of people around the world, where individuals occupying different positions often represent different social roles and show different characteristics in their behaviors. The structural hole (SH) theory demonstrates that users occupying the bridging positions between different communities have advantages since they control the key information diffusion paths. Users of this type, known as SH spanners, are important when it comes to assimilating social network structures and user behaviors. In this article, we review the use of SHs theory in social network analysis, where SH spanners take advantage of both information and control benefits. We investigate the existing algorithms of SH spanner detection and classify them into information flow-based algorithms and network centrality-based algorithms. For practitioners, we further illustrate the applications of SH theory in various practical scenarios, including enterprise settings, information diffusion in social networks, software development, mobile applications, and machine learning (ML)-based social prediction. Our review provides a comprehensive discussion on the foundation, detection, and practical applications of SHs. The insights can facilitate researchers and service providers to better apply the theory and derive value-added tools with advanced ML techniques. To inspire follow-up research, we identify potential research trends in this area, especially on the dynamics of networks.**

***Index Terms*—Applications, machine learning (ML), social networks, structural hole (SH) theory.**

## I. Introduction

NOWADAYS, it is much easier for people to connect with one another and form complex social networks in diverse scenarios. On the one hand, the booming online social networks (OSNs), such as Facebook, Twitter, and WeChat, form social connections between remote users. Media sharing networks, such as YouTube, Snapchat, and Instagram, are gaining more and more popularity. On the other hand, there are many kinds of real-world offline social networks, such as friendship networks [1], criminal networks [2]–[4], and collaborative innovation networks [5]. Studies about social network structures and user profiling have been prosperous these years, with the aim to better understand various user groups, model information flows, and improve friends or news recommendations.

It is worth noting the advantages brought about by social connections. As a leading sociologist, Burt [6] put forward the point of view in the competitive fields that social structure is a key factor in determining investment returns. Connections to diverse communities of a social network increase the social capital one player could use in the competitive fields, while the closed networks with homogeneous and repetitive information will not bring such advantages. This standpoint serves as the core of the structural hole (SH) theory. According to this theory, groups of people who are unconnected form holes in the social structure. The lack of connection is referred to as an SH. Individuals acting as bridges or intermediaries between them fill the holes, called SH spanners [7], as shown in Fig. 1. These individuals occupying the holes benefit from getting access to more different kinds of opinions and ideas, synthesizing more potentially feasible methods, and better coordinating multiple tasks of diverse communities.

The idea behind SH theory owes to the weak tie theory developed by Granovetter [8]. In the weak tie theory, the overlap of two contacts in their friendship network increases if the strength of their tie is stronger. Weak ties act as bridges to diffuse novel ideas between different groups.

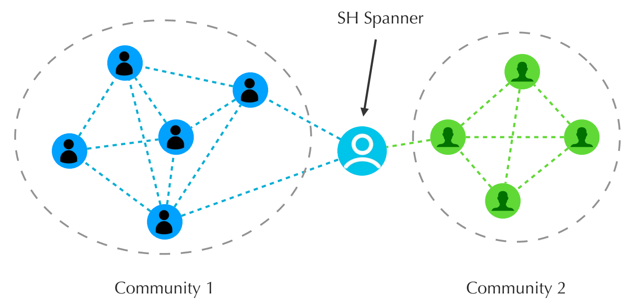



Fig. 1. SH spanners in social networks.

Manuscript received December 6, 2020; revised March 10, 2021; accepted March 17, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62072115, Grant 71731004, and Grant 61602122; in part by the project "PCL Future Greater-Bay Area Network Facilities for Large-Scale Experiments and Applications" under Grant LZC0019; and in part by the iSafe project funded by the TU Delft Safety & Security Institute. *(Zihang Lin and Yuwei Zhang are co-first authors.) (Corresponding authors: Qingyuan Gong; Yang Chen.)*

Zihang Lin, Yuwei Zhang, Qingyuan Gong, and Yang Chen are with the School of Computer Science, Fudan University, Shanghai 200433, China, also with the Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, China, and also with the Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: zihanglin18@fudan.edu.cn; zhangyuwei18@fudan.edu.cn; gongqingyuan@fudan.edu.cn; chenyang@fudan.edu.cn).

Atte Oksanen is with the Faculty of Social Sciences, Tampere University, 33100 Tampere, Finland (e-mail: atte.oksanen@tuni.fi).

Aaron Yi Ding is with the Department of Engineering Systems and Services, TU Delft, 2628 BX Delft, The Netherlands (e-mail: aaron.ding@tudelft.nl).

Digital Object Identifier 10.1109/TCSS.2021.3070321

While Granovetter argued that the strength of a tie determines whether it plays the bridging role, Burt considered that the cause lies in the SH it spans. According to the SH theory proposed by Burt [6], the advantages brought about by weak ties for SH spanners are information benefits. The SHs, which generate information benefits, gain control benefits as well, focusing on the privileges to negotiate with others.

In the SH theory, there are several indicators proposed to measure whether one node is an SH spanner in a network. Especially, the ego network is used to represent the one-hop network of a central node. A series of SH metrics is further defined to show whether one node is probably acting as an SH spanner, such as effective size, efficiency, constraint, and hierarchy. In the ego network, the given node serves as the ego and all its neighbors are the alters. All edges between these nodes are considered as the edges of the ego network. SH spanners tend to have higher values of effective size and lower values of constraint.

Efforts have also been made by researchers to detect SH spanners from several perspectives. We roughly classify the existing SH spanner detection algorithms into two categories. One is related to information dissemination, trying to identify the SH spanners as the most important nodes whose removal will block the maximum information flows or the most important nodes connecting as many communities as possible [7], [9], [10]. The second category is based on network centrality, which reflects the position and importance of a node in a social network. Some researchers developed heuristic algorithms on dynamic OSNs based on the weak tie theory [8] or came up with the concept of inverse closeness centrality to tackle the problem of finding top-$k$ SH spanners [11]–[13]. The above algorithms all consider the entire social network.

SH theory has been applied to other social science theories in social networks such as triads [14] and network oscillation [15], also in various practical scenarios. We classify the applications into several categories according to the research fields. In enterprise settings, SH theory provides valuable insights in diverse fields, including management science and innovative performance [16]–[19]. In the field of information diffusion, researchers have conducted experiments to testify the belief that SH spanners play a key role in spreading information in social networks and employed the concept to maximize social influence [7], [20]–[23]. In software development and mobile applications, the SH theory has also been adopted to understand both requirement identification [24] and defects prediction processes of software development [25], as well as interrelationships between different mobile service sectors [26]. The advancements of SH theory also help machine learning (ML) tasks with the background of social networks.

SH theory is important for both sociology and computational social science. It plays an important role in the aspects of studying critical positions in social structure [27], information dissemination [9], and link prediction [7].

To summarize, this article presents a comprehensive review of SH theory in social network analysis with the following contributions.

First, we provide a detailed review on the development of SH theory, along with the metrics measuring the importance of SH spanners in ego networks. We present the applicability of SH theory in combination with triads, network oscillation, and other social science theories in social networks.

Second, we systematically classify the related detection algorithms of SH spanners into information flow-based algorithms and network centrality-based algorithms. We categorize the applications of SH theory in practical fields into enterprise settings, information diffusion in social networks, software development and mobile applications, and ML-based scenarios.

Finally, we propose future expectations and potential research directions for SH spanner detection in dynamic social networks and the applications of SH theory in more practical fields, including dynamic networks, nonhuman networks, and integration with graph neural networks.

This article reviews the SH theory, along with the related detection algorithms of SH spanners and the applications of SH theory in practical fields in sequence. Literature review was conducted by using several major databases, including ScienceDirect, ACM Digital Library, IEEE Xplore, and SpringerLink. The rest of this article is structured as follows. In Section II, we review the SH theory. SH spanner detection algorithms are introduced then in Section III. Furthermore, we summarize the extensive applications of SHs in social networks in Section IV. We review the ML-based social predictions leveraging the SH theory in Section V and conclude this article in Section VI.

## II. Foundations of the Structural Hole Theory

In this section, we review the foundations of the SH theory. In Section II-A, we illustrate how SHs are formed in a social network and how SH spanners receive both the information and the control advantages. In Section II-B, we review the metrics of ego networks for measuring the importance of SH spanners in terms of network connectivity. In Section II-C, we introduce related social science theories that incorporate the SH theory.

### A. Illustration of Structural Hole Theory

SH theory measures the interpersonal relationship between users in social networks, especially the gains people could enjoy from their connections. Related studies that should not be ignored include the weak tie theory developed by Granovetter in 1973 [8]. In the weak tie theory, relationships between users are called ties. The relationships of a person have two categories of strength, where strong ties refer to most frequent and close contacts and weak ties refer to less frequent and less close contacts. The strength of a tie is weighted based on the intimacy, emotional intensity, the amount of time invested, and the extent of the mutual assistance in the relationship. The investigation by Granovetter shows that if there is a strong tie between persons A and B, it is possible that A and B have many ties in common, which makes A and B hard to be disconnected, as shown in Fig. 2. The study

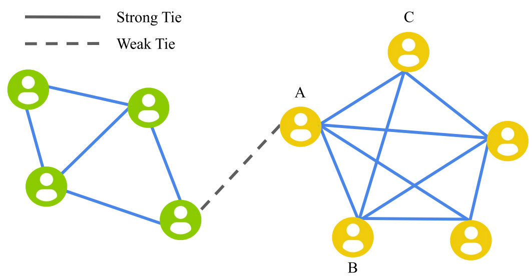



Fig. 2. Weak ties and strong ties in a social network.

of triads makes it more specific that if there is a strong tie between persons A and B and another strong tie between A and C, then there is at least a weak tie between B and C. If the tie strength is stronger between persons A and B or A and C, then the greater the probability that the tie strength is also stronger between B and C. If removing a tie will increase the length of the shortest path between two nodes, this tie acts as a bridge, which is more likely to be the source of novel information than other ties. In most cases, the weak tie theory holds that all bridges are weak ties in social networks. However, Burt [6] pointed out that a relationship will generate information benefits when it acts as a bridge over an SH, regardless of the strength of the tie. An SH is the lack of connection between two contacts, which is bridged by brokers.

In the scenario of social networks, Burt [6] first proposed the nonredundant relationship between two contacts as an SH, where the nonredundant relationship is the lack of both direct and indirect redundant connections. Individuals who fill the SHs are called SH spanners [7]. Unlike the weak tie theory that focuses more on the strength of ties for information propagation in the network structure, SH theory starts from the social capital which players could gain in competitive fields. Burt [6], [28] explored more on the benefits that SH spanners could enjoy by occupying the bridging positions in social networks. Because of the refined work divisions, people gather into different groups. People communicate much more closely in the same group than across groups, thus gradually forming terminology barriers between different communities. In a closed network, people can obtain higher credibility information at a lower cost. These closed networks, or communities, can promote internal communications and community development and generate redundant and overlapping information. Network structure serves as a proxy to learn the distribution of the sticky information, which is difficult to move to other groups. Burt *et al.* [27] and Stovel and Shaw [29] studied the correlation between advantages brought about by network structure and the achievements made by those people who contact diverse communities.

The potential value of SHs lies in the information benefits and control benefits. Information benefits indicate that SH spanners are able to access multiple nonredundant information sources, be early informed, and get referrals from their contacts which offer future opportunities, since they occupy the unique connecting positions between communities. SH spanners are able to access more information from nonredundant sources, where the information has better dissemination than repetition. Katz and Lazarsfeld [30] found that information, ideas, and innovations usually flow first to opinion leaders, and then to a wider range of people from those opinion leaders. People who bridge SHs are more inclined to connect with opinion leaders to get valuable information faster. Control benefits are from the distinct aggregative information of the connected communities, realizing the privileges for SH spanners to negotiate with users who used to interact more within communities. Establishing and negotiating relationships with contacts who are not in the same group can also lead to better returns, especially when it comes to detecting and developing profitable opportunities. Studies in social networks [27] show that brokers will receive corresponding compensation and benefits in processing information from different communities, such as bonus compensation for investment bankers, industry recognition for stock analysts, and early promotion for managers.

In a word, SH spanners occupying positions between different communities in social networks are rewarded with information benefits and control benefits: diffusing information from one group to another, negotiating and synthesizing different ideas, and promoting cooperation in diverse fields.

A recent study by Burt [31] proposed the reinforced structural holes (RSHs), showing that an SH is reinforced by the cohesion within a community and the exclusion of others. The more an SH is reinforced, the more difficult it is to bridge it, while the more likely a successful bridge would diffuse novel ideas and propagate valuable information between communities.

### B. Metrics of the Importance of SH Spanners

SH theory is described by the ego network, with the node set including one node as the ego and the surrounding nodes to whom the ego is directly connected as the alters, and the edge set including all the edges between these nodes. Here, Burt [6] proposed some metrics to figure out SHs and distinguish SH spanners in the ego network. From the metrics which can be classified as external measures for individual actors [32], Burt showed that advantage was highly related to information breadth, timing, and arbitrage. People can gain better evaluations, recognition, and salary by taking these advantages. In exploring the impact of connectivity on SH spanners, he proposed some metrics for detecting SHs, including effective size, efficiency, constraint, and hierarchy. The effective size of a node is an indicator to measure the nonredundant connections of a node. Formally, the effective size of a node $i$, denoted $e(i)$, is defined by

$$e(i) = \sum_{j \in N(i)} \left(1 - \sum_{q} p_{iq} m_{jq}\right), \quad q \neq i, j \tag{1}$$

where $N(i)$ is the set of neighbors of $i$, each $q$ is a node different from $i$ and $j$ in the ego network, and $p_{iq}$ is the mutual weight of the edge linking $i$ and $q$ from the matrix of network ties. Also, $m_{jq}$ is the mutual weight of $j$ and $q$ divided by the largest weight between $j$ and any of $j$'s neighbors. If contact $j$ is isolated from all other primary contacts, it indicates that $j$ provides one nonredundant contact in the ego network. When the relationships between $j$ and other contacts strengthen, the value from $j$ in the calculation will decrease, indicating that $j$ is gradually redundant in $i$'s network. Furthermore, efficiency is the effective size divided by the number of alters in the ego network.

As for the constraint, it is a measure of the extent to which a node's entrepreneurial opportunities are constrained within the ego network. The original definition of constraint in a node $i$, denoted $c(i)$, is

$$c(i) = \sum_{j \in N(i)} \left( p_{ij} + \sum_{q} p_{iq} p_{qj} \right)^2, \quad q \neq i, j \tag{2}$$

where the definitions of $N(i)$ and $p_{ij}$ are the same as (1). A node with constraint of 1 indicates that it has only one contact. When the constraint of a node is closer to 0, there are fewer connections between the node's contacts.

Hierarchy is an indicator that measures the extent to which the aggregate constraint on ego is concentrated in a single contact. Here, we denote the hierarchy of a node $i$ as $h(i)$, and the definition is as follows:

$$h(i) = \frac{\sum_{j \in N(i)} \left( \frac{c_{ij}}{C/N} \right) \ln \left( \frac{c_{ij}}{C/N} \right)}{N \ln(N)} \tag{3}$$

where $N$ is the number of contacts in node $i$'s network, $c_{ij}$ is the constraint between node $i$ and contact $j$, and $C$ is the sum of constraint across all $N$ contacts. The measure $h(i)$ that equals 0.0 indicates that the constraint is the same for $i$'s relationship with each neighbor, while $h(i)$ that equals 1.0 indicates that all the constraint is concentrated in a single contact.

Overall, the redundancy measures, such as effective size, mainly capture the features of connections. While the constraint measure is derived from the concept of dependence, indicated by exclusive access. These measures can help us better discover and evaluate SHs.

### C. Structural Hole Theory, Triads, and Network Oscillation

Recent studies have also demonstrated the applicability of SH theory in combination with other social science theories in social networks, validating the robustness of the theory. Here, we present two examples. One applied SH theory to study triads [14] and the other built a new theory called network oscillation on top of SH theory [15].

A triad refers to a group of three people. It is one of the simplest forms of human groups and forms the most basic structures of sociological analysis. Triads can be closed ones, where any two persons are connected, or open ones, where two of the three people are unconnected. The problem of triadic closure process, how a closed triad develops from an open one, is fundamental in the evolvement of dynamic networks. This mechanism is of particular interest to researchers and has application in sociology as well as computer science.

In their study, Huang *et al.* [14] adopted the SH theory along with a series of other metrics to analyze the triadic closure patterns of the users. They tested whether occupying SH positions would affect the triadic closure pattern. The results indicated that the existence of SH spanners in the two unconnected users greatly increases the triadic closure probability, while the middle person being an SH spanner is linked with a lower closure possibility. For illustration, suppose that A and C are both connected to B, whereas A and C are initially unconnected. If one of A and C occupies an SH position, they are over ten times more likely to get connected to gain social resources. However, if B is an SH spanner, the open triad is less likely to become closed when compared with ordinary users, as A and C are likely to be in separate communities and B may also be reluctant to connect them and lose the network advantage. Based on the observation, they integrated SH spanning in the model they proposed for triadic closure prediction, among other valuable network properties. In the experiment, their TriadFG model achieved a much better prediction performance than the benchmark algorithms.

While triad is a well-studied concept, network oscillation is a newly proposed theory based on the SH theory. Network structures are known to be related with particular advantages, and SH spanners are believed to enjoy information diversity, timing, and arbitrage advantages. Furthermore, Burt and Merluzzi [15] suggested that the evolvement of the network over time also influences the advantages it provides. Among different dimensions of network volatility, they found the oscillation between closure and brokerage highly related to network advantages. In other words, network oscillation means alternating between deep involvement in a community and connecting across different communities. By analyzing a group of investment bankers in a financial organization, they found that oscillation strongly enhanced the advantage brought by SH positions. They also gave out three possible mechanisms behind this relation, which were left for future work to verify. This work combined SH theory with other social science theories and constructed a new concept on top of them. The findings also shed light on a new understanding of how to build networks that provide advantages.

## III. Structural Hole Spanner Detection

In this section, we review the SH spanner detection algorithms. SH theory shows that SH spanners are positively related to social success by bridging different communities. Considering the impact of SH spanners on social networks through both the information and control advantages, it is essential to detect SH spanners in social networks. SH spanners cannot be directly inferred only from the relationships between nonredundant contacts. In order to better understand the problem of detecting top-$k$ SH spanners, we classify the algorithms into two categories. In Section III-A, we study the information flow-based algorithms. Some of these algorithms focus on discovering key nodes that expand the scope of information dissemination [7], [10], [33], [34], and some focus on discovering key nodes whose removal will maximally cut off information propagation [7], [9]. In Section III-B, we study the network centrality-based algorithms [11]–[13], [35], focusing more on whether a node occupies the advantageous positions in network structures. All these algorithms are summarized in Table I. We highlight open issues and directions in Section III-C.

### *A. Information Flow-Based Algorithms*

In this section, we mainly review the problem of top-$k$ SH spanner detection from the aspect of information propagation. There are algorithms of expanding information transmission, as well as methods to maximally cut off information dissemination to find SH spanners.

Considering the reverse process of strategic network formation with SHs [37], Lou and Tang [7] proposed two models to tackle the task of mining top-$k$ SH spanners in large-scale social networks, assuming that community divisions were given. They mainly studied three different social networks: coauthor network, which contains co-authorships of papers published in 28 major conferences of computer science; Twitter social network, which is a widely used microblogging system, containing relationships of following and being followed; inventor network, which is a network of inventors and contains co-inventing relationships. Based on the assumption that SH spanners are more inclined to build relationships with opinion leaders in different communities, to whom ideas usually flow first, they designed the first model named HIS. By defining the importance score and the SH score of one node, they quantified the importance and influence of each node who plays the roles of both SH spanner and opinion leader. Also, the model was derived from the intuition that getting a higher score means a user can receive more information flows from her neighbors. Besides, its convergence was also proved. Focusing on information diffusion, the second model named MaxD was designed by computing the minimal cut of a network, which maximized the decrease of minimal cut after removing $k$ nodes to get the top-$k$ SH spanners. Considering the NP-hardness of the minimal cut optimization, they adopted an approximate algorithm to achieve the goal of maximizing the decrease of the minimal cut as much as possible. In addition, these SH spanner detection models could make great improvements in community kernel detection [38] and link prediction [39]. Based on the intuition that SH spanners tend to connect with kernel members of different communities, the former (community kernel detection) aims to detect kernel members of different communities with the highest important scores, where WeBA [38] algorithm performed better. The latter (link prediction) is to predict the types of social relationships with the help of SH analysis, where partially labeled factor graph (PFG) [39] algorithm performed better. The main challenge is that sometimes we cannot get community labels in advance, which makes the data analysis process more difficult.

Since both community detection and SH spanner detection need the topological structure in a network, the two tasks can be put together and considered simultaneously. He *et al.* [10] applied a harmonic function to jointly detect community and SH spanners based on the topological nature between them, and the detecting scheme was named harmonic modularity (HAM). They first proposed the harmonic function, where the score of one node is defined as the average score of its neighbor nodes. Based on the harmonic analysis, they also introduced the $l_{2,1}$-norm penalty and orthogonality constraint to detect SH spanners and communities more effectively. They dealt with the optimization problem by both measuring the smoothness of community structure and distinguishing SH spanners simultaneously through matrix operations and then proved its convergence and computational complexity. They also defined a new metric, known as the structural hole influence index (SHII) to evaluate information diffusion leveraging linear threshold model and independent cascade model [40]. The SHII considers the proportion of affected nodes outside the community to all affected nodes in the process of information dissemination so that it can distinguish SH spanners from center nodes within the community. They achieved high performance in both SH spanner detection and community detection tasks in both synthesized and real-world data sets, including DBLP (a co-authorship network), Karate Club (a friendship network in a karate club), and YouTube (a video-based social network). The performance was measured by diverse indicators, such as accuracy (ACC), normalized mutual information (NMI), and average cluster entropy (ACE). The deficiency of this algorithm is that the time complexity is difficult to be reduced and the scale of calculation is hard to be expanded since the algorithm is based on the matrix computation process.

Another study related to SH spanner detection worth mentioning is a spectral graph embedding method proposed by Jiang *et al.* [34]. Based on their idea that a good graph embedding algorithm should preserve both macroscopic structures (such as community structure) and microscopic structures (such as SH spanners), they designed a spectral framework NOn-Backtracking Embedding (NOBE) and its graph approximation algorithm NOBE-GA. NOBE makes use of spectral graph embedding technique and nonbacktracking random walk to jointly capture both community and SH spanner structures. The main idea of their method is to transform the original graph to an oriented line graph by converting each edge into a node and defining a nonbacktracking transition matrix. The advantage of the oriented line graph is that SH spanners would be placed into critical positions with more concentrated edges' weights so that the confidence of being SH spanners will be higher. They then embedded nodes in the oriented line graph by minimizing the loss functions, with the consideration of Rayleigh quotient to tackle the objective following the idea of [41]. The main process of NOBE is divided into two steps: eigenvector decomposition and summation of embeddings of the incoming edges. Furthermore, they presented an eigenvector decomposition algorithm on a smaller scale matrix with provable approximation guarantees. The performance of their method showed that it could tackle both clustering and SH spanner detection tasks well. In SH spanner detection, they used SHII proposed in [10] as an evaluation metric and also proposed a metric called relative deviation score to measure node rankings in the embedded subspace. Their model outperformed several SH spanner detection algorithms, including HIS, AP_BICC, and HAM under linear threshold model and independent cascade model [40] on Karate, YouTube, and DBLP data sets.

TABLE I
SH SPANNER DETECTION ALGORITHMS AND THEIR PROS AND CONS

| Category | Reference | Year | Method | Key Idea | Pros and Cons |
|---|---|---|---|---|---|
| Information Flow-Based Algorithms | [7] | 2013 | HIS<br>MaxD | SH spanners are more inclined to build relationships with opinion leaders in different communities, to whom ideas usually flow first | Require community labels in advance |
| | [10] | 2016 | HAM | The authors proposed the harmonic function, where the score of one node is defined as the average score of its neighbor nodes | Jointly detect community and SH spanners; hard to apply HAM to large-scale social networks |
| | [34] | 2018 | NOBE<br>NOBE-GA | Both macroscopic structures (such as community structure) and microscopic structures (such as SH spanners) must be preserved at the same time for a good graph embedding | Jointly detect community and SH spanners; relatively less interpretable than other algorithms |
| | [9] | 2019 | maxBlock<br>maxBlockFast | The authors identified the most important nodes whose removal would block the maximum information flows | Faster and less computational cost and the detected SH spanners block more than 24% of the information propagations |
| | [33] | 2019 | ESH | The authors introduced an entropy-based method from a factor diffusion process, and applied distributed parallel computing for efficient calculation | Achieve faster performance on large-scale networks since it applies distributed parallel computing |
| Network Centrality-Based Algorithms | [11] | 2015 | WeakTie-Local<br>WeakTie-Bi | Weak ties [8] are important in the novel information dissemination between remote users [36]. Users with many weak ties are more likely to fill the structural holes | Focus on dynamic social networks as well as remote information sources; only rely on the topological structure of the network |
| | [12] | 2015 | BICC<br>AP_BICC | The average shortest path of a graph will increase significantly when structural holes are removed | Efficiently calculate the proposed approximate inverse closeness centrality for each node in the articulation point set; only rely on the topological structure of the network |
| | [13] | 2017 | Greedy<br>AP_Greedy | The authors estimated the upper bound of the function which maximizes the average distance of the graph after removing the node to filter out unlikely solution earlier | Further optimize the accuracy of the AP_BICC algorithm and increase its efficiency; only rely on the topological structure of the network |
| | [35] | 2020 | FSBCDM | SH spanners can change the current trend of community expansion and belong to multiple communities | Jointly detect community and SH spanners; relatively higher computational complexity than other algorithms except HAM |

Motivated by the discovery mentioned in [27] that different SH spanners show significantly different performance and returns, Xu *et al.* [9] devised a randomized algorithm as well as a fast, scalable, and heuristic algorithm, called maxBlock and maxBlockFast. The basic idea came from the findings by Burt [42] that one person not only builds more bridges between otherwise unconnected communities to gain more rewards but also needs to establish strong ties with her connected communities. Therefore, maxBlock dealt with the problem of top-$k$ SH spanner detection by identifying the most important nodes whose removal would block the maximum information flows, with consideration of tie strength of nodes in social networks. Xu *et al.* adopted the independent cascade model [40] to calculate the probability of information diffusion. Besides, they considered an approximate solution to block more information propagations. Due to the challenge of calculation in large-scale networks, they then proposed a fast heuristic algorithm called maxBlockFast, fully using the property of the dominator tree derived from the live-edge graph model through Monte Carlo simulations. Through extensive and complete experiments, the detected SH spanners can better disseminate information compared with existing algorithms. Moreover, the heuristic algorithm achieved shorter time cost without much loss of ACC. They found that the detected SH spanners block more than 24% of the information propagations.

Some existing methods [7], [10], [12] face huge challenges in large-scale networks, limited by the computational complexity of the algorithms. In order to achieve faster performance on large-scale graphs with billions of nodes and edges, Li *et al.* [33] proposed an SH detection algorithm named ESH based on a distributed parallel graph processing framework called PowerGraph. Unlike previous methods focused on the community structure, their proposed model introduced an entropy-based method from a factor diffusion process and applied distributed parallel computing. Since an SH spanner is more likely to collect more diverse factors diffused from different communities, the entropy of its factors tends to be higher. Meanwhile, an interior node in a community is likely to collect homogeneous factors from other nodes within the same community, resulting in low entropy. Therefore, they tried to distinguish SH spanners by evaluating the likelihood through the entropy of the factor distribution process. They then conducted several experiments by comparing three different methods. The results showed that ESH is a good choice for SH detection for its capability of dealing with large-scale networks with billions of nodes and edges.

### *B. Network Centrality-Based Algorithms*

Different from the SH spanner detection algorithms that concentrate on information dissemination in Section III-A, the algorithms discussed in this section mainly detect SH spanners from the centrality aspect of network structures, considering whether the nodes occupy advantageous positions in social networks.

Song *et al.* [11] developed a heuristic algorithm to detect top-$k$ brokers based on the weak tie theory and mainly focused on dynamic social networks, as well as remote information sources. Similar to the previous methods, they also defined the problem of top-$k$ broker detection and showed its NP-hardness by reducing it to the $k$-densest subgraph problem. The novel idea of their model is derived from the evidence that weak ties [8] are important in the novel information dissemination between remote users [36]. Users with many weak ties are more likely to fill SHs between remote users. Therefore, the authors first devised some preliminary experiments to show that brokers are correlated with the number of weak ties. They also defined the tie strength of an edge, leading to the definition of path strength. Then, the strongly connected groups have been calculated by applying Tarjan's algorithm [43], given the threshold of path strength. Furthermore, they proposed incremental algorithms to deal with the high complexity and handle the dynamic nature of social networks. The WeakTie-Local algorithm addresses unidirectional network problems, while the WeakTie-Bi algorithm addresses bidirectional network problems. They also enumerated several possible situations after inserting or deleting an edge dynamically and gave corresponding solutions. They compared the performance of their proposed algorithms with the existing methods, such as PageRank and betweenness centrality (BC), to validate the effectiveness on both DBLP and Twitter data sets. In addition, they conducted several experiments to measure the sensitivity and scalability of their algorithms. As for applications, they discussed that top-$k$ brokers can also be used to mention recommendations to expand the spread of tweets.

Rezvani *et al.* [12] came up with the new concept of inverse closeness centrality to tackle the problem of top-$k$ SH spanner detection. The main idea is that when SHs are removed, the average shortest path of the induced subgraph will increase significantly. Therefore, they first devised a basic algorithm called ICC. Due to the high time complexity of this algorithm, they then improved it and developed an efficient algorithm BICC, namely the bounded inverse closeness centrality, inspired by the sparsity and the small world law. In addition, they designed a more accurate algorithm by considering both bounded inverse closeness centrality of vertices and articulation points (APs) of the network, namely AP_BICC algorithm. Here, the APs are referred to these SH spanners, with the instinct property that tends to connect multiple isolated communities. They also showed the process of finding APs. Furthermore, using a Depth-First Search traversal on the graph, they efficiently calculated the proposed approximate inverse closeness centrality for each node in the AP set.

Xu *et al.* [13] further optimized the ACC of the AP_BICC algorithm and improved its efficiency. Comparing with the former work [12], they supplemented a detailed proof of the NP-hardness of top-$k$ SH spanner problem and proposed two novel algorithms, i.e., Greedy and AP_Greedy, which can filter out unlikely solutions earlier. As for the greedy algorithm, in each iteration, it identifies a node from the graph by estimating the upper bound of the function, which maximizes the average distance of the graph after removing the node. Besides, the filtering techniques are based on the consideration of APs. After defining the quality of solution based on the finding by Burt [6] that an influential SH spanner has a larger ratio of the number of ego's communities to the number of its neighbors, the novel algorithms showed much better performance in the SH spanner detection task than those existing ones, such as AP_BICC, Central, PathCount, 2-Step, PageRank, Constraint, HAM, HIS, and MaxD. They also confirmed that the model can well capture the features related to SH spanners.

Since the above models in this subsection are based on the average distance of a social network for detecting SH spanners, removing SH spanners will maximally increase the mean distance of vertices in the residual network. Therefore, one of the distinguishing features worth mentioning is that their models only rely on the topological structure of the network, without given community labels.

Zhang *et al.* [35] recently proposed a new algorithm named FSBCDM to discover SH spanners based on diminishing marginal utility and the previously designed community forest model [44]. According to the community forest model, a community gradually increases from the core backbone, and the expansion of the community becomes smaller as more nodes join the community, which complies with the law of diminishing marginal utility. In the context of community reconstruction, FSBCDM continues to add neighboring nodes with the maximum sum of backbone degree centrality. Some nodes break the law of diminishing marginal utility, making them the first type of SH spanners since they can change the current trend of community expansion. Another type of SH spanners is defined as the nodes at the intersections between communities, and they belong to multiple communities. The FSBCDM algorithm sorts the above two types of SH spanners according to the SHII metric proposed in [10] to find the final top-$k$ SH spanners. The authors confirmed that FSBCDM performed slightly better than HIS, MaxD, AP_BICC, and HAM.

### *C. Discussion and Future Expectations for Structural Hole Spanner Detection*

The problem of SH spanner detection is derived from the reverse process of strategic network formation with SHs. Kleinberg *et al.* [37] designed a dynamic and strategic game to fill the SHs and form links to bridge previously unconnected communities in a social network, with the aim of studying the process of network formation with SHs. Ghaffar and Hurley [45] also derived a new centrality measure called SH centrality, to recognize actors with high social capital in the process of strategic network formation. However, considering the reverse direction of their research, it is still significant to detect who plays the role of SH spanner in large social networks, which leads to all the aforementioned SH spanner detection algorithms introduced in this section.

The SH spanner detection algorithms mentioned above have their own pros and cons. First, HIS and MaxD need to know the ground truth community labels in advance, which will make the data availability conditions more challenging. From the perspective of time complexity, the ESH can achieve faster performance on large-scale networks since it applies distributed parallel computing. The time complexity of AP_BICC, AP_Greedy, HIS, and MaxD is less than HAM and FSBCDM, as the former algorithms (AP_BICC, AP_Greedy, HIS, and MaxD) only detect SH spanners, whereas the latter ones (HAM and FSBCDM) detect both communities and SH spanners. Since the time complexity of HAM is hard to be reduced due to the matrix computation process, it is difficult to apply HAM to large-scale social networks. As for interpretability, the spectral graph embedding methods, such as NOBE or NOBE-GA, are relatively less interpretable than other algorithms.

The current research on SH spanner detection is mostly on static networks. However, many real-world networks, especially social networks, are not static but constantly changing. Individuals keep developing their connections with others in social networks, while the models neglecting the dynamics in a social network will fail to capture sufficient and reliable information. Different evolution patterns of individuals in social networks may reflect their personality traits, as well as changes in their social statuses and influence in dynamic networks. Similar to dynamic network embedding [46], future works of SH spanner detection may consider proposing new methods in dynamic networks. A particularly interesting direction for future works on SH spanner detection is to extend static SH spanner detection models to dynamic settings, with the definitions of related functions for updating other nodes after dynamically adding or removing nodes or edges in a social network. As the triadic closure process is a basic unit during the evolution of networks, we may introduce triads to help measure the dynamic changes of network structures, such as the proposed model DynamicTriad in [47]. Furthermore, we can also incorporate the dynamic representations of nodes on dynamic social networks to detect SH spanners, which can provide more sufficient and real-time features for SH spanner detection.

## IV. Structural Hole-Related Applications in Practical Fields

In this section, we will present the development of applications of SH theory in practical settings, while Section V will be dedicated to introduce the applications in ML-based social prediction. All the applications discussed are summarized in Table II. We organize this section based on the fields of the applications. In Section IV-A, we present the applications of SH theory in enterprise settings. Next, we introduce applications in information diffusion in Section IV-B. In Section IV-C, we review the applications of the SH theory in the field of software and mobile application development. Finally, Section IV-D discusses the trends of applications of SH theory and makes anticipation for future studies.

### A. Applications in Enterprise Settings

In enterprise settings, SH theory offers a new perspective to understand how social connections affect employees' well beings and their decisions, as well as the mechanisms lying behind. Taking SH spanners into consideration provides valuable insights in analyzing the performance of employees and their collaborative behaviors. Such insights can further provide individuals with advice on how to develop their social networks.

*1) Performance and Innovativeness:* Studies in SH theory have suggested a robust relationship between SH positions and higher managerial performance. According to Burt *et al.* [60], managers occupying SH positions are generally evaluated to have better performance than those whose connections are densely interconnected. These managers also tend to be more highly paid, promoted faster, and more likely to be recognized as leaders. Rodan [48] analyzed the underlying mechanisms behind this relationship and identified innovativeness as the key factor. Moreover, Wang *et al.* [18] found that leader–member exchange is important when out-group weak ties contribute to innovative behaviors. In another study, Choi and Lee [17] observed a positive relationship between the inequality in the level of SHs between group members and the innovative performance of the group. Also, Ye *et al.* [16] leveraged the SH theory to identify new employees with high potential.

Rodan [48] conducted a survey-based study to analyze the driving mechanisms of the widely observed higher managerial performance achieved by SH spanners. Among the five theoretical mechanisms suggested by the SH theory: autonomy, competition, information brokering, opportunity recognition, and innovativeness, he found innovativeness playing the key role in mediating network structures and better performance, rather than information brokering. As SH spanners enjoy positional advantages by playing a bridging role in their social networks, they get access to nonredundant and diverse information and perspectives, so they can leverage advantages among multiple groups and are more likely to develop creative new ideas, which then contribute to better evaluate the performance.

As weak tie theory describes a similar phenomenon with SH theory, a work revealing the mechanism why out-group weak ties contribute to higher innovation performance is also worth mentioning here. In the study, Wang *et al.* [18] conducted an investigation on data collected from a high-tech firm. Their result indicated that a special kind of strong tie, leader–member exchange, greatly contributes to the positive relationship between the existence of weak ties connecting different groups and high innovation performance. Such a finding can help expand the theory by revealing the underlying mechanism through which SHs may contribute to innovative behaviors.

TABLE II
APPLICATIONS OF THE SH THEORY

| Field | Reference | Description |
|---|---|---|
| Enterprise settings | [48] | Underlying mechanism of higher managerial performance achieved by SH spanners |
| | [18] | Contribution of leader-member exchange to the positive relationship between out-group weak ties and higher innovation performance |
| | [17] | Inequality in level of structural holes between group members contribute to the innovative performance of the group |
| | [19] | Structural holes' effect of on a country's innovative performance |
| | [16] | Identification of high potential talents from newly-enrolled employees of a company |
| | [49] | How people use network information to choose new collaborators |
| | [50] | Recommendation of possible weak ties to employees in an organization |
| Information diffusion | [21] | Influence maximization (IM), maximization of the influence spread in an online social graph |
| | [22] | Social-role aware emotion contagion for emotion prediction |
| Software development and mobile applications | [24] | Analyzing requirement identification from the perspective of stakeholders' network structures |
| | [26] | Relationship between application service categories as well as individual apps |
| Individual-level ML tasks | [16] | Identification of high potential talents from newly-enrolled employees of a company |
| | [51] | Self-disclosure behaviors of OSN users |
| | [52] | Discriminating functional users from social users |
| | [22], [53] | Predicting users' emotional statuses utilizing social structures |
| | [54] | Predicting future tourist arrival of countries / regions using social network predictors |
| | [55] | Proposing an improved network constraint coefficient for measuring importance of nodes |
| Connection-level ML tasks | [50] | Recommendation of possible weak ties to employees in an organization |
| | [56] | Predicting future co-investment behaviors of venture capitals |
| | [57] | Proposing a sampling-based method to find the most similar nodes |
| | [58], [59] | Classifying different types of social ties |

Choi and Lee [17] also studied the relationship between SHs and innovative performance, but on a group level. They collected data from ten teams of an international company and adopted NBD[1] of patented products to quantitatively measure the innovation performance of the teams. Their results showed that the inequality in the level of SHs in a team increases over time, having a positive effect on the innovation performance. In addition, the results demonstrated that higher job grades were related to greater SHs, and people in the sales group were observed to have higher SH values than other job functionalities. Such findings can be inspiring to future research and applications in the business field. While prior studies had only covered descriptive analyses of social networks, Choi and Lee collected data from an enterprise and validated the applicability of SH theory in promoting innovation performance in NBDs.

Furthermore, Choi and Zo [19] approached the problem at a national level. They analyzed the characteristics of the global knowledge spillover network, where countries are viewed as nodes instead of humans, and found a positive effect of SHs on a country's innovative performance.

The studies mentioned above validate the advantage gained by spanning an SH in business social networks, contributing valuable insights to companies when it comes to evaluating and cultivating employees as well as managers. Uncovering the mechanism behind this advantage offers further understanding and foresees various applications in human resource management. For example, Ye *et al.* [16] adopted the theory to identify high potential talents (HIPOs) from newly enrolled employees of a company. The details of the model would be further introduced in Section V. Besides, these findings provide insights on how to gain network resources to achieve advantages and improve performance, for not only individuals but also organizations and countries.

*2) New Connections and Collaboration:* Aside from individual performance, SH theory also casts light on the establishment of new connections and collaborative relationships within the social network of an enterprise. Gao *et al.* [49] adopted the SH theory to understand how people select new collaborators based on network information, while Ghaffar *et al.* [50] aimed to recommend potential weak ties to employees to help them gain network advantages.

Collaboration is universal within organizations and is central to success. As organizations have become more distributed and teams tend to change more frequently, individuals are constantly facing the choice of selecting new collaborators. To understand how people leverage network information in this decision-making process, Gao *et al.* [49] conducted a scenario-based survey on the U.S. and Chinese employees of a global company and studied the difference in the strategies they adopted in collaboration seeking. Their findings

[1]New business development (NBD) is a key innovation index of healthy business sustainability.

suggested that Chinese respondents more closely followed the closure model, favoring candidates with shared contacts. In contrast, U.S. respondents partially followed the structure holes model in that they valued all kinds of resources equally, but did not adhere to it in their response to shared contacts. A surprising finding was that neither group of respondents entirely followed the SH theory, which had been supported by multiple prior studies in U.S. organizations. This may be explained by the fact that not much research on SHs had studied decision-making based on the social network information, so they also pointed out the need for further study to compare online versus offline collaboration choices. This work indicated the need for incorporating other theories as supplements to the SH theory when adopted to model people's behaviors in practical settings.

In addition, as both SH theory and strength of weak ties theory indicate that bridging roles and weak ties can provide individuals with valuable resources related to success, building new weak ties in the network is believed to be beneficial to individuals. Based on such inspiration, Ghaffar *et al.* [50] recommended possible weak ties to individuals to help them gain benefits in information access.

### B. Applications in Information Diffusion in Social Networks

Since SH spanners occupy crucial structural positions connecting different communities, they are believed and observed to play a key role in spreading information in social networks, especially between communities. Researchers have conducted experiments to testify this belief and employed the concept to maximize influence.

First, SH theory suggests that SH spanners are crucial in the diffusion of information and also generate information benefits both for the whole network and for themselves. According to Lou and Tang [7], 1% of users occupying the SH positions control almost 80% of the information spread between communities and 25% of all the information diffusion in Twitter. Such intuition was then utilized by them to effectively identify the top-$k$ SH spanners in large social networks, which has been introduced in Section III. Moreover, a survey-based study by Fritsch and Kauffeld-Monz [20] verified that SH spanners contributed to the subnetworks they help to connect by transferring information. Such findings fuel the application of SH theory in the field of influence maximization (IM).

IM tackles the problem of identifying a set of seed nodes to maximize the influence spread in a social network. The problem used to be costly and inefficient using conventional greedy algorithms that fail to include network structure features. To this end, Zhu *et al.* [21] proposed a structure-hole-based algorithm (SHIM) for IM in large OSNs. The intuition is that while opinion leaders are important for information diffusion within communities, SH spanners play a central role in spreading information between different communities. By first identifying the SH spanners and then choosing those with high values in both SH and influence measures as seeds, the scale of the problem can be greatly reduced by filtering out the non-SH users. They proposed an algorithm called structure hole value calculate (SHVC) to compute the structure hole values for users and identify SH spanners. In the experiment, they reported that their proposed detecting method was the fastest one compared with the existing algorithms, such as HIS and MaxD. By further combining SH values and influence values, their IM algorithm performed the fastest among all compared methods and yielded the most influential results as well.

Yang *et al.* [22] further extended the study by studying the role of SH spanners in emotion contagion. They investigated how users' positions in the social network affected their influence on emotional statuses of their friends. Surprisingly, their results showed that in emotion contagion, SH spanners may be less influential than ordinary users on the whole, different from what was observed in information diffusion [6], [7]. They also found users with social roles of opinion leaders and SH spanners more influential than ordinary users in positive emotion contagion while less influential in negative emotion contagion. Based on the observation, they proposed a new model to predict users' emotional statuses utilizing social structures, learning the influence strength between friends by considering their social roles. Their model achieved a strikingly high improvement compared with methods that do not consider correlation features.

### C. Applications in Software Development and Mobile Applications

Softwares are now everywhere in people's daily life. In recent decades, we have witnessed the fast growth of mobile application services. As a result, SH theory has also been adopted to understand the requirement identification process of software development, as well as interrelationships between different mobile service sectors.

Requirement engineering (RE) is the process of determining, documenting, and maintaining requirements in the engineering design process. It is a common role in systems engineering and software engineering. Requirements identification is one of the major objectives in the RE process, involving multiple stakeholders. As a result, stakeholders' interactions have been proved a crucial factor in success. Bhowmik *et al.* [24] explored the SH theory to analyze requirement identification from the perspective of stakeholders' network structures. As the SH theory suggests that SH spanners can produce new ideas, they verified in their analysis that stakeholders occupying SH positions did contribute a greater number of new requirements compared with ordinary users. Nevertheless, they found some exceptional cases such as project leaders, who were commonly found on SH positions but typically did not contribute many new requirements. As a result, they demonstrated the need to take into consideration the roles of stakeholders. By mapping people in a social network to stakeholders and new ideas to new requirements, they confirmed the applicability and effectiveness of the SH theory in RE. They further suggested the need of modifying SH theory to adapt to the RE process by considering the stakeholders' roles. This also provided valuable insights for future studies or potential applications of the theory in new fields.

Mobile apps have also become a fundamental part of people's lives in recent decades. Kim *et al.* [26] studied the relationship between application service categories as well as individual apps by constructing networks based on keywords similarity. They adopted SH theory in analyzing such interrelationships. They calculated the BC value as the quantitative index indicating the position and influence of a category or an individual app. In this way, they took into consideration apps' importance as brokers. In terms of the category network, they found business and travel categories have high BC values and suggested that they may play intermediary roles in the network. They also concluded that the relative importance of a category does not rely on its scale. They then constructed the micro app networks of each category and grouped categories based on the network characteristics. One of the clusters, including entertainment and utilities, is observed to have low network density and concentration but high node centrality values, especially BC values. These categories are believed to mediate the flow of information between the remaining clusters.

### D. Discussion and Possible Future Expectations

Recent works have proved the applicability of SH theory in various fields, and we are convinced that the theory will be employed in more fields in the future. Also, the applications in existing fields will continue to evolve over time. We anticipate that the observations reported by prior studies will be put into more practical uses in the real world. For example, the observed relationship between SH spanners and higher performance and innovation can be leveraged to spot employees with prospective high performance or the ones playing a crucial role in the organization, just as Ye *et al.* [16] did in their work. The key role SHs play in spreading information also inspired the new IM algorithm [21].

Besides, structure hole positions can help analyze the dynamic evolvement of network structures, which is common in real settings. Choi and Lee [17] have explored this by analyzing the inequality in the level of SHs between team members. How SH positions emerge, change, and disappear can also help reveal the nature of a network.

Meanwhile, we also expect more applications in networks where nodes are nonhuman, such as in [19], [26], since the observation in social networks can also hold true when the nodes become organizations, countries, or even items.

## V. Structural Hole Theory-Aware Machine Learning

ML is an important aspect of modern business and research and has been known as an essential tool in analyzing networks. Researchers have developed numerous ML models for purposes, such as generating user embeddings, classifying users and connections, and identifying potential connections. SH theory has also been integrated into many models as a fueling social theory, helping obtain useful results.

Basically, ML-based applications in social networks can be divided into individual-level tasks and connection-level tasks. In individual-level tasks, in order to understand individual behaviors and properties, ML-based applications of SH theory are implemented in both enterprise settings [16] and OSNs [51], [52], [61], as well as leveraged by researchers to predict user emotions in image social networks [22], [53]. In connection-level tasks, related works mainly focus on link prediction and edge classification. Link prediction aims to predict future links based on the current network structure, such as recommending possible weak ties to employees in an organization [50] and predicting future co-investment behaviors [56]. The edge classification task is also beneficial since different types of social ties tend to influence people differently [58], [59]. With the help of SH properties used in ML methods, social theory-based features are of crucial contribution to the prediction performance.

In this section, we will present several recent works adopting the SH theory in the field of ML-based social prediction. In Section V-A, we focus on individual-level tasks, concentrating on node properties and behaviors, such as node classification. In Section V-B, we introduce works concentrating on connection-level tasks, which care about distinguishing connection types as well as link prediction problems. After that, we discuss the trends and our expectations for future studies adopting the SH theory for ML in Section V-C.

### A. Individual-Level Tasks: Analyzing Individual Behaviors and Properties

As one of the major node properties to be considered when investigating social networks, SH property has been employed as a fundamental measure to analyze social graphs to understand users' behaviors. Studies have also found a relation between SH positions and user types or user behaviors in OSNs. By paying attention to users' structural positions, we can better understand and predict some of these properties. The following works mainly formulated the problem as a node classification task.

To begin with, Ye *et al.* [16] adopted the SH theory to identify high potential talents (HIPOs). HIPOs refer to employees who are believed to have high competency and are regarded as potential future leaders. Therefore, identifying the HIPOs among new employees is a central concern in human resource management. However, the current identifying method relying on manual selection is subjective and prone to bias. Driven by such a motivation, Ye *et al.* proposed an ML framework to identify HIPO employees by modeling their behaviors in social networks within the organization, inspired by the intuition that HIPOs usually manage to gain more social capital than average employees. They modeled the social capital of employees, combining both the local and global social information, with the global social information designed to indicate whether an employee plays a crucial role in the network (for example, spanning an SH). They adopted nine metrics in total reflecting node centralities, including network constraint and BC fueled by the SH theory. They then feed the representation result along with ego network representation generated by GCN[2] to an LSTM model with a global attention mechanism. Altogether, their model achieved significantly better results than all the benchmark methods. Also, their experiment validates the contribution by both the local and the global social network information.

On OSNs, users can decide whether or not to expose their personal information to the public. Intending to understand users' self-disclosing behaviors and their relations with the overall network structure, Kwon *et al.* [51] conducted a study at both ego networks and community networks, and they presented the possible relation between the self-disclosure behaviors of users and the SH theory. For all the possible network advantages mentioned in Section II, it is believed that SH spanners may be more likely to disclose their personal information on the platform to leverage these advantages by playing the bridging role. On the local scale, they first analyzed users' self-disclosure behaviors with regard to their ego networks and found a significant positive relationship between the openness of a user and the effective size of her network, which is an indicative measure of SH spanners. They then moved to the community scale and found that open users have significantly higher BC. These results all confirmed the essential role SHs play in the self-disclosure behaviors of users. Finally, they conceived a task to predict the self-disclosure levels of users using only network properties and chose random forest (RF) as a classifier, which could calculate the importance of each feature. Their model achieved a significant improvement of performance compared with the benchmark by 12% in the F1-score, and in the model, they identified BC and effective size among the top three most important features.

Nowadays, OSN platforms have seen the emergence of functional users, such as online business runners. These users tend to have more friends, often sparsely connected with each other, and they have not been studied much by existing studies. To distinguish these functional users from social users, Ying *et al.* [52] adopted concepts from the SH theory to develop metrics for measuring the diversity of a user's friend circle in their model. Their work was based on the notion that functional users tend to have a more diverse ego network, so they adopted effective size from SH theory as a measure to indicate the diversity of a user's friend circle. They were then inspired to give out a more general definition of ego network diversity, with actual size and effective size being two special cases of the broad definition. By further changing the factors in the definition, they proposed two new diversity metrics. They formulated a binary classification problem discriminating functional users from social users to evaluate the performance of the four measures. Their model achieved the best performance with the measure expected number of communities (ENC) in all evaluation metrics. Their work sheds light on the SH theory by introducing ENC as a new metric to measure SHs, which may be of inspiration to future researchers who are faced with a similar situation where access to the whole graph is not granted.

Moreover, the SH theory is also leveraged by researchers to help predict user emotions in image social networks. Founded on the investigation on emotion contagion, which has been introduced in Section IV-B, Yang *et al.* [22] proposed a new model to predict users' emotional statuses utilizing social structures. While previous attempts to predict users' emotional statuses treated individuals independently and failed to consider the interactions among them, they improved by taking the social role-aware emotion contagion into consideration. Besides learning from historical emotions and images posted, their model was also designed with the ability to learn the influence strength between friends by considering their social roles. They combined all the three aspects in a factor graph model as three different layers. By taking social roles into consideration, they let users with the same social roles share the same parameters in the influence model to reduce the complexity and improved the practicability of the model. In the experiment, the proposed model achieved a 44.3% improvement on average compared with methods that do not consider correlation features.

Cai *et al.* [53] also aimed at inferring emotion based on image social network information. They leveraged group information in their proposed factor graph model. In particular, they found out that groups with a higher ratio of SH spanners also exhibited a higher emotion homophily. The observation suggested that potential positive effect SH spanners have on emotion diffusion and the metric was adopted to construct group features. By combining image content, user personalization, and group information, their model achieved the best performance among all their compared methods. Though the observation drawn from two different data sets by the two studies might not seem consistent, they both suggested that SH spanners play a distinct role in spreading emotions. Also, their result indicated that taking into consideration such roles makes a great contribution to the prediction models, just as it did with all the other studies mentioned.

While humans are the examined social entities in most existing models, Yuan [54] adopted a tourism social network, where countries/regions are actors such as in [19] introduced in Section IV-A. It was the first work to leverage social network information, namely degree centrality and SHs measured by network constraint, as predictors to predict future tourist arrivals. The result of the LSSVR model[3] showed that the use of social network predictors achieved better performance than traditional economic predictors.

Network constraint is a widely used SH indicator in ML methods. However, Lu [55] pointed out that many nodes in the networks, especially in power networks, lack a triangular structure, making it difficult to calculate. Thus, Lu proposed an improved network constraint coefficient, taking the influence of neighborhood nodes into consideration. Lu then combined the prospect theory [62] with improved TOPSIS, a multiattribute decision-making method, to evaluate the importance of network nodes. Lu finally confirmed the effectiveness of

[2]Graph convolutional network (GCN) and long short-term memory (LSTM) are both neural network models in ML.

[3]Least-squares support-vector machines, a version of support-vector machines (SVMs).

the SH-based indicators with experiments, in order to provide inspiration to future studies.

### *B. Connection-Level Tasks: Link Prediction and Classification*

Link prediction aims to predict future links based on the current network structure or infer the missing links from a partial network. It is a fundamental problem in network science with abundant real-world applications. As mentioned in Section IV, Ghaffar *et al.* [50] formulated a link prediction problem to recommend possible weak ties to employees in an organization. While prior work had been done with link prediction in OSNs, they were the first to address link prediction in the enterprise social networks (ESNs), i.e., social networks inside enterprises, which are nowadays utilized by employees for various purposes such as sharing information and searching for experts. For optimization in the link prediction process, they introduced a social-organization overlap factor to favor candidates from different communities and discourage those in the same team with the ego. Their results also demonstrated high performance in terms of AUC and precision.

Another work in link prediction aimed to predict future co-investment behaviors. Venture capital (VC), i.e., financial capital provided for start-up companies, is of great importance in the high-tech industry and has benefitted many of the major companies in the field. Motivated by the fact that over 80% of the VC investments are related to at least two investors, Wang *et al.* [56] studied the VC co-investment behavior. They employed a total of 81 features to design a model for predicting future co-investment behaviors. In particular, they designed a group of 20 features related to the SH theory to indicate the centrality of the nodes, involving network constraint and BC. They then conducted feature selection and identified two of the betweenness features as the top ten prominent features. They found that larger values of BC of both sides were related to a higher possibility of future co-investment, which is consistent with the intuition suggested by the structure hole theory. They implemented the prediction task by proposing a structural balance-based factor graph model (SBFG), and the model was able to achieve a satisfactory prediction performance and out-performed all the compared methods. Furthermore, with only the top ten selected features, they achieved an ACC of around 90%, which dropped by only 0.18% compared with using all the features.

Similarity search is a common method used in link prediction tasks. Zhang *et al.* [57] proposed a sampling-based algorithm called Panther to retrieve the most similar top-$k$ nodes. When measuring the performance of the algorithm and other baselines, they adopted top-$k$ SH spanner detection as a task to evaluate the ACC, fueled by the intuition that SH spanners share the same structural patterns. They used network constraint as ground truth and then fed a few seed users into the model to find other SH spanners through the sampling method. In the experiment, Panther achieved consistently better performance in terms of accurately finding SH spanners than other compared methods. They also built a system recommending similar authors based on their algorithm. This work further validated the applicability of SH theory in link prediction tasks.

Aside from link prediction, the classification of links also has numerous applications in social network analysis. In a network, there exist different types of social relationships. A simple example is family, colleague, friend, and acquaintance relationships in a social network. Different types of social ties tend to influence people differently, so it is useful to identify the types of them. Tang *et al.* [58] incorporated the SH theory among several basic social science theories, including social balance theory, social status theory, and the theories of strong ties and weak ties, in their transfer-based factor graph model for classifying social ties. Their analysis of the data revealed that users are more likely (with a 20%–152% higher chance) to have the same type of social tie with an SH spanner, especially those unconnected users. Unlike domain-specific features, such observation based on social science theories holds true in different networks in diverse domains, making transfer learning more effective. Hence, they defined six features to indicate SH properties in the model. In the experiment, their TranFG model was able to achieve great performance and significant improvement over alternative methods, and their further examination validated that social science theory-based features made crucial contribution to the performance.

Similarly, Chen *et al.* [59] also worked on a transfer learning framework for social tie prediction and leveraged the SH theory among several social theories to make it possible to transfer the knowledge learned from a well-labeled graph to the target graph. In the model, they defined the existence of SHs as an attribute of each edge as part of the input. Their evaluation showed that their model outperformed traditional methods and was also less time-consuming than the TranFG model.

### *C. Discussion and Future Expectations*

SH theory offers quantitative measurement features serving as input features of ML models to help improve the performance. On the individual level, measures, such as effective size, network constraint, and BC, are effective in predicting user type and behaviors. On the connection level, SH theory has been widely adopted in combination with other social science theories. For example, social balance theory is the theoretical foundation of the model in [56] and is employed in parallel with SH theory in [58], [59]. A weak tie theory is also adopted synchronously in [50], [58], [59]. Functioning as a supporting social science theory, the SH theory can also cut down on the number of parameters and simplify the learning process, improving learning efficiency and effectiveness. We believe that this will continue to dominate the application of the theory in the ML field and expect to see it applied in more scenarios. To name a few possible directions, bridging roles that connect different regions of a city can help analyze mobility networks; the role SH spanners play in the information diffusion can help with information recommendation; the special social position of SH spanners can also contribute to recommender systems. There is much more to be explored.

Besides, we anticipate more works incorporating SH theory into graph-level tasks, such as graph classification, graph isomorphism, and graph partitioning, as SHs can imply a lot about the overall structure of the whole graph as well as the subgraphs.

Moreover, graph neural networks [63] have been developing rapidly in recent years. The intuition behind it is that nodes aggregate information from their neighbors. We believe that, owing to the special role SH spanners play in the information diffusion process, it will be valuable to consider and integrate SH theory in the propagating process of graph neural networks.

## VI. Conclusion

This article reviews the use of the SH theory in social networks from different perspectives: foundations of the theory, SH spanner detection, and applications of the theory. SHs refer to the critical bridges between communities or groups. Individuals occupying these positions are believed to possess network advantages. Hence, a number of metrics and algorithms have been carefully designed to identify these users, i.e., SH spanners. SH theory has been widely applied in social network analysis, resulting in applications in a wide range of practical scenarios as well as ML-based social prediction. We believe that a deeper exploration of SHs will further produce a number of interesting and valuable questions and findings in the field of social network analysis. In particular, we expect more work related to dynamic network evolvement in the near future.

## References

[1] B. Hendrickson, D. Rosen, and R. K. Aune, "An analysis of friendship networks, social connectedness, homesickness, and satisfaction levels of international students," *Int. J. Intercultural Relations*, vol. 35, no. 3, pp. 281–295, May 2011.

[2] V. Krebs, "Mapping networks of terrorist cells," *Connections*, vol. 24, no. 3, pp. 43–52, 2002.

[3] K. Basu and A. Sen, "Identifying individuals associated with organized criminal networks: A social network analysis," *Social Netw.*, vol. 64, pp. 42–54, Jan. 2021.

[4] A. Oksanen, J. Hawdon, and P. Räsänen, "Glamorizing rampage online: School shooting fan communities on YouTube," *Technol. Soc.*, vol. 39, pp. 55–67, Nov. 2014.

[5] P. A. Gloor, M. Paasivaara, D. Schoder, and P. Willems, "Finding collaborative innovation networks through correlating performance with social network structure," *Int. J. Prod. Res.*, vol. 46, no. 5, pp. 1357–1371, Mar. 2008, doi: 10.1080/00207540701224582.

[6] R. S. Burt, *Structural Holes: The Social Structure Competition*. Cambridge, MA, USA: Harvard University Press, 1992.

[7] T. Lou and J. Tang, "Mining structural hole spanners through information diffusion in social networks," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 825–836.

[8] M. Granovetter, "The strength of weak ties," *Amer. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, 1973.

[9] W. Xu, T. Li, W. Liang, J. X. Yu, N. Yang, and S. Gao, "Identifying structural hole spanners to maximally block information propagation," *Inf. Sci.*, vol. 505, pp. 100–126, Dec. 2019.

[10] L. He, C.-T. Lu, J. Ma, J. Cao, L. Shen, and P. S. Yu, "Joint community and structural hole spanner detection via harmonic modularity," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 875–884.

[11] C. Song, W. Hsu, and M. L. Lee, "Mining brokers in dynamic social networks," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2015, pp. 523–532.

[12] M. Rezvani, W. Liang, W. Xu, and C. Liu, "Identifying top-K structural hole spanners in large-scale social networks," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2015, pp. 263–272.

[13] W. Xu, M. Rezvani, W. Liang, J. X. Yu, and C. Liu, "Efficient algorithms for the identification of top-$K$ structural hole spanners in large social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1017–1030, Jan. 2017.

[14] H. Huang, J. Tang, L. Liu, J. Luo, and X. Fu, "Triadic closure pattern analysis and prediction in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3374–3389, Dec. 2015.

[15] R. S. Burt and J. Merluzzi, "Network oscillation," *Acad. Manage. Discoveries*, vol. 2, no. 4, pp. 368–391, Dec. 2016.

[16] Y. Ye, H. Zhu, T. Xu, F. Zhuang, R. Yu, and H. Xiong, "Identifying high potential talent: A neural network based dynamic social profiling approach," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 718–727.

[17] E. Choi and K. C. Lee, "Relationship between social network structure dynamics and innovation: Micro-level analyses of virtual cross-functional teams in a multinational B2B firm," *Comput. Hum. Behav.*, vol. 65, pp. 151–162, Dec. 2016.

[18] X.-H.-F. Wang, Y. Fang, I. Qureshi, and O. Janssen, "Understanding employee innovative behavior: Integrating the social network and leader-member exchange perspectives," *J. Organizational Behav.*, vol. 36, no. 3, pp. 403–420, Apr. 2015.

[19] H. Choi and H. Zo, "Network closure versus structural hole: The role of knowledge spillover networks in national innovation performance," *IEEE Trans. Eng. Manage.*, early access, Feb. 21, 2020, doi: 10.1109/TEM.2020.2972347.

[20] M. Fritsch and M. Kauffeld-Monz, "The impact of network structure on knowledge transfer: An application of social network analysis in the context of regional innovation networks," *Ann. Regional Sci.*, vol. 44, no. 1, pp. 21–38, Feb. 2010.

[21] J. Zhu, Y. Liu, and X. Yin, "A new structure-hole-based algorithm for influence maximization in large online social networks," *IEEE Access*, vol. 5, pp. 23405–23412, 2017.

[22] Y. Yang, J. Jia, B. Wu, and J. Tang, "Social role-aware emotion contagion in image social networks," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 65–71.

[23] Q. Zhang, D. D. Zeng, F.-Y. Wang, R. Breiger, and J. A. Hendler, "Brokers or bridges? Exploring structural holes in a crowdsourcing system," *Computer*, vol. 49, no. 6, pp. 56–64, Jun. 2016.

[24] T. Bhowmik, N. Niu, P. Singhania, and W. Wang, "On the role of structural holes in requirements identification: An exploratory study on open-source software development," *ACM Trans. Manage. Inf. Syst.*, vol. 6, no. 3, pp. 1–30, 2015.

[25] T. Zimmermann and N. Nagappan, "Predicting defects using network analysis on dependency graphs," in *Proc. 13th Int. Conf. Softw. Eng. (ICSE)*, 2008, pp. 531–540.

[26] J. Kim, Y. Park, C. Kim, and H. Lee, "Mobile application service networks: Apple's app store," *Service Bus.*, vol. 8, no. 1, pp. 1–27, Mar. 2014.

[27] R. S. Burt, M. Kilduff, and S. Tasselli, "Social network analysis: Foundations and frontiers on advantage," *Annu. Rev. Psychol.*, vol. 64, no. 1, pp. 527–547, Jan. 2013.

[28] R. S. Burt, "Structural holes and good ideas," *Amer. J. Sociol.*, vol. 110, no. 2, pp. 349–399, Sep. 2004.

[29] K. Stovel and L. Shaw, "Brokerage," *Annu. Rev. Sociol.*, vol. 38, pp. 139–158, 2012.

[30] E. Katz and P. F. Lazarsfeld, *Personal Influence, The Part Played by People in the Flow of Mass Communications*. Mumbai, India: Free Press, 1955.

[31] R. S. Burt, "Reinforced structural holes," *Social Netw.*, vol. 43, pp. 149–161, Oct. 2015.

[32] S. P. Borgatti, C. Jones, and M. G. Everett, "Network measures of social capital," *Connections*, vol. 21, no. 2, pp. 27–36, 1998.

[33] F. Li, Z. Zou, J. Li, Y. Li, and Y. Chen, "Distributed parallel structural hole detection on big graphs," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2019, pp. 519–535.

[34] F. Jiang, L. He, Y. Zheng, E. Zhu, J. Xu, and P. S. Yu, "On spectral graph embedding: A non-backtracking perspective and graph approximation," in *Proc. SIAM Int. Conf. Data Mining*, 2018, pp. 324–332.

[35] Y. Zhang *et al.*, "Finding structural hole spanners based on community forest model and diminishing marginal utility in large scale social networks," *Knowl.-Based Syst.*, vol. 199, Jul. 2020, Art. no. 105916.

[36] E. Bakshy, I. Rosenn, C. Marlow, and L. A. Adamic, "The role of social networks in information diffusion," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 519–528.

[37] J. Kleinberg, S. Suri, É. Tardos, and T. Wexler, "Strategic network formation with structural holes," in *Proc. 9th ACM Conf. Electron. Commerce*, 2008, pp. 284–293.

[38] L. Wang, T. Lou, J. Tang, and J. E. Hopcroft, "Detecting community kernels in large social networks," in *Proc. IEEE 11th Int. Conf. Data Mining*, Dec. 2011, pp. 784–793.

[39] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogenous networks," in *Proc. 5th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2012, pp. 743–752.

[40] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 137–146.

[41] F. Chung, "Laplacians and the cheeger inequality for directed graphs," *Ann. Combinatorics*, vol. 9, no. 1, pp. 1–19, Apr. 2005.

[42] R. S. Burt, "Network-related personality and the agency question: Multirole evidence from a virtual world," *Amer. J. Sociol.*, vol. 118, no. 3, pp. 543–591, Nov. 2012.

[43] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM J. Comput.*, vol. 1, no. 2, pp. 146–160, Jun. 1972.

[44] Y. Xu, H. Xu, D. Zhang, and Y. Zhang, "Finding overlapping community from social networks based on community forest model," *Knowl.-Based Syst.*, vol. 109, pp. 238–255, Oct. 2016.

[45] F. Ghaffar and N. Hurley, "Structural hole centrality: Evaluating social capital through strategic network formation," *Comput. Social Netw.*, vol. 7, no. 1, pp. 1–27, Dec. 2020.

[46] L. Du, Y. Wang, G. Song, Z. Lu, and J. Wang, "Dynamic network embedding: An extended approach for skip-gram based network embedding," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 2086–2092.

[47] L.-K. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, "Dynamic network embedding by modeling triadic closure process," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 571–578.

[48] S. Rodan, "Structural holes and managerial performance: Identifying the underlying mechanisms," *Social Netw.*, vol. 32, no. 3, pp. 168–179, Jul. 2010.

[49] G. Gao, P. Hinds, and C. Zhao, "Closure vs. structural holes: How social network information and culture affect choice of collaborators," in *Proc. Conf. Comput. Supported Cooperat. Work*, 2013, pp. 5–18.

[50] F. Ghaffar, T. S. Buda, H. Assem, A. Afsharinejad, and N. Hurley, "A framework for enterprise social network assessment and weak ties recommendation," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 678–685.

[51] Y. D. Kwon, R. H. Mogavi, E. U. Haq, Y. Kwon, X. Ma, and P. Hui, "Effects of ego networks and communities on self-disclosure in an online social network," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2019, pp. 17–24.

[52] Q. Ying, D. M. Chiu, and X. Zhang, "Diversity of a user's friend circle in OSNs and its use for profiling," in *Proc. Int. Conf. Social Informat.*, 2018, pp. 471–486.

[53] W. Cai, J. Jia, and W. Han, "Inferring emotions from image social networks using group-based factor graph model," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[54] F.-C. Yuan, "Intelligent forecasting of inbound tourist arrivals by social networking analysis," *Phys. A, Stat. Mech. Appl.*, vol. 558, Nov. 2020, Art. no. 124944.

[55] M. Lu, "Node importance evaluation based on neighborhood structure hole and improved TOPSIS," *Comput. Netw.*, vol. 178, Sep. 2020, Art. no. 107336.

[56] Z. Wang, Y. Zhou, J. Tang, and J.-D. Luo, "The prediction of venture capital co-investment based on structural balance theory," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 537–550, Feb. 2016.

[57] J. Zhang *et al.*, "Fast and flexible top-$K$ similarity search on large networks," *ACM Trans. Inf. Syst.*, vol. 36, no. 2, pp. 1–30, 2017.

[58] J. Tang, T. Lou, J. Kleinberg, and S. Wu, "Transfer learning to infer social ties across heterogeneous networks," *ACM Trans. Inf. Syst.*, vol. 34, no. 2, pp. 1–43, Apr. 2016.

[59] Y. Chen, Y. Xiong, B. Liu, and X. Yin, "TranGAN: Generative adversarial network based transfer learning for social tie prediction," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[60] R. S. Burt, J. E. Jannotta, and J. T. Mahoney, "Personality correlates of structural holes," *Social Netw.*, vol. 20, no. 1, pp. 63–87, Jan. 1998.

[61] Q. Gong, J. Zhang, X. Wang, and Y. Chen, "Identifying structural hole spanners in online social networks using machine learning," in *Proc. ACM SIGCOMM Conf. Posters Demos*, 2019, pp. 93–95.

[62] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econ., J. Econ. Soc.*, vol. 47, no. 2, pp. 263–291, 1979.

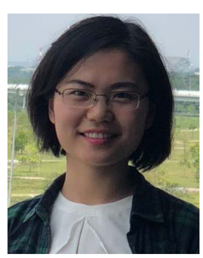
[63] F. Scarselli, M. Gori, A. Chung Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

**Zihang Lin** is currently pursuing the bachelor's degree with the School of Computer Science, Fudan University, Shanghai, China.

He has been a Research Assistant with the Mobile Systems and Networking (MSN) Group, Shanghai, since 2019. His research interests include social computing and machine learning.

**Yuwei Zhang** is currently pursuing the bachelor's degree with the School of Computer Science, Fudan University, Shanghai, China.

She has been a Research Assistant with the Mobile Systems and Networking (MSN) Group, Shanghai, since 2019. Her research interests include social network analytics and machine learning.



**Qingyuan Gong** received the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2020.

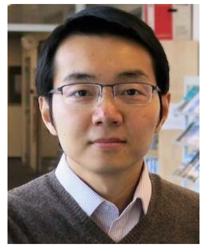
She had been a Visiting Student with the University of Göttingen, Göttingen, Germany, in 2015 and 2019, respectively, and the University of Chicago, Chicago, IL, USA, in 2018. She is currently working as a Post-Doctoral Researcher with Fudan University. She has published referred articles in *IEEE Communications Magazine*, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING (TDSC), *ACM Transactions on the Web (TWEB)*, Springer *World Wide Web: Internet and Web Information Systems (WWW)* journal, the Conference on Information and Knowledge Management (ACM CIKM), and International Conference on Parallel Processing (ICPP). Her research interests include network security, user behavior analysis, and computational social systems.

**Yang Chen** (Senior Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2004 and 2009, respectively.

He visited Stanford University, Stanford, CA, USA, in 2007 and Microsoft Research Asia, Beijing, as a Visiting Student, from 2006 to 2008. From September 2009 to April 2011, he had been a Research Associate and the Deputy Head of the Computer Networks Group, Institute of Computer Science, University of Göttingen, Göttingen, Germany. From April 2011 to September 2014, he was a Post-Doctoral Associate with the Department of Computer Science, Duke University, Durham, NC, USA, where he has served as a Senior Personnel for the NSF MobilityFirst project. He was a Nokia Visiting Professor with Aalto University, Espoo, Finland, in 2019. He is currently an Associate Professor with the School of Computer Science, Fudan University, Shanghai, China, and leads the Mobile Systems and Networking (MSN) Group, Fudan University. His research interests include online social networks, Internet architecture, and mobile computing.

Dr. Chen has served as an OC/TPC member for many international conferences, including the Symposium on Operating Systems Principles (SOSP), SIGCOMM, International World Wide Web Conference (WWW), International Joint Conference on Artificial Intelligence (IJCAI), AAAI Conference on Artificial Intelligence (AAAI), European Conference on Artificial Intelligence (ECAI), the International Conference on Database Systems for Advanced Applications (DASFAA), International Symposium on Quality of Service (IWQoS), International Conference on Computer Communications and Networks (ICCCN), Global Communications Conference (GLOBECOM), and International Conference on Communications (ICC). He also serves as an Associate Editor-in-Chief for the *Journal of Social Computing*, an Associate Editor for IEEE ACCESS, and an Editorial Board Member for the *Transactions on Emerging Telecommunications Technologies (ETT)*.

**Atte Oksanen** is currently a Professor of Social Psychology with Tampere University, Tampere, Finland. He has published 100 peer-reviewed journal articles and he has over 270 publications to his name. Publications include articles in flagship journals, such as *Criminology*, *Addiction*, and *Pediatrics* and monographs published by Routledge and Springer. His research interests include emerging technologies and social interaction online. He has led major research projects funded by different foundations, including the Academy of Finland, the Kone Foundation, the Finnish Cultural Foundation, the Finnish Foundation for Alcohol Studies, the Finnish Work Environment Fund, the Helsingin Sanomat Foundation, and the Aaltonen Foundation.



**Aaron Yi Ding** (Member, IEEE) received the M.Sc. and Ph.D. degrees (Hons.) from the Department of Computer Science (Birthplace of Linux), University of Helsinki, Helsinki, Finland, in 2009 and 2015, respectively.

He is currently a Tenured Assistant Professor with the Department of Engineering Systems and Services, TU Delft, Delft, The Netherlands, and an Adjunct Professor (Dosentti) of Computer Science with the University of Helsinki. Prior to TU Delft, he worked with TU Munich, Munich, Germany, Columbia University, New York City, NY, USA, the University of Cambridge, Cambridge, U.K., and the University of Helsinki. His research interests include edge computing, the Internet of Things, and mobile networking services.

Dr. Ding was a two-time recipient of the Nokia Foundation Scholarships and awarded the Best Paper of ACM EdgeSys 2019 and the ACM SIGCOMM Best of CCR.