Location-Based Social Network Data Generation Based on Patterns of Life

Joon-Seok Kim[†], Hyunjee Jin[†], Hamdi Kavak[‡], Ovi Chris Rouly[#],
Andrew Crooks[‡], Dieter Pfoser[†], Carola Wenk[#], Andreas Züfle[†]
[†]*Geography and Geoinformation Science, George Mason University*
jkim258,hjin6,dpfoser,azufle@gmu.edu
[‡]*Computational and Data Science, George Mason University*
hkavak,acrooks2@gmu.edu
[#]*Computer Science, Tulane University*
orouly,cwenk@tulane.edu

***Abstract*—Location-based social networks (LBSNs) have been studied extensively in recent years. However, utilizing real-world LBSN data sets yields several weaknesses: sparse and small data sets, privacy concerns, and a lack of authoritative ground-truth. To overcome these weaknesses, we leverage a large-scale LBSN simulation to create a framework to simulate human behavior and to create synthetic but realistic LBSN data based on human patterns of life. Such data not only captures the location of users over time but also their interactions via social networks. Patterns of life are simulated by giving agents (i.e., people) an array of "needs" that they aim to satisfy, e.g., agents go home when they are tired, to restaurants when they are hungry, to work to cover their financial needs, and to recreational sites to meet friends and satisfy their social needs. While existing real-world LBSN data sets are trivially small, the proposed framework provides a source for massive LBSN benchmark data that closely mimics the real-world. As such, it allows us to capture 100% of the (simulated) population without any data uncertainty, privacy-related concerns, or incompleteness. It allows researchers to see the (simulated) world through the lens of an omniscient entity having perfect data. Our framework is made available to the community. In addition, we provide a series of simulated benchmark LBSN data sets using different synthetic towns and real-world urban environments obtained from OpenStreetMap. The simulation software and data sets, which comprise gigabytes of spatio-temporal and temporal social network data, are made available to the research community.**

***Index Terms*—Data Generation, Location-Based Social Networks, Temporal Social Network Data, Social Simulation, Patterns of Life, Trajectory Data Generation, Social Network Data Generation**

I. Introduction

A social network is a structure consisting of individual users connected by a social relationship such as friendship. Social networking services build on real-world social networks through online platforms, providing ways for users to share ideas, activities, events, and interests. For example, users can share location-tagged images with their friends (e.g., Instagram), rate restaurants/bars, recommend them to their friends (e.g., Foursquare), or log jogging and bicycle trails for sports analysis and experience sharing (e.g., Bikely). This dimension of location bridges the gap between the physical world and online social networking services. As location is one of the most important components of user context, extensive knowledge about an individual's interests, behaviors, and relationships with others can be learned from their locations. These kinds of location-tagged and location-driven social structures are known as location-based social networks (LBSNs). Research on LBSNs has become a vivid topic in the last decade, enabled by many practical applications (surveyed in Section II) and rooted in the mobile data management community (e.g., [37], [49], [59]–[61], [67], [73], [73], [91], [103], [107]). Publicly available real-world data sets have been the driving force for LBSN research in recent years, but such data sets exhibit certain weaknesses:

- **Data sparsity:** LBSN data exhibits an extreme long-tail distribution of user behavior. In all existing available data sets, the vast majority of users has less than ten check-ins [43]. Besides, the number of locations visited by a user is usually only a small portion of all locations that user has visited. This results in the density of the data used in experimental studies on LBSNs to be only usually around 0.1% [43].
- **Small data sets:** Existing data sets used to train models are small, as detailed in Section II-B. They tend to only cover a short period of time, a small number of users, or a small number of check-ins.
- **Privacy Concerns:** Most LBSN data was published by users and consented for public use. However, some users may revoke this consent, for instance, by deleting their LBSN account. Such changes will not be reflected in existing LBSN data sets and thus creating severe privacy concerns.
- **No ground-truth:** There is no way to assess, in existing LBSN data, whether check-ins are missing or if the social network is correct and complete. Without knowing the ground truth, it is difficult to assess the accuracy and robustness of existing experimental results using LBSN data.

Furthermore, a recent study [33] has shown that the lower bound of predictability of the human spatio-temporal behavior (defined in [33]) is as low as 27%. They conclude that "Researchers working with LBSN data sets are often confronted by themselves or others with doubts regarding the quality or the potential of their data sets." and that "it is reasonable to be skeptical" [33].

To overcome these weaknesses, we developed a location-based social network simulation capable of creating multiple artificial but socially plausible, large-scale LBSN data sets as envisioned in [27]. These large and dense data sets will allow the broader social and data science research communities to test LBSN-related hypotheses without encountering issues pertaining to data sparsity, privacy concerns, and lack of ground-truth. Our simulation and the generated synthetic data sets enable investigation of what is now not currently possible even if there existed provenance of complete, high-fidelity, real-world LBSN data (i.e., 'perfect' Foursquare data). In the remainder of this paper and beginning in Section II, we provide an overview of LBSNs along with application areas before turning our attention to existing LBSN data sets. In Section III, we enumerate what we see to be the shortcomings of current data sets and attempt to explain our rationale for building an agent-based location-based social simulation to address those shortcomings. This is followed by our simulation results (Section IV) which demonstrates its utility for LBSN research. Section V outlines some research applications that can benefit from our simulated data, and Section VI provides a brief conclusion to the paper.

II. Related Work

Users [102] and locations [104] are the two major subjects that interact with each other in LBSNs. As such, we can observe three types of networks that constitute an LBSN, (i) a user-user social network, (ii) a location-location spatial network, and (iii) a user-location bipartite network. Figure 1 gives a schematic overview of these networks and their interaction. Like in a traditional social network, users are connected via relationships such as friend, family, or co-worker. However, when locations are added, spatial network connections can also be defined by proximity in terms of path distance, they may also introduce connections between locations that have similar semantic properties (e.g., of the same location type). Finally, the core feature of a LBSN is the user-location network, which bridges users and locations (Figure 1). This bipartite network between users and locations captures events of users visiting a physical location. Such so-called "check-ins" may be enriched with qualitative information, such as user recommendations, which may be explicit (e.g., on a scale from one to five stars), or implicit, (e.g., by observing that a user frequently checks-in at the same location).

A. Applications of LBSNs

LBSN data has been leveraged for a plethora of applications, an overview of which is given in Table I. The goal of this section is not only to show the wide variety of applications that are using LBSN data, but also to show that applications are very vivid research topics. The data generated by our proposed simulation framework will directly benefit all these applications and all researchers by providing large and ground-truth enriched data sets for a more thorough experimental analysis and thus foster more informed decision making.

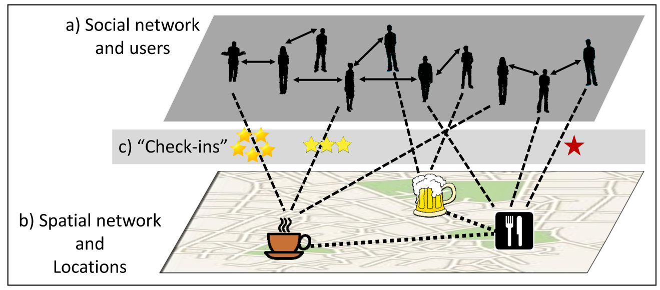


Fig. 1: LBSN Overview.

The initial work with respect to LBSN's focused on modeling and describing **human mobility patterns** (e.g., [13], [15], [52], [63], [64], [79], [80]), analyzing these patterns (e.g., [12], [53], [73]), and explaining why users choose locations and how social ties affect this choice (e.g., [69], [107]). Another application is that of **location recommendation**, which leverages check-ins of users and their ratings in the user-location network to recommend new locations to users. This research area not only has applications in helping a specific user to explore new places within a city, but can also help third-parties such as advertisers to provide specific advertisements. Therefore, a plethora of recent research, surveyed in [6], has studied this application (e.g., [8], [9], [17]–[20], [24], [30]–[32], [34], [37]–[39], [41]–[44], [51], [62], [71], [72], [81], [85], [86], [89]–[100]). A closely related application area is that of **Location Prediction** (e.g., [5], [10], [22], [36], [40], [59], [67], [84]), which attempts to predict the future check-ins of users. Another active research field is LBSNs is **Friend Recommendation** or **Social Link Prediction** (e.g., [11], [23], [48], [56], [65], [70], [75], [77], [82]), which suggests new friends to users based on similar interests at similar locations, while also having similar social connections. Other research topics with respect to LBSNs include efficient query processing (e.g., [1], [3], [4], [35], [60], [106],) finding user communities (e.g., [74], [88], [101]), and estimating the social influence of users (e.g., [76]).

While there has been a plethora of recent and interesting research with respect to the LBSN applications as noted above, the impact of LBSN research relies heavily on the quality of data, which we turn to next (Section II-B) in order to show that there is a considerable shortage of rich data sets. We will show that the data sets most commonly used have in experiments are relatively small, are limited in sample size for individual users, and cannot provide authoritative ground-truth knowledge for a meaningful evaluation of all these methods.

B. Existing LBSN data sets

Real-world LBSN data sets are a scarce resource due to the privacy implications of making such data public. Also, service providers consider such data sets invaluable when it comes to providing a competitive product and are thus somewhat unwilling to give researchers even sizable data sets. Table II summarizes the main characteristics of publicly available data sets that are intensively used by the LBSN research community (see Table I for topics and references).

TABLE I: State-of-the-Art Related Work in LBSNs and Utilized Data Sets.

References	Main topic	Gowalla	Foursquare	Yelp	Synthetic
[69], [107]	LBSN Analysis	✓			
[52], [53], [63], [79], [80]	LBSN Analysis		✓		
[12], [13], [15], [64], [107]	LBSN Analysis	✓	✓		
[73]	LBSN Analysis			✓	
[9], [32], [37], [41], [44], [71]	Loc. Recommendation	✓			
[19], [20], [38], [61], [62], [72], [81], [85], [86], [89], [90]	Loc. Recommendation		✓		
[18], [34], [51], [91]–[94], [97] [8], [17], [31], [39], [42], [96], [100]	Loc. Recommendation	✓	✓		
[24], [99]	Loc. Recommendation			✓	
[95], [98]	Loc. Recommendation		✓	✓	
[43]	Loc. Recommendation	✓	✓	✓	
[30]	Loc. Recommendation		✓		✓
[5], [36], [40], [84]	Loc. Prediction	✓			
[50], [59]	Loc. Prediction		✓		
[10], [22], [67]	Loc. Prediction	✓	✓		
[11], [23], [48], [56], [65], [70]	Social Link Prediction	✓			
[77], [82]	Social Link Prediction		✓		
[75]	Social Link Prediction	✓	✓		
[35], [106]	Query Processing	✓			
[1]	Query Processing			✓	
[3], [4], [60]	Query Processing	✓	✓		✓
[83]	Community Detection	✓			
[74], [88], [101]	Community Detection		✓		
[76]	Social Influence	✓			

TABLE II: Publicly Available Real-World LBSN Data Sets.

Data set	#Users	#Locations	#CheckIns	#Links	Period
Gowalla	319K	2.8M	36M	4.4M	20mo
BrightKite	58K	971K	4.49M	214k	30mo
Foursquare	2.7M	11.1M	90M	0	5mo
Yelp	1.00M	144K	4.10M	0	36mo

Gowalla: Collected and retrieved from the LBSN Gowalla [41], which was launched in 2007 and closed in 2012. This data set has the largest social network of any public LBSN data set while the majority of users are inactive. After removing users with less than 15 check-ins and removing locations with less than ten visitors, more than half of the visitors are eliminated [41]. A similar data set is that of *Brightkite*, which is available at SNAP [66]. As can been seen in Table II, Brightkite is smaller than the Gowalla data in every aspect.

Foursquare: In terms of the number of users and check-ins, the largest publicly available LBSN data set was collected from Foursquare [78]. However, this data set provides no social network information.

Yelp: A large data set is published by Yelp as part of the Yelp data set Challenge [87]. This data set provides additional information, such as user-location ratings, user comments, user information, and location information. Again, this data set provides no social network information.

Synthetic Data: The problem of using sparse and noisy real-world LBSN data has already been identified in previous work (e.g., [3], [4], [30]). However, none of these works have proposed a way to obtain plausible check-in data. For example, [3], [4], [30] generated user-location check-ins at random using parametric distributions without considering the semantics of the movement. While [60] created additional check-ins by replication of Gowalla and Brightkite data, thus creating more data for run-time evaluation purposes but without creating more information.

One could, therefore, question whether the experimental results of existing work on LBSNs may be considered conclusive, both in terms of scalability and effectiveness due to a lack of large scale available data sets [33]. This paper aims at closing this gap by proposing the means to generate large scale and ground-truth based synthetic data sets through simulation, which we turn to next. Synthetic data would allow insights into what is possible concerning new and improved geoinformation systems, but also in terms of privacy and anonymization research without raising any privacy concerns.

III. Location-Based Social Simulation

In order to address the limitations of LBSN data sets as discussed in Section II-B, we have created a LBSN simulation framework for generating very large, high fidelity, and socially plausible (yet not real) LBSN data. The framework simulates individuals (i.e., agents) who live and travel in an urban environment. Agents exhibit socially plausible behavior which is based on well-respected psychology and social science theories (e.g., [2], [47]). The agents' needs and preferences guide the choice of locations they visit, which may yield to new friendships, depending on the type of locations where they meet. Following the underlying spatial network, agents visit specific locations which also results in specific travel patterns emerging. Simulation parameters can be adjusted to create social settings similar to the real-world allowing us to create massive sets of simulated LBSN data. Such high fidelity data sets contain all individuals of our simulated world with certainty while not impacting the privacy of any human subject in the real-world. As a deliverable, this research yields synthetic LBSN data sets of hundreds of thousands of users, scaling to years of observed user data, and thus creating gigabytes of meaningful check-in and social interaction data (as will be shown in Section IV).

A. Model Logic: Patterns of Life

The computational framework follows up our preliminary work demonstrated in [29], which creates simulated worlds in which agents move and interact with the environment and with each other. In the work presented in this paper, we provide a more comprehensive needs and behaviors which result in more complex patterns of life and introduce social networks and there formation. In addition, each simulation instance of our framework is based on (real or synthetic) spatial networks with locations and social networks of users. The agent model logic used to generate the data is constructed based on people's daily patterns of life (PoL) supported by well-respected psychology and social science theories (e.g., [2], [47]). Each individual is equipped with the first three levels of the Maslow's [47] Hierarchy of Needs: (1) physiological, (2) safety, and (3) belongingness and love needs. To satisfy their needs, agents travel to sites on the underlying road network. Agents travel on shortest paths at a constant walking speed of $1.4m/s$, regardless of traffic.

The most basic and highest prioritized needs are physiological needs, which make agents rent an apartment, eat food at home or at a restaurant when they are hungry, and sleep when their internal circadian rhythm kicks-in according to their individually computed wake-up times. The agent's are driven to reduce their physiological needs first. The second-level needs begins with the agent engaging in a process instantiated by the agents as they try to attain financial stability. This second-level drive leads them into a process that begins by them trying to find and keep a job. From there, they seek to have enough income to pay for their own individually desired, temporally projected, and fiscally anticipated (i.e., budgeted) monthly costs in the world. Among the simulated costs we require the agents to pay for things like rent, food at home or in restaurants, education expenses for offspring, and voluntary personal expenses associated with recreation. All costs must be paid for by the individual agent from its own 'earned' income. An agent's job and pay scale is dependent on its education level and momentary job market availability. Each agent has a home, can move to a new home, has a job/work location, and may change job/work over time. The underlying logic is that an agent will usually stay in their job and home unless their financial stability is broken due to an unexpected event such as a roommate moving out. In such cases, they either move into a different apartment with a lower cost or switch to a better paying job.

Again, following Maslow's [47] Hierarchy, agents must satisfy physiological and safety needs first. Only after these needs are met can higher levels of self-actualization corresponding to belongingness and love (i.e., relationships and friendships) be computed and executed. When the more basic needs from the hierarchy are met, an agent may then choose to visit a recreational site for the purpose of socialization. Recreation sites are figurative 'hubs' for socialization. At these places agents may meet new peoples, create budding friendships, and or improve their existing friendship bonds with others. In our simulation, social relationships are simulated using a weighted and directed *social network*. Agents are attracted to recreation sites because of the agent's individual age, their individual income level, their own interests and the interests of others who visited the site in recent times, and their proximity to the site when the decision to visit the site is taken. When an agent visits a recreational site and has no friends there, there is a slight chance that the agent will establish a friendship with a stranger (i.e., focal closure). This chance slightly increases if the stranger is actually a friend's friend (i.e., cyclic closure). Thus, co-location (i) increases one's chances of becoming friends with other people and (ii) to maintain existing friendships. However, the lack of co-location, thus the lack of social interaction, decreases one's friendship strength, which may eventually lead to a disappearing friendship. We note that over years of simulation time, the agent's aim is to maximize attributes such as *happiness*, which is related to making friendships and *money balance*, which is related to job choice. With different agents having different goals, they maximize different attributes. Restaurants are another type of site used by agents to satisfy their physiological needs caused by hunger. Agents who visit restaurants have the chance to meet with their friends without coordination. Such meetings increase the strength of an existing friendship. Each agent has choices, such as preferring a certain type of restaurant, cafe, or bar. The simulation stores and writes spatial and social information for each agents into large, shared log-files which can be processed and analyzed offline.

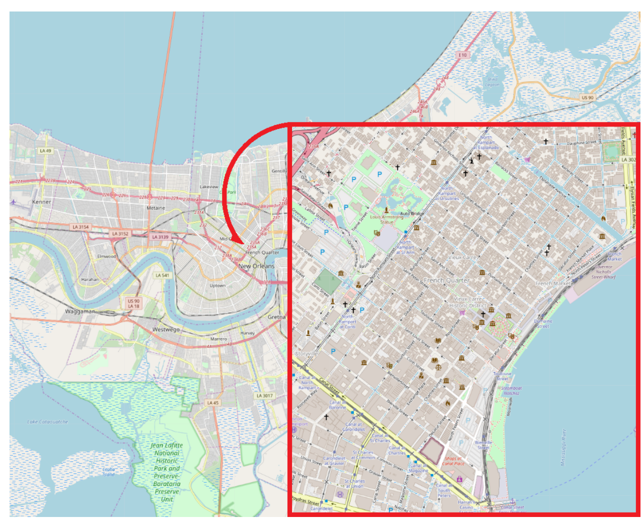


Fig. 2: New Orleans, Louisiana (NOLA), Mississippi River, Lake Pontchartrain, and the 'French Quarter'.

B. Framework Implementation

Our framework utilizes and extends the MASON (Multi-Agent Simulation of Neighborhoods) open-source simulation toolkit [45] and its GIS extension, GeoMASON [68]. MASON is a fast discrete-event multi-agent simulation library developed in Java. It was designed to be the foundation for sizeable custom-purpose Java simulations by providing the basic run-time infrastructure for simulation development.

Fig. 3: George Mason University (GMU), Fairfax, VA.

MASON has been used in the past to develop agent-based models to describe complex social interaction that is based on the agent's location in space and time, including models for riot prediction [57], simulating the spread of disease [14], and the emergence of slums in urban environments [55]. Such simulations consider only a limited number of agents, over only a few days of simulation time. One of the main challenges we addressed is scaling the system, by memorizing all previously computed shortest paths to avoid expensive recomputation. While this version of our framework supports only a single thread application, we plan to create a massively scalable implementation building upon MASON's upcoming distributed option [46]. Our Java implementation can be found at: https://github.com/gmuggs/pol.

IV. Simulation Results

Having defined our overall approach to instantiating an agent-based model that generates plausible location-based social network data (Section III), we now provide details to specific instantiations of our model, as well a discussion of the respective quantitative data generated. All of the generated data sets can be found at OSF (https://osf.io/e24th/). Due to the excessive size of some of these LBSN data sets we are about to introduce, we recommend that researchers interested in using the data re-run the simulation locally instead of downloading the data directly. Correctly parameterized executables are available for download, and our simulation is fully serialized and deterministic, such that the data generated locally is guaranteed to be identical to the downloadable data.

A. Initialization

The goal of the generated data sets discussed in the following is to act as benchmark data sets for the LBSN community. Therefore, we generated a mix of real and synthetic urban settings. Real road network and point-of-interest data was obtained from OpenStreetMap (OSM) [54], where we downloaded data for the greater New Orleans, LA (NOLA) metropolitan area (cf. Figure 2) and the George Mason University (GMU) office for Geo-Information Systems (GIS) Facilities Archives [21] provided us with data for the Fairfax, VA campus of GMU (cf. Figure 3). We also generated two synthetic urban data sets differing in size and layout denoted as *Small* (TownS) and *Large* (TownL) as shown in Figure 4. These synthetic urban components were created using a spatial network and place generator based in a generative grammar similar to L-systems described in [28].

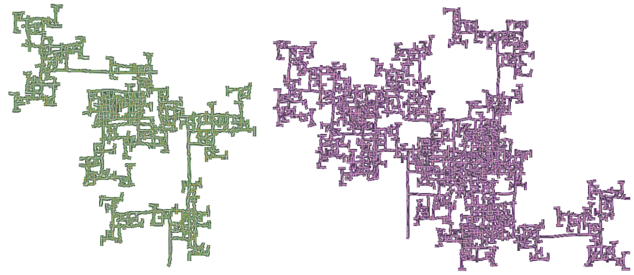


Fig. 4: Synthetic Villages - Small (Left) and Large (Right).

Within NOLA, the area we concentrated on was the historic French Quarter (FQ). The GMU area captures the main campus in Fairfax, VA. Both areas were prepared using QGIS desktop GIS software [58] and JOSM [26]. Data preparation involved editing the data sets to produce three separate shapefiles [16]: (i) building footprints, (ii) transportation networks (road and sidewalk layers), and (iii) building purpose (i.e., residential, commercial, etc.).

For reproduction of the following data generation and evaluation, we have compiled our Java source code into an executable Jar (https://github.com/gmuggs/pol/releases/tag/0.1) including dependencies (libraries and maps). Our Java code can be used to create a new simulation experiment that will consider any village, town, city, or other region of interest by loading an appropriate shapefile [16]. The computing hardware used for this work was a dual 10-core 2.8 GHz Intel Xeon E5-2680 v2 CPUs, 64 GB of RAM, and dual Xeon Phi 7120P co-processors. System utilization rarely exceeded 18 GB even when running simultaneous scenarios. The Java version used was 1.8.0_212. Finally, MASON and GeoMason versions 19 and 1.5.2 were utilized respectively.

The simulations described here were initialized by individual ASCII-text parameter files at program start. For consistency, we decided to populate our simulations of the GMU, NOLA, and the smaller and larger synthetic urban components with 1,000 agents each (cf. Fig. 5). For comparison, an additional simulation run had the larger synthetic area populated with 3,000 agents and NOLA with 5,000 agents. A note here about our experiment's data-grounding, empirical validity, and simulation plausibility: The Nonprofit Knowledge Works Data Center of New Orleans reported a fluctuating population in NOLA of between 4,176 people (estimated 2,908 households) in the year 2000 down to 3,813 people (estimated 2,635 households) in 2010 [7]. Their empirical census counts are comfortably bracketed by the population numbers we simulated.

In all simulations, the agents were initialized at random locations within their respective environments. In general, our experimental protocol involved creating simulations that ran for 15 simulated months. These were preset to halt and log

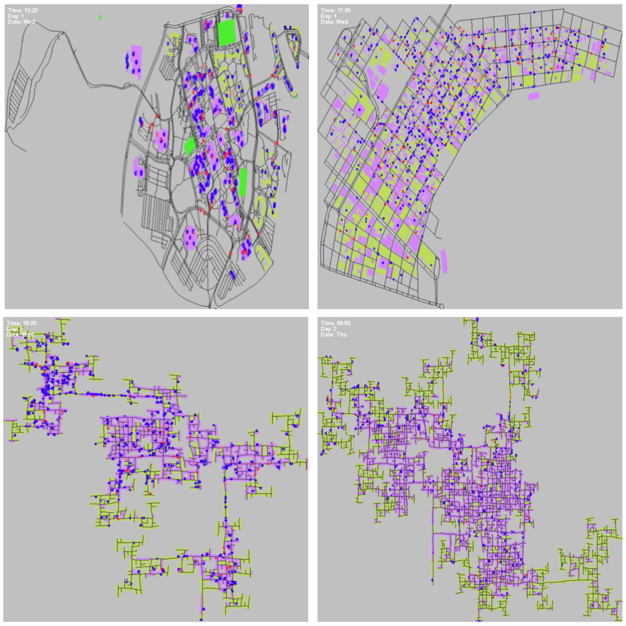


Fig. 5: Environments Populated with Agents. Clockwise from Top Left: GMU, NOLA, Large and Small Synthetic Villages.

output data. As shown in Figure 6 that there is typically a 1 to 3 month ‘settling’ time when the simulation starts. This is because the agent’s need to find their homes, work places, and then agents start to develop dynamic social networks based on temporal, spatial, and social needs. Thus, for comparative purposes, it may be useful to think of this 15-month interval as a 3-months ‘settling’ then 12-months experimental treatment. However, for the purpose of abstract comparison, we also chose to run GMU and NOLA with 1K-agents for a little over 10-years and 18-year, each, respectively. Table IV offers some particulars regarding the run-time output from our simulations.

At our download site: https://github.com/gmuggs/pol, we have provided run scripts, configuration files, fully compiled Jar binaries, and four urban environments. The Java source for the PoL simulation is located there as well. These files allow one to re-run our simulation to reproduce the data.

There were multiple settings that were used with each of the respective study areas. Table III specifies the run-time settings in detail: the area simulated, the type and number of sites simulated, the number of neighborhoods, and the count of agents. In terms of the number of sites, we simulated five types of sites: Schools, Pubs, Workplaces, Restaurants, and Apartments. The actual number of each respective site type and the number of neighborhoods simulated is the result of an internal computational process indirectly derived from the user choice of parameters but not directly accessible to the user at setup time.

B. *Results and Discussion*

Table IV gives an overview of the generated output data from the location-based social network simulation. It shows the number of agent check-ins and the number of social links attributed to each of the scenarios. We observe that the number of check-ins increases, for all study areas, linear with the number of agents. This is plausible, as the number of hours per day that agents can spend to satisfy their needs and visit sites is independent of other agents. However, we do see that the number of social links increase super-linear in the number of agents. This can be explained by more agents leading to larger co-locations of agents, creating chances for each *pair* of agents in the same co-location to become friends. We note that the generated temporal social network may have more edges than we have agent pairs. This is due to the temporal nature of the network. It reports changes over time and as such a single pair of agents can have multiple friends and unfriend events. The number reported corresponds to the number of new edges added to the temporal social network, regardless of the duration of these events. The super-linear growth of the social network also explains the super-linear run-time to create each data set, ranging from less to one hour for the 1000 agent instances to 10.5 hours for 5000 agents. Besides (i) number of check-ins and (ii) social links, we also report (iii) the run-time of each simulation and (iv) the resulting data size in Table IV. It is interesting to see that even small simulations can create sizable results given a longer duration. For example, the smaller synthetic urban component that had the longest (221 month) simulation period produced a 5.5GB result data set. However, the actual run-time of this simulation is shorter than a simulation with more agents that had a shorter simulation period, e.g., NOLA-5K - 15mo produced only a 2.3GB data set and took 1233min to run vs. NOLA-1k - 221mo produced a 5.5GB data set with a run-time of 774min. Increasing the number of agents results in more complex data structures, e.g., social networks, which in turn increases the run-time of the algorithms to process them at each step of the simulation.

Figure 6 shows the average number of friendships, over simulation time measured in 5-minute steps, for all the 1K networks. In all cases there is a three months (one month is equal to $12 * 24 * 30 = 8640$ steps) settling time during which agents establish friendships (starting with an empty social network). After this phase, friendship degrees fluctuate around mean values for each simulation. We observe that the two real networks exhibit a denser social network, due to a more uneven distribution of agents, leading to large groups of agents to co-locate at sites to become friends.

For a more detailed view of the resulting social network, Figure 7 shows two visualizations of the social networks of 1K agents exemplary for the large synthetic network and NOLA at the end of the 15 months simulation. These visuals show different types of network structures, such as two to three large social communities for the synthetic TownL, and one large community for NOLA. Since it is hard to describe the evolution of a social network over time, we have created a video for each of the four spatial areas showing the social network evolution over the $129,600$ steps within the 15 months simulation time. These videos can be found at: https://mdm2020.joonseok.org, and show how the social networks evolve from small isolated cliques into a large and complex network showing different sub-structures. The video

TABLE III: Location-Based Social Network Simulation Settings.

Settings	Maps	Area (km^2)	# of Sites						# of Neighborhoods	# of Agents
			Total	School	Recreation	Workplace	Restaurant	Apartment		
GMU-1K	GMU	3.36	1,781	1	10	250	20	1,500	1	874
GMU-3K	GMU	3.36	5,341	1	30	750	60	4,500	1	2,589
GMU-5K	GMU	3.36	8,901	1	50	1250	100	7,500	1	4,648
NOLA-1K	NOLA	6.49	1,781	2	10	250	20	1,500	2	863
NOLA-3K	NOLA	6.49	5,342	2	30	750	60	4,500	2	2,720
NOLA-5K	NOLA	6.49	8,904	4	50	1,250	100	7,500	2	4,728
TownS-1K	Town Sm	58.41	1,788	4	12	252	20	1,500	4	876
TownS-3K	Town Sm	58.41	5,348	4	32	752	60	4,500	4	2,645
TownS-5K	Town Sm	58.41	8,908	4	52	1,252	100	7,500	4	4,349
TownL-1K	Town Lg	126.20	1,789	6	12	253	18	1,500	6	853
TownL-3K	Town Lg	126.20	5,346	6	30	750	60	4,500	6	2,550
TownL-5K	Town Lg	126.20	8,904	6	48	1,248	102	7,500	6	4,216

TABLE IV: Data Sets Resulting from Location-Based Social Network Simulation

Settings	Period (mo.)	# of CheckIns	# of Links	Runtime (min)	Size (MB)
GMU-1K	15	2,082,788	9,114,337	53	398
GMU-1K	121	16,210,909	75,747,439	372	3,235
GMU-3K	15	6,229,293	27,650,685	342	1,247
GMU-5K	15	11,189,377	54,250,961	1,099	2,389
NOLA-1K	15	2,099,867	9,160,459	52	400
NOLA-1K	221	29,597,885	141,425,945	774	5,502
NOLA-3K	15	6,886,573	27,284,999	362	1,282
NOLA-5K	15	12,007,415	48,710,881	1,233	2,284
TownS-1K	15	2,101,620	7,643,374	44	359
TownS-3K	15	6,454,785	26,364,057	319	1,227
TownS-5K	15	10,760,008	45,118,825	867	2,093
TownL-1K	15	2,030,688	6,418,473	49	320
TownL-3K	15	6,340,360	22,655,915	400	1,109
TownL-5K	15	10,548,956	40,431,579	942	1,937

also can help explain the patterns observed in Figure 6 in which agents make friends during the day, while loosing some at midnight at which time we periodically lower the weights of the social network.

These videos also show, at each step of the simulation, the distribution of the number of friends per agent, an exemplary one is shown in Figure 8 for the small synthetic network (TownS-1K). We observe that all case studies exhibit realistic long-tail distributions of the number of friends: While most agents have 5-25 friends, there are outliers having 50+ friends, but also agents having only three or fewer friends. This observation agrees well with a limit to the number of people with whom one can maintain stable social relationships, e.g., Dunbar number [105]. In our simulated world such a limit would exist at around 35 friends.

V. Research Applications

Our location-based social network simulation directly benefits many research endeavors concerned with location-based social network study, including social link discovery, location recommendation, community detection, 'next-location' prediction, and potential 'social-link' prediction, among others. This section provides a non-exhaustive brainstorming of applications of our data generator that can benefit the LBSN research community.

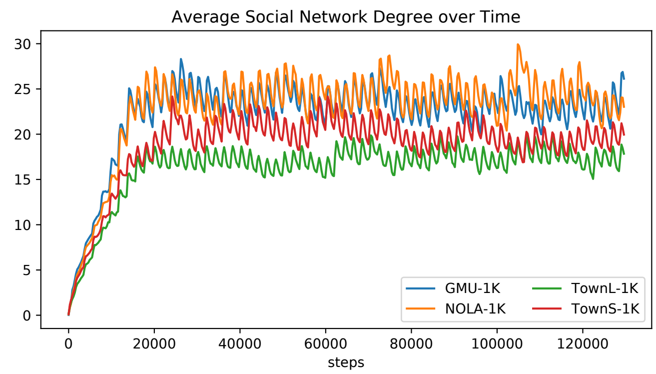
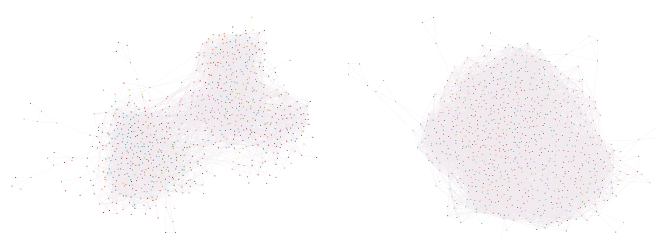


Fig. 6: Average Social Network Degree over Time (1K).



(a) TownL-1K (b) NOLA-1K

Fig. 7: Social Network.

LBSN Benchmark Data. For a fair comparison of different research methods, it is paramount to compare solutions on the same data sets. As discussed in Section II-B, publicly available data sets lack volume, temporal information and ground-truth to reliably generalize knowledge that can be mined from them. As an alternative to using these existing data sets, the spatio-temporal database community has proposed solutions to efficiently crawl data from LBSN data providers [25]. However, this data is intellectual property of the respective LBSN providers, and publishing their data for research benchmarks will violate their license agreements. As it is not possible to crawl data from the past, researchers will ultimately find themselves comparing their solutions on similar, but not identical data sets crawled at different times. Our simulation and generated data repository fills this gap. It allows different research groups, at different times, to evaluate their algorithms

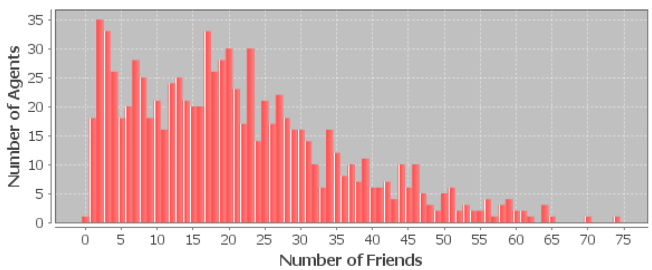


Fig. 8: Social Network Degree Distribution - TownS-1K

on the same data sets. Furthermore, our benchmark data is extensible. If researchers choose to use our simulation to generate a new data set for their particular application, then the corresponding parameter file can be added to our repository. However, for very large LBSN data sets, which may exceed 10GB of filesize, we may only provide the self-executable simulation for local re-generation of data due to bandwidth constraints.

Social Link Prediction. Traditionally, the quality of existing link prediction methods is evaluated by removing a fraction of links from the social network, and testing how well existing solutions can predict these links using the remaining links for training. A major problem with this approach is that it is unclear how accurately the LBSN reflects the real-world. Are there missing links? Are there false links? How much correct signal per noise do these data sets yield? Are existing solutions overfitting to this noise? Unfortunately, these questions are challenging to answer, as there is no way to validate whether two friends that are reported in any of the real-world LBSN data sets (see Section II-B) are actual friends. This is not the case in our simulation, all links in the social network are accurate with no uncertainty. The simulation can be easily extended to create noise, such as giving a chance for ground truth friends to not represent themselves on the social network, or for agents to fail to remove each other from the social network after they unfriend each other. The resulting obfuscated social network can than be used for link prediction, and be evaluated against the ground truth social network.

Location Recommendation. To recommend locations, our simulation allows to have agents *rate* sites (on a five "star" scale - as illustrated in Figure 1). This rating is determined by a deterministic function of the agents' preferences and the locations' attributes. To leverage this simulation for location recommendation, our simulation can be extended to obfuscate ratings by random noise (of parameterizable degree). This obfuscation can be deliberately biased, such as giving low ratings a higher chance to appear, with medium ratings more likely to be omitted. Such data would allow researchers to experimentally compare existing methods and evaluate the effect of bias between observed and ground truth ratings. Such comparison enables us to answer the question of recommendation systems' generalizability to the whole population, or if they overfit their models towards a sub-population of individuals that use the recommendation service.

Community Detection. For the tasks of community detection and social network clustering, we can extend the simulation to impose circles of friends (i.e., strongly connected groups) in our social network. Then, by observing co-locations from the data, we can see which existing solutions are able to best approximate the imposed ground truth social networks. This data generation allows to obtain a ground-truth of communities which can be used to evaluate the accuracy of community detection algorithms.

VI. Conclusions

Our research has demonstrated plausible location-based social simulation generating large-scale and high-fidelity location-based social network (LBSN) data sets. Our Patterns of Life discrete-event simulation addresses the need for sizable LBSN data sets brought on by the lack of real-world data sets that are arguably insufficient in terms of size and data reliability. This current lack of data inhibits data-driven social and data science research. Thus, our version of a location-based social simulation tackles two of LBSN's challenges head-on: (i) plausibility, in terms of generating data that exhibits realistic social behavior, to reason based on the data about real-world phenomena, and (ii) scalability, to simulate plausible numbers of agents over years of simulation time and potentially generate LBSN data for entire generations. Using a location-based social simulation provides for large data sets that enable highly-detailed socio-temporal research on LBSN, including link prediction, next-location prediction, location recommendation, and community detection, etc. We have made our simulation code, executable binaries, generated data sets and visualizations available to the LBSN research community. We hope that these tools will give new life and vitality to this impactful research field. We hope that our data sets can benefit many LBSN applications as exemplary described in Section V by providing massive social network data sets to social scientists and data scientists alike with access to authoritative provenance of social and spatial ground-truth in location-based social network research.

Acknowledgment

This work is partially supported by DARPA cooperative agreement No.HR00111820005 and NSF-CCF 1637541. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

[1] R. Ahuja, N. Armenatzoglou, D. Papadias, and G. J. Fakas. Geo-social keyword search. In *SSTD*, pages 431–450. Springer, 2015.
[2] I. Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.
[3] N. Armenatzoglou, R. Ahuja, and D. Papadias. Geo-social ranking: functions and query processing. *VLDB Journal*, 24(6):783–799, 2015.
[4] N. Armenatzoglou, S. Papadopoulos, and D. Papadias. A general framework for geo-social query processing. *Proc. of the VLDB Endowment*, 6(10):913–924, 2013.
[5] R. Assam and T. Seidl. Check-in location prediction using wavelets and conditional random fields. In *ICDM*, pages 713–718. IEEE, 2014.
[6] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel. Recommendations in location-based social networks: a survey. *GeoInformatica*, 19(3):525–565, 2015.

[7] U. S. census 2000 summary file from 2013-2017 American Community Survey and the Nonprofit Knowledge Works Data Center of New Orleans. French quarter statistical area. https://www.datacenterresearch.org/data-resources/neighborhood-data/district-1/french-quarter (accessed: 01.31.2020).

[8] X. Chen, Y. Zeng, G. Cong, S. Qin, Y. Xiang, and Y. Dai. On information coverage for location category based point-of-interest recommendation. In *AAAI*, pages 37–43, 2015.

[9] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Aaai*, volume 12, pages 17–23, 2012.

[10] C. Cheng, H. Yang, M. R. Lyu, and I. King. Where you like to go next: Successive point-of-interest recommendation. In *IJCAI*, volume 13, pages 2605–2611, 2013.

[11] R. Cheng, J. Pang, and Y. Zhang. Inferring friendship from check-in data of location-based social networks. In *ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 1284–1291. ACM, 2015.

[12] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring millions of footprints in location sharing services. *ICWSM*, 2011:81–88, 2011.

[13] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *KDD*, pages 1082–1090. ACM, 2011.

[14] A. T. Crooks and A. B. Hailegiorgis. An agent-based modeling approach applied to the spread of cholera. *Environmental Modelling & Software*, 62:164–177, 2014.

[15] T.-N. Doan and E.-P. Lim. Modeling location-based social network data with area attraction and neighborhood competition. *Data Mining and Knowledge Discovery*, 33(1):58–95, 2019.

[16] I. Environmental Systems Research Institute. ESRI Shapefile Technical Description. Technical report, Environmental Systems Research Institute, Inc, 1998.

[17] S. Feng, X. Li, Y. Zeng, G. Cong, Y. M. Chee, and Q. Yuan. Personalized ranking metric embedding for next new poi recommendation. In *IJCAI*, pages 2069–2075, 2015.

[18] G. Ference, M. Ye, and W.-C. Lee. Location recommendation for out-of-town users in location-based social networks. In *CIKM*, pages 721–726. ACM, 2013.

[19] H. Gao, J. Tang, X. Hu, and H. Liu. Exploring temporal effects for location recommendation on location-based social networks. In *Proc. 7th ACM Conf. Recommender systems*, pages 93–100. ACM, 2013.

[20] H. Gao, J. Tang, X. Hu, and H. Liu. Content-aware point of interest recommendation on location-based social networks. In *AAAI*, pages 1721–1727, 2015.

[21] George Mason University campus data . http://librarygeodata.gmu.edu.mutex.gmu.edu/campus.html, 2009.

[22] J. He, X. Li, L. Liao, D. Song, and W. K. Cheung. Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In *AAAI*, pages 137–143, 2016.

[23] H.-P. Hsieh, R. Yan, and C.-T. Li. Where you go reveals who you know: analyzing social ties from millions of footprints. In *CIKM*, pages 1839–1842. ACM, 2015.

[24] B. Hu and M. Ester. Spatial topic modeling in online social media for location recommendation. In *Proc. 7th ACM conference on Recommender systems*, pages 25–32. ACM, 2013.

[25] S. Isaj and T. B. Pedersen. Seed-driven geo-social data extraction. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, pages 11–20, 2019.

[26] JOSM: An extensible editor for OpenStreetMap . https://josm.openstreetmap.de/wiki/Download, 2020.

[27] H. Kavak, J.-S. Kim, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Location-based social simulation. In *SSTD*, pages 218–221, 2019.

[28] J.-S. Kim, H. Kavak, and A. Crooks. Procedural city generation beyond game development. *SIGSPATIAL Special*, 10(2):34–41, 2018.

[29] J.-S. Kim, H. Kavak, U. Manzoor, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Simulating urban patterns of life: A geo-social data generation framework. In *SIGSPATIAL*, pages 576–579, 2019.

[30] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel. Lars: A location-aware recommender system. In *ICDE*, pages 450–461. IEEE, 2012.

[31] H. Li, Y. Ge, R. Hong, and H. Zhu. Point-of-interest recommendations: Learning potential check-ins from friends. In *KDD*, pages 975–984, 2016.

[32] H. Li, R. Hong, S. Zhu, and Y. Ge. Point-of-interest recommender systems: A separate-space perspective. In *ICDM*, pages 231–240. IEEE, 2015.

[33] M. Li, R. Westerholt, H. Fan, and A. Zipf. Assessing spatiotemporal predictability of lbsn: a case study of three foursquare datasets. *GeoInformatica*, pages 1–21, 2016.

[34] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy. Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In *SIGIR*, pages 433–442. ACM, 2015.

[35] Y. Li, R. Chen, J. Xu, Q. Huang, H. Hu, and B. Choi. Geo-social k-cover group queries for collaborative spatial computing. *TKDE*, 27(10):2729–2742, 2015.

[36] D. Lian, V. W. Zheng, and X. Xie. Collaborative filtering meets next check-in location prediction. In *WWW*, pages 231–232. ACM, 2013.

[37] G. Liao, S. Jiang, Z. Zhou, C. Wan, and X. Liu. Poi recommendation of location-based social networks using tensor factorization. In *MDM*, pages 116–124. IEEE, 2018.

[38] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *ACM SIGKDD*, pages 1043–1051, 2013.

[39] B. Liu, H. Xiong, S. Papadimitriou, Y. Fu, and Z. Yao. A general geographical probabilistic factor model for point of interest recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 27(5):1167–1179, 2015.

[40] Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *AAAI*, pages 194–200, 2016.

[41] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *CIKM*, pages 733–738. ACM, 2013.

[42] Y. Liu, C. Liu, B. Liu, M. Qu, and H. Xiong. Unified point-of-interest recommendation with temporal interval assessment. In *KDD*, pages 1015–1024, 2016.

[43] Y. Liu, T.-A. N. Pham, G. Cong, and Q. Yuan. An experimental evaluation of point-of-interest recommendation in location-based social networks. *Proc. VLDB Endowment*, 10(10):1010–1021, 2017.

[44] Y. Liu, W. Wei, A. Sun, and C. Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *CIKM*, pages 739–748. ACM, 2014.

[45] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. Mason: A multiagent simulation environment. *Simulation*, 81(7):517–527, 2005.

[46] S. Luke, R. Simon, A. Crooks, H. Wang, E. Wei, D. Freelan, C. Spagnuolo, V. Scarano, G. Cordasco, and C. Cioffi-Revilla. The mason simulation toolkit: Past, present, and future. In P. Davidsson and H. Verhagen, editors, *Multi-Agent-Based Simulation XIX*, pages 75–86, Cham, 2019. Springer International Publishing.

[47] A. H. Maslow. A theory of human motivation. *Psychological review*, 50(4):370, 1943.

[48] G. S. Njoo, M.-C. Kao, K.-W. Hsu, and W.-C. Peng. Exploring check-in data to infer social ties in location based social networks. In *PAKDD*, pages 460–471. Springer, 2017.

[49] A. Noulas, C. Mascolo, and E. Frias-Martinez. Exploiting foursquare and cellular data to infer user activity in urban environments. In *MDM*, volume 1, pages 167–176. IEEE, 2013.

[50] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *ICDM*, pages 1038–1043. IEEE, 2012.

[51] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. A random walk around the city: New venue recommendation in location-based social networks. In *PASSAT (socialcom)*, pages 144–153. Ieee, 2012.

[52] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An empirical study of geographic user activity patterns in Foursquare. *ICWSM*, 11:70–573, 2011.

[53] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. *The social mobile web*, 11(2), 2011.

[54] OpenStreetMap. https://www.openstreetmap.org/, 2019.

[55] A. Patel, A. Crooks, and N. Koizumi. Slumulation: an agent-based modeling approach to slum formations. *Artificial Societies and Soc. Simulation*, 15(4), 2012.

[56] H. Pham, C. Shahabi, and Y. Liu. Ebm: an entropy-based model to infer social strength from spatiotemporal data. In *ACM SIGMOD*, pages 265–276. ACM, 2013.

[57] B. Pires and A. T. Crooks. Modeling the emergence of riots: a geosimulation approach. *Computers, Environment and Urban Systems*, 61:66–80, 2017.

[58] QGIS. Qgis: A free and open source geographic information system. https://qgis.org/en/site/forusers/download.html, 2019.

[59] M. A. Saleem, F. S. Da Costa, P. Dolog, P. Karras, T. B. Pedersen, and T. Calders. Predicting visitors using location-based social networks. In *MDM*, pages 245–250. IEEE, 2018.

[60] M. A. Saleem, X. Xie, and T. B. Pedersen. Scalable processing of location-based social networking queries. In *MDM*, volume 1, pages 132–141. IEEE, 2016.

[61] M. Sarwat, A. Eldawy, M. F. Mokbel, and J. Riedl. Plutus: leveraging location-based social networks to recommend potential customers to venues. In *MDM*, volume 1, pages 26–35. IEEE, 2013.

[62] M. Sarwat, J. J. Levandoski, A. Eldawy, and M. F. Mokbel. Lars*: An efficient and scalable location-aware recommender system. *TKDE*, 26(6):1384–1399, 2014.

[63] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora. Distance matters: Geo-social metrics for online social networks. In *WOSN*, 2010.

[64] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. *ICWSM*, 11:329–336, 2011.

[65] S. Scellato, A. Noulas, and C. Mascolo. Exploiting place features in link prediction on location-based social networks. In *ACM SIGKDD*, pages 1046–1054, 2011.

[66] Stanford network analysis project. https://snap.stanford.edu/index.html.

[67] Y. Su, X. Li, W. Tang, J. Xiang, and Y. He. Next check-in location prediction via footprints and friendship on location-based social networks. In *MDM*, pages 251–256. IEEE, 2018.

[68] K. Sullivan, M. Coletti, and S. Luke. Geomason: Geospatial support for mason. Technical report, George Mason University, 2010.

[69] F. Wang, G. Wang, and S. Y. Philip. Why checkins: Exploring user motivation on location based social networks. In *Data Mining Workshop (ICDMW)*, pages 27–34. IEEE, 2014.

[70] H. Wang, Z. Li, and W.-C. Lee. PGT: Measuring mobility relationship using personal, global and temporal factors. In *ICDM*, pages 570–579. IEEE, 2014.

[71] H. Wang, M. Terrovitis, and N. Mamoulis. Location recommendation in location-based social networks using user check-in data. In *ACM SIGSPATIAL*, pages 374–383, 2013.

[72] W. Wang, H. Yin, L. Chen, Y. Sun, S. Sadiq, and X. Zhou. Geo-sage: A geographical sparse additive generative model for spatial item recommendation. In *ACM SIGKDD*, pages 1255–1264, 2015.

[73] X. Wang, L. Wang, and P. Yang. Prevalent co-visiting patterns mining from location-based social networks. In *MDM*, pages 581–586. IEEE, 2019.

[74] Z. Wang, D. Zhang, X. Zhou, D. Yang, Z. Yu, and Z. Yu. Discovering and profiling overlapping communities in location-based social networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(4):499–509, 2014.

[75] Y.-T. Wen, Y. Y. Fan, and W.-C. Peng. Mining of location-based social networks for spatio-temporal social influence. In *PAKDD*, pages 799–810. Springer, 2017.

[76] Y.-T. Wen, P.-R. Lei, W.-C. Peng, and X.-F. Zhou. Exploring social influence on location-based social networks. In *ICDM*, pages 1043–1048. IEEE, 2014.

[77] G. Xu-Rui, W. Li, and W. Wei-Li. Using multi-features to recommend friends on location-based social networks. *Peer-to-Peer Networking and Applications*, pages 1–8, 2016.

[78] D. Yang, B. Qu, J. Yang, and P. Cudre-Mauroux. Revisiting user mobility and social relationships in lbsns: A hypergraph embedding approach. In *Proc. Web Conference*, WWW'19, 2019.

[79] D. Yang, D. Zhang, L. Chen, and B. Qu. Nationtelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications*, 55:170–180, 2015.

[80] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1):129–142, 2015.

[81] G. Yang and A. Züfle. Spatio-temporal site recommendation. In *Data Mining Workshops (ICDMW)*, pages 1173–1178. IEEE, 2016.

[82] G. Yang and A. Züfle. Spatio-temporal prediction of social connections. In *ACM Workshop on Managing and Mining Enriched Geo-Spatial Data (GeoRich)*. ACM, 2017.

[83] K. Yao, D. Papadias, and S. Bakiras. Density-based community detection in geo-social networks. In *Proc. 16th International Symposium on Spatial and Temporal Databases*, pages 110–119, 2019.

[84] J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activity prediction in location-based social networks. In *SIAM Data Mining (SDM)*, pages 171–179. SIAM, 2013.

[85] M. Ye, P. Yin, and W.-C. Lee. Location recommendation for location-based social networks. In *ACM SIGSPATIAL*, pages 458–461, 2010.

[86] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, pages 325–334. ACM, 2011.

[87] Yelp dataset challenge. round 9. https://www.yelp.com/dataset_challenge. Accessed: 2017-07-30.

[88] H. Yin, Z. Hu, X. Zhou, H. Wang, K. Zheng, Q. V. H. Nguyen, and S. Sadiq. Discovering interpretable geo-social communities for user behavior prediction. In *ICDE*, pages 942–953. IEEE, 2016.

[89] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen. Lcars: a location-content-aware recommender system. In *ACM SIGKDD*, pages 221–229, 2013.

[90] H. Yin, X. Zhou, Y. Shao, H. Wang, and S. Sadiq. Joint modeling of user check-in behaviors for point-of-interest recommendation. In *CIKM*, pages 1631–1640. ACM, 2015.

[91] F. Yu, Z. Li, S. Jiang, and S. Lin. Point-of-interest recommendation for location promotion in location-based social networks. In *MDM*, pages 344–347. IEEE, 2017.

[92] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *SIGIR*, pages 363–372. ACM, 2013.

[93] Q. Yuan, G. Cong, and A. Sun. Graph-based point-of-interest recommendation with geographical and temporal influences. In *CIKM*, pages 659–668. ACM, 2014.

[94] J.-D. Zhang and C.-Y. Chow. igslr: personalized geo-social location recommendation: a kernel density estimation approach. In *ACM SIGSPATIAL*, pages 334–343, 2013.

[95] J.-D. Zhang and C.-Y. Chow. Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *SIGIR*, pages 443–452. ACM, 2015.

[96] J.-D. Zhang and C.-Y. Chow. Spatiotemporal sequential influence modeling for location recommendations: A gravity-based approach. *TIST*, 7(1):11, 2015.

[97] J.-D. Zhang, C.-Y. Chow, and Y. Li. Lore: Exploiting sequential influence for location recommendations. In *ACM SIGSPATIAL*, pages 103–112, 2014.

[98] J.-D. Zhang, C.-Y. Chow, and Y. Zheng. Orec: An opinion-based point-of-interest recommendation framework. In *CIKM*, pages 1641–1650. ACM, 2015.

[99] K. Zhao, G. Cong, Q. Yuan, and K. Q. Zhu. Sar: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In *ICDE*, pages 675–686. IEEE, 2015.

[100] S. Zhao, T. Zhao, H. Yang, M. R. Lyu, and I. King. Stellar: Spatial-temporal latent ranking for successive point-of-interest recommendation. In *AAAI*, pages 315–322, 2016.

[101] Y.-L. Zhao, Q. Chen, S. Yan, D. Zhang, and T.-S. Chua. Community understanding in location-based social networks. In *Human-Centered Social Media Analytics*, pages 43–74. Springer, 2014.

[102] Y. Zheng. Location-based social networks: Users. In *Computing with spatial trajectories*, pages 243–276. Springer, 2011.

[103] Y. Zheng, Y. Chen, X. Xie, and W.-Y. Ma. Geolife 2.0: a location-based social networking service. In *mobile data management: systems, services and middleware*, pages 357–358. IEEE, 2009.

[104] Y. Zheng and X. Xie. Location-based social networks: Locations. In *Computing with spatial trajectories*, pages 277–308. Springer, 2011.

[105] W.-X. Zhou, D. Sornette, R. A. Hill, and R. I. Dunbar. Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 272(1561):439–444, 2005.

[106] Q. Zhu, H. Hu, J. Xu, and W.-C. Lee. Geo-social group queries with minimum acquaintance constraint. *arXiv:1406.7367*, 2014.

[107] W.-Y. Zhu, Y.-W. Wang, C.-J. Chen, W.-C. Peng, and P.-R. Lei. A bayesian-based approach for activity and mobility inference in location-based social networks. In *MDM*, pages 152–157. IEEE, 2016.