Received: 4 January 2017 | Revised: 30 March 2017 | Accepted: 11 April 2017

DOI: 10.1111/tgis.12285

RESEARCH ARTICLE

WILEY Transactions in GIS

How friends share urban space: An exploratory spatiotemporal analysis using mobile phone data

Yang Xu[1] | Alexander Belyi[1] | Iva Bojic[1,2] | Carlo Ratti[2]

[1] Senseable City Laboratory, SMART Centre, Singapore

[2] Senseable City Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

Correspondence
Yang Xu, Senseable City Laboratory, SMART Centre, 1 Create Way, Singapore.
Email: yangxu@smart.mit.edu

Funding information
MIT Senseable City Lab Consortium; National Research Foundation, Prime Minister's Office, Singapore; Singapore-MIT Alliance for Research and Technology (SMART) Future Urban Mobility (FM) IRG; Croatian Science Foundation, Grant/Award No. UIP 1120138813.

Abstract

This study proposes a framework to investigate the roles of urban spaces in connecting social contacts (i.e., "friends"). The framework is applied to a Call Detail Record (CDR) dataset collected in Singapore. First, a comparative analysis is performed to understand how friends share urban space differently from random people. Then, we derive two metrics to quantify the "bonding" and "bridging" capabilities of places in the city. The two metrics reflect the potential of a place in connecting friends and random people (e.g., chance encounters), respectively. Finally, we examine the temporal signature of the places' bonding capabilities, and associate the results with various types of Points of Interest (POIs). We find that: (1) friends are more likely to share urban space than random people, and they also share more locations; (2) a place could play different roles in connecting friends vs. random people, and the relationship (between bonding and bridging) varies depending on the time and type of a day (weekdays vs. weekends); (3) the temporal signature of bonding capability is strongly related to the semantics of a place; (4) certain POI types (e.g., shopping malls) tend to have a much higher impact on bonding capability than others (e.g., sports centers).

1 | INTRODUCTION

We are living in an age where human social interactions are becoming increasingly convenient. People are able to approach others at any time, from anywhere, through phones, emails, and social networking tools (e.g., Facebook and Twitter). The communications evolution has enriched the ways people socialize, and triggered extensive discussions on the impact of geography on social network structures (Cairncross, 2001; Goldenberg & Levy, 2009). Although we are no longer bounded when interacting with others, physical space still serves as a primary channel where social ties form and evolve. We tend to build our social capital at and around places where we live, work and entertain. Understanding how social networks are tied to human activities in geographic space has great implications for urban design, economic development, and social well-being, among others.

Recent advancements of information and location-aware technologies have generated many new data sources (e.g., location-based social networks, mobile phone data) that capture human movements and social interactions simultaneously. These datasets enable researchers to better understand the interplay between human mobility patterns and social network structures. During the past few years, several studies have been conducted to infer human movement patterns based on social relations (Backstrom, Sun, & Marlow, 2010; Cho, Myers, & Leskovec, 2011), or to predict new social links based on mobility similarities among individuals (Eagle, Pentland, & Lazer, 2009; Wang, Pedreschi, Song, Giannotti, & Barabási, 2011). These studies offer valuable insights into how mobility and social network structures influence each other. However, few studies attempt to link these two dimensions within the context of urban space. Our knowledge of how people share urban space with their social contacts ("friends") is still limited. A systematic investigation of this question could distinguish the roles that urban spaces play in connecting people, and suggest how cities can be better designed to accommodate people's social interactions.

Using a Call Detail Record (CDR) dataset collected in Singapore, this study proposes a framework to gain insights into the spatiotemporal characteristics of friends' use of urban space. First, we introduce a spatial co-location measure to quantify how likely a given pair of cellphone users appears at the same location at approximately the same time. By extracting a city-scale social network from the mobile phone dataset, we apply the measure to both friend pairs and random user pairs in the network. A comparative analysis is then performed to better understand how friends share urban space differently from random people. Based on the co-location patterns observed from the two sets of user pairs, we further derive two place-based co-location metrics to quantify the "bonding" and "bridging" capabilities of the places, respectively. The "bonding" capability reflects the potential of a place in bringing friends together, while the "bridging" capability describes how likely two random people tend to co-locate at a given location. Finally, we apply a hierarchical clustering algorithm to investigate the temporal signature of the places' bonding capabilities. The clustering results are associated with various types of Points of Interest (POIs) to reveal the relationships between the semantics of the places and their bonding capabilities.

2 | LITERATURE REVIEW

2.1 | Social ties and distance decay

Internet and mobile technologies have enriched the ways people socialize with each other. Since the turn of the century, debates have emerged over how these new technologies will redefine the role of geography in shaping our daily communication patterns (Cairncross, 2001; Graham, 1998). Inspired by these discussions, scholars started examining the geographic properties of various human social networks, with considerable focus on the distance decay effect. For example, using data from LiveJournal network, Liben-Nowell, Novak, Kumar, Raghavan, and Tomkins (2005) found that the probability of friendship and the geographic distance (which separates the friend pairs) follows a decay function $P(d) \sim d^{-1.2}$. Similar patterns are discovered in Facebook communities (Backstrom et al., 2010; Goldenberg & Levy, 2009) that friendship probability is inversely proportional to the geographic distance $P(d) \sim d^{-1.0}$. The findings illustrate that geography plays an important role in manifesting online social network structures.

In recent years, mobile phone data have emerged as a new data source for modeling such social-spatial relationships. In particular, Lambiotte et al. (2008) and Krings, Calabrese, Ratti, and Blondel (2009) find that the intensity of inter-city telecommunications in Belgium can be well approximated by a gravity model with a scaling exponent of $\alpha = 2$. A similar distance-decay effect is observed in an inter-city mobile communication network in China (Kang, Zhang, Ma, & Liu, 2013), but with a different scaling exponent ($\alpha = 0.5$). These studies have generated valuable insights into how spatial proximity affects social network structures. However, the interplay between human social interactions and their daily activity patterns remains unclear. The variations in the scaling exponents indicate that the impacts of physical space on social network structures are intertwined. Questions regarding how social relations are tied to human mobility patterns need to be better addressed.

2.2 | Coupling human mobility and social network analysis

Human movements exhibit a high degree of spatial-temporal regularity (González, Hidalgo, & Barabási, 2008; Song, Qu, Blumm, & Barabási, 2010). Yet such regularity depicts a partial aspect of human behavioral patterns. Friendship and social ties, for example, motivate people to travel and participate in activities beyond their daily routines. By analyzing mobile phone data and online location-based social networks (LBSNs), Cho, Myers and Leskovec (2011) conclude that social relationships explain 10 to 30% of human movements, especially for long-distance travel. Similarly, Backstorm, Sun and Marlow (2010) find that social network structures can be leveraged to better predict the physical locations of Facebook users. These studies suggest that social relations could lead to a better forecast of human mobility patterns in space and time.

Instead of inferring movement patterns from social ties information, many scholars focus on predicting new links in social networks by measuring mobility similarities among individuals (Crandall et al., 2010; Eagle et al., 2009; Wang et al., 2011). Besides their contributions to social link predictions, these studies reveal that people who are closer to each other in social space are more likely to have co-location patterns in geographic space. The correlations between social proximity and mobility similarity enable researchers to model human movement choices more precisely (Toole, Herrera-Yaqüe, Schneider, & González, 2015).

2.3 | Embedding social networks in urban space

Although the preceding studies have integrated the analysis of human mobility and social networks, the role of urban space in connecting the two has not been addressed explicitly. Cities create massive opportunities of human social interactions. Yet the spatial distance that separates people is always a constraint on face-to-face communications. There have been continuous discussions on whether human travel and telecommunications complement or substitute for each other (Albertson, 1977; Choo, Lee, & Mokhtarian, 2007; Mok, Wellman, & Carrasco, 2010). To better answer these questions, survey data are used to assess the geographic properties of social interactions that occurred in different forms (e.g., face-to-face, email and telecommunications) and among different social-economic groups (Carrasco, Miller, & Wellman, 2008; Larsen, Axhausen, & Urry, 2006).

These attempts are followed by scholars who rely on other emerging datasets to investigate the relationships among social ties, mobility and physical locations. In particular, Cranshaw, Toch, Hong, Kittur, and Sadeh (2010) introduce location entropy to quantify the diversity of social interactions at a particular place, using location traces of 489 users collected by laptop and cell phones. The location entropy measure is used to better distinguish the co-location patterns among friends and among strangers. However, as 93.7% of the location data are collected when people are with their laptops, the findings are biased towards social interactions that take place at certain locations (e.g., home and office).

Instead, Calabrese, Smoreda, Blondel, and Ratti (2011) used a large mobile phone dataset in Portugal to assess the interplay between telecommunications and face-to-face interactions. They find that more than 90% of users who have called each other have also shared the same place, and the expected number of co-locations decreases with distance between homes. The study brings a new perspective of examining social interactions within urban context. However, when and where people tend to co-locate with their friends was not discussed. By using a mobile phone dataset in Jiamusi, China, Wang, Kang, Bettencourt, Liu, and Andris (2015) find that friends are more likely to share urban space than a random pair of users, and downtown is used by many social groups while each suburb only hosts one or two. Using the same type of dataset, a more recent study quantifies the potential effects of space and place on social network structures (Shi, Wu, Chi, & Liu, 2016). These studies advance our understanding of how social contacts share urban space, and suggest the importance of embedding social network analysis in geographic information systems (Andris, 2016). However, few studies have attempted to quantify the roles of urban spaces and their characteristics in fostering social engagement. Given this as a critical issue in urban planning and policy making (Holland, Clark, Katz, & Peace, 2007; Madanipour, 1999), it is important to ask questions about how effectively urban space and places bring together people, the nature of those places, and what kinds of patterns arise in cities.

3 | RESEARCH DESIGN

3.1 | Establishing the social network structure from mobile phone data

Singapore is a city-state that covers a total area of 719 km^2. It had a population of 5.18 million in 2010. The country has achieved rapid economic growth in the past half century and it is now a global finance and transport hub. As a city recognized by its livable urban environment, Singapore has many characteristics that facilitate social bonding and community engagement. For example, the city is deployed with affordable transit services, allowing people to travel among destinations conveniently during most time periods in a day. Various shopping malls, community clubs, and food hawker centers – hierarchically clustered across different regions – enable people to socialize through a diverse set of leisure and recreational activities. Understanding how urban spaces are used and shared by people in Singapore would thus provide insights into its social and cultural environment, and inspire new thoughts on societal well-being and sustainable urban development.

The CDR dataset used in this study is collected by a major mobile phone carrier in Singapore.[1] The anonymized dataset covers 4.4 million cellphone users during a period of 50 days (21 March to 9 May, 2011). Each mobile phone record tracks the unique ID of the caller and the callee, the communication type (i.e., call/SMS), as well as the date, time and the phone users' connected cell tower when the phone communication starts. There are about 5,000 cellphone towers that are densely distributed across the whole of Singapore, and the average nearest distance between them is about 100 m. Considering that CDRs are passively collected during phone call communications, to control the issue of data sparsity, this research focuses on a subset of cellphone users, who have at least 25 active days of phone usage (e.g., made or received a call/text message). This allows us to mitigate the data sparsity issue by filtering individuals: (1) who are short-term subscribers (e.g., tourists), and (2) who have inactive phone usage during the study period. Thus, the choice of 25 active days provides a good balance between including subscribers who seldom use a phone and excluding too many users. The resulting dataset after removing these individuals consists of 2.1 million cellphone users.

To establish the social network structure, we measure the communication patterns between pairs of users. Let $G(V, E)$ denote an undirected graph, where V denotes the set of cellphone users, and E denotes the set of edges which correspond to social ties. Given two cellphone users x and y, we add a link (x, y) to G if we observe at least one reciprocal call between them during the study period. The undirected graph after performing the link generation consists of $|V|=2{,}131{,}285$ nodes and $|E|=9{,}071{,}808$ edges. The average node degree (i.e., number of social contacts per user) is 8.5, and the first, second (i.e., median) and third quartiles are 1.0, 5.0 and 12.0, respectively. In the remainder of this article, we refer to the pairs of phone users that are connected in G as friend pairs.

3.2 | Inferring individual home and work locations

To evaluate whether the dataset reflects the population distribution in Singapore, we estimate each individual's home (and work) location. The estimated home locations are then aggregated by planning area where census data are available for comparison. The process of inferring home location is still an open question (Bojic, Massaro, Belyi, Sobolevsky, & Ratti, 2015) and there have been many studies which discuss how home and work locations can be inferred from CDR data (e.g., Ahas, Silm, Järv, Saluveer, & Tiru, 2010; Isaacman et al., 2011; Xu et al., 2015, 2016). In this study, we estimate each individual's home location as the most used cellphone tower before 06:00 and after 19:00, whereas work location is identified similarly for time between 08:00 and 17:00 on workdays. After applying the location detection algorithm, we calculate the total number of cellphone users with home location in each planning area.[2] As shown in Figure 1a, we find that the total number of cellphone users sampled in each planning area is strongly correlated with the population distribution recorded by the census data (Department of Statistics Singapore 2010; http://www.singstat.gov.sg/), with a Pearson's correlation coefficient of 0.98.

Some previous studies find that the probability of friendship in a social network decays with geographic distance (Backstrom et al., 2010; Goldenberg & Levy, 2009; Lambiotte et al., 2008). To explore whether our dataset exhibits similar characteristics, we compute: (1) the total number of connected nodes C_d; and (2) the number of all pairs of nodes N_d separated at a distance d (between their home locations) in graph G. We repeat this procedure for different distance values and calculate the probability of friendship at these distances $p(d)=C_d/N_d$. As shown in Figure 1b, the probability that two people are connected at a distance follows a power law function $P(d) \sim d^{-0.95}$. The scaling exponent is very close to the value (i.e., -1.0) derived from Facebook user communities (Backstrom et al., 2010; Goldenberg & Levy, 2009).

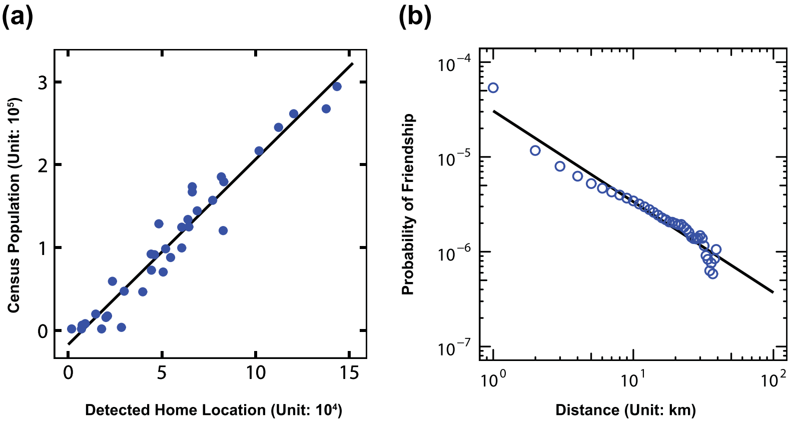

FIGURE 1 (a) Correlation between the number of detected home locations vs. census population (by planning area in Singapore). (b) The probability $p(d)$ of friendship among people who live at a distance d follows $P(d)\sim d^{-0.95}$. The fluctuation of the last several points (i.e., when $d \geq 35$ km) is caused by the boundary effect when the distance d approaches the diameter of Singapore. After removing those points, the exponent changes from -0.95 to -0.91

3.3 | Measuring co-location patterns between individuals

People share urban space with others for a variety of reasons. Some are planned (e.g., face-to-face meetings) and some are not (e.g., encountering strangers in metro stations). To examine how friends share urban space differently from random people, we need to measure how likely a cellphone user pair tends to appear at the same location at approximately the same time. To achieve this, we first divide the study area into a 500 m regular grid, and map cellphone users' location traces on to grid cells. We choose 500 m because on the one hand, it is relatively fine-grained to capture the dynamics of people's space usage. On the other hand, since cellphone towers can be densely distributed in populated areas (e.g., two cellphone towers that are vertically distributed in a building), using 500 m regular grid enables us to aggregate cellphone towers that are very close to each other, so that we can systematically evaluate phone users' co-location patterns.

Specifically, for a given cellphone user, we assign each of his/her location records to the grid cell where the corresponding serving cellphone tower locates. Although we are aware that Voronoi Polygons are often used to approximate a cellphone tower's service area (Xu et al., 2015; Yuan & Raubal, 2012), which could overlap with multiple grid cells, it appears that no consensus has been reached by existing studies on how cellphone tower service areas can be reasonably reconstructed. Moreover, given a very high density of cellphone tower deployment in Singapore, some Voronoi Polygons are too small to represent any meaningful location/place in the city. Hence, the current approach considers each grid cell as a collection of cellphone towers which reflect the general area where the cellphone users have stayed. One advantage of this approach is that when measuring co-location patterns of two cellphone users, if they appear at the same cellphone tower, they will always be assigned to the same grid cell.

We next introduce how co-location patterns can be measured between a given pair of cellphone users. As cellphone users' locations are only available when they engage in phone call/SMS communications, it is not appropriate to directly measure the number of times or the duration that two individuals co-locate with each other. Hence, we adopt the concept of spatial co-location rate used in Wang et al. (2011) to describe the probability that two individuals are co-located in space during a certain period of time.

Given the cellphone trace of an individual x as a list of tuples $\{(l_1,t_1),(l_2,t_2),\ldots(l_n,t_n)\}$, where l_i denotes the user's location (i.e., grid cell) at time point t_i, the probability that user x stays at a location L during a defined time period T is:

$$p_x(L,T)=m_x(L,T)/n_x(T) \tag{1}$$

where $m_x(L,T)$ denotes the total number of times x is observed at location L during T, and $n_x(T)$ denotes the total number of times x is observed during time period T at any location. Note that:

$$\sum_{L \in Loc_x(T)} m_x(L,T) = n_x(T) \tag{2}$$

where $Loc_x(T)$ is the set of all locations visited by x during time period T.

Note also that an individual's cellphone communications do not take place regularly over time. People could make several phone calls in a short period of time and then none for hours (Barabási, 2010; Candia et al., 2008). Hence, $p_x(L,T)$ could be biased due to the "bursty" nature of CDRs. To control this effect, when we measure $m_x(L,T)$, if an individual x is observed multiple times at location L during a one-hour time window (e.g., 07:00 – 08:00), we only consider them as one entry.

Given two individuals x and y, assuming that the probabilities of their visits to the same location are independent, we can calculate the probability that they are co-located at a given location L during time period T:

$$Col_{x,y}(L,T) = p_x(L,T) \times p_y(L,T) \tag{3}$$

Iterating this process through all the locations $L \in Loc_{x,y}(T)$ gives us a spatial co-location vector $[Col_{x,y}(L_1,T), Col_{x,y}(L_2,T), \ldots, Col_{x,y}(L_m,T)]$, where m denotes the total number of grid cells in the study area. The spatial co-location rate between x and y is calculated as the sum of the values in the spatial co-location vector:

$$SCol_{x,y}(T) = \sum_{L \in Loc_{x,y}(T)} Col_{x,y}(L,T) \tag{4}$$

The total number of locations shared by the user pair during T is:

$$NCol_{x,y}(T) = \sum_{L \in Loc_{x,y}(T)} [Col_{x,y}(L,T) > 0] \tag{5}$$

where $[Col_{x,y}(L,T) > 0]$ takes the value of 1 if $Col_{x,y}(L,T) > 0$, and 0 otherwise. We refer the spatial co-location rate and number of shared locations as the *user-pair based co-location metric*.

3.4 | Defining the bonding and bridging roles of places

The spatial co-location rate describes how likely a cellphone user pair shares the urban spaces. The spatial co-location vector, instead, distinguishes the roles that different locations play in bringing people together. Here we introduce a place-based metric, $PCol_U(L,T)$, to measure the average co-location rate of a set of user pairs U (which can be a group of friend pairs or a collection of random user pairs) at location L during time period T.

Figure 2 illustrates how the place-based metric is derived, using two pairs of individuals and four distinct locations (i.e., grid cells) as an example. Let U denote a set of user pairs (here $|U|=2$), given a user pair $(x,y) \in U$, we can compute the probabilities of their stay at different locations during time period T. By deriving the spatial co-location

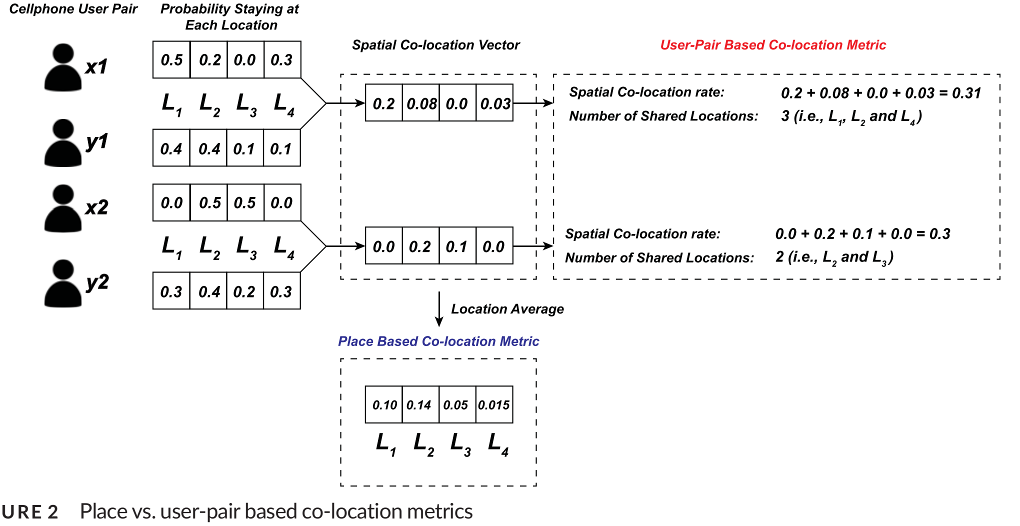

FIGURE 2 Place vs. user-pair based co-location metrics

vectors, we can generate the spatial co-location rate and number of shared locations for each user pair (i.e., user-pair based co-location metric shown in the figure).

Based on the derived spatial co-location vectors, we further aggregate the vectors in space, and calculate the average co-location rate of all user pairs at each place:

$$PCol_U(L,T) = \sum_{(x,y)\in U} Col_{x,y}(L,T)/|U| \tag{6}$$

The meaning of the placed-based metric is intuitive. For example, if the metric is computed over all friend pairs in a city, it refers to the average probability that each place is shared by the friend pairs. A place with a high value indicates that this place has a large potential of bringing friends together in the city. Instead, if we compute the metric over random user pairs, it will describe the average chance that two random people tend to co-locate at a place.

In sociology, two important concepts – *bonding* and *bridging* – are used to describe distinctive characteristics of social capital and civil engagement (Putnam & Goss, 2002). In particular, *bonding* networks tend to link people "who are like one another in important aspects", whereas *bridging* refers to "social networks that bring together people who are unlike one another" (Putnam & Goss, 2002, p. 11). Thus, we refer to the two metrics derived in this section – one for friend pairs and the other for random user pairs – as the *bonding* and *bridging* capabilities of places, respectively. The two place-based metrics can be used to compare co-location patterns of friends and random people from the perspective of urban space.

3.5 | Comparing co-location patterns: Friend pairs vs. random user pairs

We propose a framework to compare co-location patterns of friends and random people from the perspectives of: (1) user pairs and (2) urban space. As shown in Figure 3, based on the social network G extracted from the CDR dataset, we generate two sets of user pairs by: (1) drawing all friend pairs (denoted as set U_s), and (2) drawing random user pairs no matter if they are friends or not (denoted as set U_r). We then compute the spatial co-location vectors for each user pair in U_s and U_r, which are then used to derive the user-pair based and place-based co-location metrics.

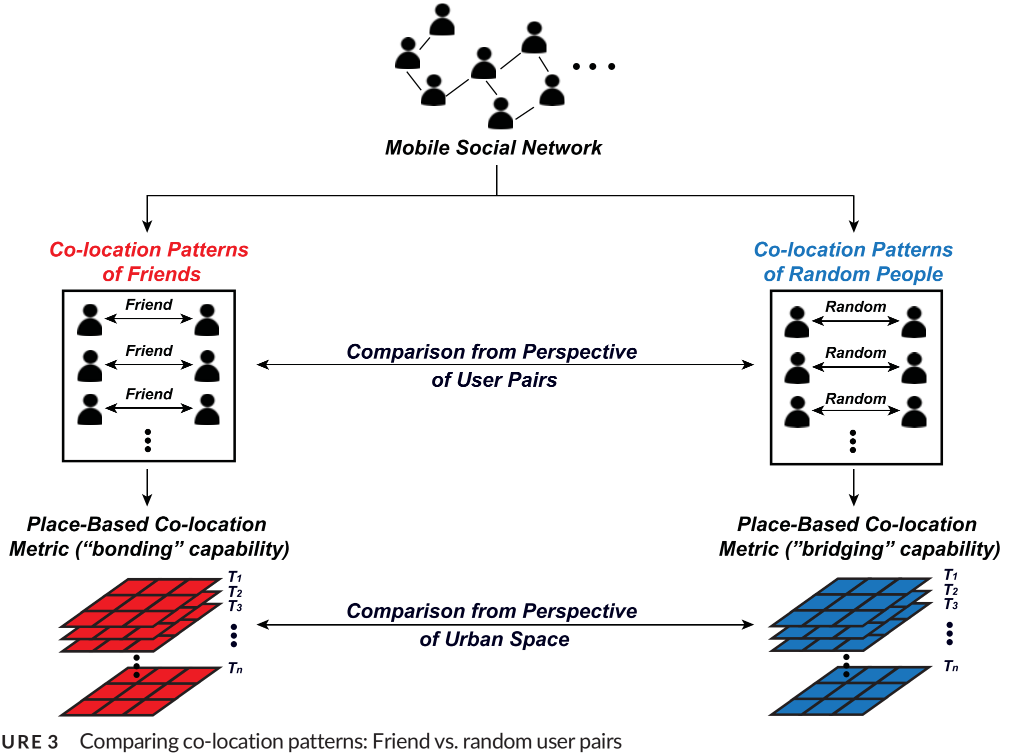

FIGURE 3 Comparing co-location patterns: Friend vs. random user pairs

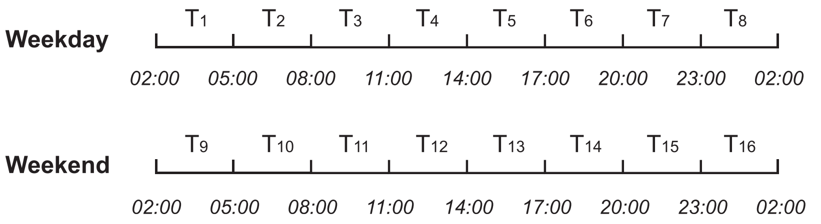

FIGURE 4 Temporal granularity for the co-location analysis

Since people who live or work at the same place are more likely to produce a high co-location rate at these places, in this research, we only consider user pairs who do not share the same home or work location (measured at the grid cell level). In particular, we extract 7,211,717 friend pairs with distinct home and work locations from 9,071,808 social ties in G. For U_r, we randomly sample 10,000,000 user pairs from the dataset and evaluate their co-location patterns.

To reflect the temporal variations of phone users' co-location patterns, as shown in Figure 4, we separate weekdays and weekends, and each type of day is divided into eight 3 hr time windows. Thus, the spatial co-location vectors of each user pair are computed over the 16 time windows. For example, given a user pair x and y, $Col_{x,y}(L, T_4)$ reflects the probability that they tend to co-locate at L during 11:00–14:00 on weekdays. Note that for a user pair x and y, $Col_{x,y}(L, T)$ is an estimate of the generic weekdays or weekends. In some situations, a pair of phone users could have a high co-location rate at a place even if they did not always share this place on the exact same day. For example, given a time window T_4 (i.e., 11:00–14:00), if the phone records of a user x usually appeared at a location L on Mondays, Tuesdays and Fridays, while that of another user y often appeared at the same location on Tuesdays, Wednesdays and Thursdays, they still tend to have a high co-location rate at L. However, we believe that calculating the spatial co-location vectors on generic weekdays and weekends is reasonable since: (1) CDRs are sparse in time and people's calling patterns could vary significantly from each other; and (2) it represents an interaction potential and reflects the fact that social contacts have similar experiences, preferences, and place-based knowledge.

4 | RESULTS

4.1 | Co-location patterns from the perspective of user pairs: Friends vs. random people

Figure 5 illustrates the mean (and standard deviation) of the spatial co-location rate for two sets of user pairs (U_s and U_r). On weekdays, as shown in Figure 5a, the average co-location rate of friend pairs exhibits notable fluctuations and the curve peaks at time window 11:00–14:00. On weekends, however, the temporal variation is less obvious, suggesting that the potential of the city in bringing friends together remains relatively stable over time. For random user pairs, the two curves show similar temporal patterns, indicating that the city's potential in bridging chance encounters are roughly the same on weekdays and weekends (Figure 5b).

By comparing Figures 5a and b, we find that friends are more likely to share urban space than random user pairs. For each time window T, we perform a Welch's T-Test to evaluate whether the average co-location rate of friend pairs ($\overline{SCol_{x,y}(T)}$, where $(x, y) \in U_s$) is larger than that of random user pairs ($\overline{SCol_{x',y'}(T)}$, where $(x', y') \in U_r$). As shown in Table 1, the p-values for all the 16 time windows are below 0.001, suggesting a significantly higher co-location rate for friend pairs.

We also notice that the average co-location rate of friend pairs on weekdays is generally higher than that of the corresponding time window on weekends (Figure 5a). By further computing the average co-location rate, for each friend pair $(x, y) \in U_s$, on weekdays as well as on weekends:

$$SCol_{x,y}(weekdays) = \frac{1}{8}\sum_{i=1}^{8} SCol_{x,y}(T_i) \tag{7}$$

$$SCol_{x,y}(weekends) = \frac{1}{8}\sum_{i=9}^{16} SCol_{x,y}(T_i) \tag{8}$$

we find that:

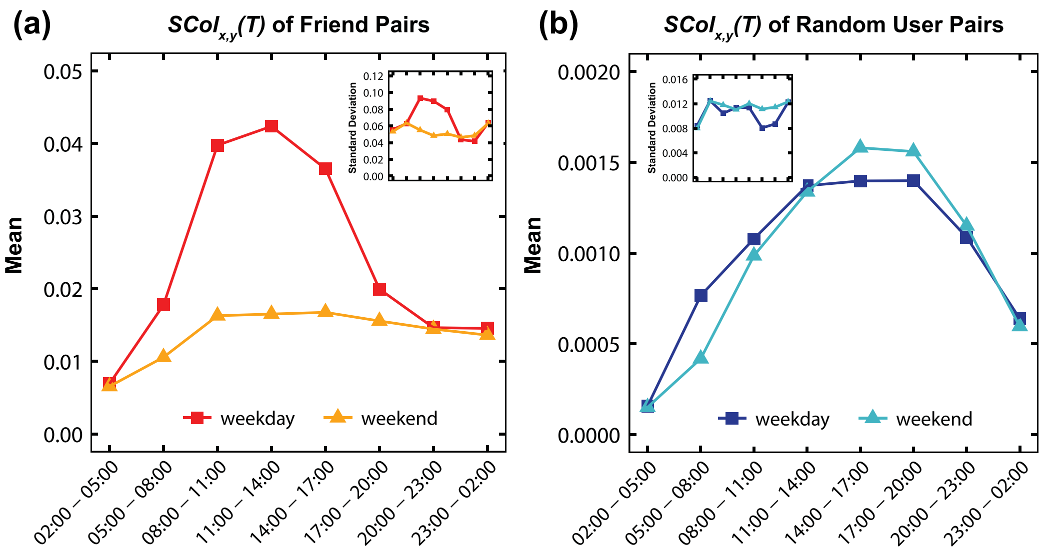

FIGURE 5 Spatial co-location rate of user pairs: (a) mean and standard deviation for friend pairs; and (b) mean and standard deviation for random user pairs

TABLE 1 Results of Welch's t-test on the average co-location rates of friend and random user pairs

Time Window	$\overline{SCol_{x,y}(T)}$	$\overline{SCol_{x',y'}(T)}$	t	df	P Value
T_1	692.2×10^{-5}	15.8×10^{-5}	323.5	7452200	<0.001
T_2	1777.1×10^{-5}	76.6×10^{-5}	717.5	7624400	<0.001
T_3	3975.3×10^{-5}	107.8×10^{-5}	1108.3	7341900	<0.001
T_4	4236.2×10^{-5}	137.1×10^{-5}	1220.8	7379100	<0.001
T_5	3650.9×10^{-5}	139.7×10^{-5}	1179.3	7425700	<0.001
T_6	1994.7×10^{-5}	139.8×10^{-5}	1129.0	7567700	<0.001
T_7	1464.1×10^{-5}	108.8×10^{-5}	860.8	7667700	<0.001
T_8	1454.6×10^{-5}	63.9×10^{-5}	578.2	7595100	<0.001
T_9	655.5×10^{-5}	15.1×10^{-5}	319.3	7448900	<0.001
T_{10}	1057.0×10^{-5}	41.8×10^{-5}	423.7	7606100	<0.001
T_{11}	1628.7×10^{-5}	98.5×10^{-5}	733.6	7703800	<0.001
T_{12}	1652.1×10^{-5}	133.7×10^{-5}	828.7	7746500	<0.001
T_{13}	1675.1×10^{-5}	158.0×10^{-5}	790.4	7797400	<0.001
T_{14}	1556.1×10^{-5}	155.9×10^{-5}	794.4	7817100	<0.001
T_{15}	1445.1×10^{-5}	115.0×10^{-5}	722.5	7808700	<0.001
T_{16}	1361.9×10^{-5}	59.6×10^{-5}	545.6	7618100	<0.001

Small p-values suggest that the difference is statistically significant.

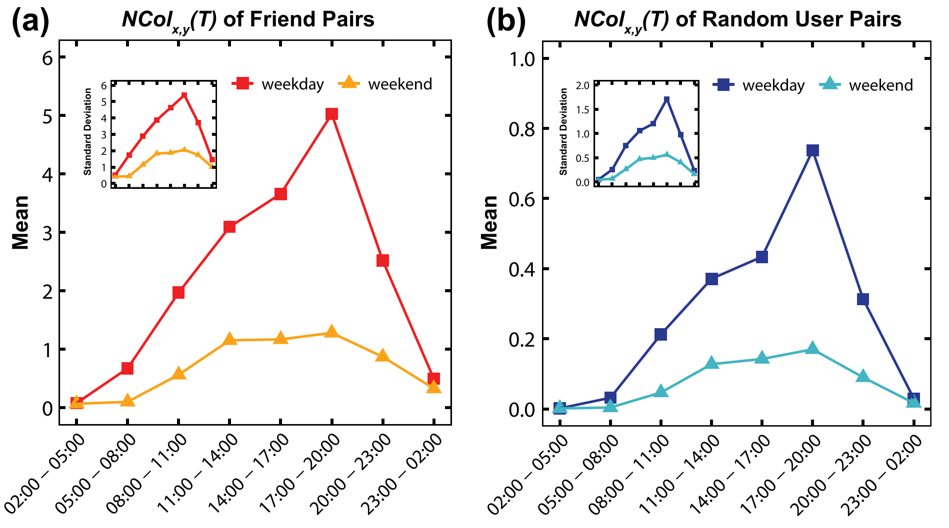

FIGURE 6 Number of shared locations of user pairs: (a) mean and standard deviation for friend pairs; and (b) mean and standard deviation for random user pairs

- 73.9% of the friend pairs have shared urban space on both weekdays and weekends, i.e., $SCol_{x,y}(weekdays)>0$ and $SCol_{x,y}(weekends)>0$.
- 19.6% of the friend pairs have shared urban space only on weekdays.
- 2.5% of the friend pairs have shared space only on weekends.
- 4.0% of the friend pairs have neither shared urban space on weekdays or weekends.

By further comparing $SCol_{x,y}(weekday)$ and $SCol_{x,y}(weekend)$ for the first category, we find that 55% (out of the 73.9% of the friend pairs sharing urban space on both weekdays and weekends) are more likely to share urban space on weekdays than on weekends, and the mean difference between $SCol_{x,y}(weekday)$ and $SCol_{x,y}(weekend)$ is 0.03. For the remaining 45%, the mean difference between $SCol_{x,y}(weekday)$ and $SCol_{x,y}(weekend)$ is -0.015. The findings enable us to better explain higher co-location rates for friends on weekdays (Figure 5a). First, the percentage of friend pairs who are more likely to share urban space on weekdays accounts for the majority of the population ($73.9\% * 55\% + 19.6\% = 60.25\%$). Furthermore, for those user pairs with a higher co-location rate on weekends, the absolute difference between weekdays and weekends is relatively small.

Figure 6 illustrates the mean (and standard deviation) of number of shared locations ($NCol_{x,y}$) for two sets of user pairs. For friend pairs, the mean $NCol_{x,y}$ reaches its peak at time window 17:00–20:00 on weekdays, suggesting a high diversity of co-location in space after normal work hours. On weekends, the mean $NCol_{x,y}$ is generally lower, with the values slightly above 1.0 during the peak periods (i.e., 11:00–14:00, 14:00–17:00 and 17:00–20:00). It appears that friends tend to share more unique locations on weekdays than on weekends. For each friend pair $(x, y) \in U_s$, we derive the total number of locations shared on weekdays and weekends as the union of the shared locations during the corresponding time windows:

$$NCol_{x,y}(weekdays) = |L(T_1) \cup L(T_2) \ldots \cup L(T_8)| \tag{9}$$

$$NCol_{x,y}(weekends) = |L(T_9) \cup L(T_{10}) \ldots \cup L(T_{16})| \tag{10}$$

Here $L(T_i)$ denotes the set of unique locations shared by (x, y) during time window T_i. Figure 7 shows the histogram of the difference between $NCol_{x,y}(weekdays)$ and $NCol_{x,y}(weekends)$ for friend pairs. We find that 81.95% of the friend pairs shared more unique locations on weekdays than on weekends. The mean and median difference (for all friend pairs) are 6.8 and 5.0, respectively.

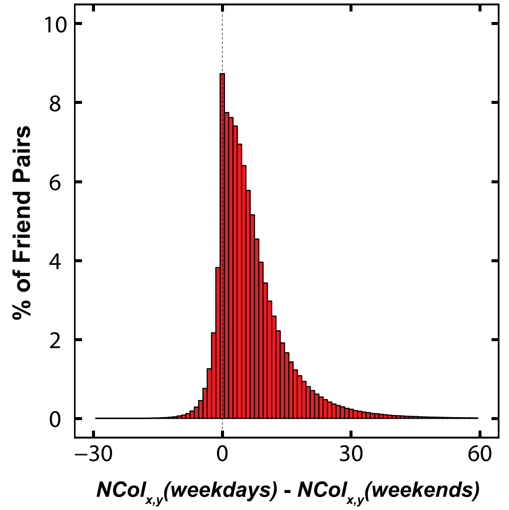

FIGURE 7 Histogram of the difference between the number of unique locations shared by friend pairs on weekdays and weekends

As compared with friends, random user pairs on average shared less unique locations (see Figure 6b). For each time window, we perform the same Welch's T-test to evaluate whether the average number of locations shared by friend pairs ($\overline{NCol_{x,y}(T)}$) is greater than that of random user pairs ($\overline{NCol_{x',y'}(T)}$). As shown in Table 2, the p-values of

TABLE 2 Results of Welch's t-test on the average number of locations shared between friend and random user pairs

Time Window	$\overline{NCol_{x,y}(T)}$	$\overline{NCol_{x',y'}(T)}$	t	df	P Value
T_1	80.5×10^{-3}	1.8×10^{-3}	400.5	7310300	<0.001
T_2	675.6×10^{-3}	31.8×10^{-3}	984.8	7441300	<0.001
T_3	1972.0×10^{-3}	212.7×10^{-3}	1601.6	7929000	<0.001
T_4	3095.8×10^{-3}	371.3×10^{-3}	1840.0	7989800	<0.001
T_5	3660.7×10^{-3}	434.0×10^{-3}	1826.4	7918000	<0.001
T_6	5028.0×10^{-3}	737.3×10^{-3}	2061.4	8268500	<0.001
T_7	2519.3×10^{-3}	313.0×10^{-3}	1556.5	7939000	<0.001
T_8	498.8×10^{-3}	23.0×10^{-3}	849.0	7493000	<0.001
T_9	69.5×10^{-3}	1.6×10^{-3}	434.2	7337000	<0.001
T_{10}	102.1×10^{-3}	4.1×10^{-3}	586.9	7477400	<0.001
T_{11}	566.3×10^{-3}	46.9×10^{-3}	1186.8	7769700	<0.001
T_{12}	1155.4×10^{-3}	127.8×10^{-3}	1472.0	7910100	<0.001
T_{13}	1171.1×10^{-3}	142.6×10^{-3}	1435.8	7951600	<0.001
T_{14}	1282.2×10^{-3}	170.1×10^{-3}	1418.0	8000400	<0.001
T_{15}	872.0×10^{-3}	90.0×10^{-3}	1180.5	7778700	<0.001
T_{16}	330.0×10^{-3}	16.9×10^{-3}	822.3	7483700	<0.001

Small p-values suggest that the difference is statistically significant.

the 16 time windows are all below 0.001, suggesting a significant difference between the number of locations shared by the two sets of user pairs.

4.2 | Co-location patterns from the perspective of urban space: Bonding vs. bridging

Two place-based co-location metrics (bridging vs. bonding) are derived to better understand how different places are shared by the two sets of user pairs. As the overall intensities of the two metrics are different, to make them comparable, for each time window T, we first calculate the relative bonding capability of each grid cell L as the average co-location rate of friends at this place, normalized by the maximum value (denoted as max$PCol_{U_s}$) across all places during the sixteen time windows:

$$NorPCol_{U_s}(L,T)=\frac{PCol_{U_s}(L,T)}{\max PCol_{U_s}} \tag{11}$$

Similarly, the relative bridging capability of a grid cell L is calculated as:

$$NorPCol_{U_r}(L,T)=\frac{PCol_{U_r}(L,T)}{\max PCol_{U_r}} \tag{12}$$

Figure 8 shows the relative bonding and bridging capability of the grid cells during selected time windows. For time window T_4 (i.e., 11:00–14:00 on weekdays), grid cells with a high level of bridging capability (see Figure 8e) mainly concentrate in an area in the southern part of Singapore, which refers to the downtown core. The downtown core serves as the financial hub of Singapore where numerous corporations and government agencies are located. A high population density in this area during the day time on weekdays creates massive opportunities for random people to co-locate with each other. The co-location patterns of friends, however, are very different from the aforementioned patterns for the same time window (Figure 8a). There are many places other than the downtown core that have a relatively high level of bonding capability, which indicates that these places, although not as "crowded" as downtown Singapore, seem to be important to friends and their social interactions. For example, the relative bonding capability at some places such as Nanyang Technological University (NTU) and the National University of Singapore (NUS) are relatively high (Figure 8a).

We also notice that places with a high level of bonding capability tend to spread out across different planning areas in Singapore. However, these grid cells become less attractive to friends during time window T_6 (i.e., 17:00–20:00 on weekdays, see Figure 8b), indicating that some places can be important to friends only during the day time. On weekends, the bonding capability of the places on 11:00–14:00 (Figure 8c) and 17:00–20:00 (Figure 8d) are different from that of the same time window on weekdays, suggesting that the role of a place in bringing friends together could vary between weekdays and weekends. For the four maps demonstrated for weekends, the place with the highest bonding or bridging capability corresponds to the same grid cell. The grid cell intersects with part of the Orchard road, which is the retail and entertainment hub of Singapore.

We notice that during 11:00–14:00 on weekends (Figures 8c, g), grid cells that are more attractive to friends tend to have a higher level of bridging capability. However, during the same time window on weekdays (Figures 8a, e), the spatial patterns of the two metrics appear to be less similar.

To assess the similarities of the two spatial patterns quantitatively, for each time window T, we compute the Spearman's rank correlation between $NorPCol_{U_s}(L,T)$ and $NorPCol_{U_r}(L,T)$. As we are only interested in places that are meaningful to friends, those grid cells with a very small value of bonding capability throughout the study days are not considered. Specifically, we calculate the mean value of $NorPCol_{U_s}(L,T)$ for each grid cell across the 16 time windows, and grid cells with a mean relative bonding capability equal or greater than 0.01 (369 in total) are included in the correlation analysis (Figure 9a).

As illustrated in Figure 9b, the correlation between the bonding and bridging capabilities of the places is higher on weekends than on weekdays. On weekdays, the correlation is higher in early morning and during the night, while the coefficient is below 0.6 during normal working hours (i.e., 08:00–17:00). On weekends, the correlation remains relatively stable and reaches the peak at time window T_6 (i.e., 17:00–20:00). The result indicates that, compared with weekdays, the bridging capability of the places on weekends, especially during the daytime, is more indicative of their bonding capabilities. One potential explanation is that on weekends, people's social interactions are strongly tied to

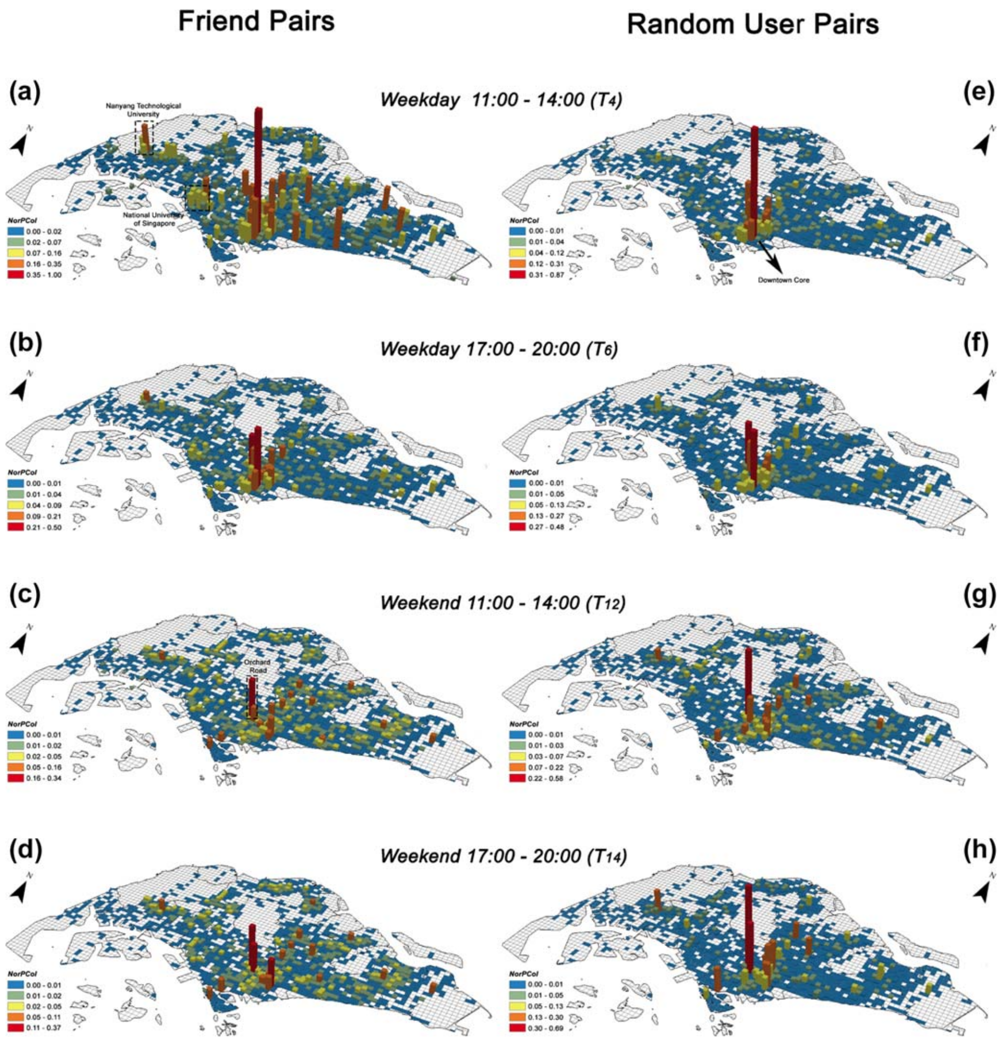

FIGURE 8 Place based co-location metrics during selected time windows: (a–d) relative bonding capabilities of the places; and (e–h) relative bridging capabilities of the places. The height of each bar illustrates the value of the corresponding grid cell. Five colors are used in each map, based on the Jenks Natural Breaks Algorithm, see De Smith, Goodchild, and Longley (2007) to distinguish grid cells with different values

recreational activities (e.g., shopping). Places which facilitate such purposes tend to attract many pairs or groups of friends. Hence, places which host more people are likely to attract more friends at the same time.

4.3 | Temporal signature of the places' bonding capabilities

For each grid cell L, there is a time series which illustrates how the bonding capability of this place changes over time. To better understand the temporal characteristics of a place's bonding capability, we first represent each grid cell L by this time series in a form of vector $V_L(q_1, q_2, \ldots, q_{16})$ normalized by the maximum value over the sixteen time windows:

$$q_i = \frac{PCol_{U_s}(L, T_i)}{\max_{1 \le i \le 16} PCol_{U_s}(L, T_i)} \tag{13}$$

We then apply an agglomerative hierarchical clustering algorithm with Ward's linkage to group the grid cells into clusters using Euclidian distance between vectors (Han, Pei, & Kamber, 2011; Yuan & Raubal, 2014). First, we perform the clustering into 20 clusters several times. At the end of each run, the clusters with a very small size (i.e., less than

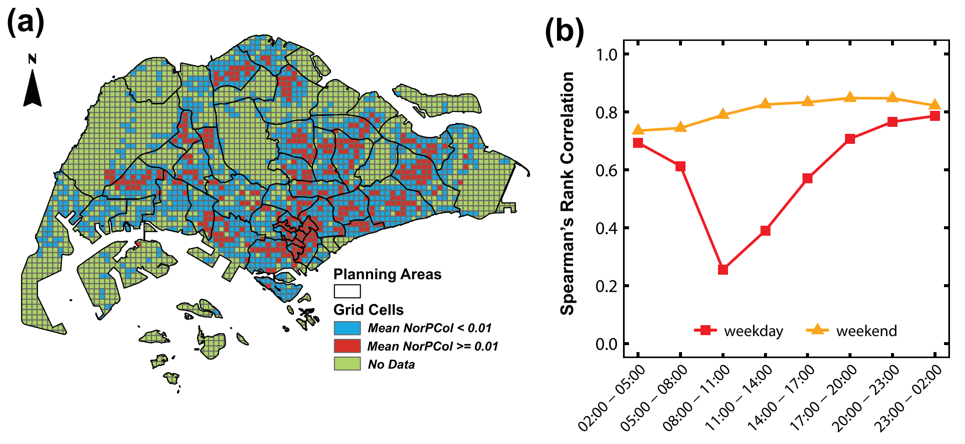

FIGURE 9 (a) Grid cells that are used in the correlation analysis (i.e., mean relative bonding capability equal or greater than 0.01); and (b) Spearman's rank correlation between $NorPCol_{U_s}(L, T)$ and $NorPCol_{U_r}(L, T)$ during each time window

1% of the total number of grid cells) are removed as "outliers". The purpose of this multi-level hierarchical clustering is to remove small clusters with relatively unique patterns. The clustering process terminates when there is no cluster with size smaller than the defined threshold (i.e., 1%).

By performing the multi-level hierarchical clustering over 1,423 grid cells, we group 1,386 of them into six clusters. The remaining 37 grid cells are identified as outliers in the clustering process. Figure 10 shows the mean center (i.e., average values of V_L of the corresponding grid cells) of each cluster. According to the results, *C*1 and *C*2 have much higher bonding capabilities on weekdays, and they cover 26.0% and 13.7% of the grid cells, respectively. The two bell

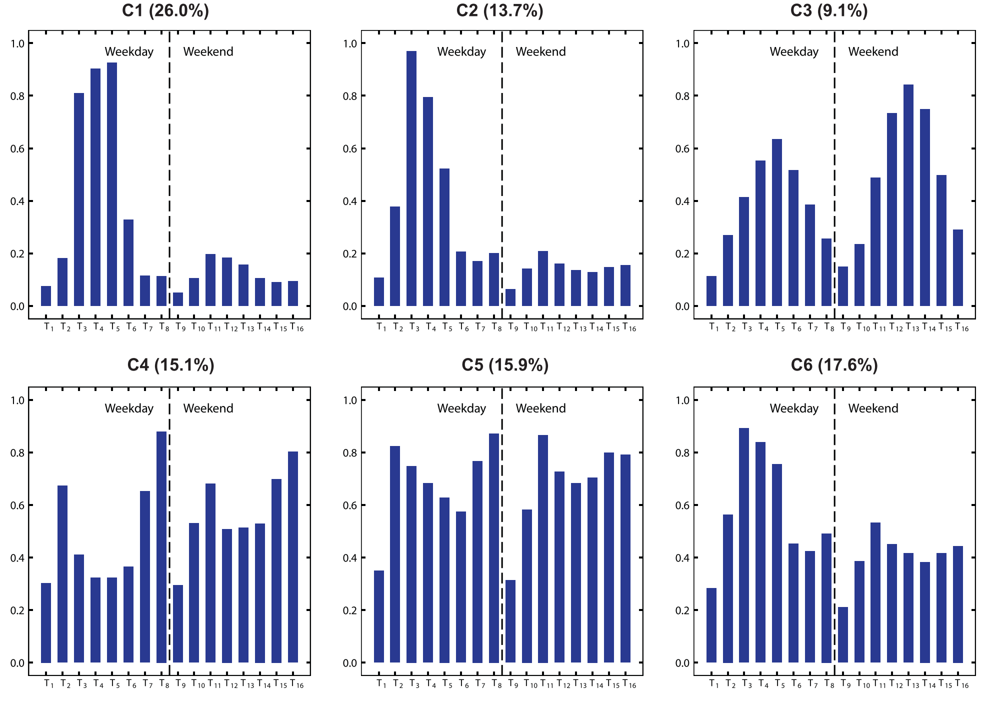

FIGURE 10 Temporal signature of the places' bonding capabilities. The x-axis of each graph denotes the time windows, and the y-axis denotes the average value of V_L (normalized bonding capability) of the corresponding cluster

curves in *C3* indicate that grid cells in this cluster are more attractive to friends around noon and afternoon, and the bonding capabilities of these places are higher on weekends than on weekdays. Clusters *C4* and *C5* show some interesting temporal patterns. The grid cells in these two clusters, especially on weekdays, have higher potentials of connecting friends before 08:00 and after 20:00. Cluster *C6* covers 17.6% of the grid cells, of which the bonding capabilities are higher on weekdays, and exhibit patterns similar to *C2*. However, the overall difference between weekdays and weekends is smaller as compared to *C1* and *C2*.

4.4 | Relationships between bonding capability and semantics of the places

The clustering analysis and the spatial patterns illustrated in Figure 8 have shown that: (1) the bonding capability of places could vary significantly from each other; and (2) the ability of a place to connect friends also changes over time. In this section, we further associate these characteristics with the semantics of the places, which allow us to have a better idea of the nature of places where friends are brought together.

To achieve this, we introduce five types of POIs,[3] which are relevant to people's daily social interactions, and analyze how they are distributed across the grid cells with different bonding capabilities and temporal signatures:

- Public commercial buildings (2,479 in total)
- Education institutes (564 in total)
- Shopping malls (283 in total)
- Sports centers (247 in total)
- Community centers & clubs (162 in total).

First, we examine the associations between place semantics and the temporal signature of the grid cells. Unlike McKenzie, Janowicz, Gao, and Gong (2015) who investigated what temporal signatures some specific POIs exhibit, we try to determine whether certain types of POIs are tied to specific temporal signatures. Figure 11 illustrates the

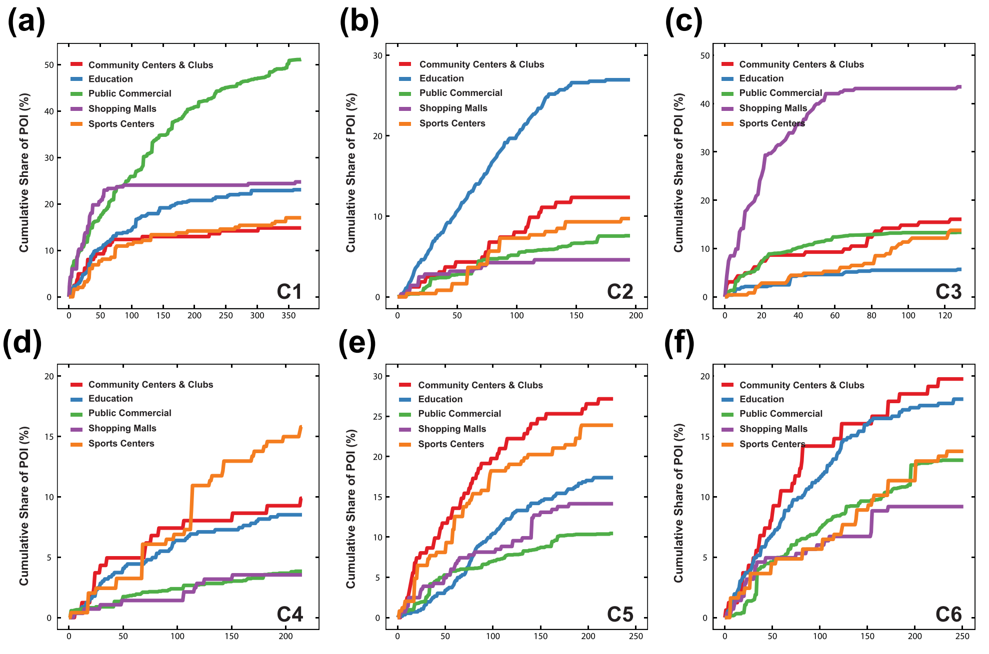

FIGURE 11 Cumulative share of POI by each cluster. The x-axis of each graph illustrates the cumulative number of grid cells in the corresponding cluster (sorted by mean bonding capability in descending order). The y-axis shows the cumulative percentage of POI coverage by type

percentages of POI covered by the grid cells in the six clusters. The x-axis denotes the number of grid cells in a cluster sorted by the mean bonding capability over the 16 time windows (in descending order). The y-axis shows the cumulative percentage of POI coverage by category. As shown in Figure 11a, grid cells in *C*1 account for over 50% of the public commercial buildings in Singapore. Moreover, as the number of grid cells reaches 100, the other four types of POI start to increase very slowly, which indicates that *C*1 is strongly associated with public commercial buildings.

Furthermore, public commercial buildings are mainly tied to work-related activities (e.g., business meetings), which potentially explains the high bonding capabilities of the grid cells during work hours on weekdays (Figure 10a). Since the friendship identified in this research encompasses multiple types of interpersonal relationships (e.g., business partners, family members, and other acquaintances), it is likely that during work hours on weekdays, places in *C*1 are shared mainly by user pairs who maintain some kinds of work relationships, although their estimated work locations are somewhere else. Cluster *C*2 includes only 14.1% of the grid cells (Figure 10b). However, it covers 27% of the education institutes in Singapore (Figure 11b). We notice that for grid cells in *C*2 (Figure 10b), their bonding capabilities reach the peak during 08:00 – 11:00, while the places in *C*1 are attractive to friends during the entire daytime (Figure 10a). Such differences between the two clusters are reflected by their POI coverages.

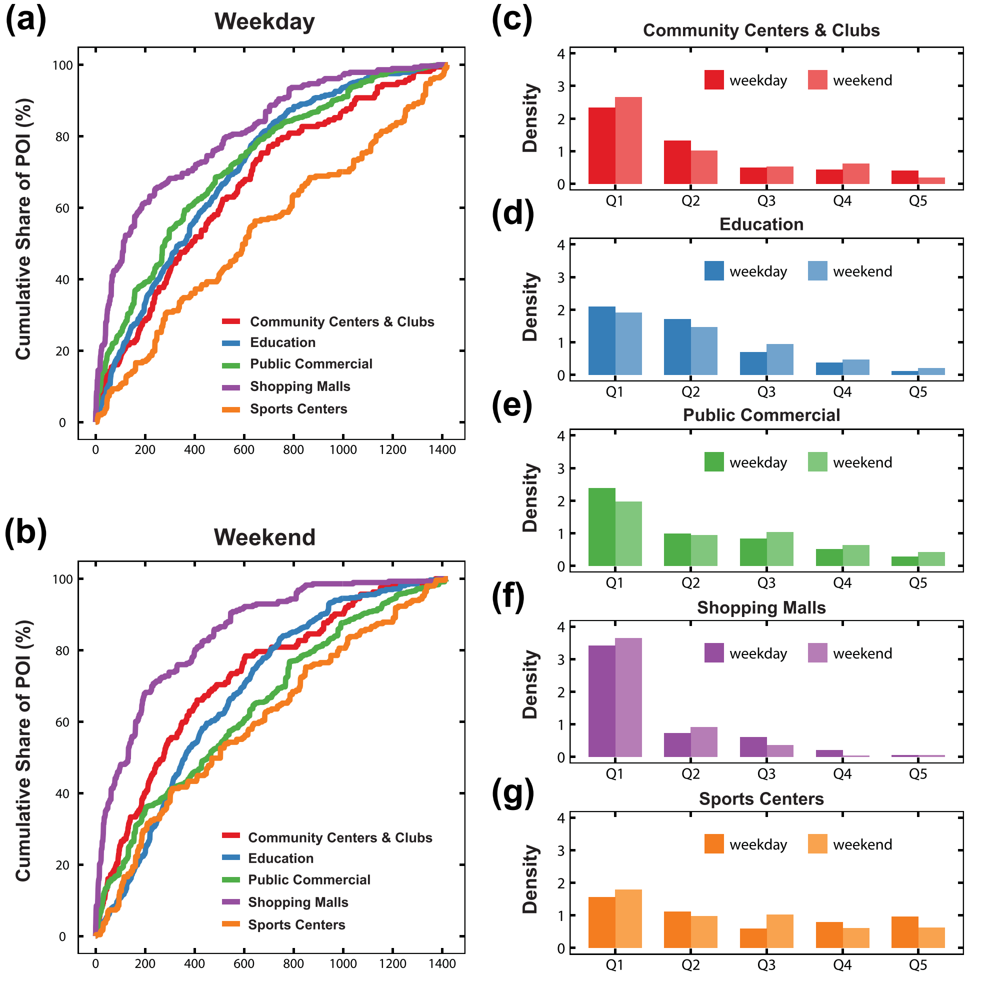

FIGURE 12 (a–b) Cumulative coverage of POIs by grid cells (sorted by average bonding capability in descending order) on weekdays and weekends; and (c–g) Q1–Q5 are groups of the grid-cells divided by 5 quantiles, i.e., Q1 includes 20% of the most attractive grid-cells, Q2 – next 20% of grid-cells and so on. Heights of the bars represent average relative density of POIs per grid-cell, i.e., average POI density in a group divided by overall average density of the corresponding type of POI

Cluster $C3$ includes 9.1% of the grid cells in the study area (Figure 10c). However, this cluster covers 44% of the shopping malls in Singapore (Figure 11c). It is very likely that shopping malls play an important role in connecting friends on both weekdays and weekends. Although we find that friends on average are more likely to share urban space on weekdays (Figure 5a), the grid cells in $C3$ are more attractive to friends on weekends, where human social interactions are more related to recreational activities.

Grid cells in $C4$ and $C5$ have higher bonding capabilities before 08:00 and after 20:00 (Figures 10d, e). We find that these two clusters cover a larger percentage of sports centers and community centers & clubs than other POI types (Figures 11d, e). It is apparent that friends are more likely to share these places during non-work hours on weekdays. However, even on weekends, the bonding capabilities of these two clusters are higher in early morning and during the night. One potential reason is that in Singapore, sports centers and community centers & clubs are associated with many outdoor facilities (e.g., swimming pools, basketball courts, tennis courts). These facilities are more attractive to friends and the general public when weather conditions (e.g., less sunshine) are better for physical and social related activities. For grid cells in $C6$, the top two POI types by coverage are community centers & clubs and education institutes (Figure 11f). We believe that this cluster corresponds to the places bonding people who do not work, mainly older (meeting at community centers & clubs) and younger (meeting at educational institutes).

To find out which places are the most attractive for friends in Singapore, we explore the overall POI coverage in a similar way separately for weekdays and weekends. Specifically, we compute the average bonding capability of grid cells on weekdays as well as on weekends, and sort them in descending orders. By exploring the cumulative percentage of POIs covered by these grid cells, we find that the overall pattern is rather similar on weekdays (Figure 12a) and weekends (Figure 12b). In general, areas that are more attractive to friends tend to cover a larger percentage of POIs. However, certain POI types seem to have a higher impact on bonding capability than others. In Figures 12c–g we further compare weekdays and weekends for each type of POI separately. In particular, we divide the grid cells into five quantiles (Q1 – Q5) and for each quantile, we compute the relative density of POIs per grid cell (i.e., average POI density in the group divided by overall average density of the corresponding type of POI). As one would expect, areas of Education Institutes (Figure 12d) and Public Commercial Buildings (Figure 12e) are more attractive on weekdays than on weekends (for the first two quantiles), compared to three other types of POIs that are more popular on weekends. We also notice that Sports Centers are almost equally distributed across all five quantiles, whereas the density of Shopping Malls in the first 20% of grid-cells is more than three times higher than average.

5 | CONCLUSIONS

Urban spaces, manifested by the streets, buildings, and various types of infrastructure, play an important role in connecting people and strengthening their social relations. For many years, urban planners have been pursuing better ways of designing cities to facilitate human social interactions. Yet such goals cannot be easily achieved without a prior knowledge of how people share urban space with each other. This research proposes a framework to gain insights into the spatiotemporal characteristics of friends' use of urban space. By applying the framework to a 50-day CDR dataset in Singapore, we find that friends are more likely to share urban space than random user pairs, and they also share more locations. On weekdays, the average co-location rate of friend pairs exhibits notable fluctuations, whereas on weekends, the temporal variation is less obvious, suggesting that the potential of the city in connecting friends remains relatively stable. By deriving the two place-based co-location metrics (bonding and bridging), we find that a place could play different roles in connecting friends vs. random people (e.g., chance encounters). On weekdays, the bonding and bridging capabilities of the places show different spatial patterns. However, the two indicators become highly correlated on weekends, indicating that places which host more people tend to attract more friends at the same time. The comparisons reveal two interesting aspects of social dynamics in the urban environment. On weekdays, people's social interactions do not always arise at places where crowds congregate (e.g., metro stations). On weekends, the places' abilities to connect friends and chance encounters become more compatible, suggesting similar destination choices among the social contacts.

A hierarchical clustering algorithm is then applied to examine how the bonding capabilities of places change over time. We derive several clusters with distinctive temporal patterns and associate them with five types of POIs. The result suggests that the temporal signature of bonding capability is strongly related to the semantics of the places. Places which are dominated by: (1) shopping malls, (2) education institutes, (3) public commercial buildings, (4) community centers & clubs, (5) sports centers show distinctive temporal patterns. Finally, we investigate which places are more attractive to friends. The result suggests that, in general, the density of POIs increases as places become more attractive. However, by further breaking down the POIs, we find that certain POI types (e.g., shopping malls) tend to have a much higher impact on bonding capability than others (e.g., sports centers).

A cohesive city is built upon healthy interpersonal relationships. Such relationships can be better nurtured within a vibrant and livable urban environment. We believe our research could benefit urban planning and social cohesion in several ways. First, the metrics derived in this study enable urban planners to better evaluate the social roles of various places in a city, which leads to better adjustment and improvement of urban infrastructures. Second, the association between bonding capability and POIs could help decision-makers anticipate the social effect of a new facility (e.g., community center) being built. Moreover, the social roles of places and their diurnal patterns could help merchants come up with business strategies (e.g., dining promotions for friends, taxi sharing services) which not only boost their profits, but also benefit the general public.

This research has several limitations. First, CDR data are passively generated during certain types of cell phone communications. These data capture partial aspects of human activity patterns, which lead to an underestimate of the co-location rate of user pairs. Second, the grid-based method does not consider the neighborhood effects, which means that friend pairs who appear at nearby (adjacent) grid cells are not taken into account. Third, the social network extracted in this study does not include the dyads who did not call each other during the study period, or who were subscribers of different cellular operators. A more comprehensive view of the social roles of urban spaces and places can be obtained by applying the framework to other mobile phone datasets or location-based social networks (LBSNs). Also, how the results are affected by the defined spatiotemporal resolutions is worth further investigation. Nevertheless, we believe this analytical framework is a new way of integrating human mobility and social network analysis in GIScience, and some of the current restrictions will be removed as we move forward and collect more detailed data about human behavior.

ACKNOWLEDGMENTS

The authors thank the Amsterdam Institute for Advanced Metropolitan Solutions (AMS), Allianz, Ericsson, Liberty Mutual Research Institute, Philips, the Kuwait-MIT Center for Natural Resources and the Environment, Singapore-MIT Alliance for Research and Technology (SMART), the Société nationale des chemins de fer français (SNCF), UBER, Volkswagen Electronics Research Laboratory, and all the members of the MIT Senseable City Lab Consortium for supporting this research.

NOTES

[1] By the time the data were collected, Singapore had about 7.38 million mobile subscribers, representing a penetration rate of 145.5%.

[2] Planning areas, also known as DGPs, are the primary census divisions of Singapore created by the Urban Redevelopment Authority (URA). There are a total of 55 planning areas in Singapore with each served by a town center and several neighborhood commercial/shopping centers (https://www.ura.gov.sg/uol/master-plan/Contacts/View-Planning-Boundaries).

[3] The POI dataset was acquired from the Singapore Land Authority (SLA). Here we only chose five major types of POIs that are most relevant to human social interactions. Other types, such as Auto Bank (ATM), Fire Station, Hotel, and Medical & Health, are not included.

REFERENCES

Ahas, R., Silm, S., Järv, O., Saluveer, E., & Tiru, M. (2010). Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology*, *17*(1), 3–27.

Albertson, L. A. (1977). Telecommunications as a travel substitute: Some psychological, organizational, and social aspects. *Journal of Communication*, *27*(2), 32–43.

Andris, C. (2016). Integrating social network data into GISystems. *International Journal of Geographical Information Science*, *30*(10), 2009–2031.

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: Improving geographical prediction with social and spatial proximity. *Proceedings of the 19th International Conference on World Wide Web* (pp. 61–70). Raleigh, NC: ACM.

Barabási, A.-L. (2010). *Bursts: The hidden patterns behind everything we do, from your e-mail to bloody crusades*. New York, NY: Penguin.

Bojic, I., Massaro, E., Belyi, A., Sobolevsky, S., & Ratti, C. (2015). Choosing the right home location definition method for the given dataset. In T.-Y. Liu, C. N. Scollon, & W. Zhu (Eds.), *Social Informatics: 7th International Conference, SocInfo 2015, Beijing, China, December 9–12, 2015, Proceedings* (pp. 194–208). Cham, Switzerland: Springer.

Cairncross, F. (2001). *The death of distance: How the communications revolution is changing our lives*. Cambridge, MA: Harvard Business Press.

Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLoS One*, *6*(7), e20814.

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G., & Barabási, A.-L. (2008). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, *41*(22), 224015.

Carrasco, J., Miller, E., & Wellman, B. (2008). How far and with whom do people socialize?: Empirical evidence about distance between social network members. *Transportation Research Record*, *2076*, 114–122.

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: User movement in location-based social networks. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1082–1090), San Diego, CA: ACM.

Choo, S., Lee, T., & Mokhtarian, P. (2007). Do transportation and communications tend to be substitutes, complements, or neither?: US Consumer expenditures perspective, 1984–2002. *Transportation Research Record*, *2010*, 121–132.

Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., & Kleinberg, J. (2010). Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences, USA*, *107*(52), 22436–22441.

Cranshaw, J., Toch, E., Hong, J., Kittur, A., & Sadeh, N. (2010). Bridging the gap between physical location and online social networks. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing* (pp. 119–128). Copenhagen, Denmark: ACM.

De Smith, M. J., Goodchild, M. F., & Longley, P. (2007). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools*. Leicester, UK: Troubador Publishing Ltd.

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences, USA*, *106*(36), 15274–15278.

Goldenberg, J., & Levy, M. (2009). Distance is not dead: Social interaction and geographical distance in the internet era. arXiv preprint arXiv:0906.3202.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453* (7196), 779–782.

Graham, S. (1998). The end of geography or the explosion of place? Conceptualizing space, place and information technology. *Progress in Human Geography*, *22*(2), 165–185.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Oxford, UK: Elsevier.

Holland, C., Clark, A., Katz, J., & Peace, S. (2007). *Social interactions in urban public places*. Bristol, UK: Policy Press.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying important places in people's lives from cellular network data. In K. Lyons, J. Hightower, & E. M. Huang (Eds.), *Lecture Notes In Computer Science. Proceedings of the 9th International Conference on Pervasive Computing* (pp. 133–151). San Francisco, CA.

Kang, C., Zhang, Y., Ma, X., & Liu, Y. (2013). Inferring properties and revealing geographical impacts of intercity mobile communication network of China using a subnet data set. *International Journal of Geographical Information Science*, *27* (3), 431–448.

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban gravity: A model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory & Experiment*, *2009*(7), L07003.

Lambiotte, R., Blondel, V. D., De Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., & Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and Its Applications*, *387*(21), 5317–5325.

Larsen, J., Axhausen, K. W., & Urry, J. (2006). Geographies of social networks: Meetings, travel and communications. *Mobilities*, *1*(2), 261–283.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences, USA*, *102*(33), 11623–11628.

Madanipour, A. (1999). Why are the design and development of public spaces significant for cities?. *Environment & Planning B*, *26*(6), 879–891.

McKenzie, G., Janowicz, K., Gao, S., & Gong, L. (2015). How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment and Urban Systems*, *54*, 336–346.

Mok, D., Wellman, B., & Carrasco, J. (2010). Does distance matter in the age of the internet?. *Urban Studies*, *47*(13), 2747–2783.

Putnam, R. D., & Goss, K. A. (2002). Introduction. In R. D. Putnam (Ed.), *Democracies in flux: The evolution of social capital in contemporary society*. New York, NY: Oxford University Press.

Shi, L., Wu, L., Chi, G., & Liu, Y. (2016). Geographical impacts on social networks from perspectives of space and place: An empirical study using mobile phone data. *Journal of Geographical Systems*, *18*(4), 359–376.

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021.

Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., & González, M. C. (2015). Coupling human mobility and social ties. *Journal of the Royal Society Interface*, *12*(105), 20141128.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabási, A.-L. (2011). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1100–1108). San Diego, CA: ACM.

Wang, Y., Kang, C., Bettencourt, L. M., Liu, Y., & Andris, C. (2015). Linked activity spaces: Embedding social networks in urban space. In M. Helbich (Ed.), *Computational approaches for urban environments* (pp. 313–336). Berlin, Germany: Springer.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation*, *42*(4), 625–646.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Lu, F., Chen, J., . . . Li, Q. (2016). Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, *106*(2), 489–502.

Yuan, Y., & Raubal, M. (2012). Extracting dynamic urban mobility patterns from mobile phone data. In N. Xiao, M.-P. Kwan, M. F. Goodchild, & S. Shekhar (Eds.), *Geographic Information Science: 7th International Conference, GIScience 2012, Columbus, OH, USA, September 18–21, 2012. Proceedings* (pp. 354–367). Berlin, Germany: Springer. Lecture Notes in Computer Science Vol. 7478.

Yuan, Y., & Raubal, M. (2014). Measuring similarity of mobile phone user trajectories: A spatio-temporal edit distance method. *International Journal of Geographical Information Science*, *28*(3), 496–520.

How to cite this article: Xu Y, Belyi A, Bojic I, Ratti C. How friends share urban space: An exploratory spatio-temporal analysis using mobile phone data. *Transactions in GIS*. 2017;21:468–487. https://doi.org/10.1111/tgis.12285