Mining Emerging User-Centered Network Structures in Location-based Social Networks

Konstantinos Pelechrinis
University of Pittsburgh
kpele@pitt.edu

Theodoros Lappas
Stevens Institute of Technology
tlappas@stevens.edu

***Abstract*—The digitization of social networks has enabled the passive collection of large scale data, which in turn have fostered social studies that have been traditionally dependent on small scale, interview-based data. During the last years, a new class of digital social networks has emerged, namely, location-based social networks (LBSNs). The main interaction between users of an LBSN is location sharing, i.e., declaring their presence to specific places. The latter ties the virtual, online world with the real space that users interact in. Thus, except from the social graph, a number of implicit network structures emerge. As an example, two people can be considered to be connected if they have been to at least k common places. Similar structures play crucial role in fields such as epidemiology and urban planning, while they can have implications in communication networks as well (e.g., mobile peer-to-peer content delivery). In this study, we examine the characteristics and the evolution of these structures using two LBSN datasets. As our analysis indicate, (i) these structures can deviate significantly from the pure social network and, (ii) they are highly dynamic (i.e., these implicit connections are ephemeral).**

I. Introduction

The proliferation of digital social networks has given rise to large-scale social studies. Traditional social network analysis has been based on small scale data obtained through interviews. This includes many challenges, such as the difference between the declared and actual social behavior of the subjects. On the contrary, online social networks provide an abundant source of information, which can be passively collected and captures the online behavior of people over a number of dimensions. Online trails of people's activities have enabled studies of network structures and network formation, information diffusion through a social network etc. However, this information, while massive, it can also include various sources of bias, since actions in online and real world can differ dramatically. For instance, Sproull and Kiesler [1] have found that a lot of the information conveyed through electronic communication would not have been diffused through another medium. As a result, online, virtual, interactions are not necessarily representative of real-world interactions between people. This can have implications in studies that involve data from "traditional" virtual social networks.

Nevertheless, during the last years, fostered by advancements in mobile handheld devices (e.g., smartphones), a new class of digital social networks, namely, geo-social or location-based social networks (LBSNs for short), has enjoyed rapid proliferation. The ability of these devices to estimate their position has enabled mobile systems to consider and integrate another dimension within the digital social networks, that of ***location***. LBSN users still interact online but this interaction captures their behavior in real space via location sharing.

Location is loosely delineated and there are different granularities for its definition. It can be a country, a city, a neighborhood or a street. In all of these, context information is absent. If we want to add semantic information, location can be a specific *place* such as a park, a monument, a mountain or even a restaurant or a university building. The last definition adds context to the location information and can enable a number of novel and innovative studies. This information can be either tracked continually by the system through a continuous spatial trajectory or it can be provided by the user in a voluntary and discrete fashion through *check-ins*.

Participants of these systems are not only *socially* connected, through the friendships declared between them. They can also be associated through their geographic location (e.g., co-locations), which can further record indirect bonding (e.g., similar interests or affiliations). In particular, LBSNs consist of two components; (i) the social and (ii) the spatial. The former is no different from any other digital social network and is represented through an affinity graph. By considering the spatial trails of LBSN users, a number of various network structures can emerge, each of which records different information. These implicit networks can capture ***associations*** between either places or people (or both). For instance, if Y number of users go to place w after visiting place z, a directed edge from z to w connects the two places in a "flow graph". Alternatively, people can be thought to be connected if they have been to at least k common places (in other words having k common check-in locations) obtaining a *k-check-in* graph. Furthermore, location socio-affiliation networks can be developed by considering the places as the affiliations.

In this paper we are interested in the network structures created among people, and in particular in the k-check-in graph. Using two LBSN datasets we examine its topology and its evolution through metrics such as degree distribution, transitivity and edge similarity. We further compare it with the social graph. Our main goal is to study the similarity of these structures and whether they can be used interchangeably when only one of them is available. We also want to examine/highlight what are the implications that can arise when using LBSN datasets to drive applications/studies that require (co-)location information. Our main findings indicate that:

- The k-check-in graph might not be a good *proxy* for the pure social graph (and vice versa), since these two structures can vary significantly and,
- The k-check-in graph exhibits *low temporal edge similarity*, since the undelying implicit connections appear to not be long-lasting.

978-1-4799-3088-3/14/$31.00 ©2014 IEEE

Scope of our work: Traditional online social networks have enabled large scale sociology studies, related to the structure and formation of social connections. However, these structures do not place networks in the real world context, since they mainly capture online virtual interactions. LBSNs can fill this gap due to their extension in the real space that people interact. In particular, the k-check-in graphs essentially record similarities in (spatial) interests and mobility and can significantly enhance sociology studies related to social ties formation (e.g., why, how and where do social ties generate? How strong are these ties?). Furthermore, we believe that these network structures will play a crucial role in other fields as well, such as urban planning and epidemiology, where people's location, rather than their actual social affinities, play a central role. Specific communication networks can also benefit; delay tolerant or mobile peer-to-peer content delivery networks can optimize their performance based on knowledge from the k-check-in graphs. Of course, when using similar dataset from digital social media, we always need to consider the various sources of bias the are included (e.g., demographic biases).

The rest of the paper is organized as follows. In Section II we describe the datasets we analyzed and we formally introduce the k-check-in structure. Section III presents the static analysis of the k-check-in graph and its comparison with the social graph. Section IV analyzes the temporal evolution of the k-check-in structure. Finally, Section V discusses related to our work literature, while Section VI concludes our study.

II. Analysis Setup

For our analysis we will use two geo-social network datasets collected by Cho *et al.* [2]. In particular:

Gowalla dataset: The dataset consists of 6,442,892 public check-in data performed by 196,591 Gowalla users in 1,280,969 distinct places, between February 2009 and October 2010. Every check-in log includes a tuple in the form `<User ID, Time, Latitude, Longitude, Venue ID>`. The users also participate in a friendship network with reciprocal relations, which consists of 950,327 links.

Brightkite dataset: The dataset consists of 4,491,143 public check-in data performed by 58,228 Brightkite users in 772,966 distinct places, during the period between April 2008 and October 2010. The check-in information is in exactly the same format as above. Brighkite users also participate in a friendship network, which consists of 214,078 links. The friendships in Brightkite are directed. However, the 214,078 links in the dataset correspond to reciprocal ties [2].

A. Notations and Definitions

A network G consists of a set of n entities (vertices/nodes) V connected through a set of m edges E. For the social network graph, V is the set of users, i.e., $u \in V$ represents a user, while E is the set of friendships, i.e., an edge $\{i, j\} \in E$ represents a friendship between users i and j. Every user $u \in V$ is further associated with a set ℓ_u, which contains the places that u has visited. We *define* the **k-check-in** graph as follows:

- $V = \{1, \ldots, n\}$ is the set of users (just as in the social network)
- $E = \{\{i, j\} : i \in V, j \in V, |\ell_i \bigcap \ell_j| > k\}$

In other words, the k-check-in graph represents a network among people who have been to more than k same places. Hence, people that are not friends can be connected in this structure due to *similar interests* in the places they visit. Conversely, friends might not be connected in the k-check-in graph, since they never visit same places. In what follows we will perform a two-mode analysis of the k-check-in network. We will begin by studying its static properties, that is, describing the graph structure we obtain by considering the network in aggregate through the period that our datasets cover (Section III). We will then continue, by analyzing its temporal properties. More specifically, we split our dataset in 4 temporal snapshots and analyze the k-check-in graph over these snapshots tracking their evolution (Section IV).

III. Static Structures

Our static comparative analysis includes metrics related to (i) the node degree distribution, (ii) the transitivity of the network and (iii) the edge similarity. We further consider k; $k \in \{0, 5, 10, 50, 100\}$.

A. Degree Distribution

We begin by examining the node degree distribution, $P(d)$, of the social and the k-check-in graphs. Figure 1 presents our results (in log-log scale). Starting from the social graph, we observe the typical heavy-tailed distribution exhibited by many social networks [3]. For small values of k, the k-check-in structure exhibits an even longer tail. Table I presents the maximum degree observed in the structures. The 0-check-in graph, i.e., an edge connects two users if they have been to at least one same venue, exhibits much higher maximum degree for both Gowalla and Brightkite. However, even a slight increase in the threshold k, reverts the situation. In particular, the more stringent constraint we place on the check-in association between two users, the less probable is to find highly "connected" individuals, that is, people that share many same places with many others.

However, note here that, for all values of k, the probability mass for small (non-zero) node degrees is significantly lower as compared to the social graph. This might seem counterintuitive, especially for $k = 0$ where we saw above that one can identify much more highly connected individuals. In order to understand this subtle point, recall that in LBSNs users volunteeringly share their locations. Hence, if they do not share there whereabouts (often) the mass probability for $d = 0$ (not shown in the log-log scale Figure 1) will be very large in the k-check-in graphs, reducing essentially the mass for $d > 0$. On the contrary, users that are more enganged in checking-in they will appear highly connected, possibly even more as compared to their social space for small values of k.

To quantitatively support the above observation, we have computed the fraction of Gowalla and Brightkite users in our datasets that have zero check-ins. This is 0.46 and 0.12 respectively. This fraction serves as a lower bound for $P(0)$ in the k-check-in graph - the probability that a node has no neighbor in the k-check-in network. This is true since users that do not share any of their locations, will not have any check-in association with other users. Table II presents the probability $P(0)$ for the different networks.

-	Gowalla	Brightkite
Social	14730	1134
0-checkin	23819	12705
5-checkin	4892	421
10-checkin	2095	114
50-checkin	144	8
100-checkin	31	2

TABLE I: Maximum node degree.

-	Gowalla	Brightkite
Social	0	0
0-checkin	0.47	0.24
5-checkin	0.77	0.92
10-checkin	0.88	0.97
50-checkin	0.99	0.99
100-checkin	0.99	0.99

TABLE II: Probability of singleton nodes.

-	Gowalla	Brightkite
Social	0.2367	0.1723
0-checkin	0.4563	0.3774
5-checkin	0.1312	0.0360
10-checkin	0.0614	0.0111
50-checkin	0.0036	$3.3\cdot 10^{-4}$
100-checkin	$7.4\cdot 10^{-4}$	$5.1\cdot 10^{-5}$

TABLE III: Mean local cc.

As we can see $P(0)$ is getting significantly higher values for the k-check-in networks as compared to the social graph. For very large values of k, $P(0)$ is almost equal to 1, meaning that the network is practically comprised of singletons. Again, more strigent constraints, will inevitably lead users that do not share their locations often, to have no neighbors in the k-check-in graph. For instance, when $k = 50$, users with less than 50 check-ins will have no neighbor in the 50-check-in graph.

Empirical fact 1: The above analysis clearly indicates that *the node degree distributions of the social graph and the k-check-in graphs are significantly different*. While part of this difference can be attributed to the diversity of users with regards to their engangement in the check-in process as aforementioned, it also captures to a large extent the different information encoded by the two networks. The k-check-in network captures the association/similarity between people with regards to their location and the strength of this affinity (different k), while the social graph captures the declared friendship between users. The latter might have been created due to similarities in different planes (e.g., work, family etc.) and hence the friendship affinity graph might not be useful for situations where location association is important. For instance, when studying epidemics the social graph cannot be used instead of a graph that captures co-locations. In other words, while the social graph can be thought as capturing the similarity between users at an aggregate level across a variety of dimensions, the k-check-in graphs are specifically tailored to spatial trails of people.

B. Edge Density

Next we examine the connectance or edge density ρ:

$$\rho = \frac{m}{\frac{n\cdot(n-1)}{2}} \tag{1}$$

The denominator is essentially the maximum number of possible edges for a graph with n vertices. Hence, ρ captures the fraction of these edges that actually exist in the network. While characterizing a network as dense or sparse formally requires identifying the value that ρ converges to at the limit of large n ($n \to \infty$), here we are simply interested in comparing the value of edge density among the different networks. Figure 2 presents our results. For easier visual obeservation of the differences we use log-scale. Note here that since $0 \leq \rho \leq 1$, the edge density will be negative in log-scale and "smaller" bars at the figure represent denser network.

For small values of k ($k \leq 5$), the k-check-in graph can be much more dense compared to the social graph. However, as we increase k the network becomes sparser. Nevertheless, this can be (partially) attributed again to the engagement of the users with location sharing. Hence, we slightly modify the definition of ρ in Equation 1 and instead of using the total number of vertices n for the k-check-in graphs, we use the *effective vertices* n^k_{eff}. The latter is the number of users that have performed at least k check-ins, and thus, it is possible for them to have connections in the k-check-in graph. Figure 3 depicts our results (in log-scale again). As we can observe, now even for moderately larger values of k, the k-check-in graph can be more dense compared to the social graph.

Empirical fact 2: *The k-check-in network can be orders of magnitude denser compared to the social graph for small and moderate values of k, while it transitions to sparser structures when k takes extremely high values.* This translates to a large number of user pairs exhibiting some level of similarity (small and moderate values for k) with regards to their location trails/preferences. However, not many of them lead to social connections as the much lower edge density of the social affinities reveals (e.g., some of the co-check-in edges might be *trivial* due to places people have to visit such as, airports, subway stations etc.). On the contrary, as shown from the low edge density values of the k-check-in graphs for large k, there are only a few user pairs among all possible ones that exhibit very high similarity. In what follows, we will delve into the details of this observation and further examine the edge overlap between the different networks.

C. Transitivity

One of the main characteristics of a large number of real-world networks is their transitivity as captured from the network's clustering coefficient (cc for short). The local clustering coefficient of node i (with $d_i > 1$) is defined as:

$$cc_i = 2 \cdot \frac{x}{d_i \cdot (d_i - 1)} \tag{2}$$

where x is the number of pairs of neighbors of i that are connected. Then the average clustering coefficient C_G of a graph G is the mean of cc_i over all the nodes of the network.

C_G measures how likely it is for vertices with a common neighbor to be neighbors themselves. Table III presents the results for the social and the k-check-in networks for both datasets. As we can notice, the average clustering coefficient of the k-check-in graph is significantly higher than that of the social graph for small values of k, while it drops drastically with an increase in k. However, this number alone is not completely informative. C_G should be compared to the probability that randomly selected pairs of vertices are neighbors. This probability is exactly ρ_G, the edge density of graph G. Hence, the network is said to be transitive if $C_G \gg \rho_G$, and the ratio $\frac{C_G}{\rho_G}$ can be thought as quantifying this transitivity[1]. For instance, the decrease of the average clustering coefficient for the k-check-in network as k increases, might be attributed

[1]Some additional details on cc and transitivity are included in the Appendix.

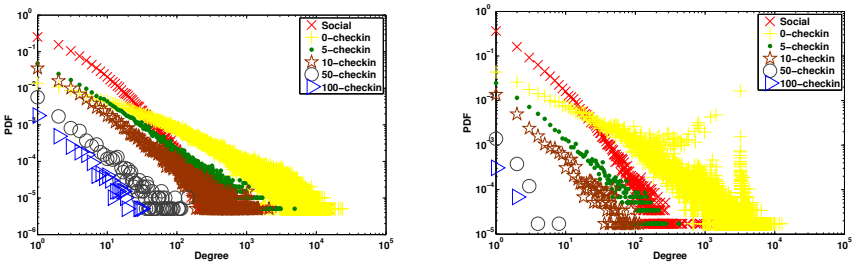

FIG. 1: Degree distributions for the social graph and the $k-$check-in friendship graph (Gowalla: left; Brightkite: right).

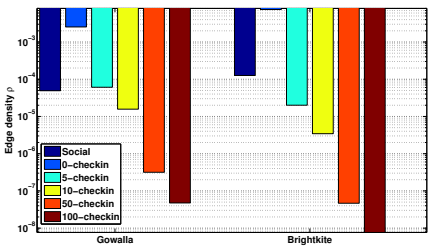

FIG. 2: Edge density (in log-scale)

to the fact that the network itself is sparser for an increased value of k. This inevitably leads to less vertices with common neighbors connect to each other. However, the network can still be transitive if C_G is significantly higher than ρ_G.

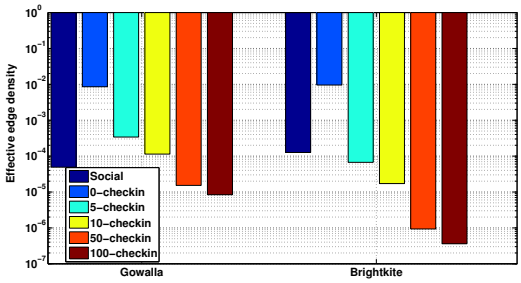

FIG. 3: Effective edge density (in log-scale)

In order to compare the *degree of transitivity* of the social graph and the k-check-in networks we define the following metric, which is applied on two graphs G_1 and G_2:

$$d_t(G_1, G_2) = \frac{C_{G_1}/\rho_{G_1}}{C_{G_2}/\rho_{G_2}} \quad (3)$$

If $d_t(G_1, G_2) > 1$, G_1 is "more transitive" as compared to G_2. The table at Figure 4 depicts $d_t(G_1, G_2)$, where G_1 is the social graph and G_2 is the corresponding k-check-in network. For small values of k the social network exhibits much higher degree of transitivity, while for increased k the case is reversed. As k grows larger, the edges at the graphs that are obtained appear to be highly correlated with each other. In other words, for large k the k-check-in graphs identify tightly connected groups with regards to their spatial trails. The more stringent the constraint on k is, the more knit connected nodes we find.

Empirical fact 3: *For large values of k, the k-check-in structure exhibits higher transitivity compared to the social network.* While for small k the network is much more dense (empirical fact 2), larger k creates edges among groups of people (not only dyads) that have a strong connection with each other. On the contrary, the edges for the case of small k are more close to random, since people can have a few common check-ins with a large number of random people (think of how many places we visit at the course of a day and how many other people go to these places; *trivial* edges).

D. Edge Similarity

The final static property that we examine is the edge similarity between the social and the k-check-in networks. In particular, we are interested in studying (i) the ability of the k-check-in graphs to recover the social links and (ii) the *redundancy* of the k-check-in edges. With the term redundant we loosely refer to edges that exist in the k-check-in graph but not in the social network. Nevertheless, this is not necessarily a drawback of the k-check-in structures, since it can reveal potential pairs of users with a large overlap in interests.

For our purposes we introduce the Social Edge Recall (SER) and Social Edge Precision (SEP) metrics. Let $e^G_{\{i,j\}}$, be one if an edge $\{i,j\}$ exists in graph G and zero otherwise. SER captures the percentage of the actual social connections that exist in the k-check-in graph, while SEP expresses the percentage of the k-check-in graph edges that are actual social links. Formally:

$$SER = \frac{\sum_{\{i,j\}\in E_s} e^k_{i,j}}{m_s} \qquad SEP = \frac{\sum_{\{i,j\}\in E_s} e^k_{i,j}}{m_k} \quad (4)$$

where E_s is the set of social links ($|E_s| = m_s$) and the super/sub-script k is used to refer to the k-check-in graphs.

If both SER and SEP obtain high values simultaneously, then we can say that the two structures are very similar; the k-check-in graph recovers a high number of the social edges, while not generating additional edges that do not exist in the friendship network. Figure 5 depicts our results. In particular, we examine the SEP vs SER curves obtained for different values of k. As we can observe k controls a tradeoff between SER and SEP. Small values of k result in low SEP and high SER, while as we increase k, SER reduces while SEP increases.

When the constraint on the number of common check-ins is loose, the k-check-in graph is able to recover a larger portion of the social links (high SER), but these are accompanied with a larger number of non-social links as well (low SEP). However, we would like to emphasize on the fact that the maximum values of SER are only 0.185 and 0.33 for Gowalla and Brightkite respectively. Less than half of the social edges are being recovered even with the least conservative value for k. This supports our aforementioned argument that similarity in location/mobility trails is only one of the social dimensions that cause friendships to form. Only 18% (33%) of the friends pairs exhibits the lowest possible spatial similarity for Gowalla (Brightkite). On the contrary, as we set more tight constraint on the threshold k, the number of non-social links that are generated are significantly reduced (higher SEP), but also the Social Edge Recall is significantly lower; an artifact of the graph sparcity for large k (Empirical fact 2). Nevertheless, there are still edges in the k-check-in graphs (e.g., $\approx$ 65% of them in Gowalla) that do not correspond to friendships. This can potentially serve as an indicator of strong similarity and a

-	Gowalla	Brightkite
0-checkin	30.25	27.23
5-checkin	2.24	0.76
10-checkin	1.23	0.4232
50-checkin	0.42	0.1927
100-checkin	0.31	0.21

FIG. 4: The k-checkins graph are less transitive compared to the social graph for smaller values of k ($d_t(G_s, G_k)$).

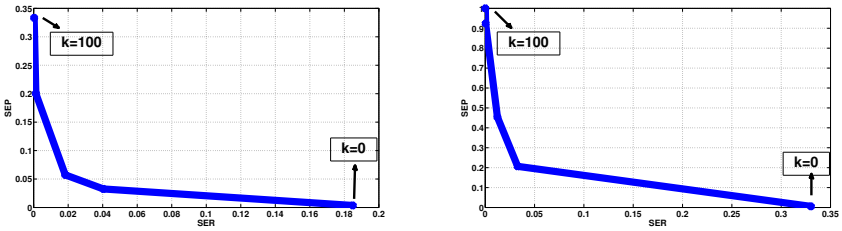

FIG. 5: SER vs SEP (Gowalla: left figure; Brightkite: right figure).

possible baseline for friendship recommendation.

Empirical fact 4: *The similarity of the social and k-check-in graphs is low as captured by SER and SEP.* The two networks encode different information. This finding further supports the fact that social link formation is a complex process that takes place over a large number of dimensions, with location/mobility similarity being only one of them.

IV. Temporal Evolution

For our temporal analysis we split the check-in trails of the users in four snapshots that span equal time periods and we build the corresponding graphs for each one of them.

A. Temporal Edge Density

We first examine the temporal evolution of the edge density. Figure 6 presents our results. As we can see, *across the different snapshots, for both Gowalla and Brightkite and for a given k, the edge density is fairly stable* (**Empirical fact 5**). Also as expected from our static results, for a given snapshot, larger k corresponds to lower edge density.

While the invariance of the edge density across the snapshots is clear, this does not necessarily mean that the k-check-in networks are also stable across the snapshots. Edge density captures only an aggregate property of the network. Therefore, in what follows we examine every edge in the k-check-in graph and its appearance across snapshots.

B. Temporal Edge Stability

A k-check-in edge is characterized as *stable* if it appears in all four snapshots considered. In order to quantify the edge stability, we first consider the union E^k of all k-check-in edges identified over all the snapshots; $E^k = \bigcup_{t\in\{1,2,3,4\}} E^{k,t}$, where $E^{k,t}$ is the set of k-check-in edges in snapshot t. Then for every edge $\{i,j\} \in E^k$, we compute its *stability score* $s_k(\{i,j\})$:

$$s_k(\{i,j\}) = \frac{\sum_{t\in\{1,2,3,4\}} e^{k,t}_{\{i,j\}}}{4}, \forall \{i,j\} \in E^k \quad (5)$$

where extending the notation used before, $e^{k,t}_{\{i,j\}} = 1$ iff edge $\{i,j\}$ appears at the k-check-in graph of snapshot t. The stability score can take four discrete values (i.e., 0.25, 0.5, 0.75 and 1), with a stable edge e having $s_k(e) = 1$.

Figure 7 depicts the probability mass function for $s_k(\{i,j\})$ for different values of k. As we can see the stability score exhibits a single mode distribution (with the exception of Brightkite for $k = 100$). Most importantly, this mode is at $s = 0.25$, which means that every edge appears only at one snapshot! This can have significant implications in services that rely on opportunistic contacts between people and their mobile devices such as delay tolerant network applications.

Empirical fact 6: *The edges of the k-check-in network appear to be highly unstable regardless of the values of k.* Edges are ephemeral and not retained across temporal snapshots of the network, even for loose constraints on k. While there can be many reasons behind this observation (e.g., users do not check-in at the same location many times, users do not check-in to all the places they go, etc.), the implications on systems that rely on opportunistic contacts can be important. These implications can range from purely architectural (i.e., opportunistic networks are not feasible) to less restrictive ones (e.g., LBSN datasets might not be the appropriate source to drive such applications).

V. Related Studies

In this study we have identified and analyzed implicit network structures that emerge in location-based social networks. There exists literature that estimates *hidden* network structures in various contexts. An implicit network structure that appears often in a variety of settings is that of **"call graphs" (or "who-calls-whom" graphs)**. These structures are created through Call Detail Records and two people are assumed to be connected if they have called each other for at least a predefined number of times. For instance, Nanavati *et al.* [4] examine in great detail the structure of call graphs utilizing data from mobile telecom operators. Many times connections identified through the above process are considered as a proxy of the social network when the latter is not known and hence, can be used for a number of sociology studies. For instance, Onnela *et al.* [5] have utilized call graphs to study the infamous Granovetter's hypotheses from traditional social network theory for the strength of weak ties. In different contexts, who-calls-whom graphs have used to study mobility models (e.g., [6]–[9]) as well as urban planning related problems (e.g., [10]).

Implicit (social) network structures can also be identified through **affiliation networks**. The latter integrate "foci" of social interactions with the social graph. They have been used for inferring ties in a variety of settings; WikiPedia article and editor page editing [11], event-based online services [12], company directors [13] etc. In our context, venues can be thought of as the foci, and an affiliation edge exists between a user and a venue, if the former has checked-in the latter (we have used such models in our previous work [14]).

Closest to our work are the recent studies from Brown *et al.* [15] [16]. The authors define the notion of "placefriends", which is essentially the 0-check-in ties, and utilize it to study the formation of online communities, exhibiting the importance

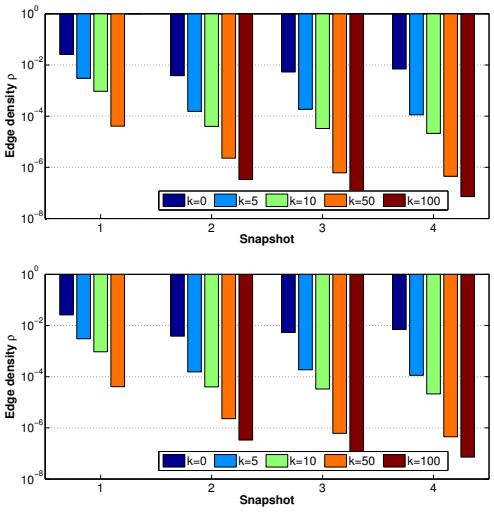

FIG. 6: The edge density of k-check-in graphs is fai stable over time, for a given k (Gowalla: top; Brightkite: bottom).

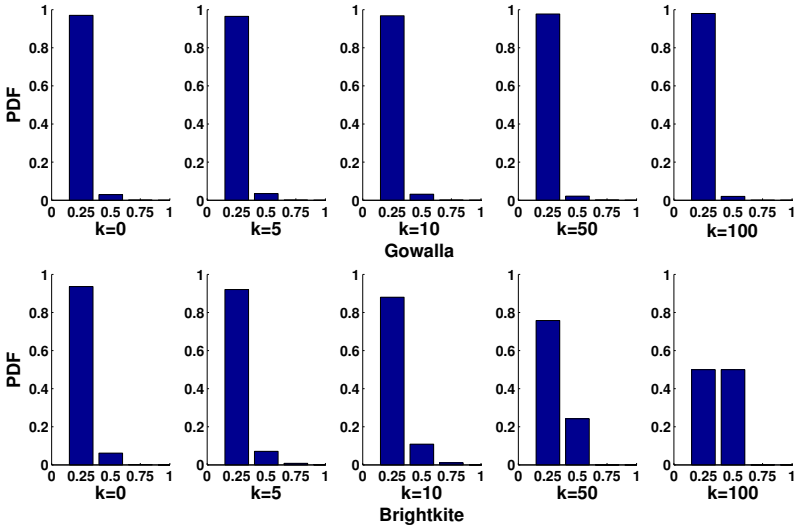

FIG. 7: Edges in the k-check-in graphs are ephemeral; they appear only in one snapshot with high probability.

of indirect structures. Our work complements the aforementioned studies by examining the actual static and temporal properties of a "generalized placefriend graph".

VI. CONCLUSIONS

In this paper, we study the k-check-in graph, a user-based, implicit network structure that appears in location-based social networks. In particular, a tie between two people is assumed, if they have checked-in to more than k same places. Using two LBSN datasets, we study the static and temporal properties of this network and we find that it varies significantly from the actual social graph. The study of these structures can have significant implications on services and studies that rely on user's co-locations. In the future, we plan to examine (i) other hidden network structures (also ones that are created among venues), and (ii) their application on epidimiology, urban informatics and p2p mobile content delivery.

REFERENCES

[1] L. Sproull and S. Kiesler, "Reducing social context cues: Electronic mail in organizational communication," in *Management Science, 32(11)*, 1986.
[2] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *ACM KDD*, 2011.
[3] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
[4] A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi, "On the structural properties of massive telecom call graphs: Findings and implications," in *ACM CIKM*, 2006.
[5] J. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.-L. Barabasi, "Structure and tie strengths in mobile communication networks," in *PNAS, 104(18):7332-7336*, 2007.
[6] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *ACM KDD*, 2011.
[7] M. Gonzalez, C. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," in *Nature, 453(7196):779-782*, 2008.
[8] C. Song, T. Koren, P. Wang, and A.-L. Barabasi, "Modelling the scaling properties of human mobility," in *Nature Physics*, 2010.
[9] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," in *Science, 327(5968):1018*, 2010.
[10] R. Becker, R. Caceres, K. Hanson, J. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky, "A tale of one city: Using cellular network data for urban planning," in *IEEE Pervasive Computing, 10(4)*, 2011.
[11] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *ACM KDD*, 2008.
[12] X. Liu, Q. He, Y. Tian, W. Lee, J. McPherson, and J. Han, "Event-based social networks: Linking the online and offline social worlds," in *ACM KDD*, 2012.
[13] G. Davis, M. Yoo, and W. Baker, "The small world of the american corporate elite," in *Strategic Organization, vol. 1 no. 3 301-326*, 2003.
[14] K. Pelechrinis and P. Krishnamurthy, "Location affiliation networks: Bonding social and spatial information," in *ECML/PKDD*, 2012.
[15] C. Brown, V. Nicosia, S. Scellato, A. Noulas, and C. Mascolo, "Where online friends meet: Social communities in location-based networks," in *AAAI ICWSM (Poster Session)*, 2012.
[16] ——, "The importance of being placefriends: Discovering location-focused online communities," in *ACM WOSN*, 2012.

VII. APPENDIX

The average local clustering coefficient tends to overestimate the network clustering coefficient, since it is dominated by vertices with small degree, which have a large local cc. Another definition has been proposed for calculating the network's clustering coefficient (e.g., see [3]):

$$C_G = \frac{6 \cdot (number\ of\ triangles)}{(number\ of\ paths\ of\ length\ two)} \tag{6}$$

Furthermore, in order to quantify the transitivity of a network, we mentioned in Section III-C that we compare the network's cc with the edge density ρ. Nevertheless, this comparisson holds completely true if we have a graph with a Poisson degree distribution. Hence, for a graph with a different degree distribution, we need to compare the measured cc of the network with the following quantity:

$$\frac{1}{n}\frac{[<\kappa^2> - <\kappa>]^2}{<\kappa>^3} \tag{7}$$

where $<\kappa^m>$ is the m-th moment of the degree distribution of the graph. Equation 7, reduces exactly the edge density when $P(\kappa)$ follows a Poisson distribution. In our case, we do not have a Poisson degree distribution but we keep the comparisson with the edge density for simplicity reasons. We also use the mean local cc, since this is followed by the majority of the relevant literature for calculating C_G.