Modeling City Locations as Complex Networks: An initial study

Lu Zhou, Yang Zhang, Jun Pang and Cheng-Te Li

Abstract Analyzing data collected from location-based social networks can reveal complex structure in human social relations. It can also lead to deep understandings of human mobility and help characterize city locations and their connectivity. In this paper, we construct location networks for six cities using a large-scale Instagram dataset. We find that these location networks share many topological features as in other different types of networks, along with properties specific to their cities. By mapping locations to their geographical coordinates, we further show that (1) our construction method can effectively reveal popular city locations, and (2) for two locations there is no clear correlation between their network distance and geographical distance. Moreover, all six location networks contain three or four large communities covering almost all locations in a city and the large communities in each city often exhibit clear spatial differences in geographical space.

1 Introduction

With the advancement of urbanization process, more and more people live in cities. The United Nation published a report in 2014 stating that 6 billion people will live in cities by 2050 (double the current amount). On one hand, city life brings people a lot of convenience. For instance, people can taste different cuisines and buy products from all over the world. On the other hand, it also results in many problems such as traffic jam and heat island effect. Many efforts have been taken to tackle these

Lu Zhou, Yang Zhang, Jun Pang (corresponding author)
Faculty of Science, Technology and Communication & Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg, Luxembourg
e-mail: `firstname.lastname@uni.lu`

Cheng-Te Li
Department of Statistics, National Cheng Kung University, Tainan, Taiwan
e-mail: `chengte@mail.ncku.edu.tw`

problems and improve people's life quality, such as the techniques of smart city. In particular, one fundamental component of cities, i.e., location, attracts a number of attentions in academic research. Existing work exploited location information to infer users' friendship [16], recommend new locations [7] and measure urban deprivation [18]. However, most of these work treat locations separated from each other, while the interactions among locations are often not studied.

Online social networks have been the most successful application during the past decade, major companies including Facebook and Instagram have attracted a large number of users. With the development of mobile devices, social networks have been extended to geographical space. Nowadays, more and more social network users are sharing their photos or statues labeled with geographic information, namely *check-in*. The large-scale check-in data can be naturally used to describe human mobility, and provide us means to study location relations.

Network is one of the most common perspectives to study interactions in complex systems, it has attracted academia a particular interest in recent years, e.g., social networks, biological networks, transportation and computational history [17]. Locations can also be organized into networks, and such networks can be used to study connectivity among locations. Most of existing studies construct location networks based on user transitions among the locations [13, 10]. Namely, a location network is built as a directed weighted graph, whose nodes are locations and each edge is formed between two locations if a user directly moved between such two locations during a pre-defined time period. However, all these networks only reflect users' movement from one location directly to another in a short time, but they cannot describe the overall connections between locations. In our work, we present a different approach to construct location networks for six cities, including New York, Los Angeles, London, Paris, San Francisco and Tokyo, with more than 15 million check-in data collected from Instagram, and conduct empirical analyses on these networks to reveal their network features and properties related to location geographical coordinates. Our contributions can be summarized as follows:

- We propose a new method for constructing location networks (Section 2). In our construction, we treat each location as a node (similar to [13, 10]) and define a weighted edge between two nodes, which describes users' check-in behaviors and measures the strength of the connectivity between these two nodes.
- We adopt four measurements to describe the constructed location networks, including mean degree, degree distribution exponent, weighted clustering coefficient and average shortest path length. We find that location networks have similar topological features as in other complex networks. In addition, location networks also exhibit differences specific to their city (Section 3).
- We rank location popularity based on the PageRank algorithm [15], and show that our construction method leads to more effective rankings of locations when compared to rankings, e.g., based on location entropy [10] (Section 4).
- We reveal the relation between geographical distances and network distances between any two nodes (Section 4). To our surprise, we discover that there is no clear linear correlation between these distances.

Table 1 Summary of the original dataset

City	Users	Locations	Check-ins
New York	95,624	21,646	2,566,328
London	56,663	10,423	1,199,500
Paris	22,409	6,916	458,291
Los Angeles	85,788	19,412	2,055,290
Tokyo	35,487	19,610	835,896
San Francisco	25,374	7,302	585,727

- We detect network communities in our location networks, and discover that there only exist three or four large communities containing almost all the locations in each city. Such communities also exhibit differences in geographical space (Section 5).

2 Network Construction

2.1 Dataset

Instagram is a photo-sharing social network service, with a fast growing user number. Instagram allows users to label locations when publishing photos. It is worth noticing that the location information on Instagram is imported from Foursquare, a social network that concentrates on location sharing. In addition, the authors of [11] have shown that Instagram users are much more willing to share locations than other social network users (e.g., 31 times more than Twitter users), which makes Instagram a suitable source for collecting check-in data.

To find users in the six cities, we start from querying Foursquare's API to collect the location IDs in each city, together with each location's rating. Then we use Instagram's API to collect users' check-ins in Instagram at the corresponding location IDs. In our experiment, we focus on users' check-ins in year 2015. To resolve the data sparseness issue, for each city, we concentrate its users with at least 10 check-ins. Moreover, locations that are visited only by one user are filtered out. In the end, we use more than 7M check-ins in total to construct six location networks (one for each city). Table 1 summarizes our dataset.

2.2 Nodes, Edges and Weights

Our general goal is to study city locations from the network perspective, for instance, to understand why locations are connected to each other and how strongly the locations are connected. Along with a (complex) topological structure, many real

networks display a large heterogeneity in the capacity and intensity of connections. Thus, it is important to have a measurement which reflects the relevancy between any two locations in a city. Our intuition is to take city locations as nodes in the location network, locations are connected through users' check-in behaviors. In order to measure the strength of the connections, we need a measure to estimate and summarize how different users behave on the connections.

First, if two locations have been visited by one user, we consider that these two locations are associated with each other and there exists an edge between them. Let $L(u)$ denote the set of all locations that user u has visited. We define $E(u) = \{(\ell_i, \ell_j) \mid \ell_i, \ell_j \in L(u) \wedge \ell_i \neq \ell_j\}$ as the set of edges constructed from u's check-in data. Based on this definition, we build an undirected graph, in which nodes denote (check-in) locations, an edges between any two nodes mean that the two locations have been visited by one user. Let V be the set of all check-in locations for a given city. We construct a location network $G = (V, E)$, where $E = \{(\ell_i, \ell_j) \mid (\ell_i, \ell_j) \subseteq V \times V\}$ captures all existing connections between two locations through users' check-in behaviors.

The next step is to quantify the strength for each edge (ℓ_i, ℓ_j) in G. An edge connecting two locations can be visited by many users, whose check-in behaviors can vary differently: some may visit many different locations while others visit only a few. Thus, we need to take users' active levels into account to measure edge weights. We adopt the *Shannon entropy* to quantify a user's active level (similar to the definition of location entropy to measure location's popularity [6]):

$$entropy(u) = - \sum_{\ell_i \in L(u)} p_{\ell_i} \cdot ln(p_{\ell_i}) \tag{1}$$

where $p_{\ell_i} = u_{\ell_i} / \sum_{\ell_j \in L(u)} u_{\ell_j}$ describes the probability that user u visited location ℓ_i. It is easy to see that *entropy*(u) depends on both the diversity of locations and the frequency how u visited those different locations.

Then, we use users' active levels to define two locations' (ℓ_i and ℓ_j) edge weight as the following:

$$W_{\ell_i \ell_j} = \sum_{u \in U_{\ell_i \ell_j}} entropy(u) \cdot \frac{\sqrt{u_{\ell_i} u_{\ell_j}}}{\sum_{(\ell_s, \ell_t) \in E(u)} \sqrt{u_{\ell_s} u_{\ell_t}}} \tag{2}$$

where $U_{\ell_i \ell_j}$ represents the set of users who visited both locations ℓ_i and ℓ_j, u_{ℓ_i} denotes user u's number of check-ins at ℓ_i. The factor $\sqrt{u_{\ell_i} u_{\ell_j}}$ reflects how often u visits both ℓ_i and ℓ_j, and it is normalized by all pairs of locations u has visited. By multiplying *entropy*(u) and $\frac{\sqrt{u_{\ell_i} u_{\ell_j}}}{\sum_{(\ell_s, \ell_t) \in E(u)} \sqrt{u_{\ell_s} u_{\ell_t}}}$, it shows how much a user u's check-in behavior can be cast onto the edge (ℓ_i, ℓ_j). Furthermore, in order to reflect the contributions made by all the users who visited both ℓ_i and ℓ_j, we sum up each user's contribution to have the final weight for the edge (ℓ_i, ℓ_j).

In the end, we construct a weighted graph, representing a location network for each city: $G = (V, E)$, with a function $W : E \to R^+$ assigning a positive value to

Table 2 Statistics for the six constructed location networks

City	Nodes	Edges	Min Weight	Max Weight	Mean Weight
New York	21,646	5,697,507	8.9673e-05	286.2492	0.0383
London	10,423	1,861,304	2.1687e-04	165.5303	0.0647
Paris	6,916	655,793	1.7152e-04	655.1800	0.0665
Los Angeles	19,412	3,881,191	8.5379e-05	381.1223	0.0468
Tokyo	19,610	2,635,335	1.0492e-04	157.8897	0.0287
San Francisco	7,302	1,334,240	1.7223e-04	100.6024	0.0407

Table 3 Basic information for each city

City	Population	Size	Population density
New York	8.41	789	10,659.06
London	8.67	1,572	5,515.27
Paris	12.29	12,012	1,023.14
Los Angeles	9.82	1,214	8,088.96
Tokyo	13.62	2,188	6,224.86
San Francisco	0.84	121	6,942.15

every edge in G. For each $(\ell_i, \ell_j) \in E$, we have $W(\ell_i, \ell_j) = W_{\ell_i \ell_j}$ as defined in Equation 2. Table 2 summarizes the numbers of nodes and edges, the minimal, maximal and mean weights, for the six constructed location networks. For all the location networks, most edges have small weights (this can be concluded from the mean weights), while, there exist a few edges with large weights, which, we suppose, often connect the most popular locations in the cities. In order to confirm our hypothesis, we obtain each pair of locations which are connected by an edge with the largest weight for each city. For example, there exists an edge with the largest weight connecting Rockefeller Centre and Times Square, which are known as two of the most famous locations in New York City. Similarly, the other location pairs are also the most famous places in the corresponding cities: Tower Bridge and Tower of London in London, Musee du Louvre and Notre-Dame de Paris in Paris, Staples Centre and Dodger Stadium in Los Angeles, Tokyo Disneyland and Tokyo Dome in Tokyo, AT&T Park and Union Square in San Francisco.

In addition, it is easy to see that New York city has the largest location network, followed by Los Angeles and Tokyo, and then by London. Paris and San Francisco have relatively small location networks. The reason for the different sizes of location networks are mostly due to the difference between the population and the size of each city. We list all the basic information (i.e., population, size and density) of these six cities in Table 3. From Tables 1, 2 and 3, we find that New York has the largest population density, and its number of Instagram users, check-ins, nodes and edges are also the highest. Meanwhile, all the corresponding parameters in Paris are the smallest. Thus, we observe a positive correlation between the size of our location networks and city's population density.

3 Graph Measurements and Analysis

After constructing the (weighted) location networks, we analyze their properties to have a representative description of these networks. In the following, we present four measurements, as discussed in [12].

Mean degree (MD). Even though our network is weighted, to compare with other networks, we consider the unweighted mean degree here. That is, for each node, we compute its number of degrees and calculate the mean of all the nodes' number of degrees as mean degree.

Degree distribution exponent (DDE). In general, the degree distribution is the probability distribution of these degrees over the whole network. For our location networks, we need to construct a histogram of the degrees. Like other power-law degree distributions, the histogram is highly right-skewed. It means that its degree distribution has a long right tail of values that are far above the average, indicating that more nodes have smaller degrees while less ones have much larger degrees. To describe the degree distribution of each location network, we compute the exponent α for each degree distribution curve: $p_k \sim k^{-\alpha}$, where p_k denotes the number of each degree k, and α means the exponent for each degree distribution curve.

Weighted clustering coefficient (WCC). Clustering coefficient captures the degree to which nodes in a graph tend to cluster together. In other words, it is related to the number of closed triangles in the neighborhood of a node. Here, we apply the local clustering coefficient algorithm [19, 14]. For a node, its clustering coefficient is the fraction of the number of present links over the total number of possible links between its neighbors. Therefore, the outcome strictly ranges between 0 and 1, where 0 denotes that no links exist between the neighbors, and 1 if all possible links exist. The equation for clustering coefficient of any node in a location network is given as

$$WCC_{\ell_i} = \frac{\sum_{\ell_j,\ell_k \in Cnei(\ell_i)} W_{\ell_i,\ell_j} + W_{\ell_i,\ell_k}}{\sum_{\ell_m,\ell_n \in Nei(\ell_i) \wedge \ell_m \neq \ell_n} W_{\ell_i,\ell_m} + W_{\ell_i,\ell_n}} \tag{3}$$

where $Cnei(\ell_i)$ denotes a set of pairs of locations which are both neighbors of ℓ_i and are also connected in the network, namely closed triplets. For each node, we sum the value of the closed triplets that are centred on the node and divide it by the total value of all triplets centred on the node. The larger coefficient of one location implies that user who visits this location will also visit its neighbors more frequently.

After obtaining the coefficient for each location, we compute the average clustering coefficient value for the whole network. To some extent, the average value reflects the density of the whole network.

Analysis. Table 4 summarizes the computed four measurements for the six location networks. We compare their features with other typical real-life networks, such as biological networks and technological networks. First, we measure the proportion of existed edges to all possible edges in each location network (New York: 0.024, London: 0.034, Paris: 0.027, Los Angeles: 0.021, Tokyo: 0.014, San Francisco: 0.050).

Table 4 Properties comparison with six cities

City	Nodes	Edges	MD	DDE α	WCC	CRC
New York	21,646	5,697,507	526.426	3.818	0.011	-9.5e-4
London	10,423	1,861,304	357.153	3.116	0.013	0.0448
Paris	6,916	655,793	189.645	2.185	0.019	0.0670
Los Angeles	19,412	3,881,191	399.875	3.500	0.014	0.0560
Tokyo	19,610	2,635,335	268.775	3.493	0.014	0.0427
San Francisco	7,302	1,334,240	365.445	3.970	0.012	0.0339

Compared with the given statistics of a number of published networks [12], we can find that the edge proportions in our location networks are higher than other networks, where most edge proportions are less than 10^{-4}. This implies that most locations in our networks have more connections with others. Furthermore, the nodes with more connecting edges are often popular locations in the cities. Our location networks' mean degrees are much bigger than other networks [12] (e.g., WWW Altavista network: 10.46, physics coauthorship: 9.27, metabolic network: 9.64), which is mostly due to the large numbers of edges in our location networks. On the other hand, clustering coefficient values of our location networks are much lower than other networks [14]. The main reason for this result is that, due to Equation 3, the denominator can be influenced largely by the number of edges in networks. As our networks have a plenty of edges, the values of clustering coefficient arrive at a low level [1]. Since the clustering coefficient measures the density of triangles in a network, higher value means the networks are much denser and the neighbors of a node in the network are more likely to be connected. The computed clustering coefficients in Table 4 imply that the neighbors of one location are not necessarily connected and influenced by other locations. When considering the exponent α, we find that our network and other networks have similar values. It means the degree distribution of our location networks also follows a power law form. In general, we can conclude that location networks share basic topological features of complex networks while having their own characteristics.

Table 4 also shows differences among the six cities. London's location network has the largest mean degree, even though the location network of New York has a much larger number of edges than London. This is due to that the edge proportion in New York's location network is much lower than London's, and its degree distribution curve also has a longer tail than London's. This is also reflected in the clustering coefficients: London's location network has a higher clustering coefficient than New York's, meaning that its edge weights are relatively larger. For another example, Tokyo has many more locations than Paris, and the numbers of nodes and edges in its location network are much larger than Paris' network. However, the clustering coefficient in Tokyo's location network is smaller than Paris'. Meanwhile, Tokyo's

---

[1] Clustering coefficient for unweighted graphs will increase when there are more edges. However, WCC considers not only the number of edges, but also the edge weights. Since most of the edges in our location networks have small weights, WCC will decrease when there are more edges.

location network's exponent α is higher than Paris', which means most nodes in Tokyo's network have much smaller degrees than nodes in Paris' location network. These also explain why the mean degree of Tokyo's location network (i.e., 7.717) is much smaller than Paris's (i.e., 12.612). We also observe that the shortest path length of location network in Paris is the largest while New York has the smallest shortest path length. From these, we can conclude that each location network has features specific to their city.

4 Location Ranking and Distances

4.1 Location Ranking

Location popularity has received many attentions in recent years, it is considered as an essential part for building real-world applications such as location recommendation and friendship prediction. With our location networks constructed, we want to measure each location's popularity as well. To do so, we adopt one of the most classical algorithms on measuring the popularity of nodes in networks, i.e., PageRank.

To further validate our popularity measurement, we take location entropy [6] for comparison. Location entropy is one common measurement for location popularity. If a location has a high entropy, it shows that many different users have visited the location, thus the location is popular. To compare popularity ranking on locations, based on the results of executing PageRank on our networks and computing the location entropy of each node in our networks, we adopt a quantitative way with the help of location ratings. As mentioned in Section 2, due to the connection between Foursquare and Instagram's APIs, when collecting check-in data at a certain location, we are able to get the location's rating from Foursquare (between 1 and 10). We treat these ratings as a ground truth to rank location popularity.

We compute the correlation coefficients between ratings and both PageRank scores and location entropies for all locations in each city. The results in Table 5 show that PageRank scores are much more correlated with ratings than location entropies. With the assumption that high ratings indicate popular locations, we conclude that PageRank scores, computed on our location networks, are very effective in evaluating location popularity. This further demonstrates the usefulness of our network construction method (note that location entropy only concentrates on the check-ins of each location, but not on the relations among locations).

4.2 Location Distances

Besides location popularity, locations have other important properties, i.e., their geographical coordinates. Where a location is geographically located, in many cases,

Table 5 Comparison between correlation coefficients obtained by PageRank and location entropy

City	PageRank	location entropy	City	PageRank	location entropy
New York	0.4356	0.1420	Los Angeles	0.4714	0.1885
London	0.3774	0.2039	Tokyo	0.4169	0.1720
Paris	0.3952	0.1482	San Francisco	0.4449	0.2079

can determine its fundamental properties such as its functionalities and its value in the real estate market. In this subsection, we study the relation between locations organized as a network and their actual geographical coordinates.

To our knowledge, human mobility is probably constrained geographically by the distance one can travel within a day [9]. Thus, we assume that the geographic distance (Geo_{dist}) between any two locations has an influence on the connectivity between them in location networks. In other words, Geo_{dist} has an impact on user's judgement on the choice of visiting different locations. Thus, we need to obtain the check-in correlation coefficient between Geo_{dist} and Net_{dist} (i.e., network distance, that is, the shortest path length in the network between any two locations), in order to determine whether there exists any correlation between such two kinds of distance. In particular, Geo_{dist} means the actual geodesic distance computed by the Euclidean distance using latitude and longitude values, while the weights on edges in the constructed location network can be considered as Net_{dist}. Finally, we get the correlation coefficient between these two kinds of distance for each city, shown in last column (CRC) in Table 4, where most of the correlation coefficient values are close to zero. This suggests that there is no linear relation between geographical distance and edge weight defined in the paper.

5 Community Analysis

To further understand the structure of our location networks, we perform analysis on their community structures. In simple terms, a community is a subset of nodes in a network with links among the community members are much more than between the community and the rest of the network. According to [20], the community structure is one of the most useful granularity to study networks, it has been used by researchers, e.g., to study interactions between modules [1] and predict unobserved connections [4].

5.1 Network Community

We adopt one of the classic approaches, namely the fast greedy modularity optimization algorithm [5] (fast greedy), to detect network communities in our location

Table 6 Summary of detected largest communities for each city

| City	Comm. 1	Comm. 2	Comm. 3	Comm. 4	#. Comm.	%. Top communities
New York	7,679	5,775	5,678	2,473	18	99.8%
London	3,626	3,450	3,299	-	15	99.5%
Paris	2,775	2,333	1,646	-	11	97.7%
Los Angeles	6,309	5,326	4,611	3,133	14	99.8%
Tokyo	6,865	5,845	4,884	1,918	21	99.5%
San Francisco	2,524	2,238	2,192	-	16	95.2%

networks. The algorithm is essentially a fast implementation of the first community detection algorithm based on modularity optimization [8]. Starting from a set of isolated nodes, the fast greedy algorithm adds edges from the original graph to maximize the modularity [5] of the newly generated graph at each step.

We obtain multiple network communities for each city's location network with many communities only containing a few locations while several large (three or four) ones containing most of the locations in the network. Table 6 presents the statistics on the sizes of the largest communities in each city. For instance, the top 4 communities in New York and Los Angeles contain more than 99.8% of the locations in these two cities. This observation corresponds well to other complex networks [5] for containing only a few large components in their network structures.

5.2 Geographical Community

Next, we project network communities into a geographical space, and find that different communities are associated with different geographical signatures. Figure 1 exhibits the network communities in New York. The locations in the black community in New York (top left in Figure 1) are located throughout the city, including Manhattan, Brooklyn and Queens, while the blue and green communities mainly concentrates on Manhattan. Moreover, most locations of the blue community are in midtown, downtown and upwest while the green community has many locations in upeast side and the central park. Meanwhile, the locations in the red community are distributed more uniformly compared to other three communities with an interesting concentration in Jersey city. Similar observations can be made in all other five cities.

Recall the conclusion in Section 4.2 that there is no obvious correlation between geographical distances and edge weights. Meanwhile, in this section we find that different communities (partitioned based on the densely connected edges) distribute at different geographical spaces. This seems to be a contradiction. However, we need to mention that locations of different communities are not located completely differently, and most communities are overlapped in city centers.
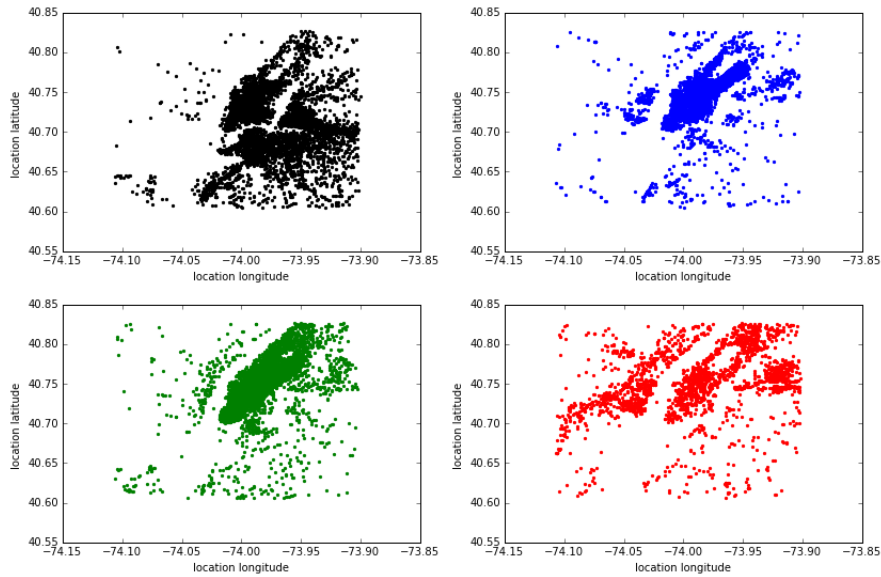
Fig. 1 The largest communities of New York shown on the map separately

6 Conclusion and Future Work

In this paper, we have constructed weighted location networks based on check-in behaviors of millions of Instagram users across six cities in the world. Our initial study of the constructed location networks has focused on their basic features as defined for complex networks. Moreover, we mapped locations in each city to their corresponding geographical coordinates, and discovered that our construction method is effective in revealing popular locations. We also discovered that there is no linear correlation between geographical distance and our edge weight. For each location network, we found a few largest communities covering almost all locations in a city, as well as such communities have an obvious distribution geographically.

The way how we constructed location networks has an emphasis on quantifying a user's over-all check-in behavior and then distributing it to all edges connecting two locations that the user has visited. It is interesting to compare our construction methods with other methods for constructing location networks from location check-ins. Our next step is to further analyze community structure of each location network, e.g., through the use of different community detection algorithms and the study of their location category distributions. For instance, we want to apply the recently proposed clustering method in [3] to check whether location networks also consist a higher-order organization. Based on a larger Instagram dataset we have collected, we will also investigate growth models for location networks, for example, by following the application of preferential attachment to the growth of the Internet [2].

References

1. Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed membership stochastic block-models. Journal of Machine Learning Research **9**, 1981–2014 (2008)
2. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
3. Benson, A., Gleich, D., Leskovec, J.: Higher-order organization of complex networks. Science **353**(6295), 163–166 (2016)
4. Chang, J., Blei, D.M.: Relational topic models for document networks. In: Proc. 12th International Conference on Artificial Intelligence and Statistics (AIStats), *JMLR Proceedings*, vol. 5, pp. 81–88. JMLR.org (2009)
5. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E **70**(6), 066,111 (2004)
6. Cranshaw, J., Toch, E., Hone, J., Kittur, A., Sadeh, N.: Bridging the gap between physical location and online social networks. In: Proc. 12th ACM Conference on Ubiquitous Computing (UbiComp), pp. 119–128. ACM (2010)
7. Gao, H., Tang, J., Hu, X., Liu, H.: Exploring temporal effects for location recommendation on location-based social networks. In: Proc. 7th ACM Conference on Recommender Systems (RecSys), pp. 93–100. ACM (2013)
8. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences **99**(12), 7821–7826 (2002)
9. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
10. Hristova, D., Williams, M.J., Musolesi, M., Panzarasa, P., Mascolo, C.: Measuring urban social diversity using interconnected geo-social networks. In: Proc. 25th International Conference on World Wide Web (WWW), pp. 21–30. ACM (2016)
11. Manikonda, L., Hu, Y., Kambhampati, S.: Analyzing user activities, demographics, social network structure and user-generated content on Instagram. CoRR **abs/1410.8099** (2014)
12. Newman, M.E.J.: The structure and function of complex networks. SIAM Review **45**(2), 167–256 (2003)
13. Noulas, A., Shaw, B., Lambiotte, R., Mascolo, C.: Topological properties and temporal dynamics of place networks in urban environments. In: Proc. 24th International Conference on World Wide Web (WWW Companion), pp. 431–441. ACM (2015)
14. Opsahl, T., Panzarasa, P.: Clustering in weighted networks. Social Networks **31**(2), 155–163 (2009)
15. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab (1999)
16. Scellato, S., Noulas, A., Mascolo, C.: Exploiting place features in link prediction on location-based social networks. In: Proc. 17th ACM Conference on Knowledge Discovery and Data Mining (KDD), pp. 1046–1054. ACM (2011)
17. Schich, M., Song, C., Ahn, Y.Y., Mirsky, A., Martino, M., Barabasi, A.L., Helbing, D.: A network framework of cultural history. Science **345**(6196), 558–562 (2014)
18. Venerandi, A., Quattrone, G., Capra, L., Quercia, D., Saez-Trumper, D.: Measuring urban deprivation from user generated content. In: Proc. 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW), pp. 254–264. ACM (2015)
19. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998)
20. Yang, J., McAuley, J.J., Leskovec, J.: Detecting cohesive and 2-mode communities indirected and undirected networks. In: Proc. 7th ACM International Conference on Web Search and Data Mining (WSDM), pp. 323–332. ACM (2014)