Stratification structure of urban habitats

Aonghus Lawlor, Cathal Coffey, Rory McGrath, and Alexei Pozdnoukhov

National Centre for Geocomputation
National University of Ireland Maynooth
alexei.pozdnoukhov@nuim.ie
http://ncg.nuim.ie

Abstract. This paper explores the community structure of a network of significant locations in cities as observed from location-based social network data. We present the findings of this analysis at multiple spatial scales. While there is previously observed distinct spatial structure at inter-city level, in a form of catchment areas and functional regions, the exploration of in-city scales provides novel insights. We present the evidence that particular areas in cities stratify into distinct "habitats" of frequently visited locations, featuring both spatially overlapping and disjoint regions. We then quantify this stratification with normalized mutual information which shows different stratification levels for different cities. Our findings have important implications for advancing models of human mobility, studying social exclusion and segregation processes in cities, and are also of interest for geomarketing analysts developing fidelity schemes and promotional programmes.

Keywords: community detection, location-based social networks, urban mobility

1 Introduction

State-of-the-art mobility models based on gravity laws [1], intervening opportunities [2], competing destinations [3] or the recent parameter-free radiation model [4] operate with flow data aggregated over distinct spatial areas. Meanwhile, the mobility of individuals is still relatively unexplored. The recent growth of interest to studying human mobility spans from availability of detailed datasets of individuals movements. These data present irregularly sampled (both in time and space) locations of individuals either via cell phone usage logs [5], manifested by users themselves by "check-ins" in location-based social network services [6], or allowing geo-tagging feature in Twitter [7, 8]. Empirical evidence suggests that regular commutes is a dominating mobility pattern [5] which also governs occasional fluctuations as people tend to arrange their travel plans by considering accessibility and convenience with regard to their primary locations such as home and work. This rational paradigm in human mobility puts foundations to many models used in transportation research and urban planning.

Predictive modelling of human movement beyond regular commutes is a challenging task. Some studies suggest the presence of the so-called "habitats" [9], formation of which is likely to be related to spatial choice processes in human decision making. The habitats representation allows certain improvement upon the Levý flight model, particularly, it features slow temporal dynamics of exploration of all but non-primary habitat. Both common sense and empirical observations [10, 7, 11] suggest the importance of social influence on the formation of atypical patterns of mobility. People tend to follow recommendations of their friends in planning travel, or joining them on a trip to explore new areas and visit particular places for recreation, leisure or tourism.

1.1 Contributions of the paper

Social networks possess distinct community structure which often show geographical patterns both at inter-city [12] as well as intra-city scales [13]. As the popularity of places in a city spreads over social networks these patterns may manifest themselves in the structure of network of locations. In other words, we hypothesize that locations become popular within different population groups formed either by social or geographical proximity, and that community structure of social networks translates into a similar structure of popularity of places in a city within these groups.

This implies the existence of "urban habitats", i.e. the disjoint groups of locations possibly spread over the same spatial area but popular within different and non-overlapping user groups. In this work we present an evidence of such stratification observed in location-based social network data by defining a similarity measure over a pair of locations through the number of their common visitors and applying community detection on the resulting network of locations. We propose to use the normalized mutual information as a measure that quantifies this stratification.

2 Location-based social network data

Dataset used in this study was collected from an online location-based social network [11]. It consists of all the public check-in data at Brightkite between April 2008 to October 2010. Brightkite also contain an explicit social network. The friendships are directed and only bi-directional friendships were considered as edges in the study below.

The following pre-processsing was applied to the dataset. The Brightkite data contained 51,406 unique users, 772,966 unique locations across the world and a total of 4,747,281 check-ins. It was also observed that no one user had over 2,100 check-ins. We found that this cut-off point was due to the limitation of the Brightkite API. This API only allows for the 2,100 most recent check-ins to be retrieved. Despite this cap being applied we found that the sampling time still spanned over several months. Several steps of filtering were applied to the Brightkite data before being used. We found that a total of 9.88% of the check-in data had a precision of less than 10^{-3}. This was due to users being able to specify the resolution of there location. Some users opted to only specify the

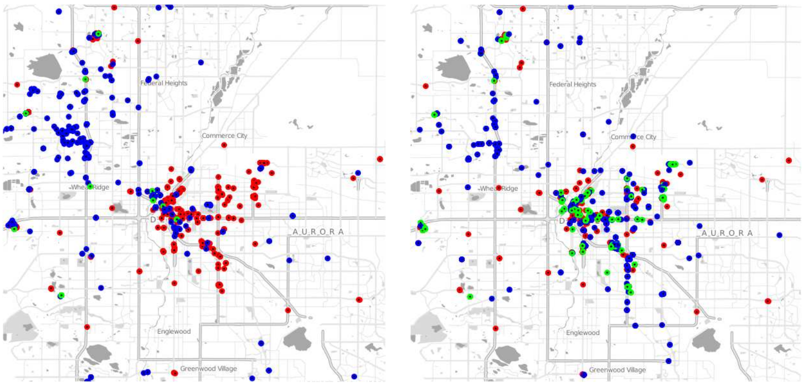

Fig. 1: Two Brightkite users connected by friendship ties visit a larger number of common locations (green dots, on the right), as opposed to two non-friends in the same region (left).

state or country they were in. As we are dealing with significant locations in cities these values were removed. We primarily focused on the cities of Denver, Los Angeles, San Francisco and Tokyo. Bounding boxes were placed around each of these cities: if over 50% of a persons check-in locations fell within the cities bounding box they were assigned to that city. This resulted in 1,273 people in Denver, 2,439 people in Los Angeles, 1,913 people in San Francisco and 1,979 people in Tokyo.

Figure 1 illustrates spatial locations of the check-ins of two users in Denver region. Locations shown in green were visited by both users. The users on the right of Figure 1 are "friends" on the Brightkite network, and visited a larger number of common locations as opposed to the two users on the left of Figure 1 who have no friendship ties. This finding motivated us to investgate this pattern at large using community detection techniques.

3 Community detection

Community detection has become a popular approach in structural analysis of complex networks [14]. It involves finding a partition which minimizes the density of links within groups relative to the density of links between groups, thus finding closely connected groups of nodes. This exact metric, known as modularity [15], is a popular objective function in community detection. Particular challenges concerned with community detection are computational efficiency and scalability of the methods when applied to large networks. Exact algorithms are often inefficient and various heuristics and greedy approaches has to be employed to detect communities in the networks of millions of nodes. Various extensions of basic community detection problem deal with weighted and directed networks, as well as consider the task of finding overlapping communities and identifying hierarchical structures.

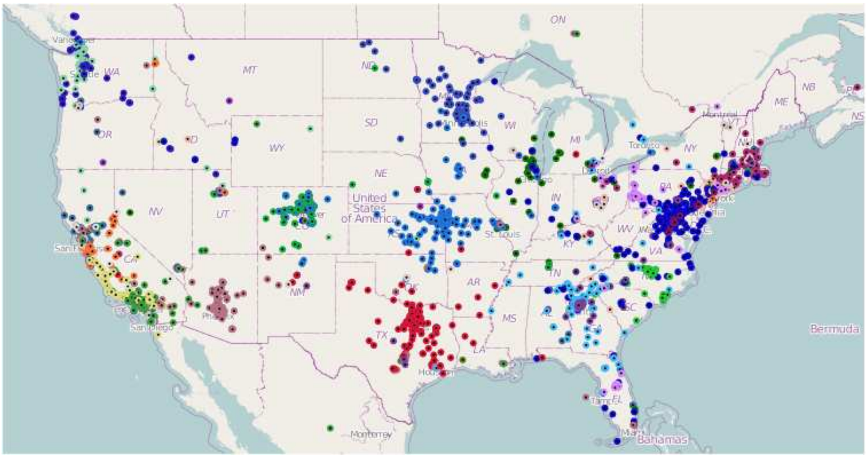

Fig. 2: Major communities detected in the check-ins data reveal catchement areas of large cities.

3.1 Infomap and Oslom

Of the numerous community detection methods available we chose a combination of Infomap [16] and OSLOM [17]. Previous methods of community detection [15] have focused on optimising the modularity, an approach which although useful, has some serious issues. In contrast, Infomap takes a different approach by looking for probability flows of random walks over the links of the network- the communities it finds are groups of nodes among which information flows easily. This method is ideally suited to studying the community structure of mobility networks [9], and it deals naturally with directed and weighted networks.

Recently, Fortunato et al [17] introduced a new technique for community detection called OSLOM. This method optimises the local statistical significance of communities. OSLOM has many advantages over traditional modularity optimisation techniques, including a firm basis for assessing the statistical significance of communities and the ability to detect overlapping communities, which Infomap does not. The OSLOM method is well suited to dynamic networks, and we use it in a refinement step after first finding the communities using the Infomap method. The degree of refinement is quantified by the Normalised Mutual Information [18] (NMI)- a measure of the similarity of community structure which can account for overlapping communities. As we refine the communities using OSLOM we see a decrease in the NMI indicating the stratification in the network as the communities successively merge and overlap with each other.

3.2 Network of significant locations

We built a weighted network of locations by placing an edge between locations i and j if there exist users that visited both places, and assigning it the following weight:

$$w_{ij} = \frac{N_{i \to j}}{N_j} + \frac{N_{j \to i}}{N_i}, \tag{1}$$

where N_i, N_j is a total number of users that visited i and j and $N_{i \to j}$ and $N_{j \to i}$ is a number of those users from i that also visited j and vice versa. Various factors, including geographical proximity of locations, segregation and social influence may impact the community structure in this network leading to possibly overlapping communities. This assumption justified our choice of community detection methods.

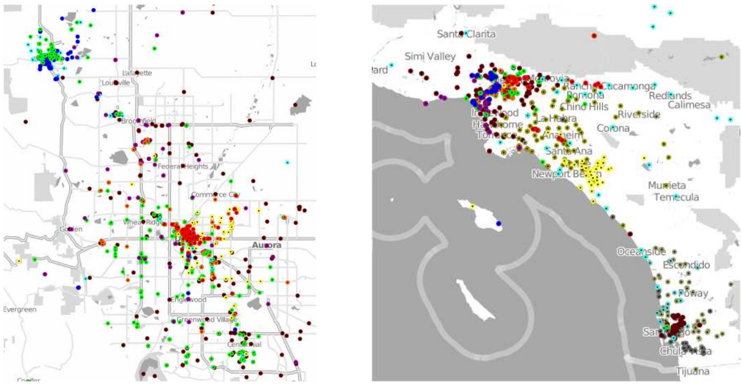

Fig. 3: Geographical spread of the 10 largest communities of locations in Denver (incl. Boulder) and Los Angeles (incl. San Diego) areas.

4 Experimental results

The largest 30 communities of US detected by Infomap in the full network of 772,966 unique locations are shown in Figure 2. They reveal an expected pattern of catchment areas formed by large cities. This pattern is relatively well explained by spatial interaction theories [1–4]. However, the patterns of interaction within cities are explored less.

4.1 From inter-city to intra-city patterns: evidence of stratification

We applied the methods to the network of locations at several metropolitan regions. We observed communities (which we hereby call “habitats”) at various surburbs, many of which also include locations within city centre areas. Figure 3 presents an example of the geographical spread of identified habitats (shown in different colors) around Denver and Los Angeles. A reasonable question to ask is whether inhabitants of different surburbs prefer visiting the same or distinct locations in the city centre. This exploration requires a community detection method which uncovers overlapping communities.

We refined the derived partitioning by applying the OSLOM method and found that communities do overlap at several distinct locations. Figure 4 presents such an example for Los Angeles, highlighting the overlap regions with a spatial kernel density estimate of the locations belonging to more than one community. Closer exploration showed that they typically feature major attractions, shopping malls, and some nightlife areas. We then studied if the level of overlap varies in different localities by computing the normalized mutual information.

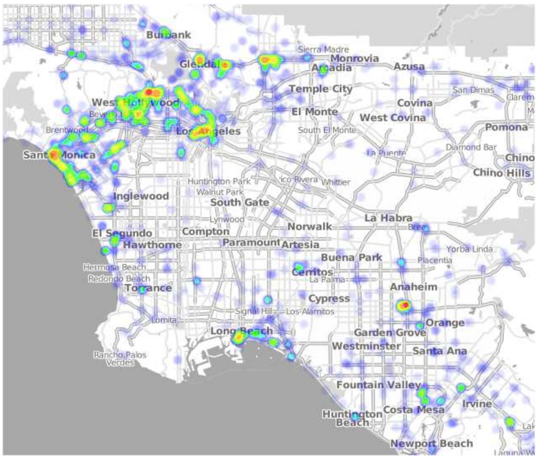

Fig. 4: Spatial density of check-in locations shared by all of the communties in LA feature major attractions, shopping malls and some nightlife areas.

4.2 Quantifying and understanding stratification

Figure 5 shows the NMI (normalised mutual information) [18] for successive refinement runs of the OSLOM method. We start by finding the community structure at t=0 using Infomap. These are non-overlapping and represent the hard community structure present in the co-location network. As we apply the OSLOM method to the communities found by Infomap, it refines the node membership based on their statistical significance and also finds nodes which belong to multiple communities. We then measure the NMI (normalised mutual information) relative to the original community structure. After some small number of runs the NMI settles down to its final value and no more stratification takes place. It is evident that different cities have varying degrees of stratification. Further research is required to understand if cultural, racial, or socio-economic segregation can be the reason for this.

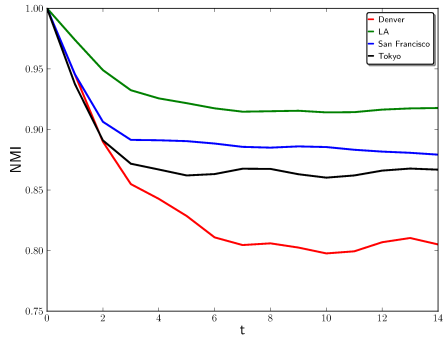

Fig. 5: Normalised mutual information plot for successive refinement runs of the OSLOM method shows different stratification levels for selected cities.

5 Discussion and Conclusions

Spatial contingency of catchment areas has been previously observed at inter-regional levels and laid foundations to a variety of spatial interaction models. An analysis undertaken in this paper has revealed novel structures at the shortest intra-city scales available from location-based social network data analysis. A distinct pattern of stratification of the communities of frequently visited locations into so-called “urban habitats” was observed. We used mutual information to characterize the degree of stratification [18]. The presense of stratification is most likely related to the social influence processes and tend to be closely related to the structure of the social network of the visitors as we observed in our preliminary results.

This approach can be used by urban planners and sociologists to explore and quantify segregation in modern cities given that sampling biases concerned with typical profiles of the users of social network services are accounted for. Our results also highlight a possible direction to refine spatial interaction and urban mobility models to better account for social infuence and segregation patterns, particularly at short spatial scales of intra-city interactions.

Acknowledgments. Research presented in this paper was funded in part by Science Foundation Ireland through Stokes Lectureship, Strategic Research Cluster grant 07/SRC/I1168 and 11/RFP.1/CMS/3247 award, Complexity-NET programme Complexity in Spatial Dynamics (COSMIC) supported by IRCSET, and IBM PhD Fellowship programme. The authors gratefully acknowledge this support.

We would also like to thank Fergal Walsh and Felix Kling for their help with software and Joerg Reichardt and Dirk Helbing for fruitful discussions.

References

1. Zipf, G.K.: The p1 P2/D hypothesis: On the intercity movement of persons. American Sociological Review **11**(6) (1946)
2. Stouffer, S.A.: Intervening opportunities: A theory relating mobility and distance. American Sociological Review **5**(6) (1940) 845–867
3. Fotheringham, A.S.: A new set of spatial-interaction models: the theory of competing destinations. **1**(15) (1983) 15 – 36
4. Simini, F., Gonzalez, M.C., Maritan, A., Barabasi, A.L.: A universal model for mobility and migration patterns. Nature advance online publication (2012)
5. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968) (February 2010) 1018–1021
6. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C.: A tale of many cities: universal patterns in human urban mobility. arXiv:1108.5355v4 [physics.soc-ph] (2011)
7. Sadilek, A., Kautz, H., Bigham, J.P.: Finding your friends and following them to where you are. In: Proceedings of the fifth ACM international conference on Web search and data mining. WSDM '12, New York, NY, USA, ACM (February 2012) 723–732
8. Pozdnoukhov, A., Kaiser, C.: Space-time dynamics of topics in streaming text. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks. LBSN '11, New York, NY, USA, ACM (2011) 8:1–8:8
9. Bagrow, J.P., Lin, Y.R.: Spatiotemporal features of human mobility. (February 2012)
10. Axhausen, K.W.: Social networks, mobility biographies, and travel: survey challenges. Environment and Planning B: Planning and Design **35** (2008) 981–996
11. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. KDD '11, New York, NY, USA, ACM (2011) 1082–1090
12. Expert, P., Evans, T.S., Blondel, V.D., Lambiotte, R.: Uncovering space-independent communities in spatial networks. Proceedings of the National Academy of Sciences **108**(19) (May 2011) 7663–7668
13. Walsh, F., Pozdnoukhov, A.: Spatial structure and dynamics of urban communities (June 2011) The First Workshop on Pervasive Urban Applications (PURBA).
14. Fortunato, S.: Community detection in graphs. Physics Reports **486** (January 2010) 75–174
15. Newman, M.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences **103**(23) (2006) 8577
16. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences **105**(4) (2008) 1118–1123
17. Lancichinetti, A., Radicchi, F., Ramasco, J.J., Fortunato, S.: Finding statistically significant communities in networks. PLoS ONE **6**(4) (April 2011) e18961+
18. Esquivel, A.V., Rosvall, M.: Comparing network covers using mutual information. arXiv:1202.0425v1 [math-ph] (2012)