

Spatial structure and dynamics of urban communities

Fergal Walsh and Alexei Pozdnoukhov

National Centre for Geocomputation
National University of Ireland Maynooth
{fergal.walsh, alexei.pozdnoukhov}@nuim.ie
<http://ncg.nuim.ie>

Abstract. This paper explores the evolution and spatial organization of urban communities in a daily cycle of a city using scalable methods of social network analysis. Using mobile phone network records, we have observed spatial patterns in community structures at fine spatial scales which are most likely related to the activities typical to the different functional regions of a city. It is known that interactions between different parts of the city are subject to a distance decay, moreover, our findings show that regardless of the time of the day or current activity, people tend to communicate within a spatially enclosed community. Furthermore, we have investigated transitions and transformations of these communities at the temporal scales of several hours related to the cycles of human movement and activities in a city.

Keywords: community detection, complex networks, smart city, mobile phone data

1 Introduction

There is an established tradition in geographical and regional sciences to study communication data as a reflection of spatial interaction processes between different regions. Landline phone calls were found to be an indicator of inter-region economic ties' strength. Novel communication technologies produce datasets that add new dimensions to these studies. First, they provide much more detailed data and hence increase the spatial resolution of the analysis, and second, they add a temporal element to the study of the observed processes. This is the case for geo-referenced mobile phone communication data which we study in fine spatial and temporal details to reveal dynamics of human activities in a typical daily cycle of a city.

We combined two methods adopted from the study of complex networks to examine the time varying spatial structure of communities in a city and its surrounding suburbs. The spatial structure of these communities was observed from the analysis of the short-term aggregates of the traffic on a mobile telephone network. In a similar manner to [1], the analysis was based on an origin destination

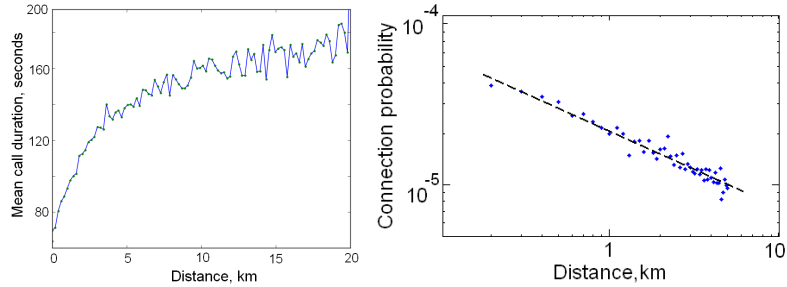


Fig. 1: Left: an average length of the calls as a function of call distance. Right: probability of users to be connected decays with distance as $d^{-0.44}$ (dashed line in log-log scale).

matrix of calls and text messages between customers. There are a number of important differences in the present approach however.

In our case we use the cells of the network as the spatial areas rather than administrative boundaries of any kind. Cells in urban areas are typically very small due to high populations and thus offer a much higher spatial resolution than the available administrative boundaries. Rather than using the customer's home address, each end of the call was geolocated to the cell used at the time of the call. The mobile phone network is by definition dynamic with phone users constantly moving through space and available to place or receive calls at any time, anywhere. We use time-of-call locations in an attempt to capture these dynamics. Finally, and most importantly, we perform community detection at multiple snapshots in time and then analyse the change in community composition.

We hypothesise that different communities will be found at different times of day, especially in urban areas where the population and communication habits vary so much with the time of day. These variations capture some fundamental properties of human behavior in terms of communication habits which give an indication to optimise resource allocation strategies and refine planning of the city infrastructures.

2 Mobile Communication Data

In this paper we analyse a dataset derived from the activity on a mobile phone network. The dataset consists of records of all calls and text messages sent and received by the customers of an Irish mobile phone network operator over the course of a week. The network is a GSM network with about 3700 cells (spatial units) of varying size across the country. There are over 1 million customers and an average of 40 million communications per day. Each Call Detail Record (CDR) contains the anonymised IDs of the caller and the callee, the date and time, type of event, duration, and the ID of the cell the customer's phone was connected to at that time. If the call was between two customers of the operator, there will be two records in the dataset - one for the caller and one for the callee.

The caller and callee records in dataset were matched by customer IDs and call times to create a single record which contains the call time and the origin and destination cell IDs.

A general statistical analysis of the network has shown typical properties for communication networks including the power-law node degree distributions. In spatial analysis we have concentrated at shortest scale available. First, the average duration of the calls grows with distance from around 60 seconds (Figure 1, left). A plateau is reached at around 250km with a call length of 6 minutes. Second, at city centre scales of 100 meters to 5 kilometers a probability for 2 users to be connected follows a power-law distance decay with an exponent of $d^{-0.44}$, consistent throughout the day. Regardless of the time of the day or current activity (work, leisure) people tend to communicate with spatially proximate counterparts. To investigate this finding, we created an origin-destination matrix by averaging the counts of calls for each hour of a typical weekday and carried out a detailed community structure analysis.

3 Methods

3.1 Community Detection

The identification of densely connected groups of nodes in a network is known as community detection. It involves the division of a network into groups such that nodes that communicate with each other the most are placed in the same group. The quality of the division can be measured using Newman’s modularity metric [5], which compares the density of links within groups to the density of links between groups. The ideal community detection method is one that maximises the modularity of the partitioning. However, this is computationally impossible for large networks so several algorithms exist which attempt to achieve good modularity in a computationally feasible manner.

One such algorithm is the Louvain method [2] which is our choice for this paper. This is a hierarchical grouping method which starts with each node as a separate group and then successively forms larger groups. At each iteration the method considers each of the neighbours of each node and assesses the possible gain in modularity achieved by moving that node into its neighbour’s group. If a positive gain is possible, the node is moved to the group that will result in the highest gain. When all nodes have been considered in this way, a new network is formed where the groups become the nodes. This two step process is then repeated on the new network until there is no change between successive iterations. By making use of heuristics this algorithm can scale to millions of nodes.

3.2 Tracking Change in Communities

When studying the change in community structure over time it is necessary to perform community detection on snapshots of the network at different points in

time. It must be possible to relate the communities found at each step to those found at previous and subsequent steps. Greene et al.[4] introduce a method which discovers dynamic communities by matching the static communities found at each time step. Dynamic communities are ones which can be found at multiple time steps. This matching is done by comparing the membership of each static community at each time step to the dynamic communities found previously. If the membership is similar within a certain threshold then the static community is assumed to be a continuation of that dynamic community. If it is not similar a new dynamic community is created. When dynamic communities merge together they continue to exist as separate communities with the same membership in subsequent time steps. When a dynamic community splits into two or more parts, it may result in the creation of new dynamic communities or it may cause the reemergence of previously seen communities. This can happen when two communities merge together and then split again after a few time steps. The method however includes a parameter which determines how long an unobserved dynamic community should be considered before it is marked dead. If the splitting occurs after this time then a new dynamic community is created instead of the resurrection of the original one.

3.3 Visualising Change in Communities

It is easy to visualise the communities detected in spatial data at a single time step on a map where we colour each area by its assigned community. From this we can visually assess the quality of the communities and interpret the results. However, it is more difficult to visualise the changes in community membership over time. Series of maps can be useful but it is also necessary to be able to see the full time line in a single visualisation. Rosvall and Bergstrom [7] introduce the concept of an alluvial diagram for visualising the changes in cluster membership over time. Here we adopt that concept and construct a similar diagram. In this diagram each node (cell) is represented by a single line. At each time step the lines are ordered vertically and coloured by the dynamic cluster to which the corresponding node belongs. By viewing this diagram side by side with the time step maps we can better understand the changes that take place.

4 Experiments and Discussion

For the purposes of this study we concentrate on the Dublin region. This area has the highest concentration of customers and therefore has a large number of small cells (about 1000). Dublin is the capital city and also the financial, commercial, services and education centre of Ireland. However, 54% [3] of the working population of Dublin city commute from outside the city, mostly from the suburbs and neighbouring counties. We expect, therefore, to find changes in the spatial structure of the detected communities at different times of the day.

Initially we run the Louvain method over a single weighted network for the entire day. The clustering of the full 24 hour network results in few large spatially contiguous clusters. However, the modularity is low at 0.2, meaning that

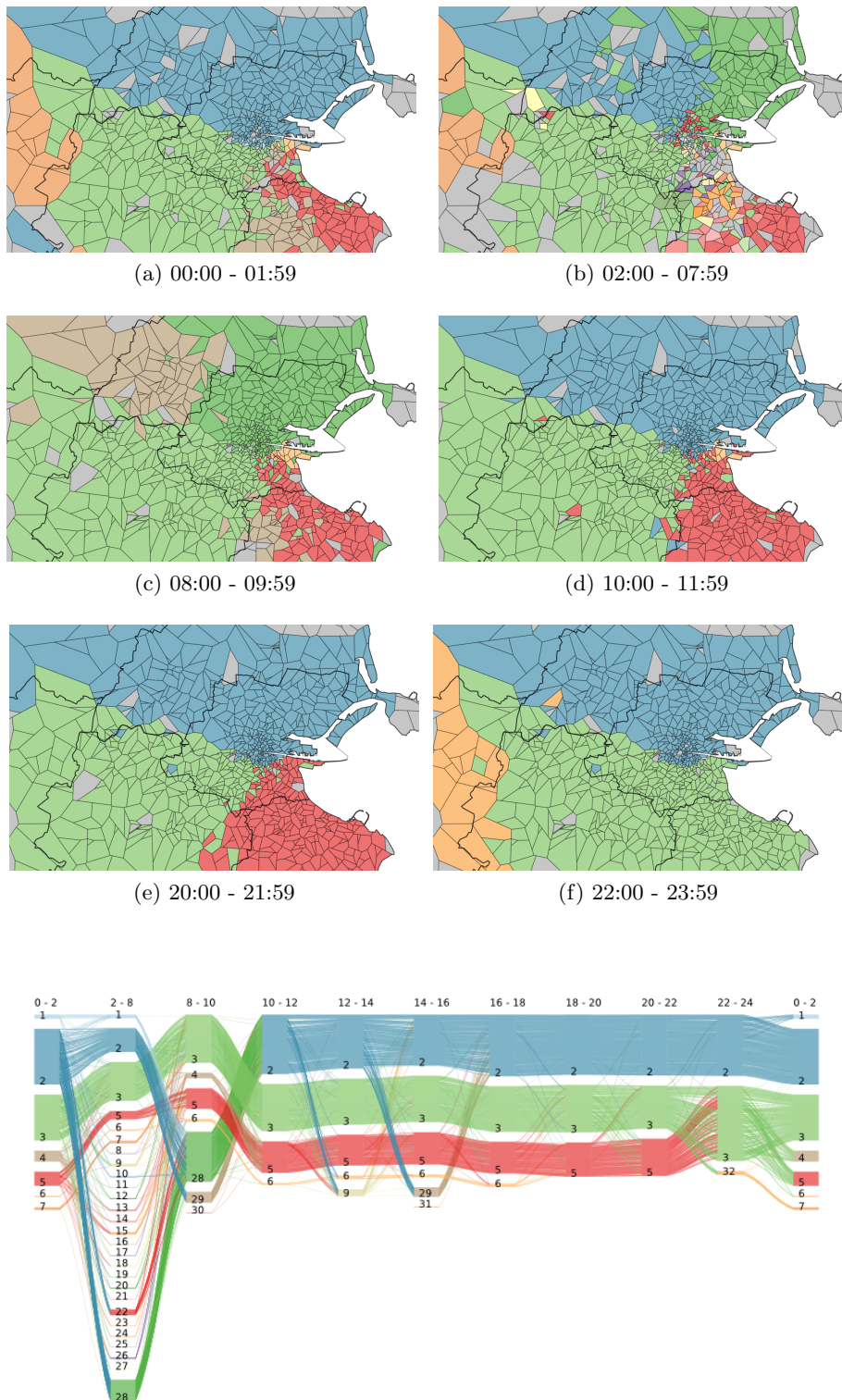
the result does not represent a good division of the network into communities. We argue that good quality communities can not be found in this network because the full day aggregation of weights creates strong links between spatially disconnected areas, resulting in a network configuration that is an average, but never actually occurs in time.

The movement of people and changing calling patterns throughout the day suggest that there should be a change in the community structure. Our main experiment attempts to discover and quantify this change. We performed community detection on each of the 24 origin-destination matrices and applied the tracking algorithm to the results. With these network snapshots we had the opposite problem to the one described above for the full day network. Due to the limited amount of data available to us (5 weekdays), the aggregated matrices at one hour resolution are quite sparse and have low weights. The resulting communities were small and spatially discontinuous. We found that it was necessary to aggregate the matrices into longer time periods to achieve good results - as measured by the modularity and a visual inspection of both the spatial contiguity and temporal continuity of communities. The temporal continuity was assessed by applying the tracking algorithm to the results to determine matching communities in adjacent time steps and then visualising the resulting community membership transitions as an alluvial diagram.

Through a visual analysis of the temporal profile of activity we determined possible groupings of the hourly matrices that would increase quality without masking the effects of the temporal phenomena which we wished to discover. We decided to perform community detection on the aggregate networks of nine two hour periods between 8am and 2am (modularity of 0.65 - 0.67) and one period between 2am and 8am (modularity of 0.49). Smaller aggregations between 2am and 7am resulted in meaningless results, as evidenced by high numbers of spatially discontinuous clusters and many disconnected cells. However when aggregated, there are clear clusters to be found during these time periods.

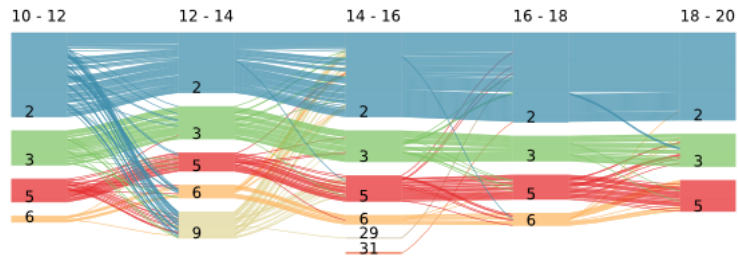
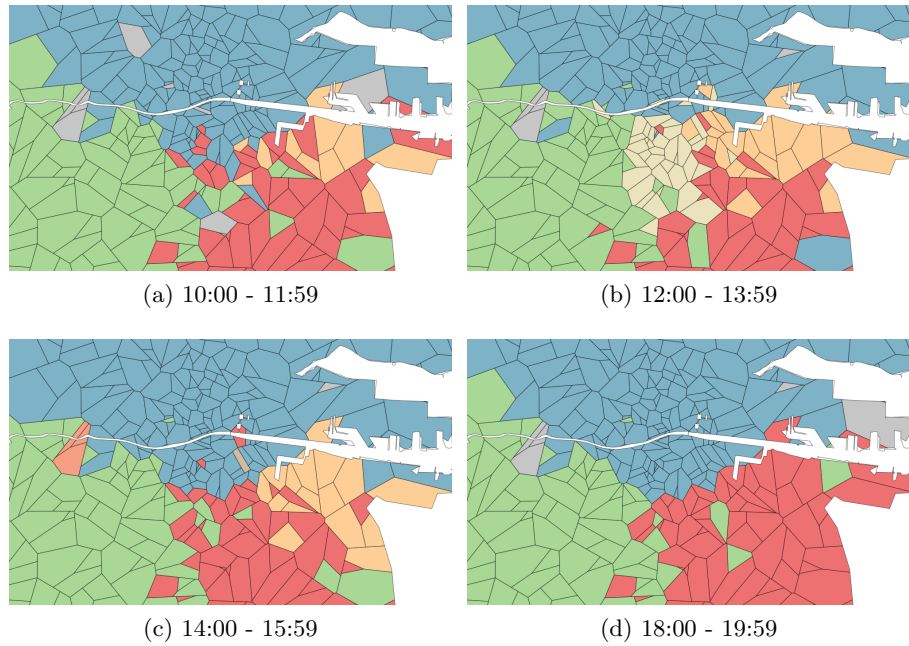
4.1 Discussion of Results

From the sequence of maps in Figure 2 we can see the evolution of the spatial structure of communities in the daily cycle of the Dublin region. Most of the communities found are nearly spatially contiguous. The discontinuities that are present are likely due to the inaccuracy in the cell tessellation and/or low levels of activity in those cells. The most striking result here is the clear north south divide that is evident in all the maps. The River Liffey creates a physical division, flowing west to east through Dublin. It is well known that there is a social divide between the north and south and now we see a clear communication divide with only a small buffer zone. We can also see that the towns to the far west of the city form their own community during the night and early morning but during the day time they merge with the other north western and south western suburbs of the city. From Figure 3 we see that there are also smaller scale changes during the working day.



(g) Alluvial timeline diagram of community membership. Each line represents a single cell. Each column represents a time period, as labeled.

Fig. 2: The daily cycle of community structure in the Dublin region.



(e) Alluvial timeline diagram of community membership. Each line represents a single cell. Each column represents a time period, as labeled.

Fig. 3: Community structure changes in the city center area during the working day

5 Conclusion

Spatial cohesiveness in communication communities has previously been observed at inter-regional levels [6]. The analysis undertaken in this paper has revealed community structure at the shortest spatial resolutions available from mobile phone network data analysis. A distinct pattern of transitions and transformations of the communities at the temporal scales of several hours and sub-

kilometer distances was observed. It is most likely related to the cycles in professional and social activities of the citizens, showing the tendency to communicate locally within professional and social groups localized at particular city districts in the course of the day.

It is likely that the results of this study could be more interesting if a longer period of data was available. Unfortunately we could not compare typical weekdays with weekends or holidays with the limited amount of data available to us at this time.

This analysis involved a combination of methods adopted from complex networks analysis applied to a dataset of mobile phone activity. There are two evident directions in which the methods can be improved. Firstly, community tracking methods need to be adapted to deal with the periodicity of human behaviour. Secondly, we have observed multiple scales at which spatial regularities emerge and further research is required to understand the hierarchy and semantics of these nested and overlapping communities, both in terms of methodologies and visualisation approaches.

Acknowledgments. Research presented in this paper was funded in part through Stokes Lectureship programme and Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland, and faculty research awards from Google and IBM. The authors gratefully acknowledge this support. We would also like to gratefully acknowledge Dr. R. Farrell and the support of Meteor for providing the data used in this paper, in particular Helene Graham, John Bathe and Adrian Whitwham.

References

1. Blondel, V., Krings, G., Thomas, I.: Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. *Brussels Studies* 42 (2010)
2. Blondel, V., Guillaume, J., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, P10008 (2008)
3. CSO: Census 2006 Place of Work - Census of Anonymised Records (POWCAR). http://www.cso.ie/census/POWCAR_2006.htm (2006), http://www.cso.ie/census/POWCAR_2006.htm
4. Greene, D., Doyle, D., Cunningham, P.: Tracking the evolution of communities in dynamic social networks. In: *Proc. International Conference on Advances in Social Networks Analysis and Mining (ASONAM'10)* (2010)
5. Newman, M.: Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103(23), 8577 (2006)
6. Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., Strogatz, S.H.: Redrawing the map of great britain from a network of human interactions. *PLoS ONE* 5(12), e14248 (12 2010), <http://dx.doi.org/10.1371/journal.pone.0014248>
7. Rosvall, M., Bergstrom, C.: Mapping change in large networks. *PLoS One* 5(1), e8694 (2010)