The role of space in social groups: Analysis and technological applications

Chloë Brown

Jesus College

A thesis submitted in June 2014 for the degree of Doctor of Philosophy

The role of space in social groups:
Analysis and technological applications

Chloë Brown

Summary

Space plays an important role in social networks, with pairs of geographically close individuals being more likely to have a connection than those far apart. However, lack of available data has meant that larger social groups have remained less well-understood. The availability of information about users' fine-grained location from increasingly geographically-aware online social services, and from the use of advanced sensing technology, now facilitates the investigation of the role of space for groups in social networks at various scales. The thesis of this dissertation is that the relationship between physical space and social groups at different scales can be exploited to create or improve technological social applications. I begin by showing that in some online social networks, communities still tend to be geographically close despite the ease of forming long-distance online connections. In others, while geography may play a less key role, there are still communities of friends in the online network who visit the same real-world places. I demonstrate how such place-based groups could be useful to location-based online social networks for improving friend recommendation. I next study groups at the scale of a city and show that there is clustering around places in the social network, and that some places around the city are far more likely to host meetings between friends than others. I compare individual and group mobility as observed in the datasets and find differences between the places people visit on their own and with their friends, which can be applied to group venue recommendation in location-based online social networks. I finally examine the role of space for groups at the scale of a single building by studying location and face-to-face interaction traces of people in workplace buildings, collected using ubiquitous sensing technology. I compare the recorded interactions within and between groups of employees in two different buildings, and assess the impact of space on intra- and inter-group communication. These studies demonstrate that spatial factors may be related to the interactions of groups of employees in workplaces, but furthermore that the automatic sensing methods used could be extended to provide applications for monitoring and maintaining beneficial levels of communication within and between workplace groups in the future.

Declaration

This thesis:

- is my own work and contains nothing which is the outcome of work done in collaboration with others, except where specified in the text;
- is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other university; and
- does not exceed the prescribed limit of 60,000 words.

Chloë Brown
June 2014

Acknowledgements

First, I must thank my supervisor, Cecilia Mascolo, for her advice, expertise, guidance, and considerable patience, during my time as a Ph.D. student. Without her encouragement, I would most likely not have begun the work leading to this dissertation, let alone finished it.

I am grateful to those with whom I have co-authored papers over the last three years: Vincent Blondel, Christos Efstratiou, Desislava Hristova, Peter Key, Neal Lathia, Ilias Leontiadis, Mirco Musolesi, Vincenzo Nicosia, Anastasios Noulas, Rosica Pachilova, Daniele Quercia, Kerstin Sailer, Salvatore Scellato, and James Scott, and to Richard Gibbens and Pietro Liò for their guidance during the progress of my research. Many thanks, too, to my other colleagues and friends, in particular, at the University of Cambridge: Jisun An, Andrius Aucinas, Petko Georgiev, Theus Hossmann, Dmytro Karamshuk, Nadia Llanwarne, Sarfraz Nawaz, Kiran Rachuri, Haris Rotsos, Sandra Servia Rodríguez, John Tang, and Amy Zhang, and at the Université Catholique de Louvain: Etienne Huens, and Corentin Vande Kerchove.

Thanks as well to Google, for generously awarding the Google Europe Fellowship in Mobile Computing to fund my Ph.D. studies.

Finally, for their unwavering belief in me and support of my endeavours, I thank my family: Alan, Gordon, Jenny, Jeoffry, Lizzie, and Tim, and two incredibly special fellow Ph.D. students: Karoliina Lehtinen, at the University of Edinburgh, and Natasha Morrison, at the University of Oxford. I hope I can help these last two in the pursuit of their own Ph.D.s as much as they have helped me in the pursuit of mine, should they require it.

My heartfelt thanks to you all.

Contents

Chapter 1

Introduction

Geographic distance and physical space have traditionally played an important role in human social groups within countries, within cities, and at the level of single buildings such as homes and workplaces. For decades sociologists have observed that spatial proximity greatly increases the probability that two individuals will be friends [Ste41]. The advent of the Internet prompted speculation that distance could cease to influence social relationships so strongly, as people became able to communicate with greater ease with others from all around the world [Cai01]. However, this has not been the case and individuals continue to communicate mostly with others located physically close to them, whether this communication is face-to-face, by telephone, or by e-mail [MWC10]. The enormous popularity of online social networking services in recent years has provided another means of online communication, and thus another opportunity for geographically remote people to connect with one another. Nevertheless, analyses of these services have shown that the probability of a connection in many online social networks also decreases with increasing geographic separation [BSM10, KKNG12, LNNK$^+$05, SMML10]. Although strangers with shared interests may form online friendships despite having never met, people currently tend to use online social networks to keep in contact with people they know in the offline world, with whom they meet in person [GWT11]. At a much smaller scale, physical proximity has also been shown to play an important role in who communicates with whom within buildings such as offices, university departments, and hospitals, so that the spatial layout of working environments is important to facilitate effective contact and communication between their occupants [SBL$^+$09, AH06].

Closely connected groups of friends, or *communities*, in online social networks have attracted much research interest and been the subject of some study [BHKL06, KNT10, MMG+07]. This is partly due to such groups having many potential applications in technological services and systems, such as obtaining coarse-grained visual representations of large networks, sorting personal online contacts into manageable groups, finding partitions to speed up system performance, and providing recommendations [LVM+12, PKVS11, PES+10]. Very recently, interest has turned to the spatial properties of social groups. For example, Onnela et al. [OAG+11] analysed the geographic properties of communities in a mobile phone communication network, and found that small groups tend to be spatially tight, with dispersion increasing as groups become larger. Expert et al. [EEBL11] noted how spatial proximity can foster social interactions, and presented a technique specifically for extracting space-independent communities from networks, by accounting for this effect during community detection. Such research into the spatial properties of groups has only now become widely possible on a large scale, due to a previous general lack of access to suitable data concerning people's physical whereabouts as well as their social connections. For this reason, little has been known about the role of space in social groups, and whether this has been affected by the introduction of new social networking technologies, or, as in the case of single connections between just two friends, has remained largely unchanged.

Mobile Internet access is growing rapidly, and combined with the all-pervasive 'socialisation' of online services, location is being widely incorporated into many online social networks. People use mobile phone applications to 'check in' (indicate their physical whereabouts and share this information with their online friends), as in Foursquare and Facebook [Fac12, The12]. Online content is increasingly tagged with an associated geographic location, from tweets in Twitter, to photographs in the image-sharing services Flickr and Instagram [Fli12, Ins12, Twi09]. The available data about the geography of these networks could be used fruitfully in applications such as providing personalised recommendations, improving mobile search results, or in targeted advertising [CML11, RCE08, dLM13]. Given the wide variety of technological uses of groups in online social networks, it seems promising that these groups should also be useful in conjunction with location data, in order to create and improve social technological applications.

There is also benefit to be gained from studying the interactions of groups at a smaller scale than the resolution that is provided by such online location data, that is, studying the behaviour of groups of people within single buildings, such as workplaces. For example, in many knowledge-based work environments, where creativity and innovation are key, it is intuitive that interactions between members of different groups or teams, with complementary expertise or skill sets, can be highly beneficial as a source of fresh perspectives, information, and ideas. Conversations between individuals who are not necessarily part of the same group have long been judged to be essential for team coordination, cohesiveness and productivity [Bur04, ITM96, JM00, WFDJ94]. One factor that could clearly affect the ease of such interactions is the physical spaces of the workplace itself; for example, high-traffic areas such as coffee machines and photocopiers may be particularly likely places for inter-group meetings [ITM96]. In general, if spaces encourage the mixing and meeting of a diverse range of people, meetings between individuals from different teams or social groups will occur more readily, which could be crucial; indeed, face-to-face communication has been shown to be more important than electronic means such as email or SMS [Pen12, SS12].

The study of the relationship between social groups and space at the building scale has also been hindered by difficulty in acquiring suitable data, due in this case to problems observing and recording people's behaviour accurately and reliably, and recent technological developments have also helped to make it possible to overcome these challenges, in this case, improvements in ubiquitous sensing technology. Studies of workplace communication have traditionally used pen-and-paper methods from the social sciences, with data being gathered through direct manual observations and participant surveys. These approaches have various disadvantages and limitations: the presence of an observer may cause people to reflect upon and change their natural behaviour [Why43], and surveys can suffer from participants giving answers they feel are socially desirable, or misremembering [BSBS78, vdM08]. Recently, developments in ubiquitous sensing technology have produced devices such as wearable badges and sensors, enabling social interaction patterns to be studied in a less obtrusive way [OOWK$^+$09]. There is now the means to conduct experiments using technology to monitor the interactions of groups in buildings such as workplaces, using sensors to detect both contact between employees and the spaces in the workplace where these contacts take place. Beyond such analysis, these advances in ubiquitous sensing also

provide scope for new applications to improve workplace communication: one can imagine making sensed data available to employees and their managers in order to allow them to assess and optimise their own interactions, as well as providing insight into which spaces in the building promote beneficial inter-group meetings.

1.1 Thesis and research contributions

The thesis explored in this dissertation is that *the relationship between physical space and social groups at different scales can be exploited to create or improve technological social applications*. In order to examine this thesis, it is necessary to analyse the use of space by groups at different scales, and to consider how this might be used by such technological applications in each case. My research makes two main contributions:

1. To examine and model the use of geographical space in social groups at different scales, through analysis of the newly available abundance of data concerning people's social connections alongside their physical whereabouts: online location-based social networks (LBSNs) covering global and intra-city scales, and interaction traces collected using modern ubiquitous sensing technology at the scale of single buildings, and

2. To show how the use of space by groups can be used in social technological applications, namely, friend recommendation in online social networks, group venue recommendation in location-based networks, and the potential improvement of workplace communication between employees through the sensing and presentation of face-to-face interactions.

More specifically, the rest of this dissertation is structured as follows. In **Chapter 2**, I outline the existing work in this area and describe how my research relates to and builds upon this.

In **Chapter 3**, I begin by analysing social groups at a potentially global scale, as made possible by online social networks allowing online connections to be established with ease between geographically remote individuals. I define measures for quantifying the extent to which members of an online group visit the same places in the physical world, and show that in some online social networks, groups are very close geographically, but in others space plays a less important role. I then outline why, in

these cases, it may be desirable to uncover location-based groups that do exist in the network, but which are difficult to find using standard community detection techniques, and present a method for performing this task. I finally show how these location-based groups could aid the application of friend recommendation commonly seen in online social networks.

In **Chapter 4**, I consider groups at the scale of single cities, and analyse the kinds of places in the city that friends have in common. I show that city-scale social networks show some clustering of social ties around places, and I examine the kinds of places where people tend to meet with their friends. I find that the different kinds of places in the city may be important for the structure of the city-level social network, including the presence of clustering and communities. I then examine how knowledge of the different kinds of places where people tend to meet with groups of their friends, as opposed to the places where people go by themselves, may be exploited to improve an important application in location-based online social networks, namely that of venue recommendation.

In **Chapter 5**, I study groups at the small scale of an individual building. I analyse the social networks and face-to-face interactions of groups of people in a workplace, and examine the impact of the building space on these interactions, by comparing the behaviour of the same group of people when they were working in two different buildings. I find that different kinds of spaces could be important for different kinds of communication within and between groups. I discuss how these results could be used in conjunction with ongoing improvements in ubiquitous sensing technology to help managers monitor the 'communication health' of their organisations, and to help employees to manage their interactions and potentially enhance innovation and productivity.

In **Chapter 6**, I review the contributions of my research and draw conclusions, as well as outlining how future research may build on the work I have described in this dissertation.

1.2 List of publications

Some of the research related to this thesis has been published in various peer-reviewed journals and conferences. These publications are as follows:

Chapter 3:

- [BNS+12b]: Where Online Friends Meet: Social Communities in Location-based Networks. Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM), Dublin, Ireland, June 2012.

- [BNS+12a]: The Importance of Being Placefriends: Discovering Location-focused Online Communities. Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. In ACM SIGCOMM Workshop on Online Social Networks (WOSN), Helsinki, Finland, August 2012.

- [BNS+13]: Social and place-focused communities in location-based online social networks. Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. In European Physical Journal B 86 (6), 290 (2013).

Chapter 4:

- [BNMB13]: A Place-focused Model for Social Networks in Cities. Chloë Brown, Anastasios Noulas, Cecilia Mascolo, and Vincent Blondel. In Proceedings of the IEEE/ASE International Conference on Social Computing (SocialCom), Washington DC, USA, September 2013.

- [BLM+14]: Group colocation behavior in technological social networks. Chloë Brown, Neal Lathia, Anastasios Noulas, Cecilia Mascolo, and Vincent Blondel. In PLOS One, 9(8):e105816, August 2014.

Chapter 5:

- [BEL+14a]: Tracking serendipitous interactions: How individual cultures shape the office. Chloë Brown, Christos Efstratiou, Ilias Leontiadis, Daniele Quercia, and Cecilia Mascolo. In Proceedings of the 17^{th} ACM Conference on Computer Supported Co-operative Work (CSCW), Baltimore, Maryland, USA, February 2014.

- [BEL+14b]: The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies. Chloë Brown, Christos Efstratiou, Ilias Leontiadis, Daniele Quercia, Cecilia Mascolo, James Scott, and Peter Key. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp). Seattle, Washington, USA, September 2014.

I acknowledge with gratitude the help and advice of my co-authors in producing the papers listed above, and point out that while I made use of their expertise, and suggestions, the results and technical contributions outlined in this dissertation are my own work, conducted towards the exploration of the thesis in the following chapters.

Chapter 2

Related work

Social groups have been studied for a long time in the social sciences [BMBL09], but it has traditionally been difficult to conduct large-scale studies of the geography of communities in social networks, due to the cost and effort involved in obtaining suitable data concerning both the social connections of individuals and their geographic locations. While it has recently become feasible, largely thanks to mobile phone data [OAG+11], the prolonged existence of this constraint has limited the possibility of studies of the spatial properties of social groups and, therefore, of potential technological applications of these groups and their geography. The relationship between the physical space provided by a single building and the groups that occupy that space has also been a subject of research interest, largely in the field of architecture [SP09]. However, these studies have also been affected by problems with data acquisition, in this case caused by the high cost and inaccuracies involved in reliance on human observers to record interactions between people and the locations of those interactions. Recently, advances in ubiquitous sensing technology have begun to make it possible to overcome these difficulties [OOWK+09], simultaneously creating the potential to use these sensing devices as the basis on which to build applications based around measuring and improving group interactions in buildings such as workplaces.

2.1 Social groups at the inter-city scale

2.1.1 Analysis

2.1.1.1 Telecoms data

Most of the previous large-scale studies of the geography of social groups have involved social networks potentially at the inter-city scale; that is, social ties may connect individuals based in different cities. This is unsurprising given that the datasets that lend themselves most easily to this kind of analysis are arguably those from telecoms operators, namely, Call Detail Records (CDRs) from mobile phone networks. A CDR contains, amongst other information, the mobile telephone numbers of the people making and receiving the call, and the identities of the cell towers involved in connecting the call. Therefore such datasets satisfy the requirements of containing information about people's social connections, in that a social network can be inferred from who calls whom, and about physical whereabouts, given that the datasets include the identities of the cell towers involved in making mobile phone calls and can thus be used to estimate the geographic position of the person placing or receiving a call.

One of the earliest studies concerning specifically the geography of mobile phone networks going beyond dyads was performed in 2008 by Lambiotte et al. [LBdK$^+$08]. They studied a dataset of billing records from a mobile phone communication network within Belgium, containing information about phone calls and text messages over a period of 6 months, and, while information about where calls were placed was not available, the authors had access to home localisation information for each user at the zip code level, giving an effective resolution of about 5km. They analysed triangles in the social network, the first generalisation of dyads towards larger groups, and found that shorter links are more likely to belong to triangles than are longer links, which indicates some relationship between clusters in the network and geographic proximity. In more detail, the authors concluded that there are two regimes of communication: between people who are geographically close, calls tend to be short, and there tends to be a reasonably high level of clustering of social network ties close to these pairs in the network topology. Conversely, some connected pairs are separated by large geographic distances. These social ties tend to be characterised by longer calls between the individuals concerned, with lower associated clustering, and more extended triangles than

would be expected given typical models of spatial networks.

The zip codes of mobile phone users were also used in the 2007 study by Palla et al. [PBV07] of the time dependency of k-clique communities at large scale. The study involved a dataset from a mobile phone network concerning the communications of 4 million users over a year, and the authors showed that communities discovered using the network topology tended to contain people living in the same neighbourhoods. This was used as validation of community detection rather than being an objective of the study per se, but it does suggest that clusters in these mobile phone communication networks tend to be geographically localised. Similarly, one of the datasets used in the 2010 study by Ahn et al. [ABL10] is from a mobile phone network. The main focus of the study is the use of link clustering, rather than the more traditional use of nodes, to detect communities in networks. However, in the case of the mobile phone network, the authors find that by plotting the spatial locations of clusters at a broad resolution, the locations of several large cities can be seen to characterise some communities. By examining clusters at a more fine-grained resolution, larger regional communities emerge, and at a still smaller resolution communities within cities can be seen. Again, the study of the geography of communities was not one of the objectives of this work, but the incidental observation about community geography suggests that there could be value in examining the spatial properties of communities more directly.

The first large-scale study specifically concerning the geographic properties of communities in a mobile phone network was done by Onnela et al. [OAG+11] in 2011. The authors studied a social network comprised of 3.4 million mobile phone users in a European country, with localisation at the cell tower level, taking as the location of each user their most frequently recorded geographic position. By running a community detection algorithm to group the users into communities, the authors directly explored the relationship between the topological positions in the network and the geographic positions of community members. The key finding of this study was that small groups were spatially tight, but that there was a disproportionate increase in the geographic span of a community once the group size exceeded about 30.

The common theme emerging from this initial research is that space does indeed determine network structure to a large extent, resulting in the tendency of network communities to be geographically close. In 2011, Expert et al. [EEBL11] proposed a method of mitigating this fact in order to discover space-independent communities in

spatial networks. Using a modified modularity metric to incorporate and effectively attempt to 'neutralise' distance effects, the authors showed that they were able to uncover Flemish- and French-speaking regions in the Belgian mobile phone network, that could not be found when correcting for distance by using the modified metric. For example, Brussels correctly appeared in the French community (80% of people in Brussels speak French), even though it is physically located in the Flemish-speaking region of the country.

2.1.1.2 Online social networks

There have also been a small number of large-scale studies in the past few years of the geography of social groups making use of location data available in online social networks. Again, this has been made possible by the combination of widespread use of online social networking services and of mobile Internet access, so that meaningful location data can be included in content posted to online social networks.

One of the first such studies was that by Scellato et al. [SMML10] in 2010, where the authors analysed datasets containing social connections and locations of users of four online social networks with geographic data: Brightkite, Foursquare, LiveJournal, and Twitter. They defined measures specifically for studying the geographic properties of these networks, and while the focus of the work was not communities, one of these measures, the geographic clustering coefficient, did go beyond dyadic characteristics to measure the tendency of users in network triangles to be located close to one another, similarly to the study of mobile phone users by Lambiotte et al. described above. The results showed that the specifically location-based social networks, Brightkite and Foursquare, had more geographically close triangles than the other two services, which might be due to the focus on location as opposed to content-sharing, which might encourage connections between users with common interests but not necessarily close location. A follow-up study [SNLM11] involved three of these location-based social networks: Brightkite, Foursquare, and a third, similar service, Gowalla. Again examining the socio-spatial properties of triangles in the social networks, the authors found that, conversely to the study by Lambiotte et al., the probability that a link in each of these networks belongs to a triangle appears to be unaffected by the geographic proximity of the connected users, which might show that communities in online social networks are less influenced by geography than those in mobile phone networks.

A 2012 study of Twitter by Takhteyev et al. [TGW12] examined the geography of the microblogging service, investigating the influence of geographic distance, national boundaries, frequency of air travel, and language, on the existence of social ties. The authors observed that many ties lie within the same metropolitan region, with 39% of ties having a geographic length of less than 100km, but that there are also a substantial number of very long ties spanning more than 1000km. This confirms the results of other research that clusters in the network comprise individuals located physically close to one another. By investigating ties between regional clusters, the authors determined that the best predictor of inter-cluster ties was the frequency of airline flights between two regions, suggesting a close link between online social ties and offline connections, but that language differences, distance, and national borders also play a role. While this work also did not examine the geographic properties of social groups directly, they are arguably at least implied by the analysis of the regional clusters.

2.1.2 Applications

Since the analysis of the geography of social groups is necessary before investigation of how it may be used in technological applications, very little work exists along these lines owing to the recency of large-scale analysis becoming feasible. Still, there have been some examples that, while not addressing this question directly, have touched on the area.

In 2010, Backstrom et al. [BSM10] showed how using the user-supplied address data and friendship network in Facebook, an algorithm can be created to predict the location of an individual using the locations of a set of their friends, outperforming IP-based geolocation. While this work was not concerned with social groups directly, it implicitly exploits the idea that at least some of an individual's friends, as a group, are likely to be located near to that individual. Also involving Facebook, in 2010 Wittie et al. [WPD+10] showed how the fact that there tends to be high overlap between the sets of friends of a person making a post and friends of the recipient of a post, and the fact that posts made my local users dominate each regional network, could be used to partition state with geographic awareness to reduce latency and bandwidth consumption in content delivery in the Facebook network.

More recently, and concerning mobile phone data rather than online social net-

works, de Domenico et al. [dLM13] showed that by considering the movements of friends, and other people with highly correlated mobility patterns, it is possible to increase the accuracy of location prediction, and therefore the prediction of where a person will be and whom they will meet. The authors argue that this in turn has potential applications in targeted advertising, dissemination of location-aware content, and mobile search. In this work, the strongest definition of a social tie that the authors had access to was the presence of two individuals in one another's mobile phone address book; they had no information about communication or explicit social ties of a real or virtual nature. However, the main idea of the work was that there tend to exist groups of individuals, including but not limited to friends, with highly correlated mobility patterns, and this fact can be exploited to improve location prediction and its associated applications in technological services.

In Chapter 3 of this dissertation, I take advantage of the abundance of data recently made available about users of location-based online social networks, their social connections, and their geographic locations, and present the results of explicit analysis of the geographic properties of social communities in large-scale datasets from online social networks, building on the foundation provided by the first exploratory study by Onnela et al. Given these results, I then show a potential technological application, namely, using geographically-based communities to improve friend recommendation in online social networks.

2.2 Social groups at the intra-city scale

2.2.1 Analysis

The analysis of social groups within cities has largely centred around connectivity properties of the social contact graph, without regard to exactly where the contacts that give the graph its structure take place. One major reason for this is that much of this research was based around the application of Delay Tolerant Networks (DTNs), the idea being that people coming into contact with one another around a city could transmit information between carried mobile devices, and so knowledge of the structure of this contact graph would be vital for effective routing decisions [HCY11]. Some major work in this area was done by Kostakos et al. [KNY+09], who considered a

city in terms of its Urban Pervasive Infrastructure (UPI), made up of human, technical, and spatial components. The authors empirically collected data about the UPI of the city of Bath, UK, and derived specific characteristics such as mobility, concerning distance travelled, speed, and so on, and social structure, concerning groups and patterns of encounter within the city. In order to map movements of people, they used Bluetooth devices such as mobile phones, and to identify locations, WiFi and GSM access points. Some outcomes of these studies did examine the kinds of spaces where groups meet in the city, finding, for example, that strong clustering and close proximity characterised an office environment, while these were not seen in street and campus networks [KOP+10].

However, these studies using mobile sensing to detect social contacts at the city scale are limited by the constraints of managing the sensing infrastructure, and therefore could not be carried out across many different cities in the manner now allowed by the widespread use of online location-based social networks (LBSNs). A 2011 study by Pan et al. [PBB11] making use of Bluetooth data from mobile phones also used data from the participants' Facebook profiles, and the authors compared the k-clique community structures of the two networks with that obtained when they are combined. The results showed that the two networks tend to complement one another, with the online network strengthening connections within cliques, but no analysis of the locations where these contacts took place was performed.

Once again, a lack of suitable large-scale data for analysing the relationship between geography and intra-city social groups has tended to limit the research possible in the area. Thanks to the increasingly widespread incorporation of location information into online social networks, and their high popularity in many places around the world, studying the use of spaces in the city by social groups has become more feasible, and my research outlined in Chapter 4 of this dissertation takes advantage of this opportunity to build on the work described above.

2.2.2 Applications

As early as 2007, the potential for the application of the idea that social groups may be associated with particular places was explored by Gupta et al. [GPJB07], who proposed an algorithm for automatically identifying social groups from contact traces and

discovering the associations of groups with places in the city from the community mobility traces. The application the authors had in mind was that of the delivery of geo-social recommendations of the kind that are provided to individuals, rather than groups, by today's LBSNs. However, the application itself was never explored, the focus of this work being the algorithm for identifying the groups.

Further concerning the application of place recommendation, the first work investigating the task of venue recommendation in LBSNs was that by Berjani and Strufe [BS11], who explored the application of recommending places to visit to users of the LBSN Gowalla in Austin, and New York City, an idea that has subsequently been examined by others [NSLM12a, NSLM12b, YYL10, YLL12]. None of this work considered actually providing recommendations for groups of users, but rather made implicit use of social groups and their geography at the city level. In particular, the fact that social groups may like to visit certain places means that it is possible to use information about a user's friends and the places they have visited to make personalised venue recommendations.

As in the case of groups potentially at the inter-city scale, it has largely been the case that analysis of the geographic properties of these groups needs to be undertaken before associated technological applications can be investigated. I present some analysis in this direction and an associated application in Chapter 4 of this dissertation.

2.3 Social groups at the intra-building scale

2.3.1 Analysis

Difficulties facing the analysis of the geography of social groups at the scale of a single building are generally not resolved by the availability of mobile phone data or the widespread usage of location-based online services, because the spatial resolution of both those kinds of data is not fine-grained enough to be able to distinguish the different kinds of spaces occupied within a building. However, technology can again help in such analysis, in this case with advances in ubiquitous sensing technology allowing unobtrusive monitoring of people's interactions with one another and their locations within a building, without some of the pitfalls of relying on self-reports and questionnaires, namely that people misremember or give answers that they feel are

socially desirable rather than the truth [BSBS78], or that they change their behaviour in response to awareness of being observed [Why43].

Some research in the field of architecture has examined the effect of the nature and layout of spaces on behavioural patterns and group interactions, since these issues are important to consider when designing buildings. One method used to analyse such phenomena is that of space syntax, used by Penn et al. [PDV99] to show that the physical space of a workplace building can directly affect how often employees meet one another and interact face-to-face. Similarly, a 2008 study by Toker and Gray [TG08] showed that the spatial configuration of a working environment can have a strong effect on the frequency and location of informal meetings between colleagues.

While the aforementioned studies did not explicitly concern the usage of building spaces by groups of people, work by Allen and Henn [AH06] has shown that the architecture of a technical workplace can be crucial for communication within and between teams, to the extent that physical space may be a management tool as important as organisational structure for today's technical organisations. They have studied the interplay between these two factors and found them to have profound effects on the process of information flow in workplace communication networks, with consequences for innovation and productivity. In particular, Allen describes in his book a study of the company Steelcase, where several departments were moved into a new building intended to stimulate communication among its occupants. Physical space in this building was allocated not departmentally but according to the products on which employees were working. Interdepartmental communication increased, showing the importance of physical space for enabling communication between groups, in this case, organisational departments.

These ideas were investigated in the field of computer science by Waber et al. [WOOKP10], who experimentally studied the effect of social group strength on productivity by using 'sociometric badges' to sense face-to-face interactions and perform indoor user localisation. They found that the strength of an individual's social group is positively correlated with productivity, but while the focus on social groups in this case was explicit, the impact of location was left implicit; the authors demonstrated that social group strength could be increased by allowing workers to take coffee breaks at the same time as their team-mates, which implies interactions taking place in informal spaces, but this was not directly studied.

In Chapter 5 of this dissertation I build on these previous analyses by presenting the results of a study designed to investigate directly, unlike the work thus far described, the impact of the availability of different kinds of building spaces on contact within and between groups in a single building, in this case a research laboratory. This study has the unique advantages that it was conducted using automatic sensing of interactions and location, rather than human observation or self-reports, at the same time as comparing the behaviour of the same group of employees of the same organisation across two different buildings, which is usually difficult due to the time, money, and other resources involved in moving an organisation from one location to another. I show that different kinds of spaces are likely to be important for different kinds of communication that may occur within and between groups.

2.3.2 Applications

In the case of building-scale groups too, very little research exists into how the role of physical space for groups may be used in technological applications. This is partly due to the fact that until recently, most research in this area was done by architects whose aims are focused around building design rather than other ways of using the results of their analysis.

One exception to this is the 'break-time barometer' system presented by Kirkham et al. [KMG+13], which is based on the idea of promoting the simultaneous use of informal 'break time' spaces by people in the workplace, thus encouraging interaction between multiple members of the group of employees in the space. The authors do not explicitly analyse the relationship between groups, but they do discuss the importance of spontaneous interactions and the fact that these are likely to happen in informal spaces, as I describe in Chapter 5 of this dissertation.

Other applications of this kind of analysis involve the automatic sensing of interactions between members of different groups in the workplace and the display of this information to employees with the intention of encouraging them to be more aware of and perhaps change their behaviour, as suggested by Kim et al. [KCHP08], who discussed the use of sociometric badges capable of sensing automatically interactions in a group during meetings, and the subsequent indication to the participants of who was tending to dominate conversation and who was less forthcoming. Such an application

specifically involving the use of space by social groups in the sensing of communication has not yet been explored, and in Chapter 5 I also discuss how this might be done, and the challenges that still need to be addressed before such an application could be widely adopted as a technological tool in workplaces, or indeed in other kinds of buildings where group communication is important.

Chapter 3

Social groups in global-scale networks

In today's technologically connected world, it is easy for geographically distant individuals to communicate with one another, especially since the advent of the Internet and its widespread use. Thanks to these technological social networks, there is much information available about the social ties between individuals and also about their location, and the study of these spatially embedded social networks has been a recent topic of interest for researchers, who have mainly studied mobile phone datasets [OSH+07, GHB08, OAG+11] and online social networks [BSM10, CS11, CCLS11, CML11, CTH+10]. Despite the ease afforded by technology of maintaining social networks over long distances, distance still matters for dyadic relationships: the likelihood that two people have a connection in these networks has been shown to follow a so-called gravity law, with the probability of a social link decreasing proportionally to the square of the increasing distance between their locations [KCRB09, LBdK+08, SNLM11]. However, to date, little research has specifically examined the role of geographical distance and physical space for groups of three or more people, beyond the observation that members of groups in social networks are often close to one another in space, as described in Chapter 2.

In this chapter, I present a more detailed analysis of the spatial properties of groups of friends in two online social networks with location information: the microblogging service Twitter, and the online location-based social network (LBSN) Gowalla. I define some measures of the extent to which a group is connected by physical places, extract communities from the social networks, and then examine how place-centric these communities are. I show that in Twitter, space appears to be less important than

in Gowalla. I then show how it is possible to find place-based communities in networks such as Twitter, by incorporating place information into community detection, and illustrate one potential application of these place-based groups in the online social networks, namely to aid friend recommendation, a common task in such services.

In relation to the thesis, this suggests that while space seems to be less important for the global-scale communities facilitated by some online social networks than for others, location-based communities of users still exist in the networks, and this fact may be exploited to improve the application of friend recommendation in online social networks.

3.1 Datasets

I first analyse the spatial properties of groups, using two datasets containing information from online social networking services where information about the places visited by the users is available. One of the datasets was collected from the microblogging service Twitter; the other was collected from the online location-based social network Gowalla. Both datasets contain a social network, and location information in the form of 'check-ins'.

3.1.1 Check-ins

Check-ins are a major component of today's location-based online social networking services. When using such services, people record their current location in terms of a specific place such as 'café', 'railway station', and so on, rather than simply geographic coordinates, using an application on their mobile phones. A check-in is a record of the form (u, p, t), where u is the ID of the user who made the check-in, p is the ID of the place where the check-in took place, and t is a timestamp. The datasets also contain for each place p its name (e.g., 'Starbucks Coffee', 'King's Cross Railway Station', and so on), its geographic coordinates (latitude and longitude), and a category (e.g., Food, Transport, and so on).

3.1.2 Twitter

Twitter is a microblogging service created in 2006 [KGA08]. Users can 'follow' other users to receive their 'tweets', that is, 140-character messages, on their personal Twitter pages, and it is widely used for content dissemination, for example, users with similar interests sharing links to news articles or other pages on the web [KLPM10]. Not being a specifically location-based social network (LBSN), Twitter itself does not have a 'check in' function, but many users of Foursquare, the most popular LBSN in 2010 when the dataset was downloaded, post their check-ins from Foursquare on Twitter [NSMP11].

The Twitter dataset I analyse here was gathered between May 2010 and September 2010, and contains all of the check-ins from Foursquare publicly posted to Twitter during that time[1]. The dataset also contains the Twitter friend lists of the users (their lists of followers and people they follow). The Twitter social graph I construct and analyse is undirected, since I include edges between users u_1 and u_2 only when each of u_1 and u_2 follows the other, although Twitter does allow directed edges. This is because I intend to analyse online friendships where both users concerned have indicated that they are friends with one another, rather than the kinds of ties that also exist on Twitter where a celebrity may be followed by many fans, but does not know these people as individuals [KLPM10].

3.1.3 Gowalla

Gowalla was a location-based social network launched in 2009, and closed in 2012 after the company had been bought by Facebook [TBL13]. The focus of the service was location-sharing, so users could add friends to form a social network, and share their check-ins with these friends. The dataset I analyse is a complete snapshot of the service downloaded on 19^{th} August 2010. It contains all of the check-ins that took place prior to that time, and the entire social network.

The numbers of users, friendships, places, and check-ins contained in each dataset are shown in Table 3.1.

[1] It is estimated that this constitutes about 25% of all the check-ins made on Foursquare during this period. Some of these publicly posted check-ins may be missing if they were deleted by the user before they could be downloaded.

	Users	Friendships	Places	Check-ins
Gowalla	165,051	765,872	13,561,773	1,541,951
Twitter	663,198	11,959,895	34,037,471	4,274,022

Table 3.1: Number of users, friendships, places, and check-ins in the Gowalla and Twitter datasets.

3.2 Measuring the place properties of groups

I study the properties of groups of friends in the two social networks, in terms of the places where members of the groups have checked in. Community structure is an important property of many networked systems, not just social networks, and so there exist many methods for extracting meaningful groups of nodes from such networks [For10]. The precise definition of community employed by these methods can differ; for example, one can require that in order to form a community a subset of nodes in a graph should have more connections between them than expected in a null model [New06], or alternatively that the community has more links between its own members than from its members to other nodes in the graph [GN02].

In the following, I have used two community detection methods: the Louvain method, proposed by Blondel et al. [BGLL08], and the DEMON algorithm, proposed by Coscia et al. [CRGP12]. The Louvain method is a greedy agglomerative method based on modularity optimisation, that works by first placing each node in its own community, and then repeatedly merging communities in such a way as to attempt to maximise the increase in modularity, which is a measure of the quality of a partition of a graph into communities [New06]. The DEMON algorithm works by having each node 'vote' for the communities that are in its neighbourhood, using a label propagation algorithm [RAK07], and merging these communities to form the final partition. Perhaps the most important difference between the two methods is that the Louvain method produces only one partition of the network into communities, meaning that each node may be assigned to one and only one group, while DEMON allows for overlapping communities. I chose these two methods because they are reasonably computationally efficient and so can be used on large graphs such as the social networks I analyse here, and because they do not require the number of communities to be specified in advance, since this would impose unnecessary constraints on the communities extracted.

3.2.1 Measures for place properties

I will now define the measures I use to assess the place properties of communities extracted from the social networks. First, it is necessary to give some preliminary definitions.

3.2.1.1 Preliminary definitions

Each social network is represented as an undirected graph $G = (V, E)$, where the set of nodes $V = \{u_1, \ldots, u_n\}$ is the set of n users in the network, and the set of edges E is composed of pairs of users present in one another's friend lists.

For each online service there is a set of places $P = \{p_1, \ldots, p_m\}$, containing all of the m places where the users of that service in the dataset have checked in. Let P_i be the subset of P containing the places where user u_i has checked in, and let V_j be the subset of V containing all of the users who have checked in to place p_j.

A subset of the social graph G^P, which I will call the *placefriends graph*, can also be defined: $G^P = (V, E^P)$, where the set E^P of *placefriends edges* is the subset of E containing only those edges (u_i, u_j) where u_i and u_j have checked in to at least one of the same places.

The definitions of the place properties of communities can now be given.

3.2.1.2 Placefriends edge density

Placefriends edge density measures the extent to which members of the same group have places in common, and is defined to be the fraction of possible pairs of members of a community C who share at least one place:

$$\frac{1}{|C|(|C|-1)} \sum_{u_i, u_j \in C} A^P_{ij}$$

where A^P_{ij} is the ij^{th} entry in the adjacency matrix of G^P (i.e., 1 if u_i and u_j share a place, 0 otherwise). The value of the placefriends edge density will be close to 1 when a community member typically shares at least one place with most other members, i.e., the community is largely place-based and does not exist solely online. Note that the shared place need not be the same for all pairs; the extent to which this is the case is captured by the next measure.

3.2.1.3 Place focus

Place focus measures the extent to which members of a group have checked into one particular place, and is defined to be the fraction of members in a community who have checked in to the most-shared place of that community:

$$\max_{p_i \in P} \left(\frac{|V_i \cap C|}{|C|} \right)$$

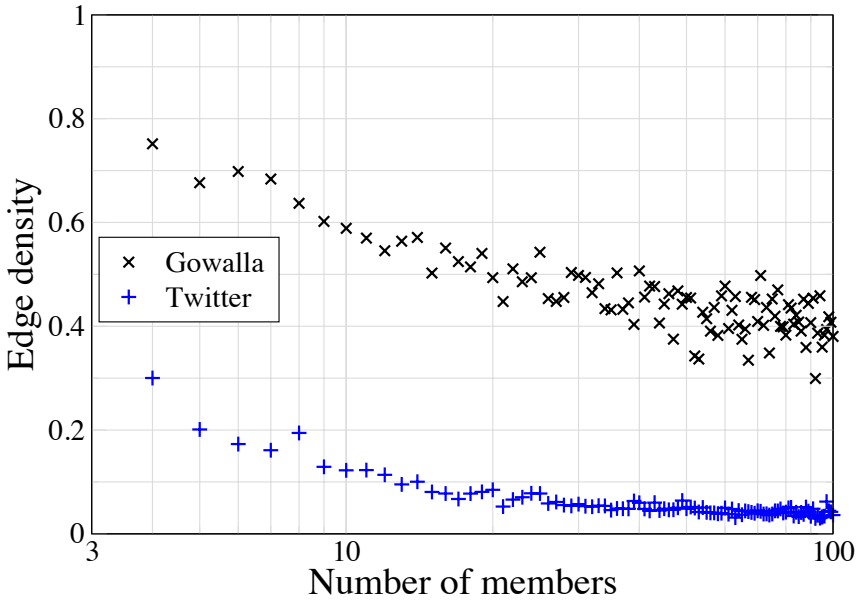
The value of the place focus will be close to 1 when there is a single place where most of the members of a community have checked in.

Together, these two measures can help to characterise communities in terms of whether their members tend to check in to the same places, and whether they all tend to check in to the same places, or alternatively have different places in common with one another.

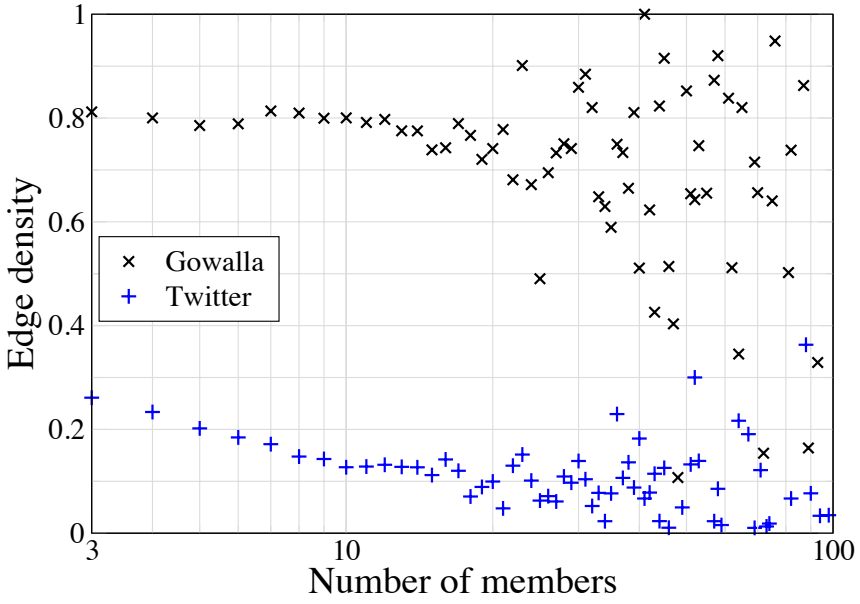
3.2.2 Results of measuring community place properties

Figure 3.1 shows the mean placefriends edge density for communities in both networks, giving an indication of whether communities tend to contain users who have checked in at the same places or not. The figure shows that when communities are extracted from the two datasets' social networks using the DEMON method and using the Louvain method, communities in Gowalla tend to have considerably higher placefriends edge density than do communities in Twitter. In Gowalla, communities with fewer than 20 members have an average placefriends edge density of 0.5 or more when extracted using the DEMON method, and of 0.7 or more when using the Louvain method. This contrasts with Twitter communities, which have an average placefriends edge density of 0.3 or below for both community detection methods. The noisy nature of the right-hand side of the Louvain plots can be attributed to the fact that the communities extracted by this algorithm do not overlap, while DEMON allows overlapping communities and therefore results in a smoother curve. The tendency of Twitter communities to have lower placefriends edge density appears independent of the algorithm chosen, however.

The place focus of communities in both networks is shown by Figure 3.2. Similarly to placefriends edge density, communities in Gowalla show a higher average value



(a) DEMON



(b) Louvain

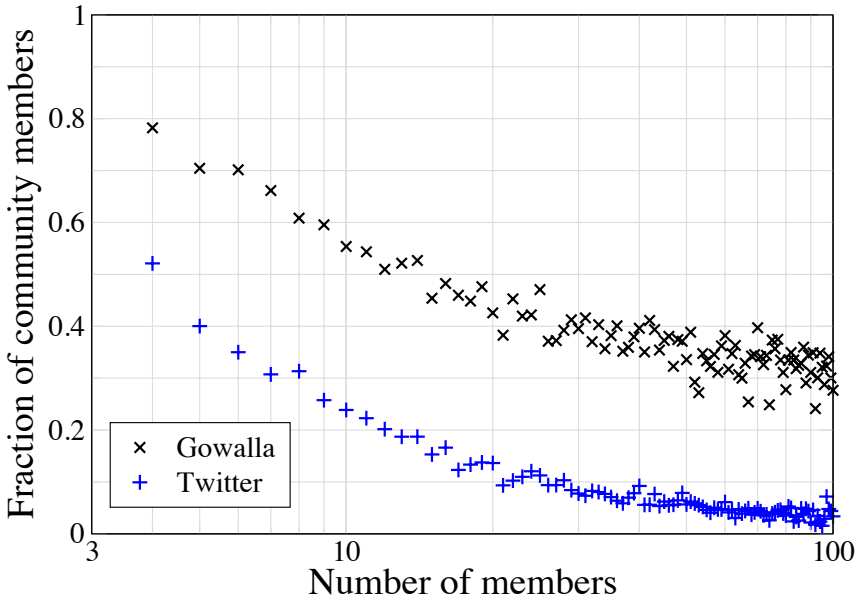
Figure 3.1: Placefriends edge density of communities: fraction of possible pairs in the community who have checked in to one or more of the same places.

of this measure than communities of the same size in Twitter. This indicates that in Gowalla, it is more common for many members of a community to have checked in to that community's most-shared place than it is in Twitter.

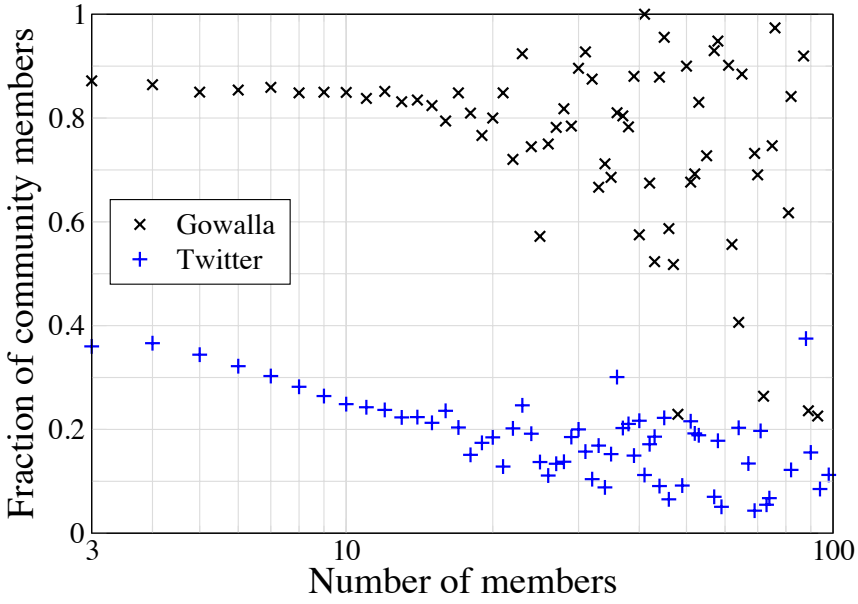
These results suggest that communities in Gowalla tend to be more place-centric than those in Twitter in terms of both the measures defined for assessing the place properties of communities in the online social networks. In the next section, I show that place-based communities do indeed exist in the Twitter social network, but they are more readily found when place information is incorporated into the community detection process. Specifically, I demonstrate that by assigning weights to edges in the social network using information from the check-ins of the users concerned, it is possible to give place-based edges greater importance for community detection and expose place-based groups that might otherwise be obscured by the more general edge structure of the social graph.

3.3 Uncovering place-based communities in online social networks

The fact that the communities examined thus far in the Twitter social network did not show a strong tendency for users in the same community to have checked in to the same places does not necessarily mean that groups of friends who do check in to the same places do not exist in the Twitter network. It could be the case that by using not only the topology of the social graph, but also including the available place information in the community detection process, it is possible to reveal such groups. One reason for this might be that many links in the Twitter social network have nothing to do with places, but connect users with similar interests who are located geographically far apart [SMML10]. These links will be treated by the community detection algorithm with equal importance to those edges that do connect friends who check in to the same places, who should be grouped in place-based communities. In this section, I present a method for distinguishing between these two kinds of edges in the social graph: I assign weights to edges using values derived from information about places where the connected users have both checked in. Later in this chapter, I will show how the place-based groups revealed using this method could be useful to location-aware



(a) DEMON



(b) Louvain

Figure 3.2: Place focus of communities: fraction of community members who have checked in to the community's most-shared place.

technological applications.

3.3.1 Method

Given the social network $G(V, E)$, the set of places P, and the associated user check-ins, I define a function f that takes an edge connecting users u_i and u_j, e_{ij}, from the social graph, and returns a non-negative real number. This number is a weight $f(e_{ij})$ derived from the information about the check-ins made by u_i and u_j. The proposed method for extracting place-based communities is then as follows:

1. Assign to each edge e_{ij} a weight $f(e_{ij})$ based on information about the check-ins made by u_i and u_j.

2. Remove from the graph all edges with weights below some threshold t.

3. Remove from the graph all nodes with no incident edges.

4. Run a community detection algorithm that works with weighted graphs. In the following, I have used the Louvain algorithm [BGLL08] to perform community detection.

3.3.1.1 Labelling functions

I have experimented with a number of ways to derive weights for the edges of the social graph from place information, as used in step 1 in the method described above. For each way of deriving weights, a weighting function f is defined, that takes an edge of the social graph and returns a weight based on the check-ins of the two users connected by the edge. These weighting functions are:

- `binary`: This results in the subgraph of the unweighted social graph G that contains only those edges e_{ij} where u_i and u_j have checked in to at least one of the same places, that is, equivalent to the *placefriends* graph G^P defined in Section 3.2.1.1. The function is so named because it will return weights valued either 1, if u_i and u_j have checked in to one or more of the same places, or 0, if they have not. Note that this function forces use of the threshold $t = 1$, as otherwise all edges in the graph would be removed in step 2 of the method described

above. Formally,

$$f_{binary}(e_{ij}) = \begin{cases} 1 & \text{if } |P_i \cap P_j| > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Recall that P_i represents the subset of P containing the places where user u_i has checked in.

- `checkins`: This function assigns to each edge e_{ij} a weight equal to the sum, over all of the places where u_i and u_j have checked in, of the lower of the numbers of check-ins that u_i and u_j have made to that place. This aims to capture the extent to which users visit the same places, since in the case of a place that only one of the users has checked in to, the summand in question will be 0. The definition also avoids giving undue weight to edges where one user has checked in many times and the other very few times. Formally,

$$f_{checkins}(e_{ij}) = \sum_{p_k \in (P_i \cap P_j)} min(c_{ik}, c_{jk})$$

where c_{ik} and c_{jk} are the respective numbers of check-ins that user u_i and user u_j have made to place p_k.

- `places`: This assigns to each edge e_{ij} a weight equal to the number of places where both u_i and u_j have checked in. Formally,

$$f_{places}(e_{ij}) = |P_i \cap P_j|$$

- `ratio`: For each place, I compute the ratio of the number of all the check-ins at a place to the number of users who have checked in there, that is, the mean number of check-ins per user for that place. The weight assigned to an edge e_{ij} by this function is equal to the maximum value of this ratio for the places that u_i and u_j have in common, or 0 if they have no places in common. Formally,

$$f_{ratio}(e_{ij}) = \max_{p_k \in (P_i \cap P_j)} \left(\frac{c_k}{|V_k|} \right)$$

  where c_k is the total number of check-ins made to place p_k, and V_k is the set of users who have checked in to place p_k as before.
  The rationale for this definition is that places where many people check in infrequently, such as an airport, will give low values and are not likely to indicate important place-based friendship ties. On the other hand, places with high values are checked into by a few users many times, for example, somebody's house, and may indicate that the sharing of a place by those two friends is in some sense more important [SNM11]. Thus, edges where users share such places are weighted more highly by this function.

3.3.1.2 Thresholding

Step 3 in the method described above aim to remove users who do not belong to place-based communities in the online social network from the social graph. After assigning weights to edges, those edges with weights lower than a threshold value t are eliminated. The users left with no ties are then removed from the graph. Choice of the threshold t is explored further in Section 3.3.2.2.

3.3.2 Assessing potential utility of extracted groups

I assess the utility of the place-based groups extracted using this method by examining whether or not their members have been colocated, that is, not just whether they have visited the same places, but whether or not they may have visited those places at the same time. I argue that it is more likely that having checked in to a common place is

more meaningful for members of a place-based group if they were checking in there close to one another in time. Since the datasets contain no 'check-out' information, I consider users to have been colocated if they have checked in to the same place within one hour of one another, in line with previous research using online location-based social network datasets [FVA+12, HSL11a, HSL11b].

I define a measure called *colocation fraction* in order to quantify how many members of a group have been colocated according to their check-ins. First of all, define the *colocation matrix* M:

$$M_{ij} = \begin{cases} 1 & \text{if } u_i \text{ and } u_j \text{ have been colocated according to their check-ins, as defined above} \\ 0 & \text{otherwise.} \end{cases}$$

The colocation fraction of a community C is then defined to be:

$$\frac{\sum_{u_i,u_j \in C} M_{ij} \cdot A_{ij}}{\sum_{u_i,u_j \in C} A_{ij}}$$

where A_{ij} is the adjacency matrix of the social graph G, that is, it has the value 1 when u_i and u_j have a connection in the social graph, and 0 when they do not. The colocation fraction is therefore equal to the proportion of the social ties between members of the community C where the users concerned have been colocated.

3.3.2.1 Results

Figure 3.3 shows the mean colocation fraction for communities of different sizes, when the threshold parameter t was set so that 90% of edges were pruned from the social graph. The communities detected having used the `places`, `checkins`, and `ratio` functions tend to have higher values of colocation fraction than communities extracted from graphs weighted using `binary`, and from the unweighted graph (labelled as 'unannotated' in the figure). The figure shows that this effect is seen in Gowalla as well as in Twitter, which suggests that even though communities detected in the unweighted graph in Gowalla tend to contain users who visit the same places, as shown in the analysis earlier in this chapter, if it is desired to find *colocated* groups of users, it could still be useful to use this method to find communities.

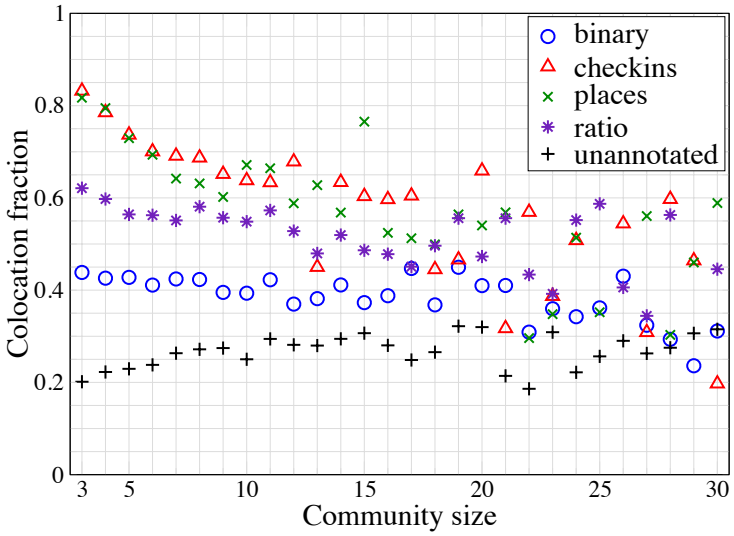
The functions for `places` and `checkins` give higher values of the colocation fraction than does the function for `ratio`. This could be because the thresholding step removes users who have checked in to fewer places, and it is possible that if users check in to enough of the same places they will eventually be colocated by chance. However, these users have declared online friendship, so the method is not simply grouping so-called *familiar strangers* [Mil92] who meet regularly but who are not friends, and I argue that colocation of these users may still be meaningful.

Note also that for `places` and `checkins`, thresholding necessarily discards users who have checked in to fewer than t places and with fewer than t check-ins, respectively. It may be that to detect place-based communities it is necessary to ignore users for whom there is not enough location information. As location continues to become more important to online social networking, data sparsity may present less of a problem. For now, the verifiably place-based communities that can be detected include the users who actively make use of location features, and these are likely to be the users for whom applications of the detection of these communities could be most useful.

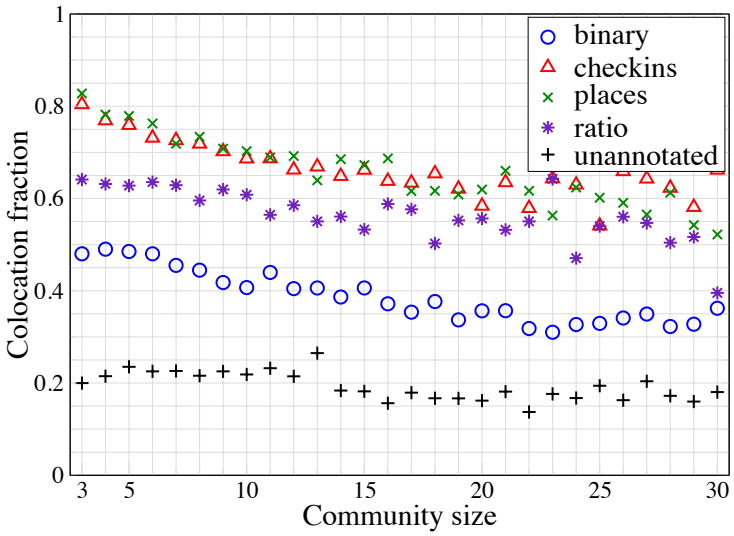
3.3.2.2 Choice of threshold parameter

The threshold value t acts as a tuning parameter, determining how aggressively the method excludes users from consideration for inclusion in place-based groups. To illustrate the effect of this parameter, Figure 3.4 shows the mean colocation fraction for communities of different sizes extracted from the Twitter network weighted using the `places` function. The figure also shows the effect of removing 90% of links from the graphs at random, rather than according to their weights, to show that it is the weighting that produces the result rather than the removal of links from the social network. Figure 3.4 shows that communities found using higher t are increasingly place-based. Pruning less spatially relevant links from the graph will increase the extent to which communities tend to be place-based, but users may be removed due to not having highly-weighted incident links.

When choosing t the needs of the intended application must be considered: it may be that any communities found need to be highly place-based, even if most users will be excluded from placement in a community, in which case t should be higher than if less place-based communities are more acceptable but it is desirable to assign as many



(a) Gowalla



(b) Twitter

Figure 3.3: Colocation fraction of communities found using the different annotation functions.

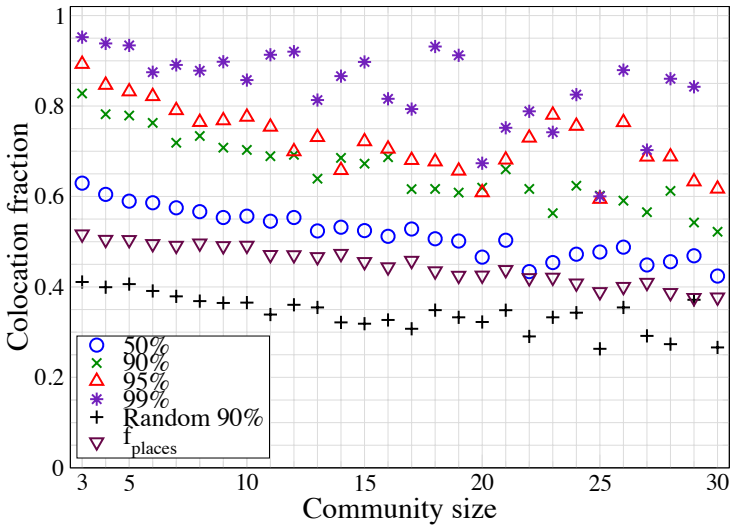


Figure 3.4: Effect of varying the threshold parameter.

users as possible to a group. For example, if using these communities for friend recommendation, as discussed in the next section, it might be better to have more users available for consideration. Alternatively, if the place-based groups were being used to sort users to aid privacy management, groups being more definitely based around places could be more important and a more restricted group of users might be an acceptable cost.

3.4 An application: friend recommendation in a location-based network

One common task for location-based social networking services such as Gowalla is that of *friend recommendation*, that is, suggesting to users of the service other users with whom they do not currently have ties in the social network, but to whom they might want to connect. One study by Scellato et al. [SNM11] found that 30% of newly placed ties in an LBSN were placed between people who had checked in to at least one of the same places. It therefore seems likely that members of the same place-based

	Nodes	Edges	Clustering
May 2010	109,115	517,113	0.17
August 2010	156,495	747,771	0.16

Table 3.2: Number of nodes, number of edges, and clustering coefficient of the two snapshots of the Gowalla social graph.

community who are not currently connected might make good candidates for friends to recommend to a user, possibly more so than members of the same group found without using place information. In this section, I demonstrate how place-based communities might be used in this way, to improve the application of friend recommendation in LBSNs.

3.4.1 Dataset

I again use the Gowalla dataset described in Section 3.1. The dataset contains two complete snapshots of the social network, one downloaded on 4^{th} May 2010, and the other on 19^{th} August 2010. Table 3.2 shows the number of nodes, number of edges, and the clustering coefficient of the social network in each of the snapshots. In particular, 230,658 edges were formed during the 3 months between the two snapshots.

3.4.2 Methodology

I investigate the potential use of place-based communities for friend recommendation by inspecting the edges formed in the social graph between the two snapshots, making the assumption that one can consider that a successful friend recommendation made using the first snapshot would be one where the user in question had indeed formed a social tie to the friend suggested by the time of the second snapshot.

I extract communities from the first snapshot of the graph using the Louvain method [BGLL08], both without weighting the graph, to extract 'social' communities, and by weighting the graph using the `binary` function described in Section 3.3.1.1 to find place-based, or 'local' communities. I then identify the ties formed in the period between the two snapshots, and consider them according to whether they are within or between both kinds of community. Note that I do not consider in this analysis nodes that were not present in the first snapshot, and that isolated pairs of nodes, single nodes, and nodes

with no ties are not considered to be in communities and are therefore excluded from the analysis.

Since many community detection methods, including the Louvain method, use heuristics in order to manage the computational complexity of the problem, they are non-deterministic and can give different output depending on the order of the input [For10]. While it was unnecessary in the analysis earlier in this chapter, because the aim was to capture general characteristics of the graph structure, which as aggregates are not greatly affected by small differences in community assignment due to randomness in the detection method, when performing *longitudinal* analysis and examining the community placement of individual nodes it is necessary to account for this effect. To this end, I use the algorithm proposed by Kwak et al. [KCE+09], which works by running a chosen community detection algorithm able to handle weighted graphs n times on the same network. The input is given in a randomised order each time, thus obtaining n potentially different community partitions of the graph. In principle, if the graph has a strong community structure then these partitions should differ only in the community placement of a relatively small subset of nodes. The network is re-weighted according to the frequency with which pairs of nodes have been place in the communities of each of the n partitions. In particular, the weight w_{ij} of an edge connecting the nodes representing users u_i and u_j is increased or decreased proportionally to the number of times that this pair of nodes have been put in the same community in each of the n partitions. The re-weighting procedure has the effect of reinforcing more robust groupings over those appearing by chance or due to a particular input ordering. This process is iterated on the re-weighted network, until a consistent placement of nodes into communities is obtained and all the partitions obtained in the n runs of an iteration are identical[2].

I compare the actual number of links created within or between communities after the first and before the second snapshots, to the corresponding expected number of links created in a null model where new links are placed both at random, with the only constraint being that self-loops and parallel edges are avoided. Specifically, if x is the observable of interest, i.e., the number of links created within or between communities, and the expected value of x in the null model is denoted by $\tilde{x}$, I compute the ratio r:

[2] Although Kwak et al. noted in their paper that the convergence of their method cannot be guaranteed for certain graphs, I did not encounter this problem with the Gowalla networks and I was able to find a stable partition of both graphs.

$$r = \frac{x}{\tilde{x}}$$

so that if the same number of edges form as expected, $r = 1$, and otherwise, r indicates how many times greater or smaller than expected is the number of edges actually formed.

For example, if computing $\tilde{x}$ for the number of edges forming between nodes in the same social community, then $\tilde{x}$ is equal to the total number of edges formed in the social graph between the first and the second snapshots, multiplied by the fraction of those edges that could be formed between nodes belonging to the same social community. This quantity can be thought of as the number of 'missing' edges within social communities, i.e., the number of pairs in the same communities that were not connected at the time of the first snapshot. In this case, the expression for $\tilde{x}$ would read:

$$\tilde{x} = \frac{K_{\circ}^{in}}{K_{\circ}} K_{+}$$

where $K_{\circ}^{in}$ is the number of edges missing between members of the same social community, $K_{\circ}$ is the total number of missing edges between members of the same social community and members of different social communities, and K_{+} is the number of edges added between the first and second snapshots.

I also compare the number of edges actually formed within or between social and local communities to the number of such edges formed when the same total number of edges as are actually formed between the two snapshots are placed according to the preferential attachment model [BA99], where the probability that an edge forms from a chosen node to a target node is proportional to the existing degree of the target node, resulting in a 'rich-get-richer' effect that produces a power-law degree distribution commonly seen in online social networks [MMG^{+}07, NRC08, ZSW^{+}12]. This will give an indication of how recommending members of the same social or local community as friends could perform in relation to recommending users according to their popularity, which is possibly a less naïve approach than recommending users at

		Random	PA
Social communities	**Between**	0.62	0.37
	Within	25.6	18.0
Local communities	**Between**	0.78	0.74
	Within	70.7	44.0

Table 3.3: Edges formed within and between social and local communities that actually formed, as a proportion of those formed using random and preferential attachment (PA) models.

random.

3.4.3 Results

The results of computing r as defined above, for edges forming within and between social and local communities, are shown in Table 3.3. The number of social ties formed between members of the same social community, that is, communities found without using place information, is 25.6 times greater than expected when ties are placed uniformly at random between disconnected users, and 18.0 times greater than when edges are added according to the preferential attachment model. This demonstrates the potential utility of communities in friend recommendation, suggesting potential benefit in recommending disconnected users within the same community as friends for one another, rather than suggesting users at random, or the most popular users.

The additional value that might be gained from considering *place-based* communities is demonstrated by the value of r for edges formed between members of the same local community; social ties are 70.7 times more likely to form between users placed in the same local community in the first snapshot than when placing edges at random, and 44.0 times more likely than when using preferential attachment. The fact that these values are higher than the corresponding values for social communities suggests that recommending disconnected users in the same local community as friends could be even more beneficial when performing the friend recommendation task.

3.5 Discussion

The relative unimportance of places in Twitter communities extracted without incorporating place information into the community detection process may reflect Twitter's

focus on content-sharing, which encourages users to connect based on shared interests that may not depend on location [SMML10]. In other networks such as Gowalla where there is an explicit focus on places, geography may be inherently more important. It might also be that this effect is seen because in online social networks based around location-sharing, people are more willing to share their location with people who would know where they were anyway, or those they know in person having met them face-to-face, which requires a geographic link, than with people they have never met [CRH11, LCW+11, PKK12]. This suggests that in these networks, geography is important to characterise communities, despite the potential for global-scale communities afforded by easy online connections.

My research into detecting place-based communities is related to the work by Expert et al. [EEBL11], who noted the tendency for communities in spatial networks such as the mobile phone communication network to have some local geographic basis, and presented a method for uncovering space-independent communities by accounting for this effect during community detection. My aim in this chapter was conceptually the opposite: to uncover place-based communities in networks where there are place-based ties included alongside those that may not have a geographic basis. The rationale behind this was that being able to find local groups could be helpful for various applications, one example of which is the task of friend recommendation in online social networks.

3.5.0.1 Limitations

There are some caveats that must be borne in mind when considering analysis of datasets such as those I have examined here. One of these is that such datasets are unavoidably biased towards representation of certain demographics, and the results should therefore not be taken to be representative of general principles but as possibly specific to the online services from which the datasets were taken. For example, a study by Mislove et al. in 2011 found Twitter users in the US to over-represent densely populated areas, to be predominantly male, and to represent a non-random sample of the population distribution of race and ethnicity [MLA+11]. The user behaviour observed in such services may be influenced by such biases. However, these inherent data biases do not mean that results such as those I have shown cannot be useful to the services that use the data, for example, the online social networks themselves; even

the data may not be representative of general mobility behaviour, characteristics of the ways that an online service's particular users behave in relation to that service could be used to improve that service. One example of this might be the potential utility of local communities for friend recommendation in Gowalla that I have outlined in this chapter. As for the general issue of data bias, it may be interesting to observe whether the situation changes if mobile Internet access and the incorporation of location into online services continues to become more widespread.

One must also bear in mind that check-ins of the kind analysed here are voluntary, and therefore the places visited by users as indicated by their check-ins are an underestimate of the places those people have actually visited. People may not check in everywhere they go for a variety of reasons, including not having time, getting bored, or having privacy concerns [LCW+11]. Caution should be exercised if seeking to draw conclusions about actual mobility patterns from check-in data, although other research has shown that there are considerable similarities between the mobility patterns observable in location-based online social networks and those seen in mobile phone data [CML11]. Of course, mobile phone datasets also have their inherent biases, and so the most reliable applications of results from the analysis of such data are most probably in areas directly relating to the use of those services, whether the mobile phone network or online social networks of the kind that I have studied.

3.6 Summary

In relation to the thesis that the relationship between physical space and social groups at different scales can be exploited to create or improve technological social applications, I have explored in this chapter the role of space in groups in two online social networks with location information: the microblogging service Twitter, and the location-based network Gowalla. Both of these networks exist online, and therefore have the potential for social ties to span between cities and even continents. In analysing communities in these social networks, I found that members of the same community tend to check in to the same places to a greater extent in Gowalla than in Twitter, possibly due to Twitter's focus on content-sharing in comparison to Gowalla as a location-based service. In Gowalla, geography seems essential to characterise communities, while in Twitter, space is less important but still plays some role, with place-based groups being

revealed in the network when information about places is taken into account when extract communities. In terms of possible applications of these properties, I have showed how the importance of geography for communities in an online social network such as Gowalla could be used to improve friend recommendation in such a service, since people seem more likely to connect to others within the same place-based community as themselves than others in the same group without considering location when extracting communities.

Chapter 4

Social groups in city-scale networks

Before the Internet age, face-to-face meetings were generally necessary for people to become acquainted and therefore to foster friendships. Although it was thought that the rise of the Internet could remove this barrier posed by geography to the formation of social networks [Cai01], so far it appears that face-to-face meetings between friends are still an important part of the social life of an urban area [MWC10]. The importance of the nature of the particular places in the city where friends meet is itself important, as different kinds of space are different in their social use [CTH+10, KOP+10], but it has been difficult until recently to study the relationship between the kinds of places around cities where people meet and the social network structure, especially in terms of groups of friends. Similarly to the global-scale social networks studied in the previous chapter, one reason for this has been the lack of easily accessible, large-scale data about social connections within a city together with the places where people go, which is now available from online location-based social networks.

In this chapter, I study social networks with accompanying location information at an intra-city scale. I first analyse some properties of places that friends have in common with another, particularly in relation to the presence of triangles in the social networks, which can be building blocks for larger communities. I also study potential face-to-face meetings at places around the city by analysing colocation between friends, and propose a model for generating a network with some of the empirically observed properties of the network.

I then focus on the behaviour of colocated friends in an online location-based social network and in a mobile phone communication network, and find that groups of

colocated friends seem to visit different places from those where people go when they are on their own. I show that this different usage of places around the city by individuals and by groups of friends can be helpful in venue recommendation, a common task in location-based online social networks. By considering groups of friends as units rather than as the sum of their parts, it could be possible to improve recommendation performance due to the difference between solo and group mobility.

In relation to the thesis, my findings suggest that the different uses and variety of places around a city is important for groups in the social network, in that some locations tend to be used for social meetings far more than other kinds of places. In other words, people go to different places for the purpose of shared activities with their friends than they go to in general, and this behaviour could aid applications such as venue recommendation in location-based online social networks.

4.1 Properties of city-scale social networks

I first analyse the social and spatial properties of city-level social networks constructed using friendship and location information from a location-based social network dataset. I examine the structural characteristics of the social graphs and show that they have properties commonly seen in social networks. These include the high level of clustering and community structure that signify social groups, as discussed in the previous chapter where I showed that there are such groups in online social networks that are place-based, in that their members visit the same places as one another. In accordance with this, I now analyse the places visited by the people in the social network, and show that it is common that members of structural triangles have visited at least one of the same places. I further examine the nature of the places in cities that these friends visit, and show that some kinds of places seem more likely to host social meetings than others.

4.1.1 Data description

The dataset I analyse in this section is comprised of all the check-ins from the location-based social networking service Foursquare that were posted publicly on Twitter between November 2010 and September 2011, where the check-in took place in one of

City	Users	Venues	Check-ins
Atlanta	28,275	18,270	368,608
Boston	23,579	13,243	296,150
Chicago	42,791	33,261	715,652
Minneapolis	13,396	12,696	235,793
Seattle	16,205	15,051	260,023

Table 4.1: Number of users, number of venues, and number of check-ins in the dataset, for each of the five cities.

five major US cities: Atlanta, Boston, Chicago, Minneapolis, and Seattle. Similarly to the dataset described in Section 3.1, a check-in contains a user ID, a venue ID, and a timestamp, thus identifying a user as having been in a certain place at a certain time. The dataset contains information for each venue downloaded directly from the Foursquare venues database, which includes the venue names, geographic coordinates, and categories indicating something about the usage of a place; these are defined by Foursquare and include Arts and Entertainment, College and University, Food and Drink, Nightlife Spot, Outdoors and Recreation, Professional, Residence, Shop and Service, and Travel and Transport. The dataset also contains the Twitter friend lists of all of the users present (see Section 3.1.2).

The numbers of users, venues, and check-ins present in the dataset for each of the five cities are shown in Table 4.1.

4.1.1.1 Social networks

I study the *placefriends graph* for each city, as defined in the previous chapter (Section 3.2.1.1). Recall that this is a social network where users are connected by a social tie when each is in the friend list of the other and have checked in to at least one of the same places on Foursquare. Formally, given:

1. The set of n users: $V = \{u_1, u_2, u_3, \ldots, u_n\}$, and

2. The sets of places for each user $u_i \in V$: $P_i = \{v \mid \text{user } i \text{ has checked in to place } v\}$

I define the graph $G^P = (V, E^P)$ where the n nodes V of the graph represent the users and the graph G^P has an undirected edge (u_i, u_j) in the set of edges E^P whenever:

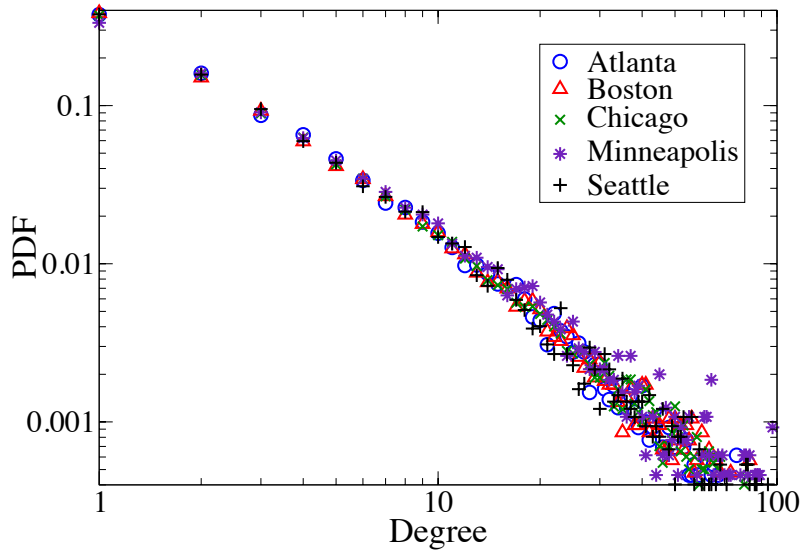


Figure 4.1: Degree distributions in the city-level social graphs.

1. the users both follow one another on Twitter, and

2. $|P_i \cap P_j| > 0$, i.e. the users have both checked in to at least one of the same places during the data collection period.

I will use this graph to study in more detail, at the intra-city level, the social network structure based around places that was revealed by my analysis in the previous chapter.

4.1.2 Structural properties

I first analyse the structures of the city-level social networks and confirm that they exhibit well-known characteristics of standard social networks: a heavy-tailed degree distribution, small-world properties (high clustering with respect to a random network, and small shortest path lengths), and strong community structure.

4.1.2.1 Degree distribution

Many networks, from those in biological systems such as metabolic networks, to technological structures such as the topology of the Internet and the page structure of the

	N	K	N_{GC}	C	C_r	d	d_r	Q	Q_r
Atlanta	13,011	46,756	11,476	0.16	0.0006	4.6	4.6	0.53	0.17
Boston	10,478	41,505	8,816	0.17	0.0010	4.3	4.0	0.45	0.15
Chicago	19,931	84,778	17,287	0.16	0.0004	4.6	4.9	0.47	0.14
Minneapolis	6,499	30,640	5,914	0.18	0.0016	4.4	4.2	0.41	0.12
Seattle	7,445	28,466	6,392	0.18	0.0008	4.4	4.6	0.46	0.16

Table 4.2: Structural properties of the city-level social networks: Number of nodes N, number of edges K, number of nodes in the giant connected component N_{GC}, clustering coefficient C, clustering coefficient in a random graph of the same size C_r, average shortest path length d, average shortest path length in a random graph of the same size d_r, modularity Q, and modularity of a random graph of the same size Q_r.

World Wide Web, exhibit a heavy-tailed degree distribution, and social networks are known also to have this property [ASBS00, BA99, GN02, WF01]. Figure 4.1 shows the degree distributions of the social networks for the cities in the dataset, which do indeed have the common heavy-tailed shape. Essentially, this means that many individuals have only a small number of social connections, but there are some people with a large number of friends.

4.1.2.2 Clustering

Another commonly observed property of social networks is a relatively high level of clustering, compared to that seen in a random graph of the same size [Wat99]. In terms of social relationships, this corresponds to the fact that many of an individual's friends are likely to be friends with one another. The level of clustering in a graph can be measured by the clustering coefficient. The clustering coefficient C_n of a node n with N neighbours is defined to be the number of links between the N neighbours, divided by the number of possible links that could exist between the neighbours, i.e., $\frac{2N}{N(N-1)}$.

The clustering coefficient C of a graph is then defined to be the mean clustering coefficient of all its nodes. Table 4.2 shows the clustering coefficients in the five city-level networks, and the corresponding clustering coefficients in an Erdős-Renyi random graph with the same numbers of nodes and edges. The values are consistent across all the five networks, being between 0.1 and 0.2. This is much higher, on the order of thousands of times, than the level observed in the random graphs.

4.1.2.3 Average shortest path length

Social networks are known commonly to have a low average shortest path length, which, together with high clustering, makes these networks 'small-world' networks [Wat99]. A shortest path from one node m to another n is defined to be the smallest number of steps that are needed to reach n when starting at m and travelling along edges in the graph. The average shortest path length is then defined to be the mean shortest path length over all pairs in the graph. This quantity is defined only for connected graphs, and so for these networks I consider only the giant component that always contains at least 80% of the nodes in the graph, the presence of which is also common in social networks [KNT10]. Table 4.2 shows that the average shortest path lengths in the giant components of the city-level social networks and the corresponding random graphs are roughly comparable, being between 4 and 5 hops in both cases.

4.1.2.4 Community structure

The final prominent structural characteristic of social networks I consider here is strong community structure. One measure of the strength of community structure in a network is the modularity Q of a partition of a network into its communities. Values of 0.3 or above are generally considered high, and to be indicative of the kind of community structure commonly seen in various biological, technological and social networks [New06]. Table 4.2 shows the values of Q for the best partitions of the city networks found by the Louvain algorithm [BGLL08] (discussed in Section 3.2). These values show that the place-based social networks in the five cities do indeed have strong community structure.

4.1.3 Spatial properties

I now investigate in more detail the spatial properties of the social networks in the dataset, to try to gain insight into the nature of place-based friendships and the places that friends have in common.

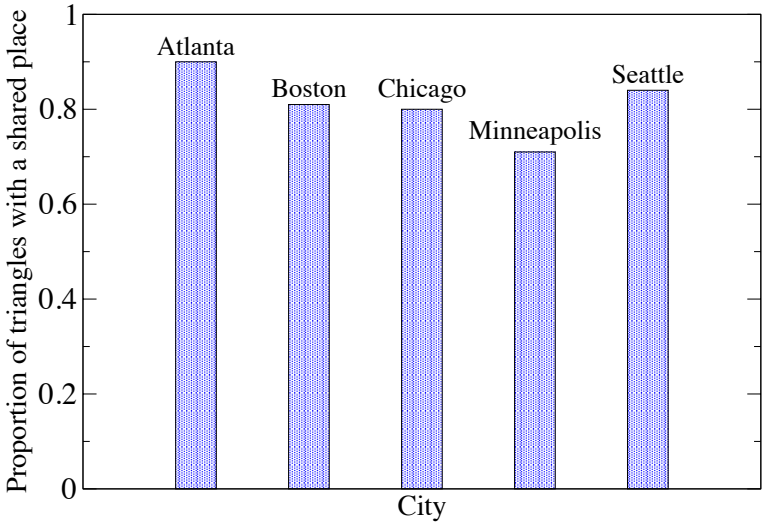


Figure 4.2: Proportions of triangles with a shared place in each city dataset.

4.1.3.1 Spatial clustering

I first consider the groups of three friends u_1, u_2 and u_3 where the ties (u_1, u_2), (u_2, u_3) and (u_3, u_1) all exist in the social graph, and examine the proportion of such triangles where at least one place has been visited by all three people. Figure 4.2 shows the proportions for each of the five cities; in each case, more than 70% of triangles in the social network are such that the three friends have all visited at least one of the same places. This indicates some level of clustering around places in the social networks.

4.1.3.2 Types of places

I now consider more directly face-to-face meetings between friends in the social networks, and examine the places where these meetings take place. Since the dataset does not contain 'check-out' information, I consider a colocation to be represented by check-ins made by each of a pair of Foursquare users within 1 hour of one another.

Figure 4.3 shows the probability that a pair of colocated users are friends according to the type of place where the colocation occurred, as given by the Foursquare category of that place. Some categories (Food, Nightlife and Residence) have a noticeably

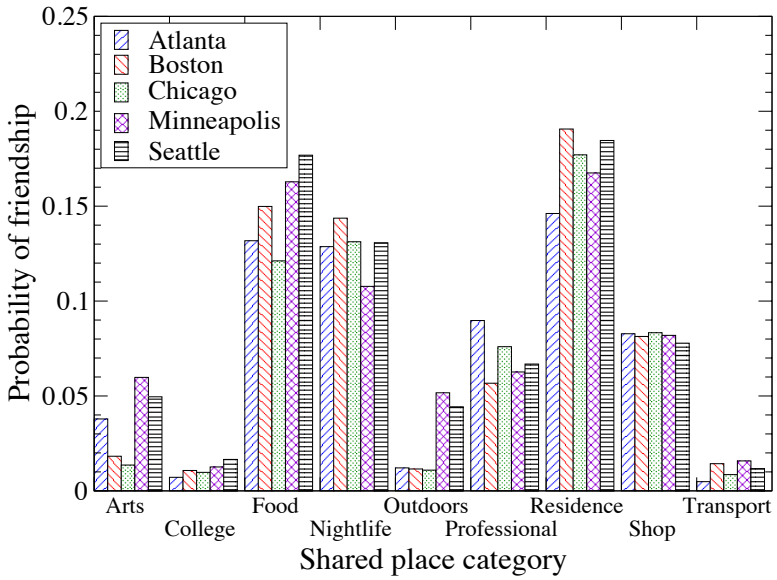


Figure 4.3: Measured probability of friendship, given the category of a shared place.

greater likelihood that colocated pairs are friends than the others do, which suggests that some places have a greater potential for fostering or for reinforcing social ties than others. Specifically, across the cities, the probability that a pair colocated at a Food, Nightlife Spot, or Residence venue are friends is between 0.1 and 0.2. The probability for the categories Professional and Other Places, and Shop and Service, is lower, being between 0.05 and 0.1, and the other categories generally have a very low probability.

One could think of this as dividing places into three categories: 'social' places, where people may tend to meet with their friends, 'semi-social' categories, where some meetings between friends may take place but there may also be many meetings with strangers, and generally 'non-social' places, where people may meet with strangers over friends to a greater degree than in other kinds of places. This is consistent with the observation of Kostakos et al. [KOP+10], in a study involving the analysis of encounters in one city (Bath, UK), that some places play host to rather 'tight' social networks where people are likely to know most of the others present, some places may have 'looser' social networks where there are often groups of friends but members of different groups are strangers to one another, and some places are such that most people located there together do not know one another.

4.1.4 A model for social networks in the city

Having analysed the Foursquare dataset, I now use the above observations to define a model for city-level social networks based around people meeting at places in the city, which produces social networks with the same structural properties as those observed in the datasets. I then explore how the places where people meet one another at places around the city might affect the structure of the social network based around those places, by running the same model without various components.

4.1.4.1 Model definition

The model takes into account three kinds of information about the places in the city:

1. The popularity of a place, i.e., the number of users who have checked in there (Figure 4.4).

2. The geographic (latitude and longitude) coordinates of a place.

3. The semantics of a place and the activities that take place there, as indicated by the Foursquare categories described in the previous section.

The first two kinds of information are relevant to ensure that the mobility of people implied by the model is consistent with the mobility patterns evident in the dataset. Previous research into human mobility patterns has shown that intra-urban mobility tends to show certain patterns [NSL+12], and the mobility implied by the model should not be contradictory to this. The manner in which place popularity and geographic location are used by the model to achieve this is described below. The information about place categories, on the other hand, is used to define how 'social' a place tends to be, given the observations in the previous section, and can influence the likelihood that social ties form within a place according to its category.

In the light of the above analysis, the main idea of the model is to generate a place-based social network, where people visiting the same places have some probability of forming a social tie, and there is also some level of clustering around places resulting in the intra-place triangles seen in the dataset. The model first assigns each person to a place, preserving the place popularity of the original dataset so that the implied visits of people to places are not inconsistent with the patterns shown by the data (Step 1 below).

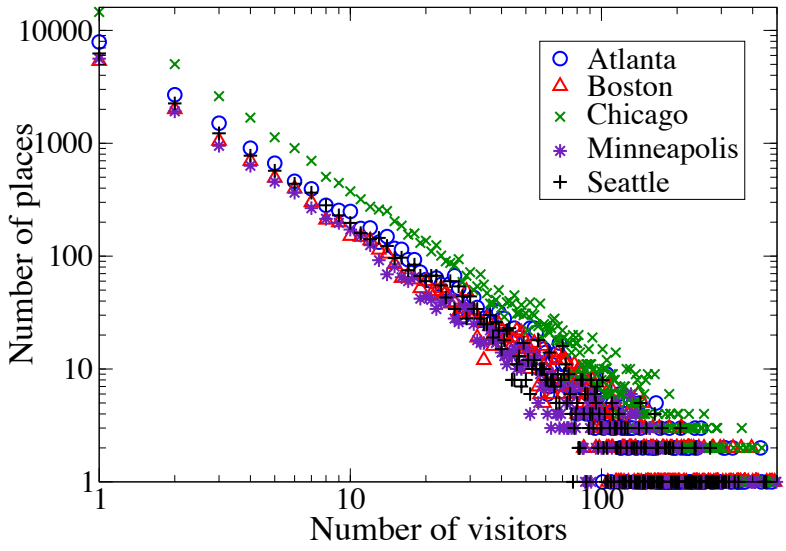


Figure 4.4: Numbers of people checking in at a place in the dataset.

Each person is then assigned to further places, preserving the distribution of places per person in the original data, and also the distribution of distance between places visited seen in actual mobility traces (Step 2 below). Ties are then placed between people who share the same place with probability dependent on the category of the shared place as suggested by Section 4.1.3.2 above (Step 3), and finally intra-place triangles are closed as suggested by Section 4.1.3.1 (Step 4). The specific procedure to generate a city-level social network is thus as follows:

1. Begin with the sets V of people and M of places in the dataset for the city.

2. For each person u, consider them to have visited n places, where n is sampled from the distribution of places per user in the dataset for the city. Assign to u an initial place from the set M of all the places in the city, chosen with probability proportional to the popularity of that place.

3. For each of the $n-1$ additional places that u has visited in the dataset, assign to u another place m from M, with probability qr^{α}, where q is proportional to the popularity of m in the dataset, and r is inversely proportional to the rank distance of the place from the first place of u. The rank distance of a place m_2 from place

m_1, $rank_{m_1}(m_2)$, is the position of m_2 in the list of places in descending order of geographic proximity to m_1. Given a set of places M in a city, the probability of a person visiting place $m_1 \in M$ given that they have visited place $m_2 \in M$ is defined to be:

$$r \propto \frac{1}{rank_{m_1}(m_2)}$$

where the rank distance function for place m_1, $rank_{m_1}$, is defined over other places m_2 as:

$$rank_{m_1}(m_2) = s + 1$$

where s is the number of places in the city closer to m_1 than m_2 is to m_1. This method of place assignment is designed to reproduce observed patterns of intra-city human mobility. I use the exponent $\alpha = 0.84$ in accordance with other research concerning human mobility in cities, which has shown that people's movements within a wide range of cities of varying sizes and densities tend to follow this distribution [NSL+12].

4. For each place m in M, for each pair of people (u_1, u_2) who have been assigned place m:

   (a) Place a social tie between u_1 and u_2 with probability

   $$p = \begin{cases} p_{cat}(m) & \text{if } pop(m) \leq 30 \\ 0.001 & \text{otherwise} \end{cases}$$

   according to the category of m: $p_{cat}(v) = 0.15$ for 'social' places (Food, Nightlife Spot, and Residence), $p_{cat}(m) = 0.08$ for 'semi-social' places (Professional and Other Places, and Shop and Service), and $p_{cat}(m) = 0.01$ for all other places, in the light of the analysis in the previous section. $pop(m)$ is the popularity of m. The choice of the threshold 30 used in this step was found to be necessary experimentally, and is discussed in Section 4.1.4.3.

   (b) For each of u_1's existing friends f in the social network who have visited m, place a link between f and u_2 with probability 0.15, in line with the 'social' probability. This implements an intra-place triangle closing mechanism as suggested by the observation in the previous section that the

City	C	d	Q
Atlanta	0.13	4.1	0.44
Boston	0.14	4.0	0.37
Chicago	0.14	4.2	0.40
Minneapolis	0.15	4.0	0.39
Seattle	0.16	4.0	0.39

Table 4.3: Structural properties of synthetic place-based networks: average values, over 10 runs of the model for each city, of the network clustering coefficient C, average shortest path length d, and modularity Q.

majority of triangles in the networks have one common place shared by all three members.

4.1.4.2 Properties of synthetic networks

I now present the properties of the synthetic networks generated using the procedure described above. I show that the model produces networks with the structural properties observed in the real data. The results given for each city are the average of 10 runs of the model. I study the effect of each piece of information by removing each one in turn, running the model, and examining the properties of the resulting networks: first, I analyse the effect of intra-place triadic closure, and then examine the effect of the place information including distance (place location), place category, and place popularity.

The results of running the full model, incorporating information about distance, place semantics, and including the triangles between people meeting friends of friends, are shown in Figure 4.5 and Table 4.3. Figure 4.5 shows that the full model produces networks with a heavy-tailed degree distribution, as seen in the actual social networks. The average exponent across the real networks was measured as 2.28, while that for the synthetic networks is 2.51. Table 4.3 gives the values of the clustering coefficient, average shortest path length, and modularity of the synthetic graphs; they can be seen to display the structural properties of the empirically observed networks: high clustering (between 0.1 and 0.2), short path lengths (around 4 hops), and strong community structure (as indicated by a modularity value Q higher than 0.3). The values for the synthetic networks can be seen to be similar to those for the networks constructed using the datasets that were given in Table 4.2. The first row of Table 4.4 shows the averages of these values, and the other rows in the table show the differences that are

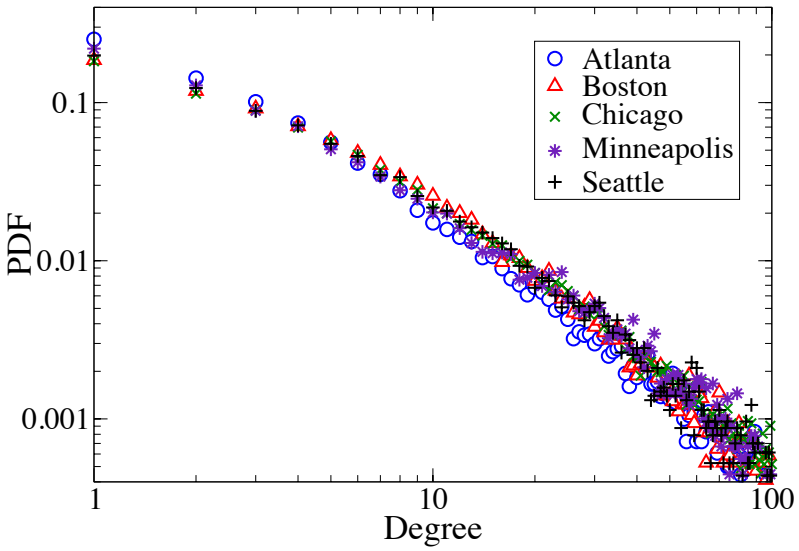


Figure 4.5: Degree distributions of synthetic graphs.

Model	C	d	Q
Full	0.14	4.0	0.40
No triadic closure	0.05	4.0	0.38
No distance	0.13	3.8	0.33
No categories	0.01	3.8	0.29

Table 4.4: Average values of the clustering coefficient C, average shortest path length d, and modularity Q for the synthetic networks, when including all information in the model, and when running it without one kind of information.

observed when running the model without one of its various components.

The second row of Table 4.4 shows that in the absence of the triadic closure mechanism, when social ties are placed only according to the probabilities of connection at given places according to their category, clustering in the graph is unsurprisingly much lower than in the actual networks. Triadic closure is a concept often explained in terms of social balance theory [CH56], in that if an individual has two friends who themselves are not friends, more psychological stress will result than if those two friends are also friends and the relevant triangle exists in the social graph. One can imagine that this effect would be particularly strong if the individuals concerned were actually meeting face to face, such as would be likely when those three individuals are

connected by the sharing of a common place.

The third row of the table shows that running the model assigning places to users using only the popularity of the place (instead of using the popularity combined with the rank distance) reduces the modularity value Q of the resulting networks, which implies weaker community structure. This might be understood in terms of people being more likely to visit, and therefore to meet with friends at, places nearby to locations they already visit. When distance is not taken into account in the model, this additional clustering effect based on geography disappears and leaves only clustering at single places produced by triadic closure, which weakens the community structure of the network.

Finally, the last row of the table shows the results of running the model without the category-based probabilities and instead using in each case a uniform probability that will produce the same number of ties in the resulting network. This dramatically reduces the clustering coefficient C of the resulting network, even when the triadic closure mechanism is present. This may be because the clusters of social ties tend to be around the more sociable places where friends tend to meet, that have higher probabilities for intra-place ties in the full model. Even when the same number of ties are placed, if these are not focused around 'social' places but spread over all the venues with equal probability, the network does not display the high clustering characteristic of a social network. This might suggest that the use of particular spaces around the city for social and non-social purposes could affect, or indeed be affected by, the group structure present in the social network. However, this is at present entirely theoretical and further research would be required in order to determine the exact mechanisms, if any, at work.

4.1.4.3 Discussion

The mechanism of assigning people to places based on popularity and forming intra-place ties has much in common with the preferential attachment model [BA99], which was proposed to produce scale-free networks by means of having new nodes connect to existing nodes with probability proportional to their degree. That is, the graphs grow in a 'rich-get-richer' fashion; the more neighbours a node already has, the more likely it is that it acquires more neighbours, which produces graphs with a power-law degree distribution. The model presented above can be seen as employing a version of

preferential attachment to places: more popular places are more likely to attract more people, and therefore more potential pairs of friends.

However, the model uses not only place popularity, but takes into account the categories of places to determine the likelihood of intra-place ties. The probabilities of social tie formation based on the categories of places are used when the places in question have 30 or fewer visitors, and a lower probability for all more popular places. In experiments, I found that this was necessary in order to be able to obtain networks with the correct community structure. This might intuitively be considered to be a manifestation of the idea of the *constraint* of a place [Fel81]. A place with high constraint forces individuals there to interact much and often (for example, a family home) resulting in more or stronger social ties than a place with low constraint where many individuals may go, but will probably not encounter one another, or where visitors to the place do not typically interact with many others (for example, a park).

The threshold of 30 is the same as that found in the study by Onnela et al. of the geographic span of communities in a mobile phone communication network [OAG+11]. The authors observed that communities of 30 or fewer members tended to be geographically tight, with a 100% increase in the geographic spread of communities occurring as the number of members increased from 30 to 40. This might suggest that communities at the intra-city level, such as those observed in the place-based social networks under study, are intrinsically constrained in size in order to be sustainable within the city. This idea is related to previous studies that have found an apparent upper limit of around 30 on the maximum number of more intense relationships that a person may be able to maintain, with contacts who are not part of this 'inner circle' being communicated with less frequently [HD03, SDBA12]. In the city-level social networks, one way in which this threshold could manifest is as a limit on the maximum number of people with whom one is able to maintain regular face-to-face contact at a place, reflected in the size of close-knit groups based around places. As suggested by the results shown here, this might affect the community structure of the resulting social network.

4.2 Face-to-face meetings at places in the city

Having studied the relationship between places around the city and ties in the social network in terms of the places that friends have in common, in this section I examine in more detail the venues where meetings between friends take place. I examine characteristics of meeting locations such as the distance from the places most frequently visited by those users, their categories, and whether those users or their friends have checked in at those places previously. I then expand my analysis to consider larger groups, of three or more friends. Due to data sparsity, I am not able to analyse the behaviour of larger groups in Foursquare, but I show that there are similarities between the behaviour of pairs seen in Foursquare and that observed for larger groups in a dataset from a mobile phone network: users are more likely to visit places near their usual locations when with friends, and are more likely to visit places where people in the wider social circle of members of the group have also been. It appears that the locations visited by people when with their friends tend to differ somewhat from those where they go on their own, and in the last section of this chapter, I will demonstrate an application of this behaviour that could be useful in location-based online social networks, namely, improving the performance of the common task of venue recommendation.

4.2.1 Pair meeting behaviour

I study the places visited by individuals and by pairs of friends as recorded by their check-ins on Foursquare. The dataset was downloaded in the same way as those described in Sections 3.1.2 and 4.1.1, and contains 2,315,350 check-ins made to 109,314 venues, by 104,266 users in New York City. The social network, constructed by taking users to be friends when each follows the other on Twitter (as in Section 3.2.1.1), is an undirected graph with 99,725 nodes and 821,948 edges.

I investigate *social check-ins* in the dataset, defined to be check-ins where the user making the check-in can be assumed to have been colocated at a venue with one of their friends in the social network. I employ the usual definiton of colocation discussed in Section 3.3.2, where a pair of friends is considered to have been colocated when they have checked in to the same venue within one hour of one another.

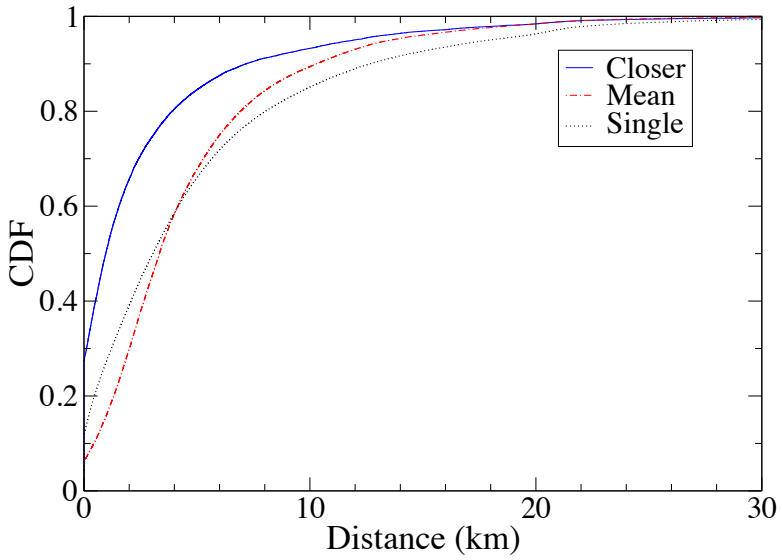


Figure 4.6: Distance of social check-in venues from the checking-in user's top locations.

4.2.1.1 Distance from frequently visited locations

I analyse the distance of the locations of check-ins from a user's most frequently visited location. I define a user's *top location* to be the Foursquare venue where they have previously checked in the greatest number of times, and compute the distance between this location and the venue where a check-in takes place. In the case that a user has more than one top location, all of those locations are taken into account, and I compute both the mean distance from the check-in venue to those locations, and the minimum distance from the check-in venue to one of those locations. Figure 4.6 shows the cumulative distribution function (CDF) of the distance of a social check-in venue from a user's top locations, and for comparison, the distribution of the distance of the venue of all check-ins (not just social check-ins) to the closest of the top locations of the user making the check-in. The results show that social check-ins tend to be closer to a top location of one of the friend pair concerned than do check-ins in general.

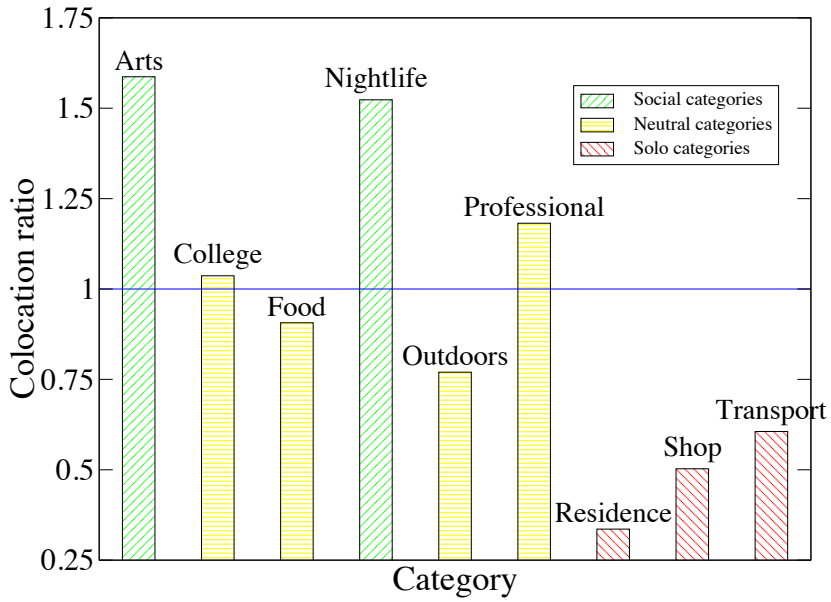


Figure 4.7: Representation of social check-ins in each category, compared to all check-ins. 'Social' categories have social check-ins over-represented by more than 25%, 'neutral' categories have social check-ins represented within 25% of the proportional value, and 'solo' categories have social check-ins under-represented by more than 25%.

4.2.1.2 Venue categories

I next use the information from Foursquare about the categories of venues to analyse the types of places where users tend to go with their friends. As described above, Foursquare defines 9 broad categories[1] for venues: Arts and Entertainment (e.g., cinemas, music venues), College and University (e.g., schools, university buildings), Food (e.g., cafés, restaurants), Nightlife (e.g. bars, clubs), Outdoors and Recreation (e.g., parks, nature spots), Professional (e.g., workplaces), Residence (e.g., homes), Shop and Service (e.g., shops, hospitals, churches), and Travel and Transport (e.g., railway stations, airports). I compute the ratio of the proportion of social check-ins in each category to the proportion of all check-ins in that category. Formally, define:

- $cat\_checkins(c)$ to be the total number of check-ins to category c in the dataset.
- $cat\_social\_checkins(c)$ to be the number of social check-ins to category c.
- $cats$ to be the set of all categories defined by Foursquare.

Then for each category c in $cats$:

$$proportion(c) = \frac{cat\_checkins(c)}{\sum_{c' \in cats} cat\_checkins(c')}$$

That is, the proportion of all check-ins that occur in category c, and

$$social\_proportion(c) = \frac{cat\_social\_checkins(c)}{\sum_{c' \in cats} cat\_social\_checkins(c')}$$

That is, the proportion of all social check-ins that occur in category c. I then quantify the propensity of each category to include venues where social check-ins take place by defining the colocation ratio:

$$colocation\_ratio(c) = \frac{social\_proportion(c)}{proportion(c)}$$

That is, the ratio of the proportion of social check-ins taking place in that category to the proportion of all check-ins taking place in that category. If this value is 1, it means that the category in question is equally likely to host both solo and social

[1] There is in fact a tenth 'Uncategorized' category for when venues have not been assigned a category, but these venues are excluded from the dataset.

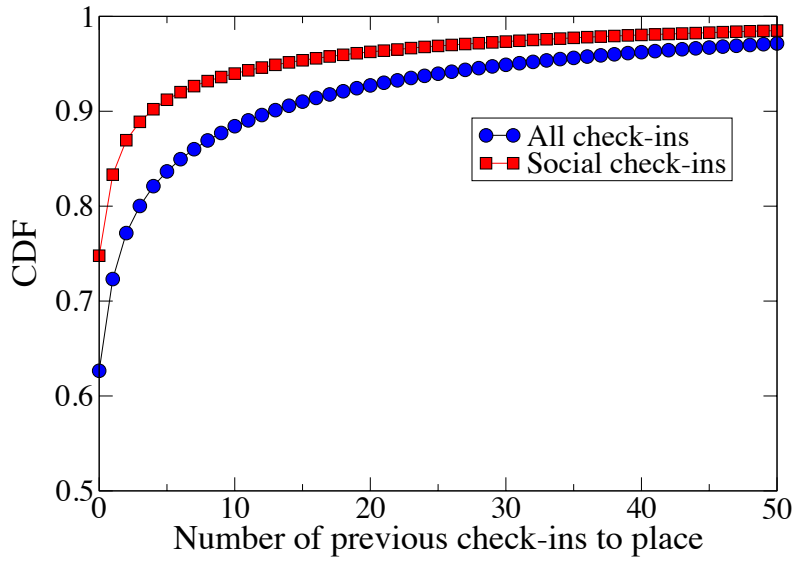


Figure 4.8: Social check-ins more often take place at venues new to the checking-in user than check-ins in general.

check-ins. If the value is markedly less than 1, the category is less likely to host social check-ins than it is to host solo check-ins, so places in this category might be less good to recommend for visits by pairs of friends. If the value is much more than 1, the category is more likely to host social check-ins than it is to host solo check-ins.

I compute $colocation\_ratio(c)$ for each category c. Figure 4.7 shows that more than 1.5 times the proportion of social check-ins are to venues in the Arts and Entertainment and Nightlife Spot categories than the proportion of check-ins in general that are in these categories. Meanwhile, less than 0.7 of the proportion of social check-ins are to Residence, Shop and Service, and Travel and Transport venues than the proportion of check-ins in general.

4.2.1.3 Previous check-ins

I examine how likely users are to visit new places with their friends. Figure 4.8 shows the cumulative distribution function (CDF) of the number of previous visits in the dataset by the checking-in user to the visited venue, for all check-ins and for social check-ins. Social check-ins are more likely to take place at new venues than check-

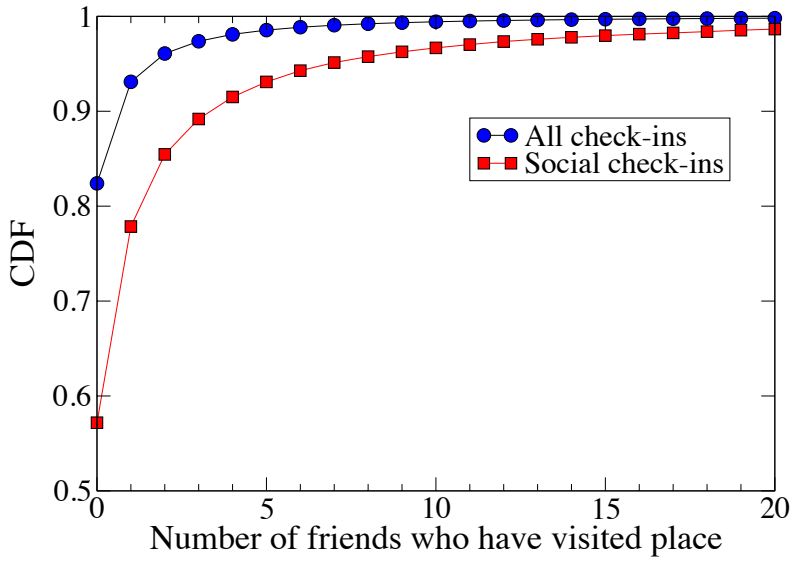


Figure 4.9: Social check-ins more often take place at venues where the checking-in user's friends have been than check-ins in general.

ins in general; the probability that a social check-in takes place at a previously visited venue is about 0.25, compared to about 0.38 for any check-in.

4.2.1.4 Check-ins by friends

Finally, I investigate how many social check-ins take place at venues where members of the friends' wider social circles have previously checked in. Figure 4.9 shows the cumulative distribution function (CDF) of the number of a user's friends who have previously checked in to the venue in question, for all check-ins and for social check-ins[2]. Social check-ins are more likely than general check-ins to take place at a venue where a user's friends have been before. About 18% of all check-ins are to a place visited by at least one friend before, but about 43% of social check-ins being to such venues. At first glance, this may seem contradictory to the observation that social check-ins tend to take place at venues where the user has not been before, but in fact it is the case that pairs of friends tend to check in together at venues that are new to

[2]Note that previous check-ins by friends are required to be at least an hour before the check-in in question, to avoid counting the first social check-in in a pair constituting a colocation event.

the pair in question, but not to their wider social circle. It tends to be not one of the pair that has checked in to the venue before, but other friends of one or both of the colocated friends.

4.2.2 Group meeting behaviour

I now examine the venues checked into by groups of 3 or more friends. There are not enough such group colocations in the Foursquare dataset, due to data sparsity, but I instead analyse a data from a mobile phone operator. Mobile phone usage is more widespread than that of location-based social networks at present, and so this dataset does not suffer from the same problem. Although the mobile phone dataset does not include venue category information and does not allow unique identification of venues, instead having localisation at the level of individual cell towers, some of the same trends can be seen as in the Foursquare dataset for pairs. In addition, previous research has shown that the mobility patterns apparent from LBSN check-ins show strong similarity to those seen in other kinds of mobility data including cellphone records [CML11].

The telecoms dataset is a large, anonymised set of billing records for over one million mobile phone users in Portugal, gathered over a twelve month period between 2006 and 2007. The dataset contains information about mobile phone calls, but does not contain text messages (SMS) or data usage (Internet). In order to preserve privacy, individual phone numbers were anonymised by the operator. Each user in the anonymised dataset is identified by a hashed ID. Each entry is a CDR (Call Detail Record), consisting of a timestamp, the IDs of the caller and the callee, the call duration, and the cell tower IDs of the caller and callee towers. The dataset also includes the latitude and longitude of the cell towers, which allows us to study the relationship between the social network constructed by placing edges between nodes representing people whenever one person has called another, and the physical location of the users. The social network has 1,954,188 nodes and 19,370,004 edges.

I extract colocated friendship groups by considering as colocated people who make calls using the same cell tower within one hour of one another, in agreement with the definition of social check-ins in the Foursquare dataset. Within a temporal window of one hour, I consider as groups the connected components of the subgraph of the social

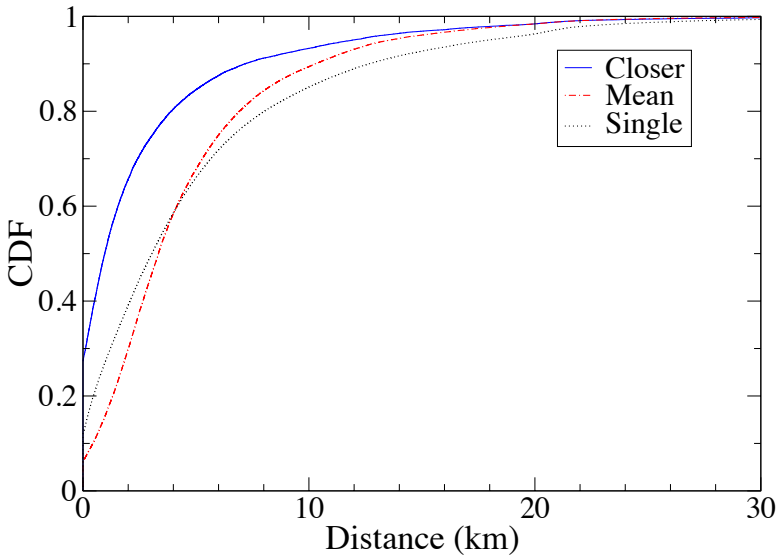


Figure 4.10: Distance from each group member's top locations of group colocations in the telecoms dataset.

network that contains only edges between people colocated during that hour.

Note that while for the purpose of these analyses the telecoms data are treated in the same way as the check-in data, the two datasets are quite different in nature, which could result in different observed behaviour based on people's usage of the respective services. For example, check-ins can be considered to be entirely voluntary, because users check make an active choice to check in and perform the action using an application on their phones. A mobile phone call, on the other hand, while resulting in a location record similar to a check-in, may not be voluntary to the same extent; the location record is a side-effect of a call having a different purpose from simply recording location as with the check-in, and calls may also be received without that person choosing to be called, in a different manner from Foursquare users having to choose to check in. Therefore it is important to remember that the mobile phone data cannot be taken to be an extension of the Foursquare data, even though the analysis performed here is the same and it is interesting to compare the findings in each dataset.

4.2.2.1 Distance from frequently visited locations

Figure 4.10 shows the cumulative distribution function of the distance of the colocation of a cellphone user with a group of their friends from that users top locations. The figure also shows the distribution of the distance of all calls from the caller's top locations (not just social colocations). Similarly to social check-ins in Foursquare, group colocations are more likely to take place near to one of the group's top locations than calls in general. Note that the higher values of the CDF at very small distances compared to Foursquare is due to the coarser-grained spatial resolution of the cellphone dataset; in Foursquare we have individual venues for check-ins, giving single-building accuracy, whereas in the cellphone dataset we are limited to the resolution of cell towers[3].

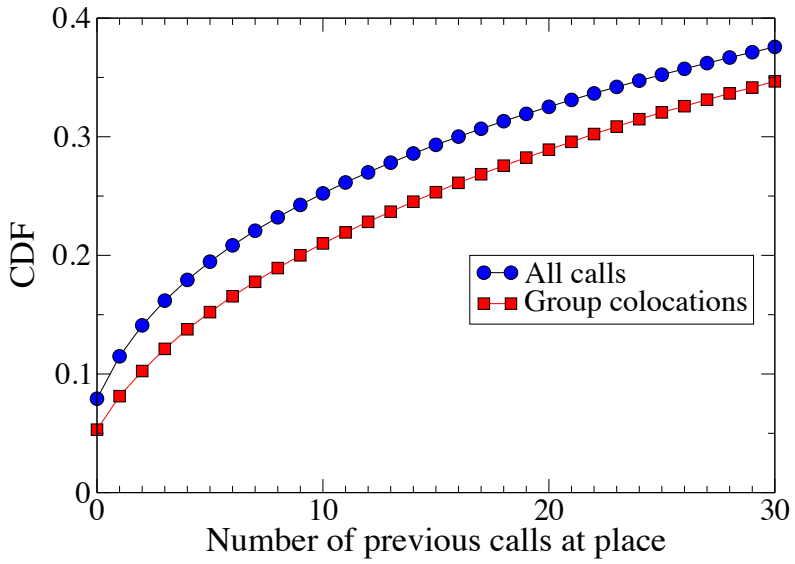
4.2.2.2 Previous visits

Figure 4.11a shows the CDF of the number of times a member of a colocated group has previously been seen at the place of colocation. Contrary to the behaviour of pairs in Foursquare, an individual is less likely to meet a group of their friends at a new place than at a place where they have been before. I investigate this phenomenon further by analysing separately groups of different sizes: pairs, trios, quartets, and quintets. Figure 4.11b shows the CDF for groups of each size. The results are in agreement with the results from Foursquare: pairs are more likely to meet at new places than people are to call from new places in general, but an individual is likely to meet a larger group somewhere they have been before.

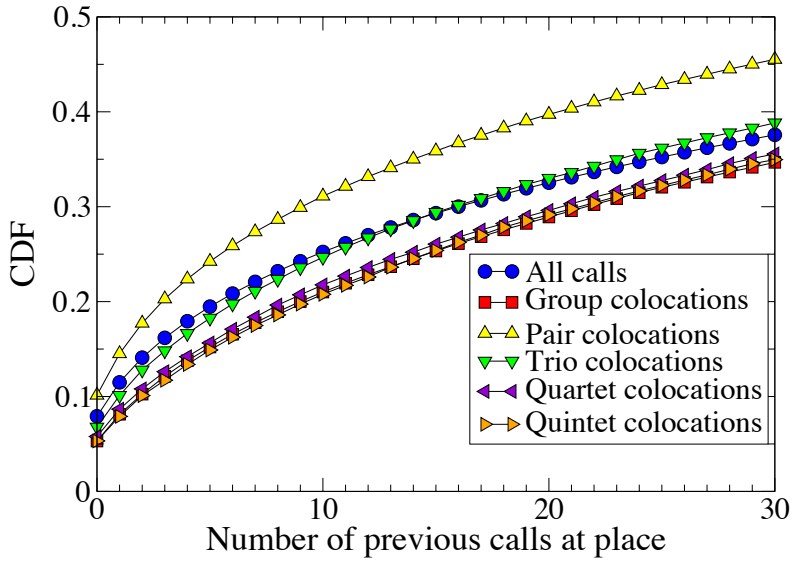
4.2.2.3 Previous visits by friends

Finally, Figure 4.12 shows the CDF of the number of an individual's friends, excluding the group with whom they are colocated, who have been at the colocation place before. Again in agreement with the Foursquare data, a user is likely to meet a group at places where their wider circle of friends have been previously, compared to somewhere where none of their friends have been.

[3] In urban areas such as Lisbon, each cell tower covers an average of 0.13 km^2.



(a) Cumulative Distribution Function (CDF) of number of previous visits by group members.



(b) Breakdown of previous visits by group size.

Figure 4.11: Previous visits by group members to colocation locations in the telecoms dataset.

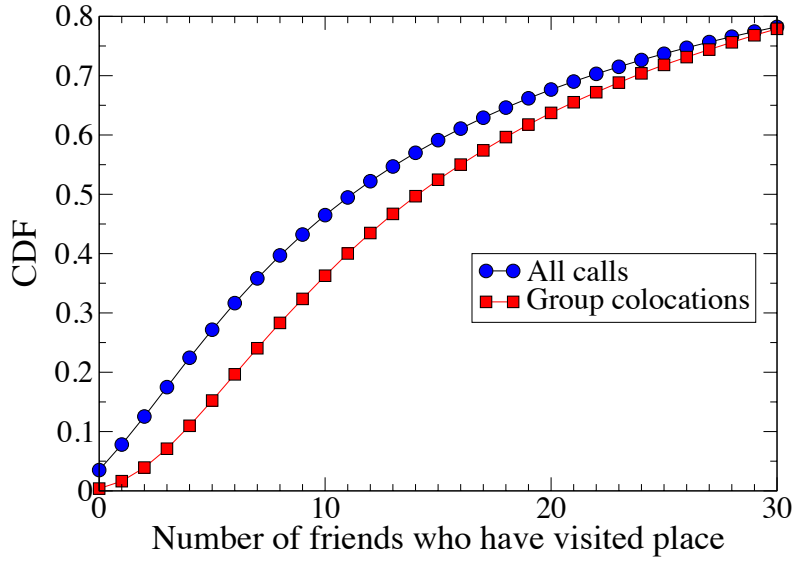


Figure 4.12: Cumulative Distribution Function (CDF) of number of previous visits by group members' friends to colocation locations in the telecoms dataset.

4.2.3 Discussion

In both datasets analysed in this section, it seems that friends are likely to meet closer to one of the individuals' familiar locations than people may go in general. This is in agreement with the finding by Calabrese *et al.* [CSBR11] that as the distance between the homes of pairs of colocated users increases, colocations take place in an area closer to one of the homes of the pair.

The finding that meetings between friends are over-represented in some categories such as Entertainment and Nightlife, and under-represented in others, such as Shop and Transport, gives some indication that people are more likely to visit some categories of places with their friends than others. This could have implications for place recommendation in online location-based social networks, as I will demonstrate in the rest of this chapter. In both datasets, it is also the case that people are more likely to be colocated with friends at places where their other friends not part of that group have been before.

My analysis has also shown differing behaviour between pairs of friends (both in Foursquare and in the mobile phone dataset) and larger groups: while an individual is

more likely to travel to a new place with a single friend than they are on their own, the opposite behaviour is seen where larger groups are concerned. One way to interpret this could be in terms of research that has shown that people are more likely to take risks when with peers [GS05], possibly including visiting a new place. Being with a friend may increase confidence and willingness to explore. However, this must be balanced, in a larger group situation, with the fact that the larger a group, the more difficult it is for that group to coordinate [ILGP74]. This could be a reflection of greater ease for a group to meet at a place familiar to all of its members, than to identify and agree on a new meeting place.

4.3 An application: group venue recommendation

An important problem for LBSNs is that of venue recommendation: suggesting to users places that they might enjoy visiting. Recent research has generated a variety of systems for performing this recommendation task [BS11, NSLM12b, YZYW13, YYL10]. However, these systems tend not to distinguish between the situations when users intend to visit a place by themselves, and when they are looking for somewhere to go out with a friend. Given the differences between solo and social check-in behaviour seen in the above analysis, it seems likely that LBSNs could benefit from exploiting information about the friend a user is with to alter venue recommendations specifically for a social meeting place.

The task of group recommendation – that is, making recommendations for a group of two or more people, rather than for an individual – is often addressed by computing recommendations for the individuals in the group, and then merging the resulting lists of items [GLRW13]. However, this approach does not consider that the behaviour and preferences of individuals may be different in a group scenario from when they are by themselves. In this section, I demonstrate how making recommendations for venues by considering groups as a whole, rather than by combining recommendations made for their individual members, could improve recommendation performance in an LBSN when users are looking for venues to visit with their friends.

4.3.1 Problem definition

The group venue recommendation problem can be formulated as a prediction problem as follows: given a group of colocated friends and a list of candidate venues, using information in the system up to the time of the colocation, produce a ranking of the candidate venues where the first in the list is the most likely colocation venue, the second the second most likely, and so on.

The aim is as frequently as possible to rank the actual colocation venue in the top N predictions, for reasonably small N, making the assumption that a correct prediction of the colocation venue would mean that having recommended that venue to the group would have counted as a successful recommendation.

4.3.2 Datasets

I use the dataset from Foursquare described in Section 4.2.1, and the telecoms dataset described in Section 4.2.2. In the following, I refer to 'check-ins' in both datasets, even though it is only the Foursquare data that contain actual check-ins made using a location-based social network application. In the case of the mobile phone network, a 'check-in' is simply the same type of information as recorded by a Foursquare check-in, i.e., user ID, location ID, and timestamp, but with reference to a phone call rather than an LBSN check-in. A phone call between two users is taken as equivalent to a check-in by each of those users at their respective locations at the call start time given by the CDR. Colocation events are inferred in the same way as described in the previous analysis: see Sections 4.1.3.2 and 4.2.2.

4.3.3 Prediction features

I use a machine learning approach to prediction, as has been demonstrated effective for single-user location prediction by other research [CS11, NSLM12b, NSLM12a]. I define prediction features based on check-ins to candidate venues. These features are of three kinds: *global features*, which are computed using information specific to the candidate venue but not to any user or group of users, *single-user features*, which are computed using information about check-ins by one user to the candidate venue, and *group features*, which are computed by combining information about check-ins to the

candidate venue by all of the users in a group. This last kind of feature has not been examined by previous work in the area, and I demonstrate here the potential benefit for venue recommendation in LBSNs of using groups in this way.

4.3.3.1 Global features

These features are defined on a candidate venue v and time t of the colocation[4], but do not take into account information about individual users:

- `pop_checkins`: The number of check-ins to v before time t, as a measure of venue popularity.
- `pop_checkins_c`: The number of social check-ins (Section 4.2.1) to v in the dataset before time t.
- `cat`: The Foursquare category of v. This feature aims to reflect the finding in the previous section that venues of some categories are more likely to host social check-ins than others. Note that this feature is not defined for the mobile phone dataset, due to lack of category information for locations.

4.3.3.2 Single-user features

These features are defined using information about a candidate venue v and time t of the colocation, as well as further information about an individual user u. These types of features have previously been used profitably in venue prediction in LBSNs [CS11, NSLM12a, NSLM12b]. In the following, I will show the difference between including in feature vectors the feature's value for each of the users in a pair of Foursquare friends for whom a prediction is made, and for combining this information into a group feature. Note that this comparison is not possible for groups in the telecoms dataset since the group size is not fixed.

- `single_dist_closer`: The distance from v to the closest of the user's top locations before time t (see Sections 4.2.1.1 and 4.2.2.1).

[4]This is taken in the case of Foursquare to be the time of the first check-in by a member of the colocated group constituting that colocation event, and in the case of the telecoms data to be the time of the start of the first call that is used to infer the colocation event.

- `single_dist_mean`: The mean distance from v to the user's top locations before time t. Together with `single_dist_closer`, this feature aims to encode information about the distance to the locations that a user visits the most, given that users are more likely to check in closer to places they usually visit.

- `single_hist_total`: The total number of check-ins to v by the user before time t. This feature encodes information about which locations the user tends to visit, which has previously been shown to perform very well in venue recommendation for individual users [NSLM12b].

- `single_social`: The total number of check-ins before time t to v by any of the user's friends. This feature encodes information about which locations the user's friends tend to visit, which has also previously been found to be helpful in making venue recommendations to Foursquare users [NSLM12b].

4.3.3.3 Group features

These features are defined on a candidate venue v and time t of the colocation, and also encode information about a *group* of friends, which aims to enable the use of differences between social and solo check-in behaviour shown by the above analysis to improve group recommendations over those one can make by combining single-user recommendations.

- `group_dist_closer`: The distance from v to the closest of any of the group's top locations before time t.

- `group_dist_mean`: The mean distance from v to the group's top locations before time t. In conjunction with `group_dist_closer`, this feature aims to characterise a target venue in terms of not only the distance from each of the users' top locations but in terms of a combination of the distances, since according to the above analysis, users stay closer to frequently visited locations when with a friend than in general.

- `group_hist_total`: The total number of check-ins to the v by users in the group before time t.

- `pair_hist_min`: The minimum number of check-ins to v by a user in the group before time t. In conjunction with `group_hist_total`, this feature aims to combine information about the users' previous visits to a target venue.

- `group_social`: The total number of check-ins to v before time t by friends of the users in the group who are not also in the group. If users in the group have a mutual friend who is not also in the group, this friend's check-ins to v are counted only once.

- `group_mutual`: The total number of check-ins to the v before time t by other users who are friends with *all* the users in the group, but who are not in the group themselves. Each of these friends' check-ins are counted only once.

In the following, I will demonstrate and quantify the benefit in using these group-based features in group venue recommendation, over the performance provided by using global and single-user features alone.

4.3.4 Methodology

I train an M5 model tree predictor [Qui92, WW97] to use in evaluation of the features. I chose this method of prediction having experimented with various machine learning algorithms provided by the WEKA machine learning framework [HFH+09] and found that M5 model trees provided the best results of these methods[5]. This is in line with the results of other research into venue recommendation as a prediction problem [NSLM12a]. M5 model trees create decision trees by dividing training examples according to the values of the features, specifically, they split training examples on the attribute that maximises expected error reduction. At the leaves are linear regression models that use some combination of features to output a numeric score that can be used to rank the candidate venues.

In the case of each dataset, I choose a time t_0 corresponding to the point 90% of the way through the data collection period for that dataset. I train prediction models

[5] In particular, I trained predictors based on WEKA's implementations of M5 trees, linear regression, random forests, and support vector machines, and tested these by training each algorithm using the training set and assessing the correlation coefficient and root mean square error (RMSE) according to 10-fold cross-validation on the training set. The M5 trees produced the best results, giving a correlation coefficient of 0.8405 and RMSE of 0.271, while the linear regression did the worst, with correlation coefficient 0.0969 and RMSE of 1.2623.

using features generated from data up to the time t_0, and leave data from after t_0 to be used in testing the trained models. For each colocation event that took place before t_0, I build a training example feature vector $\mathbf{x}$ that encodes features of the visited venue and the group of colocated friends, and give it the label 1. I build a corresponding negative training example $\mathbf{x}'$, with features defined on the same group of friends, but on a randomly chosen venue where the colocation did *not* take place, and give this the label 0. The training set consists of pairs of positive and negative examples, labelled 1 or 0 accordingly. This effectively reduces the ranking problem to a binary classification task [CSS99], and the trained predictor is able to output values for an input test example where the higher the predicted value, the more likely, according to the predictor, that the colocation of the friends in the example took place at the example venue.

I evaluate the trained models on group colocation events not used in training. For each colocation event I compute a feature vector $\mathbf{x}$ with features defined for the group of friends in question and the colocation venue. I also compute feature vectors for the same group and 1,000 other randomly chosen venues in the relevant dataset. The same candidate set of venues is always used for the same colocation event when testing different combinations of features. I present all of these feature vectors to the trained predictor, which outputs a numeric score for each vector, with higher scores indicating more likely colocations for the group according to the trained model, thus providing a ranked list of the candidate venues. Given the colocation events and their associated ranked lists of venues, I compute Average Recall @ N, that is:

$$AR(N) = \frac{\text{\# correct venues in top } N \text{ of ranking}}{\text{\# colocation events}}$$

for varying values of list size N. In the intended application scenario, reasonable performance would be desired for small N (as a guide, perhaps 10 – 20), since users on mobile devices do not want to have to search through large lists of recommendations to find a venue they would like to visit.

I also train a single-user predictor to make recommendations for individual users, in order to be able to compare recommendations made directly for groups with the results of combining individual recommendations after ranking. To do this, I compute feature vectors containing all of the global features except for `pop_checkins_c` (to evaluate the results of using a predictor not specialised for colocation check-ins to

make individual recommendations), and all of the single-user features, for individual check-ins in the dataset up to time t_0, assigning the label 1 where the user did check in to the venue concerned and 0 when the venue has been randomly chosen from the rest of the venues in the dataset. I take the same number of positive and negative example pairs as were used to train the pair-based predictor, and train a model using these to make predictions for a single user.

I quantify the extent to which the differences between solo and social check-in behaviour affect the performance of recommending venues to groups looking for somewhere to go together. I compare direct recommendation using the group features to the approach of taking the two ranked lists from the single-user predictor (Section 4.3.4) for each user of the pair and combining them, as is a common approach in group recommendation [GLRW13]. I present the same lists of candidate venues to each predictor for each of the test examples, to enable comparability of results. Specifically, I test two methods of combining individual recommendation lists, referred to as *Combined single 1* and *Combined single 2*, and defined as follows:

Combined single 1: I compute Average Recall @ N for varying values of N, where this is defined for the combination of individual user predictions as:

$$AR(N) = \frac{\text{\# correct venues in top } N \text{ for all users in a group}}{\text{\# colocation events}}$$

Combined single 2: I take the ranked lists for all of the individual users in the group, compute the mean rank of each candidate venue across these lists, create a new list with the venues in order of mean rank, and compute Average Recall @ N for varying values of N as:

$$AR(N) = \frac{\text{\# correct venues in top } N \text{ of mean-ranks list}}{\text{\# colocation events}}$$

4.3.5 Results

I now present the results of testing the trained models for colocation venue prediction. I first consider the results for pairs in Foursquare, which enables me to examine the difference between including the single-user features for both users in a pair in feature

vectors, and including a group feature intended to encode the same kind of information. I also study the contribution of different kinds of features (popularity, distance, previous check-ins, and so on) to the prediction performance, and show that recommending for pairs directly can perform better than combining individual recommendations for each of the pair.

I then consider the results for groups in the mobile phone network dataset. I again examine the contribution of different kinds of features to prediction performance, and show that in this case too, recommending for groups directly can outperform the combination of individual recommendations made for each of a group's members.

4.3.5.1 Foursquare dataset

I first assess the importance of the group features defined in Section 4.3.3.3, compared to the single-user features that can be used in single-user recommendation. I train three models:

- Using all of the features defined, that is, each feature vector includes global features, the single-user features computed for each of the users in a pair, and the group features computed for this pair.
- Using the global features, and the single-user features computed for each of the users in the pair, without the group features.
- Using the global features and the group features, without the single-user features.

Figure 4.13 shows the results of testing these three models for $N < 20$. This is the most relevant range for the mobile recommendations application; mobile users accessing applications on small screens do not want to have to scroll through a very large number of recommendations when searching for somewhere to go.

Combining information about each of the individual users in the same feature vector does not perform as well as computing the features considering the pair directly as a group. Furthermore, the group features together with the global features are able to perform just as well as when the single-user features are also included. This demonstrates that there may be benefit to be gained by exploiting the characteristics of group behaviour that the group features encode.

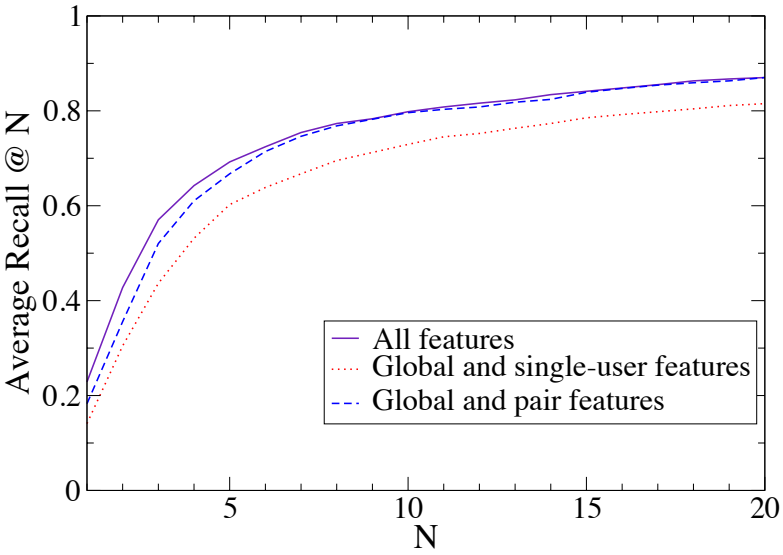


Figure 4.13: Improvement using group features over only global and single-user features for pairs in Foursquare.

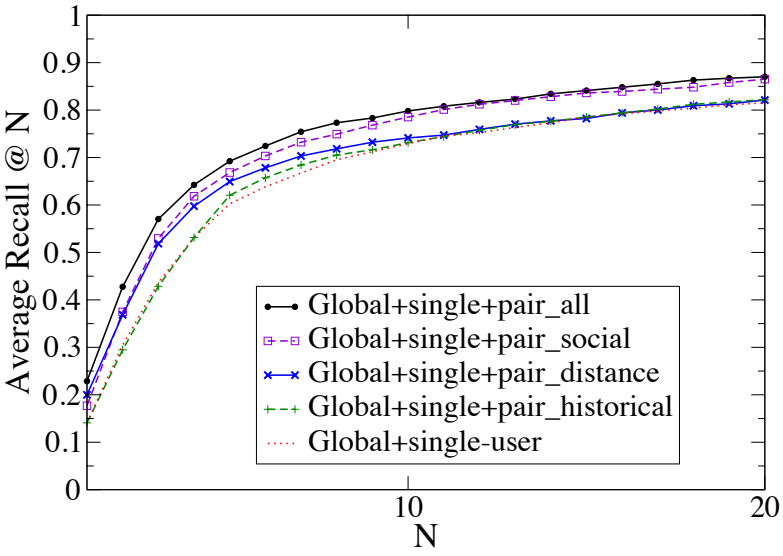


Figure 4.14: Contribution of different kinds of group features to recommendation performance for pairs in Foursquare.

Next, I analyse the contribution of individual kinds of group features to the trained model. I begin with the model using only the global and single-user features, and then assess the contribution of each kind of group features (distance, historical and social; note that popularity features can only be global), by adding each one to the model, showing the improvement in performance they provide. Figure 4.14 shows the value of $AR(N)$ for varying values of N, given models trained using:

1. Global and single-user features only: `pop_checkins`, `pop_checkins_c`, `cat`, `single_dist_closer` for each of the users, `single_dist_mean` for each of the users, `single_hist` for each of the users, and `single_social` for each of the users.

2. As 1, but with the group distance features `group_dist_mean` and `group_dist_closer`.

3. As 1, but with the group historical features `group_hist_total` and `group_hist_min`.

4. As 1, but with the group social features `group_social` and `group_mutual`.

5. The model with all of the global, single-user and group features in combination.

Each of the kinds of group features improves the performance of the model over that with just the global and single-user features. The largest improvement is gained in the addition of the social information features, `group_social` and `group_mutual`. This suggests that information about friends taking into account both users as a group is particularly useful when considering group venue recommendations, in particular the implications for venue recommendation of the finding that for pairs in Foursquare, social visits may be more likely to be to places where a user's friends have been than visits in general. The distance features also improve performance, showing the effect of users staying closer to frequently visited locations when with a friend than in general.

In contrast, the historical visits features do not appear to be as useful. This is in stark contrast to analysis by Noulas et al. [NSLM12a], who found that historical visits were one of the best performing features when making venue predictions for Foursquare users on an individual basis, and that social information, while still helpful, was less important. The lack of improvement from the addition of the historical features could be explained by social check-ins in Foursquare being more likely to be to places that are new to a user than check-ins in general. When considering only

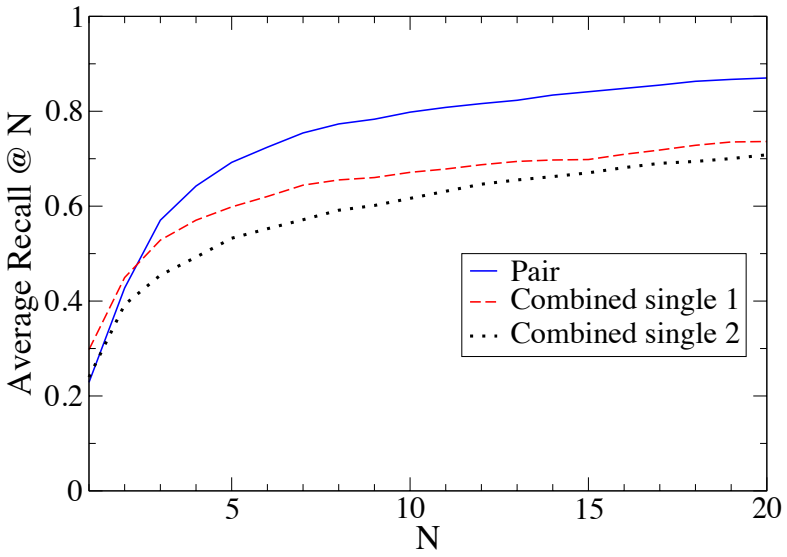


Figure 4.15: Improvement obtained using group recommendation over combining single recommendations for pairs in Foursquare.

historical visits, these new venues where the friends meet are indistinguishable from venues where the users would never go, and so these features are not as useful for recommending venues for social visits.

Figure 4.15 shows the results of comparing recommendation made directly for groups to those obtained by combining recommendations made for the individual users (Section 4.3.4). Using the features computed on friend pairs directly to perform recommendation achieves better performance than both methods of combining the individual lists, with Average Recall @ 10 of 80%, compared to 67% and 62% for the combination of single-user recommendations made using the same types of features. These results suggest that the differences between social and solo check-in behaviour mean that combining individual recommendations may not do as well as making recommendations for pairs directly.

4.3.5.2 Telecoms dataset

In the same way as for the Foursquare dataset, I assess the contribution of each kind of information as encoded by the features to recommendation performance. I begin with

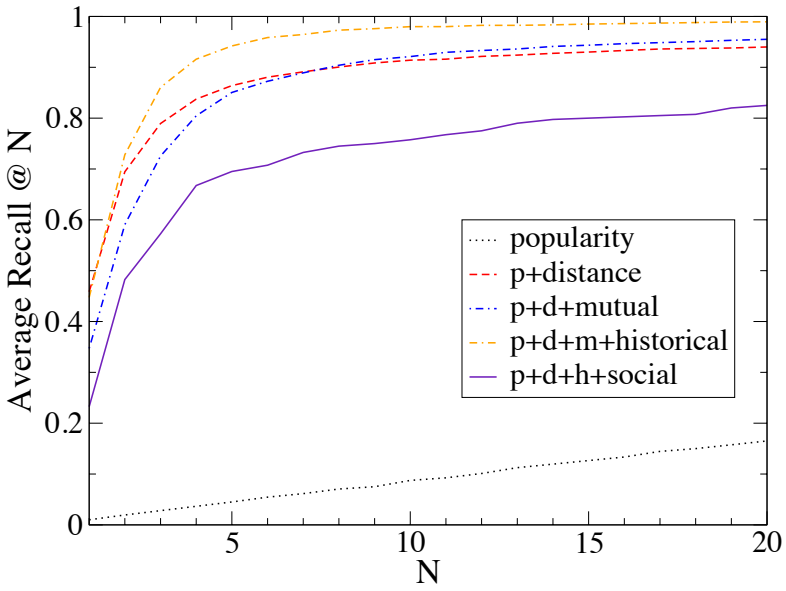


Figure 4.16: Improvement obtained by adding kinds of group features one-by-one in the telecoms dataset.

the model using only popularity features, and then assess the contribution of distance, historical and social features, by adding them one at a time to the model, and examining the improvement in performance they may provide. Figure 4.16 shows the value of $AR(N)$ for varying values of N, given models trained using:

1. Popularity only: `pop_checkins` and `pop_checkins_c`.
2. As 1, but with distance features `group_dist_mean` and `group_dist_closer`.
3. As 2, but with the group social feature `group_mutual`.
4. As 3, but with the historical features `group_hist_total` and `group_hist_min`.
5. As 4, but with the group social feature `group_social`.

The results are similar to those for pairs in Foursquare, except for the fact that the historical features show a large improvement over the model without them for groups in the mobile phone network. This could be explained by the fact that, as shown by the analysis earlier in this chapter, groups in the telecoms dataset are more likely to meet at places that are *not* new to them, in contrast to pairs in Foursquare. The use of

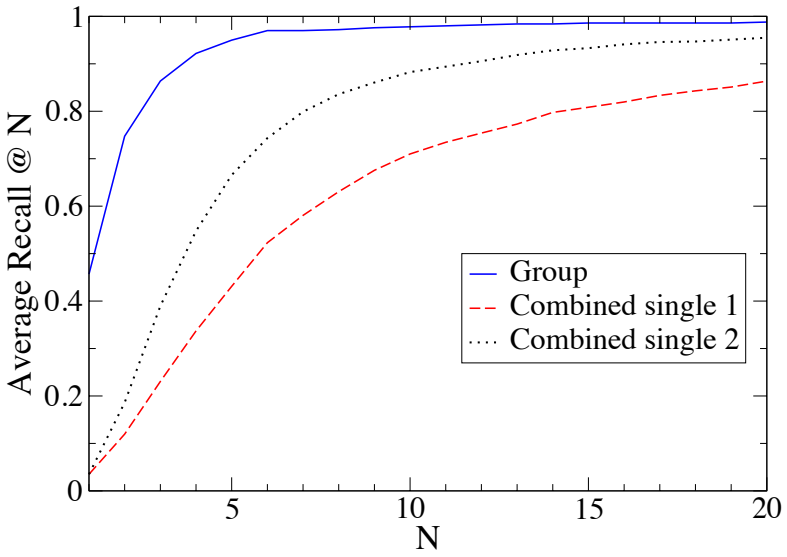


Figure 4.17: Improvement obtained using group recommendation over combining single recommendations for groups in the telecoms dataset.

information only about check-ins by mutual friends can be seen to outperform the use of information about check-ins by all friends of all members of a group; in fact the inclusion of that information makes performance noticeably worse. One could speculate that this could be due to people's social connections in the mobile phone network being more diverse than those in the Foursquare network (e.g., a wider variety of friends, family, colleagues, and so on, all of whom might tend to meet in different places and so whose inclusion while trying to make recommendation for a group of which they are not part might harm performance, while people are less likely to connect to particular kinds of acquaintances such as work colleagues in some location-based social networks [CRH11, LCW+11]), but this is only one possibility and further research would be needed to determine the cause of the observed effect. Finally, Figure 4.17 shows the results of performing recommendations for groups directly compared to the two ways of combining individual recommendations for each member define in Section 4.3.4. As in Foursquare, making recommendations for groups can be seen to outperform considerably the two combination methods. This suggests that the differences between individual and group behaviour might be used to improve venue recommendation in

the case of groups of friends looking for somewhere to go together in location-based social networks, and indeed possibly to improve location prediction of mobile phone users too, which has technological applications in areas such as targeted advertising, mobile web search, and dissemination of location-aware content [dLM13].

4.4 Discussion

The varying use of different kinds of spaces around the city for different purposes, including meetings between groups of friends, was shown in my analysis above both in terms of the finding that the likelihood that two people colocated in a given space are friends appears to be connected to the nature of the place, e.g., people colocated at a Residence venue are more likely to be friends than people at a Transport venue, and also in terms of the finding that visits to some kinds of places such as Arts and Entertainment venues and Nightlife venues are more likely to be made with friends than visits to other kinds of places. This is in agreement with the observation by Kostakos et al. [KOP+10] that some spaces, such as pubs and offices, are more likely to host encounters between groups of people who know one another than other kinds of places, and extends that work by analysing multiple cities and showing the same patterns that Kostakos et al. found in their study of Bath, UK. The model presented above for place-based social networks in cities makes use of this idea by using the category of a place to determine the probability that people with that place in common have a tie in the social network, and running the model without this aspect results in a social network without the empirically observed levels of clustering and strength of community structure. While this is not sufficient evidence to draw conclusions about causal relationships between the spatial and community structures of city-level social networks, it does suggest that there might be fruitful research to be done in this direction that might be easier, with today's easily available location data from technological social networks than it has been in the past.

My analysis of the differences between the places where people go when with groups of their friends and those where they go more generally showed similar patterns for both the LBSN dataset and the telecoms dataset, with people being more likely to stay closer to familiar locations when with friends, and being more likely to visit places with friends where the group's wider circle of friends have also been. However, in the

Foursquare dataset it appeared more likely that pairs of friends visit new places than do people on their own, while in the telecoms data the opposite behaviour was seen. Upon further investigation, considering only pairs in the telecoms dataset the same behaviour is observed as in Foursquare, while larger groups seem more likely to visit places they have been to before, which might be understood in terms of the confidence to explore new places inspired by the presence of a peer having to be balanced by the additional difficulty in coordinating a larger group. It seems likely that these differences in the use of urban spaces by individuals and by groups, and the characteristics group and individual mobility, could potentially be used to improve technological applications such as the venue recommendation task common in LBSNs that I have explored here.

4.4.0.3 Limitations

Similarly to the results for inter-city scale groups examined in Chapter 3, there are some points that must be considered when interpreting these results and drawing conclusions. As mentioned in Section 3.5, the analysis of the behaviour of Foursquare users in this chapter is only certain to be relevant to those Foursquare users it concerns, who are by their nature as users of the service not representative of a general population, and similarly for the mobile phone users whose behaviour is reflected by the telecoms data. Therefore these results should not be used to draw general conclusions about social and mobility behaviour but instead taken in the context of the technological service concerned; as I have shown, for applications *within* such services, such as venue recommendation for groups in LBSNs, the results could be useful. Additionally, it is possible that some of the patterns seen in Foursquare check-ins do extend outside the specifics of the services, since building on Cho, Myers, and Leskovec's observation of considerable similarity between the mobility patterns in some LBSN datasets and in mobile telephone network data [CML11], I have shown similarities also between the mobility behaviour of groups seen in the Foursquare dataset analysed in this chapter and the dataset from the mobile network operator.

4.5 Summary

In relation to the thesis that the relationship between physical space and social groups at different scales can be exploited to create or improve technological social applica-

tions, I have explored in this chapter the role of space in groups at the scale of cities, by analysing one dataset from the online location-based social network Foursquare and one from a mobile telephone network in Portugal. The results show differences in the use of places around a city by groups and individuals, and suggest that some kinds of places have a greater propensity to host meetings between social groups than others. I have showed how these differences in group and individual mobility behaviour might have uses for applications such as venue recommendation, which is an important task in today's location-based online social networks; in particular, it might be beneficial to make recommendations taking into account whether a user is looking to go somewhere by themselves or with a group. Given that venue recommendation is just one application that benefits from improvements in location prediction, the mobility behaviour of social groups in cities could most likely also be used to improve other such applications such as dissemination of location-aware content or returning relevant mobile search results.

Chapter 5

Social groups in building-scale networks

Having examined the role of space in social groups at inter- and intra-city scales, in this chapter I focus on groups at the scale of a single building. In the last chapter, I showed that some kinds of places around a city (e.g., venues in Foursquare's Entertainment, Food, and Nightlife categories) are particularly important for meetings between friends. Is the same phenomenon observed for the different types of spaces in individual buildings? Does more social interaction take place in informal areas such as cafeterias than in office spaces, for example?

Research in this area tends to face difficulties not because of the lack of data that has posed problems in the case of the study of the larger-scale networks, but more due to the difficulties in observing people going about their daily lives without causing them to change their behaviour [Why43], as well as indoor localisation historically having been difficult [DCC12]. Modern ubiquitous sensing technology can now help to address this, by enabling unobtrusive, automatic sensing of people's social interactions and their location within the buildings they occupy.

In this chapter, I analyse a dataset of social interaction and location traces collected using electronic badges worn by people in a workplace environment. I study the sensed behaviour of the same sample of employees in two different workplace buildings, and examine how the physical spaces available may affect face-to-face communication in such an environment. Specifically, I study in which spaces of the buildings contacts between members of different working groups take place, and likewise contacts between

members of the same groups. I examine the types of contact taking place in different kinds of space such as offices, meeting rooms, and cafeterias, and show how the availability of such spaces may have an impact on the numbers of intra- and inter-group contacts in the workplace.

In relation to my thesis, the results of this study show that space at the scale of a single building, in this case, the workplace, can play a role in the interactions within and between members of different groups due to the different activities that members of these groups are involved in. Informal spaces such as cafeterias and coffee rooms afford more opportunity for contact between groups, due to workplace teams being functionally defined according to projects or areas of expertise so that designated workspaces are most likely to host intra-group contact. Interaction between members of different groups can be beneficial for innovation and productivity in the work environment [Bur04, Pen12], and so these informal spaces may play an important role in their propensity to host unplanned encounters between employees who might not work together. It is clear that with recent and continuing advances in ubiquitous sensing that there is scope for new technological applications involving the automatic monitoring of these effects, to enable managers to understand and maintain the 'communication health' of their workplaces in terms of contact within and between groups of employees.

5.1 Effects of space on the communication of workplace groups

In the field of architecture, the effect of the nature and layout of spaces on the behavioural patterns of people is an important factor in building design. Significant effort has been put into understanding how the physical space of workplaces can directly affect how often employees meet one another and interact face-to-face [AH06]. Communication between employees is a vital factor in the operation of an organisation, and even in today's technologically connected world, face-to-face interactions remain crucial for the exchange of ideas and information [Pen12, SS12]. It is therefore unsurprising that building spaces that facilitate such interactions is a significant consideration in architectural design.

Measuring the impact of a workplace building layout on face-to-face communication is an important step, not only to validate architects' objectives, but also to enable the evaluation and reconsideration of traditional design principles. Studies in architectural design, such as the work of Thomas Allen [AH06], consider how organisational structure and spatial configuration of work environments combine to influence communication between employees. However, these studies suffer from a crucial shortcoming: they lack reliable means of measuring face-to-face interactions in the workplace. Traditional approaches to evaluating the use of spaces in buildings rely on ethnographic studies where observers track employees over a period of time, or on self reports and surveys. Both approaches can deliver biased results, either because participants alter their behaviour when they know they are being observed [Why43], or because they tend to offer socially desirable responses to surveys [BSBS78]. Furthermore, studying the impact of a building's layout on social behaviour is challenging considering the large number of variables that can affect such behaviour. For example, different types of organisational structure may affect social behaviour more significantly than space layout.

In this section, I present the results of a study that addresses these two challenges: firstly, data was collected using wearable sensing tags capable of capturing face-to-face interactions and the actual locations of people. The tags are unobtrusive and thus make it less likely that people will change their behaviour due to awareness of being monitored. Secondly, the study was performed in a research institution in the UK that moved from their old premises to a new purpose-designed building. Two data collection deployments were performed, one in the old building and one in the new. Considering that the set of additional variables, such as organisational structure, remained unchanged, the results allow examination of the impact that spatial design may have on inter- and intra-group interaction in a workplace.

The work described in this section relies on a theoretical premise established by Thomas Allen [AH06], who defined three types of communication necessary in an organisation. The first is *communication for coordination*, which takes place between people working on the same project, in order to coordinate work activities. Second, *communication for information* is necessary for people working in the same area to keep up to date with developments in their field of expertise. It is intuitive that these two kinds of communication should, in a typical office environment, take place in

offices and designated meeting rooms, since the managers of most organisations tend to be aware that these types of communication are crucial. When deciding who sits where, they tend to arrange that people working on the same projects and in related fields are near to one another.

The third type of communication is *communication for inspiration*, which, "In an organisation that relies on creative solutions to problems," Allen writes, "is absolutely critical. It is usually spontaneous and often occurs between people who work in different organisational units, on different projects." The criticality of these interactions between members of different teams, who might not normally encounter one another during their work, has been demonstrated by much other research [Bur04, Pen12]. It follows that as well as offices and meeting rooms, workplaces should include informal spaces such as coffee areas, where unplanned encounters between employees can take place, outside those meetings that would be expected given the formal organisational management structure and division into subgroups working on various projects [KFRC90, WOOKP10]. Indeed, this idea is already being put into practice by high-tech organisations such as Google, where "even the length of the lines inside the cafeteria are designed to make sure Google employees talk to others they don't necessarily work with [...] if there is no line, you won't talk to anyone, you won't interact" [Hen13].

5.1.1 Dataset description

I analyse the relationship between the positions of people in a formal organisational structure, and their interactions in the differing spaces of two different workplace buildings. The dataset consists of traces collected using RFID badges (Figure 5.1) worn by the same employees in an industrial research lab for two periods of two weeks' duration each, with one data collection period in each of two different buildings. The organisation was moving from an old building to a new, purpose-built building, which had been designed with various aims concerning interactions between different groups of people in mind, and so this analysis affords the opportunity to study the effect of the spaces available in the two buildings on face-to-face interactions, and to assess whether these aims may have been achieved.

Data collection involved the use of wearable RFID tags for the collection of face-

Figure 5.1: One of the RFID tags used in data collection, worn on the chest as in the deployments.

to-face interactions and location information. The aim was to use a technology that was not obtrusive, and would not affect significantly the normal social patterns of the participants. Measurements were captured by active RFID badges [CVdBB+10] worn on the body. The badges are lightweight radio transceivers, programmed to transmit a beacon periodically (every 1 second), and to listen continuously for beacons from other badges nearby. The badges are configured to transmit low signal strength beacons that were experimentally evaluated to have a range of 1.5m - 2m with clear line-of-sight. When worn by participants, the beacons are shielded by the body, meaning that successful communication can occur only when another badge is facing that of the participant. In this way, the tags can assess continued face-to-face proximity between users.

Data collection also involved the capture of location information. A number of RFID tags were placed on the walls of the workplace buildings and were configured to transmit beacons of greater signal strength than the badges worn by the participants, achieving a range of around 3m - 5m. Location tags were deployed in each participant's office, in meeting rooms, in laboratory spaces, and in communal areas such as cafés, kitchens, and common rooms. The dataset therefore also contains traces that can be used to approximate the locations where face-to-face interactions took place.

The dataset contains location and interaction traces collected from the same sample of 24 employees in both buildings. The first data collection deployment was performed

in November 2012, where participants were tracked for 2 working weeks, and 59 room tags were deployed across 3 floors of the target building. The deployment captured 1669 unique face-to-face contact occasions. The company moved to the new building in January 2013, and the second study was conducted in June 2013, allowing enough time for the participants to settle in the new environment. The second study again collected traces for approximately 2 weeks, using 84 room tags covering 3 floors of the new building, and the deployment captured 2693 unique face-to-face contact occasions.

5.1.1.1 Pre-processing

Before performing the analysis, a certain amount of processing of the raw data was necessary. I work under the assumption that continued face-to-face proximity is a good proxy for a social interaction between users. Specifically, I consider an indication of social interaction the presence of two individuals facing each other at a distance of no more than 2 metres, for a duration of more than 30 seconds. Defining the distance threshold for such matching to be 2m (the configured range of the radio transmission) makes the likelihood of false positives in the dataset negligible. Reducing the number of false negatives (face-to-face proximity not detected by the tags) can be controlled by using time windows within which detected beacons can be considered as indicators of proximity for that duration [PBC+11]. I use a 2-minute time window, having verified that 5-minute and 10-minute windows did not significantly alter the findings presented here. I consider as contacts only traces where at least 2 beacons are received 30 seconds apart, thus avoiding counting very short contacts, such as when two people pass one another in the corridor without stopping.

The short communication range of the tags means that the dataset as collected does not include contacts between larger groups of people interacting in a larger space (e.g., in a meeting). I compensate for this by applying a transitivity property over the original dataset: if participant $P1$ is in contact with participant $P2$, and at the same time participant $P2$ is in contact with participant $P3$, then $P1$ and $P3$ can be considered to have been part of the same group interaction.

I establish the approximate location where an interaction takes place by considering the traces of static tags received by all participants in a meeting. I apply a simple voting scheme over the number of static tag beacons received by all interacting users

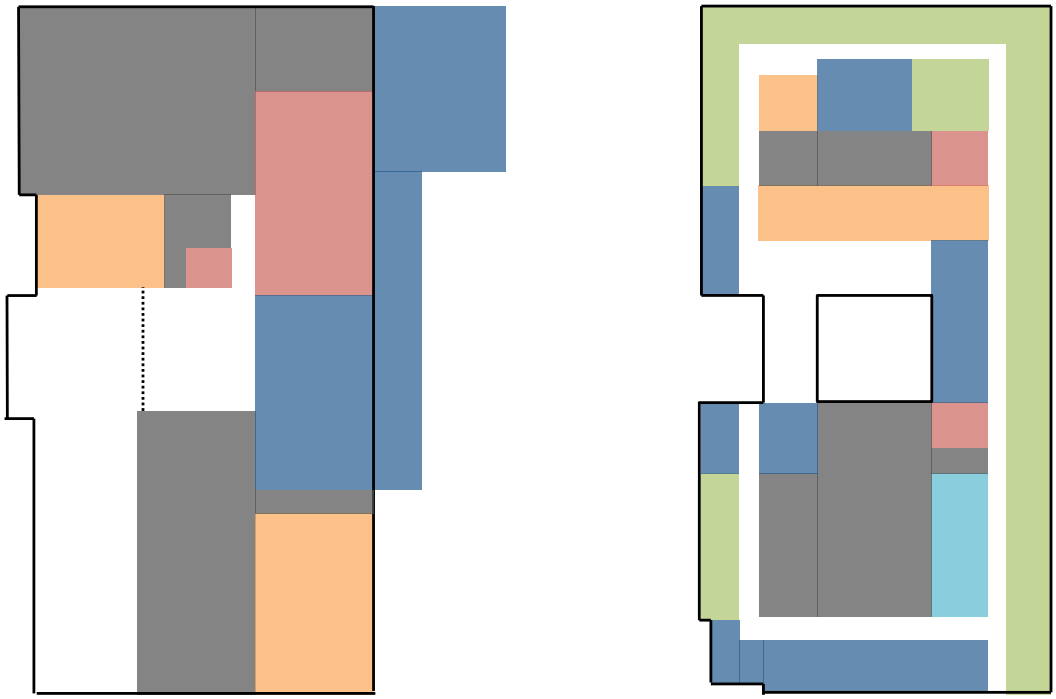
within the specified time window. Using the ID of the static tag with the highest number of beacons received, I assign the type of this location (office, meeting room, cafeteria, etc.) to that interaction. In the vast majority of cases the most probable location is clearly distinguishable in this way, with a much higher number of beacons received from one room tag than from the others. In the remaining cases, the potential alternative rooms are of the same type as the top one (this tends to occur with different offices near to one another). Using the static tag with the highest number of received beacons allows the identification of a type of location where a meeting took place, even though in some cases it is not possible to pinpoint the exact location.

5.1.2 Aims in design of the new building

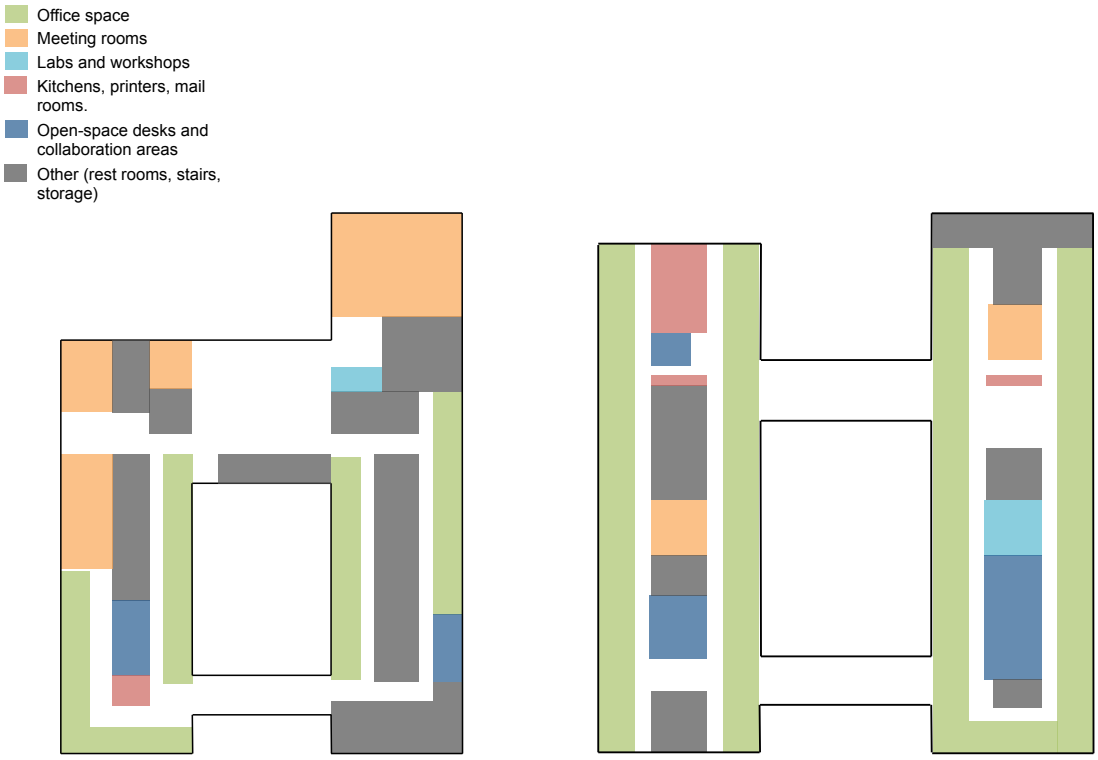
The new building was designed and built specifically for the research institution in question, with the architects having particular intentions with respect to the use of the space. The key design decisions focused on enabling more interaction between people from different research groups who might not usually encounter one another: in terms of Allen's types of communication, *communication for inspiration*.

The most obvious difference between the two buildings in this respect is the presence in the new building of a central cafeteria area located away from the office spaces, where employees can buy food at lunchtime, and meet for coffee (see Figure 5.2). The best two coffee machines, having coffee of the same quality as in commercial coffee houses, were deliberately placed on the ground floor, opposite to the main entrance, so that most people would have to walk past in the morning. It was expected that the quality factor would encourage people to gather in the cafeteria for good coffee, where they would have a greater chance of serendipitous encounters than in the smaller kitchens upstairs. The kitchens on the individual floors were not provided with equivalent quality machines to bring this into effect. In the old building, there was no cafeteria serving food as in the new building; instead people would commonly buy food from elsewhere, or bring it from home, and eat it in kitchen spaces close to their offices. There was a kitchen where many people would eat, but, as one of the participants in the study commented, "it was not the same as having a café, as people also ate in other spaces throughout the building."

To a similar end, there was also a general aim in the design of the building to



(a) New building



(b) Old building

Figure 5.2: Building layouts. The figures show the ground floor (left) and first floor (right) of the two buildings. Different colours indicate the type of space in each building, e.g. offices, meeting rooms, kitchens.

encourage increased use of shared spaces, as opposed to individual offices. Lab spaces were made bigger in the new building so that they might accommodate more people from different groups. There are lots of open areas and mini conference rooms without doors, in order to encourage groups to meet in these shared spaces, rather than in their own offices. It is probable that most meetings in these kinds of spaces would be related to work, and thus likely to be for the purpose of *communication for information* and *communication for coordination*.

5.1.3 Methodology

I investigate whether these differences in the physical space of the new building are reflected in the patterns of interactions between its occupants. I first study whether there is more communication between different subgroups in the organisational hierarchy in the new building, in accordance with the aims of the architects. I then examine in more detail the use of different kinds of spaces (offices, meeting rooms, kitchens, and so on) in the two buildings, and which spaces might be important for interactions within and between groups.

5.1.3.1 Impact of building space on face-to-face contacts

I use the management structure shown in Figure 5.3 to define subgroups, and consider three such groups, one corresponding to each of the three components present in the graph. I first quantify inter-group contact by measuring the proportion of contact pairs that are intra- and inter-group: one would expect that there will be a higher proportion of contact pairs inter-group in the new building, given the emphasis in the building's design on shared spaces to facilitate inter-group contact.

I further analyse communities in the contact graphs, since in terms of the flow of ideas and information, a less modular structure or one with more connections between communities would be advantageous [For10]. I first find k-clique communities in the network, defined to be the union of all cliques of size k in the graph that can be reached through adjacent (sharing $k-1$ nodes) k-cliques [DPV05], and compute the proportion of edges in the contact graph that exist within and between communities, for varying values of k. If the new building space promotes inter-group interaction as the architects intended, the proportion of inter-community edges should be higher for the contact

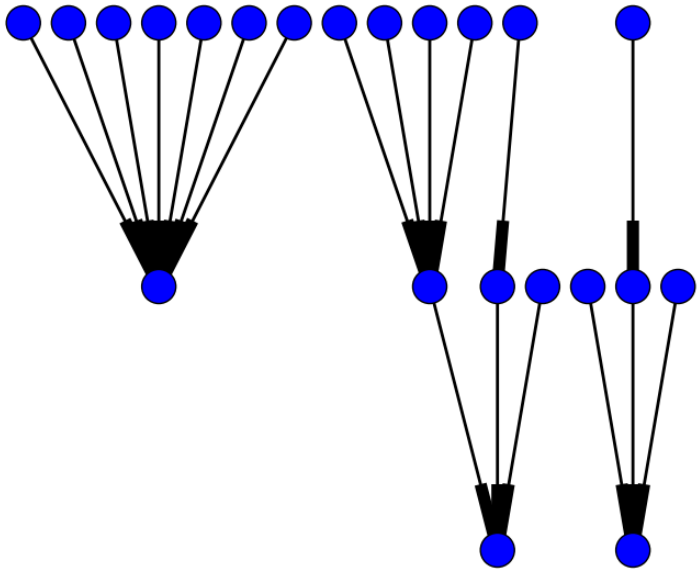


Figure 5.3: The structure of the subset of the organisational management hierarchy made up of the individuals participating in the study. Circles represent the participants, and arrows indicate the 'is managed by' relationship.

graph from the new building than that for the graph from the old building.

I also run the Louvain algorithm for community detection [BGLL08] on the contact graph, and compute the modularity Q of the best partition found by the algorithm, a measure of the strength of the community structure of the network (see Section 4.1.2.4). I choose this algorithm because, having examined the overlapping community structure of the network using k-clique analysis, it is also interesting to examine a single good partition of the network into communities, as is found by the Louvain algorithm, and to have a way of quantifying the modularity of the partition, which is in this case computed by the detection algorithm itself. If the aims of the architects were achieved, one would expect that in the new building there would be a less strong community structure to the contact graph due to increased mixing between different groups of people, and therefore the value of Q would be lower than that seen in the old building.

5.1.3.2 Different kinds of spaces and inter-group interactions

It is well-known that areas in which people can gather together to eat and drink are often important hubs for the kind of social contact and chance meetings that are so beneficial for the exchanges of ideas and information [AH06, ITM96]. Therefore it is unsurprising that the architects of the new building study envisaged that the cafeteria would promote interaction between people in different groups who might not normally encounter one another.

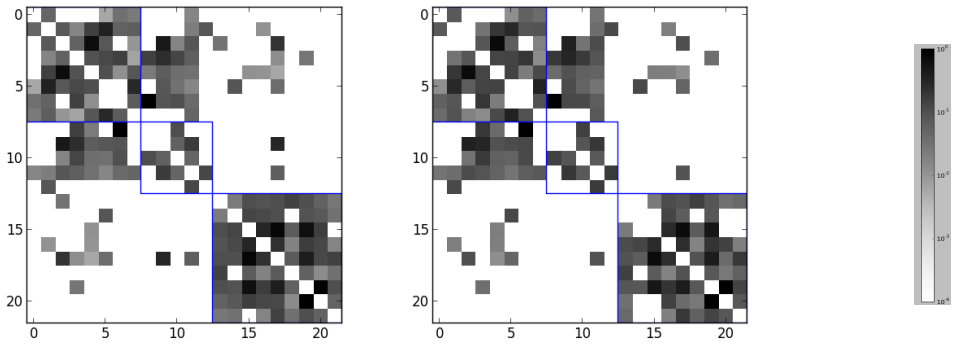
I test quantitatively the importance of food and drink spaces for inter-group interactions by computing the number of contacts taking place in different kinds of spaces (e.g., offices, meeting rooms, kitchens) within the two buildings over the course of the working day. One would expect that in the new building, the impact of the cafeteria would manifest as an increase in the number of contacts occurring over the lunch hour (12-1pm). I further investigate the importance of lunchtime for inter-group contact by comparing the proportion of contact pairs that are between individuals from different groups in all of the data for each building, and in the same data but with the lunch hour removed. In both cases, the proportion of communicating inter-group pairs should go down when lunchtime is excluded from the analysis.

I then specifically investigate the proportion of inter- and intra-group contacts that occur in different kinds of spaces in the old building and in the new building. One would expect that in the new building, a greater proportion of inter-group contacts would occur in kitchen areas (which include the cafeteria) than in the old building, and also that these proportions would reflect the inclusion of more meeting rooms in the new building to encourage colleagues to venture away from their own offices in order to hold work-related meetings.

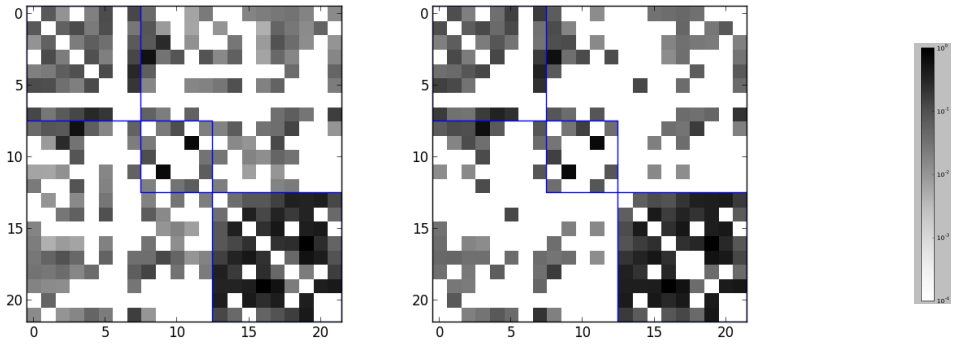
5.1.4 Results

5.1.4.1 Impact of building space on face-to-face contacts

In the new building, 76.1% of pairs were cross-group, increased from 58.8% in the old building. This suggests that the aim of the architects to design the building to promote mixing between people who might not encounter one another otherwise may have been successful, and that there might be more opportunities for *communication for inspiration* in the new building. Figure 5.4 shows a visual representation of the extent



(a) Old building, with lunchtime (left) and without lunchtime (right)



(b) New building, with lunchtime (left) and without lunchtime (right)

Figure 5.4: Inter- and intra-group contact pairs in the old and the new buildings, with and without lunchtime. Each row and column corresponds to an individual, and the ordering of individuals is by group so that adjacent rows (columns) represent colleagues in the same group. Group boundaries are shown by the blue lines. A dark square indicates that contact was recorded between the individuals concerned, and a white square indicates no recorded contact. A more 'blocky' structure suggests less inter-group mixing. The total duration of recorded contact is reflected by how dark a shaded square is, with the durations normalised by the total time that the individuals concerned were recorded as being in the building and also according to the maximum recorded duration, so that the values shown are the proportion of their time in the building that the corresponding individual spent in contact with another, as a fraction of the maximum such proportion. The increase in inter-group contacts in the new building is clear. The figures on the right show fewer inter-group contacts than those on the left, which demonstrates the importance of lunchtime for social contact in both buildings.

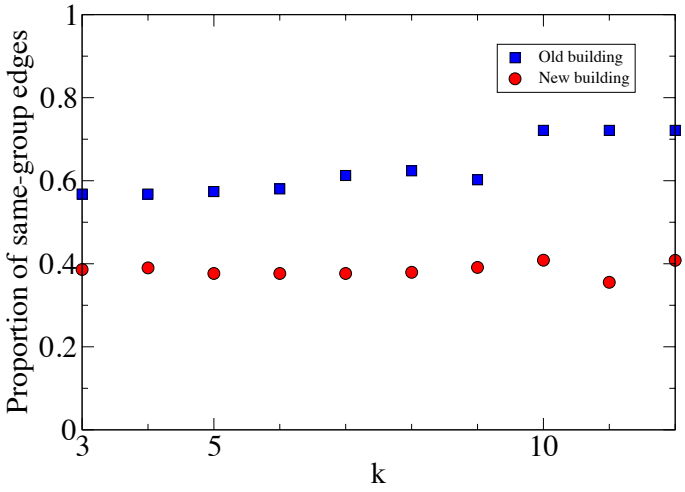


Figure 5.5: Proportion of intra-community edges that are between individuals in the same group, for varying values of k used in the k-clique algorithm for community detection. The proportion of same-group intra-community edges is lower in the new building than in the old building.

of this effect, in the form of netgraphs [AH06]. Each row and column corresponds to an individual, and the ordering of individuals is by group so that adjacent rows (columns) represent colleagues in the same group. Group boundaries are shown by the blue lines. A dark square indicates that contact was recorded between the individuals concerned, and a white square indicates no recorded contact. A more 'blocky' structure suggests a lower level of inter-group mixing. The netgraphs make clear the extent to which *more inter-group mixing is encouraged by the design of the new building*, with many more dark squares outside the blue lines indicating contact between individuals in different formal subgroups.

Figure 5.5 shows the results of k-clique analysis of the contact graph. Specifically, I plot, for varying values of k, the proportion of intra-community edges in the contact graph that are between individuals in the same subgroup. The proportion of intra-community edges connecting those in the same group is lower in the new building than in the old building, for all of the values of k. This implies that the community structure of the contact graph is less constrained by the formal group structure and therefore that *the new building space may indeed encourage more opportunities for*

	Old building	New building
Group A	Floor 1	Floor 3
Group B	Floor 2	Floor 3
Group C	Floor 1	Floor 1

Table 5.1: Office locations of groups in the two buildings.

mixing between individuals in different groups.

I further confirm this result by checking the modularity Q of the best partition of the contact graphs found by running the Louvain community detection algorithm [BGLL08]. A larger value indicates a stronger community structure, with values of around 0.3 or more being considered high.

The measured value of Q for the best partition of the contact graph in the old building was 0.324, compared to 0.145 for that in the new building. This suggests again that the community structure is less modular and that there are more contacts that are outside usual meeting groups.

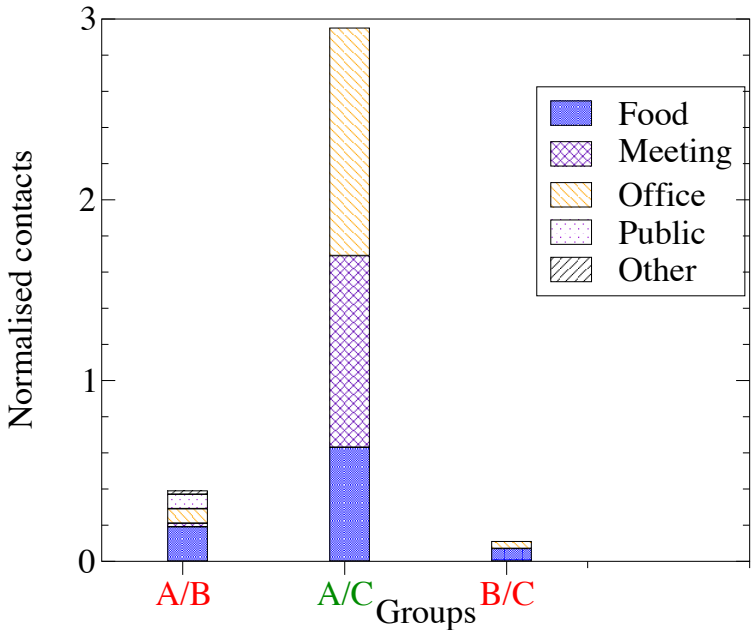
5.1.4.2 Different kinds of spaces and inter-group interactions

Before examining the impact that different location types may have on inter-group interactions, I assess the extent to which the distribution of offices across floors may affect communication. Previous work has indicated that splitting employees across floors may have a significant impact in social interactions, mostly in traditional building designs [SM11]. Of the individuals involved in this analysis, each was on the same floor as the rest of their group, but some groups shared a floor and others were on different floors. Table 5.1 shows the allocation of groups to floors in each of the two buildings: group pairs A and C in the old building, and A and B in the new building, are located on the same floor.

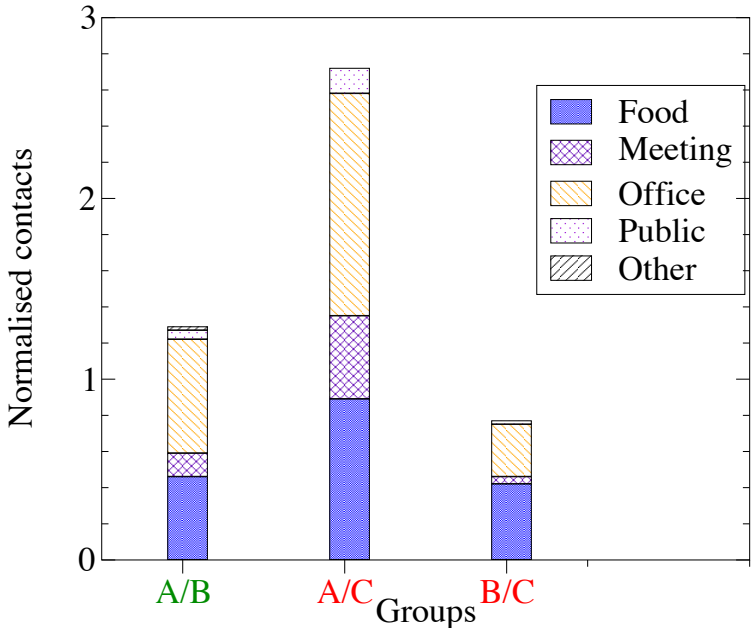
In order to assess how this distribution may have affected interactions I calculate the number of inter-group interactions for each pair of groups in both buildings. I normalised the values by the product of the sizes of the groups involved, to account for the number of possible pairs:

$$N_{AB} = \frac{C_{AB}}{|A| \cdot |B|}$$

where N_{AB} is the normalised number of inter-group contacts for groups A and B, C_{AB} is the absolute number of inter-group contacts, and $|A|$ and $|B|$ are the sizes of the two



(a) Old building



(b) New building

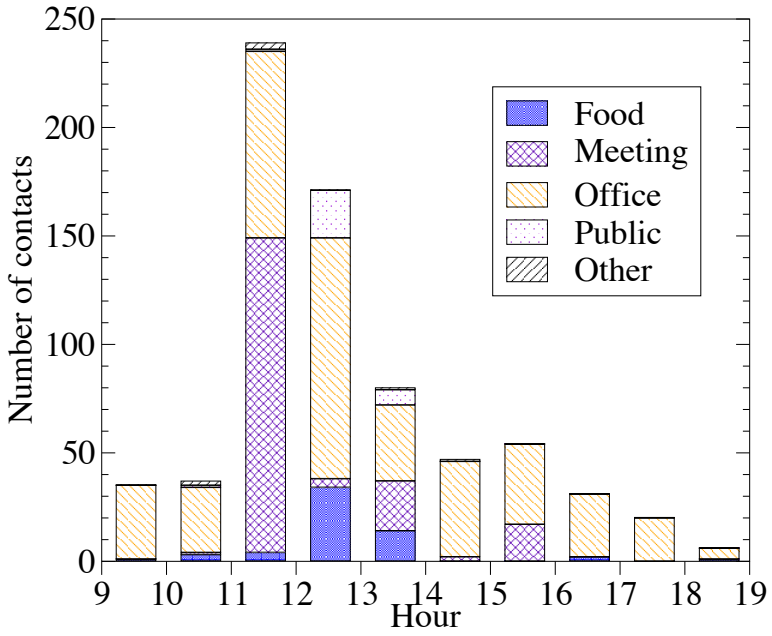
Figure 5.6: Inter-group contacts, by location. The labels on the horizontal axis show the pairs of groups, in green for groups located on the same floor, and in red for those on separate floors.

	Old building (%)	New building (%)
Food	28.0 (3.7)	37.4 (9.4)
Meeting	28.8 (24.4)	12.3 (34.0)
Office	38.7 (65.5)	45.5 (51.5)
Public	3.8 (5.1)	4.3 (3.7)

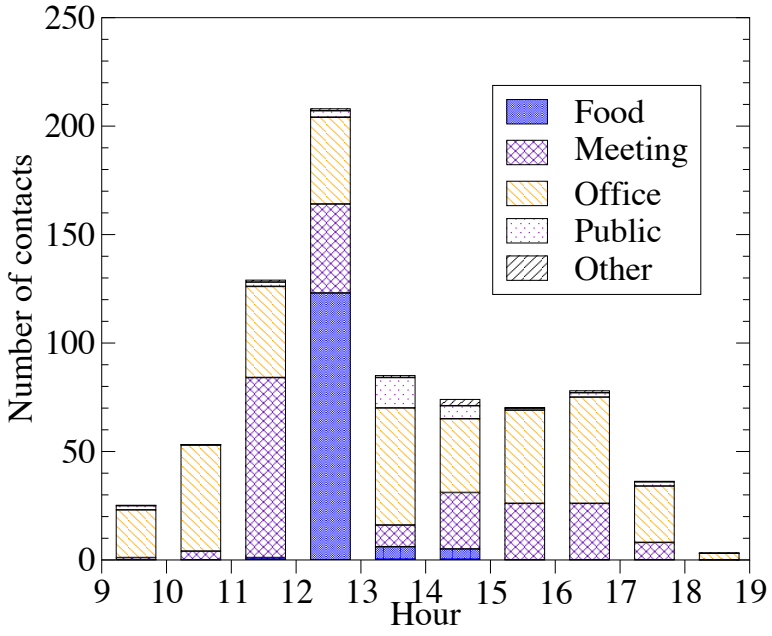
Table 5.2: % of inter-group (and intra-group, in parentheses) meetings that take place in each kind of space, in the old building and in the new building. Inter-group contacts far exceed intra-group interactions in Food spaces.

groups. Figure 5.6 shows the normalised number of inter-group contacts in the two buildings. The graphs show that in the old building, floor allocations were a strong factor in inter-group interactions. Indeed, the vast majority of inter-group interactions take place between groups A and C, with offices on the same floor. In the new building, there is an increase in inter-group interactions for all the pairs. Groups A and C have maintained the same level of interaction although they are now located on different floors, while groups B and C show a 7-fold increase compared to the old building, despite being located on separate floors. This increase is higher than the 3.5-fold increase in interactions between groups A and B, which are located on the same floor. These results suggest that in the new building the distribution of offices across floors is not such a dominant factor in determining the interactions between groups. The graphs also suggest that interactions in food areas may be more important in the new building. I investigate further the importance of food spaces by examining the impact of lunchtime on inter-group interactions, by comparing the contact pairs comprising people in different groups previously shown with the same when excluding lunchtime (12-1pm) from the analysis. The right-hand side of Figure 5.4 shows the results; in both buildings, inter-group contacts decrease when lunchtime is excluded, demonstrating that lunchtime, and therefore most likely contact in spaces for eating and drinking, is indeed somewhat important for contact between people in different groups. Figure 5.7 demonstrates this effect by showing the numbers of contacts detected between participants in different kinds of spaces (food and drink areas, meeting rooms, offices, and public areas) during different hours of the day. The impact of the cafeteria in the new building is clearly visible as a sharp increase in the number of contacts taking place in Food spaces during the hours of 12-1pm.

I finally examine directly the proportion of inter-group contacts that occur in each



(a) Old building



(b) New building

Figure 5.7: Number of contacts in different kinds of spaces, by hour of the day. The impact of the cafeteria in the new building is clearly visible, with a large increase in contacts in Food spaces between the hours of 12-1pm.

kind of space. Table 5.2 shows the results from each building. In the new building, a greater proportion of inter-group contacts take place in food areas than before, which suggests again that the intention of the architects for the cafeteria to function as somewhere where people in different groups have the opportunity to meet one another has been realised. Furthermore, the proportions of inter-group contacts that take place in Food spaces far exceed the proportions of intra-group interactions that take place in Food spaces in both buildings: in the old building, while only 3.7% of intra-group interactions take place in Food spaces, 28.0% of inter-group interactions happen there. In the new building, 9.4% of intra-group contacts occur in Food spaces, compared with 37.4% of inter-group interactions, which provides some quantification of the importance of food and drink areas for inter-group encounters. The table also shows the proportions of *intra*-group contacts that take place in each of the different kind of spaces. The majority of these contacts occur in meeting rooms and offices, which is expected given that such contacts are likely to be those comprising what Thomas Allen refers to as *communication for coordination* and *communication for information*. More of these contacts take place in meeting rooms in the new building than in the old building, which suggests that the architects' aim of encouraging more meetings in shared spaces, away from individuals' own offices, was indeed met in the realisation of the design.

5.2 Potential applications: encouraging effective workplace communication

The study reported in this chapter demonstrates the use of ubiquitous sensing technology to monitor changes in face-to-face communication patterns between employees influenced by the spaces provided by their workplace building. It is not difficult to see that such information could be used fruitfully to provide feedback to the employees themselves on their patterns of communication, or indeed to their managers, to encourage people to take advantage of the possibility of encounters with those they might not normally meet or opportunities to interact informally with their teammates.

A prototype use of sensing technology to monitor workplace communication was the use of sociometric badges by Kim et al. [KCHP08], who discussed the use of

electronic badges capable of sensing interactions between members of a group during meetings, and how this information could be presented back to the group members to make them more aware of the extent to which they dominated conversation or perhaps did not contribute as much as they would like to. Similarly, Rachuri et al. [RMMR11] implemented a platform for sensing interactions between colleagues using smartphones, which made use of a web interface to present the sensed information to the users of the application. In the light of the results that I presented above, which suggested that the space of the new building encourages inter-group communication, it would be possible to extend these location-agnostic social sensing platforms to make use of location data as well as social information. The combination of sensed social and location information could then be presented to employees and to their managers to help visualise where and with whom people communication, and to use awareness of this to balance communication within and between teams in the workplace or encourage the use of different building spaces that favour particular kinds of communication.

To give an example of such an application, it would be possible to create maps of workplace floor plans reflecting the extent to which each area hosts conversations between those at different positions in the formal organisational structure, and therefore which kinds of communication (for coordination, information, or inspiration, in terms of Thomas Allen's work) are most likely to take place in which kinds of space. Similarly, it would also be possible to monitor using the same technology the kinds of interaction in which an individual engages over the working day, and to make this information available to the employees, to encourage them to balance the three kinds of communication to meet their working needs. For example, one could create graphs showing how a research group is connecting within their own team and also how well-connected they are with other teams, both of which have been shown to be important in the workplace [Pen12]. It would then be possible to measure the impact of this feedback on organisational productivity, adding to branches of research investigating how ubiquitous computing can be used to nudge people to change their behaviour [RHM+10].

Of course, the ability to track social interactions in the workplace and the potential display of data from this monitoring raises further questions about privacy, how happy employees would tend to be about being tracked in this way with the potential for

others to see their data, and whether this in itself would cause behaviour to change. All of these issues could be ground for further research before this type of technology became widely deployed in workplaces, but these issues also make it difficult at present to study on a large scale the influence and effects that such technology could have. At present, research is ongoing into people's willingness to share their location at any one time [TLH+10, TCD+10, TKCS10], and it still remains to address these issues as they might affect the day-to-day monitoring of all employees within an organisation, but in the future, one could imagine every member of staff wearing a sensor badge, creating what Pentland calls a 'God's-eye view' of the workplace [Pen12].

5.2.1 Discussion

This particular study helps to fill a gap in the body of work involving the interaction between physical space and formal organisational structure to influence face-to-face encounters between individuals in the workplace. Specifically, I have used the advantages of modern ubiquitous sensing technology over methods such as manual observation and self-reports to measure face-to-face encounters, in a direct comparison of the communication behaviour of the same employees from the same organisation in two different physical workplace buildings. It is usually difficult to obtain suitable data for studies such as this one due to the infeasibility of simply moving an organisation from one building to another for the purposes of an experiment, owing to the high time, effort, and financial costs, and the fact that studies comparing organisations in different kinds of buildings cannot account for all the possible organisation-specific variables that might affect the validity of comparisons. Furthermore, while it is possible to study the effect of spatial configuration within a single organisation at lower cost, by simply changing office layout, for example, a fundamental aspect of this analysis is the nature of the spaces provided by the building (e.g. number and location of meeting rooms, food and drink areas, etc.), which cannot be changed without altering the building itself.

The experiment described above provides a rare example of a direct comparison of two different workplace buildings and the impact of the space on the communication of and potential for interaction between the same employees, in conjunction with the formal organisational structure. The results provide evidence building on the body of

existing work on this subject, supporting the idea that communal spaces could be important to provide opportunities for *communication for inspiration* between employees who may not work together and are in different organisational subunits, and that there is demonstrably more potential for encounters in these spaces between those who may not otherwise meet. The results also suggest that office and meeting room spaces are most likely still important for *communication for coordination* and *communication for information* between members of the same team, showing directly the value of both kinds of spaces to allow all of the forms of communication important for a thriving innovative organisation. Furthermore, the data suggest that the aims of the architects to encourage more use of shared spaces – both by members of the same organisational subunits, exemplified by the provision of more meeting rooms in the new building, and by members of different teams, as in the case of the cafeteria – were met, which also provides evidence that such architectural considerations can be of value, and that the role of space for groups at the level of a single building may be important to consider in building design.

These observations that different spaces in the building tend to have greater or lesser potential to host different kinds of interaction and communication can be seen to parallel the findings in Chapter 4, where I showed that some places around the city are much more likely to be meeting places for friends than are other places. In particular, the importance of the cafeteria in the workplace building for meetings between individuals who do not work together, but who may interact on a more socially-motivated and less work-motivated level, mirrors the social nature of food-oriented places and recreation venues in urban areas.

5.2.1.1 Limitations

There are a few points that should be remembered when interpreting the above results and drawing further conclusions; one is that the requirements for recording a contact using the active RFID badges are fairly stringent, and this may mean that while it is possible to mitigate the problem of false negatives (failing to record an encounter when one takes place) as outlined above, levels of contact occurring may have been underestimated. However, this issue is consistent across the two measurement periods in the two buildings, since the experiment was set up in the same way using the same technology, and so the comparison made is still valid, despite the fact that ab-

solute numbers of contacts should not be taken as completely accurate. One should also bear in mind that this is just one sample of one organisation, and should not be taken as representative. Different organisations might be affected differently under the same conditions; many more such studies would be needed in order to draw more general conclusions. Furthermore, since this specific organisation is a research laboratory, there may be specific characteristics of intra-organisational communication that might be different in other types of organisation, for example, those that are more commercial. Finally, this study concerned only the short-term impact of the different physical workplace environment on face-to-face communication. Other studies have shown that such face-to-face interaction between employees can have important effects on productivity and innovation [Pen12, SS12], but these are phenomena that require evaluation over a longer period. I have dealt here only with communication patterns measured over the short term, and not their longer-term consequences, but it would be possible in the future to examine such effects.

5.3 Summary

In support of the thesis that the relationship between physical space and social groups at different scales can be exploited to create or improve technological social applications, in this chapter I have studied the use of space at the level of a single building by different groups of people, in this case, colleagues in a workplace. The results of my analysis showed that informal spaces such as kitchens and cafeterias afford more opportunity for contact between groups, and that office and meeting rooms were in this case, as expected, used more for meetings between people in the same group. All of these types of communication are important in many workplace environments to enable coordination and collaboration between people involved in the same projects as well as the exchange of ideas between people who might not normally talk to one another. The evidence of the particular role that building space plays, as well as the demonstration of how it can be monitored using ubiquitous sensing technology, could be beneficially applied in the future to assess the 'communication health' of a workplace and help maintain effective interaction between employees, as well as informing the design of purpose-built workplaces. While actual implementation of large-scale systems for this purpose still faces privacy issues that are beyond the scope of my research, it is clear

that with recent and continuing advances in ubiquitous sensing technology that there is scope for such applications to help improve effective communication in the modern workplace.

Chapter 6

Conclusions

6.1 Thesis summary and contributions

In the preceding chapters I have presented and discussed the results of my research, in relation to the thesis: *the relationship between physical space and social groups at different scales can be exploited to create or improve technological social applications*. These results, and the contributions they represent to build on the existing body of work outlined in Chapter 2, are summarised below.

In Chapter 3, I analysed datasets from two online social networks with location information, which have the potential for communities in the social network to span across cities, or even around the world, given the ease of forming an online connection. The results showed that although in some online social networks such as Twitter, geographical space may not be a crucial factor constraining social connections, location still plays a role in the structure of online communities, an effect which may be more pronounced in specifically location-based services, such as Gowalla, which showed many communities whose members visit the same places offline. I then showed how even though space appears less important in Twitter communities, there are groups of friends in the Twitter network who do visit the same places, and presented a method for extracting these groups. Finally, I demonstrated that being able to find such location-based groups could potentially be applied to improve friend recommendation in online location-based social networks.

This research into the spatial properties of global-scale communities in online social networks builds on earlier work that analysed the geographic characteristics of

dyads, mostly in telephone communication networks but also online. It has historically been difficult to analyse larger groups in such networks due to lack of available data on social networks combined with fine-grained location information, but now that mobile Internet access and the huge popularity of online social networks makes data much more readily available, such research is feasible to conduct.

In Chapter 4, I studied groups in social networks at the intra-city scale, by analysing location-based social network and mobile telecoms data. I investigated clustering around places in the social network and the locations where people tend to meet their friends, and the results showed that many triangles exist in the networks such that their constituent friends visit common places, which suggests that groups within cities may be based around the places their members tend to frequent. My analysis suggests that there may be differences between the mobility of people on their own and when located with a group of their friends; in particular, the nature of the spaces in the city where people meet seems to be important, with people much more likely to meet their friends at some categories of places than others. I then showed how this idea can be used to improve the application of venue recommendation in location-based online social networks, specifically, when making recommendations for a group of friends to meet somewhere in the city, better results may be obtained by considering the group as a unit rather than making recommendations for individuals and trying to combine them.

This work on groups at the intra-city scale builds on previous work largely focused around the application of Delay Tolerant Networks, where contacts between people in a city and the community structure of the existing network are important for efficient routing, but the nature of spaces themselves such as the categories of venues in LBSNs have not generally been analysed on a large scale with respect to the groups, and explicit social ties such as those declared in an online social network are not as important as actual face-to-face contacts, even if between 'familiar strangers' [Mil92]. Previous work that has taken into account the types of spaces where people meet has also tended to be on a smaller scale than that possible with location-based social network data, due to using WiFi and GSM access points for localisation and Bluetooth devices such as mobile phones to detect people's location [KNY+09, KOP+10], making data collection more difficult in the past than it is now with location-aware online social services covering many different cities. The group venue recommendation problem has also not

previously been studied, with prior work on LBSN venue recommendation concerning only single users.

Finally, in Chapter 5, I focused on the role of space for groups at a much smaller scale, that of a single building. I studied the interplay between building space and face-to-face interactions between employees in a workplace, by directly comparing two sets of location and interaction traces collected using ubiquitous sensing technology, from the same sample of people but in two different workplace buildings. The results showed that different kinds of spaces are important for different kinds of communication within and between groups: interactions between people who work together, and which are likely to be for information and to co-ordinate work, are likely to take place in shared office spaces and meeting rooms, while informal spaces such as kitchens and cafeterias are more important for encounters between people who do not normally work together.

The investigation of the two different workplace buildings represents a first study of the interplay between building space and face-to-face interactions that used the particular unobtrusive ubiquitous sensing technology, thus overcoming some difficulties faced by previous work relying on self-reports and human observers. Furthermore, it made use of an opportunity to study the same group of people from the same organisation in two different buildings, which allows direct comparison between interactions in the two buildings without having to account for other variables that might affect comparisons involving different people working for different organisations. Opportunities to conduct experiments such as this are rare due to the time, effort, and financial cost involved in moving an organisation from one building to another, and combined with the historic lack of suitably unobtrusive technology capable of sensing reasonably reliably both face-to-face interactions and location, this has meant that such a direct comparative analysis of the interplay between two different building spaces and interactions within and between working groups has not been possible before. This research demonstrates that, since there is evidence that the factors I have studied are involved with communication in the workplace, which is known to be crucial for organisational productivity and innovation, such sensing technology could be applied in the workplaces of the future to monitor employees' communication automatically and provide them with feedback to help improve their interaction patterns. While investigation of this idea on a large scale raises many privacy issues and is beyond the scope

of my thesis, I have shown that such applications could be viable.

In summary, my research provides evidence that at least at present, space largely remains inseparable from social groups despite the ease of space-independent communication in today's technologically connected world, whether at the global scale of online social networks, or at the smaller scale of individual meeting places within a city or in the context of an office building. Different communication technologies may emerge and become obsolete, and the applications I have discussed and demonstrated here are necessarily linked to those specific technologies current at the present moment, but it does not seem implausible that the principle that the role that space plays in social groups at various scales may remain useful in technological social applications for some time into the future.

6.2 Directions for future research

The research I have presented in this dissertation builds upon the foundation of previous work, as described above, but also raises some questions and suggests directions for future research in the area.

An ongoing branch of enquiry might concern the changing role of space for social groups as location-based technology becomes more widely adopted; while the findings in this dissertation provide a picture of some aspects of the use of space by groups at different scales at the present time, technology by its nature is developing all the time and the situation may change as new ways to communicate emerge and make long-distance social connections even easier to maintain than they are at present.

There are other technological applications of the role of space in social groups besides those that I have demonstrated here that could form the focus of further research. For example, some research has been done into what affects with whom people wish to share their location [CRH11, PKK12, TLH+10, WKC+11], and also into group-based privacy management in online social networks via the automatic detection of groups [JO10, LVM+12]. Location-based groups such as those examined in Chapter 3 could potentially be used in this way to improve privacy controls in location-based online services.

At the scale of a city as investigated in Chapter 4, my analysis showed differences between individual and group mobility patterns, but the reasons behind these differ-

ences remain unclear and so in order to understand the observed phenomena further investigation is needed. It could be that by understanding the factors that drive group and solo mobility, new mobility models taking these effects into account could be derived, which would enable better location prediction and the improvement of associated applications such as targeted advertising, dissemination of relevant location-aware content, and mobile web search results.

It also still remains to investigate on a large scale the application discussed in Chapter 5, where I concluded that it seems likely that spatial factors do affect communication within and between groups in the workplace, and that it would be possible to monitor using sensing technology the interactions of employees, and their locations, in order to be able to maintain helpful levels of different kinds of communication and promote productivity and the exchange of information and ideas. Further research is required into the privacy concerns that would arise from such a monitoring system, and also of the kinds of feedback that it would be helpful to provide to employees about their communication behaviour in order to enable them to manage their interactions most effectively. Finally, it would be interesting to study the interplay between space and social groups in buildings other than workplaces such as those I discussed in Chapter 5, using the same methods of automatic data collection using ubiquitous sensing technology.

Bibliography

[ABL10] Yong-Yeol Ahn, James P. Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307), 2010.

[AH06] Thomas John Allen and Günter Henn. *The organization and architecture of innovation: Managing the flow of technology*. Routledge, 2006.

[ASBS00] Luıs A. Nunes Amaral, Antonio Scala, Marc Barthélémy, and Eugene H. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21), 2000.

[BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439), 1999.

[BEL+14a] Chloë Brown, Christos Efstratiou, Ilias Leontiadis, Daniele Quercia, and Cecilia Mascolo. Tracking serendipitous interactions: How individual cultures shape the office. In *Proceedings of the ACM Conference on Computer Supported Co-operative Work*. ACM, 2014.

[BEL+14b] Chloë Brown, Christos Efstratiou, Ilias Leontiadis, Daniele Quercia, Cecilia Mascolo, James Scott, and Peter Key. The architecture of innovation: Tracking face-to-face interactions with ubicomp technology. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014.

[BGLL08] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 2008.

[BHKL06] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*. ACM, 2006.

[BLM+14] Chloë Brown, Neal Lathia, Cecilia Mascolo, Anastasios Noulas, and Vincent D. Blondel. Group colocation behavior in technological social networks. *PLOS ONE*, 9(8):e105816, August 2014.

[BMBL09] Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. Network analysis in the social sciences. *Science*, 323(5916), 2009.

[BNMB13] Chloë Brown, Anastasios Noulas, Cecilia Mascolo, and Vincent Blondel. A place-focused model for social networks in cities. In *Proceedings of the ASE/IEEE International Conference on Social Computing*. IEEE, 2013.

[BNS+12a] Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. The importance of being placefriends: discovering location-focused online communities. In *Proceedings of the ACM Workshop on Online Social Networks*. ACM, 2012.

[BNS+12b] Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Where online friends meet: Social communities in location-based networks. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 2012.

[BNS+13] Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Social and place-focused communities in location-based online social networks. *European Physical Journal B*, 86(6), 2013.

[BS11] Betim Berjani and Thorsten Strufe. A recommendation system for spots in location-based online social networks. In *Proceedings of the ACM Workshop on Social Network Systems*. ACM, 2011.

[BSBS78] Norman M. Bradburn, Seymour Sudman, Ed Blair, and Carol Stocking. Question threat and response bias. *Public Opinion Quarterly*, 42(2), 1978.

[BSM10] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the ACM International Conference on World Wide Web*. ACM, 2010.

[Bur04] Ronald S. Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2), 2004.

[Cai01] Frances Cairncross. *The death of distance: How the communications revolution will change our lives*. Harvard Business Press, 2001.

[CCLS11] Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z. Sui. Exploring millions of footprints in location sharing services. In *Proceedings of the AAAI Internation Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 2011.

[CH56] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider's theory. *Psychological Review*, 63(5), 1956.

[CML11] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*. ACM, 2011.

[CRGP12] Michele Coscia, Giulio Rossetti, Fosca Giannotti, and Dino Pedreschi. DEMON: A local-first discovery method for overlapping communities. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*. ACM, 2012.

[CRH11] Henriette Cramer, Mattias Rost, and Lars Erik Holmquist. Performing a check-in: Emerging practices, norms and 'conflicts' in location-sharing using Foursquare. In *Proceedings of the ACM Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 2011.

[CS11] Jonathan Chang and Eric Sun. Location3: How users share and respond to location-based data on social networking sites. In *Proceedings of the AAAI International Conference on Weblogs and Social Media)*. Association for the Advancement of Artificial Intelligence, 2011.

[CSBR11] Francesco Calabrese, Zbigniew Smoreda, Vincent D. Blondel, and Carlo Ratti. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLOS ONE*, 6(7):e20814, 2011.

[CSS99] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10, 1999.

[CTH+10] Justin Cranshaw, Eran Toch, Jason Hong, Aniket Kittur, and Norman Sadeh. Bridging the gap between physical location and online social networks. In *Proceedings of the ACM International Conference on Ubiquitous Computing*. ACM, 2010.

[CVdBB+10] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-Francois Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLOS ONE*, 5(7), 2010.

[DCC12] Gabriel Deak, Kevin Curran, and Joan Condell. A survey of active and passive indoor localisation systems. *Computer Communications*, 35(16), 2012.

[dLM13] Manlio de Domenico, Antonio Lima, and Mirco Musolesi. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, 3, 2013.

[DPV05] Imre Derényi, Gergely Palla, and Tamás Vicsek. Clique percolation in random networks. *Physical Review Letters*, 94, 2005.

[EEBL11] Paul Expert, Tim S. Evans, Vincent D. Blondel, and Renaud Lambiotte. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, 108(19), 2011.

[Fac12] Facebook. Building better stories with location and friends. http://developers.facebook.com/blog/post/2012/03/07/building-better-stories-with-location-and-friends, 2012.

[Fel81] Scott L. Feld. The focused organization of social ties. *American Journal of Sociology*, 1981.

[Fli12] Flickr. The map. http://www.flickr.com/help/map/, 2012.

[For10] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486, 2010.

[FVA+12] Cristina Frà, Massimo Valla, Alessio Agneessens, Igor Bisio, and Fabio Lavagetto. Mobile sensing of users' motion and position context for automatic check-in suggestion and validation. In Tom Lovett and Eamonn O'Neill, editors, *Mobile Context Awareness*. Springer London, 2012.

[GHB08] Marta C. González, Cesar A. Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196), 2008.

[GLRW13] Jagadeesh Gorla, Neal Lathia, Stephen Robertson, and Jun Wang. Probabilistic group recommendation via information matching. In *Proceedings of the IW3C2 Conference on World Wide Web*. IW3C2, 2013.

[GN02] Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 2002.

[GPJB07] Ankur Gupta, Sanil Paul, Quentin Jones, and Cristian Borcea. Automatic identification of informal social groups and places for geo-social recommendations. *International Journal of Mobile Network Design and Innovation*, 2(3), 2007.

[GS05] Margo Gardner and Laurence Steinberg. Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental Psychology*, 41(4), 2005.

[GWT11] Anatoliy Gruzd, Barry Wellman, and Yuri Takhteyev. Imagining twitter as an imagined community. *American Behavioral Scientist*, 55(10), 2011.

[HCY11] Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 10(11), 2011.

[HD03] Russell A. Hill and Robin I. M. Dunbar. Social network size in humans. *Human Nature*, 14(1), 2003.

[Hen13] Steve Henn. 'Serendipitous Interaction' Key to Tech Firms' Workplace Design. *All Tech Considered (NPR blogs): `http://www.npr.org/blogs/alltechconsidered/2013/03/13/174195695`*, 2013. Accessed 19 February 2014.

[HFH+09] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1), 2009.

[HSL11a] T. Hossmann, Thrasyvoulos Spyropoulos, and F. Legendre. A complex network analysis of human mobility. In *Proceedings of the Workshops of the IEEE Conference on Computer Communications*. IEEE, 2011.

[HSL11b] Theus Hossmann, Thrasyvoulos Spyropoulos, and Franck Legendre. Putting contacts into context: Mobility modeling beyond inter-contact times. In *Proceedings of the ACM International Symposium on Mobile Ad-Hoc Networking and Computing*. ACM, 2011.

[ILGP74] Alan G. Ingham, George Levinger, James Graves, and Vaughn Peckham. The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology*, 10(4), 1974.

[Ins12] Instagram. How location works on instagram. `http://help.instagram.com/customer/portal/articles/183775`, 2012.

[ITM96] Ellen A. Isaacs, John C. Tang, and Trevor Morris. Piazza: A desktop environment supporting impromptu and planned interactions. In *Proceedings of the ACM Conference on Computer Supported Co-operative Work*. ACM, 1996.

[JM00] Phillip Jeffrey and Andrew McGrath. Sharing serendipity in the workplace. In *CVE*. ACM, 2000.

[JO10] Simon Jones and Eamonn O'Neill. Feasibility of structural network clustering for group-based privacy control in social networks. In *Proceedings of the ACM Symposium on Usable Privacy and Security*. ACM, 2010.

[KCE+09] Haewoon Kwak, Yoonchan Choi, Young-Ho Eom, Hawoong Jeong, and Sue Moon. Mining communities in networks: A solution for consistency and its evaluation. In *Proceedings of the ACM Conference on Internet Measurement*. ACM, 2009.

[KCHP08] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex 'Sandy' Pentland. Meeting mediator: enhancing group collaboration using sociometric feedback. In *Proceedings of the ACM Conference on Computer Supported Co-operative Work*. ACM, 2008.

[KCRB09] Gautier Krings, Francesco Calabrese, Carlo Ratti, and Vincent D. Blondel. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07), 2009.

[KFRC90] Robert E. Kraut, Robert S. Fish, Robert W. Root, and Barbara L. Chalfonte. Informal communication in organizations: Form, function, and technology. In *Human Reactions to Technology: Claremont Symposium on Applied Social Psychology*, 1990.

[KGA08] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about twitter. In *Proceedings of the ACM Workshop on Online Social Networks*. ACM, 2008.

[KKNG12] Juhi Kulshrestha, Farshad Kooti, Ashkan Nikravesh, and Krishna P. Gummadi. Geographic dissection of the twitter network. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 2012.

[KLPM10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the ACM International Conference on World Wide Web*. ACM, 2010.

[KMG+13] Reuben Kirkham, Sebastian Mellor, David Green, Jiun-Shian Lin, Karim Ladha, Cassim Ladha, Daniel Jackson, Patrick Olivier, Peter Wright, and Thomas Ploetz. The break-time barometer: an exploratory system for workplace break-time social awareness. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2013.

[KNT10] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and evolution of online social networks. In *Link Mining: Models, Algorithms, and Applications*. Springer, 2010.

[KNY+09] Vassilis Kostakos, Tom Nicolai, Eiko Yoneki, Eamonn O'Neill, Holger Kenn, and Jon Crowcroft. Understanding and measuring the urban pervasive infrastructure. *Personal and Ubiquitous Computing*, 13(5), 2009.

[KOP+10] Vassilis Kostakos, Eamonn O'Neill, Alan Penn, George Roussos, and Dikaios Papadongonas. Brief encounters: Sensing, modeling and visualizing urban mobility and copresence networks. *ACM Transactions on Computer-Human Interaction*, 17(1), 2010.

[LBdK+08] Renaud Lambiotte, Vincent D. Blondel, Cristobald de Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul van Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21), 2008.

[LCW+11] Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. I'm the mayor of my house: Examining why people use

foursquare-a social-driven location sharing application. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, 2011.

[LNNK+05] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33), 2005.

[LVM+12] Yabing Liu, Bimal Viswanath, Mainack Mondal, Krishna P. Gummadi, and Alan Mislove. Simplifying friendlist management. In *Proceedings of the ACM International Conference on World Wide Web*. ACM, 2012.

[Mil92] Stanley Milgram. The familiar stranger. In John Sabini and Maury Silver, editors, *The individual in a social world: Essays and experiments*. McGraw-Hill Higher Education, 1992.

[MLA+11] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. Understanding the demographics of twitter users. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*, volume 11. Association for the Advancement of Artificial Intelligence, 2011.

[MMG+07] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the ACM Conference on Internet Measurement*. ACM, 2007.

[MWC10] Diana Mok, Barry Wellman, and Juan Carrasco. Does distance matter in the age of the internet? *Urban Studies*, 47(13), 2010.

[New06] Mark E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 2006.

[NRC08] Atif Nazir, Saqib Raza, and Chen-Nee Chuah. Unveiling facebook: A measurement study of social network based applications. In *Proceedings of the ACM Conference on Internet Measurement*. ACM, 2008.

[NSL+12] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PLOS ONE*, 7(5), 2012.

[NSLM12a] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. Mining user mobility features for next place prediction in location-based services. In *Proceedings of the IEEE International Conference on Data Mining*. IEEE, 2012.

[NSLM12b] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. A random walk around the city: New venue recommendation in location-based social networks. In *Proceedings of the IEEE International Conference on Social Computing*. IEEE, 2012.

[NSMP11] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. An empirical study of geographic user activity patterns in foursquare. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 2011.

[OAG+11] Jukka-Pekka Onnela, Samuel Arbesman, Marta C. González, Albert-László Barabási, and Nicholas A. Christakis. Geographic constraints on social network groups. *PLOS ONE*, 6(4), 2011.

[OOWK+09] Daniel Olguín-Olguín, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex 'Sandy' Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(1), 2009.

[OSH+07] J.-P. Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18), 2007.

[PBB11] Susan Juan Pan, Daniel J. Boston, and Cristian Borcea. Analysis of fusing online and co-presence social networks. In *Proceedings of the*

Workshops of the IEEE International Conference on Pervasive Computing and Communications. IEEE, 2011.

[PBC+11] André Panisson, Alain Barrat, Ciro Cattuto, Wouter Van Den Broeck, Giancarlo Ruffo, and Rossano Schifanella. On the dynamics of human proximity for data diffusion in ad-hoc networks. *Ad-Hoc Networks*, 2011.

[PBV07] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446(7136), 2007.

[PDV99] Alan Penn, Jake Desyllas, and Laura Vaughan. The space of innovation: interaction and communication in the work environment. *Environment and Planning B*, 26, 1999.

[Pen12] Alex 'Sandy' Pentland. The new science of building great teams. *Harvard Business Review*, 90(4), 2012.

[PES+10] Josep M. Pujol, Vijay Erramilli, Georgos Siganos, Xiaoyuan Yang, Nikos Laoutaris, Parminder Chhabra, and Pablo Rodriguez. The little engine(s) that could: scaling online social networks. In *Proceedings of the ACM SIGCOMM Conference on Data Communication*. ACM, 2010.

[PKK12] Xinru Page, Alfred Kobsa, and Bart P. Knijnenburg. Don't disturb my circles! boundary preservation is at the center of location-sharing concerns. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 2012.

[PKVS11] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. Community detection in social media. *Data Mining and Knowledge Discovery*, 2011.

[Qui92] John R. Quinlan. Learning with continuous classes. In *World Scientific AI*, 1992.

[RAK07] Usha N. Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, September 2007.

[RCE08] Omer Rashid, Paul Coulton, and Reuben Edwards. Providing location based information/advertising for existing mobile phone users. *Personal and Ubiquitous Computing*, 12:3–10, 2008.

[RHM+10] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton, and Susanna Hertrich. Ambient influence: Can twinkly lights lure and abstract representations trigger behavioral change? In *Proceedings of the ACM International Conference on Ubiquitous Computing*. ACM, 2010.

[RMMR11] Kiran K. Rachuri, Cecilia Mascolo, Mirco Musolesi, and Peter J. Rentfrow. Sociablesense: Exploring the trade-offs of adaptive sampling and computation offloading for social sensing. In *Proceedings of the ACM International Conference on Mobile Computing and Networking*. ACM, 2011.

[SBL+09] Kerstin Sailer, Andrew Budgen, Nathan Lonsdale, Alasdair Turner, and Alan Penn. Evidence-based design: Theoretical and practical reflections of an emerging approach in office architecture. 2009.

[SDBA12] Alistair Sutcliffe, Robin Dunbar, Jens Binder, and Holly Arrow. Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British Journal of Psychology*, 103(2), 2012.

[SM11] Kerstin Sailer and Ian McCulloh. Social networks and spatial configuration—how office layouts drive social interaction. *Social Networks*, 34(1), 2011.

[SMML10] Salvatore Scellato, Cecilia Mascolo, Mirco Musolesi, and Vito Latora. Distance matters: Geo-social metrics for online social networks. In *Proceedings of the USENIX Workshop on Online Social Networks*. USENIX Association, 2010.

[SNLM11] Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo. Socio-spatial properties of online location-based social networks. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 2011.

[SNM11] Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Exploiting place features in link prediction on location-based social networks. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*. ACM, 2011.

[SP09] Kerstin Sailer and Alan Penn. Spatiality and transpatiality in workplace environments. In *Proceedings of the International Space Syntax Symposium*. Royal Institute of Technology, 2009.

[SS12] James B. Stryker and Michael D. Santoro. Facilitating face-to-face communication in high-tech teams. *Research-Technology Management*, 55(1), 2012.

[Ste41] John Q. Stewart. An inverse distance variation for certain social influences. *Science*, 93(2404), 1941.

[TBL13] Lisa Thomas, Pam Briggs, and Linda Little. Location tracking via social networking sites. In *Proceedings of the ACM Conference on Web Science*. ACM, 2013.

[TCD+10] Eran Toch, Justin Cranshaw, Paul Hankes Drielsma, Janice Y. Tsai, Patrick Gage Kelley, James Springfield, Lorrie Cranor, Jason Hong, and Norman Sadeh. Empirical models of privacy in location sharing. In *Proceedings of the ACM International Conference on Ubiquitous Computing*. ACM, 2010.

[TG08] Umut Toker and Denis O. Gray. Innovation spaces: Workspace planning and innovation in US university research centers. *Research Policy*, 37(2), 2008.

[TGW12] Yuri Takhteyev, Anatoliy Gruzd, and Barry Wellman. Geography of Twitter networks. *Social networks*, 34(1), 2012.

[The12] The Next Web. Foursquare hits 20 million users and 2 billion check-ins. http://thenextweb.com/socialmedia/2012/04/16/foursquare-hits-20-million-users-and-20-billion-checkins-seeing-a-million-new-signups-per-month, 2012.

[TKCS10] Janice Y. Tsai, Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Location-sharing technologies: Privacy risks and controls. *I/S: A Journal of Law and Policy for the Information Society*, 6:119, 2010.

[TLH+10] Karen P. Tang, Jialiu Lin, Jason I. Hong, Daniel P. Siewiorek, and Norman Sadeh. Rethinking location sharing: Exploring the implications of social-driven vs. purpose-driven location sharing. In *Proceedings of the ACM International Conference on Ubiquitous Computing*. ACM, 2010.

[Twi09] Twitter. Think globally, tweet locally. http://blog.twitter.com/2009/11/think-globally-tweet-locally.html, 2009.

[vdM08] Thea F. van de Mortel. Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, 25(4), 2008.

[Wat99] Duncan J. Watts. Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105(2), 1999.

[WF01] Andreas Wagner and David A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 268(1478), 2001.

[WFDJ94] Steve Whittaker, David Frohlich, and Owen Daly-Jones. Informal workplace communication: What is it like and how might we support it? In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, 1994.

[Why43] William Foote Whyte. *Street corner society: The social structure of an Italian slum*. University of Chicago Press, 1943.

[WKC+11] Jason Wiese, Patrick Gage Kelley, Lorrie Faith Cranor, Laura Dabbish, Jason I. Hong, and John Zimmerman. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. In *Proceedings of the ACM International Conference on Ubiquitous Computing*. ACM, 2011.

[WOOKP10] Benjamin Waber, Daniel Olguín-Olguín, Taemie Kim, and Alex 'Sandy' Pentland. Productivity through coffee breaks: Changing social networks by changing break structure. *Available at SSRN 1586375*, 2010.

[WPD+10] Mike P. Wittie, Veljko Pejovic, Lara Deek, Kevin C. Almeroth, and Ben Y. Zhao. Exploiting locality of interest in online social networks. In *Proceedings of the ACM Conference on Internet Measurement*. ACM, 2010.

[WW97] Y. Wang and I. H. Witten. Induction of model trees for predicting continuous classes. In *Springer ECML*, 1997.

[YLL12] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation: a generative model approach. In *Proceedings of the International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2012.

[YYL10] Mao Ye, Peifeng Yin, and Wang-Chien Lee. Location recommendation for location-based social networks. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2010.

[YZYW13] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhu Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the ACM Conference on Hypertext and Social Media*. ACM, 2013.

[ZSW+12] Xiaohan Zhao, Alessandra Sala, Christo Wilson, Xiao Wang, Sabrina Gaito, Haitao Zheng, and Ben Y. Zhao. Multi-scale dynamics in a mas-

sive online social network. In *Proceedings of the ACM Conference on Internet Measurement*. ACM, 2012.