Evolution of social networks through smart phones and radio sensors

by

© *Sipan Ye*

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of *Science*

Department of *Computer Science*

Memorial University of Newfoundland

May 2018

St. John's Newfoundland

Abstract

Social interactions have been an integral part of human civilization. It reflects human society and its evolution. The first challenge in the research of social networks is data acquisition. The high cost and low efficiency are always restrictions. In addition, most existing studies of network problems are using single datasets to build their network models. It is paramount to find a general method of obtaining a highly accurate network model to represent social interactions. Therefore, we propose a cross-platform system and strategy to collect data through radio sensors and design a combined scheme with multiple datasets in order to settle this problem. Moreover, we use complex network theory to build our network models. The next challenge is network dynamic. A larger number of real-world networks are dynamic, i.e. social networks, as the topology of a network changes over time. It is also hard to describe the topological variance of the network using a static network model where it does not have any time features. Thus, we propose a weighted temporal network model to illustrate the time effect of social network problems. In this study, we also analyze the effect of friendship on human social interactions and activities. The relationships among networks are shown as well. Furthermore, we show the combined network model provides a highly efficient way to construct social networks.

Keywords: Social Network, Complex Network, Social Interaction, Temporal Network, iBeacon.

Acknowledgements

Thanks to all colleagues for helping me in two rounds of experiments. Three labs of you provided access to deploy equipment and collect data. I express my deepest thanks to Dr. Yuanzhu Chen. His assistance in both technical and living fields, including his academic advice and patience, was integral to this research. Funding for this research was also provided by Dr. Yuanzhu Chen. I would also like to thank Dr. Ting Hu and Dr. Cheng Li for their careful and precious guidance which were extremely valuable for my study both theoretically and practically. I acknowledge with gratitude to Dr. Vander Wal who is always sincere and helpful in my study and gave me a chance to apply my research to the biology field. All your sincerely helps allows me to complete my study that would not have been possible otherwise. I also thank Chen Zhang provides helpful comments that improved the manuscript. WineMocol lab provides discussions on social network related papers which induces helpful suggestions for this study. The WineMocol group study about scientific writing also gives me strong skills to write this thesis. I perceive as the opportunity of studying in Memorial University of Newfoundland as a great milestone in my life. I will strive to use gained skills and knowledge for further improvements.

I will appreciate again to underscore the unwavering support of my supervisor Dr. Chen for these three years. I would also like to thank my parents and friends who helped me a lot in my life and study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Most things are not unique and have fellow-creatures. We, as individuals, are a part of social networks and, as creatures, are elements of biological systems. Analyzing characteristics of the population in broader terms is a mutual way to reveal problems of networks. Mathematically, the study of networks is known as graph theory [1]. It has been developed in many fields, the Konigsberg bridge problem [2], the relationships among social entities [3] and wildlife migrations [4]. Many real networks which have irregular and complex structures contain thousands or more nodes and evolve dynamically via time attracts scientists interests in the study of complex networks. It is involved in various fields. For example, transmission of infectious diseases is always a serious problem in all species [5][6]. In addition, animal migration is also an attracted topic in the field of biology [4]. Biologists focus on studying and understanding the migration and evolution of wildlife. As a result, there has been a significant increase in demand for research on the topic recently. People want to utilize these algorithms to study what the disease dynamics are, what activities of wildlife in their life are, and how they construct their own society [1][7][8][9].

On the other hand, a number of medical scientists focus on the infection and prediction of disease in human society, such as HIV [5], SARS [6] and sexually transmitted diseases [10]. Google also estimates the flu trends via people's search patterns [11]. Not only in these areas, but other scientists show interests in the study of human society. A great deal of research has been done in recent years. In 1967, Milgram held an experiment by sending mails to one person and transferring to others in a small town to study the connections and structure in human society [12]. With the development of social media platforms, Facebook, Twitter, and Weibo have become popular communication tools in human society and daily life [13][14]. By crawling users' relations and tweets through Twitter's open API, Kwak et al. proved that such interaction data can have a great potential for further research compared to Google keywords search data [15].

1.2 Motivations and Challenges

1.2.1 A multiple network data model

Social network works mainly based on human mobilities and behaviours. It also includes differently scaled actual networks and infrastructures. Online social media, airports systems and university communities are good instances. Such a social structure and interaction built on the activities, behaviours, and connections of individuals can show some aspects of personalities and even predict the behaviours. The studies on the migration and evolution of social networks show a significant increase of research demands [15][16][17]. These studies are utilized to get solutions to their specific problems.

To analyze the networks and predict behaviours, at first, we defined a network and analysis method with actual data as fundamental resources and mathematical definitions. As the social network can be constructed from various data such as online social interactions data [3], human transportation data [16], cell phone data [18] and mobile apps checking data. For example, an efficient way to get the actual data was proposed from the online social networks such as Facebook, Twitter and Weibo. Most studies only use single network data to construct the network. The limitations of such approach are obvious. For instance, the global positioning system (GPS) function cannot work accurately inside the buildings. Networks based on it cannot reflect the behaviours and interactions of individuals when they are indoor.

On the other hand, it is very hard to describe network analysis tools that can be used to resolve network structure, revert real-world behaviours and anticipate trends. Although the network is fed with actual data, it is not always easy to figure out the structure and features. A few universal models have been proposed to conduct the analysis on it [16][18].

1.2.2 Data acquisition via Bluetooth sensors

One way to obtain network data requires cooperation with different industries, for instance, geographic data from human transportation information [16], cellphone service carrier [18] and mobility data [19]. Alternatively, many researchers also propose their own data collection methods, such as crawling data from social media [13][20] and survey data [14]. Nevertheless, time and cost of these approaches are unbearable. Furthermore, plenty of customized devices have been proposed to collect the social network data, such as proximity loggers [21] and GPS collectors [22]. Still, the cost of devices and the difficulty of deployment limit the feasibility of obtaining these data.

Recently, occupancy of mobile phones increases rapidly [23]. A large number of mobile device data collection schemes have been proposed. The convenience and extensive coverage of it shows an enormous potential for agile and high efficient data collection. Thus, introducing iPhone for both data broadcasting and collecting is natural. At the same time, when using the iPhone for data collection, we can easily get the timeline of each data collection. This will help us analyze and format our contact network model.

A restriction of building a Bluetooth connection via two devices requires a couple of seconds, and even longer when two pedestrians pass by. Sometimes, devices may not have enough time to build a connection. It will cause invalid data for these two individuals encounters. To address this problem, we choose iBeacon, a Bluetooth technology based radio sensor as another data broader and design a specific data collecting scheme. All smartphones can detect such devices' signal. Additionally, the price and power consumption of it is low.

1.2.3 Social interactions comparison based on temporal network

A common solution to solve the problems of the complex system is to create a network model and design analysis tools. Traditionally, static network models, which contains nodes, edges and attributes of them, are suitable for most situations. According to the previous discussion, it has been involved in epidemiology filed [5][6], transportation research [16], social network studies [3][18] and many other different fields. Static network models indeed give a precise description and a clear visualization of the network problem on a macroscale.

However, many large complex networks are dynamic and the topologies may vary in different time periods. Specifically, in social networks, interactions between individuals change rapidly and friendship may vary through time [24]. Relations between people will break down and establish. Thus, time elements play a crucial part of the networks. The topologies of these networks are changed in different time periods [24]. Static network models cannot represent these dynamic changes and we will lose lots of information if overpass time sequence. Thus, we study a temporal network method to recover the time influence in this paper. By expanding networks on the timeline and analyzing changes of a network topology, it provides us with a time-based mathematical and computational tools to understand the networks better and observe the evolution of the networks. Thus, we design two dynamic graphs (snapshots), to transform our two social interactions networks into temporal network models. Then we also discuss centralities measurements and analyze the time influence on network topologies and social activities.

On the other hand, the ties among nodes contain a lot of information on complex networks, i.e. the capacity of transportation networks. The weight of temporal network is not well studied [25], where it plays an important role in evaluating the network. For instance, Nathan et al. pointed out that the edge weight which demonstrates the relations between two nodes will affect the network structure based on time [24]. Thus, studying the influence of weight on the network topology change is a problem. We discuss a weighted temporal network model and measurement tools in this paper to address the problem.

1.3 Contributions

In our research, we proposed a complex network approach using multiple network data to rebuild human interaction and social structure. In the meantime, a method to combine multiple network data is discussed in the study. It bridges the gaps where single type dataset limits the robust and accuracy of the network. We also evaluate the performance of our approach. To improve the scalability of data collection, we introduce iBeacon sensors as data broadcasting devices which the cost is very low [26]. Meanwhile, due to the low energy cost of such device, iBeacon devices can provide months, even more than one year working period in a small size and weight device [27]. Additionally, the relations between friendship and real social interactions are studied in this thesis. Through the method, we believe that it can reduce researchers' efforts and works radically. A small-scale experiment was conducted at Memorial University of Newfoundland (MUN). Daily movements of participants were tracked by cellphones and iBeacon sensors for two months. In this thesis, we also propose a new weighted temporal network model to incorporate the important dynamics in timeline. Moreover, we present the process of converting static network models to temporal models. The importance of time elements on social relationships and activities is demonstrated. We also evaluate the weight effects on temporal network models.

1.4 Thesis outline

In this paper, we firstly introduced the concepts and related studies of complex network. We also analyze the importance of complex network applications and research in different fields. Tools and methods to collect network data are also compared in this part.

In the next chapter, We induced a network data combined approach to extend applicability and improve the scalability of constructing networks. Especially, we focus on human social network study with our own datasets. Robust and performance of it are evaluated in this part. This study fixes the gaps of the normal single data network, for instance, the limitation of GPS data in the indoor environment.

A cross-platform Bluetooth hardware is also discussed in the next chapter. It enables us to obtain more data with low cost but a wide range of usage. Associating with our data combined approach, we generalize three network models to reflect human interactions and social structures. We also compare features of different networks and analyze the relations in static methods, such as centralities to verify our results.

Additionally, traditional methods consider all data together in a network. However, for social networks, individuals and relations are changed with time pass by. Networks may change rapidly via time. We generalize a temporal network model and discuss the time impact on different networks.

Chapter 2

Related Work

2.1 Applications of complex network

The original idea of the complex network is proposed by Radcliffe-Brown [28]. It is built for anthropologists and sociologists who expresses more interests in the concept of "social structure". In order to understand the "interweaving" and "interlocking" relations which social actions are organized, the "social network" concept was proposed. Researchers began to study and analyze the content, the actions, and the connections of the "network".

Complex network models help researchers, like Meyers and Pourbohloul, to obtain better accurate results of the SARS disease infection simulations and realistic database [6]. It has already been involved in many different areas for many years. In epidemiology field, Shweta et al. found out that contact network and homogeneous models showed more powerful and higher efficient based on their simulation [29]. In the field of epidemiology, especially infectious disease is strongly corresponding to network theory [30]. Leon found out that decades before, scientists have proved the connections between there two fields [31][32]. Epidemic spreading analysis attracts many scientists, epidemiologists, and mathematicians to model complex network framework for research [33]. Throughout the network model, biologists can generate and analyze results that the actual source and targets in disease infections [17].

2.2 Research on social network

Furthermore, researchers noticed that social relationship would significantly affect human health [34]. Major social network studies are conducted on complex networks [6][16][18][19]. Albert and et al. have proposed degree distribution and coefficient concepts to evaluate networks [7]. Furthermore, Newman has introduced centralities and defined other algorithms to estimate the networks [8][35]. Mobile device technology has been growing rapidly in recent years. A team from Stanford University proposed a social network model based on the users' locations. The model could reliably predict human movements and locations dynamically in their future activities [18]. However, such approaches consume the power of user cellphones largely, so the users are affected. Besides, Apple prohibits the ability of Wi-Fi usage for developing and researching. This would sharply reduce the data source of research, like Locaccino of Justin and et al. [36]. We utilize Bluetooth technology where almost every smartphone can easily access. iBeacon, a wireless Bluetooth low energy proximity logger, is defined as Broader for data broadcasting with UUID, major and minor value, shown in Table 2.1 for helping users distinguish devices [27]. Major value is used to tag different groups while devices will be given minor value to be identified.

Attributes	Type	Definition
UUID	String	Universally Unique Identifier
Major value	Int	Major value of device, for distinguishing in groups.
Minor value	Int	Minor value of device, for distinguishing devices in a group.
RSSI	Int	Received Signal Strength Indicator.

Table 2.1: Data format definition

In the field of social networks, many researchers pay attention to friendship and mobility [18], population mobility model [16] and predictions [19]. The results of studies on these topics can predict human mobility, develop cities and prevent disease infections. Although different networks are analyzed, such as human movement network and travel location network, various datasets are only fitted to the same network. There is lack of studies on comparisons among different networks. The deviations between varied networks within same individuals can also reflect difference between human relationship and daily activities. We limit our discussion on this specific point to dig more information from comparing different networks.

2.3 Hardware of network data collection

In biology, complex network can represent communications and living habits of animals. This enables researchers to observe the evolution [21] and transmission of disease [37] of the species. In this case, scientists have begun to track animals for research, like disease infection and migration, many years, even 100 years [4]. Traditionally, biologists would catch the animals and tag them, like Winkler et al. [38]. By observing these targets, researchers can generate sampling data [39]. However, the manually tagging and observing by a human would always cause errors and mistakes. In addition, it is high cost, laborious, and low accuracy due to the budget of hiring people and longtime experiment. With the development of science and technology, a large number of methods and techniques to solve this problem were published in these years. In Cryan's research, he and his team chose a biology approach to track long distance animals' movements. Hydrogen isotope analysis is a noted method. People measured the hydrogen isotope ratio of the bat hair to distinguish bats and determine their migratory movement [40].

Elsewhere, GPS is also introduced to animal tracking. Researchers tied the device on animals. The weight of the devices has been reduced to only about 50g. Thus, some birds scientists use such device for tracking migration of small birds [4]. However, the budget of such hardware is still pretty high. Moreover, the battery life of these devices cannot support long distance and timeline tracking. Even more, Brooks and his colleges compared two types of GPS sensor on the zebra to study their behaviour. They found out that the weight and fit, even in an accepted range or limitation, will actually take effects on animals' daily behaviour, which will also affect the research and result [22].

Doyle and his team have come up with another wireless solution by using Ultra High Frequency (UHF) radio signals on sheep [41]. These proximity loggers record data of social interactions between pairs of sheep and researchers can study the sociality of sheep flock. Some existed approaches can help researchers gather lots of important information about wildlife, but require specific hardware. Rutz et al. [42] and Hamede et al. [21] both initialize their research on the customized proximity loggers [43] for animals tracking and contact network forming, which was figured out that the network will be changed based on the mature season and the sex. P. Sikka established a new kind of wireless sensor device, Mica system. However, even during their series of experiments, they suffered a lot between their own designed two versions hardware, Mica 1 and Mica 2 [44]. The limitation is obviously detected that specific and customized hardware would restrict the potential application of reusability or even cross-cutting research possibility and increase the cost. As smartphones occupy a large proportion in human society, the trend of it also represents upward [45]. Data collection from smart devices is used in a large number of studies. To be more specific, with the development of social media network and maturity of cellphones technology, many researchers want to find a balance and combination via these methods and have proposed lots of approaches for social network and contact network studies. A team from Standford university public a social network model based on the users' locations, which the model could reliably predict human movements and locations dynamically in their future activities [18]. However, such approaches will cost and influent users' daily usage of cellphones a lot due to the large power consumption.

Additionally, based on the Bluetooth technology, all smartphones can easily access the iBeacon. Furthermore, some scientists also tried to combine online social network data with offline location data of users. They utilized the locations of users to verify, distinguish, and predict two users have a connection in online social media or not, which this is an indeed practical way to integrate advantages from two different stages [36]. Unlike these methods, the methodology we present in this thesis concentrates more on realistic data for generating human real behaviours and their connections.

2.4 Time effects on complex systems

Complex network theory is involved in many network-based phenomena in the real world, from biology, for instance, birds migrations [4], to epidemiology such as the spread of HIV [32]. In epidemiology field, Shweta et al. found out that contact network and homogeneous models showed more powerful and efficient based on their simulation [29]. Epidemic spreading analysis attracts many scientists, epidemiologists, and mathematicians to model complex network framework for research [33]. Furthermore, a large number of social network studies are conducted on complex networks [16][18].

However, these network problems are studied by static network models. In other words, the topology of the network has invariable features. The nodes and edges in the network are never changed. However, in real life, such networks are dynamic where old nodes will be removed and new nodes will be added in the network. The connectivities of nodes are also changeable. Nathan and his team have token an experiment on the campus of Massachusetts Institute of Technology [24]. They collected two kinds of data, self-report data from the survey and behavioural data from users' cell phones. Two networks are created and compared with each other from the two datasets. Nathan et al. have analyzed them by different time dimensions and found that interactions between friends will change rapidly in the different time period [24]. It brings an idea about time effects on the social networks. However, they still use static network models to process data at given time intervals. The influence of time continuity and change on networks are not captured.

To make up the gaps, Tang and his team proposed a new temporal distance metrics to evaluate the evolution of a network on time dimension [46]. In addition, Hyoungshick and Ross have studied on temporal node centrality. A dynamic network concept and a time-ordered graph have discussed and a static network is converted into a temporal network in their study [25]. Moreover, temporal centralities tools are also studied to evaluate the performance of the temporal networks. The time effects on network topology is well represented. However, an important feature on network measurements is ignored, edge weight. Thus, we consider edge weights on temporal edges in networks to improve the temporal centralities measurements.

Chapter 3

Inducing Social Interaction Through Co-location Using Fused Network Data

Data is a vital characteristic of network problems. It will effect the robustness of the network. To enhance the capability of networks, we propose a data combined method to reduce the errors of the original datasets and improve the performance of the networks in this chapter. Furthermore, by studying social interactions data and friendship data, we evaluate the diversity of social networks.

3.1 Research questions

Most social networks research relies on a single type of data, for instance, GPS, Bluetooth or online social applications interaction data, to analyze network structures and predict human mobility [18]. Fewer groups use two or more types of data to construct networks, which still has limitations. The use of GPS is limited indoor due to the signal-less. This type of data is the main part of constructing a network in many social network studies. Interactions between two smart devices [18] via Bluetooth is also a major approach. Even though, it provides stable and usable data collection, the ability and range of application of cross-platform is a strict problem. Building connections between two devices costs lots of time. In reality, people do not have enough time waiting for building connections when they pass by. Thus, using multiple data together is much necessarily needed. This method is proposed in this paper.

To show how robust of the method, an easy way to measure the networks based on combined data becomes the first challenge. In addition, the structure of a network varies a lot depending on data types. Normalization between two kinds of networks is another challenge.

Despite users' actual interactions in daily life, relationships among people put influence on their mobility [18]. Friendship relies on human's subjective opinions. Human daily activities are affected by their work and family events [47]. Human mobility networks and friendship network have been well studied respectively. However, they may vary from each other. We focus our research on the relations between friendship and social interactions, including difference and similarity between them.

3.2 Methodology

We assume data combined networks and phenomenon of human relations between actual interactions are given. The network structures are derived and difference between these networks are shown using a complex network approach. Starting with the data collection, we utilize general hardware to save cost and improve the scalability. In particular, we design an iOS-based application to monitor and collect daily activities and interactions of participants.

Representing data into networks is a crucial concern for studying and analyzing social networks. The network structures vary depending on the definition of nodes and edges. According to the dataset, we have defined three different types of networks to observe human interactions. We apply centralities and coefficient algorithms into our network models to evaluate the networks and analyze them. This process compares the difference and similarity between networks.

3.2.1 Data collection

- *Hardware:* Normally, a data collection method only works on a specific research [21][41]. To avoid this, we choose iPhone as both a data collecting and data providing tool. Bluetooth accessories in the cell phones considerably save power consumption and cost. This design can save the cost, provide the scalability and improve the reusability for future work.

- *Data Definition:*

  – Friendship data: To generate people actual relations, we propose a survey asking participants rank their friends by friendship from high to low. This data is used to observe and rebuild users actual social relation network.

  – Co-location data: When two users pass by each other, their smartphones will automatically build a connection between them. The data can be expected to clearly depict human's interaction.

  – Location data: Trace of users' daily activities is monitored. We set the accuracy of the GPS as hundreds of meters which can observe the users' locations but still power efficient.

Attributes	Type	Definition
UUID	String	Universally Unique Identifier.
Timestamp	Date	Data recorded time.
RSSI	Int	Received Signal Strength Indicator.
Longitude	Float	The longitude information of user's location.
Latitude	Float	The latitude information of user's location.

Table 3.1: Data format definition

- *WMLSIRecorder:* Collecting above data, we develop an iOS application WineMocol Social Interactions Recorder (WMLSIRecorder) to record our datasets and construct them. Every data is recorded with a timestamp. Data communication range is set to 10m based on the Bluetooth protocol. Data are recorded only when detected devices are in the range of 10m. Additionally, location information is stored when users move.

The app runs in the background mode and stores the above data. As well as, our application provides a low power cost which only increases $3\% - 5\%$ in daily use. It may be acceptable for the experiment and future usage.

3.2.2 Definition of networks

Using datasets we have collected, we characterize three variations from the real-world social networks. In particular, we define the nodes and edges of the networks. All of them can be considered as human's activities and communications with others. To our best knowledge, major social network research only builds the data model using one kind of data. We propose a data combined approach to model the diversity of network structure and enhance the accuracy of the results. Nodes V are denoted by mobile devices which represent users in all networks.

- Friendship network:

$$G_f = (V, E_f, W_f), \tag{3.1}$$

  where it is summarized from the friendship data, a directed edge E_f from node u to node v describes the relation between u and v provided by u. The weight W_f is the set of the relationship's rank. The closer friendship u has with v, the larger weight the corresponding edge will be recorded.

- Co-location network: We reconcile data from co-location dataset with location data from location dataset for constructing a more reliable network structure to reflect human social network.

$$G_c = (V, E_c, W_c), \tag{3.2}$$

where E_c is the set of weighted and undirected edges among users. The weight function $W_c : E_c \rightarrow \mathbb{R}^+$, which carries the number of times that the two involved users have been co-located during the experiment period.

- Location network:

$$G_l = (V, E_l, W_l) \tag{3.3}$$

According to the location dataset, two users, who have a close Haversine distance (great circle distance of the surface of the earth) within a short time period, will be built an edge E_l between them. Morever, every two nodes only has one edge.

3.2.3 Network comparison

Network structures are quite different depending on various data models. To prove the robust and stability of our data combined model, we use centralities and coefficient in evaluating networks. Normalization is also a challenge when comparing two different networks. We use Pearson Coefficient which suggests using expected number of each value to compare their difference and similarity.

$$r_{ij} = \frac{\text{cov}(A_i, A_j)}{\sigma_i \sigma_j} = \frac{\sum_k (A_{ik} - \langle A_i \rangle)(A_{jk} - \langle A_j \rangle)}{\sqrt{\sum_k (A_{ik} - \langle A_i \rangle)^2} \sqrt{\sum_k (A_{jk} - \langle A_j \rangle)^2}} \tag{3.4}$$

$$\sum_k A_{ik} A_{jk} - \frac{k_i k_j}{n} = \sum_k A_{ik} A_{jk} - \frac{1}{n} \sum_k A_{ik} A_{jk} \tag{3.5}$$

$$= \sum_k (A_{ik} - \langle A_i \rangle)(A_{jk} - \langle A_j \rangle) \tag{3.6}$$

where $\langle A_i \rangle$ denotes the mean $n^{-1} \sum_k A_{ik}$ of the elements of the ith row of the adjacency matrix [35]. It calculates the normalized covariance, similar to the cosine.

Thus, the range of the coefficient is limited in $[-1, 1]$, where the result is close to the 1, the high similarity it has. If the evaluation of the combined network matches up the demonstrated single data network, for instance, GPS based network and co-location based network, we can determine that the combined work can fit the gaps and errors that single data network cannot have.

3.2.4 Vertex centralities

To evaluate networks, we utilize centralities algorithms. The algorithms are main measurements in complex networks [7] to convert network models into mathematical matrix. Despite the data model and network structure, networks having similarity should have similar centralities trends.

- Degree Centrality C_d: Degree centrality is a widely used method to reflect the node interactions in networks [35]. The higher degree a node has, the more influence it may have in a network. We obtain each node in and out degree to depict their effects.

  $$C_D(u_i) = \frac{\sum_{j=1}^{n} A_{ij}}{N(n)}, \tag{3.7}$$

  where A is the adjacency matrix of the network and N is the maximum possible degree in a graph has n nodes, which is used for normalized the degree centrality.

- Closeness Centrality C_c: A person is very active in a social network, which means he/she has numerous interactions with other people. He/She will also have high influence in the network. Closeness centrality provides a nature value to monitor and reflect this situation [48].

$$C_C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(u,v)}, \tag{3.8}$$

where a high closeness value indicates larger centrality.

In our project, we use edge weight for the network. The numbers of the edges between two nodes is set as the weight. Thus, we use Dijkstra's algorithm to calculate the weight. We inverse the edge weight w rather than itself as $\frac{1}{w}$ to make nodes have higher centrality when they have larger weight.

- Betweenness Centrality C_b: A node affording data transmissions in a network plays a vital role. In our social networks, a person, such as a professor, acts as a bridge via different students. This phenomenon leads to use betweenness centrality in order to estimate each node importance and to evaluate the performance of networks [49].

$$C_B(u) = \sum_{s,t \in U} \frac{\sigma(s,t \mid u)}{\sigma(s,t)} \tag{3.9}$$

We still utilize the Dijkstra's algorithm to add edge weight to the centrality.

3.3 Experiment results

Starting with the experiment, we invited 10 colleagues to install WMLSIRecorder on their cellphones. Some of them have high collision due to same working space. The system automatically runs in the background. Data were collected for two months. We set several filter to clean up our datasets to remove redundant and invalid data.

- Set filter for GPS data with time parameter t_1 as 60 seconds.

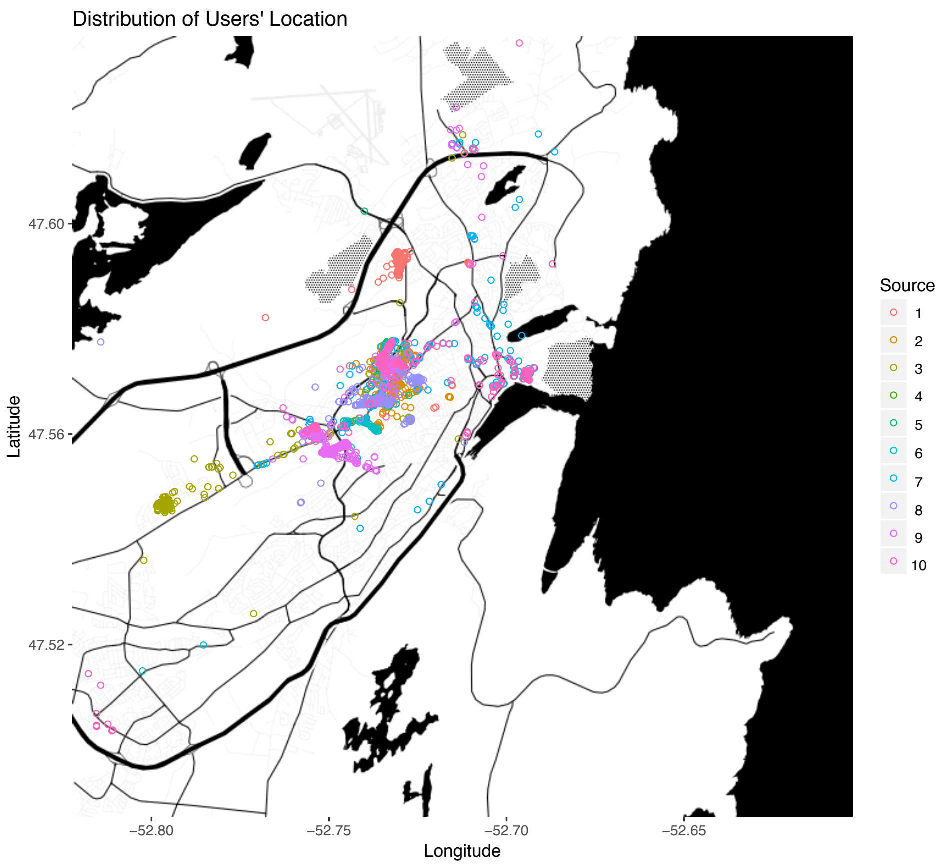


Figure 3.1: Co-location data distribution

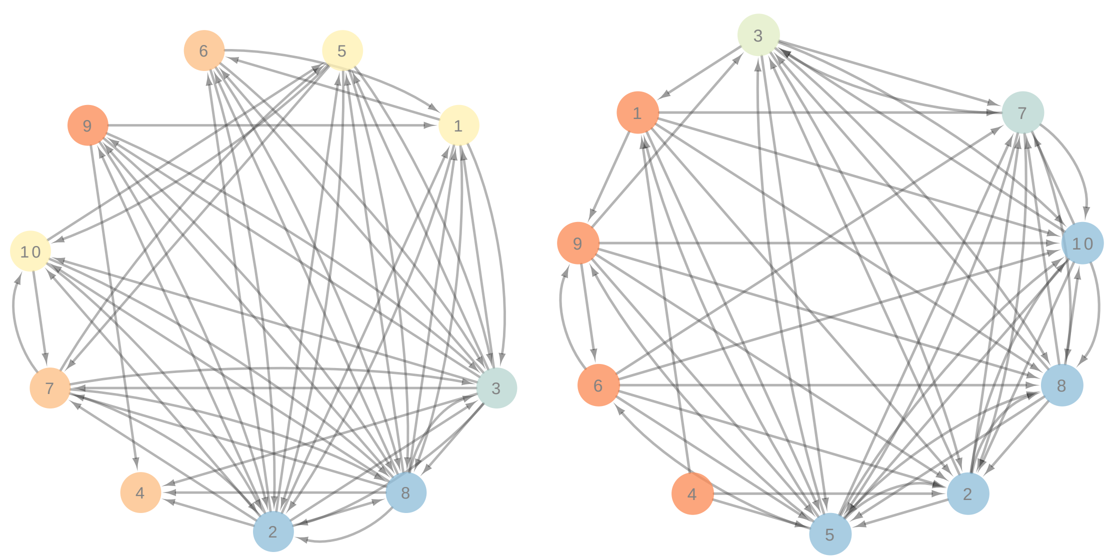
- Set filter for Device data based on filtered GPS data with time parameter t_2 as 300 seconds.

- Combine two different data set. Due to the device data being directed network, we shape and colour the nodes.

After filtering the data, we represent our data in the map shown in Figure 3.1. It is obvious that the centre of the map which is MUN is a highly social obvious happening location. Users have a large number of interactions at this position.

3.3.1 Diagrams of network structures

Graphical methods are often common selection to fit data to theoretical distributions for network characterizations. We generate our data models into graphs to provide a visualization of the networks as shown in Figure 3.2. Users are numbered and denote the nodes in the network. According to the dataset, a directed edge is built from a user considering another user as a friend in Friendship network G_f. Additionally, edges in co-location network G_c represent detections from a user to another. Degree of each node is coloured from blue to orange in ascending order.

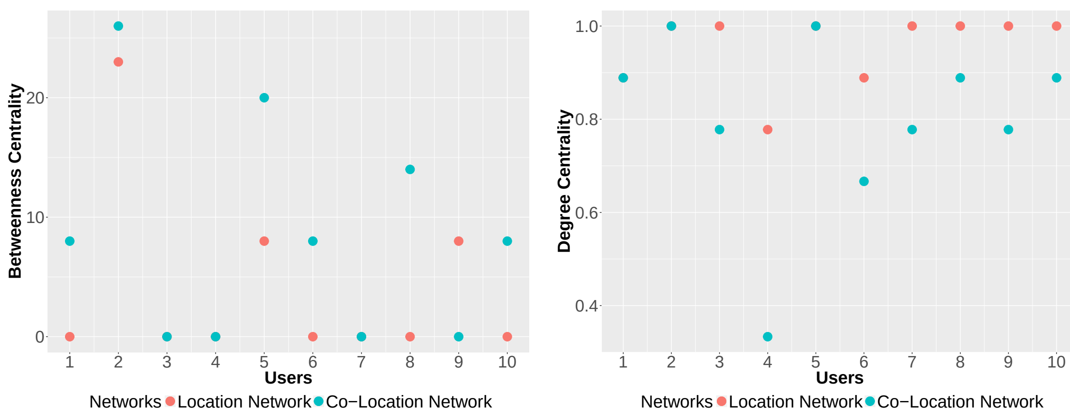


(a) Friendship Network (b) Co-location network

Figure 3.2: Network models

3.3.2 Centrality measurements and comparisons

To verify the performance of combined data network G_c, Location network G_l is fed with Location dataset. Different from G_c, G_l is an undirected network. Thus G_c is converted to undirected network by merging all weights of edges between two nodes in order to compare with G_l. Centralities of G_l and G_c are evaluated in Figure 3.3. Although a few fluctuation occurs, the limit of GPS indoor performance effects as the main trends of both networks is similar. Specifically, coefficient evaluations of these two centralities are evaluated as 0.7 and 0.8 respectively, which reveals a high similarity between two networks.



(a) Betweenness Centrality

(b) Degree Centrality

Figure 3.3: Centralities of G_l and G_c

3.3.3 Friendship & co-location networks analysis

In Table 3.2, the Pearson Coefficient reflecting a low result 0.15 which manifests poor relations on C_b. Besides, other coefficient values of centralities are unfavourable low. Namely, the similarity of G_f and G_c is exceedingly low. Specifically, compared with G_f degree centrality, those of G_c have an entirely different trend as well as the degree value is also varied, shown in Figure 3.4. Rather than friendship affects human mobility, the influence on daily interactions and events is intensely small.

C_b Coefficient	C_c Coefficient	C_d Coefficient
0.15	0.084	0.45

Table 3.2: Coeffiecient of G_f and G_c

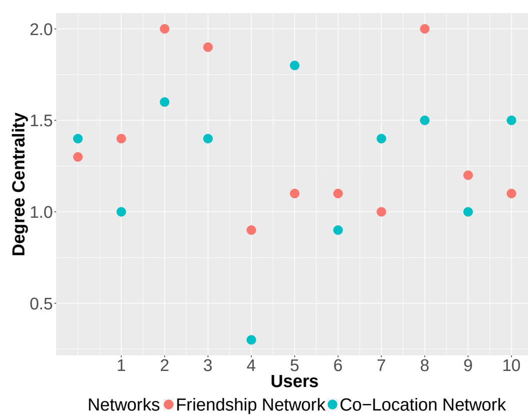


Figure 3.4: Degree centralities comparison

However, we still expect that there should have relations between friendship and social interactions. In our whole experiment, people spend most of the daytime with colleagues on campus, but the activities are quite different after work [24]. To dig into the question and capture more details of the time effects on the network, we have divided the timeline into several different periods.

- Daily time: From 9:00 am to 4:00 pm on weekdays. Our participants are all

MUN members and their normal working and studying periods are from 9:00 am to 4:00 pm. Thus, this daily time definition can indicate the common working period of all candidates in the datasets. At this time period, most candidates are more likely to stay on campus.

- Night time: From 4:00pm to 8:00am on weekdays. It defines the off-work time.
- Weekends: The whole days of Saturday and Sunday.

According to this, night time and weekends periods are considered as off-work time. We use it to compare with our G_f. The result is what just expected. The similarity of these two networks is sharply increased shown in Table 3.3. Thus, we believe that the social activities after work time of people will be influenced by their friendship.

C_b Coefficient	C_c Coefficient	C_d Coefficient
0.65	0.73	0.59

Table 3.3: Coeffiecient of G_f and G_c on off-work time

3.4 Summary

A complex network model with multiple data combination is declared in this paper: a set of Bluetooth connection data and GPS data combined as co-location network. Combining multiple network data is an essential method to construct a social network, while single data is limited by the environment. Our model reliably composes muti-data and rebuilds social networks. We also developed a system to gather data, WMLSIRecorder. The tool collects multiple data but works in a low power cost mode. On the other hand, different environments and social events cause the structure of social networks varied. We find that friendship has influence on human social activities, which is more evident during off-work hours.

Chapter 4

Adding Bluetooth Connectivity

The limitations of single data are discussed in the previous chapter. Moreover, we find that it is also hard to collect data. Therefore, a low-cost Bluetooth based data collection approach is discussed in this chapter. The performance of this method is proved. The advantages of our design are also demonstrated.

4.1 Research and limitations of current data acquisition

Data collection is a classic problem of network constructions. Building an accurate dataset is a pivotal issue in network analysis. Craig and his groups design a questionnaire to collect information from volunteers [14]. It is an effective way to obtain data. However, the cost of sending people to do the survey is high. Another common solution asks researchers to code and crawl data from online social media [14][15]. A huge amount of information can be generalized by this method which gives us more details and features to observe the social network. But in most cases, it is not supported by the official online social media, even though Twitter has shut down its API to disallow users to access these data. Any changes of these companies will force this approach to be failed and have updates to fix the problems. It highly relies on third parties.

In addition, mobile phones have become an important part of human's life. Grabbing data from it is also a delicate way [23]. GPS location data and Bluetooth data via building connections between two users are well developed in many studies. The power consumption and connection building time limit its performance. To avoid this, many proximity loggers were studied. WiFi is widely used in social network research [36]. However, it requires heavy design and improvements for the buildings already existed, such as airport and shopping mall [50]. Some mobile systems restrict the access to WiFi scan. For example, Apple has prohibited publics to acquire WiFi scan API until 2015. Even now, only parts of developers with its permissions can touch this. We want to propose a cross-platform scheme to allow every smartphone benefitting it. Thus, iBeacon becomes the only choice. Some researchers proposed customized loggers [42], but have to use specific devices for collecting. We decided to use iBeacon sensors as additional data broader which can be agilely deployed in the environment. The cost of it is only cheap. My group members and I have also proposed a fast radio sensor deployment algorithm which minimizes the number of iBeacons sensors in a space with given size but maximizes the accuracy. It is affordable and we extended out *WMLSIRecorder* for collecting and tracking users' interaction data by smartphones.

4.2 WMLSIRecorder extension

We extend our *WMLSIRecorder* to collect Proximity data from iBeacon. It contains all data from Table 2.1 and 3.1. We also include a time frame for future use. To diversify the situations of human activities, we propose two types of beacons deployment, *Fixed Beacon* B_f, a beacon at a fixed place, and *Moveable Beacon* B_m, a beacon in a moveable place. We will talk about it in next section.

Furthermore, we also design a bar chart view for data observing in the foreground as shown in Figure 4.1. Each detected device or sensor will be represented in a single cell. The live RSSI data updates and the nearby devices are represented in the cell. This can let researchers know the exact data update frequency.

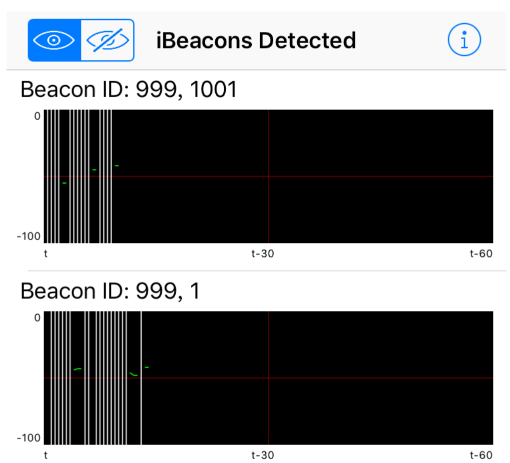


Figure 4.1: *WMLSIRecorder* iOS application

4.3 Data collection

As we mentioned above, we create an iOS application, *WMLSIRecorder*, for data collection. It allows us to be more convenient and fast collect data than manually recording. The format of the data is represented in Table 2.1

4.3.1 Beacon deployment

In order to diversify the situations of human activities, we divide the beacon deployment into two main parts, home and office. This will let us obtain data to observe different activities of a human when they at home and university. In this way, we assume a series of patterns of the beacon deployment to simulate different kinds of situation that human might meet in their daily life. Moreover, offline iBeacon which was previously detected can also be represented in the view. Additionally, there are two designed types of beacons deployment shown below:

- B_f: Beacon at a fixed place, which means the devices always stay in the same place, such as the bedroom of the home and the office at the university.
- B_m: Beacon in move place, which means the beacon has high mobility, for instance, user's pocket and/or bag.

B_f can figure out that the volunteer is in this fixed location or not, which we can observe that the interactions of testers in two different parts, home and office. For the $B_m$pattern, the devices sometimes will travel with the tester. It will also provide us more accurate data the place volunteers visited and other testers they met due to the long connection time between two phones, which we discussed in the previous chapter.

Under this regulation, we can observe and record participants' access to any fixed locations via B_f. On the other hand, B_m is an evidence to prove the collision and meeting up for different individuals. Thus, we can get logs when these participants encounter.

4.3.2 Data format and process

Symbols and definitions of data collection process are given in Table 4.1. To collect data, we design and format our own dataset for collection. UUID, major and minor value are collected. Moreover, signal strength data is recorded as RSSI. Due to the band channel, the signal of Bluetooth devices is easily interfered by other devices in the environment, such as Wi-Fi, wireless earphones and mouse which use the same band with Bluetooth [51]. In this situation, when a user gets close to the iBeacon sensor, sometimes, the application may not detect the iBeacon or RSSI is recorded as 0. These data are defined as noise. They will affect our dataset. To address this problem, we restrict the detecting time should be longer than 500 millisecond(ms), which means that only if the application detects the signal from the iBeacon longer than 500ms, the data of this collision will be recorded into the dataset.

Furthermore, the connection distance will also cause similar errors. While two devices lived a little bit far from each other, the RSSI might be always 0 or smaller than -100. Thus, we only consider the RSSI value between -100 and 0 as valid data in our project.

Attributes	Symbol	Details
User Devices	I	iPhone
Users	U	Colleagues in the lab.
Hardware	B	iBeacon sensors.
Colloision Proximity	D_p	Collision between device and beacons, which will record the meet up event between fixed and moved nodes.
Co-Location	D_c	Collision between devices, which could increase nodes meet up chance in data and the GPS based location information.
Time Frame	T	Eight weeks period.
Simulated Network	N	Small social network of a lab.

Table 4.1: Data collection define

4.3.3 Experiment design

In order to build our dataset, we applied for WMLSIRecorder extension in our previous data collection. Users were required to open their Bluetooth and GPS functions on their cellphones to ensure the WMLSIRecorder work during this period. Scheme of the experiment is showing below:

- *Data collection via iBeacon:* The *WMLSIRecorder* is installed on each user's smartphone. Each participant gets one iBeacon fitted in his/her home as B_f. We define the action of a user detecting B_f when he/she comes back home as *check-in*. *Check-out* data denotes the status of leaving the house. The *check-in & out* status of users in & out home will be observed. Additionally, office places of participants will also be fited with iBeacon sensors. As our volunteers come from five different labs, thus, we fit only one iBeacon sensor in each office. This approach can let us build several small social groups in our experiment with diverse environments and observe interactions in such small groups and connections via all of them.

- *Data types:* The *WMLSIRecorder* records three types of data as we mentioned, Bluetooth directed connection data, proximity data and GPS location data. The radio signal broadcasted by iBeacon sensors will be recorded when a user reaches the range of the sensors. On the other hand, in the mod of location data, while users get close to each user, their radio data communication will be recorded. Here, we defined the *close* as the distance between each volunteer should be less than 10 meters and the communication time should be larger than 1 second, which can filter errors of the system. The GPS information is collected as well.

- *Time and estimation:* This test is held for eight weeks. According to our design, the individual at least has two status, *check-in & out* for each his/her home location every day. Thus, every volunteer will have at least 2 B_f status data in daily. As we plan to set each data as a node in our proximity network model in next section, the total numbers of nodes will be at least 2400. Users have more chance to *check-in* & *out* in their office. For example, meeting or courses in other places and leaving for food and toilets. This will double, even triple the number of nodes in the network. Hence, we can have enough data for our experiment to analyze our models.

4.4 Model of proximity data

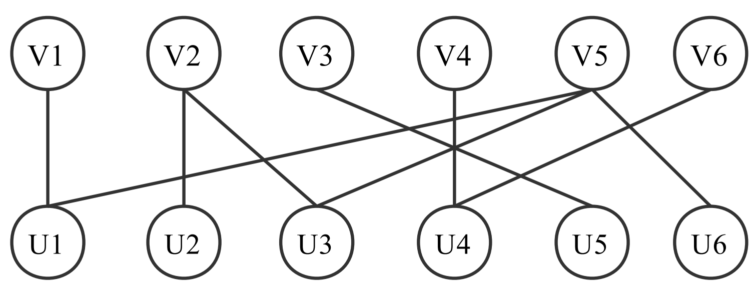


Figure 4.2: Bipartite graph design for proximity network G_p

Unlike the previous chapter, proximity data adds a new kind of node into the network, beacon node V. Hence, a new network model using bipartite graph is discussed in this section. A bipartite graph can clearly represent the relations between two different types of nodes [35]. A directed bipartite graph G_p, shown in Figure 4.2, is proposed to analyze data and build our network model in this section.

$$G_p = (U, V, E_p, W_p). \tag{4.1}$$

We consider that each user, or called collecting device, is denoted by a node u. Node v represents each iBeacon. The edge E_p between u and v represents the detection for users to iBeacon sensors. The direction is always from u to v due to only the user detecting iBeacon. Here, E_p is the edge set weighted by the number of times that a user is in the proximity of a beacon.

4.4.1 Conversion of network

4.4.1.1 One mode projection

In Chapter 3, we have presented our G_f and G_c networks as directed graphs. However, their only have one type mode. Comparing them with our new G_p model will be a challenge. To address the problem, G_p is converted to a one mode graph G_u which only contains nodes U in the network as same as G_f and G_c by the following steps:

- All nodes B are removed from the G_p network.
- If a node B is detected by multiple nodes U, an edge E_u will be built between every two node U. Moreover, the edge weight of E_p will be summed to the new edge E_u.

Although, the information of the network will be lost during the conversion and relations between node B and node U are removed, summarising the weight of the

E_p to the E_u can reduce the loss. Furthermore, we added a time filter to improve the precision. The weight of the edge E_u will only be counted when the two nodes U detected the same node B in a specific time shift. It can avoid overvaluing the weight of the whole network G_u.

4.4.1.2 Directed to undirected network

After one mode projection, the original directed bipartite graph is converted to an undirected network G_u. However, G_c is a directed network which represents the detection from one user to another user. Comparing this two network, we firstly should convert G_c to an undirected network. To simplify the problem, the directions of the edge in G_c are removed. Weights of two directions between two nodes are added together.

4.5 Result and Evaluation

Dataset is built and collected by our WMLSIRecorder for a period of two months. Although, the RSSI data will have lots of error and gaps [27], and in order to generating an accurate and stabilized network model in performing and reconstructing the real behaviors of people, which the model can be a very effient way to reproduce people real life in society, we user our G_c network to verify the new network model G_p as we have proved the performance of G_c in the previous chapter.

- **Mathematics approaches**: Centralities are chosen to evaluate and compare two networks. We keep using three types centralities, betweenness centrality, closeness centrality and degree centrality. Additionally, normalized Pearson coefficient is adapted too.

- **Experiment setup**:
  - The broadcasting gap of the iBeacon sensor is set to 500ms. It can provide a high-frequency signal refresh rate to prevent users missing any iBeacon sensors and extend the battery life to 1 year long.
  - A large number of errors will occur due to the previous setting. We added a Time parameter t_3 as 300s. In every 300s, only one detection of the same iBeacon sensor will be recorded for each user.
  - To reduce the overvaluation of the network, we set another Time parameter t_4 as the 60s. It only allows for an E_u existed when two nodes detected the same iBeacon sensor within t_4.

4.5.1 Network representation

Figure 4.3 demonstrates the proximity data of the network G_c. Each participant is marked by numbers. iBeacons sensors are merged into our G_c via one mode projection.

4.5.2 Network comparison

A natural way to compare two networks is to quantify the networks. We evaluate three types of centralities for both G_c and G_p. Obviously, a user in both two networks has similar centrality value in all measurements shown in Figure 4.4. To be more

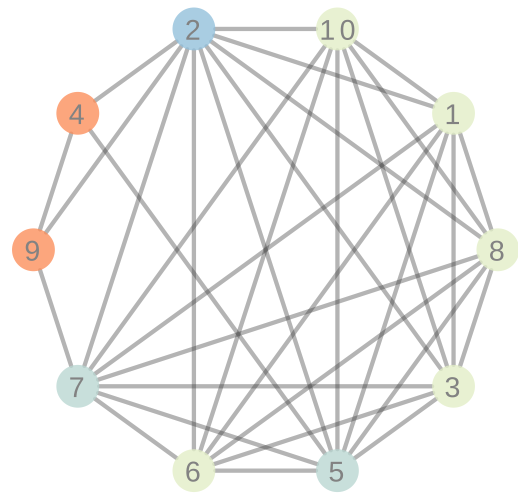


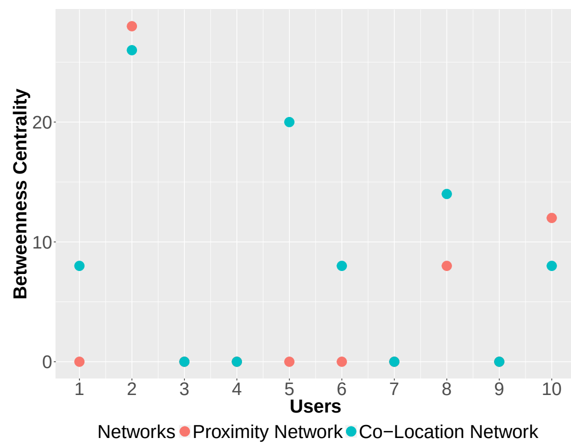
Figure 4.3: One-mode proximity network G_u. Colour from orange to blue represents the degree of each node from low to high.

important, some nodes even has exactly same centralities, for instance, node 3 and 4 obtain same centralities in both Figure 4.4(a) and 4.4(b).

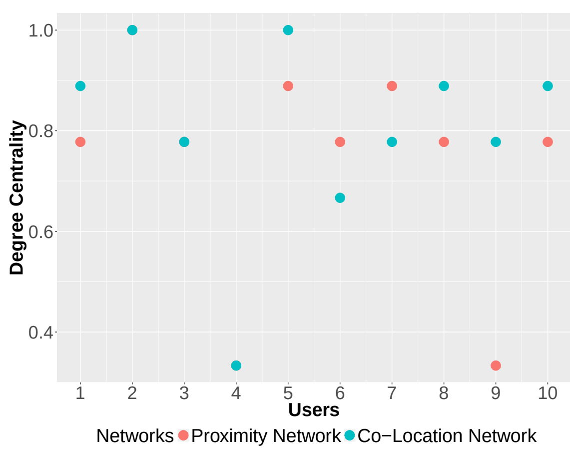
C_b Coefficient	C_c Coefficient	C_d Coefficient
0.70	0.71	0.72

Table 4.2: Coeffiecient of G_c and G_p

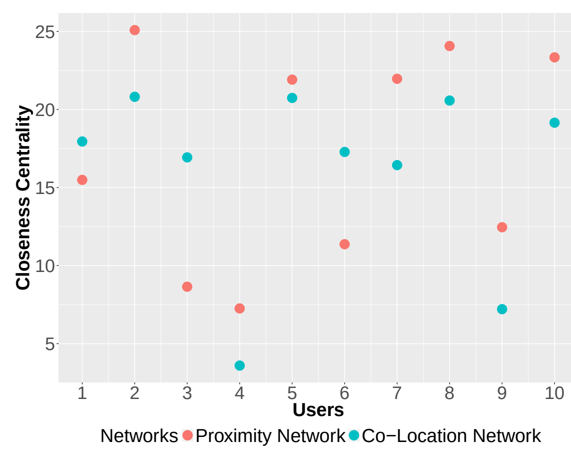
Furthermore, the similarity between G_c and G_p is also studied to measure the performance of the G_p. As expected, the coefficients of all centralities are close to 1. The result shows a proximity-based network model can demonstrate the relations and connections between people accurately and precisely.



(a) Betweenness Centrality



(b) Degree Centrality



(c) Closeness Centrality

Figure 4.4: Centralities of G_c and G_p. 4.4(a) illustrates the betweenness centrality of G_p and G_c. In 4.4(b), degree centrality of two networks is shown. 4.4(c) represents the closeness centrality.

4.5.3 Remarks

In this chapter, we propose a cross-platform hardware data collection solution by using iBeacon sensors and mobile devices. The price of the sensors is cheap. Moreover, smartphones own a great part of the cell phones market. The performance of such method is also proved. Thus, our approach can rapidly reduce the cost of the research and widely used in different scenarios.

Chapter 5

Analysis Between Co-location Interactions And Proximity Using Temporal Network

We take a superficial look at the time effects on the co-location network G_c topologies and discussed its similarity compared with friendship network G_f in the previous chapter. It brings a distinct result compared with non-time influence network. However, it still uses the static networks methods and roughly intercepts parts of the networks. Thus, we decide to use temporal network models to build and analyze our datasets.

5.1 Gaps between static and dynamic network on social network research

Static network is a mature way to study many network problems, such as protein network, transportation network and food chain network. However, time is an essential element of some networks, especially large networks. For example, interactions between individuals change rapidly in the social network. Moreover, the relationship is also varied based on different time periods. There will be always new one built and old one broke up. Thus, time elements play a crucial part of the networks. The topologies of these networks are changed in different time period [24]. In this situation, the static network model cannot represent such kinds of information. Thus, we study a temporal network method to recover the time influence in this paper.

On the other hand, the weight of temporal network is not well studied [25], where it plays an important value in evaluating the network. For instance, Nathan et al. pointed out that the edge weight which demonstrates the relations between two nodes will affect the network structure based on time [24]. Thus, studying the influence of weight on the network topology change is a problem. We discuss a weighted temporal network model and measurement tools in this paper to address the problem.

5.2 Temporal network & dynamic algorithm

In this section, we model our data by using temporal graph tools on two networks, co-location network G_c and proximity network G_p from the definitions in Chapter 3. Furthermore, we also add time elements to the network to study the time effect on the network.

5.2.1 Temporal network model

To build our temporal network model, we use the method from Hyoungshick and Ross [25]. In their study, they add time elements to the network model and construct temporal graphs. The temporal graph is a directed graph that describes both structures and changes of the structures. The temporal network model allows us to study the evolution of users interactions (e.g. co-location of each other and proximity to landmarks) over time.

5.2.1.1 Temporal co-location network

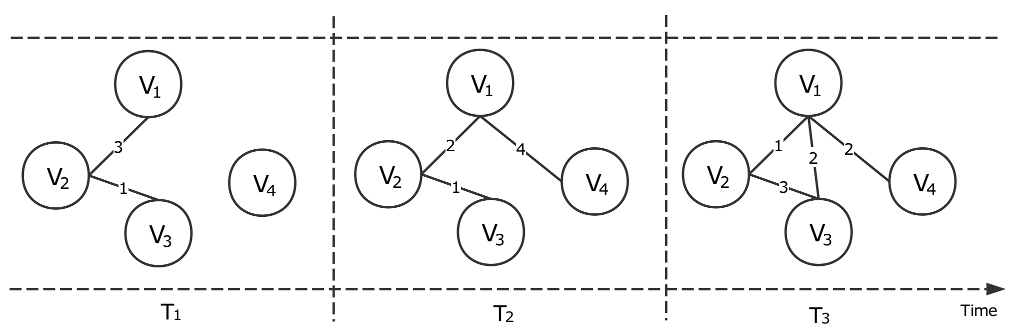


Figure 5.1: Snapshots of the network

Creating a temporal network out of a static network with the edge event log has two steps. This is a similar procedure to Hyoungshick [25]. First, we divide the entire time period of the event log $[0, T]$ into $\lceil T/\Delta T \rceil$ intervals of ΔT each. With each interval T_i ($i = 1, 2, \ldots, \lceil T/\Delta T \rceil$), we can create a static weighted graph as in 4.1

co-location network, denoted

$$G_i = (V, E_i, W_i). \tag{5.1}$$

G_i is called the i-th snapshot of G (i.e. Figure 5.1). See Algorithm 1 for details.

Next, we build the temporal graph using the snapshots G_i ($i = 1, 2, \ldots, \lceil T/\Delta T \rceil$). The temporal graph is a weighted directed graph on $|V| \times (\lceil T/\Delta T \rceil + 1)$ vertices. These vertices are organized in $\lceil T/\Delta T \rceil + 1$ stages of $|V|$ each, and each snapshot G_i defines the connectivity from stage i and stage $i+1$. Specifically, we denote

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W}) \tag{5.2}$$

and each vertex $v_{i,j}$ is the i-th vertex of the snapshots in stage j. From stage j to stage $j+1$, there are two types of edges:

1. horizontal edges $(v_{i,j}, v_{i,j+1})$ for $i = 1, 2, \ldots, |V|$, and

2. cross edges $(v_{i,j}, v_{i',j+1})$ and $(v_{i',j}, v_{i,j+1})$ if and only if $\exists (v_i, v_{i'}) \in E_i$.

Note that all edges in $\mathcal{G}$ are directed and each edge in the snapshots introduces two cross directed edges in $\mathcal{G}$. $\mathcal{W}$ is the temporal edge weight defined as the number of connections from node v_i at stage j to node $v_{i'}$ at stage $j+1$. Besides, the horizontal edges will have an edge weight defined as $\mathcal{M}$ where $\mathcal{M}_j$ is the mean of all nodes $\mathcal{W}$ from stage j to stage $j+1$. A temporal co-location network model is shown in Figure 5.2 We also provide the pseudocode of creating temporal networks models in Algorithm 2.

5.2.1.2 Temporal proximity network

We can also create a temporal version of the proximity network (Equation 4.1) using a similar workflow.

Algorithm 1 Create snapshots

Input:

$G = (V, E, W)$: Static weighted graph

L : Time sequence of vertex interactions in $[0, T]$

ΔT : Time interval length

Output: $G_1, G_2, \ldots, G_{\lceil T/\Delta T \rceil}$, where $G_j = (V_j, E_j, W_j)$

INIT

for $j = 1, 2, \ldots, \lceil L/\Delta t \rceil$ **do**

$V_j = V$

$E_j, W_j = \phi$

$G_j = (V_j, E_j, W_j)$

end for

CREATE SNAPSHOTS

for each log entry r in L **do**

$t_r =$ time stamp of r.

$u =$ detecting device.

$v =$ detected device.

$j = \lceil r/\Delta t \rceil$

if $(u, v) \notin E_j$ **then**

add (u, v) to E_j

$W_j(u, v) = 1$

else

$W_j(u, v) = W_j(u, v) + 1$

end if

end for

Algorithm 2 Snapshots to Temporal Graph

Input: $G_1, G_2, \ldots, G_{T/\Delta T}$, where $G_i = (V_i, E_i, W_i)$

Output: $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$

INIT

$\mathcal{V} = \{v_{i,j} | i = 1, 2, \ldots, |V| \text{ and } j = 0, 1, \ldots, \lceil T/\Delta T \rceil$

$A, C, \mathcal{M} = \phi$

ADD CROSS EDGES

for $j = 1, 2, \ldots, \lceil T/\Delta T \rceil$ **do**

  for each edge $(v_i, v_i') \in E_j$ **do**

    add edge $(v_{i,j-1}, v_{i',j})$ to $\mathcal{E}$

    add edge $(v_{i',j-1}, v_{i,j})$ to $\mathcal{E}$

    $\mathcal{W}_{(v_{i,j-1}, v_{i',j})} = W_j(v_i, v_i')$

    $\mathcal{W}_{(v_{i',j-1}, v_{i,j})} = W_j(v_i, v_i')$

    $A_{j-1} = \mathcal{W}_{(v_{i,j-1}, v_{i',j})} + \mathcal{W}_{(v_{i',j-1}, v_{i,j})}$

    $C_{j-1} = C_{j-1} + 2$

  end for

end for

ADD HORIZONTAL EDGES

for $j = 1, 2, \ldots, \lceil T/\Delta T \rceil$ **do**

  $C_{j-1} = C_{j-1} + |V|$

  $\mathcal{M}_{j-1} = A_{j-1}/C_{j-1}$

  for $i = 1, 2, \ldots, |V|$ **do**

    add edge $(v_{i,j-1}, v_{i,j})$ to $\mathcal{E}$

    $\mathcal{W}_{(v_{i,j-1}, v_{i,j})} = \mathcal{M}_{j-1}$

  end for

end for

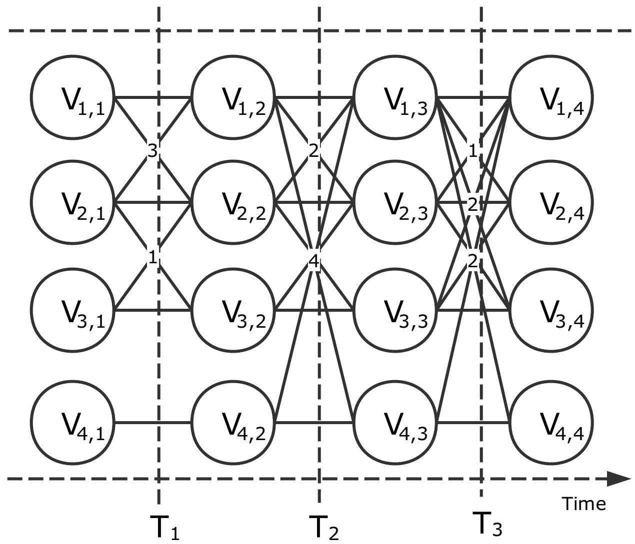


Figure 5.2: Temporal co-location network model.

1. Give a bipartite graph $G_p = (U, V, E_p, W_p)$ for the proximity network, the event log file, time period $[0, T]$, and a ΔT, we can generate a series of snapshots $G_j = (V_j, V_j', E_j, W_j)$, for $i = 1, 2, \ldots, \lceil T/\Delta T \rceil$ using Algorithm 1.

2. We next project each G_j to the vertex set V_j (e.g. the users). The weight will sum all edge weights of two nodes v who have interactions with the same node v' in its snapshot.

3. We use the same method defined in temporal co-location network, Algorithm 2, to build our temporal proximity network.

We make two points here. First, using Algorithm 1 for bipartite graphs is nature. The only difference is that the bipartite one has two vertex sets, respective to co-location snapshot. Second, the information of the network will be lost during the conversion and relations between node V and node V' are transformed, our edge weights can reduce the loss.

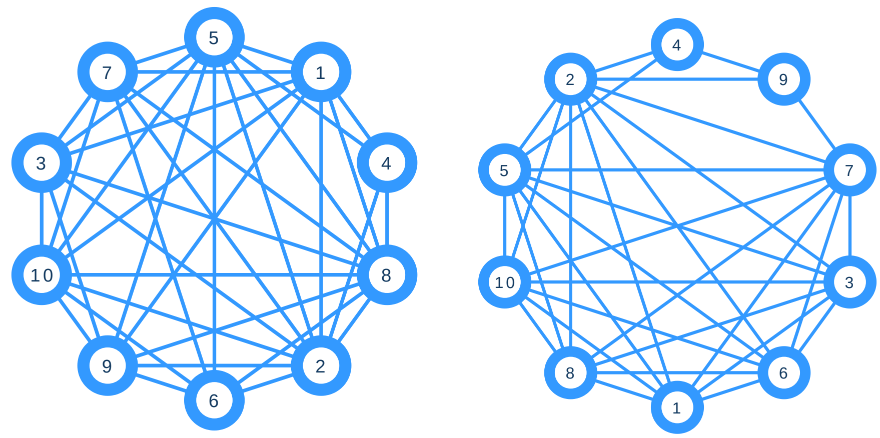
5.2.1.3 Centrality measurements

According to our temporal network models, Equation 5.2, we expand the static graphs into time-based graphs. The temporal graph $\mathcal{G}$ is a series of stages connected by snapshots (i.e. Figure 5.2), which can be processed algorithmically as static graphs but shows more details than the static ones. Thus, it is natural to use standard centrality tools to evaluate these graphs.

On the other hand, we are curious about the influence of a person in a social network, i.e. a person, who is a professor, acts as a bridge via different students. Betweenness centrality can represent this phenomenon. Additionally, a person is very active in a social network, which means he/she has numerous interactions with other people [35]. He/She will also have high influence in the network. Closeness centrality provides a nature value to monitor and reflect this situation.

- Closeness centrality C:
$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(u,v)}, \tag{5.3}$$
where a high closeness value indicates larger centrality. $d(u,v)$ is the distance from u to v. We inverse the edge weight w rather than itself as $\frac{1}{w}$ to make nodes have higher centrality when they have larger weight.

- Betweenness centralities B:
$$B(u) = \sum_{s,t\in U} \frac{\sigma(s,t \mid u)}{\sigma(s,t)}, \tag{5.4}$$



(a) Co-location Network (b) Proximity network

Figure 5.3: Static co-location and proximity network

where $\sigma(s,t)$ is the number of all shortest paths from node s to t and $\sigma(s,t \mid u)$ is the path from s to t that goes through u. The larger edge weight will give more influence on betweenness centrality [49].

5.3 Evaluation and analysis

In our experiment, we used the datasets from our previous chapters, proximity data and co-location data. We analyze the change of social interactions on the timeline in this section. We also discuss the difference between snapshots and static networks by mean and variance of edge weights. Furthermore, centralities of temporal co-location and proximity networks are evaluated, along with their correlations.

5.3.1 Influence of different time periods

In Figure 5.4, we present a one-month period to observe the frequency of our participants' interactions in co-location network. Interactions among all volunteers within one hour is considered as the frequency of the hour in our network. The gray scale is changed from white to black according to the frequency updating from low to high.

We can observe that the frequency of the day 7, 14 and 21 are higher than other days. The reason is that these three days are Monday, where people have more interactions. However, the next Monday on day 28 is different, where the Winter semester is over and many participants are off campus.

An important feature which indicates that most interactions have happened between 9:00 am and 5:00 pm is a key point to demonstrate the time effects on social activities. Participants can also have interactions after their works, even at midnight, such as 2:00 am to 5:00 am on day 10 and 1:00 am on day 12, although it is not as high as daytime hours. (It only has few influence on the whole network. But this captured phenomenon cannot be observed on static graphs.) Thus, time can give us more information about social dynamics that static networks cannot.

5.3.2 Time variance of co-location and proximity networks

To understand the time influence on social networks, we study the difference between snapshots and the static networks. In addition, we also discuss temporal centralities to demonstrate how the network topology changes with time.

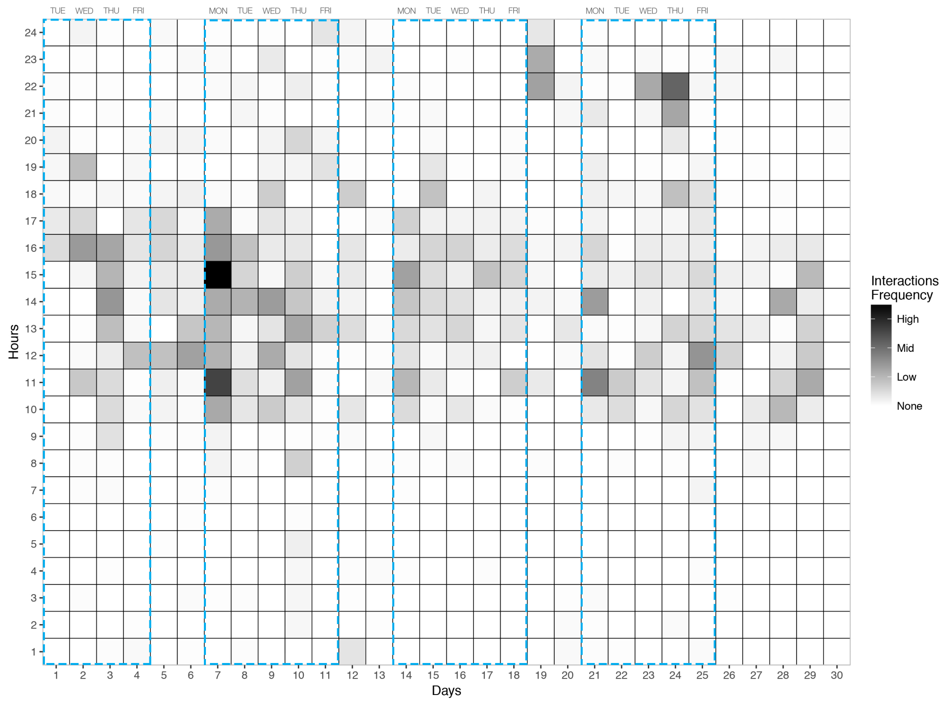
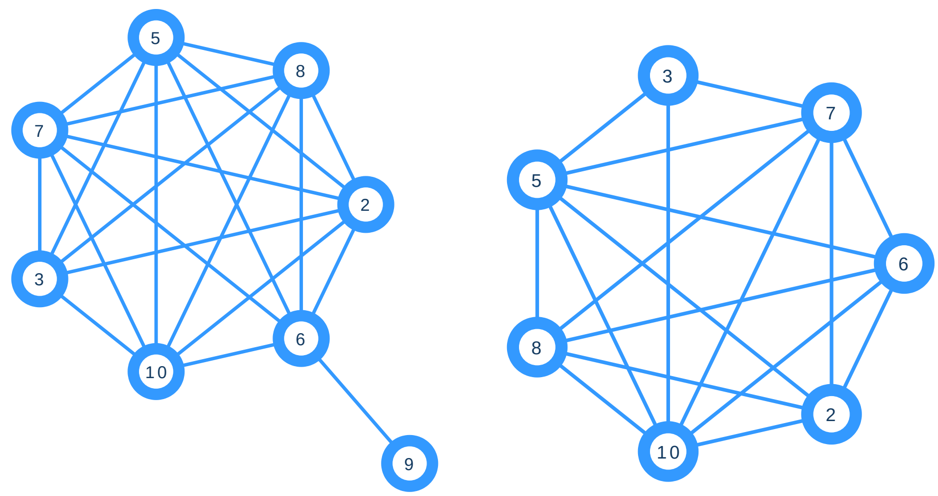


Figure 5.4: Frequency of total interactions, from 2017/04/04 to 2017/05/03

5.3.2.1 Snapshot of the network topology

We extract one-day social interactions in our dataset. From Figure 5.4, the day 7 has a high frequency of social activities, which we believe that the network topology of it will be most similar to the whole network structure.

In Figure 5.3, two static network diagrams represent the whole time period topology of co-location network and proximity network, respectively. Besides, we find that topologies of the two networks on a single day varies a lot. In the static co-location network, the node 9 has interactions with all nodes, except node 4. But node 9 only interacts with node 6 in the corresponding snapshot graph shown in Figure 5.5(a). Moreover, node 1 and 4 even do not exist in this snapshot. Similar to the co-location network, three nodes do not show up in the proximity snapshot graph. Edges are also reduced, for instance, there is no edge between node 2 and 3 in the snapshot graph, while the corresponding edge exists in the static network. Thus, we can clearly observe that the topology varies via time.

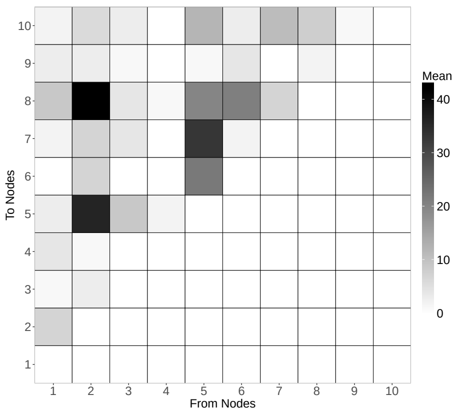


(a) Co-location Network (b) Proximity network

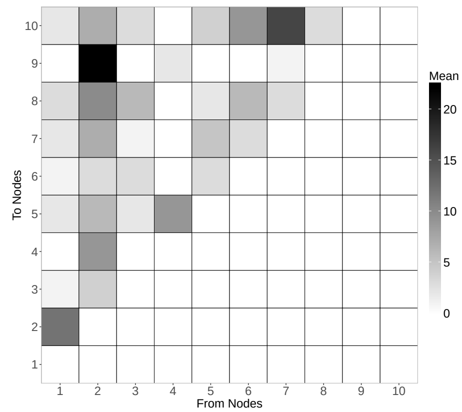
Figure 5.5: One day snapshot of the network topology

5.3.2.2 Numbers of interactions on timeline

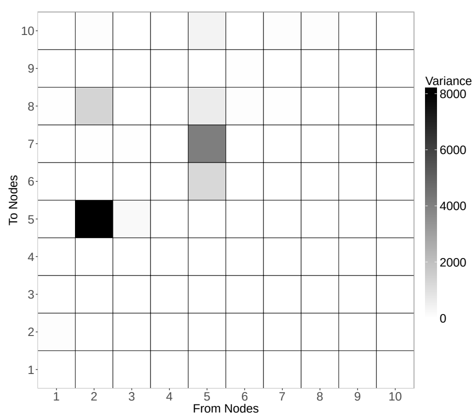
Before evaluating the networks, we analyze the changes of interactions among users, edge weight, on the timeline, as shown in Figure 5.6. Based on two one-mode snapshots, co-location and proximity, we can obtain a series s of edge weight between two users over the one-month period. Thus, the mean and variance of s is calculated. For co-location snapshots, the mean value of edge weight, i.e. numbers of encounters, fluctuates between 0 to 40's (i.e. Figure 5.6(a). Only triangle is showing because of undirected graphs.). However, the variance of it goes into a state of high volatility (i.e. Figure 5.6(c)). For projected proximity snapshots, the fluctuation of the mean value is in the range of 0 and 20's. In Figure 5.6, the highest variance of it is 75 times more than the mean one.



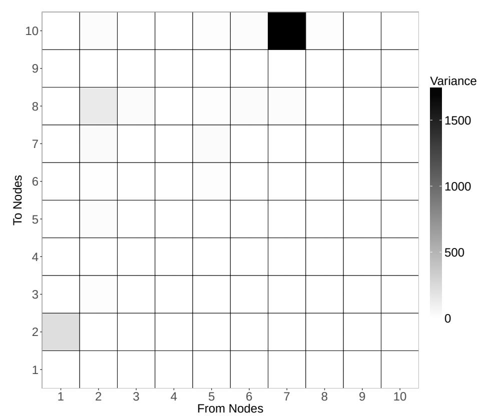
(a) Mean edge weight of co-location snapshots



(b) Mean edge weight of proximity projected snapshots



(c) Variance edge weight of co-location snapshots



(d) Variance edge weight of proximity projected snapshots

Figure 5.6: Edge weight changes of co-location snapshots and proximity projected snapshots

5.3.2.3 Temporal centrality evaluation

In this section, we use centrality measurements to reveal the evolution of networks. We use two variants of the temporal networks. One is unweighted graphs shown in Figures 5.7(a) and 5.7(b). The other is weighted shown in Figures 5.7(c) and 5.7(d), which can reflects the influence of encounter frequencies. Betweenness centralities of each node on different time intervals are presented. The higher betweenness centrality a user has the darker shade in the figure. We can observe that the centralities of each node change over time. For instance, in Figure 5.7(a), user 2 obtains the highest centrality on day 6 but has no value from day 21 to 23. In addition, each column demonstrates centralities of all nodes on a single day interval. It is obvious that the topology of the network is varied over time.

Furthermore, we compare two temporal networks, co-location and proximity, to illustrate the weight effect on network models, as shown in Tables 5.1 and 5.2. We use Pearson coefficient to process our evaluation. It suggests using expected number of each value to compare two arrays and calculates the normalized covariance, similar to the cosine. The range of the coefficient is $[-1, 1]$, where the closer to 1 the result is, the higher similarity they have. B_c and B_p represent the betweenness centrality of unweighted temporal co-location network and proximity network, respectively. Betweenness centralities of the two weighted temporal networks are denoted by B'_c and B'_p. Closeness centralities of two networks are following the similar definitions as B. Compared the network itself, the weight has less effect on closeness centrality, i.e. the coefficient of C_c and C'_c is 0.92 and the value of C_p and C'_p is 0.9. The reason is that we use the inverse of the weight w when calculating the closeness centrality (Equation

5.3) and it has a small dynamic range, which have little impact on the mean distance from a node to another. However, the result for between closeness is distinct. The weight attribute is considered to bring more changes to a network, for instance, 0.16 and 0.32. That is, a greater edge weight can modify the betweenness centrality in a greater deal using our definition (Equation 5.4). From this phenomenon, we can obtain that nodes with high betweenness centrality varies greatly and the influence of a person are also changed over time in social networks. On the other side, the similarity of two different networks is also changed dramatically, for instance, comparing B_c with B_p and B'_p, the coefficient is changed from 0.64 to 0.06, and comparing C_c with C_p and C'_p, the difference of them is 0.3. This change illustrates that the edge weight also takes a massive impact on the performance of networks.

Coefficient	B_c	B'_c	B_p	B'_p
B_c		0.32	0.64	0.06
B'_c			0.10	0.72
B_p				0.16
B'_p				

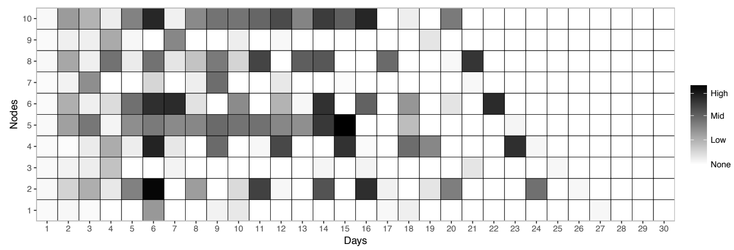
Table 5.1: Coeffiecient of betweenness centrality on temporal networks.

Coefficient	C_c	C'_c	C_p	C'_p
C_c		0.92	0.62	0.3
C'_c			0.64	0.44
C_p				0.90
C'_p				

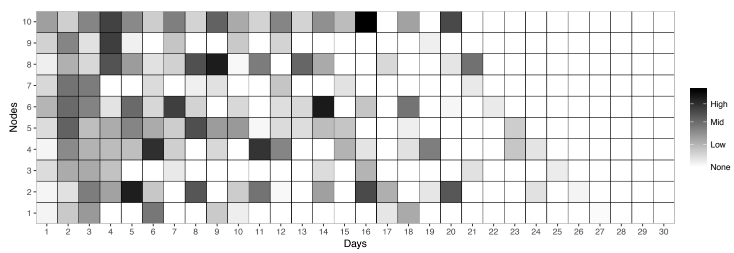
Table 5.2: Coeffiecient of closeness centrality on temporal networks

5.4 Consequence

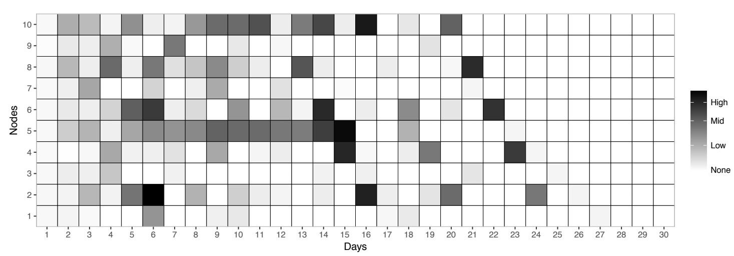
In this chapter, we discuss the effects of time on social interactions, i.e. co-location network and proximity network. We first propose two static models to reproduce two social networks. Second, we study two types of snapshots, one-mode and bipartite, to extend static models by adding the time factor. Then, we propose a weighted version of the temporal network model and present the procedures of converting static network models to temporal models. Moreover, the measurement tools of temporal graphs are also studied. We find that the variation of social activities among people on different time periods. Additionally, we confirm the time variance of social networks and the evolution of network topologies, and we show the effectiveness of edge weight on temporal network models.



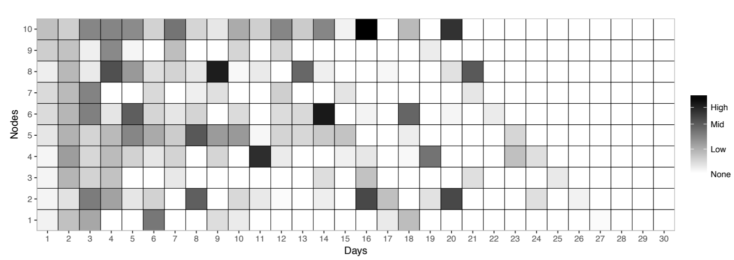
(a) Unweighted temporal co-location Networkk



(b) Unweighted proximity network



(c) Weighted temporal co-location network



(d) Weighted temporal proximity Network

Figure 5.7: Time variance of temporal co-location and proximity network betweenness centrality

Chapter 6

Contributions & Future Work

Through our research, we hope that we would have contributions to trajectories of social network research and temporal network models. The network models we built and analyzed can help other researchers find out social connections of people more clearly. We also propose a multi-data combined network approach. It can solve the limitation of single data network and improve the robustness. At the same time, a Bluetooth-based cross-platform data collection method is dicussed in this thesis. By using the iBeacon sensors, our system provides a low-cost and high performance data collection approach. Moreover, it can not only use for social network research, but other network problems data collection. We also studied the similarity between friendship network and social interactions network at the side of static network. Although friendship network topology is different, relative to co-Location network, they still have high similarity on off-work periods. On the other word, social interactions of people will be effected by their friendship after their working time. Furthermore, the time variance on the social networks is also analyzed. We generalize a weighted version of the temporal network model to unveil the impact of time elements on co-location and proximity network.

However, there are still some insufficient points in our study. The number nodes in our dataset is not large enough. Thus, we plan to make a large-scale simulation in future. Furthermore, we find some other interesting and potential ideas that can extend our current research.

6.1 Large-scale simulation

Our models and methods can easily migrate and fit for the large-scale scenes, such as country-wide human society and the site of wild animals. The large-scale human activities simulation would let scientists observe the daily social events and interactions with others of people. The Network Simulator 2, NS2, could be an effective tool for us to simulate a large-scale social network based on the data we collected in the research. This tool is originally built for simulation of wired or wireless networks [52]. It is also a proper instrument for our simulation. Furthermore, all nodes during in the small-scale experiment will be randomly times. The large-scale human activities simulation will be hold based on the data and model we obtain. This simulation would let scientists observe the daily social events and interactions with others of people.

6.2 Prediction

Predicting human activities or social structure change is also a popular topic these days. It can help scientists to prevent disease infection and city development. Based on our model, we believe the data structure and connections of human could also help us to predict possibility of people future behavior after a very long observing time period and more specific algorithm.

6.3 Living laboratory & large scale experiment

Complex network is also widely used in the field of bioglogy. We could also apply our research and model on wildlife. We have a potential opportunity to work with some biologists at MUN to put to research on wild animal, Caribou, at Fogo island. Parts of caribou will be targeted with the devices. This experiment will be a long distance and scale test. We will both correct our data from caribou by biologists and tourists. Then we would look forward to fill the data into our network model to provide data structure and further information for helping biologists with their research. Biologists could use this model to predict the migration of wild animals. Even more, it may also help virologist to predict virus infection.

6.4 Temporal network development

The temporal network studied in this paper opens abundant space to explore for future research. Temporal weights can be further studied, which it does not only one attribute, the number of interactions between two nodes, but has more features. We also plan to build bipartite temporal graphs which can describe the network data precisely where there are two types of nodes, for instance, proximity network. In addition, we believe that the temporal network can be also appled in other fields. For instance, it can be developed in economics, i.e. the evolution of relations between inverstors and investees.

References

[1] Stefano Boccaletti, Vito Latora, Yamir Moreno, Mario Chavez, and Dong-Uk Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424:175–308, February 2006.

[2] Norman Biggs, E Keith Lloyd, and Robin J Wilson. *Graph Theory, 1736-1936.* Oxford University Press, 1976.

[3] Moira Burke and Robert E. Kraut. Growing closer on facebook: changes in tie strength through social network site use. In *Proceeding CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 4187–4196, May 2014.

[4] Martin Wikelski, Roland W. Kays, N. Jeremy Kasdin, Kasper Thorup, James A. Smith, George W. Swenson, and Jr. Going wild: what a global small-animal tracking system could do for experimental biologists. *Journal of Experimental Biology*, 210:181–186, 2007.

[5] Samuel R. Friedman and Sevgi Aral. Social networks, risk-potential networks, health, and disease. *Journal of Urban Health*, 78(3):411–418, 2001.

[6] Lauren Ancel Meyers, Babak Pourbohloul, M.E.J. Newman, Danuta M. Skowronski, and Robert C. Brunham. Network theory and sars: predicting outbreak diversity. *Journal of Theoretical Biology*, 232:71–81, Jan 2005.

[7] Reḱa Albert and Albert-LaśzloB́arabaśi. Statistical mechanics of complex networks. *Review of Modern Physicis*, 74:47–97, January 2002.

[8] M. E. J. Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics*, 45(2):167–256, 2003.

[9] Sergey Dorogovtsev and José Fernando Ferreira Mendes. *Evolution of Networks: From Biological Nets to the Internet and the WWW*. Oxford University Press, 2003.

[10] Adaora A. Adimora and Victor J. Schoenbach. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *The Journal of Infectious Diseases*, 191:S115–S122, 2005.

[11] Google.org. Google flu trends. (`https://www.google.org/flutrends/about/`), 2014.

[12] Stanley Milgram. The small world problem. *Psychology Today*, 1(1):61–67, May 1967.

[13] Sidharth Muralidharan, Leslie Rasmussen, Daniel Patterson, and Jae-Hwa Shin. Hope for haiti: An analysis of facebook and twitter usage during the earthquake relief efforts. *Public Relations Review*, 37:175–177, Jun 2011.

[14] Craig Ross, Emily S. Orr, Mia Sisic, Jaime M. Arseneault, Mary G. Simmering, and R. Robert Orr. Personality and motivations associated with facebook use. *Computers in human behavior*, 25(2):578–586, Mar 2009.

[15] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceeding WWW '10 Proceedings of the 19th international conference on World wide web*, pages 591–600, Apr 2010.

[16] Xiaoyong Yan, WenXu Wang, Ziyou Gao, and Yingcheng Lai. Universal model of individual and population mobility on diverse spatial scales. *Nature Communications*, 8(1639), 2017.

[17] Julian A. Drewe. Who infects whom? social networks and tuberculosis transmission in wild meerkats[. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1681):633–642, Nov 2009.

[18] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, August 2011.

[19] Filippo Simini, Marta C. Gonzaĺez, Amos Maritan, and Albert-LaśzloB́arabaśi. A universal model for mobility and migration patterns. *Nature*, 484:96–100, February 2012.

[20] Nicole B. Ellison, Jessica Vitak, Rebecca Gray, and Cliff Lampe. Cultivating social resources on social network sites: Facebook relationship maintenance be-

haviors and their role in social capital processes. *Journal of Computer-Mediated Communication*, 19:855–870, July 2014.

[21] Rodrigo K. Hamede, Jim Bashford, Hamish McCallum, and Menna Jones. Contact networks in a wild tasmanian devil (sarcophilus harrisii) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology letters*, 12(11):1147–1157, Aug 2009.

[22] C. Brooks, C. Bonyongo, and S. Harris. Effects of global positioning system collar weight on zebra behavior and location error. *The Journal of Wildlife Management*, 72(2):527–534, 2008.

[23] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10, 2015.

[24] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.

[25] Hyoungshick Kim and Ross Anderson. Temporal node centrality in complex networks. *Physical Review E*, 85(2):026107, 2012.

[26] Nic Newman. Apple ibeacon technology briefing. *Journal of Direct, Data and Digital Marketing Practice*, 15(3):222–225, 2014.

[27] Apple Inc. iOS: Understanding iBeacon. (`https://support.apple.com/en-us/HT202880`), Jun 2016.

[28] John Scott. *Social network analysis.* Sage Publications, third edition edition, Dec 2012.

[29] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, 4:879–891, Oct 2007.

[30] Leon Danon, Ashley P. Ford, Thomas House, Chris P. Jewell, Matt J. Keeling, Gareth O. Roberts, Joshua V. Ross, and Matthew C. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.

[31] Alden S. Klovdahl. Social networks and the spread of infectious diseases: the aids example. *Social Science & Medicine*, 21:1203–1216, 1985.

[32] Robert M. May and Roy M. Anderson. Commentary transmission dynamics of hiv infection. *Nature*, 326(12):137–142, Mar 1987.

[33] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, Aug 2015.

[34] Lisa F. Berkmana, Thomas Glassb, Ian Brissettec, and Teresa E. Seeman. From social integration to health: Durkheim in the new millennium. *From social integration to health: Durkheim in the new millennium*, 51:843–857, 2000.

[35] Mark Newman. *Networks: an introduction.* Oxford University Press, 2010.

[36] Justin Cranshaw, Eran Toch, Jason Hong, Aniket Kittur, and Norman Sadeh. Bridging the gap between physical location and online social networks. In *Proceeding UbiComp '10 Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 119–128, Sep 2010.

[37] Ana Perisic and Chris T. Bauch. Social contact networks and disease eradicability under voluntary vaccination. *PLoS Computional Biology*, 5(2), Feb 2009.

[38] David W. Winkler6, Peter H. Wrege, Paul E. Allen, Tracey L. Kast, Pixie Senesac, Matthew F. Wasson, Paulo E. Llambías, Valentina Ferretti, and Patrick J. Sullivan. Breeding dispersal and philopatry in the tree swallow. *The Condor*, 106(4):768–776, 2004.

[39] Thomas Richard Edmund Southwood and Peter A. Henderson. *Ecological methods*. John Wiley & Sons, third edition edition, 2000.

[40] Paul M. Cryan, Michael A. Bogan, Robert O. Rye, Gary P. Landis, and Cynthia L. Kester. Stable hydrogen isotope analysis of bat hair as evidence for seasonal molt and long-distance migration. *Journal of Mammalogy*, 85(5):995–1001, 2004.

[41] Rebecca E. Doyle, John C. Broster, Kirsty Barnes, and William J. Browne. Temperament, age and weather predict social interaction in the sheep flock. *Behavioural Processes*, 131:53–58, Oct 2016.

[42] Christian Rutz, Michael B. Morrissey, Zackory T. Burns, John Burt, Brian Otis, James J. H. St Clair, and Richard James. Calibrating animal-borne proximity loggers. *Methods in Ecology and Evolution*, 6(6):656–667, 2015.

[43] Suzanne Prange, Trevor Jordan, Colin Hunter, and Stanley D. Gehrt. New radiocollars for the detection of proximity among individuals. *Wildlife Society Bulletin*, 34(5):1333–1344, Dec 2006.

[44] Pavan Sikka, Peter Corke, and Leslie Overs. Wireless sensor devices for animal tracking and control. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, Nov 2004.

[45] Sophie Curtis. Quarter of the world will be using smartphones in 2016. (`http://www.telegraph.co.uk/technology/mobile-phones/11287659/Quarter-of-the-world-will-be-using-smartphones-in-2016.html`).

[46] John Tang, Mirco Musolesi, Cecilia Mascolo, and Vito Latora. Temporal distance metrics for social network analysis. In *Proceedings of the 2Nd ACM Workshop on Online Social Networks*, WOSN '09, pages 31–36, New York, NY, USA, 2009. ACM.

[47] Marta C. González, César A. Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453:779–782, June 2007.

[48] Linton C.Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

[49] Ulrik Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, May 2008.

[50] Google Inc. Google indoor position system.

[51] Basic rate/enhanced data rate (br/edr). (https://www.bluetooth.com/what-is-bluetooth-technology/bluetooth-technology-basics/br-edr), 2016.

[52] The network simulator. (http://www.isi.edu/nsnam/ns/).