Student thesis series INES nr 335

Compute a Crowdedness Index on Historical GIS Data- A Case Study of Hög Parish, Sweden, 1812-1920

Fei Lu

2015
Department of
Physical Geography and Ecosystems Science
Lund University
Sölvegatan 12
S-223 62 Lund

Fei Lu (2015).
Compute a Crowdedness Index on Historical GIS Data- a Case study of Hög Parish, Sweden, 1812-1920
Master degree thesis, 30 credits in *Geomatics*
Department of Physical Geography and Ecosystems Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2014 until *June* 2014

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily represent those of the instit

Compute a Crowdedness Index on Historical GIS Data - A Case Study of Hög Parish, Sweden, 1812-1920

Fei Lu

Master thesis, 30 credits, in *Geomatics*
December, 2014

Supervisor
Ali Mansourian, Lund University

Exam committee:
Lars Harrie, Lund University
Finn Hedefalk, Lund University

Department of Physical Geography and Ecosystem Science

Lund University

Abstract

As the development of human civilizations, the classic definition of crowdedness index or population density is not able to describe population dynamic conciliations properly. Furthermore, the detailed population information is normally only stay on urban or country levels. Methods to compute crowdedness index of a small scale-area are lacked. Demographers need some new methods to compute human crowdedness index. In recent years, geographical information techniques have played an important role in historical demography studies. Most researches are based on historical GIS database. The aim of this project is to evaluate current crowdedness indexes and develop new indexes based on historical demographical and geographical data. In this project, two current crowdedness indexes are reviewed in the study area Hög parish. It also develop some new crowdedness index of buildings, two population's crowdedness indexes, and also two crowdedness indexes of divided property unit in the study area. In addition, all the indexesare evaluated and discussed based one a set of test evaluations. Conclusions are made based on those case studies and evaluations. From the study results and evaluations, a building crowdedness index method is suitable for this case; all of the three population crowdedness index (social interaction index, betweenness centrality and church index) are suitable for this case study; and finally both two property unit population indexes are able to compute results but also need to be developed as well. However, the test analyses are only based on a small study area with limited buildings, populations. Some developed index methods do not provide satisfactory results. For the further study, more cases should be tested and development of every method should be done.

Keywords: Physical Geography and Ecosystem analysis, GIS, historical GIS, demography, crowdedness index

Acknowledgement

It is a pleasure to thank all those who gave me the possibility to accomplish this degree project thesis.

First of all I would like to express the heartfelt thanks to my supervisor Ali Mansourian, who has supported me through the process of this thesis work with his patience and knowledge while allowing me to work in my own way. Without his enlightening instruction, impressive kindness and patience and valuable hints, I could not complete this thesis. Furthermore, I would like to thank the teachers and PHD students of GIS Center Lund University, from which I gained the necessary useful data and helps during this thesis work.

At last, I would like to appreciate my family and friends who always supported me warmly and silently as well.

Fei Lu

November 2014

Table of Content

1. Introduction

1.1 Background

The fast development of geographical information and computing technologies (geo-ICT) has enabled geospatial tools for researchers such as geographical information system (GIS) (Van et al., 2009). Those geo-ICT tools can be used in many geographical studies especially in e.g. historical GIS (store, display and analyze historical data by using Geographical information system techniques) and demography researches. In the recent years, geo-GIT tools are playing a more and more important role in historical and (historical) demography research. Because historical data could be combined, linked, and analyzed by these tools easily.

Traditionally, the historical GIS project mainly focus on the administrative spatial units (like countries, wards, or communities) to combine with demographic and other socio-economic data (Knowles, 2008). However, during recent years, a trend is shown that more attentions of demographic studies have been shifted from traditional large-scale process to analyze longitudinal micro data in the form of "life histories". Because demography deals with the fates and choices of individual, the micro-level is optimally suited to study chains of causation (Alter et al., 2009). Linking the longitudinal data with detailed geography data enables researchers to trace individuals in both space and time. Demographers can now analyze the process of family formation and the change that span lifetime and be even followed across generations.

During the last decades of years, several historical longitudinal databases have been created (Holdworth, 2003). Individuals' or families' information or characteristics are included in those databases as the most important elements. However, in order to study the datasets at micro level, the researches not only need to be constructed on individual's life course events but also the corresponding geographical locations (Ekamper, 2010). Thus, effects of the individual's life can be analyzed in a geographical environment. To be able to correlate individual's life with the surrounding geographical environment, the identification and visualization of geographic locations on maps is critical. Using suitable (in appropriate scale) historical maps and linking them with demographic database are the key factors for analysis.

Back to the 18th to 20th centuries in Sweden, living geographic environments normally affected people's life since most citizens were working in agricultures (Bengtsson and Dribe, 2002).This would enable demographers locate an individual's life line on property unit level by analyzing how the surrounding land/soil condition affect the production, how the economics of a community could be affected by individuals' proximity and gathering, and how individual's health was affected by the population density in a property or several neighbor properties. This thesis reviews some common methods for measuring crowdedness index. Some new methods for crowdedness indexes are developed as well. All of those methods are tested and test results are discussed as well. A conclusion is made at the end.

1.2 Research questions

This project focuses on the crowdedness index of the study area. Some research questions are presented by reviewing and developing some crowdedness index methods.

The major questions that in this study attempted to address in the study are:
1) What is the buildings closeness index of the study area?
2) What is the population crowdedness index of the study area?
3) What is the population crowdedness situation of each property unit?
4) How will crowdedness index of a property unit change when population data is not accuracy enough?

1.3 Aim

The main purpose of this project study is to test and develop crowdedness indexes methods. The specific aims are as followed:
1) Test one current buildings crowdedness index method and one current population crowdedness index method.

2) Develop and evaluate one method for buildings crowdedness index and two methods for population crowdedness index.

3) Develop some methods for measuring crowdedness index on property unit level.

4) Evaluate all the results to identify the most suitable ones.

1.4 General methodology

Figure1.1 shows the outline of the whole study. A literature review study is carried out to obtain a general theoretical background of the crowdedness index based on historical data. The whole study is divided into three steps: the buildings crowdedness indexes are measured firstly, then the measurements of population crowdedness indexes, and lastly population crowdedness indexes of property unit level. There are three methods to measure the buildings crowdedness indexes: Density-Based clustering method, crowdedness index method for buildings, and betweenness centrality. The density-based clustering and betweenness centrality methods are found in the literatures. The crowdedness index for buildings is developed in this thesis study. All of these methods are tested and evaluated by defined criteria. From the literature review results, lack of population index measurements of a small area or property unit studies are found. Two kinds of population crowdedness indexes are developed in this thesis study: social activity index and travel index. These two indexes are tested and evaluated by defined criteria. Two kinds of crowdedness index measuring of populations on property unit levels are developed: crowdedness index of property units on individual property unit level and crowdedness index of property units based on buffer zone selection. All the outcomes are evaluated by defined criteria.

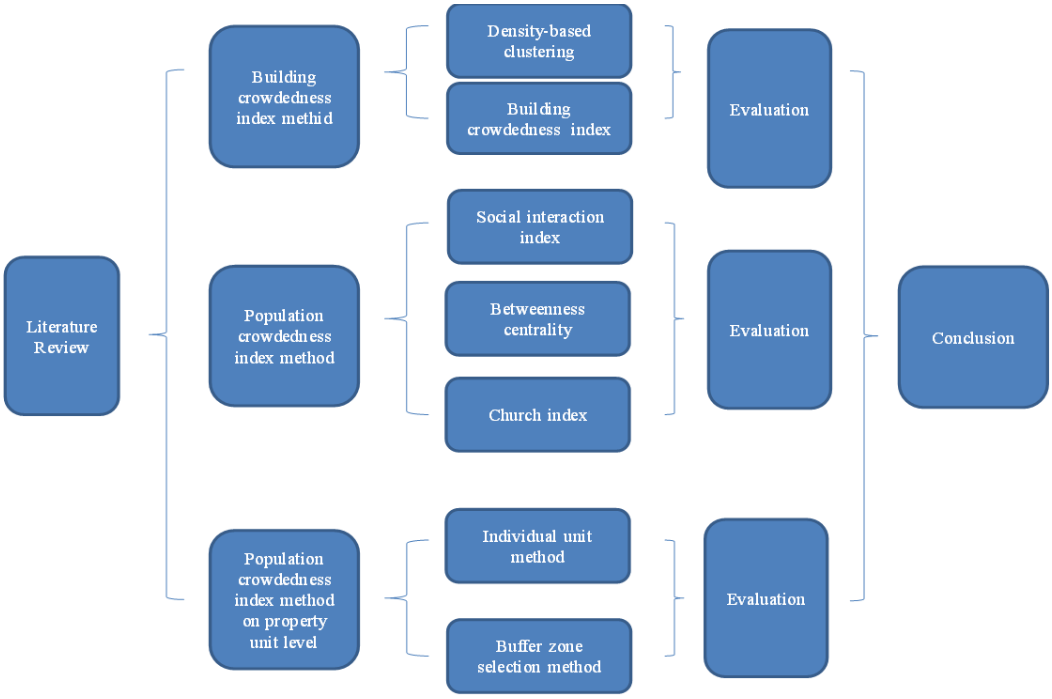


Figure 1.1, Outline of the thesis study

1.5 Organization

Chapter 2 reviews presented studies on measuring the crowdedness for small indoor areas, measuring of crowdedness for larger scale areas based on various GIS and demographic techniques, and historical GIS study based on historical databases.

Chapter 3 reviews and presents three building crowdedness index measurement methods. The theoretical background and case study work is introduced as well. Evaluations are made based on study results and pre-defined criteria.

Chapter4 presents two population crowdedness index measurement methods. Both the theories and implementations of those two methods are introduced. Same as section 4, evaluations are made based on presented criteria as well.

Chapter5 introduces two methods to measuring population crowdedness index on property unit level. Two small cases are studied and evaluations of the study results are evaluated too.

Chapter6 and 7 discuses all the evaluation results and conclusions of the whole project study are made.

2. Literature review

Some basic ideas of crowdedness and previous related studies are reviewed in this Chapter. In section 2.1, some basic ideas about crowdedness in demography are introduced. It also reviews some current methods to measure crowdedness index in a small indoor area. Section 2.2 reviews some methodologies to estimate population value of an area when the population data is in low accuracy level. It includes both traditional demographical methods and modern GIS methods. Since this study focuses on the measurement of crowdedness indexes of historical data, the section 2.3 reviews the development of historical GIS techniques for demographic studies and some presents historical GIS databases.

2.1 Crowdedness

There are few directions that related with the issue of crowding. For demographers, the experiment on human use of space is the study field (Sommer, 1967). In demographics, the word “crowdedness” or “population density” is defined as a measurement of density or physical crowing (Frank, 1990).The classic method to measure population density in an administration area (unit per area indicants) is using total population value divided by the area size (Stokols, 1972). However, this method has severe shortcomings since the population density is only a measurement of physical proximity and the actual spatial orientation is disregarded. Descriptions of population dynamic conciliations cannot able to seen (Westover, 1989). Sommer (1971) presented that there is a distinctions between the word crowdedness and density. The crowding could be seen as a stressful consequence of population density. As the development of city civilization, individuals are living and working more socialized. Since people are closer to each other, the casing of stress-related and physical effects is increased as well (McCain et al, 1976). The old population density method cannot describe the population crowdedness situation properly. A new method to describe crowdedness and physical distance in a specific area is needed.

The measurements of crowdedness in small areas public facilities such as offices, train stations, hospitals are developed during the last decades. O’Brien (1990) presented a method based on a Euclidian distance formula to measure crowdedness index for populations. This formula can be used to measure relative population density of an indoor small space area such as offices, submarines, etc. The crowdedness index measurement is applied in medical files as well. There are also some real time crowdedness index measurement methods existed. Melissa et al (2011) compared various methods to measuring crowdedness index of ED (emergency department) in hospitals. All of those methods are based on frequency if ambulance diversion or by provider opinions in a time line. Another field to measuring crowdedness index, by using real time line, is in public transportation facilities. SoonKi et al. (1993) presented a so called “real time crowdedness measuring system” for train stations. It based on real time line and CCD cameras. This method combines time line and CCD cameras facial recognition to measure real time crowdedness index of train and bus stations. It was tested as fully functions in many large trains and bus stations.

2.2Population estimation on large geographic areas

Urban areas are usually divided into several blocks or districts. In some geographical studies, these divided areas could be defined as property units (Stone et al, 2007). Knowledge of population situations of these property units is important for many socioeconomic, political and environmental problems. This detailed population information supports necessary planning practices for both public agencies and private sectors (Rees et al, 2004). Although the detailed population information is important for many aspects, such detailed data usually not accurate enough or only available for limited years (e.g. in the US national demographic database, the some population information in only stay on cities level). Furthermore, in this thesis study, the population data is only stay on property unit level as well. It is necessary to present techniques to derive population estimates for property units when the detailed population data is not available. There are mainly two types of techniques to estimate population for a region area: demographical and geographical techniques.

There are three popular types of demographical methods to estimate and measure population crowdedness (Ghosh and Rao, 1994). Smith and Mandell (1984) presented a so-called Component II (CM-II) technique. The formula can be calculated by adding total number of births and net migration at all ages and subtracting the estimated number of deaths from population totals at a specific year. The second popular method is called ratio correlation (RA); it builds up a relationship between a change share and different types of symptomatic variables and uses these relationships to estimate the population distribution (Martin and Serow, 1978). The Housing unit (HU) method is the most common technique using today (Smith and Cody, 2004). This method multiply the number of housing units in a property units, estimating from tax payments or electric bills, by person per household (PPH), then add up their product with the group quarters population (Smith and Cody, 2004). The limitation of the traditional demographical methods is that, due to the data constrains and policy concerns, the scale may only stay on country or city levels.

As the development of geographical information techniques (GIT) and remote sensing (RS), geographical methods have been applied since 1950s (Watkins, 1984). With the help of RS technique, the HU method is adopted more accuracy. The high resolution satellite images could recognize house units with low uncertainties and fast processing speed. Then the total population and their distributions could be estimated through a special demographical survey. The limitations of using remote sensing data are that parameters extracted from satellite images have indirect linkage with the built up environment and the resolution of images are often too coarse to recognize HUs. To implement the small-area population crowdedness, a new method was presented by Deng and Wu (2013). The method is working through with combing high resolution satellite images, parcel data, and land use data together to estimate and distribute population information of a small area.

2.3 Demographic and Historical GIS

As the review of the 2.2 section, there are effective methods to estimate and distribute population information of one property unit by combining traditional demographic and GIS techniques. However, since the GIS techniques only started at 1950s, those methods cannot be used for historical demographic study (e.g. studies based on data of 19^{th} or 20^{th}century) any more. A national historical geographic information system (NHGIS) database is a suitable data source for that kind of studies. The amount of NGGIS database has increased for many Europe countries (Knowles, 2005). Generally, these databases covered from 19^{th} to 20^{th} century's period. Changings boundaries of administrative units, changings of land uses, and changings of human geography during the covering periods are recorded by these databases as well (Gregory et al, 2002). National Demographical Database is another powerful database that could be used for historical demographic studies. Longitudinal information on individual and their family information are included in this kind of database.

During the recent years, more attentions were shifted from studying demographic regimes and large-scale process to micro data in the form of life histories (Alter et al, 2009). By using national historical and demographic database; demographer could be able to analyze the process of family formation and change that span life time (Edvisson, 2000). Ekamper (2010) presented a method to link historical cadastral map data, population census data, and population register data of a Dutch city during the 19^{th} century. This method provide additional tools to analyze historical demographic on micro level. Ruggles et al. (2008) created the database Integrated Public Use Micro data Series Project (IPUMS). Census records all over the world of individual level are available in this population database. This database is encouraging many new researches with historical demographic data.

3. Materials

3.1 Study area

The study area is located in the province Skåne on the southern Sweden. Figure 2.1 shows the location of the target area. The study area belongs to the European hardwood vegetation zone. Agriculture land and forests were the primary land use type. The study area is located between latitude 55.76 °N to 55.85 °N, and longitude 12.97 °E to 13.33 °E in WGS 84 coordinates system. The total coverage of the study area is around $50km^2$.

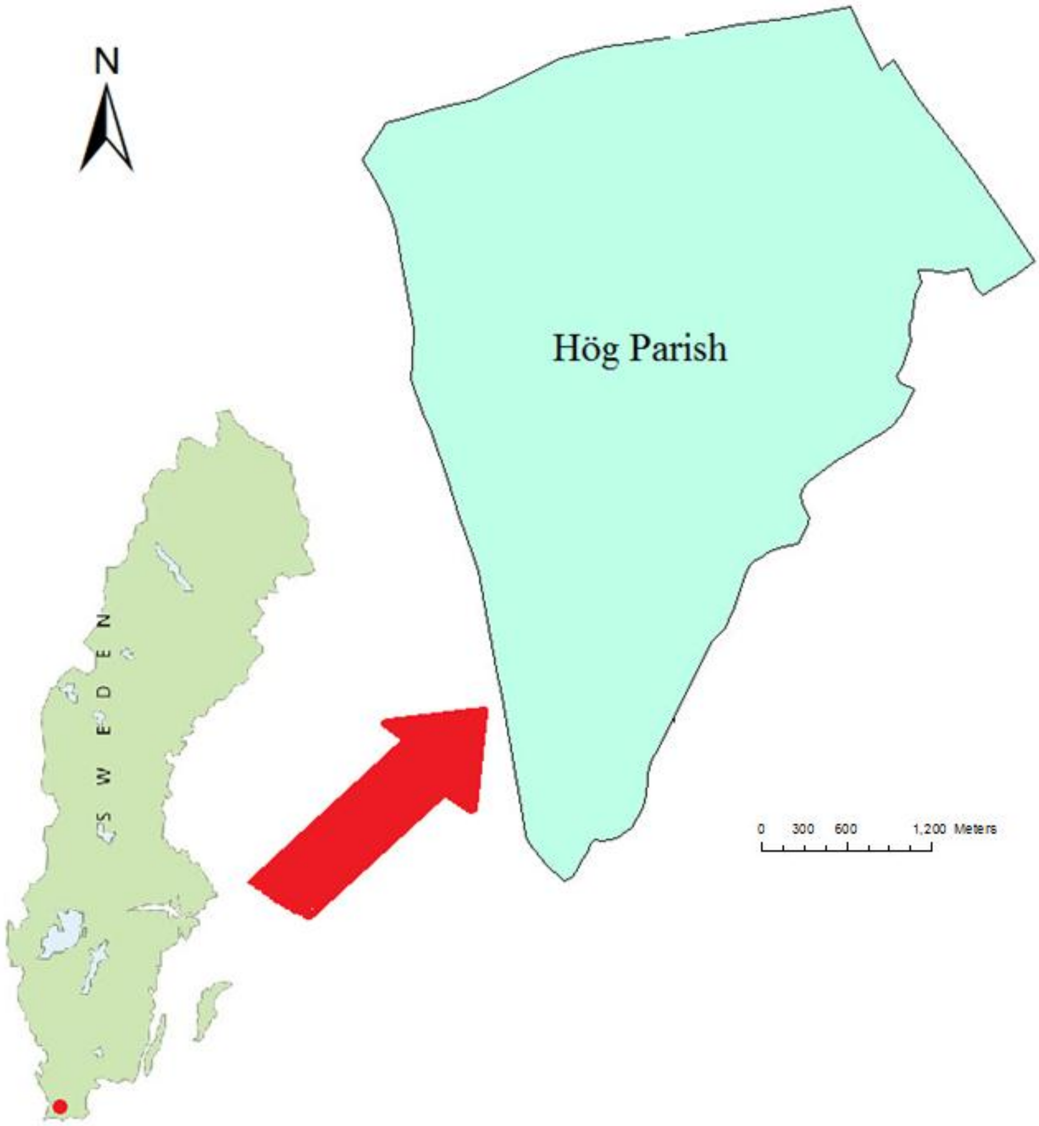


Figure 2.1, Study area

3.2 Material Data

There are two types of data in this study: demographic data and historical maps. The demographical data is provided by the Scandia Economic Demographic Database (SEDD). It contains longitudinal demographic and economic information about all persons that lived in the target area Hög parish from time serious 17^{th} century and onwards. These data are created and developed by Center for Economic Demography (CED) and Lund University during the last decades (Bengtsson et al, 2012). In this database, individual's birth location and living time are recorded. However, for the birth location, often only parish or vague address is given. Thus person's birth or living location is on low resolutions level (property unit level). The historical maps cover time series from year 1757 to 1915. The maps are provided by the Swedish National Land Survey Agency (Lantmäteriet). These maps provide five parishes areas; Hög parish is one of them. Land uses, administrative and topographical details of the study areas are included in the data as well. All of the paper maps data are digitized, geo-referenced, linked with demographic data, and stored in a graphical database.

4. Crowdedness Index for Buildings

This chapter introduces two algorithms to measure crowdedness index of buildings: DBSCAN clustering methods and building crowdedness index. The DBSCAN clustering algorithm is a statistical method to describe patterns distributions in an area; the building crowdedness index algorithm is a new method that developed in this study. The measurement of building's crowdedness index can be seen as a preparation for measuring population crowdedness. Both methods are introduced and tested in this section.

4.1. Index methods

Two main algorithms are introduced in this chapter: density-based clustering methods and a building crowdedness index methods. Both of them are based on Euclidian distance of a rectangular coordinate system. A case study will be presented to evaluate these two building crowdedness indexes.

4.1.1 Clustering method

Clustering analysis is a popular multivariate statistical approach. By using this approach, spatial patterns are identified by distinguishing the similarities and dissimilarities between objects. Normally, the clustering analysis approach can be broadly divided into four methods (Malik et al, 2013): hierarchical, partitioning, density-based and grid-based methods. The choice of a method is related to the characteristics of data. In this case study, density-based method is used since this method can create clustering area without shape limitations. And this method can give results both visually and mathematically to analyze the clustering.

Density-based method could discover the underlying data by grouping observations through specific density functions (Gaonkar and Sawant, 2013). DBSCAN (Density-based Spatial Clustering of Application with Noise) is one algorithm of this method (Ester et al., 1996). The two input parameter of this method are: Minimum number of points (*MinPts*) and Neighborhood distance (*Eps*-neighborhood). The general process of DBSCAN is: firstly, checking each point in the dataset, if the *Eps*-neighborhood of an object *p* has more than *MinPts* objects, a new cluster is created with a *p*as core object; the second step is the procedure repeats looking for the objects, which are directly density-reachable from those core objects; and then the procedure will not be ended until no new clusters are created to the current clusters.

4.1.2 Building crowdedness index

A buildings crowdedness index equation measures building crowdedness index in a two dimensional rectangular coordinate system. The calculation process is based on the Euclidian distance between each building object. In a given coordinate system, object *O* is the original point (*O*), given a distance *D*; a circle buffer zone based on the center of the circle *O* and radius D could be drawn. The area inside the generated buffer zone is called the target area. Every building, which is inside the target, is O_i. Distance between each O_i and O is defined as d_i. Figure 4.1 shows the relationships between each object.

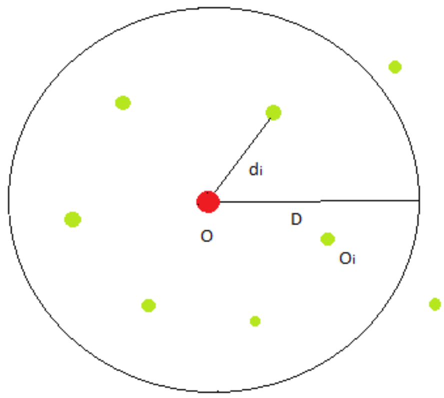


Figure 4.1, building crowdedness index

The index of object O is calculated by the following equation:

$$C_{index} = \frac{n}{\sum_{i=1}^{n} d_i}$$

Where the C_{index} is the building crowdedness index going to be measured, n is numbers of buildings inside the buffer area; d_i is the distance between each target point to the original point. Theoretically, by following this equation, crowdedness index of the original point will be increased by decreasing the sum of distance d_i. In other words the closer objects are, the larger index value of the original point is. This is a method to calculate crowdedness index of each building.

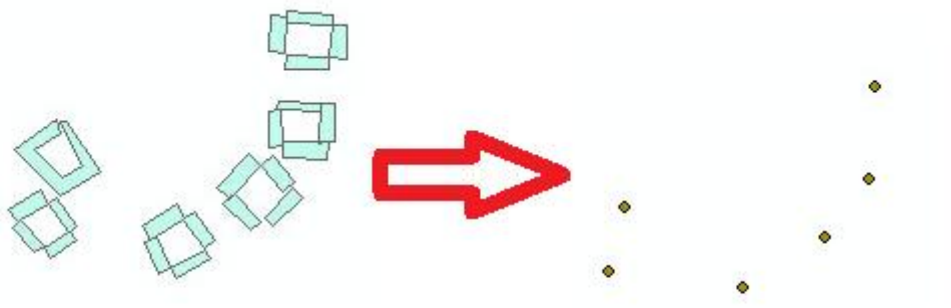
4.2 Case study I

In this case study, both methods mentioned above is tested in the study area. Results of each approach are presented separately as well. The evaluation of index methods for the data is finalized at the end of this section.

4.2.1 Method

Clustering method

In order to make the original building data suitable for DBSCAN analysis, all the building polygons data are transformed into points by using ArcGIS 10. (Figure 4.2)



.

Figure 4.2, transformation procedure: from polygons to points

After the transformation process, all the building points are laid in a geographical projection system (Swedish National Reference System: SWEREF 99) with *x* and *y* coordinates values. Then all of those coordinates were exported as table formats files.
In the software RStudio, a DBSCAN function is used to perform the DBSCAN algorithm. During the clustering process, two parameters should be decided: *Eps* and *MinPts*. In order to decide those two parameters, a *k-dist* plot can be used as a heuristic so that the k^{th} nearest neighbor are graphically presented. A distance matrix is created by *Matlab* and a curve is simulated due to the points. Figure 4.3 shows the simulation result. In the figure 4.3, the horizontal axis shows the ID of each point, and the vertical axis shows the distances between each point to its N^{th} point. In this case, the N^{th} was testes as 3, 4, 5, and 6.

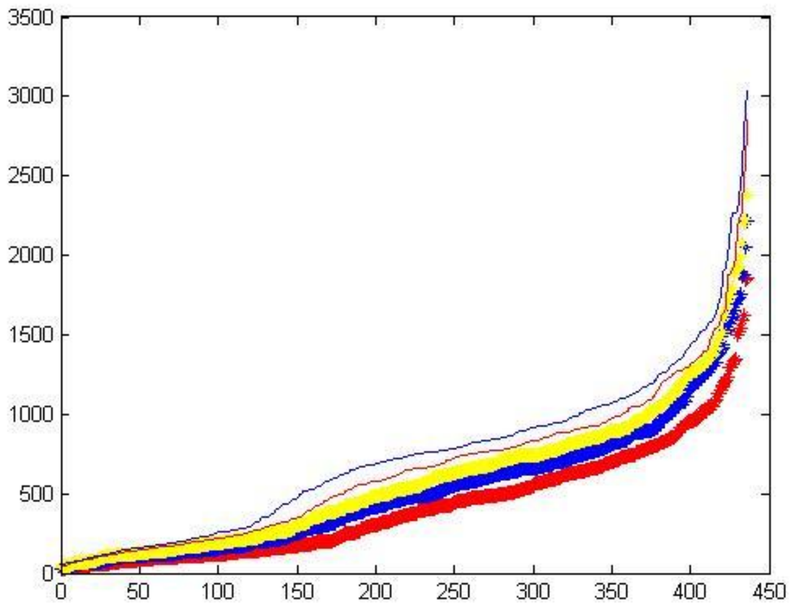


Figure 4.3, the plot for the points with different values of k. The *y* axis is the distance from each point to its k^{th} nearest neighbor while the x axis is the point ID

From this plot, levels of density could be detected, thus identifying whether the clusters have varied densities. The vertical axis value of the point where a sharp change occurs is the optimal value for Eps (Gaonkar and Sawant, 2013). As presented by Liu et al. (2007), through the value of Eps, relies on k, it does not change dramatically as the *k* increases. Thus, based on the initial value of k, k+1 is used as minimum number of objects in a cluster. In this case, the k value is deceived as 6. Then after the clustering process, the results would be plotted as graph.

Building crowdedness index equation

To perform the building crowdedness index equation, the buffer radius should be decided first. As presented in the literature review section, there are few related studies focusing on property unit levels. In this particular case, 1000 meters is used as the distance *D*. Same as the DBSCAN clustering method, all the building points are distributed in the same reference coordinate system with *X* and *Y* coordinate. Distances from one building to all other buildings can be calculated by equation:

$$d_i=\sqrt{(x_i-x_0)^2+(y_i-y_0)^2}$$

where the d_i is the Euclidian distance between building o to building i, x_o is the x-coordinate of building o, x_i is the x coordinate of building i, y_o is the y coordinate of the building o, and y_i is the y-coordinate of the building i. After the distance calculation process, the buildings crowdedness index equation is processed by *Matlab*. In ArcGIS, an attributed table was built for buildings and all the crowdedness indexes of corresponding buildings are stored in this table as well. To enhance the visibility of the result, index values are classified into five classes and then exported as a map.

4.2.2Results

DBSCAN

The DBSCAN algorithm is performed by setting k value to6, and the *MinPts* is determined as $k+1$ which equals to 7. Figure 4.4 shows the plot clustering from DBSCAN algorithm when k is equal to 6 and *Eps* equals to 58.1. Objects with triangular shape and colors are classified with clustering (yellow, blue, black, light blue, red, green, and purple); objects cannot be classified in to clustering groups are defined as noise which shows with white color and circle shape.

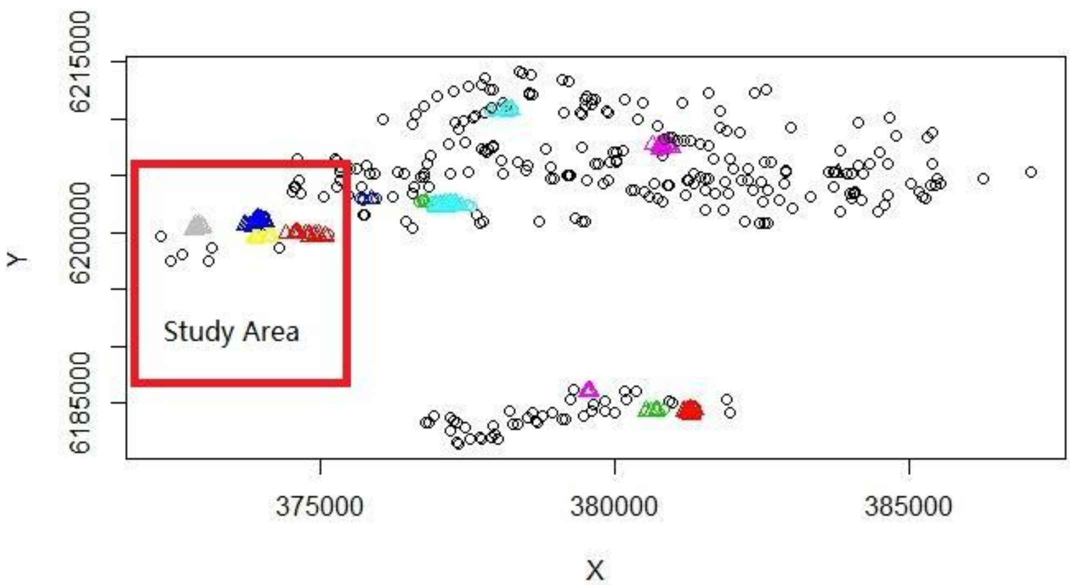


Figure 4.4, the plot clustering from DBSCAN algorithm when k=6 and *Eps*=58.1

Building crowdedness index

Figure 4.5 shows the result of building crowdedness index algorithm in study area Hög. The indexes are classified to 5 classes, grey color shows the smallest index value and red color shows the largest index value; yellow, green and blue are in between of the extreme values. Table 4.1 shows the *x* and *y* coordinates and index value of each input building.

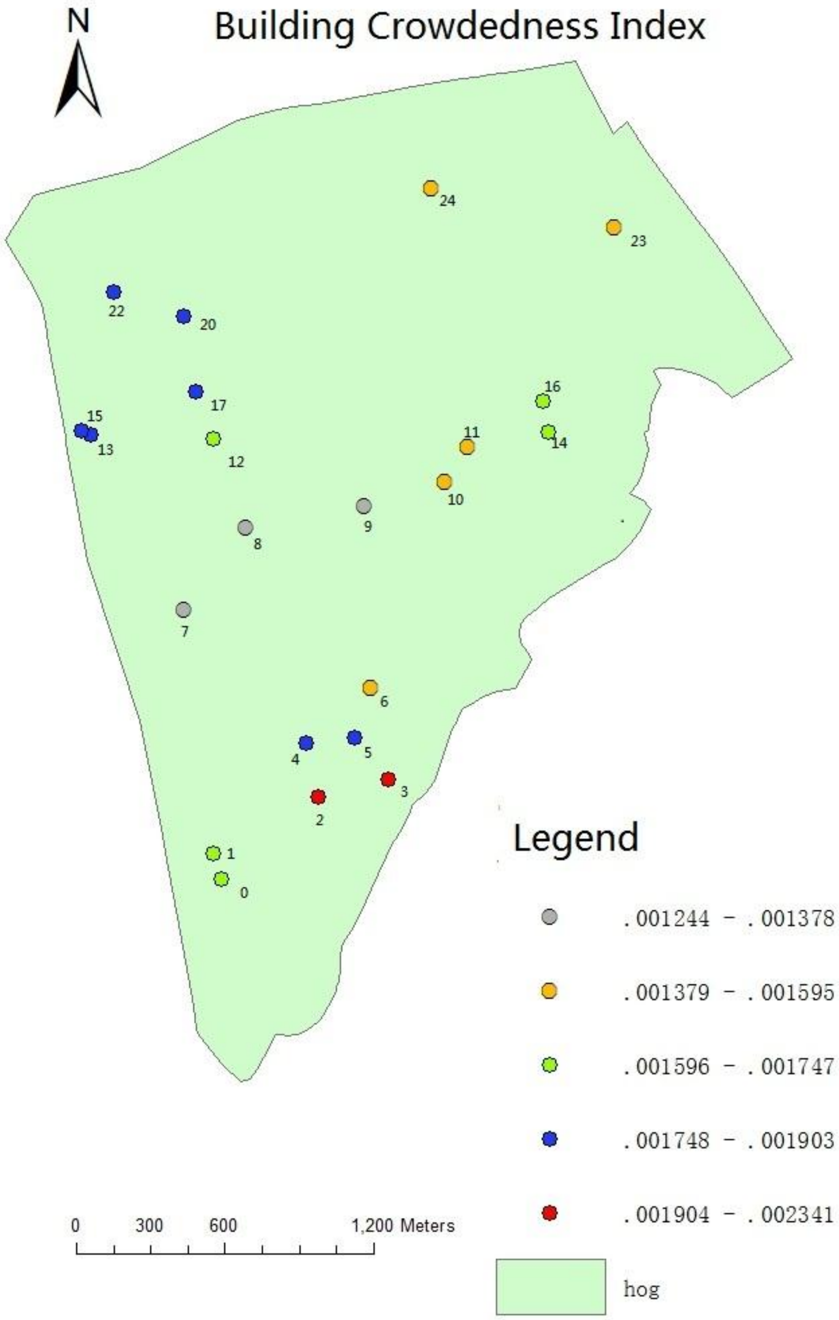


Figure 4.5, the plot of index value in study area

Table 4.1, *x* and *y* coordinate and corresponding index value

ID	POINT_X	POINT_Y	C_index
1	378019	6181885	0.002133
2	378933	6184107	0.001591
3	376913	6183849	0.001855
4	377194	6183751	0.001858
5	377246	6183445	0.001877
6	377316	6183256	0.001747
7	376820	6183272	0.001883
8	376784	6183292	0.001816
9	377196	6182566	0.001244
10	377445	6182897	0.001378
11	377923	6182985	0.001351
12	378248	6183086	0.001479
13	378341	6183223	0.001595
14	378668	6183284	0.001671
15	378647	6183410	0.001701
16	377689	6182027	0.001903
17	377740	6181810	0.002341
18	377885	6182052	0.001782
19	377951	6182253	0.001527
20	377348	6181479	0.001673
21	377313	6181583	0.001632
22	378192	6184265	0.001452

4.2.3 Evaluation

In this section, both methods for measuring crowdedness index of buildings are evaluated. The evaluations focus on two aspects: visual and quantitative. The visual evaluation focuses on maps layout. The quantitative evaluation on the other hand focuses on the output index values, and analyzes the results statistically. For the DBSCAN clustering methods, clusters should be able to be seen clearly, and noises should not take large proportions. For the building crowdedness index logarithm, more buildings close to each other should have larger closeness index values.

DBSCAN Clustering

One may argue that the clustering method could not provide the specific crowdedness index value of each building. In this case study, if many buildings are classified into the same cluster, a conclusion can be drawn that this cluster has higher index value than other clusters. This conclusion would be used as a reference for other spatial index method. On the plot graph of DBSCAN clustering, one can see that objects in the input data set are classified into 7 clusters. Buildings are classified into a cluster can be seen as have larger crowdedness index value than buildings are not been clustered. Clusters with larger numbers of buildings have larger index values than clusters with less numbers of buildings. However, buildings inside the study area are only clustered into4classes (not include noise class). It is not good enough for testing the other index algorithm. Furthermore, since the input buildings data is not only including buildings in the Hög parish but also buildings from other parishes. The result may be affected by buildings from other parishes as well. Thus, the DBSCAN method is not suitable for this case study.

Buildings crowdedness index equation

It is clearly to see in the output table that all buildings' index values are calculated. On the map layout, all index values are classified to 5 classes. For the visual evaluation, points which are close to each other have similar index values and classified into same group. However, building 24 and 23 have long distance compare with building 7 and building 8. But the crowdedness index values of building 24 and 23 are larger than building 7 and 8. These two buildings show unreasonable result.

In order to test this equation quantitatively, an evaluation test is used. In the evaluation test, one building is selected randomly as the original building *o*. The radius distance *D* of buffer zone is set as extreme value infinitely to make sure all other buildings in the study are would be included in this test. After setting up the test condition, the only variable of this test *di* is increased by the increment of distance between the original point *o* and each *i* point. Figure4.6 shows the increment of *di*. Point 1 is tested firstly, and then after exporting the index value O_{1}, point 2 is added to the simulation, and then output the index O_3, repeat the steps before till all points *i* are tested.

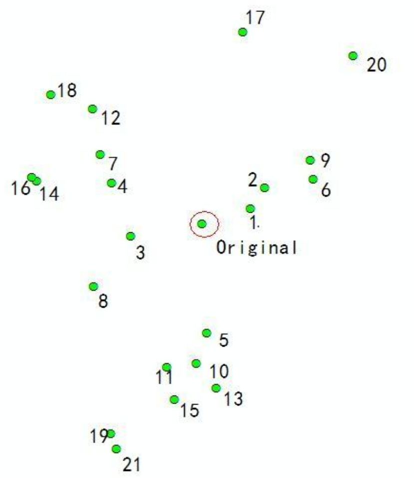


Figure 4.6, Test order of crowdedness index

After the evaluation test, a curve can be created by those output index values. Figure 4.7 shows the test index of the study area. The horizontalaxis is the number of points tested in the simulation; vertical value is the exported index value.

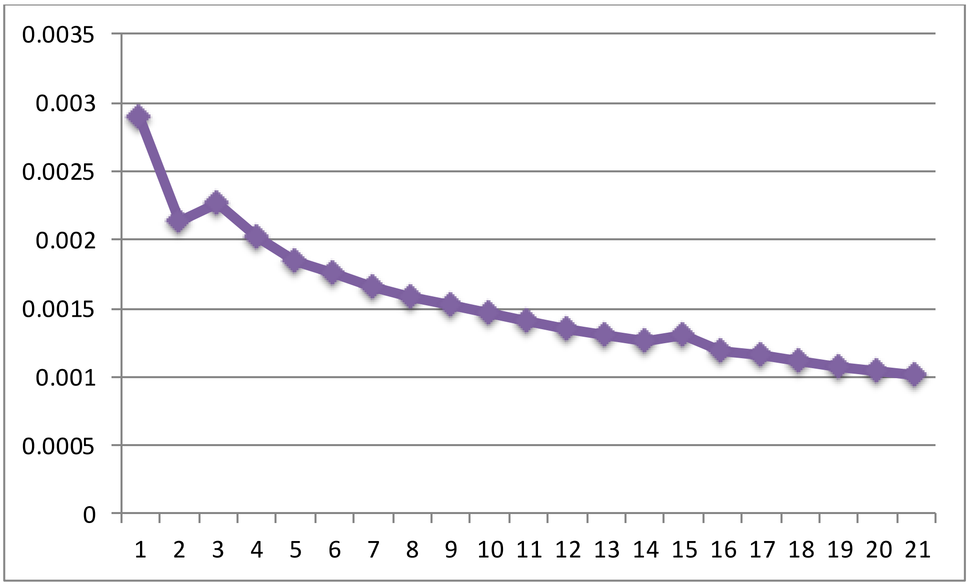


Figure 4.7, evaluation test index curve of study area

From the curve, the crowdedness index value is decreased by following the increase of numbers of buildings to be calculated. However, at the point 3 and point 15, the index values are increased which against the curve's trend. The evaluations are tested 3 more times by selecting the original point randomly, and getting similar type curve graphs. Thus, although from the whole curve, the index is inversely proportional to the increasing number of buildings. But some diverse situations exist. The equation is not stable enough for this case study.

5. Population crowdedness index

After the process of buildings crowdedness measurements, population crowdedness index is going to be measured for the next step. This section introduces two main algorithms to measure the population crowdedness index of the study area: the so called "social interaction index" and "travel index". For the travel index, two sub index methods "betweenness centrality" and "church index" are introduced. All of these algorithms are working with urban road network system. A case study is used for evaluating these methods.

5.1.1 Social interaction index

Nowadays, people could contact each other in different ways: by mails, by telephone and by Internet network, etc. Back in the 19th century, the most common way that people could connect to each other was through meeting at public locations (such as the markets) and private residences. The social interactions processes are done during the meeting time. The social interaction index can measure the index value of an area. This method is working with urban network system. By using distances between buildings along a road network and the number of persons living in each building, the equation can calculate an index value of each building in a road network system.

In a given road network, locations of buildings are known and all buildings are connected to the network as well. Given an original building *O*; given a distance *D*, we create a buffer zone along the given network by using the *O* and *D*. In the created buffer zone, each building inside the buffer zone is building *i*, the distances between building *O* and other buildings along the given road network are defined as d_i, the number of persons lives in the original building is called M_0, the number of people living in building *i* is M_i. Figure 5.1 shows the working theory of social activity index. Each building which is inside the buffer zone along the network is selected.

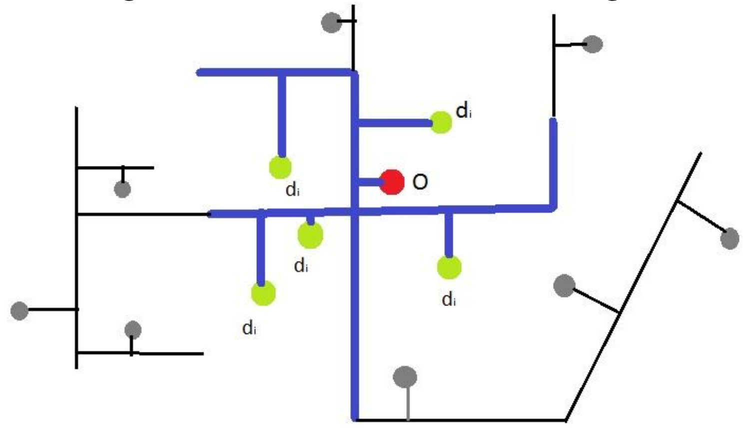


Figure 5.1, social interaction index

The index of building *O* could be calculated by the following equation:

$$S_{Index} = \frac{M_0}{50} + \sum_{i=1}^{n} \frac{M_i}{d_i}$$

Where the S_{Index} is the social interaction index that going to be measured, $M_0$is the number of people living in building *O*, M_i is the number of people living in building *i*, d_i is distance along the road network between building *O* and building *i*, *n* is the number of buildings in the buffer area.

According to the equation, the social activity index is depended on three variables: M_0, M_i, and d_i. Since both population information and distance are considered in this method, the index value is realistic. However, it is necessary to know the detailed population information of each building. Collection processes of the population information may take large time or large amount of funds.

5.1.2 Travel index

The traveling index includes two sub methods: betweenness centrality and church index. The betweenness algorithm is focusing on road network system. It measures the connectivity between each individual road. The church index is focusing on buildings and working with road network as well. It measures a kind of crowdedness index which is affected by people's churching behavior.

5.1.2.1 Betweenness centrality

In a road network system, the whole network can be divided into various vertices and nodes and edges. The betweenness index is focusing on edges and nodes. Before the calculation process, a street network should be transferred to a connectivity map. Figure 5.2 shows an example of transformation process from street network map to a connectivity graph.

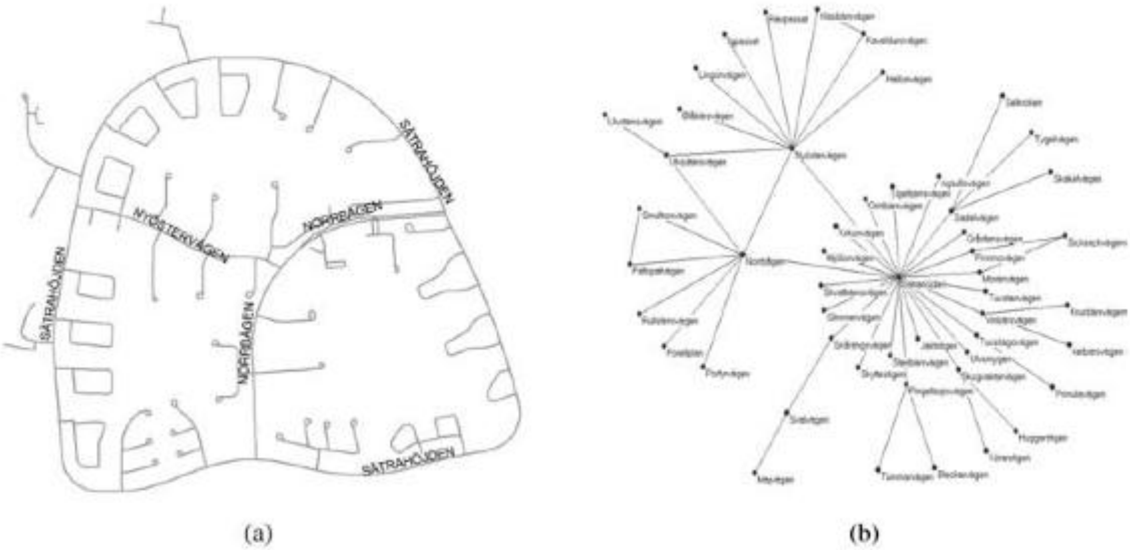


Figure 5.2, street map (a) and its connectivity graph (b)(Jiang and Claramunt, 2004)

The betweenness centrality measures the extent which node is located between to paths that connect pairs of nodes. The algorithm is defined by the following equation:

$$C_B(v_i) = \sum_{j=1}^{n} \sum_{k=1}^{j-1} \frac{P_{ikj}}{P_{ij}}$$

Where the $C_B(v_i)$ is the betweenness centrality which going to measured, P_{ij} is the number of shortest route from node i to j, and P_{ikj} is the number of shortest route from node i to k passing through node k. In spite of those indexes mentioned above, the betweenness centrality is working on road line level. The output value of betweenness centrality is for an individual road in a network.

5.1.2.2 Church index

As mentioned before, during the old days, the most common way people meet each other is by meeting at public places regulated. Church is one of the representative public locations which people are visiting periodically. When people are going to visit a church, some private residences might be passed by. A so called church index method is used to calculate an index which affected by locations of church and one's residence. This method is working with urban road networks.

Figure 5.3 shows the working theory of the social activity index. On the shortest route from one house to a church; numbers of residences are passed by. From the blue point house to the red point church: the shortest route is shown in the map; R1, R2, R3 are the residences which are passed by; the shortest length is divided by these residences to L1, L2, L3 and L4. Assume building M_i will pass through residence R_i to a Church. Total numbers of people lives in buildings M_i is N_i. The social activity index of the residence can be calculated by following equation:

$$CH_{index} = \frac{\sum N_i}{100}$$

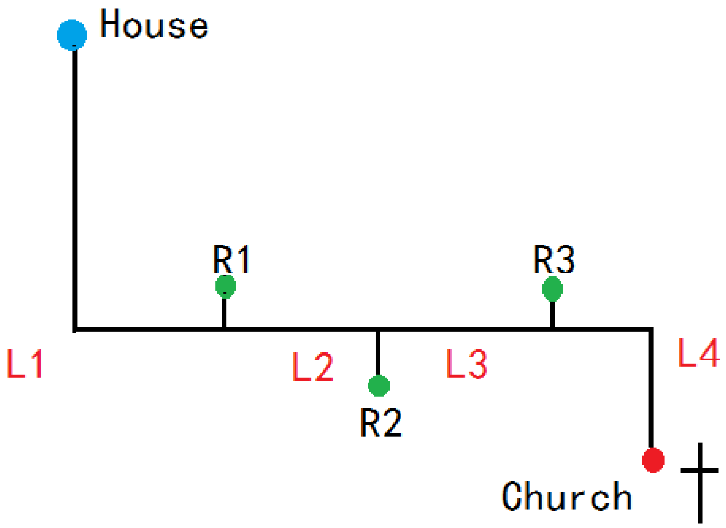


Figure 5.3, Church index

5.2 Case study II

In this case study, the three population crowdedness index methods mentioned above are tested in the study area. Results of each approach are presented separately as well. The evaluation of index methods for the data is finalized at the end of this section.

5.2.1 Method

Social interaction index

As presented in the theory section, the social activity index equation is working on road network system. Thus, before the calculation process, a street network should be built. Figure 5.4 shows the built up network and related buildings.

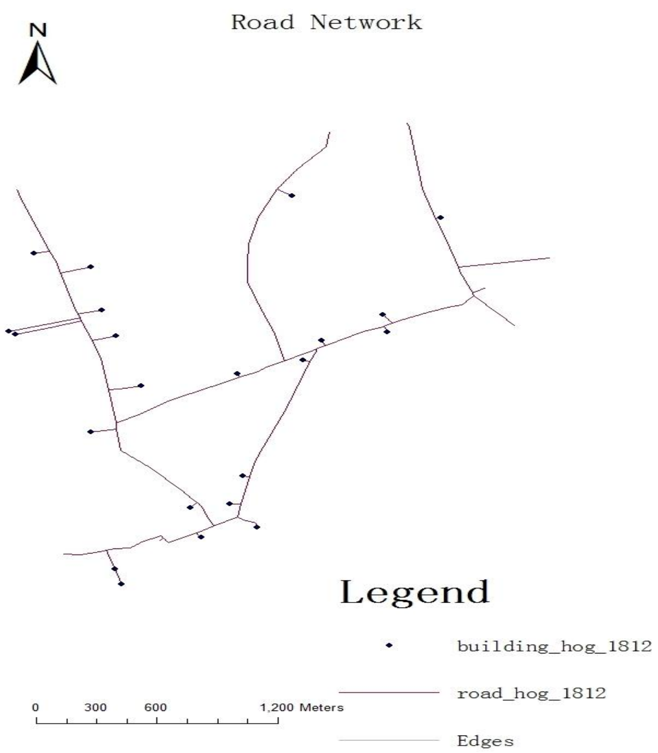


Figure 5.4, Road network and related buildings

Distances along the road network from one building to all the other buildings are stored in a distance matrix. However, there is no accurate data to show detailed population information of each specific building. The study area is divided in to 20 property units, the population data only stay on property unit level. To solve this problem, a so-called "family-staying" method is used: in each property unit, people with same family name are regarded as living in same building. Then another population matrix was built up as well as carrying population data of each building. Both matrixes are set as input data for the social activity index equation, the calculation is operated by Matlab and the output index is stored in a table as well. In ArcGIS, attribute tables of each building are created to store the corresponding social interaction index value. The final output is classified into 5 classes for visibility enhancement and exported as map.

Betweenness centrality method

To performance the betweenness centrality algorithm, a connectivity map of the network should be created. Before that, roads in the network are named by number ID so that they could be identified in the connectivity graph. Figure 5.5 shows the road network with encoded names. The connectivity map is created by open source software package Pajek (Batagelj and Mrvar, 1997). After the creation process of connectivity graph, distance values of each connection are given and stored in a matrix. The distance value of connection in the connectivity graph is given based on corresponding road length. The distance value has inverse relative with roads length. Based on the connective graph, the equation can be run. Figure 5.6 shows the connective graph of existed road network based on edges as nodes. The distance value of each connection is shown as well. It should be noticed that in the figure 5.6, nodes are equal the corresponding road edges in figure 5.5. The arcs in figure 5.6 are representing the connective relationships of each node.

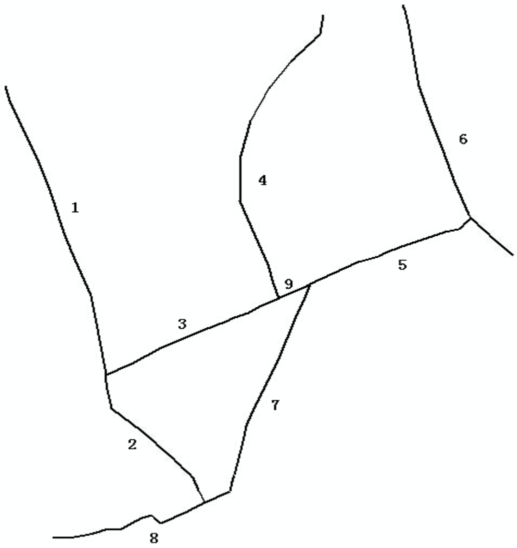


Figure 5.5, road network with numerical names

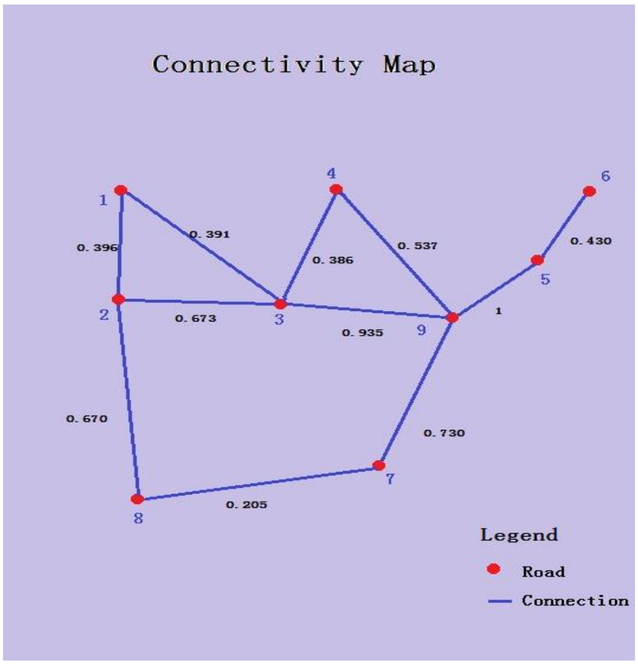


Figure 5.6, connectivity graph with distance value

It should be noted that in this case study, some nodes might not have any betweenness centrality values since they are vertexes nodes in a connective graph. After the calculation process, the centrality values of each road are stored in attribute table as well. For the visual exhibition, those index values are reloaded to ArcGIS.

Church index method

To perform the church index equation, a road network of the study area is used. Since there is no church information in the study area, one building in the road network is selected randomly as a church. The figure 5.7 shows the road network and selected church location.

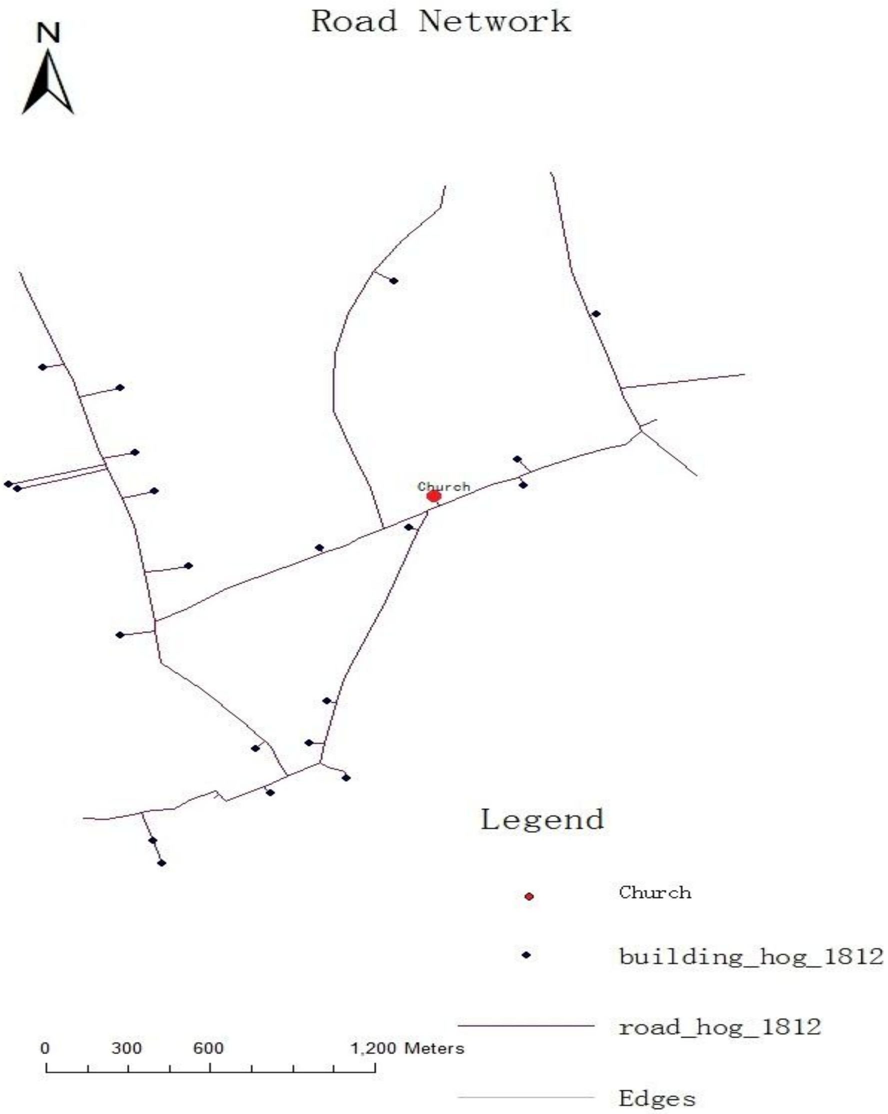


Figure 5.7, road network and related building, church

After all the preparations, the shortest routes from each building to the created church location are found out by ArcGIS. These routes are divided into arcs by passed buildings. By using these selected shortest routes, the total number of buildings and people lived in can be counted and used for the church index equation. The calculation process is operated by MatLab, and the output index value is stored in attribute table. Same as others index methods, the output index results are presented by ArcGIS for visual presentation as well.

5.2.2Results

Social interaction index

Figure 5.8 shows the Social activity index of study area. The result map is classified into 5 classes. Gray color shows the smallest index value and red color shows the largest index value; yellow, green and blue are in between the extreme values. It should be noted that the index is not only decided by the spatial location of the road network but also the number of persons living in. Table 5.1 shows the numbers of person living in and index value of each input building.

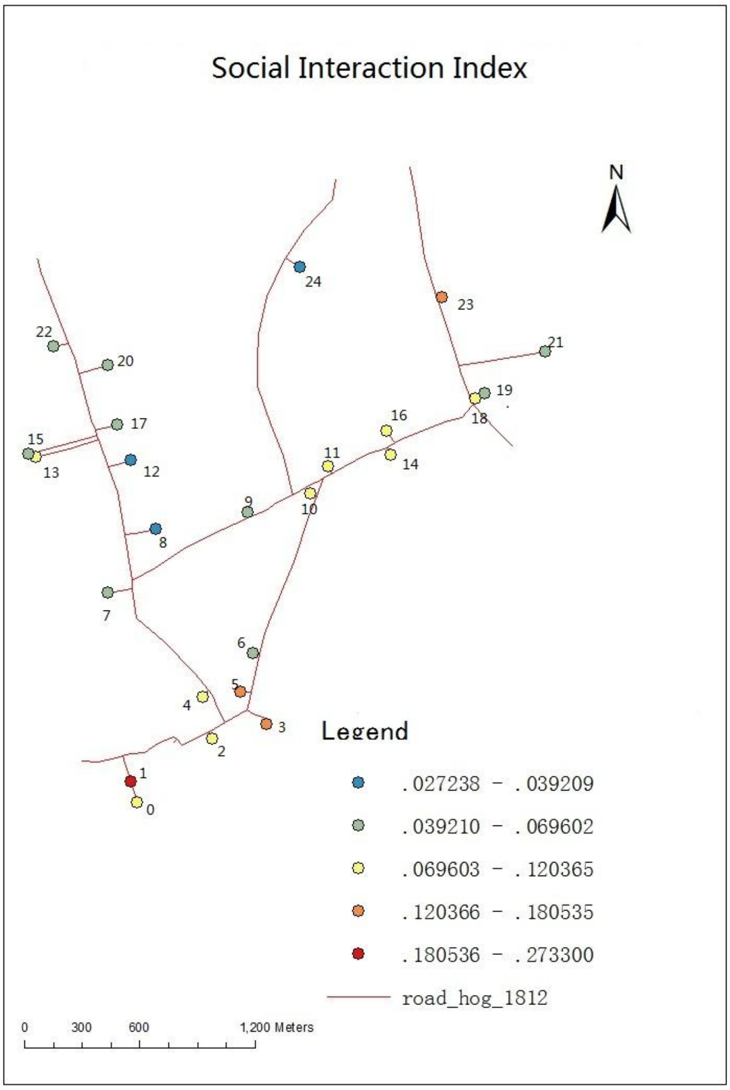


Figure 5.8, "Social activity index" of the study area

Table 5.1 shows the social activity index of each building and also numbers of persons living in.

Table 5.1, Social activity index of study area

ID	Social Index	Pop No.
0	60.021613	3
1	60.021330	6
2	80.034644	4
3	120.018763	6
4	100.014311	5
5	120.034353	6
6	20.030335	1
7	20.034050	1
8	20.007238	1
9	40.008941	2
10	40.041527	2
11	60.038634	3
12	20.007915	1
13	60.060365	3
14	40.053837	2
15	60.000000	2
16	40.053117	2
17	40.029602	2
18	60.019844	3
19	60.007879	3
20	20.038244	1
21	20.034249	1
22	40.021991	2
23	20.160535	1
24	20.019209	1

Betweenness centrality method

Table 5.2 and figure5.9 show the results of betweenness centrality method in the study area. The centrality value is describing betweenness centrality value of each road. For the road 1, 4, 6 and 8, the betweenness values are 0. That is because these roads are vertex roads of the network. They are not playing a bridge role in the street network. Road 9 is connected with 4 other roads (3, 4, 5, and 7) and the road itself has very short length which means high distance value, so that it has the largest value of betweenness 36. The other roads show reasonable results as well.

Table 5.2, Betweenness centrality

Road ID	Betweenness centrality
1	0
2	6
3	26
4	0
5	14
6	0
7	7
8	0
9	36

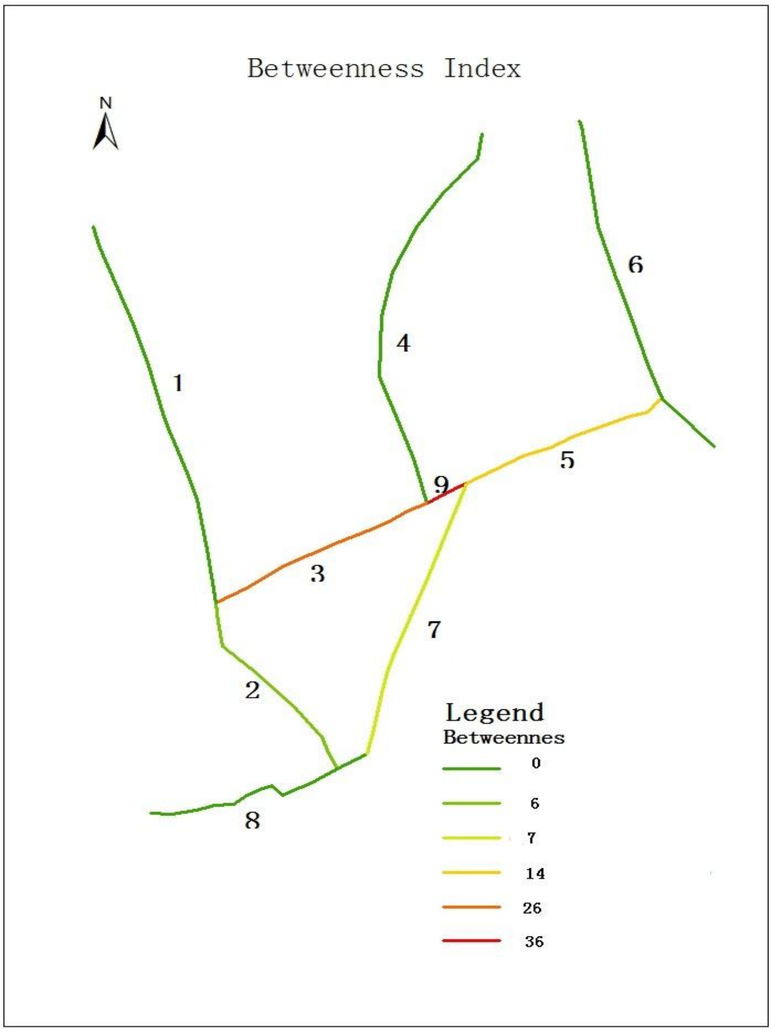


Figure 5.9, betweenness centrality of study area

Church index method

Figure 5.10 shows the church index of the study area. The result is shown in colors from green to red: the lowest index value with color dark green and the highest value with dark red. The selected church point is point 11 which is lying on the middle of the network system. It could be seen clearly that buildings closer to the church point have larger index value compare to buildings have longer distances away. The church point itself has the largest social activity index value. However, Point 7 is an exception: comparing with point 8 and 12, point 7 has shorter distance value to the church point but with much lower index value. That is because no other points pass through point 7 to the church in this road network, so that the S-index value is only depended on population value of itself. The similar reason and result is shown on building point 24 as well.

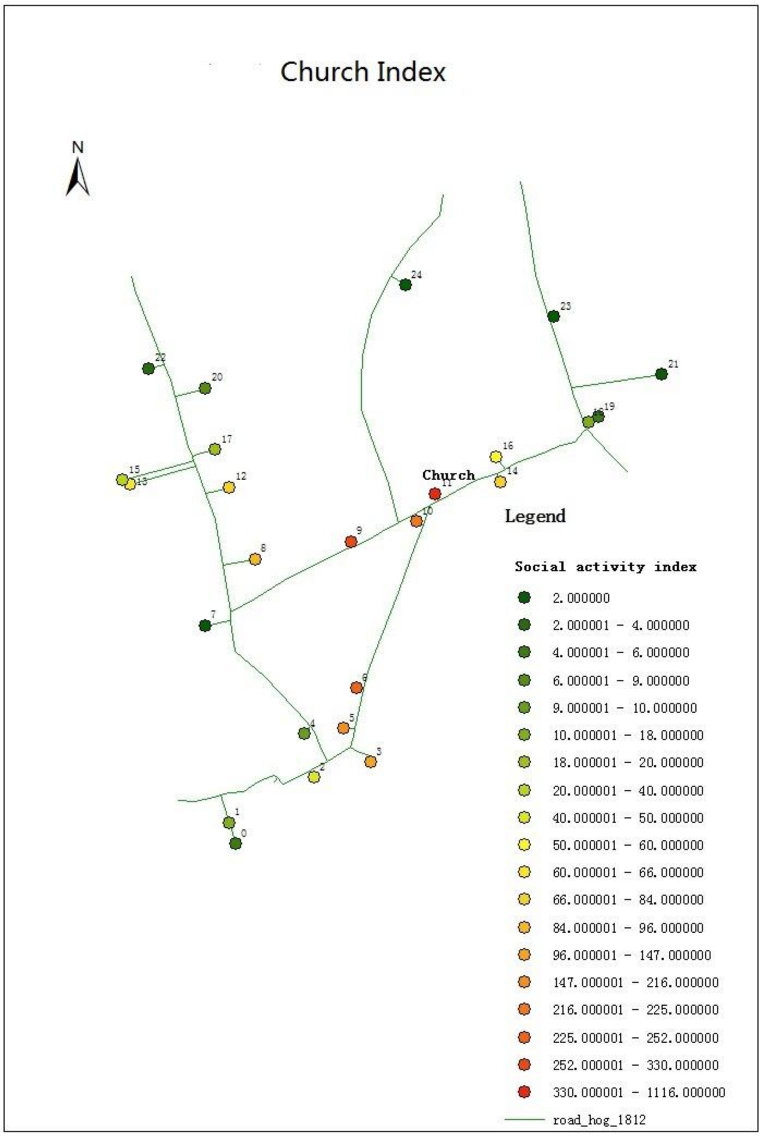


Figure 5.10, Church index of the study area

5.2.3 Evaluation

In this section, all the results of three population index measurement methods are evaluated. The evaluations are mainly based on two aspects: visually and quantitatively. The visual evaluation focuses on maps layout. The quantitatively evaluation on the other hand focus on the output index values, and analyzes the reasonability statistically. For the social interaction index, areas with more buildings exist and larger population density should be able to have larger crowdedness index values. For the betweenness centrality, road segments with more connections with other road segments should have larger centrality value. Betweenness centrality of vertex road segments should be zero. For church index, buildings with more people are passed through to the church point should have larger index values. This index value has no correlations with the distance from buildings to church point.

Social interaction index

From the social interaction index output, that the index value is depended on both populations value and distance. The building with ID number 24 has very large distance values to all other buildings. The population value of this building is small as well (population value: 1). This building point has the lowest crowdedness index value (showing on the map with color blue). The building with ID number 1 is very close to building 0 and this building also has the highest population value (population value: 6). The crowdedness index value of this building is the largest (showing on the map with red color). For buildings 15 and 13, although the Euclidian distance of these two buildings is very small, but they have very large road network distance. The index values of these two buildings are small. All other buildings points are showing reasonable visual results as well.

In order to test the equation quantitatively, a similar evaluation test with building crowdedness index equation is used. In the evaluation test, one random point is selected as the original point in the exist network. After setting up the test condition, then start to search the closest building *i* by increasing the selection distance *D* from 0 to infinite. After the closest building *i* being founded, this point and its population value are imported to the social interaction crowdedness equation. After the calculation process, an output $S_1$is created. Keep the first index value and repeat the selection process to calculate the second index value S_2. Then repeat all the steps till all points *i* are tested. Figure 5.11 shows the evaluation test. The point inside the red circle is the random original point *O*; all other points are the test variable points ordered by distance along the network. Table 5.3 shows the output index value of this evaluation test.

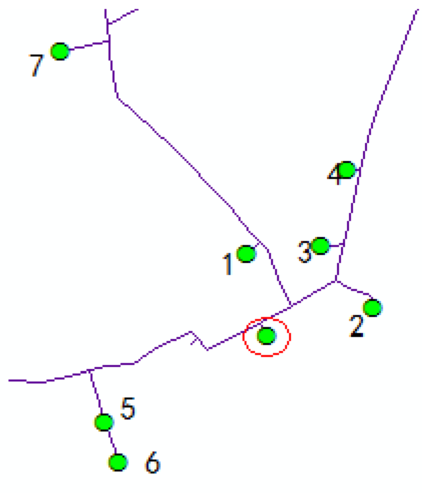


Figure 5.11, evaluation test of social interaction index

Table 5.3, evaluation test of social activity index

Index	Pop	ID
0.013298221	5	1
0.028322651	6	2
0.042398718	6	3
0.044064012	1	4
0.048356442	3	5
0.050842991	2	6
0.051717276	1	7

After the evaluation test, a curve could be created by those output index value. Figure 5.12 shows the evaluation test result of the study area. The X axis is the number of points tested in the evaluation; Y value is the output S_index values, th e number of each is the population value of individual building.

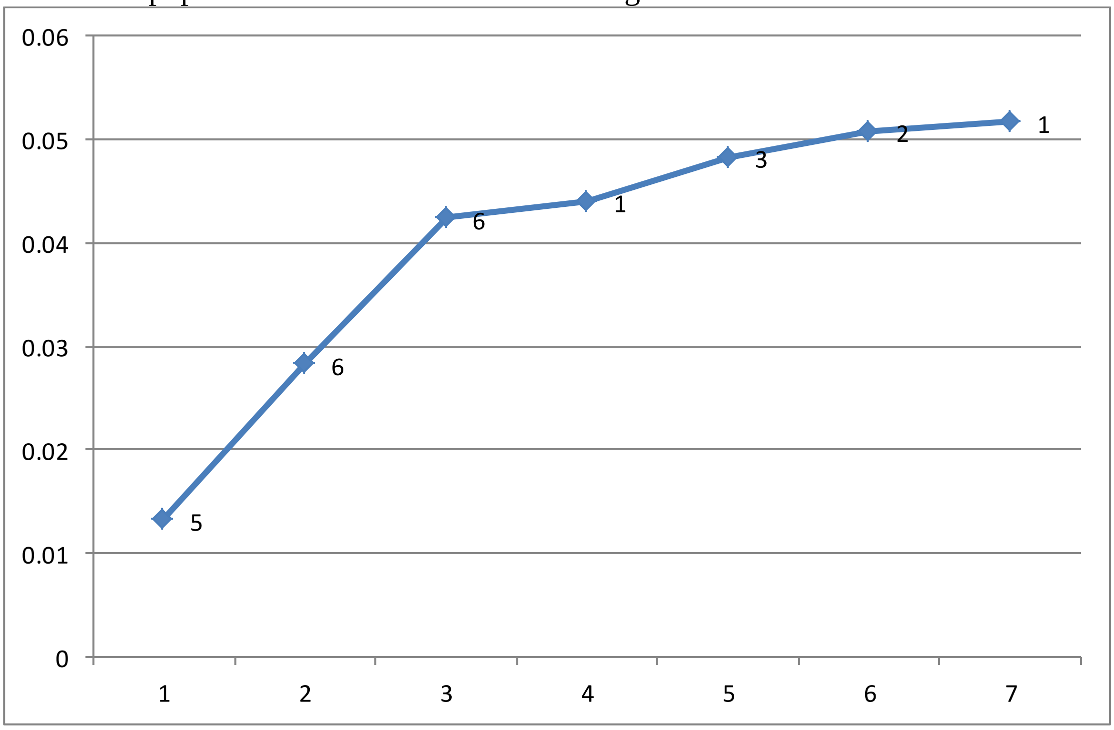


Figure 5.12, evaluation test result curve of study area

From the figure 5.12, the index value is increased by following the increase of numbers of buildings are included. The slop is becoming sharper when population value is large and becoming gentle when population value is small. The same evaluation tests are tested 3 more times by selecting the original point randomly. All the tests show similar results.

Betweenness centrality method

From the result table 5.2 and figure 5.9, betweenness centrality value of each road segment has been calculated. The shortest and most connected road segment No. 9 has the largest betweenness value. Since there are some edges (e.g. No. 1, 8, 6) which are not in the between of other roads. So that betweenness centrality value of those road segments are zero. The result of this case study is reasonable.

Church Index method

The church index is depending on both building's location of a road network and population values. From the case study results, buildings lying on a route with many other buildings pass through (more people pass through this building point to church) are having large index values. The church index value is not depending on the distance between original buildings to the church.

To evaluate this result, an evaluation test is done. Figure 5.13 shows the evaluation test condition: a simple route from building point 6 to church point. All blue points with ID 1 to 6 are buildings lying on one route to the church in a road network system. Population data is stored as attribute value of each building point. Table 5.4 shows the population attribute value of each building point. At the beginning, only point 1 and church are used to test the church index of both points. Then the index values are outputted and stored as the first result. After this step, adding point 3 to the test as well and repeat the first step. The new test result is stored as second result. Repeat the last step again till all the six buildings points are tested in the evaluations. Table 5.5and figure 5.14 show the index changing condition of the church point. The table 5.6 shows the index value of each church points when all the six buildings are added in the evaluations.

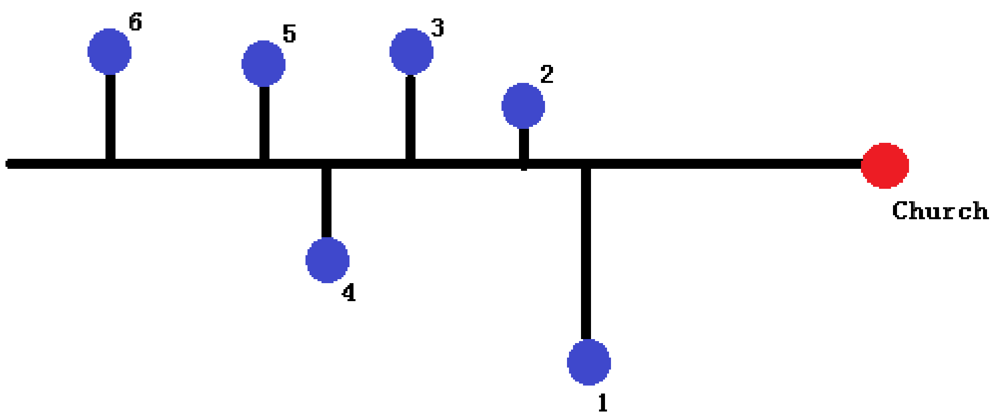


Figure 5.13, simulation test of case 1

Table 5.4, building ID and population

Building ID	Population
1	10
2	2
3	5
4	3
5	4
6	6

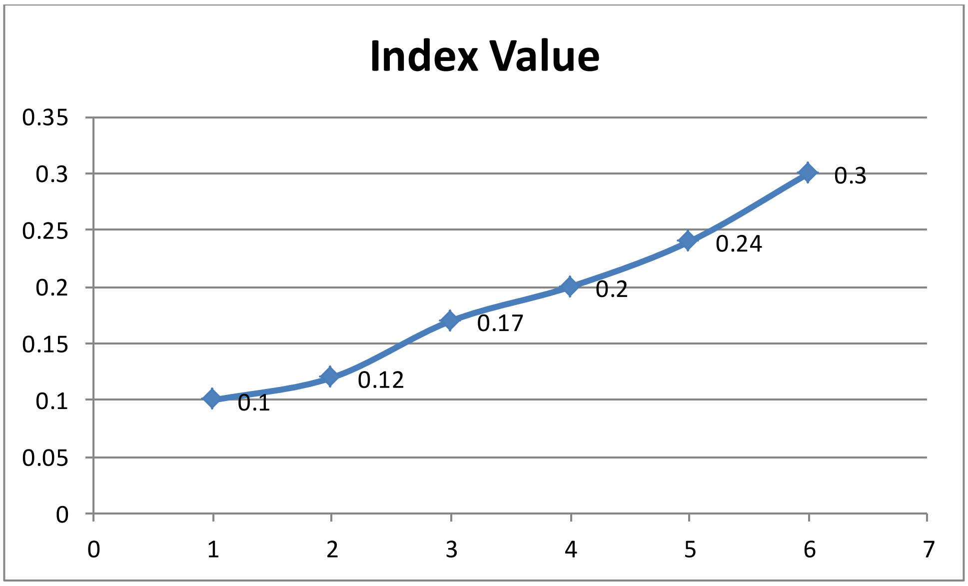


Figure 5.14, test result-index change of church point

Table 5.5, Index changing of church point

Added buildings	Index Value of Church Point
1	0,1
1,2	0,12
1,2,3	0,17
1,2,3,4	0,20
1,2,3,4,5	0,24
1,2,3,4,5,6	0,30

Table 5.6, Index value of blue buildings when all points being added

Building ID	Index Value
1	0.3
2	0.2
3	0.18
4	0.13
5	0.10
6	0.06

From the table 5.5 and figure 5.14, the index value of the church point has positive relationship with number of buildings lying on the route. The more buildings on the route, the larger church index value the church point will have. After the simple simulation test, a conclusion could be made: the social activity index is depending on both building location in shortest route and its population value.

6. Crowdedness index for property unit level

In this section, two main methods will be introduced. The first algorithm introduces a method to measure population crowdedness index of each property unit. The index value is only measured based on each unit individually. The second algorithm introduces a method to measure crowdedness index based on a buffer zone selection.

The index value of each property unit is depended on both the property unit itself and the surrounding units.

6.1 Crowdedness index on individual property unit

Because of the low accuracy of historical population data, population data is only on property unit level. Populations of each individual building are unknown. There are several different combinations between buildings and populations in one property unit. The figure 6.1 shows the property units borders with road networks and buildings. The whole study area Hög is divided into 20 property units, populations and buildings of each unit are known. Table 6.1 shows detailed information of property units, buildings, and populations. It could be seen that in some property units, there are no road networks crossing through or no buildings existing. During the following part of this section, a new method will be presented to measure property crowdedness index of each property unit with all the possible combinations between populations and buildings. This method will be tested as well.

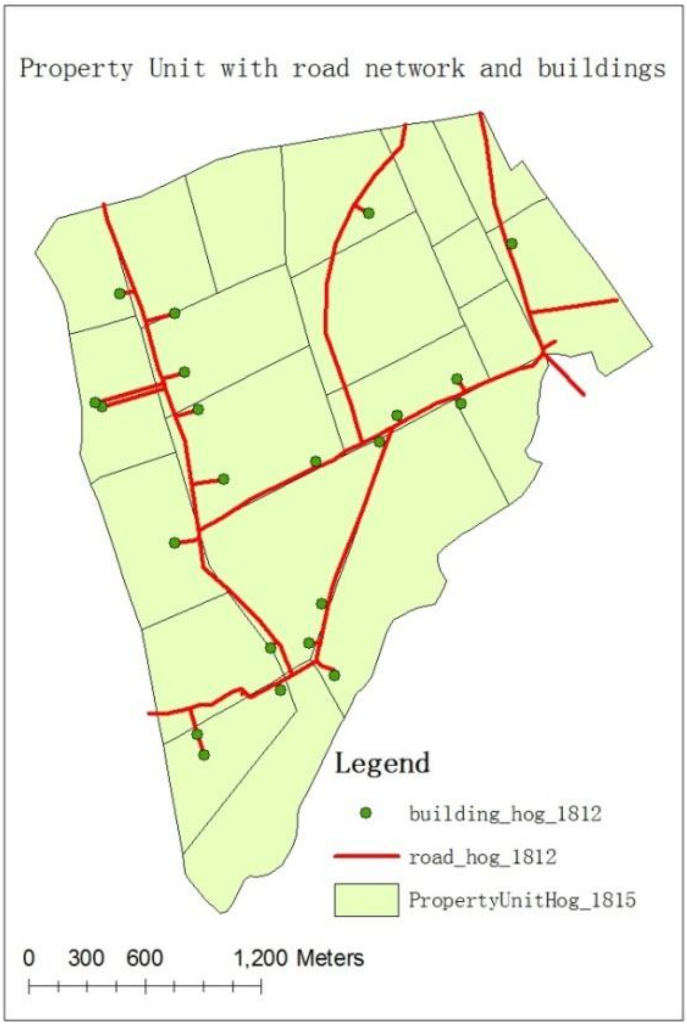


Figure 6.1, Property unit with road network and buildings

Table 6.1, Property units list and their corresponding buildings and population

Property unit	Buildings	Population
1	2	6
2	3	9
3	3	6
4	0	3
5	1	9
6	3	4
7	1	5
8	1	11
9	0	3
10	2	6
11	4	7
12	1	2
13	1	6
14	1	2
15	1	6
16	0	2
17	0	3
18	1	5
19	0	1
20	0	1

6.1.1 Method

The case study can be divided into two parts: firstly list all the possible combinations between buildings and population in each property unit; and then calculate the crowdedness index of each property unit (*P-index*) by created equation. Figure 6.2 shows the flow chart of this simulation test.

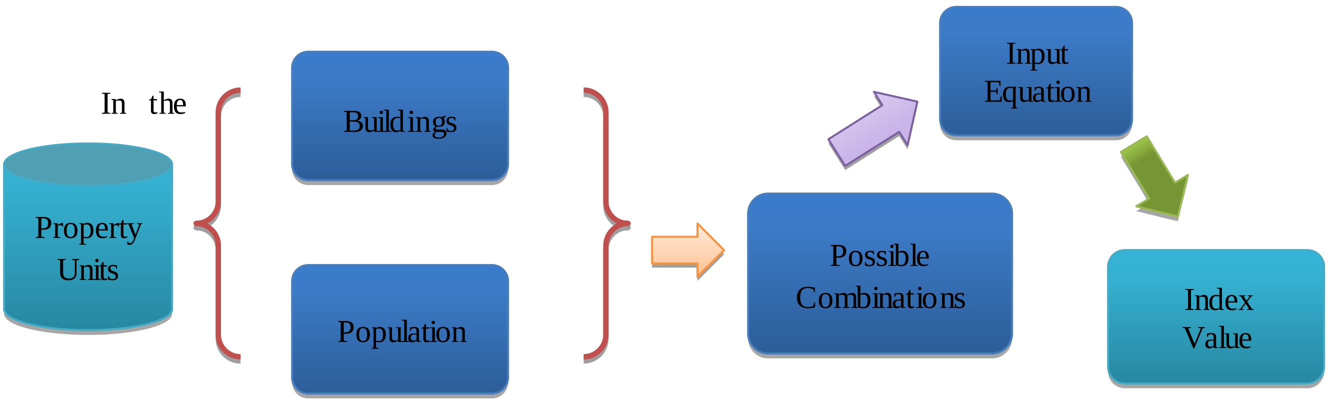


Figure 6.2, workflow of simulation experiment

In the case study, all people living in a property are allocated to the corresponding buildings randomly and each possible situation should be listed. This can be done in Matlab. Figure 6.3 shows the situation of property unit 11 as an example. The number of people living in this unit is 7; building number is 4 (named as building 1, 2, 3, 4);

distance along the network between 1 and 2 is 847 meters, 1 and 3 is 666 meters, 2 and 3 is 686 meters, 1 and 4 is 730 meters, and finally 3and 4 is 43 meters.

The Euclidean distance between 1 and 2 is 609 meters, 2 and 3 is 382 meters, 1 and 4 is 546 meters, 3 and 4 is 52 meters. Table 6.2 shows all the possible combinations in this property unit.

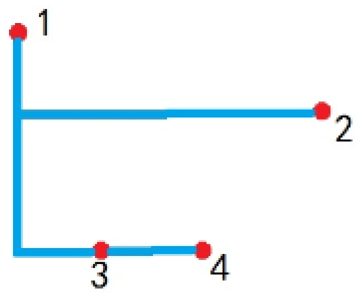


Figure 6.3, an example of unit 11

Table6.2, possible combination between buildings and population of the example unit

Building1	Building2	Building3	Building4
1	1	1	4
1	1	2	3
1	1	3	2
1	1	4	1
1	2	1	3
1	2	2	2
1	2	3	1
1	3	1	2
1	4	1	1
2	1	1	3
2	1	2	2
2	1	3	1
2	2	1	2
2	2	2	1
2	3	1	1
3	1	2	1
3	2	1	1
4	1	1	1

After listing all the possible combinations, the *P-Index* could be calculated by the following equation:

$$P_{Index} = \frac{\sum w_i x_i}{\sum W_i}$$

Where the w_i is population of each building, x_i is crowdedness index of building i, $\sum W_i$is number of populations of the target property unit. The $x_i$is calculated based on the social interaction index but without buffer distance selection. The distances between two buildings are calculated in both road network distance and Euclidian distance. In some property units, there is only one building exist. The *P-Index* value of this property unit is equal to social interaction index of this building.

6.1.2 Results

After all the possible combinations have been calculated, the *P-index* of each property can be calculated as well. Table 6.3 shows the *P-index* of each property unit in the study area. The *p-index1* is the property unit index calculated by road network distance, the *P-index 2* is the property unit index calculated by Euclidean distances. It should be noticed that since there are many combinations of some property units, the P-Index values in this table are mean values of all combinations.

Table 6.3, P-Index of each property unit

Property unit	Buildings	Population	P-Index1	P-Index2
1	2	6	0.039	0.04
2	3	9	0.089	0.22
3	3	6	0.054	0.057
4	0	3	-	-
5	1	9	0.018	0.018
6	3	4	0.070	0.12
7	1	5	0.010	0.01
8	1	11	0.022	0.022
9	0	3	-	-
10	2	6	0.076	0.084
11	4	7	0.065	0.135
12	1	2	0.040	0.04
13	1	6	0.012	0.012
14	1	2	0.040	0.04
15	1	6	0.012	0.12
16	0	2	-	-
17	0	3	-	-
18	1	5	0.010	0.01
19	0	1	-	-
20	0	1	-	-

For property unit 4, 9, 16, 17, 19 and 20, due to the data problems, there are no building data for those property units so that no values are given to them. From those existed *P-index*, it is clear to see that property units with same population have larger *P-Index* if there were more buildings inside (e.g. from property 1 and 2). Property units with the same building numbers have larger *P-Index* if the total population of this property unit was larger (e.g. from property unit 2 and 3). Property units with only one building normally have very high *P-Index* value compared with others. And these situations happened in both road network distances case and Euclidean distances case. Property units with only one building have the same *P-index* value in both cases. *P indexes* of properties units with more than one building are different in road network distance and Euclidean distances case. *P-index* in the same property unit of Euclidean distances case has larger index value than road network distance case.

6.1.3 Evaluation

The main propose of this case study is to test how the *P-Index* of a property might be changed when population data is not accurate enough. In order to analyze the stability of the *P-Index* value, the comparative analysis between each possible combination should be done. In this case, all the test values of possible combinations should be calculated and then the stability of them is analyzed. If the analysis result is stable enough, then the equation could be seen suitable for this case study.

Test Results

The following table 6.4 and figure 6.4 show the test results using property unit 3 as an example. In the table 6.4, result 1 is the test result calculated by road network distance, the result 2 is the test result calculated by Euclidean distances. The mean and standard deviation (SD) are available at the bottom of the table. From the figure 6.4, it is clear to see that the test results are shocking around an interval. Similar figures are shown for all other property units as well.

Table 6.4, Test results of Property 3

Pu-3	Result 1	Result 2
Case1	0.060	0.058
Case2	0.050	0.059
Case3	0.053	0.052
Case4	0.054	0.057
Case5	0.050	0.048
Case6	0.045	0.069
Case7	0.060	0.066
Case8	0.051	0.067
Case9	0.062	0.064
Mean	0.054	0.06
SD	0.006	0.007

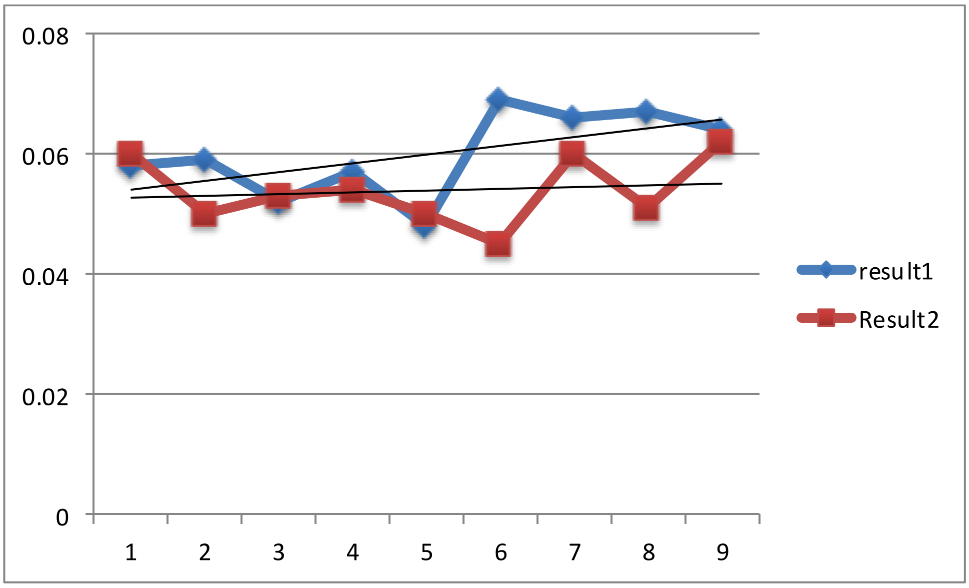


Figure 6.4, test of property unit 3

Test Analysis

The next step is to evaluate the result in quantitatively statistical ways. The standard deviation (SD) is the first factor to be evaluated. In this example evaluation test, the SD values of both results are low (0.006 and 0.007) which is low enough to be accepted. Then the next step is mainly testing the amplitude of variation of the whole result values. If the results are stable, then this method could be seen suitable for this case study. In this test process, the set of results are analyzed by a statistical method called Shapiro-wilk test (Razali et al, 2010). The Shapiro-wilk test is a kind of statistical test of normality in frequentist statistics. It uses the null hypothesis to check whether the input sample is normally distributed. In this test, the alpha value was set to 0.05 (5%), and the test results values *P* of result 1 is 0.0001497 and for result 2 is 0.0003393. Both *P* values are smaller than the set-up alpha value 0.05;hence the samples are not normally distributed. This test result means the differences of each P-index values in the output dataset are overlarge.

6.2. Building index of each property unit based on buffer zone selections

In this section, a method to calculate index value based on a selected buffer distance from the border of each property unit will be introduced. In some cases, like the figure 6.5 shown below, buildings may lay on closely to property unit's boundary. As shown in the figure, the property unit 1 is sharing its boundary with property units 3, 2, 8, 7 and 20. Buildings in property 1, 2 and 7 are distributed along property units' boundaries. In this case, the index value of these units may be affected by each other. Since the population data is still on property unit level, combinations between buildings should be considered as well. During the following part of this section, a simulation experiment are performed as well to calculate and analyze each possible property unit index of every combination between population and buildings in both Euclidian and road network distances.

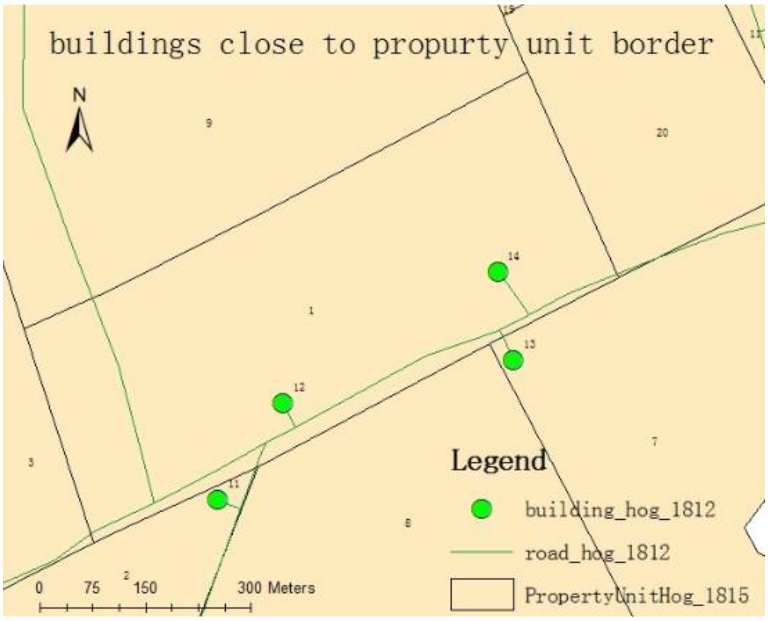


Figure 6.5, property units buildings lay on borders

6.2.1 Method

This simulation test could be divided into two parts as well. The first step is to decide suitable buffer distance from unit's boundary. The buffer areas should be able to cover and only cover those surrounding property units which are sharing border with the target one. In this case, considering the size of each unit, the buffer distance is considered as 200 meters. After the buffer selection of property unit border, during the second step, two types of property unit indexes are calculated individually: indexes of property units based on Euclidian distances between buildings and indexes of property units based on distances along road networks.

The first step is to determine a buffer distance along the road network. The crowded index can be calculated by using the surrounding buildings, which selected by the buffer distance. Same with case 6.1, all the possible combinations between buildings and populations in each property unit are listed. It should be noticed that, in this case study, combinations are done between several different properties units since the selected buildings in a buffer zone are coming from different property units. After these steps mentioned above, the property unit index is calculated by the same equation *P-Index*. Figure 6.6 shows the workflow of this case.

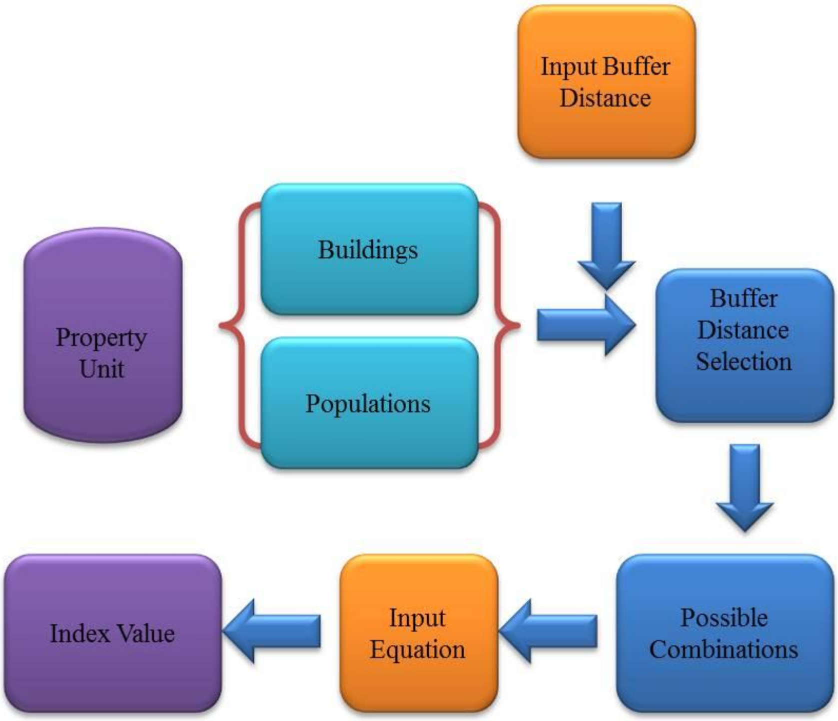


Figure 6.6, working flow of case II

The figure 6.7 below shows the property unit 1 as an example. For this property unit, there are 5 other units surrounded as sharing boundaries: property unit 20, 9, 3, 2, 8, and 7. The area inside the blue line is the buffer zone which created by 200 meters from boundary of unit 1. Buildings 12 and 14 are belonging to the property unit 1; buildings 10, 11, and 13 are belonging to units 3, 20, and 11. The Euclidian distance from building 12 to building 14, 13, 11 and 10 are 357 m, 332 m, 166 m, and 481 m; and distances along the road network are 443M, 368M, 184M, and 506M.

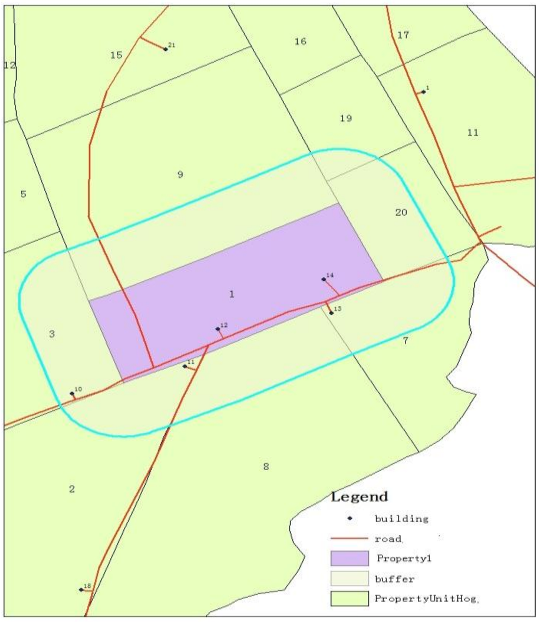


Figure 6.7, Buffer distance selection property 1 as an example

As mentioned in the case 6.1, since the precision of population data is only on property unit level, all people living in a property will be allocated to the corresponding buildings randomly and each possible situation could be listed by the *MatLab* program. In this particular case study, since the numbers of buildings and population in one property unit is small. It is possible to list all the possible combinations between buildings and population. However, if too many buildings and populations stay in one property unit or lying on surrounding units (e.g. 2000 people lives in 100 buildings in same property unit), it is lowly efficient or even impossible to list all the possible combinations by Personal Computers (PC). A so called "Monte Carlo experiment" (Anderson, 1986) could be used to solve this problem. The Monte Carlo method is a kind of computational algorithm to sample randomly in a large sample size dataset. The number of samples can be set manually.

After listing all the possible combinations, the *P-Index* is calculated by the same equation as well. In some property units, there is only one building existing after the buffer distance selection. The *P-Index* value of this property unit is equal to the social interaction index of this building. Table 6.5 and 6.6 show the buffer distance selection results of each property unit on Euclidian and road network distance. In these tables, the first row *O.Building* is building ID of the original building which the property unit index is going to be calculated; the second row *Property Unit* is the property units that the original building belongs to; the third row population is number of people living in the property unit; and the last row represents all the selected building (*S.Building*) by the buffer distance, and the property unit they belong to (*Prop*) and finally the distances between them (*Dist*). It should be noticed that due to the data problems, there are no populations or buildings in some property units.

Table 6.5, buffer selection (Euclidian distance)

O.Building	Property Unit	population	S.Building&(Prop)&[Dis]
12	1	6	14(1)[357];10(3)[481];11(2)[166];13(7)[322]
14	1	6	12(1)[357];10(3)[388];11(2)[513];13(7)[127]
11	2	9	18(2)[883];17(2)[1095];0(8)[1222];16(6)[1372] 8(14)[1173];12(1)[166];10(3)[339];15(13)[1196]
18	2	9	11(2)[883];17(2)[211];0(8)[374];16(6)[490] 8(14)[817];12(1)[1045];10(3)[732];15(13)[346]
17	2	9	18(2)[211];17(2)[214];0(8)[281];16(6)[202] 8(14)[859];12(1)[1257];10(3)[933];15(13)[380]
5	3	6	9(3)[381]; 10(3)[614]; 4(5)[201]; 8(14)[700]
9	3	6	5(3)[381]; 10(3)[485]; 4(5)[583]; 8(14)[414]
10	3	6	9(3)[485]; 5(3)[664]; 4(5)[819]; 8(14)[839]-
-	4		-
4	5	9	3(12)[309]; 2(18)[522]; 5(3)[201]-
16	6	4	20(6)[483];19(6)[513];15(13)[222]
19	6	4	20(6)[109];16(6)[483];15(13)[645]
20	6	4	19(6)[109];16(6)[493];15(13)[581]
13	7	5	14(1)[127]
0	8	11	15(13)[359];17(2)[214];18(2)[2374];11(2)[1222] 12(1)[1376];14(1)[1648];13(7)[1542]
0	9	-	-
6	10	6	6(10)[41];2(18)[572];2(12)[615];4(5)[486]; 5(3)[533]
7	10	6	7(10)[41];2(18)[854];2(12)[607];4(5)[459]; 5(3)[496]
1	11		-
3	12	2	2(18)[297]
15	13	6	17(2)[197];16(6)[222];20(6)[581];19(6)[645]
8	14	2	9(3)[414]-
21	15	6	-
-	16	-	-
-	17	3	-
2	18	5	3(12)[297]-
	19		-
-	20		-

Table 6.6, buffer selection (Road network distance)

O.Building	Property Unit	population	S.Building&(Prop)&[Dis]
12	1	6	14(1)[433];10(3)[506];11(2)[184];13(7)[368]
14	1	6	12(1)[433];10(3)[944];11(2)[616];13(7)[161]
11	2	9	18(2)[943];17(2)[1162];0(8)[1343];16(6)[1479] 8(14)[1434];12(1)[184];10(3)[594];15(13)[1579]
18	2	9	11(2)[938];17(2)[280];0(8)[462];16(6)[599] 8(14)[1466];12(1)[1053];10(3)[1462];15(13)[698]
17	2	9	11(2)[1159];18(2)[285];0(8)[288];16(6)[426] 8(14)[1293];12(1)[1309];10(3)[1683];15(13)[525]
5	3	6	9(3)[656]; 10(3)[1450]; 4(5)[446]; 8(14)[897]
9	3	6	5(3)[656]; 10(3)[1121]; 4(5)[853]; 8(14)[565]
10	3	6	9(3)[1121]; 5(3)[1450]; 4(5)[1645]; 8(14)[888]-
-	4		-
4	5	9	3(12)[588]; 2(18)[672]; 5(3)[446]-
16	6	4	20(6)[699];19(6)[804];15(13)[372]
19	6	4	20(6)[109];16(6)[804];15(13)[1087]
20	6	4	19(6)[109];16(6)[699];15(13)[988]
13	7	5	14(1)[167]
0	8	11	15(13)[511];17(2)[288];18(2)[460];11(2)[1340] 12(1)[1484];14(1)[1893];13(7)[1829]
-	9	-	-
6	10	6	6(10)[721];2(18)[937];3(12)[849];4(5)[507]; 5(3)[656]
7	10	6	7(10)[721];2(18)[940];2(12)[860];4(5)[510]; 5(3)[609]
1	11		-
3	12	2	2(18)[399]
15	13	6	17(2)[519];16(6)[373];20(6)[993];19(6)[1096]
8	14	2	9(3)[576]-
21	15	6	-
-	16	-	-
-	17	3	-
2	18	5	3(12)[400]-
	19		-
-	20		-

After all the preparations processes, the index value of each property unit in this case is calculated by the equation:

$$P_{Index} = \frac{\sum w_i x_i}{\sum W_i}$$

Where the w_i is population of each building, x_i is the crowdedness index of building i, W_i is sum of populations of all buildings. In some property units, there is only one building existed. The P-Index value of this property unit is equal to social activity index of this building.

6.2.2 Result

After the calculation process, the final result of each property unit in this case study had been calculated. Table 6.7 shows the final *P-Index* value of each property in Euclidian and road network distances case. It should be noticed that the result values of each P-Index are mean values as well.

Table 6.7, P-Index in 500 M and 1000 M cases

Propurty Unit	Index(Euclidain)	SD	Index(Road network)	SD
1	0.03963	0.00217	0.03246	0.00337
2	0.03866	0.00225	0.03088	0.00350
3	0.02592	0.00347	0.02413	0.00243
4	-	-	-	-
5	0.08583	0.00276	0.07975	0.00159
6	0.01810	0 00309	0.01576	0.00086
7	0.06181	0.00556	0.05898	0.00423
8	0.08309	0.00099	0.09167	0.00121
9	-	-	-	-
10	0.03673	0.00152	0.02584	0.00260
11	0.14	-	0.14	-
12	0.02336	-	0.01501	-
13	0.05317	0.00361	0.05021	0.00147
14	0.01534	0.00090	0.01478	0.00064
15	0.12	-	0.12	-
16	-	-	-	-
17	-	-	-	-
18	0.00573	-	0.0075	-
19	-	-	-	-
20	-	-	-	-

It could be seen that for property unit 4, 9, 16, 17, 19 and 20; due to the data problem, there are no building data for those property units so that no index values for these units as well. From those existed p-index, property units 11, 12 and 18 have no standard deviation value. That is because there is only one building in these property units, the indexes of these units are equal to crowdedness index values of these buildings. From those results with both index values and standard deviations, it is clear to see that all the standard deviations are low. The differences between indexes from Euclidian distances and road network distance are small (the largest difference is belong to property unit 10 is 0.01089), and index values of Euclidian distances are normally larger than index value from road network distance except unit 8 and 18.

6.2.3 Evaluation

The main purpose of this case study is to test how the index value of a property could be changed when population data is not accurate enough. Two sub-cases are tested in order to test the result in different type of distances. Same with every other case, all the possible property unit index value are tested. In this case, all the test values of possible combinations should be calculated and then analysis the stability of them. If the analysis result is stable enough, then the equation could be seen as suitable for this case study.

Test Results

The following table 6.8, figure 6.8 and figure 6.9 show the test results with property unit 3 as an example.

Table 6.8, index value of property unit 3 in both Euclidian distance and road network

ID	Index Value(Euclidian)	Index value(road Network)
1	0.020684	0.026071
2	0.024961	0.022045
3	0.024971	0.022519
4	0.020712	0.027482
5	0.023834	0.022463
6	0.025942	0.020779
7	0.023782	0.023601
8	0.027032	0.023404
9	0.025966	0.024067
10	0.026276	0.028879
Mean	0.024416	0.024131
SD	0.002098	0.002436

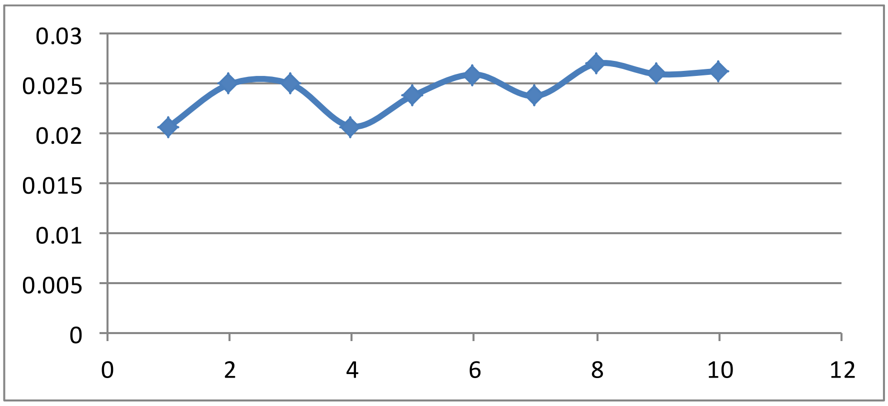


Figure 6.8, index value change of property unit 3 (Euclidian distance)

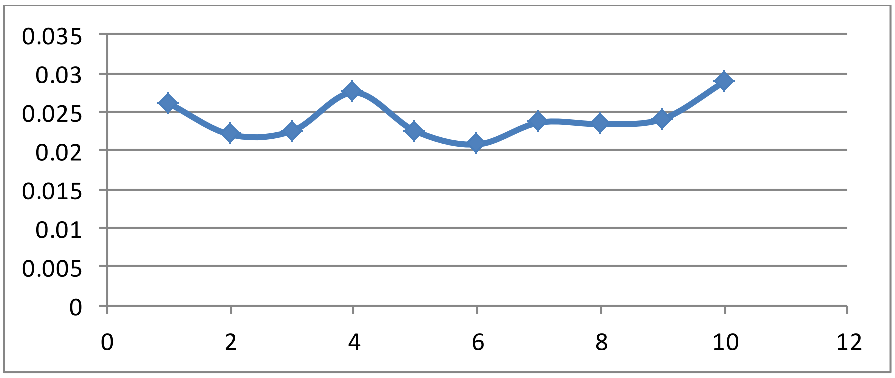


Figure 6.9, index value change of property unit 3 (Road network)

In the table, the first row is the situation for each case in the test, the second and third rows are showing test results of each corresponding case.The mean and standard deviation are available at the bottom of the table. From the figure 6.8 and figure 6.9, it is clear to see that the test results are within a certain interval. The standard deviation of each case is a very low value. Similar figures are shown for all other property units as well.

Test analysis
Same with case I, the SD value of each dataset is mainly evaluated primarily. The SD value of each test result is 0.00298 for Euclidian distance and 0.002436 for road network distance. Both of these two SD value are in very low level. The SD values are acceptable for evaluation. Shapiro-Wilk test is going to use in the next step.

After the test process, the *p* value of Euclidian distance is 3.052e-05, and *p* value of road network is 0.0003495. Both of these two *p* values are smaller than 0.05. From the result, both of these datasets are not normally distributed. The similar results show on all other property units as well. Same conclusion with the last case study could be made: the differences of each P-index values in the output dataset are overlarge.

7. Discussion

This section discusses the results of all the three main methods. Based on the case studies, each algorithm has been performed and all the study results have been evaluated. Evaluation results are reviewed and discussed in the following section.

7.1 Building crowdedness index method

As mentioned in section 4.2.3, the DBSCAN clustering method cannot measure a specific index value of buildings in the study area. However, this clustering algorithm could give some clusters of buildings. More buildings stay in a cluster have higher crowdedness index. But for this specific case study, the clustering groups are not good enough. Noises are taking high proportions of the whole input data.

The advantage of building crowdedness index method is that it could measure the index value of each individual building with high efficiency. The building crowdedness index method provided visually satisfactory results for both case study and test analysis. Most buildings are closely to each other's with higher index values. However, this method did not pass the quantitative statistical test. In the evaluation test, the closeness index values curve does not show a satisfactory result by increasing sample size.

7.2 Population crowdedness index method

Comparing with building crowdedness index method, the social interaction index method is able to measure population crowdedness index for a specific building. The index value is depending on both building's spatial distribution and its population value. Same with the building crowdedness index algorithm, the social interaction index method provided visually satisfactory results on both case study and evaluation test. For the quantitative statistical test, the social interaction index value curve shows a satisfactory result by increasing sample size as well. However, since the index value is depending on both spatial distribution and population value, all detailed information (such as population value of each building) is required.

Similar with DBSCAN clustering method, the betweenness centrality algorithm cannot provide a detailed index value of each building either. It measures betweenness centrality of each divided road segment. Buildings belonging to the same road segment have same index value. The advantage of this method is less computationally complex. It is not necessary to measure every buildings index value. However, since it could not work with road network directly, a connectivity map has to be created before the calculation process. Because road segments without playing bridge roles will have zero values, all buildings belonging to those kinds of road segments will have no index values.

The church index algorithm provides an algorithm which depends on both building's location and population values. Same with the social interaction index, this method generates visually satisfactory results on both case study and evaluation test as well. The index value is not depending on the distance between the target buildings to church point. Satisfactory evaluation result was shown for the quantitative test as well. A reasonable curve was shown in the evaluation test. Positive correlation relationship was detected.

7.3 Population crowdedness index of property units

Both property units crowdedness index are able to compute index value of the study area. The advantage of the first algorithm is its high efficiency since it only considers the buildings and populations inside one property unit. Fewer factors are considered and less calculation is used by this algorithm. In the evaluation test, all the standard deviation vales of the output property unit indexes are in very low level. However, the test result shows all the possible property unit indexes are not accepted as normal distributed. The differences between each possible population crowdedness value on each property unit are overlarge. This algorithm should be improved or developed in the further study.

Comparing with the last algorithm, the buffer distance selection algorithm considers not only population and buildings inside the target unit but also those buildings and population surrounded. This gives us more realistic crowdedness indexes. But, at the same time, since more factors are considered, more calculation processes and time will be used. The algorithm has lower efficiency. Same with the last method, although acceptable SD values are provided, but the Shapior-wilk test is failed as well. The differences between each possible population crowdedness value on each property unit are overlarge. This algorithm should be improved or developed in the further study.

8. Conclusions

This study attempts to develop and evaluate crowdedness index for historical demography. Two indexes are based on current study and techniques (e.g. DBSCAN Clustering, Betweenness centrality), the rest of them are developed in this thesis study.

Every index has been tested and evaluated by case studies. Based on the results and further analysis, the following conclusions could be drawn. Firstly, the building crowdedness index could be computed by the building crowdedness index method. But this method is not reliable enough; some errors or uncertainties may exist. Secondly, both the social interaction index and travelling index are suitable for population crowdedness index of the study area. The social interaction index is focusing on crowdedness index of people's daily life. The travelling index is suitable for measuring crowdedness index when some events happen (e.g. churching time). Thirdly, both two methods are able to measure crowdedness index values of property units. The method to measure index value by considering property unit individually gives satisfying results and efficiency. The other method gives us more realistic results since it considers not only people living in the target unit but also the surroundings. However, both of them also give results with overlarge shifting range which is needed to be improved in the further.

All the aims are achieved and all the research questions have been answered. The study results are basically reliable, even if there may be some uncertainties and limitations during the study. For further studies working on this field, more study cases are suggested to be tested so that the closeness index algorithm and population crowdedness index of property unit algorithms could be developed. Based on the conclusions, the suitable algorithms may be used in other relevant demographical studies.

Reference

Anderson, H. L., (1986), Metropolis, Monte Carlo and the MANIAC, Los Alamos Science 14: 96-108

Alter,G., Mandemakers, K., and Gutmann, M. P. (2009). Defining and Distributing Longitudinal History Data in General Way Through in Intermediate Structure. *History Social Research-HistoricheSozialforschung, 34(3), 78-114*

BatageljV., and Marvar A. (1997). Network /pajek: program for Large Network Analysis, available at http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Bengtsson, T., Dribe, M., and Scensson, P. (2012).*The Scanian Economic Demographic Database, version 2.0.*Retrieved from Lund University, Center for Economic demography.

Bin J., and Claramunt C., 2002, A Structural Approach to Model Generalization of an Urban Street Network, Geographic Information Science pp56-69

Deng C., and Wu. C., 2013, Improving Small-Area Population Estimation: An Integrated Geographic and Demographic Approach, Annals of the Association of American Geographers, 103:5, 1123-1141

Edvinsson S., 2000, The Demographic Data Base at Umeå University- A resource for historical studies. In P. H. Hall, R. McCaa, and G. Thorvaldsen (Eds), Handbook of international historical microdata for population research (pp. 231-248), Minnesota population Center.

Ekamper, P. (2010). Using Cadastral Maps in Historical Demographic Research: Some Examples from the Netherlands.*History of the Family*, 15(1), 1-12.

Ester, M., Kriegel, H.P., Sander, J., Xu, X.W. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd international conference on knowledge discovery and data mining.KDD.96.

Gaonkar, M.N. and Sawant, K. (2013).AutoEpsDBSCAN: DBSCAN with Eps automatic for large dataset. ISSN:2319-2526, vol.2.

Ghosh M., and Rao J. N. K., 1994, Small area estimation: An Appraisal, Statistical Science 9(1): 55-56

Gregory I. N., Bennett C., Gilbam V. L., and Soutball H. R., 2002, The Great Britain Historical GIS Project: From Maps to Changing Human Geography, The Cartographic Journal, Vol. 39,NO.1, 37-49

Holdsworth, D. W., (2003). Historical Geography: New Ways of Imaging and Seeing the Past. *Progress in Human Geography*, 27, 486-93.
Frank O'Brien, 1990, A Crowing index for finite population

Knowles A. K., 2005, Reports on National Historical GIS Projects, Historical Geographiy, 33:293-314

Knowles, A. K., (2008). GIS and History. In A. K. Konowles (Ed.), *Placing history: How maps spatial data and GIS are changing historical scholarship*. 1-13.

Malik W. A., Möhring J., and Piepho H. P., 2013, A Clustering-based test for non-additivity an unreplicated two-way layout. Communications in Statistics

Martin J. H., and Serow W.J., 1978, Estimating demographic characteristics using the ration correlation method, Demography 15:223-33

McCain, G., Cox, V. C., & Paulus, P. B., 1976, The Relationship between Illness Complaints and Degree of Crowding in a Prison. Environment behavior, 1976, 8, 283-290

Melissa L.M., Ru D., Jesse M. P., 2011, Comparison of Methods for Measuring Crowding and Its Effects on Length of Stay in the Emergency Department, Academic Emergency Medicine, 2011, 01, 1553-2721

Razali, N. M., and Wah,Y. B., 2011, Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnow, Lillief-ors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics 2(1): 21-33

Ruggles S., Sobek M., Alexander T., Fiteh C. A., Geeken P., Kelly H. P., King M., Ronnander C., 2008, Integrated Public Use Micro data Series: Version 4.0 [Machine readable database], Minneapolis, MN: Minnesota Population Center.

Smith S.K., and Mandell M., 1984, A comparison of population estimation methods: Housing unit virus Component II, ratio correlation, and administrative records. Journal of the American Staticistical Association 79: 282-89

Smith S. K., and Cody S., 2004, An Evaluation of population estimates in Florida : April 1, 2000, Population Research and Policy Review 23:1-24
Stokols, D., 1972, On the Distinctions between Density and Crowding: Some Implications for Further Research Vol. 79 No.3, 275-272.

Stone,G., A. Lekht, N. Burris., and Williams C., 2007. Data Collection and Communications in the public health response to disaster: Rapid Population Estimate Surveys and the Daily Dashboard in Post-Katrian New Orleans. Journal of public health management and practice 13(5):453-60

Sommer, R., 1967, *Small group ecology.* Psychological Bullelin, 1967, 67, 145-152

Sommer, R., 1971, Room Density and User Satisfaction. Environment and Behavior, 1971, 3, 412-417

Van, M. N., Scholten, H., J., and Van de Velde, R., (2009). Geospatial technonlgy and the role of location in science.*The GeoJournal Library, 96. 1-13*

Watkins J. F., 1984, The Rffect of residential structure variation on dwelling unit enumeration from aerial photographs. Photogrammetric Engineering and remotes sensing 50(11): 1599-1607

Westover, T. N., 1989, Perceived Crowding in Recreational Settings: an Environment-behavior Model. Environment and Behavior, 258-276

Appendix

Appendix 1: MatLab code to calculate building crowdedness index

This method works on x and y coordinates of objects. For a input matrix [m,n] with x and y coordinates in a buffer distance 1000 meters, the crowdedness index value of each point is last (j).

```
for j= 1:1:m
for s=1:1:n
d(j,s)=((x(j)-x(s))^2+(y(j)-y(s))^2)^(1/2);
detect(j,s)=(d(j,s)<=1000 && d(j,s)>0) ;
k=sum(detect);
end
end
for j = 1:1:m
di(j)=0;
dm(j)=0;
for s = 1:1:n
if 1000>=d(j,s)>0
l(j,s)=d(j,s);
di(j)=di(j)+d(j,s);

end
end
for j = 1:1:m
dm(j)=0;
for s = 1:1:n
if l(j,s)>0
dm(j)=dm(j)+1/l(j,s);
end
end
last(j)=dm(j)*k(j);
end
```

Appendix 2: Matlab code to calculate social interaction index

This method works with two input matrixes: building coordinate matrix and population value index. For input coordinate matrix A and a population matrix B with buffer distance 1000 M, the social interaction index of each building is index (j).

```
for j=1:1:A

for s=1:1:A
if 1000>=A(j,s)>0
detect(j,s)=A(j,s);
end
end
end

for j=1:1:B
k(j)=0;
for s=1:1:B
if detect(j,s)>0;
          k(j)=k(j)+(B(j,s)/detect(j,s));
end
end
index(j)=B(j,j)/50+k(j);
end
```

Appendix 3: Matlab code to calculate betweenness centrality

This method works on weighted and weighted directed graphs.

```
function [bc,E] = betweenness_centrality(A,varargin)
[trans check full2sparse] = get_matlab_bgl_options(varargin{:});
if full2sparse && ~issparse(A), A = sparse(A); end

options = struct('unweighted', 0, 'ec_list', 0, 'edge_weight', 'matrix');
options = merge_options(options,varargin{:});

% edge_weights is an indicator that is 1 if we are using edge_weights
% passed on the command line or 0 if we are using the matrix.
edge_weights = 0;
edge_weight_opt = 'matrix';

ifstrcmp(options.edge_weight, 'matrix')
% do nothing if we are using the matrix weights
else
edge_weights = 1;
edge_weight_opt = options.edge_weight;
end

if check
% check the values
ifoptions.unweighted ~= 1 &&edge_weights ~= 1
check_matlab_bgl(A,struct('values',1,'noneg',1));
else
check_matlab_bgl(A,struct());
end
ifedge_weights&& any(edge_weights< 0)
error('matlab_bgl:invalidParameter', ...
'the edge_weight array must be non-negative');
end
end

if trans
   A = A';
end

weight_arg = options.unweighted;
if ~weight_arg
weight_arg = edge_weight_opt;
else
weight_arg = 0;
end
ifnargout> 1
    [bc,ec] = betweenness_centrality_mex(A,weight_arg);

    [i j] = find(A);
if ~trans
temp = i;
       i = j;
       j = temp;
end

ifoptions.ec_list
       E = [j i ec];
else
       E = sparse(j,i,ec,size(A,1),size(A,1));
end

else
bc = betweenness_centrality_mex(A,weight_arg);
end
```

Appendix 4: Matlab code to sampling population values of buildings

This method works on sampling populations to buildings. For input number of population A and numbers of buildings B, the possible combination shows in result matrix.

```
%% input
m = input('How many persons: ');
n = input('How many houses: ');

N=1:m;
for i=2:n
   N=[N,1:m];
end
k=nchoosek(m*n,n);
l=nchoosek(N,n);
a=sum(l,2);
for i=1:k
if a(i)==m
h(i,:)=l(i,:);
end
end
ld=sum(h<=0,2);
h(ld>0,:) = [];
[newmat,index] = unique(h,'rows','first');
result=newmat;
```

Institutionen för naturgeografi och ekosystemvetenskap, Lunds Universitet.

Student examensarbete (Seminarieuppsatser). Uppsatserna finns tillgängliga på institutionens geobibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på LUP student papers (www.nateko.lu.se/masterthesis) och via Geobiblioteket (www.geobib.lu.se)

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (www.nateko.lu.se/masterthesis) and through the Geo-library (www.geobib.lu.se)

300 Johan Westin (2014) Remote sensing of deforestation along the trans-Amazonian highway
301 Sean Demet (2014) Modeling the evolution of wildfire: an analysis of short term wildfire events and their relationship to meteorological variables
302 Madelene Holmblad (2014). How does urban discharge affect a lake in a recreational area in central Sweden? – A comparison of metals in the sediments of three similar lakes
303 Sohidul Islam (2014) The effect of the freshwater-sea transition on short-term dissolved organic carbon bio-reactivity: the case of Baltic Sea river mouths
304 Mozafar Veysipanah (2014) Polynomial trends of vegetation phenology in Sahelian to equatorial Africa using remotely sensed time series from 1983 to 2005
305 Natalia Kelbus (2014) Is there new particle formation in the marine boundary layer of the North Sea?
306 Zhanzhang Cai (2014) Modelling methane emissions from Arctic tundra wetlands: effects of fractional wetland maps
307 Erica Perming (2014) Paddy and banana cultivation in Sri Lanka - A study analysing the farmers' constraints in agriculture with focus on Sooriyawewa D.S. division
308 Nazar Jameel Khalid (2014) Urban Heat Island in Erbil City.
309 Jessica, Ahlgren & Sophie Rudbäck (2014) The development of GIS-usage in developed and undeveloped countries during 2005-2014: Tendencies, problems and limitations
310 Jenny Ahlstrand (2014) En jämförelse av två riskkarteringar av fosforförlust från jordbruksmark – Utförda med Ekologgruppens enkla verktyg och erosionsmodellen USPED
311 William Walker (2014) Planning Green Infrastructure Using Habitat Modelling. A Case Study of the Common Toad in Lomma Municipality
312 Christiana Marie Walcher (2014) Effects of methane and coastal erosion on subsea-permafrost and emissions
313 Anette Fast (2014) Konsekvenser av stigande havsnivå för ett kustsamhälle- en fallstudie av VA systemet i Beddingestrand
314 Maja Jensen (2014) Stubbrytningens klimatpåverkan. En studie av

stubbrytningens kortsiktiga effekter på koldioxidbalansen i boreal barrskog

315 Emelie Norhagen (2014) Växters fenologiska svar på ett förändrat klimat - modellering av knoppsprickning för hägg, björk och asp i Skåne

316 Liisi Nõgu (2014) The effects of site preparation on carbon fluxes at two clear-cuts in southern Sweden

317 Julian Will (2014) Development of an automated matching algorithm to assess the quality of the OpenStreetMap road network - A case study in Göteborg, Sweden

318 Niklas Olén (2011) Water drainage from a Swedish waste treatment facility and the expected effect of climate change

319 Wösel Thoresen (2014) Burn the forest - Let it live. Identifying potential areas for controlled forest fires on Gotland using Geographic Information System

320 Jurgen van Tiggelen (2014) Assimilation of satellite data and in-situ data for the improvement of global radiation maps in the Netherlands.

321 Sam Khallaghi (2014) Posidonia Oceanica habitat mapping in shallow coastal waters along Losinj Island, Croatia using Geoeye-1 multispectral imagery.

322 Patrizia Vollmar (2014) The influence of climate and land cover on wildfire patterns in the conterminous United States

323 Marco Giljum (2014) Object-Based Classification of Vegetation at Stordalen Mire near Abisko by using High-Resolution Aerial Imagery

324 Marit Aalrust Ripel (2014) Natural hazards and farmers experience of climate change on highly populated Mount Elgon, Uganda

325 Benjamin Kayatz (2014) Modelling of nitrous oxide emissions from clover grass ley – wheat crop rotations in central eastern Germany - An application of DNDC

326 Maxime Rwaka (2014) An attempt to investigate the impact of 1994 Tutsi Genocide in Rwanda on Landscape using Remote Sensing and GIS analysis

327 Ruibin Xu (2014) Spatial analysis for the distribution of cells in tissue sections

328 Annabelle Finck (2014) Bird biodiversity in relation to forest composition in Sweden

329 Tetiana Svystun (2015) Modeling the potential impact of climate change on the distribution of Western Corn Rootworm in Europe”

330 Joel Forsmoo (2014) The European Corn Borer in Sweden: A Future Perspective Based on a Phenological Model Approach

331 Andrew Ekoka Mwambo (2015) Estimation of Cropland Ecological Footprint within Danish Climate Commissions 2050 Scenarios for Land use and Bioenergy Consumption

332 Anna Lindstein (2015) Land- atmosphere exchange of carbon dioxide in a high Arctic fen: importance of wintertime fluxes

333 Karla Susana Markley Vergara (2015) Present and near future water availability for closing yield gaps in four crops in South America

334 Klara Århem & Fredrik Fredén (2015) Land cover change and its influence on soil erosion in the Mara region, Tanzania: Using satellite remote sensing and the Revised Universal Soil Loss Equation (RUSLE) to map land degradation between 1986 and 2013

335 Fei Lu (2015) Compute a Crowdedness Index on Historical GIS Data- A Case Study of Hög Parish, Sweden, 1812-1920

336 Lina Allesson (2015) Impact of photo-chemical processing of dissolved organic carbon on the bacterial respiratory quotient in aquatic ecosystems