An Interactive Web Application For Exploring and Analysing Airbnb Listings In Singapore

Valencia TAN
School of Computing & Information Systems
Singapore Management University
Singapore
valenciatan.2020@scis.smu.edu.sg

Yashicaramya KUMARAPANDIAN
School of Computing & Information Systems
Singapore Management University
Singapore
yashicaram.2020@scis.smu.edu.sg

Derek POH
School of Economics
Singapore Management University
Singapore
derek.poh.2021@economics.smu.edu.sg

ABSTRACT

In this research paper, we will be discussing the application our team developed to allow tourists, policy makers and community members to visualise and analyse the distribution of Airbnb listings on a map. The app can be used to help better understand the impact of Airbnb on the local housing market, improve urban planning and enhance tourism development. With our app, users can select and modify their inputs to conduct both spatial point pattern analysis (SPPA) and network constrained spatial point patterns analysis (NetSPPA).

SPPA studies the distribution of the points, whether the distribution is random or clustered. This form of analysis can be very useful in the evaluation of events and we would be able to investigate whether there are any dependency relationships between different point distributions to make a comparison and conclusion. For SPPA, users can use Kernel Density Estimation, G-Function, F-Function, Cross K-Function and Cross L-Function.

NetSPPA allows us to analyse if the distribution of the spatial point events are affected by a network or whether the spatial point events occur alongside a network. For NetSPPA, users can use Network Kernel Density Estimation, K-Function and Network Cross K-Function.

1 INTRODUCTION

Airbnb, a popular online platform for short-term accommodation rentals, has gained significant attention worldwide for its disruptive impact on both the hospitality industry and local communities. Singapore, a global financial hub and popular tourist destination, has not been exempt from this phenomenon. With the rise of Airbnb listings in Singapore in recent years, there is a growing need to analyse the areas where Airbnbs are commonly found in Singapore and derive further insights on the spatial characteristics of these Airbnb listings.

The insights derived from our application can help policy makers, urban planners and tourists to better understand the implications of Airbnb on the local housing market, improve urban planning and enhance tourism development. For example, knowing that Airbnbs are normally found in clusters near amenities like convenience stores could help urban planners and companies to take advantage of this desire for convenience stores and build more convenience stores in specific areas to facilitate the growth of Airbnb rentals in that area - which would in turn boost the revenue of local Singapore Airbnb landlords.

As of now, many web applications used to visualise Airbnbs globally usually do not focus on Singapore (likely due to our relatively small geographical size). As such, having an application like ours that visualises and explores the Airbnb listings and amenities would be highly valuable.

This paper aims to analyse our application which explores the use of both spatial pattern analysis and network constrained spatial analysis on Airbnb listings in Singapore zones (mainly Kallang, Downtown Core, Outram, Rochor, Jurong West, Sembawang, Pasir Ris).

This paper reports on our development efforts in designing and implementing a geospatial application that can be utilised by urban planners, policy makers, companies and tourists to analyse Singapore Airbnb listings. The first few sections of the paper outlines our motivation and objectives. This is followed by a quick review on current research papers specific to applying geospatial techniques on Airbnb listings and past works of other geospatial web applications. Then, the approach and statistical analysis used in developing our application, would be explained before moving on to provide an overview of the application interface. Finally, the results obtained from analysing Singapore's Airbnb listing would be discussed. This paper concludes by highlighting the areas for improvement and future development work for extending our geospatial application.

2 MOTIVATION

Our team decided to develop an application that allows users to analyse Airbnb listings in Singapore and their relationships with other points of interest (MRTs, bus stops etc.) as there is a lack of such apps for Singapore Airbnb listings. After the COVID pandemic, with travelling becoming the norm and Singapore as one of the most popular cities to travel to in the world [1], Singapore

Airbnb listing information would deem to be extremely valuable. Moreover, with the high cost of living in Singapore [2], many tourists would turn to cheaper accommodation options like Airbnb rentals. In order to provide a good accommodation experience for such tourists, it is crucial to analyse the current Airbnb listings in Singapore to study the regions where Airbnb listings are most prevalent and why.

3 OBJECTIVES

In this project, we would like our Shiny web application to help users:

1) Visualise distributions and realise the benefits of spatial point patterns analysis
2) Conduct spatial point patterns analysis on Airbnbs in Singapore (Kernel density estimation, G-Function, F-Function, Cross K-Function, Cross L-Function)
3) Conduct network constrained spatial point patterns analysis on Airbnbs and other points of interests in Rochor (NetKDE and K-Function, Cross K-Function)
4) Use all the insights gathered from the analysis and models to make practical decisions

4 RELATED WORKS

4.1 Senior's Works - Spatial Pointers

Statistical Functions Seniors Employed

When we explored the theme of Spatial Point Patterns Analysis, we were inspired by one of the seniors' Shiny application that allows users to upload their data for SPPA and NetSPPA [3]. For their application, they did SPPA mainly on McDonald's outlets in Singapore and NetSPPA mainly on childcare centres in Punggol, Singapore. We referenced the SPPA section where 4 main types of statistical analysis were used:

1. ***Kernel Density Estimation:*** Visualise which area(s) have the highest intensity of points density
2. ***G-Function:*** Determine if the point events resemble clustered, dispersed or random distribution using cumulative frequency distribution of the nearest neighbour distance of a point pattern
3. ***F-Function:*** Determine if the point events resemble clustered, dispersed or random distribution by calculating the minimum distance from a random point to the original points
4. ***Cross L-Function:*** Determine if there is correlation between two types of point events by calculating the number of points in a given distance

Statistical Functions Our Team Employed

For our application, we similarly implemented the following functions:

1. ***Kernel Density Estimation:*** Visualise which location(s) in the selected zone have the highest intensity of Airbnb points density
2. **G-Function:** Determine if the Airbnbs in the selected zone resemble clustered, dispersed or random distribution using cumulative frequency distribution of the nearest neighbour distance of a point pattern
3. ***F-Function***: Determine if the Airbnbs in the selected zone resemble clustered, dispersed or random distribution by calculating the minimum distance from a random point to the original points
4. ***Cross K-Function***: Determine if there is correlation between two types of point events (eg. Airbnbs & MRTs, Airbnbs & Bus stops etc.) by calculating the number of points in a given distance
5. ***Cross L-Function***: Determine if there is correlation between two types of point events (eg. Airbnbs & MRTs, Airbnbs & Bus stops etc.) by calculating the number of points in a given distance (Standardised version of Cross K-Function)

Seniors' Function Implementations VS Our Teams Function Implementations

For the seniors' project, they did not include Cross K-Function which was understandable since Cross K-Function and Cross L-Function are extremely similar - given that the latter was derived from the former through standardisation.

However, for our project, we wanted to include more functions so that technically inclined individuals who are keen to explore the minute differences between the 2 mentioned functions are free to do so.

4.2 Senior's Works – Signal

Statistical Network Functions Seniors Employed

When developing our application, we were inspired by one of the seniors' Shiny applications which explored the use of Network-constrained Spatio-temporal Analysis for Traffic Accidents in the UK [4]. In that application, 4 main types of statistical analysis were used:

1. ***Network-Constrained Kernel Density Estimation***: Visualise which segment(s) of the road have the highest intensity of traffic accident points or casualty points along the network
2. ***Network-Constrained K-Function:*** Determine if there is spatial correlation between accident points on a linear network using geometrically corrected K-Function
3. ***Network-Constrained Cross K-Function***: Determine if there is correlation between accident points and variables selected
4. ***Network-Constrained Cross Pair Correlation Function:*** Determine if there is correlation between accident points and variables selected

Statistical Network Functions Our Team Employed

For our application, we similarly implemented the following Network-related functions:

1. ***Network-Constrained Kernel Density Estimation:*** Visualise which segment(s) of Rochor's street network have the highest intensity of Singapore Airbnb listing points or amenity points (MRT, bus stops etc.) along the network
2. ***Network-Constrained K-Function***: Determine if there is spatial correlation between Rochor Airbnb listing points on a linear network using geometrically corrected K-Function
3. ***Network-Constrained Cross K-Function:*** Determine if there is correlation between Airbnb listing points in Rochor and amenity variables selected

We focused on Rochor as Rochor has a significant number of Airbnbs since it is the zone with the 4th most number of Airbnbs. Additionally, Rochor has enough point events for the other amenity points (MRT, bus stops, etc.) which will allow us to draw statistical conclusions with the Airbnbs in Rochor. There were other zones with higher numbers of Airbnbs but we did not choose them as they only had 1 point event for the other amenity points which is not enough to draw statistical conclusions especially for Network-Constrained Cross K-Function.

<u>Seniors' Network Function Implementations VS Our Team's Network Function Implementations</u>

For the seniors' project, since they were analysing accidents, they focused on analysing factors that could affect accident rates in specific regions of the UK. In some of their network features, they had filters to control the time, environment (weather) and casualty.

For our project, since we were analysing Singapore Airbnb listing points, we focused more on studying the different types of amenity points around. This is because for our project, isolating and studying factors like time and weather would be less appropriate as Airbnb listings tend to be much more influenced by factors like amenities instead.

For the seniors' implementation, they implemented less technical adjustments like time, environment (weather) and casualty. Since we wanted our application to be well-used by technically inclined experts, we introduced more technical adjustments. For instance, for the Network-Constrained Kernel Density Estimation under our app's NetSPPA tab, besides allowing users to select different variables (Airbnbs, MRTs, bus stops), we allowed users to control more technical features such as 'Kernel Smoothing Input' and 'Method'. At the bottom of the Network graphs generated, we've also additionally included descriptions to aid the users in interpreting the graphs. This helps less technical users to comprehend the results more easily.

5 ANALYSIS METHODS

Different spatial statistical analysis methods can be employed to conduct spatial analysis of Singapore Airbnb listings. In our application, we've split the analysis methods under 4 different tabs as outlined below. The 4 tabs are:

1. ***Visualisation***: where different zones can be selected to see the Airbnbs and amenity points scattered
2. ***Tmap***: where Airbnbs and amenity points can be viewed in the whole of Singapore (users can select and deselect accordingly)
3. ***SPPA***: where first-order spatial point analysis techniques are used to study the distribution of Airbnbs in Singapore zones and second-order techniques are used to analyse the distribution of Airbnbs and spatial independence between two types of points (Airbnbs & MRTs, Airbnbs & bus stops etc.) in Singapore zones
4. ***NetSPPA***: where network-constrained statistical functions are used to view the distribution of Airbnbs and amenity points over the street network of Rochor and also analyse the distribution of between two types of point events in Rochor

5.1 Visualisation (Visualise by zone)

Under this tab, the user can select one of the 7 zones in the dropdown and visualise the spread of the following points in the zone selected: Airbnbs, tourist attractions, bus stops, hotels, malls, MRTs, 7-11 convenience stores, universities.

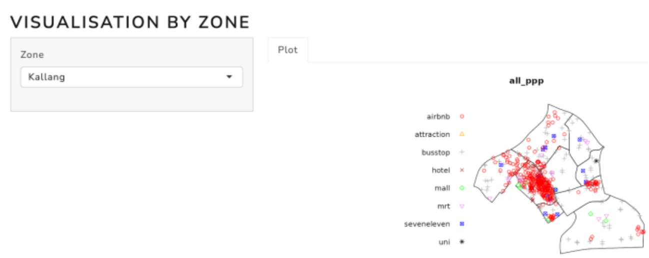


Fig 5.1.1: Visualisation Tab - Kallang zone selected

5.2 TMap (Visualise entire Singapore)

Unlike the 'Visualisation' tab in 5.1 where users can only see the points zone by zone, under this tab, users can see points in all of Singapore (across all zones). Moreover, users can easily select and deselect specific types of points as they wish.

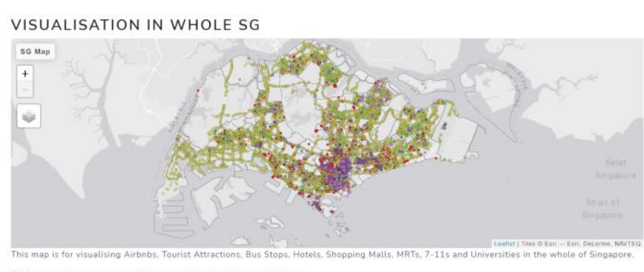


Fig 5.2.1: Tmap Tab - all points selected

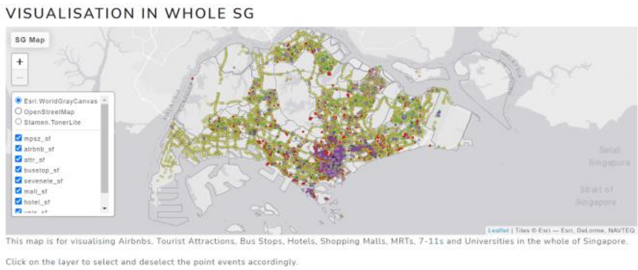


Fig 5.2.2: Tmap Tab - how to select/ deselect points

5.3 Spatial Point Pattern Analysis

In this section, first-order spatial point analysis techniques are used to study the distribution of Airbnbs. For first-order spatial point analysis techniques, we have Kernel Density Estimation.

In this section, we've also implemented second-order spatial point analysis techniques to analyse the distribution of Airbnbs and the relationships between two types of points (Airbnbs & MRTs, Airbnbs & Bus stops etc.) in Singapore zones. For second-order spatial point analysis techniques, we used G-Function, L-function, Cross K-Function and Cross L-Function.

5.3.1 Kernel Density Estimation

$$\hat{\lambda}_\tau(s) = \frac{1}{\sigma_\tau(s)} \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right)$$

Fig 5.3.1.1: Kernel Density Estimation formula

Kernel Density Estimation (KDE) is a first-order method to compute the intensity of a point distribution. In our application, we have implemented 4 different types of kernels - Gaussian, Epanechnikov, Quartic and Disc.

In KDE, calculation of intensity is based on Euclidean distance search bandwidth and does not take into consideration the presence of road network structures.

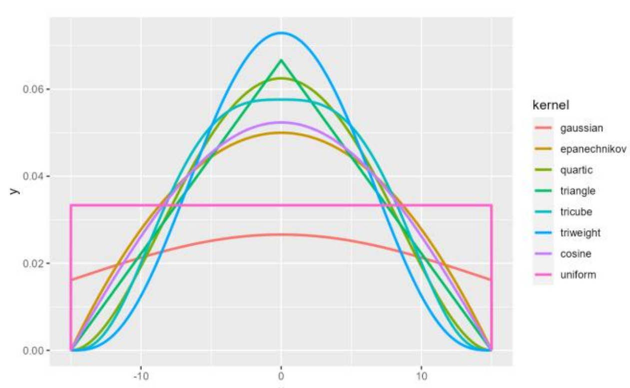


Fig 5.3.1.2: Kernel Density Estimation Graphs for the different kernels [5]

Different kernel types will affect the intensity of the point events differently. For example, referring to Fig 5.3.1.2, comparing between Gaussian and Epanechnikov, we can see that Epanechnikov has a higher peak. This means that the intensity of the point events will reduce faster within a shorter distance as compared to the Gaussian kernel type that has a flatter and shorter peak.

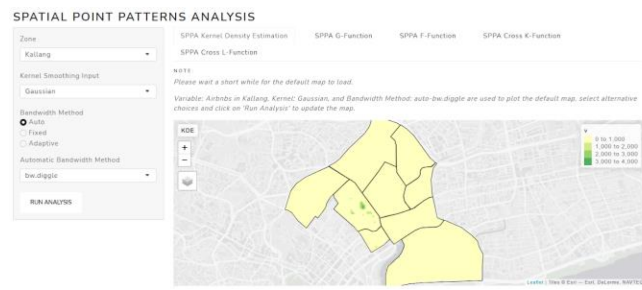


Fig 5.3.1.3: KDE plot generated for Airbnbs in Kallang (Gaussian, Auto bandwidth method - bw.diggle)

In Fig 5.3.1.3, we can see that there are regions in green. These are areas of higher intensity of Airbnbs. While KDEs are useful in spotting general areas of intensities, if we want a more specific view of higher intensity regions of Airbnb points along the streets of Kallang, we should use a Network-Constrained Kernel Density Estimation map (which we will cover under 5.4.1, which lies under the NetSPPA tab in the interface).

5.3.2 G-Function

Up till now, we only focused on first-order spatial point analysis. While first-order methods are good for determining the general spatial arrangement of Airbnb points in specific zones or in Singapore, it is important to introduce second-order spatial point analysis to unravel the distribution of Airbnbs and spatial relationships between two types of points.

$$G(r) = \frac{\#[r_{\min}(s_i) < r]}{n} = \frac{\#\,\text{point pairs where } r_{\min} \leq r}{\#\,\text{of points in study area}}$$

Fig 5.3.2.1: G-Function Formula

The G-Function is a second-order method to tell us the way the Airbnbs are spaced in a zone. G-Function is distance-based analysis that can be applied to point data to determine if there is a tendency of the locations of the data to exhibit a systematic pattern over an area as opposed to being randomly distributed [6]. The number of simulations has to be specified for the G-Function as it uses Monte Carlo to test for Complete Spatial Randomness (CSR). The G-Function will perform a number of independent simulations of n events (i.e. 999) in the study region. For each simulated point pattern, estimate G(r) and use the maximum (95th) and minimum (5th) of these functions for the simulated patterns to define an upper and lower simulation envelope. If the estimated G(r) lies above the upper envelope or below the lower envelope, the estimated G(r) is statistically significant.

5.3.3 F-Function

$$F(d) = \frac{\#[d_{\min}(p_i, s) < d]}{m} = \frac{\#\,\text{of point pairs where } r_{\min} \leq r}{\#\,\text{sample points}}$$

Fig 5.3.3.1: F-Function Formula

The F-Function is a second-order method to determine the minimum distance from each point to any event in the study area. F-Function randomly selects m points, calculates dmin(pi,s) as the minimum distance from location pi to any event in the point patterns. Finally, with the points and dmin(pi,s) values, F(d) is calculated. The F-Function can be used to let us know if the points are clustered, dispersed or at random.

The F-Function will perform a number of independent simulations of n events (i.e. 999) in the study region. For each simulated point pattern, estimate F(r) and use the maximum (95th) and minimum (5th) of these functions for the simulated patterns to define an upper and lower simulation envelope.

5.3.4 Cross K-Function

For any pair of types i and j, the multitype K-function Kj(r), also called the bivariate or cross-type K-function, is the expected number of points of type j lying within a distance r of a typical point of type i, standardised by dividing by the intensity of points of type j:

$$K_{ij}(r) = \frac{1}{\lambda_j}\mathbb{E}\left[t(u, r, \mathbf{X}^{(j)}) \mid u \in \mathbf{X}^{(i)}\right]$$

Fig 5.3.4.1: Cross K-Function Formula

for r > 0, where X(i) is the sub-process of points of type j, with intensity Ai, and t (u, r, x) is the count of r-neighbours. Dividing by A allows us to compare point patterns with different intensities.

5.3.5 Cross L-Function

The Cross L-Function is very similar to Cross K-Function as it's derived from the Cross K-Function through standardisation. The explanations in 5.3.4 for Cross K-Function are still applicable to Cross L-Function.

5.4 Network-Constrained Spatial Point Pattern Analysis

Why use Network-constrained spatial point pattern analysis? Instead of just analysing between different types of points in SPPA, we take into consideration the network structure for our analysis since the network structure can affect the distribution of Airbnb listing points and provide us with more insights.

5.4.1 Network-Constrained Kernel Density Estimation

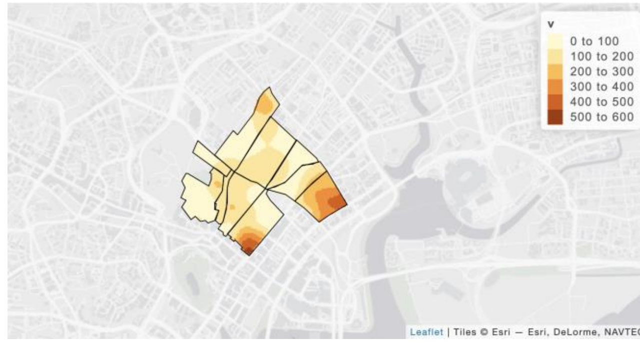


Fig 5.4.1.1: KDE of Airbnbs in Rochor

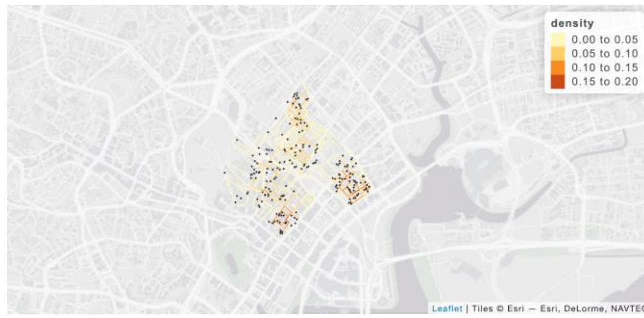


Fig 5.4.1.2: Network-constrained KDE of Airbnbs in Rochor

As seen in Fig 5.4.1.1, with the conventional KDE (from section 5.3.1) used in the SPPA tab, we're only able to see general regions with higher densities of Airbnb listing points indicated in darker orange. The KDE in Fig 5.4.1.1 and the Network-constrained KDE in Fig 5.4.1.2 both used the same technical settings - 'Quartic' option for Smoothing Kernel and 'Simple' option for Method on the UI. From Fig 5.4.1.2, we can see the road constraints and precise points of the Airbnbs along the street network of Rochor.

From Fig 5.4.1.2, we can see that there are higher densities of Airbnb points towards the outer-ends of Rochor where larger roads are seen. This could suggest that Airbnbs in Rochor tend to be near larger, main roads.

5.4.2 Network-Constrained K-Function

The Okabe-Yamada Network Constrained K-Function defines the Network Constrained K-Function, by adapting Ripley's K-Function through replacing the Euclidean distance with the shortest path distance [4]. With a point v on the network, all locations in the network that can be reached from v by a path of length shorter than or equal to a radius r, defined by the user, would be considered [4]. It is defined by the function below, with λ(L) denoting the total length of the linear network [4]:

$$K^L(r) = \frac{1}{\lambda}\mathbb{E}\left[\sum_j \frac{1\{0 < d_L(u, x_j) \leq r\}}{m(u, d_L(u, x_j))} \;\middle|\; u \in \mathbf{X}\right]$$

Fig 5.4.2.1: Network-Constrained K-Function Formula

With the Network-Constrained K-Function, we can determine if there is spatial correlation between Airbnb points on a linear network using geometrically corrected K-Function.

5.4.3 Network-Constrained Cross K-Function

What's the difference between Network Constrained K-Function and Network Constrained Cross K-Function?

Network Constrained K-Function deals with points of the same type while Network Constrained Cross K-Function processes two different sets of points. Estimation is based on measuring pairwise distances from all points of type i to all points of type j [4].

$$K_{ij}(r) = \frac{1}{\lambda_j}\mathbb{E}\left[t(u,r,\mathbf{X}^{(j)}) \mid u \in \mathbf{X}^{(i)}\right]$$

Fig 5.4.3.1: Network-Constrained Cross K-Function Formula

With the Network-Constrained Cross K-Function, we can determine if there is correlation between Airbnb listing points in Rochor and amenity variables selected.

6 DATA APPROACH & APPLICATION ARCHITECTURE

6.1 Data Collection

Spatial Point Patterns Analysis

- URA 2019 Master Plan Planning Subzone Boundary Data (.shp) - By Prof Kam in In Class Ex 9
- Airbnb Locations in Singapore (.csv) - We will get the point events via the longitude and latitude
- MRT Stations in Singapore (.shp)
- Bus Stops in Singapore (.shp)
- Tourist Attractions in Singapore (.shp)
- Shopping Malls in Singapore (.csv) - We will use this list of shopping mall names to get the point events using OneMapSG API
- 7-11 Stores in Singapore - We will extract the list of stores (name, address, postal) into xlsx, then locate their point events using OneMapSG API
- Hotel Locations in Singapore (.kml)
- Universities and Colleges in Singapore - We will extract the list of universities and colleges (name) into xlsx, then locate their point events using OneMapSG API

Network-Constrained Spatial Point Patterns Analysis

- Road Network of Singapore (Road Section Line) (.shp)
- Airbnb locations in Singapore (.csv) - We will get the point events via the longitude and latitude

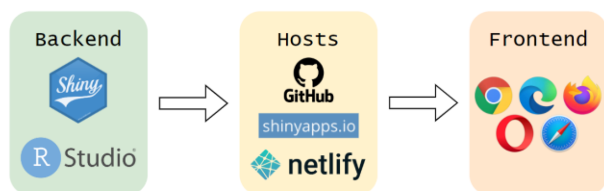
6.2 Data Cleaning

1. Transformation of CRS to SVY21 and EPSG Code 3414
2. Dropping of Z Dimension for hotel dataset
3. Making invalid geometries valid
4. Checking of missing values
5. Removal of irrelevant columns
6. Removal of irrelevant points (eg. bus stops in Malaysia)
7. Jittering to remove duplicate points by adding a small perturbation to the duplicate points so that they do not occupy the same space

6.3 Data Transformation

1. Creation of ppp objects for visualisation
2. Rescaling of ppp object from m to km for KDE map in SPPA
3. Transformation of multilinestring geometry to linestring geometry for NetSPPA
4. Extraction of Airbnbs and amenity points that are in Rochor

6.4 Application Architecture



R libraries used: shiny, bslib, shinycssloaders, readxl, httr, jsonlite, maptools, sf, sfdep, raster, spatstat, spNetwork, rgdal, sp, tmap, tidyverse

7 APPLICATION INTERFACE

Visualisation

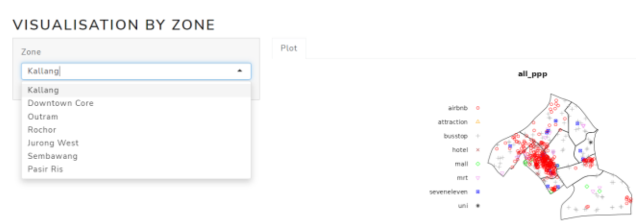


Fig 7.1: 'Visualisation' Tab

In Fig 7.1, we can see that there are 7 zones to analyse in Singapore. By selecting a zone, all the points in that zone will be displayed. Each point type has its own unique symbols. For instance, Airbnb points are red circles and attraction points are orange triangles.

Tmap

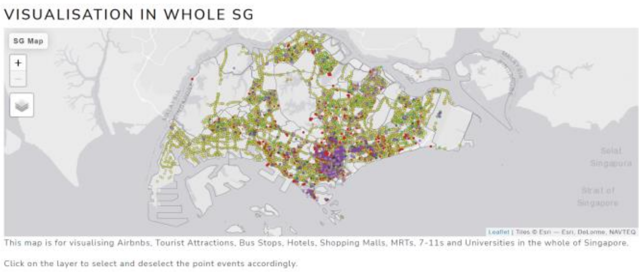


Fig 7.2: 'Tmap' Tab

In Fig 7.2, we can see different point types in the whole of Singapore. The user can easily select and deselect different point types.

SPPA

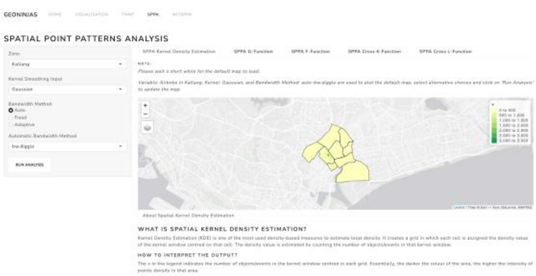


Fig 7.3: 'SPPA' Tab – KDE

As seen in Fig 7.3, this KDE plot is generated for the selected zone Kallang. The kernel smoothing input chosen is "Gaussian", bandwidth method is "Auto" and automatic bandwidth method is "bw.diggle".

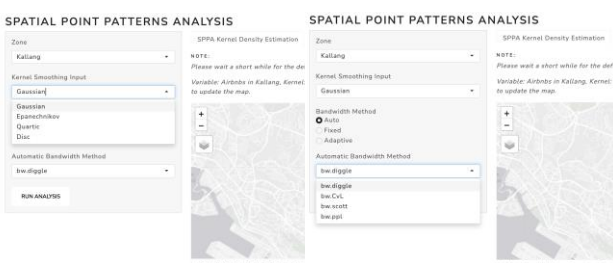


Fig 7.4: 'SPPA' Tab - KDE Options To Select

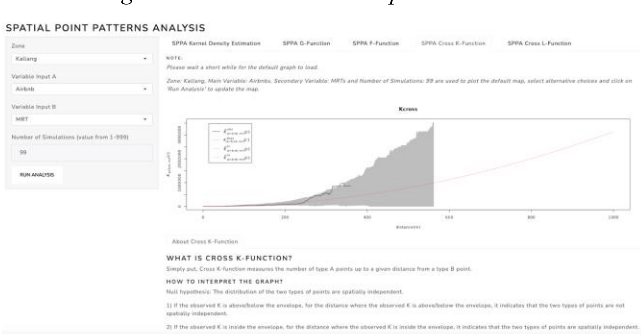


Fig 7.5: 'SPPA' Tab - SPPA Cross K-Function Layout

As seen in Fig 7.5, the Cross K-Function graph is generated when the user selects:

1. 1 Zone
2. 1 Variable Input A
3. 1 Variable Input B
4. Number of Simulations

Under the graph, a description to help users interpret the graph is provided.

NetSPPA

Please note that for the NetSPPA section, the zone being used for analysis is Rochor. The team decided to fix the analysis solely on Rochor as Rochor has a significant number of Airbnbs since it is the zone with the 4th most number of Airbnbs. Additionally, Rochor has enough point events for the other amenity points (MRT, bus stops, etc.) which will allow us to draw statistical conclusions with the Airbnbs in Rochor. There were other zones with higher numbers of Airbnbs but we did not choose them as they only had 1 point event for the other amenity points which is not enough to draw statistical conclusions especially for Network-Constrained Cross K-Function.

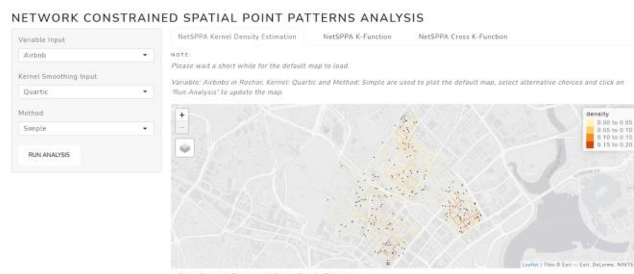


Fig 7.6: 'NetSPPA' Tab - NetSPPA KDE (Rochor)

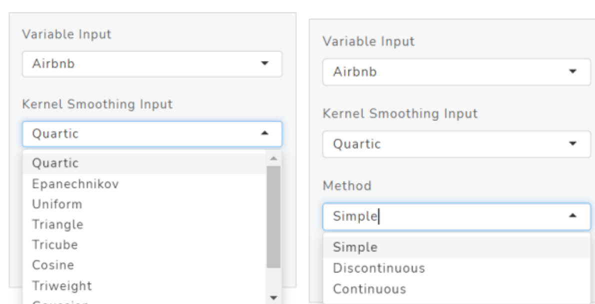


Fig 7.7: 'NetSPPA' Tab - NetKDE Options To Select

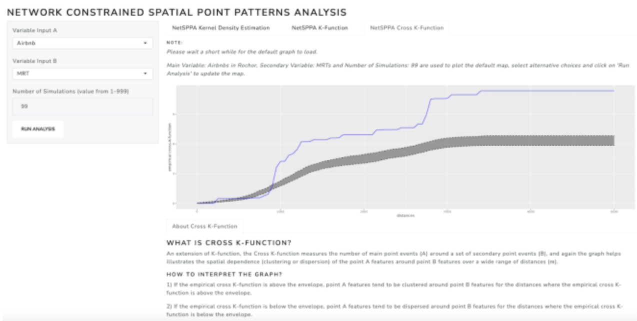


Fig 7.8: 'NetSPPA' Tab - NetSPPA Cross K-Function (Rochor)

The main difference between SPPA Cross K-Function and NetSPPA Cross K-Function is that for SPPA Cross K-Function, users can still select their own zone to analyse. However, for NetSPPA Cross K-Function, the team has decided to solely focus on Rochor.

8 DEMONSTRATION & RESULTS

Let's gather our findings from the various tabs.

<u>'Visualisation' Tab</u>

We can see that the zone with the highest number of Airbnbs is Kallang. This is followed by Downtown Core, Outram and Rochor. Let's analyse these zones.

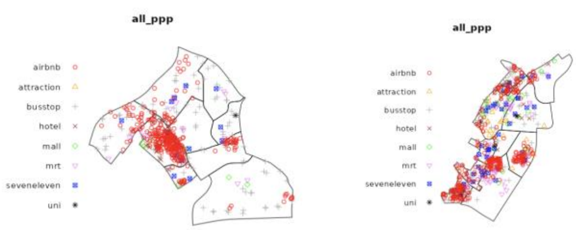


Fig 8.1 Left - Kallang; Right - Downtown Core

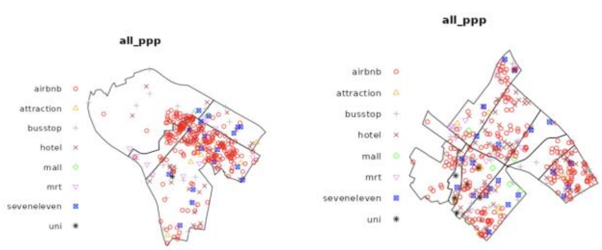


Fig 8.2 Left - Outram; Right - Rochor

From Fig 8.1 and 8.2, we can see that for all top 4 zones above, areas with many Airbnb listing points tend to have many 7-11 stores, MRTs and bus stops nearby. However, for some zones specifically, there are more unique spatial patterns. For instance, while we do see some malls being near regions with high numbers of Airbnb points in other zones, specifically in Downtown Core, many malls are seen near Airbnb points - this is especially the case towards the Northern part of Downtown Core. Moreover, in Rochor, Airbnb points are usually found near Universities.

<u>'Tmap' Tab</u>

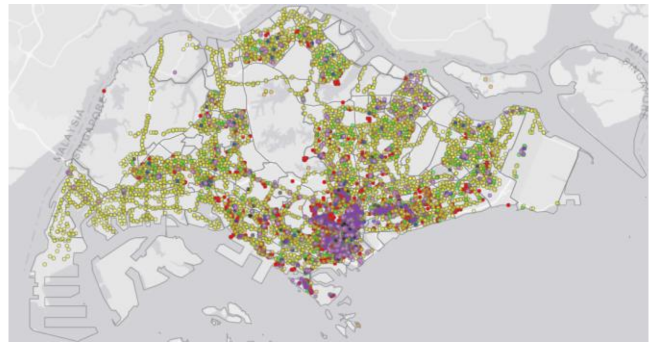


Fig 8.3: Tmap (Whole Singapore)

From Fig 8.3, we can see that the central region of Singapore has the highest concentration of Airbnb points. This aligns with our common knowledge as the central region in Singapore (eg. Orchard Road, Marina Bay) is the main working location and has the greatest number of amenities. As such, international employees and tourists would usually prefer living around the city centre in Singapore. Cognisant of this growing demand, Airbnb rentals have exploded in the central region of Singapore - which nicely explains what we see in Fig 8.3.

<u>'SPPA' Tab</u>

KDE

Let's explore the top 4 zones with the high numbers of Airbnb listings and analyse their density spread across each of the 4 zones.

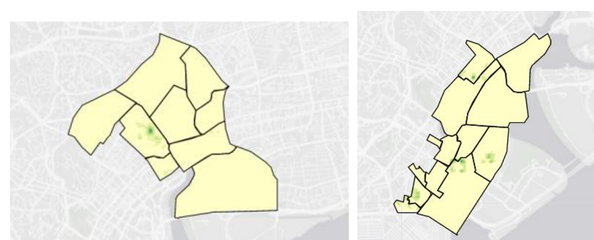


Fig 8.4 Left - Kallang; Right - Downtown Core

Fig 8.5 Left - Outram; Right - Rochor

Interestingly, the plots of the top 4 zones with the highest numbers of Airbnb points in Fig 8.4 and 8.5 resemble the top 4 plots in Fig 8.1 and Fig 8.2. This is because in Fig 8.1 and 8.2, the areas with higher concentrations of Airbnb points are now shaded in different colours to indicate higher Airbnb point density for that zone.

SPPA G-Function

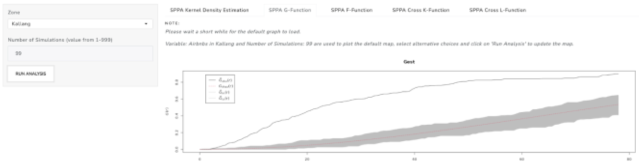


Fig 8.6: SPPA G-Function for Kallang

From the G-Function graph, we can infer that since the observed G is above the envelope, it indicates that the Airbnbs in the Kallang Zone are clustered. We can reject the null hypothesis (Null hypothesis: the Airbnbs in the Kallang are randomly distributed) as the value is statistically significant. This means that the high density regions that we saw for Kallang in Fig 8.4 and 8.1 were really clusters of Airbnbs.

SPPA F-Function

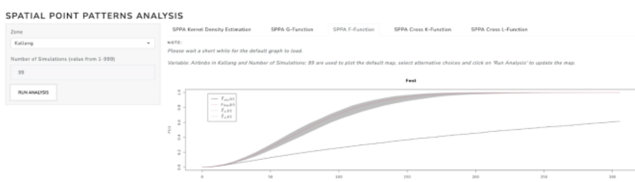


Fig 8.7: SPPA F-Function for Kallang

From the F-Function graph, we can infer that since the observed F is below the envelope, it indicates that the Airbnbs in the Kallang are clustered. We can reject the null hypothesis (Null hypothesis: the Airbnbs in the Kallang are randomly distributed) as the value is statistically significant. Essentially, either G-Function or F-Function can be used to determine if the Airbnb points are clustered. In this application, we included both functions to give users more options to determine whichever function they preferred to use.

SPPA Cross K-Function

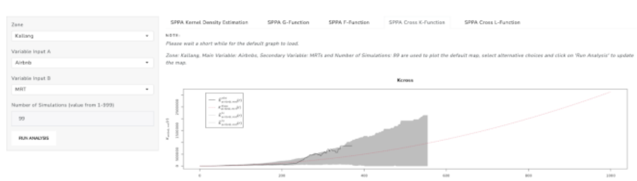


Fig 8.8: SPPA Cross K-Function for Kallang

From the graph in Fig 8.8, we can see that up to 300m, the observed K is inside the envelope. This indicates that up to 300m, Airbnb and MRT points are spatially independent. This is an interesting concept - from the 'Visualisation' Tab and 'Tmap' Tab for Kallang, we established that regions with higher densities of Airbnbs tend to be near MRTs. After doing G/F-Function, we've established that the regions of higher Airbnbs in Kallang are indeed clusters.

Logically, we would expect that the closer an Airbnb is to the MRT the better and that from 0-300m, there would be no spatial independence between Airbnbs and MRTs. This is a generally sound assumption as many people would like to be near MRTs out of convenience so the demand for Airbnbs that are much closer to MRTs would be higher, leading to a rise in Airbnb points that are situated very close to MRT points.

However, this does not seem to be the case - in reality according to the graph, there's spatial independence for Airbnb and MRT points from 0-300m. It could be due to the fact that being too near MRTs is seen as a nuisance due to the loud noises generated by the trains. Cognisant of this, landlords may not typically convert units that are extremely near MRTs to Airbnb rentals due to the lower demand.

SPPA Cross L-Function

Since SPPA Cross L-Function is a standardised version of Cross K-Function and are very similar, similar results would be generated.

<u>'NetSPPA' Tab</u>

NetSPPA Kernel Density Estimation

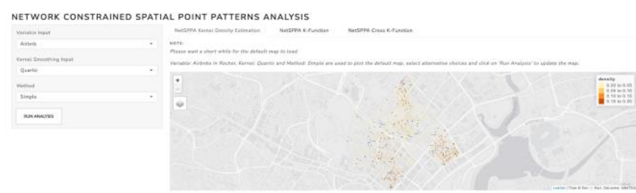


Fig 8.9: NetSPPA Kernel Density Estimation for Rochor

Please refer to Section 5.4.1 for the analysis obtained on Rochor and comparison between a conventional SPPA KDE and a NetSPPA KDE.

NetSPPA K-Function

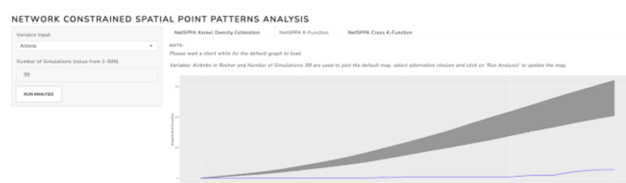


Fig 8.10: NetSPPA K-Function for Rochor

From the graph in Fig 8.10, we can see that the blue empirical network K-Function of the Airbnb point events in Rochor is below the envelope. This indicates that the Airbnbs in Rochor are more dispersed than what we can expect from a random distribution. We can reject the null hypothesis (Null hypothesis: The distribution of point events is uniformly distributed over the street network in Rochor) as the value is statistically significant. This means that generally for all distances between Rochor Airbnb points, we can safely conclude that the Rochor Airbnb listing points are not uniformly distributed across the street network.

NetSPPA Cross K-Function

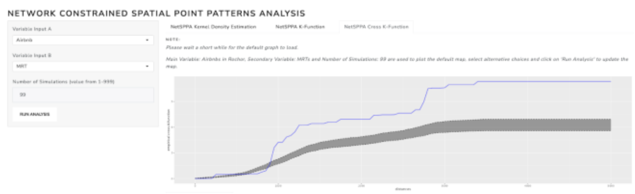


Fig 8.11: NetSPPA Cross K-Function for Rochor

From the graph in Fig 8.11, we can see that from 700-850m the empirical Cross K-Function is below the envelope. This indicates that in Rochor the Airbnb points tend to be dispersed around MRT points. From 850-900m, the empirical Cross K-Function is within the envelope. This indicates that in Rochor, Airbnb points tend to be randomly located around MRT points. From 900m onwards, the empirical Cross K-Function is above the envelope. This signifies that in Rochor Airbnb points tend to be clustered around MRT points from 900m onwards.

This analysis is similar to the one we did for the SPPA Cross K-Function we did for Kallang under Fig 8.8. Similarly for Rochor, we would expect that the closer to the MRT the better and that from 0-900m, there would be no spatial independence between Airbnbs and MRTs. This is a generally sound assumption as many people would like to be near MRTs out of convenience so the demand for Airbnbs much closer to MRTs would be higher, leading to a rise in Airbnb points that are situated very close to MRT points.

However, this does not seem to be the case - in reality according to the graph, there's spatial independence for Airbnb and MRT points that are closer together. It could be due to the fact that being too near MRTs is seen as a nuisance due to the loud noises generated by the trains. Cognisant of this, landlords in Rochor may not typically convert units that are extremely near MRTs to Airbnb rentals due to the lower demand.

9 DISCUSSION

From the demonstration and results section, we picked up many interesting insights pertaining to the Airbnbs and the amenity points in Singapore. Logically speaking, most of us would assume that since people generally prioritise convenience, Airbnbs would be strategically located near MRTs. However, from our analysis obtained from both SPPA Cross K-Function and NetSPPA Cross K-Function, Airbnbs and MRTs are spatially independent at closer distances. This observation holds true for different zones explored (eg. for Kallang and Rochor). It could be due to the fact that being too near MRTs is seen as a nuisance due to the loud noises generated by the trains. Cognisant of this, landlords in both Kallang and Rochor do not typically convert units that are extremely near MRTs to Airbnb rentals due to the lower demand.

From Fig 5.4.1.2, we can see that there are higher densities of Airbnb points towards the outer-ends of Rochor where larger roads are seen. This could suggest that Airbnbs in Rochor tend to be near larger, main roads.

10 LIMITATIONS

The datasets used may not be the most updated version. For users who are using this web application and want real-time updates, this application is still unable to automatically provide real-time feedback on the real-time data of Airbnb listing points and amenity points in Singapore. Moreover, we have yet to incorporate the uploading of SHP and CSV files by users for them to do analysis on the topics they are interested in.

11 FUTURE WORKS

Prices of Airbnbs and hotels can also be added to further prove the positive relationship of prices between the Airbnbs and hotels nearby, as the Airbnb prices will increase as the hotel prices increase, due to common amenities like MRT, 7-11, and bus stops. Adding the individual prices of Airbnbs would be interesting to visualise, as it will complement the distance to amenities like MRT, 7-11, tourist spots, and bus stops. Additionally, a search system can also be implemented, where users will be able to search for their desired Airbnb listing.

12 ACKNOWLEDGEMENTS

Thank you Dr. Kam Tin Seong, Associate Professor of Information Systems at School of Computing & Information Systems, Singapore Management University for his guidance and support throughout this course.

13 REFERENCES

[1] Wong, D. (2018, September 26). *Singapore is 2nd most visited city in Asia-Pacific, 5th in the World: MasterCard*. The Straits Times. https://www.straitstimes.com/singapore/singapore-is-2nd-most-visited-city-in-asia-pacific-5th-in-the-world-mastercard

[2] Government of Singapore. (2022, December 7). *Singapore: is it really the most expensive place to live*. Gov.sg. https://www.gov.sg/article/singapore-is-it-really-the-most-expensive-place-to-live

[3] Teo, J. P., & Yu, Y. (2021, November 19). *Spatial Pointers*. Shinyapps.io.https://yiling-yu.shinyapps.io/is415_project_shiny_app/

[4] Ang, J. Y., Chong, S., & Tan, Y. L. (2019, April 15). *Signal Research Paper*. SMU Wikipedia. https://wiki.smu.edu.sg/1819t2is415/Signal_Research_Paper

[5] Gelb, J. (2023, April 11). *Network Kernel Density Estimate*. CRAN R Project. https://cran.r-project.org/web/packages/spNetwork/vignettes/NKDE.html

[6] Tejera, M., Carella, G., & Baptiste, F. (2021, December 15). *Distance-based functions for the spatial analysis of Point data in BigQuery*. Carto. https://carto.com/blog/distance-based-functions-bigquery