Web Analytics Project Assignment

Project Name: Urban Planning (Singapore)



Group Member:

Hon Jia Yu Samuel (Part Time/ G1902020G)

Sen Poh Swee (Part Time/ G1901954D)

Sim Shu Biao Desmond (Full Time/ G1901940G)

Leong Yew Wing (Part Time/ G1902405D)

Contents

1.	Introduction	3
2.	Background	4
3.	Data Sourcing	5
4.	Motivation	6
5.	Setting	8
5.1.	. Estate Selection	8
5.2.	. Network Graph	8
5.3.	. Programming Language	g
6.	Methodology	10
6.1.	. Raw Dataset	10
6.2.	. Initializing of Network	10
6.3.	Pre-processing of Networks	10
6.4.	. Network Graph Establishment	11
6.5.	. Community Detection	11
7.	Results and Discussions	12
7.1.	. Main Graph – Hougang (Mature) and Sengkang (Non Mature)	12
7.2.	. Main Graph – Ang Mo Kio (Mature) and Bukit Batok (Non Mature)	13
7.3.	. Main Graph – Tampines (Mature) and Woodlands (Non Mature)	13
7.4.	. Mature VS Non Mature Estate	14
7.5.	. Community Detection	15
8.	Experiment	17
8.1.	. Ang Mo Kio Vs Bukit Batok	17
8.2.	. Woodlands s Tampines	18
8.3.	. Hougang Vs Sengkang	18
9.	Recommendation & Conclusion	20
10.	Appendices	21
Α	Appendix A: Main Graph - Estate Networks	21
Α	Appendix B: Main Graph - Degree Distribution Charts	21
Α	Appendix C: Subgraph - Estate Networks	21

1. Introduction

Urban planning is a technical and political process concerned with the development and design of land use and the built environment, including air, water, and the infrastructure passing into and out of urban areas, such as transportation, communications, and distribution networks.

Urban planning deals with physical layout of human settlements. It is closely related to the field of urban design and some urban planners provide designs for streets, parks, buildings and other urban areas. Urban planning is also referred to as urban and regional planning, regional planning, town planning, city planning, rural planning, urban development or some combination in various areas worldwide.

The importance of the urban planning is increasing in the 21st century, as modern society begins to face issues of increased population growth, climate change and unsustainable development especially to the major cities in the world such as Singapore, Manila, Mumbai and Paris etc.

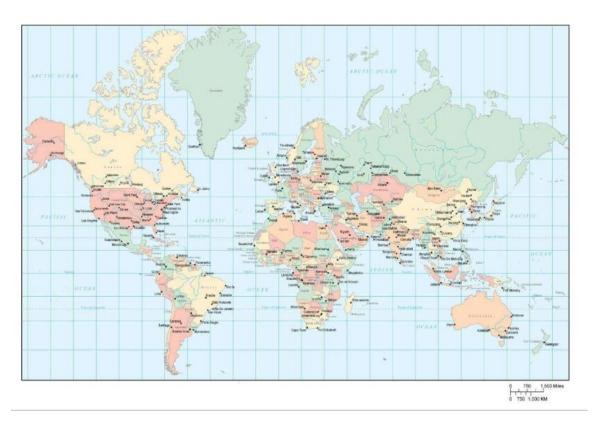


Figure 1: Major Cities in the World Map

2. Background

The population in Singapore has jumped more than five times since independence on the year of 1965. Its population density is the second highest (Hong Kong is the first) in the world. Singapore population will be expecting to grow to keep up with the economy competitiveness within the region and to tackle for ageing population due to the low fertility rate.



Figure 2: Growth of Singapore Population

A Sustainable Population for a Dynamic Singapore: Population White Paper, or simply known as the Population White Paper (PWP), is a controversial white paper released by the government of Singapore in January 2013 that projects Singapore's population as 6.9 million by the year 2030.

Many Singaporeans have attributed the government's population and immigration policy as the cause of overcrowding and falling reliability of its public transportation system, increasing property prices for housing, suppressed wage level, increased competition for jobs (especially for professionals, managers, executives and technicians and education, increasing income inequality and other social problems .Academics have also criticized the PWP as being "overly mechanistic, economically simplistic and astonishingly sociologically and politically naïve.

Although Land Transport Authority (LTA) have introduced more buses on the road and opened up MRT lines in the recent years, it is still not deniable the fact is overcrowding, and traffic jam are still common during peak hour.

At the same time, Housing Development Board (HDB) is acting first to build more flat and develop new estates which is also known as non-mature to accommodate for the another 1 million of population growth in next decade (in the year of 2030). HDB offer 'Build-To-Order' or in short ('BTO') scheme for young couple in Singapore to own the new flat but required to wait for 4 years. The new flats are launched more at the non-mature estate than mature estate.

3. Data Sourcing

While the group is exploring for the relevant transport data to study about the network in Singapore, the group has come crossed LTA DataMall which is part of Singapore Smart Nation initiative. The link to the LTA DataMall is found below:

https://www.mytransport.sg/content/mytransport/home/dataMall.html



Figure 3: LTA DATAMAL

LTA DataMall is a push from the government to try and improve the efficiency of Singapore, as a country, using technology. Among other projects, there has been a push to make real-time data openly available so that everyone, from individuals to corporations, can use it creatively to make solutions to day-to-day municipal problems. It is launched on 2016.

This dataset provides both offline geographical data on roads & public transport, as well as real-time data on things like bus arrivals and taxis. Using this dataset, the Python programming language, and basic programming and data-science techniques, we will build a trip planner to find the shortest bus commute from A to B, but powered by real data and bounded by real-world.

Some of the projects that based on LTA DataMall can be easily found by google search. One of the notable works is the python coding that written by the current Singapore Prime Minister's Son – Li Haoyi. His related works can be found in the link below:

http://www.lihaoyi.com/post/PlanningBusTripswithPythonSingaporesSmartNationAPIs.html#improvements

4. Motivation

With the understanding of the background and the datasets can retrieve from LTA DataMall using the API (Application Programming Interface), the group is motivated to study about the bus network in Singapore based on population distribution in the estate. Below is the table of the population distribute.

It will be interesting to discover how the bus network is being connected within the estate. The best way to understand the impact of population growth compare to the public facilities (in the case will be the transportation network), which usually takes place in non-mature estate, it will be an ideal way to make a comparison between mature and non-mature estate, to see whether the current non-mature is catching up with mature estate based on the similar population.

As there is no proper definition of Mature and Non-Mature Estate from HDB, the group has decided to determine it by having the conservative population growth from 1999 to 2009 and 2009 to 2019. The mature and non-mature can be seemed in same the table.

Table 1:Population Distribution

Name	Area	Population 30/6/1999	Population 30/6/2009	Population 30/6/2019	Estate
Ang Mo Kio	NE	180,112	179,297	163,950	Mature
Bedok	E	284,318	294,519	279,380	Mature
Bishan	С	89,746	91,298	88,010	Mature
Bukit Batok	W	126,200	144,198	153,740	Non-Mature
Bukit Merah	С	148,299	157,122	151,980	Mature
Bukit Panjang	W	96,031	128,734	139,280	Non-Mature
Choa Chu Kang	W	136,105	173,291	190,890	Non-Mature
Clementi	W	90,864	91,874	92,420	Mature
Geylang	С	118,565	120,690	110,200	Mature
Hougang	NE	203,402	216,697	226,240	Mature

Name	Area	Population 30/6/1999	Population 30/6/2009	Population 30/6/2019	Estate
Jurong East	W	88,883	88,118	79,240	Mature
Jurong West	W	203,838	267,524	264,860	Mature
Kallang	С	92,099	99,559	101,520	Mature
Pasir Ris	Е	116,245	133,863	148,020	Non-Mature
Punggol	NE	-	59,386	170,560	Non-Mature
Queenstown	С	97,684	98,502	96,340	Mature
Sembawang	N	31,418	72,732	95,920	Non-Mature
Sengkang	NE	60,870	167,054	244,600	Non-Mature
Serangoon	NE	122,205	124,782	116,310	Mature
Tampines	Е	252,448	261,743	256,730	Mature
Toa Payoh	С	121,004	124,653	120,650	Mature
Woodlands	N	186,805	245,109	254,730	Non-Mature
Yishun	N	176,689	185,214	220,320	Non-Mature

5. Setting

5.1. Estate Selection

The group has selected 3 x mature and 3 x non-mature estate to research. The basic rule to select the estate is based on the similar population.

The reason to compare Hougang is compared with Sengkang is because both estates are located next to each other. Next, Ang Mo Kio and Bukit Batok is the gateway between North to Centre and North to West respectively. Lastly, both Tampines and Woodlands are the regional hub for East and North respectively.

Mature (M)	Non-Mature (NM)
Hougang (NE)	Sengkang (NE)
Pop: 226,240	Pop: 244,600
Ang Mo Kio (NE)	Bukit Batok(W)
Pop: 163,950	Pop: 153,740
Tampines (E)	Woodlands (N)
Pop: 256,730	Pop: 254,730

Table 2: Mature VS Non-Mature Estate

5.2. Network Graph

Bus stop shall be considered as 'Node'. Bus route shall connect with bus stop together as a 'Edge'. In addition, the 'Edge' come with Weight as Distance.



Figure 4: Illustration of Network

5.3. Programming Language

Python language has been used in the research. Below are the 3 main modules:

1) Requests: Extract HTTP from DataMall

2) JSON: Create readable data format from Javascrip

3) Networkx: Network Graph





Figure 5: Main Python Module

6. Methodology

6.1. Raw Dataset

An account shall be registered at the DataMall to request for API. 2 set of data are being extracted: Bus Stop and Bus Routes to JSON format.

6.2. Initializing of Network

Pair up the Bus Stop based on the Stop Sequence. If the stop sequence is 5, the next stop sequence is 6, provided the bus service is the same number. The distance will also be calculated since is given as cumulative distance in data. The pair up data will write to CSV file. Please see the example below.

Edge Point A Road Name A Description A Point B Road Name B Description B Bus No. **Bus Direct** Dist (km) 75009 TampinesCtrl 1 Tampines Int 76059 Tampines Ave 5 Tampines Int 10 0.6 76059 76069 0.5 Tampines Ave 5 Opp Our Tampines Hub Tampines Ave 5 Our Tampines 10 76069 Blk 147 96289 Sime i Ave Blk 147 Tampines Ave 5 10 1.2 Changi General Hoso 96289 Simei Ave 96109 Sime i Ave angi General Ho 0.4 10 Opp Blk 3012 96109 Simei Ave 85079 Upp Changi Rd Opp Blk 3012 10 0.6 85079 Upp Changi Rd Flextronic 85089 Upp Changi Rd Flextronic 10 0.2 Upp Changi Rd Aft Sungei Bedok Bedok Rd lft Sunge i Bedol 0.3 85069 Bedok Rd Bedok Mkt PI 85059 Bedok Rd Bedok Mkt P I 10 0.3 Bedok Rd Bedok Rd 85059 Bef Jln Chempaka Kuning 85049 In Chempaka Ku 10 0.4 85049 Bedok Rd Opp Man Fatt Lam Tp 85039 Bedok Rd p Man Fatt Lam 10 0.3 85039 Bedok Rd Opp Excelsior Gdns 85029 Bedok Rd pp Excelsior Gdi 10 0.2 85029 Bedok Rd Aft Bedok Meth Ch Bedok Rd ft Bedok Meth C 0.2 85019 Bedok Rd Eastwood Ctr 94079 Upp East Coast Rd Eastwood Ctr 10 0.3 Upp East Coast Rd Upp East Coast Rd Bedok Camp 2 94079 Bedok Camp 2 94069 10 0.6 94069 Upp East Coast Rd Opp Kew Green Condo 94059 Upp East Coast Rd) Kew Green Cor 10 0.3 94059 Upp East Coast Rd Temasek Sec Sch 94049 Upp East Coast Rd Temasek Sec Sci 0.4 10 94049 Upp East Coast Rd Aft Bedok Sth Ave 3 94089 Upp East Coast Rd it Bedok Sth Ave 0.2 94089 Upp East Coast Rd Calvary Ably Of God Ch 94039 Upp East Coast Rd /ary Ably Of God 10 0.3 Opp The Summit 94029 Upp East Coast Rd Dpp The Summir 94039 Upp East Coast Rd 10 1 0.4

Table 3: Example of the Estate Data that extracted from JSON format.

6.3. Pre-processing of Networks

Aft Sennet Rd

Opp Evergreen Ave

94029

94019

Upp East Coast Rd

Upp East Coast Rd

Estate name is used as the Substring of the road name such as "Ang Mo Kio" to find road name for: Ang Mo Kio Street 81, Ang Mo Kio Ave 3 etc. Common way for HDB to name it.

94019

93099

Upp East Coast Rd Aft Sennet Rd

Upp East Coast Rd pp Evergreen Av

10

10

0.2

0.4

As bus route will only be considered within the estate, there will not be discounted graph in the model. Below is the illustration.

Only Green Nodes & Edge that <u>arewithin</u> the estate are conside<u>r</u>ed in the Study

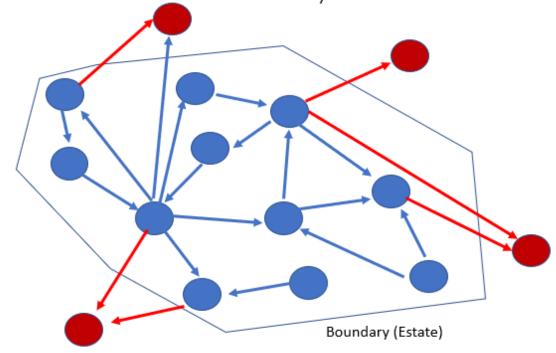


Figure 6: Boundary of an estate

6.4. Network Graph Establishment

Read from CSV file to define nodes, edges and weights. The graph will be drawn as directed.

6.5. Community Detection

There are limited algorithms to detect community on a directed graph. Even for the network module, the only available algorithm to use is strongly connected component (SCC). It is logical to use the SCC for the detection because it will be easy to visual the bus loop within the estate. SCC is useful to explore the efficiency in terms of bringing the commuter around the estate. The concept of SCC is based on detecting for the looping within the graph. If the commuter wants to travel to other estate, the prefer option will be taking a MRT.

7. Results and Discussions

All the results and network graphs are generated from python based on the module call Networkx.

7.1. Main Graph – Hougang (Mature) and Sengkang (Non-Mature)

Despite having the similar population, Sengkang has about half of nodes and edges of Hougang. It might be due to the LRT is the alternative for bus in Sengkang. When compare to density, Sengkang is performed better than Hougang due to having lesser edges and nodes but is more utilised in terms between edges and nodes. It should be typical to have highest nodes in interchange, however, for Sengkang, the highest node is found community hub. Both Hougang and Sengkang has fair the same in Clustering Coefficient. All the top clustering coefficient nodes are located near to MRT/ interchange.

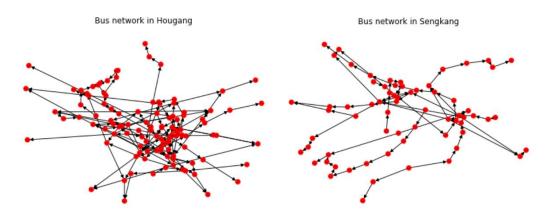


Figure 7: Bus Network for Hougang and Sengkang

Table 4: Result Comparison based on Main Graph for Hougang and Sengkang

Estate	Nodes	Edges	Density	Density	Node with Highest Degree	Node with Highest Clustering Coefficient		
Hougang	110	142		0.01	Hougang Ctrl Int	Blk 512 (0.17)		
Tiougang	110	0.01	0.01	0.01	0.01	0.01	(8)	Blk 831 (0.17)
Sengkang	67	72		0.02	Sengkang Community Hub (6)	Sengkang Int (0.17)		
Scrighting	07	0.02	0.02		Block 323B (6)	Blk 234 (0.17)		

7.2. Main Graph – Ang Mo Kio (Mature) and Bukit Batok (Non-Mature)

Both graphs are considered dense. Although Ang Mo Kio have more edges and nodes than Bukit Batok, but their density is about the same. In the case for Ang Mo Kio, the highest node is found at Yio Chu Kang interchange because the Yio Chu Kang is located along Ang Mo Kio Ave 5. In terms of Clustering Coefficient, both estates score the same. It is surprised that Ang Mo Kio Depot is located at the border of Ang Mo Kio, which has become one of the 3 highest in terms of clustering coefficient within Ang Mo Kio. Based on google map, it is besides the Ang Mo Kio Industry park and is no near to any MRT. Aft Bt Batok West Ave 5 is also located out of nowhere next to a nature park without any MRT. It means it is having more neighbours' nodes in the particular of the area.

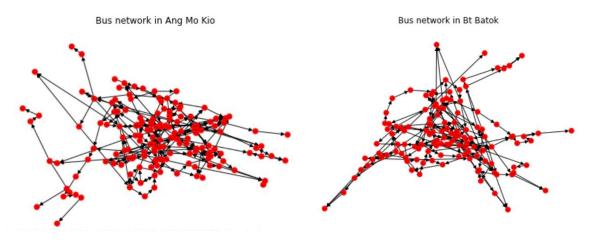


Figure 8: Bus network for Ang Mo Kio and Bukit Batok

Estate	Nodes	Edges	Density	Node with Highest Degree	Node with Highest Clustering Coefficient
Ang Mo Kio	157	215	0.01	Yio Chu Kang Int (10)	Ang Mo Kio Depot (0.167) Opp Yio Chu Kang Stadium (0.167) Aft Ang Mo Kio Stn Exit A (01.67)
Bt Batok	129	182	0.01	Bt Batok Int (8)	Blk 644 (0.167) Blk 225 (0.167) Aft Bt Batok West Ave 5 (0.167)

Table 5: Result Comparison based on Main Graph for Ang Mo Kio and Bukit Batok

7.3. Main Graph – Tampines (Mature) and Woodlands (Non-Mature)

Both graphs are much dense that the other 4 graphs due to having a higher population. Both are considered as regional hubs which is obvious the mature estate for Tampines is slightly better developed than Woodlands in terms of slightly higher number of nodes and edge. The node with

the highest clustering coefficient is Tampines east station which is far from the Tampines interchange. It can be due to it is next to a living estate – Pasir Ris (extreme east in Singapore) which might have more buses that travelled from Pasir Ris to Tampines. For Woodlands, Blk 825 and 823 is located some distance from the MRT station. It has notice that north of Blk 825 and 823 are Senoko industry park and Sembawang Shipyard. Thus, it can say that the 2 bus stops are the gateway from the MRT to industry park.

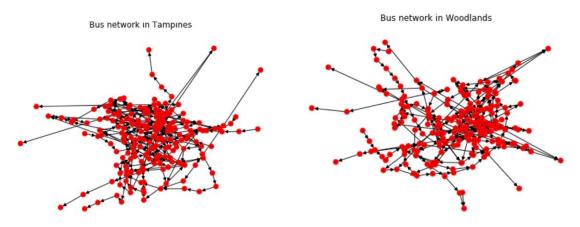


Figure 9: Bus network for Tampines and Woodlands

Estate	Nodes	Edges	Density	Node with Highest Degree	Node with Highest Clustering Coefficient
Tampines	200	276	0.01	Tampines Int (10)	TAMPINES EAST STN EXIT B (0.167)
Woodlands	210	266	0.01	W'Lands Temp Int (12)	Blk 825 (0.08)

Blk 803 (0.08)

Table 6: Result Comparison based on Main Graph for Tampines and Woodlands

7.4. Mature VS Non-Mature Estate

In general, Mature Estate has more bus stop and routes than non-mature estate. The degree probability charts for all the 6 estate can be found in the Appendix which has shown as scale free. As Singapore is increasing the number of buses and stops particularly for non-mature estate, its chart will continue to grow based on the power law distribution for the next decade.

For Sengkang case, it is even significant lesser when compare to Hougang which the main root of the cause might be due to LRT. But Sengkang has scored the best in terms of density than the rest of the estate. Both Mature and Non-Mature Estate have the highest degrees at or near to centre area where interchange and MRT is.

In terms of highest Clustering coefficient node, it is seen that more mature estate can also find at or near to interchange or MRT whereas non-mature estate is usually not near to any of these, thus this might be an area of improvement for the government to investigate it.

7.5. Community Detection

The section will be discussing about detecting the community within estate. In graph theory term, it can be explained to find the subgraph out from the main graph. The subgraph will use to discover for the key player, in the context of bus stop.

It is common for commuter to use Google Map for navigating on the shortest travel from one location to another location. However, while the commuter check for google map again during the travel, the result might be different. This is because the network is not strongly connected. As explain in the previous section, the detect community algorithm is known as Strong Connected Component (SCC) which is to find the subgraph looping from the main graph. All the single nodes will be eliminated and focused on the main loop of the subgraph.

With the reduction of the nodes and edges in the graph, more measurements can be performed such as the diameter and average path length of the network. The key player will also be access and measure in the detection. The experiment of the result comparison will be illustrated on Section 8.

The table below summaries the reduced nodes and edges from SCC algorithm.

Table 7: Nodes and Edges based on subgraph

Estate	Nodes	Edges
Hougang	28	36
Sengkang	3	3
Ang Mo Kio	104	148
Bt Batok	81	113
Tampines	147	223
Woodlands	139	182

Below is the subgraph of the 6 estates that generated out from python:

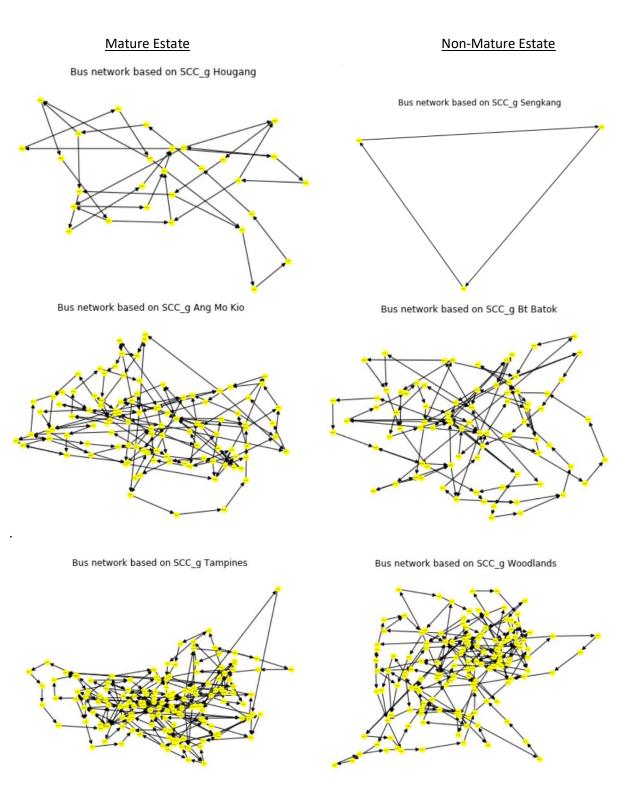


Figure 10: Subgraph Network

8. Experiment

The following are the result comparison between Mature VS Non-Mature Estate:

8.1. Ang Mo Kio Vs Bukit Batok

- Measure: Path Length & Diameter
- Objective: To analyse how quickly a commuter can travel from one bus stop (node) to another bus stop

Maximum diameter is 22 links for Ang Mo Kio and 24 Links for Bukit Batok. Both number of bus stops number are almost identical. It can be due to lesser nodes and edges for Bukit Batok which make it appear to be similar with Ang Mo Kio.

Average shortest path length is maximum is 9.75km and 9.45km which means that it has almost the same distance let say from point A to point B. It can be concluded that both estates have developed similar inter bus routes within the subgraph. Inter bus route is refer to buses that operate within the area. As Singapore is having hot season for whole year round, walking for more than 15minutes might be a challenge for elderly and children especially during the noon time, therefore it is important for the inter bus for travel around the estate specially to interchange or MRT.

However, as Bukit Batok will continue to face the raising of population, it is important to have more strong connected nodes and edges to keep up with Ang Mo Kio. If compared based on the current population of the both estates, Bukit Batok still have room for improvement on strongly nodes and edges.

Table 8: Comparison of Path Length & Diameter

Estate	Nodes	Edges	Diameter (Link/ Bus Stop)	Path Length (Weight Considered; km)
Ang Mo Kio	104	148	22	9.75
Bt Batok	81	113	24	9.45

8.2. Woodlands s Tampines

- Measure: Clustering Coefficient
- Objective: To analyse the connectivity of each bus stop (node) to another bus stop

Clustering coefficient is measured the node link to the maximum number neighbour node that connected as the clique. This is also important for the inter bus route concept that explain in Section 8.1. If the commuter took a wrong bus, it will be important for how the commuter can deviate back to the original point that want to go. It is also very common to have high traffic along the road during the peak hour. Having a more flexibility of the bus option will give commuter a choice to decide which bus route to take. In this comparison, it is obvious that Tampines has the higher Clustering Coefficient than woodlands when both are considered as regional hub. Woodlands still have a lot to catch up with Tampines while the population is continuing to increase for Woodlands.

Estate Average Clustering Coefficient

Tampines 0.004

Woodlands 0.0006

Table 9: Comparison for Average Clustering Coefficient

8.3. Hougang Vs Sengkang

- Measure: Between Centrality
- To identify the key player and analyse whether the bus stop (node) is sufficiently equipped for high traffic flow

The degree centrality is simple measurement based on the maximum based on number of edges on the node which is more useful for the main graph. Closeness centrality is based on closeness by the number of links which is not so suitable to directed graph with distance.

Betweenness centrality is used to identify for player which measures the control of bus stop (I) on 2 non-adjacent bus stop (J) and (k). If Bus Stop (I) is on the path of many such interactions, then bus stop (I) is an important player.

The table below is the result. The location of the bus stop is all found near to the bus interchange. Hougang as being the mature estate has a higher score than non-mature estate – Sengkang.

Estate	Betweenness Centrality	Location
Sengkang	0.5	Sengkang Station Sengkang East Rd
		Sengkang Sq.
Hougang	0.61	Hougang Central

Table 10: Key Player and Betweenness Centrality

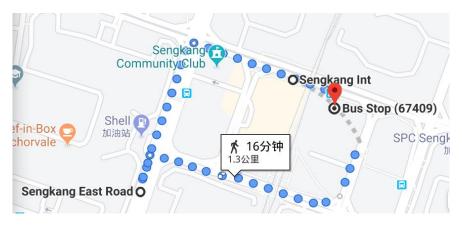


Figure 11: Key Player at Sengkang



Figure 12: Key Player at Hougang

Betweenness centrality in the case will use to find for bus stop that commuter will perform for bus transfer on key player. It is important to know whether:

- Bus stop has the capacity to accommodate the number of commuters
- Likely prone for heavy traffic

The key player is useful for the following suggestions on:

- Bus operator may consider double or single deck or bus lane to allocate during the peak hour.
- Traffic police can also be considered to station for control of the traffic.
- Study for the number of crowds with the design of the bus stop shelter

9. Recommendation & Conclusion

Network measurement such as Degree, Nodes, Density, Clustering, Diameter, Path length etc. are useful to measure the efficiency of the network. Hypothesis studies for network measurements can perhaps be further explored and come out with some confidence level for bus network, particularly important in city like Singapore. Strong Connected Component (SCC) has worked very well in detecting the community. Random walk can also be used to create another subgraph which might give a different result than SCC. Node relationship can also be performed based on the bus frequency or number of commuters. Sensitivity studies of the bus network in the scenario of MRT breakdown should be an interesting topic.

The report has addressed on urban planning based on background and data sourcing that has motivated the group to study about the bus network for mature and non-mature estate in the context of estate's population in Singapore. The report has also illustrated the use of API to request the raw data from LTA DataMall, pre-processing of data by pairing up of bus stop to establish the network graph on 6 different estates. Main Graph and Subgraph (based on Community Detection) are also being analysed and discussed which have bring out different insights and suggestions for the bus network system in Singapore. Overall, mature estate has scored better than non-mature estate in terms of bus network. To meet for challenge of population growth, it is important to understand the bus network and develop the bus network in an effective way where network graph is one of the good tools to access it.

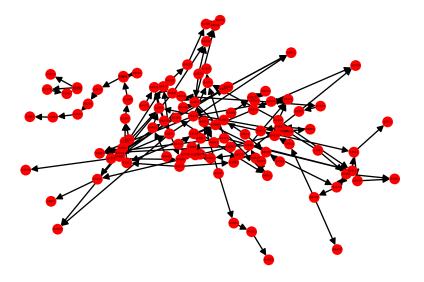
10. Appendices

Appendix A: Main Graph - Estate Networks

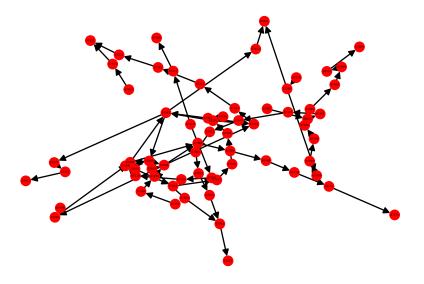
Appendix B: Main Graph - Degree Distribution Charts

Appendix C: Subgraph - Estate Networks

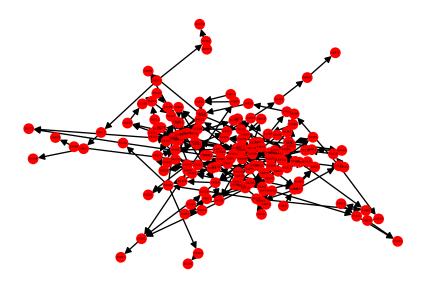
Bus network in Hougang



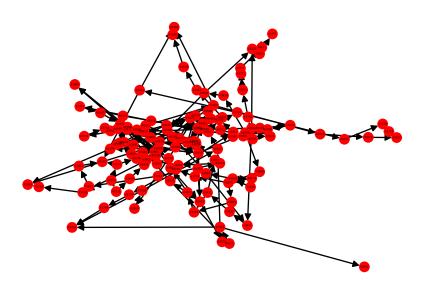
Bus network in Sengkang



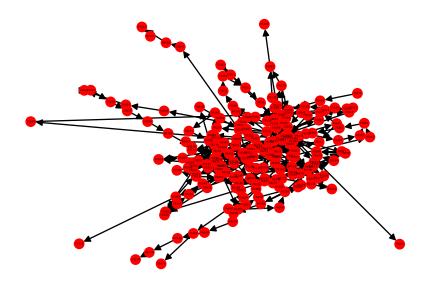
Bus network in Ang Mo Kio



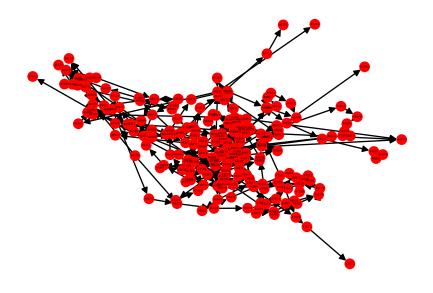
Bus network in Bt Batok

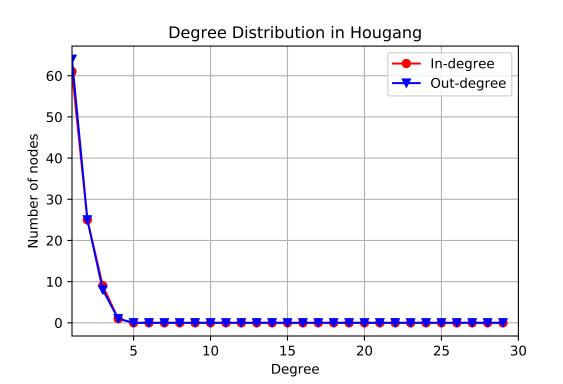


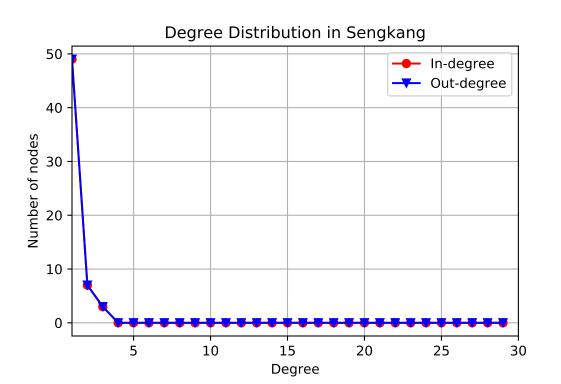
Bus network in Tampines

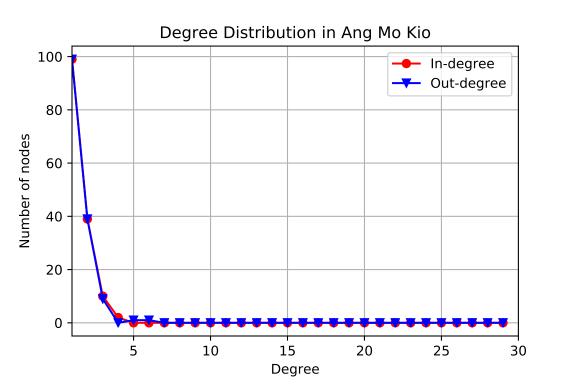


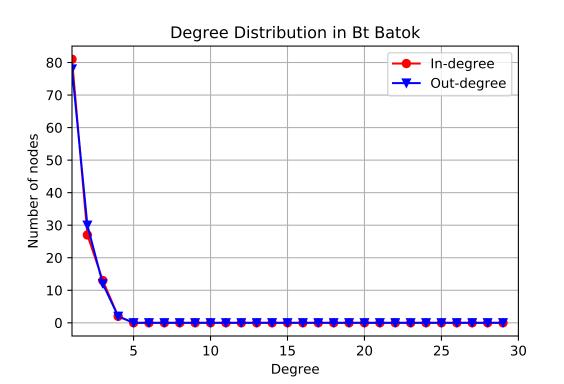
Bus network in Woodlands

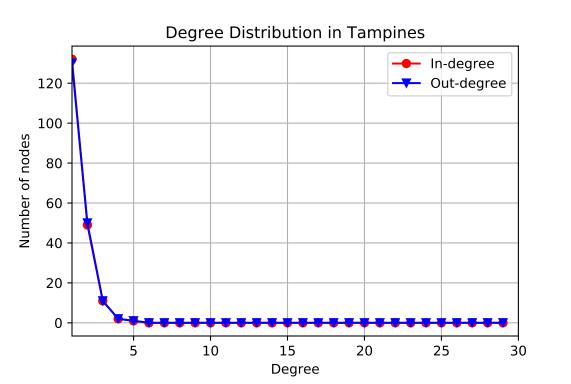


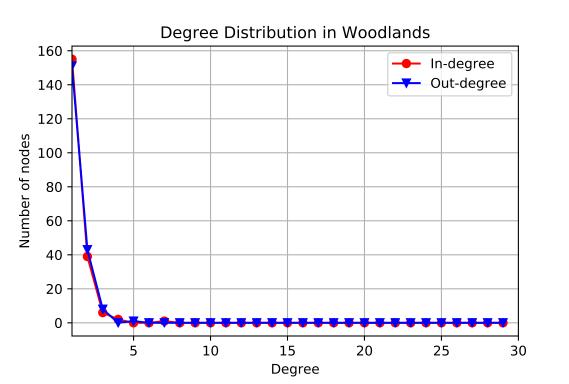




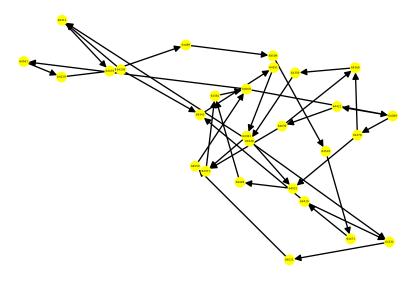




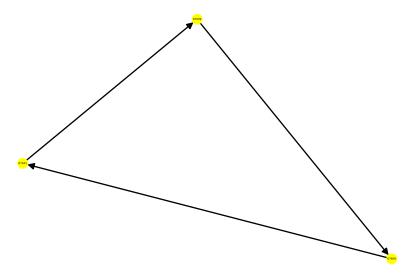




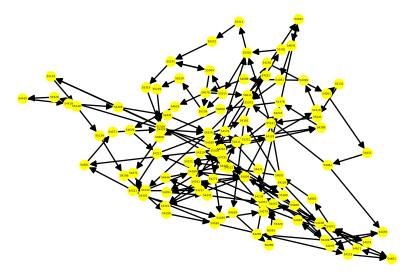
Bus network based on SCC_g Hougang



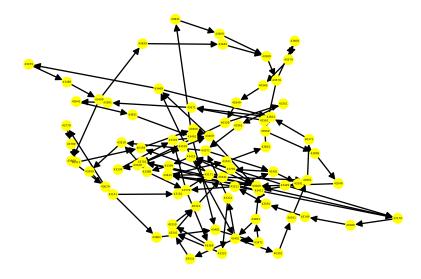
Bus network based on SCC_g Sengkang



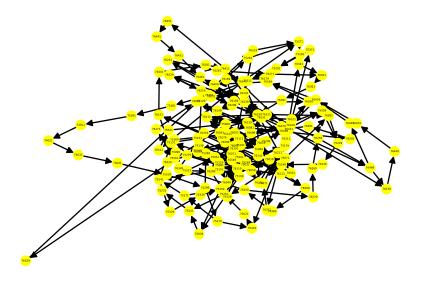
Bus network based on SCC_g Ang Mo Kio



Bus network based on SCC_g Bt Batok



Bus network based on SCC_g Tampines



Bus network based on SCC_g Woodlands

