Analysis of the Commuting Network in the Metropolitan Area of Sao Paulo (MASP), a COVID-19 Hotspot

Shiv Yucel

March 26 2021

1 Introduction

1.1 Overview

The health and economic strain caused by COVID-19 has spurred research into the mobility-based mechanisms of disease spread (Chang et al., 2020; Peixoto, Marcondes, Peixoto, & Oliva, 2020). For respiratory pathogens such as COVID-19, environmental risk factors including traffic and transportation are of great importance to understand the flow of disease within urban areas (Hauge & Meijerink, 2020). The Metropolitan Area of Sao Paulo (MASP), a COVID-19 hotspot with 109,698 cases and 7,615 deaths by May 2020, will be the focus of this analysis (de Souza et al., 2020). The purpose of this study is to answer two primary questions:

- Question 1: What mechanisms in the network determine commuting flow to a region?
- Question 2: Which regions in the MASP commuting network are most at risk of disease transmission?

This study will begin by discussing the motivation for an intra-state understanding of human traffic in MASP, followed by a description of the commuting network data. Following this, the un-weighted and weighted degree distribution (node strength) will be analysed, and assessed for power-law trends. The subsequent analysis involves using a Multiple Regression Quadratic Assignment Procedure to understand how characteristics of the network relate to daily commuting flows within the region. Question 2 is answered using an updated betweenness centrality statistic which incorporates both the strength of commuting flow and number of intermediary nodes, identifying regions at potential risk (Opsahl, Agneessens, & Skvoretz, 2010).

1.2 Background

The first case of COVID-19 in South America was documented on February 25, 2020, growing to over 3 million cases by August 12 (Ribeiro, Sunahara, Sutton, Perc, & Hanley, 2020). The diagnosis occurred in the Metropolitan Area of Sao Paulo, the most populous region in the Southern Hemisphere, with 23 million people spread across 39 municipalities (Candido et al., 2020; Marques et al., 2018). A phylogenetic analysis of the disease, which sequenced genomes from a spatially representative sample of Brazil, tracked viral dynamics across various stages of the pandemic (Candido et al., 2020). After international introduction, the phylogenetic analysis showed that lineages increasingly began spreading within states, 5.1x more frequently than inter-state transmission. (Candido et al., 2020). The disproportionate impacts of intra-state transmission highlight the importance of explaining factors contributing to the baseline commuting network in MASP, and of identifying regions most at risk.

2 Data: MASP Commuting Network

The 2017 Origin and Destination Survey is a questionnaire developed by the Sao Paulo Urban Transportation Department (COMPANHIA DO METROPOLITANO DE SÃO PAULO (METRÔ-SP), 2017). It provides detailed information on commuting patterns in MASP, based on 100,000 interviews across 510 research zones, dispersed throughout all 39 municipalities (Figure 1). The municipality of Sao Paulo contains 66% of the research zones. The data is in the form of an origin destination matrix, well-suited to be represented as a weighted, directed network. The nodes in the network are the numbered research zones, which have assigned coordinates, populations, and areas. The directed edges represent the existence of flow from one area to another, and edge-weights are represented by the magnitude of commuter flow and mean travel time. There are 29,648 directed edges, and the shortest path between the two most distant nodes (diameter) is of length 5. The network density is 0.11 implying that 11% of all potential connections exist. It is also weakly connected, communicating that some nodes cannot be reached in both directions to each other. Other summary network statistics are presented in Table 1 (Bloodgood, Hornsby, Rutherford, & McFarland, 2017).

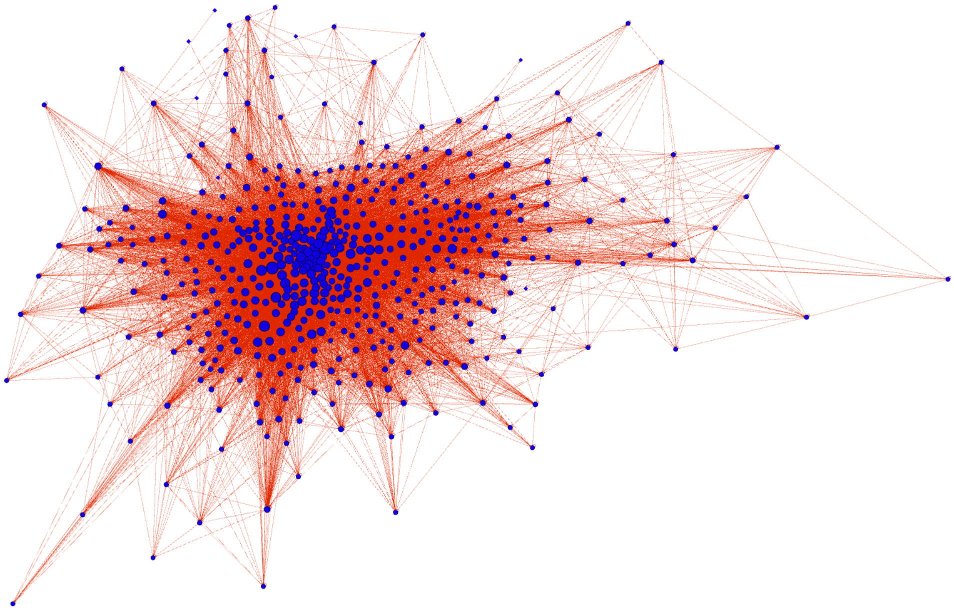


Figure 1: Geographically Representative Graph of Sao Paulo Commuting Network – Node Size Determined by In-Degree

Metric	**Value**
# of Nodes	510
# of Edges	29648
Avg. Shortest Path Length	2.04
Avg. Shortest Path Length (mean time)	66.86
Density	0.11
Diameter	5
Global Clustering Coef.	0.34
Avg. Commuting Time	66.86mins
Avg. Daily Commuters to Region	1447.34
Weakly Connected	Yes

Table 1: Summary Network Statistics of MASP Commuting Network

3 Degree Distribution Analysis

3.1 Un-Weighted In-Degree Distribution

The in-degree of a node is the sum of the directed edges which flow into the node. In the context of disease transmission, in-degree represents the diversity of regions which flow into a particular commuting zone. Given an adjacency matrix where a value of 1 indicates the existence of an edge between two nodes, and 0 represents no edge, the in-degree k of node i is calculated as the sum of row i across all columns

j. In mathematical notation, this is calculated as:

$$k_i^{\text{in}} = \sum_j a_{ij} \tag{1}$$

The in-degree distribution (Figure 2) initially slopes upwards, with a maximum count of nodes between 40 and 50 in-degrees. After this peak, the distribution displays a right-ward skew, with the number of observed nodes for a given degree dropping as the degree increases. For the 510 regions, the median in-degree distribution is 67.15, with a standard deviation of 43.77 in-degrees. The minimum in-degree is 1, and the maximum in-degree of 280 lies 4.95 standard deviations from the mean (Table 2).

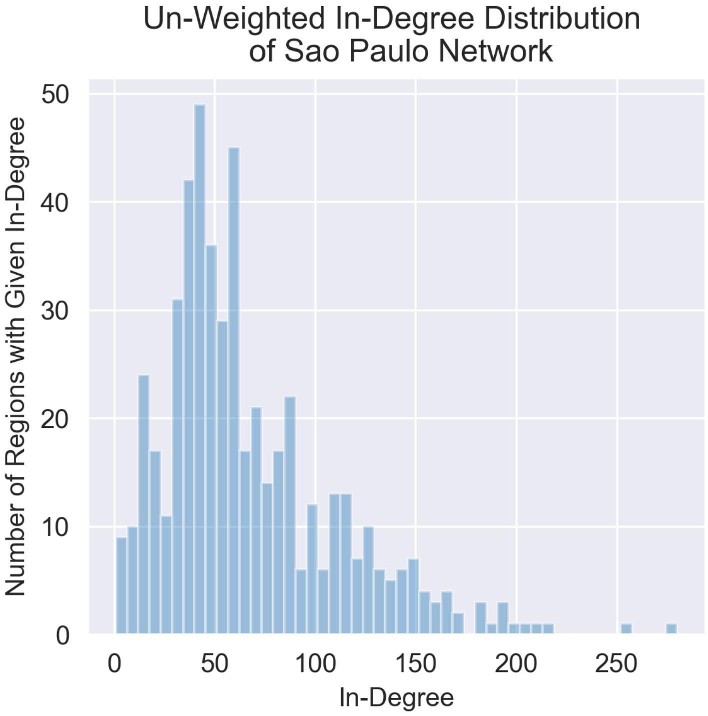


Figure 2: Un-Weighted In-Degree Distribution of MASP

	MASP In-Degrees
std	43.771452
min	1.000000
median	56.000000
max	280.000000

Table 2: Summary Statistics of Un-Weighted In-Degree Distribution

At the right extreme of the distribution, there are six commuting regions which do not share in-degrees groupings with any other regions. These nodes represent the commuting regions with the most topological in-flow, receiving edges from 200 or more regions, presented in Table 3. From an epidemiological perspective, these nodes are of substantive interest as regions that regularly receive individuals from many other regions. The six regions with the highest in-degrees, along with their municipalities, are displayed in Table 3.

Region ID	**In-Degree**	**Municipality**
2	280	Sao Paulo
82	255	Sao Paulo
93	216	Sao Paulo
5	211	Sao Paulo
79	204	Sao Paulo
300	201	Sao Paulo

Table 3: Regions with highest 6 in-degrees, including municipality

3.1.1 Test for Scale-Free Properties

The heavy tails of degree distributions in large, complex networks have been observed to display self-organizing properties in the form of scale-free distributions, where the degree k of a given node follows a predictable power-law (Barabási & Albert, 2019; Lewis, 2009). Observing the MASP network to be scale-free, by having a power-law tailed in-degree distribution, may help to identify a generalisable organisation pattern. A power-law tail has been theorised to capture mechanisms of preferential attachment and growth in the network – both common properties in transportation networks (Barabási & Albert, 2019). Preferential attachment refers to 'rich get richer' mechanism, by which the probability that a new node attaches to a target node i is an increasing function of the degree of i (Alstott, Bullmore, & Plenz, 2014; Simpson, 2001). A distribution following a power-law does not provide evidence of linear preferential attachment, but that it is *consistent* with that mode of attachment. A conclusion would depend on the generative mechanism of the network (Alstott et al., 2014). A distribution can match a scale-free distribution if the following power-law equation fits the data, for some best-fit parameter γ:

$$P(k) \sim k^{-\gamma} \tag{2}$$

To identify this visually, the complementary cumulative distribution function (CCDF) of the in-degree distribution on a log-log scale must be described by a best fit power-law function. Figure 3 is an example from word frequency data, which fits to both a power-law and lognormal distribution.

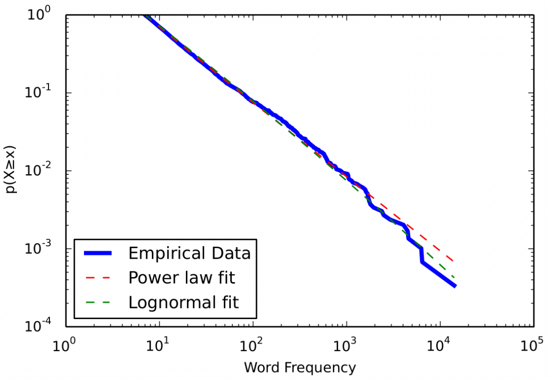
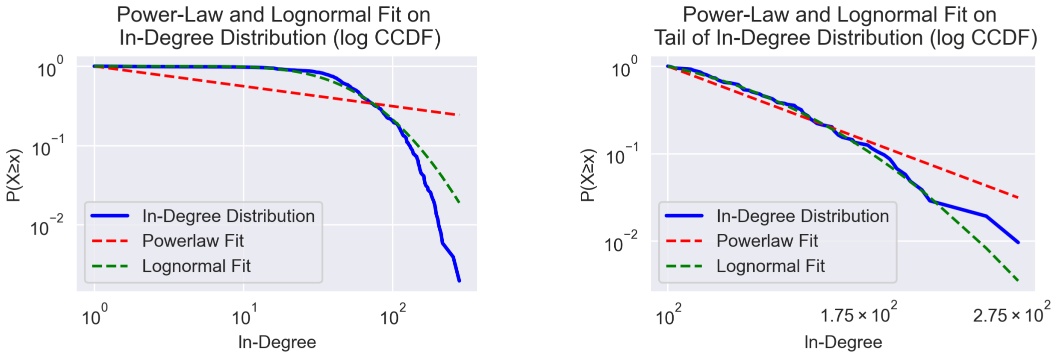


Figure 3: Example of Power-Law Degree Distribution in Word Frequency Data (Alstott, 2014).

In Figure 4(a), a power-law function is fit to the entire CCDF of the in-degree distribution, while in Figure 4(b) it is fit only to the tail, where power-laws are often best followed (Mateos, 2016). Both plots reveal that a power-law function does not capture the in-degree distribution. The lognormal distribution, a heavy-tailed distribution that is not scale-free, is shown by the green line in Figures 4(a) and 4(b). It provides a closer approximation to the in-degree distribution than the power-law, and a likelihood ratio test favours the lognormal fit with a p-value of 0.00 in both plots.



(a) Power-Law/Lognormal Fit on CCDF of In-Degree Full Distribution

(b) Power-Law/Lognormal Fit on CCDF of In-Degree Distribution Tail (xmin=100)

Figure 4: Power-Law and Lognormal Fit on Un-Weighted In-Degree Distribution (CCDF)

3.2 Node Strength Distribution

In the Sao Paulo commuting network, the weighted in-degree of a given node (node strength) is defined as the total number of individuals from all regions which commute to that node on a daily basis. The count of nodes with a given number of commuters initially rises to a maximum between 50,000 and 100,000 commuters, before a strong right skew and outliers which receive a substantively larger daily inflow of commuters (Figure 5)

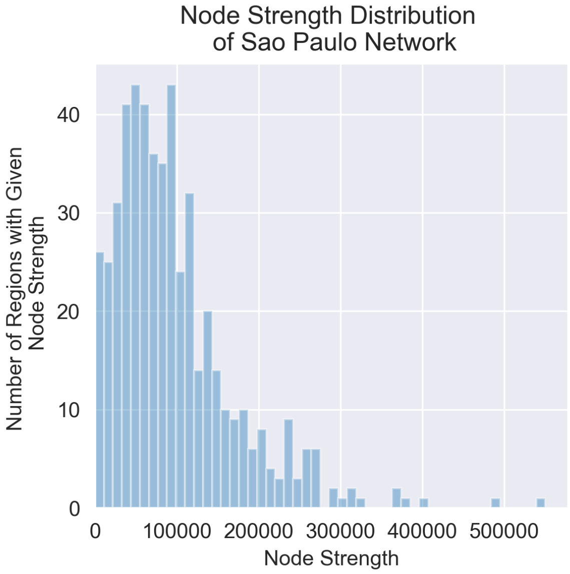


Figure 5: Node Strength Distribution (Total Daily Commuters to a Region)

Summary statistics on the node strength distribution are presented in Table 4. The median inflow is 81,865 daily commuters, with a standard deviation of 75,181. The region with the most inflow has 550,320 daily commuters, over 6 standard deviations from the mean. The 6 regions with the greatest daily inflow, and their municipalities, are presented in Table 5.

	MASP Node Strength
min	124.908355
median	81865.250093
max	550320.773386
std	75181.881687

Table 4: Summary Statistics of Node Strength Distribution (Total Daily Commuters to a Region)

Region ID	**Node Strength**	**Municipality**
82	550320.77	Sao Paulo
422	487018.11	Santo André
93	403274/08	Sao Paulo
377	379749.06	Guarulhos
2	368763.71	Sao Paulo
436	329216.43	Mauá

Table 5: Regions with highest 6 node strength, including municipality

The correlation between a node's unweighted in-degree and node strength based on total commuting flow is 0.415. A scatterplot of unweighted in-degree versus node strength is presented in Figure 6. Evidently, topological edge connections are not entirely predictive of commuting flow. For this reason, it is especially important to understand the factors that influence edge weight, as inflow to a region cannot be explained solely by the physical structure of the network.

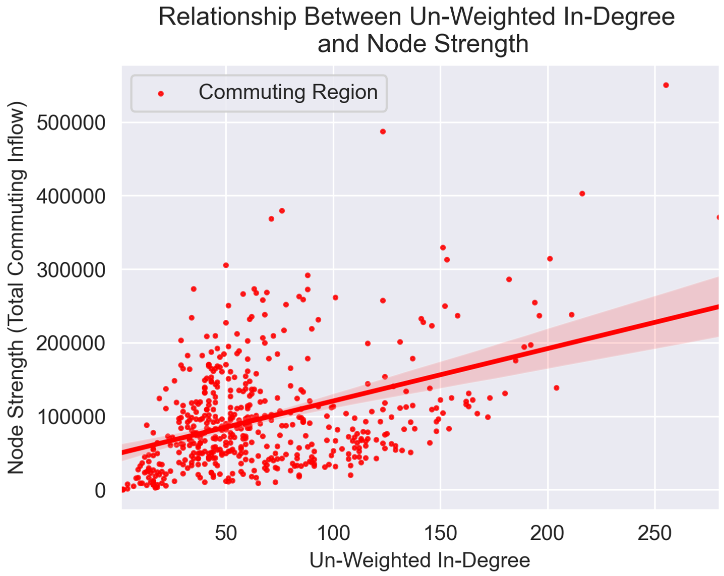
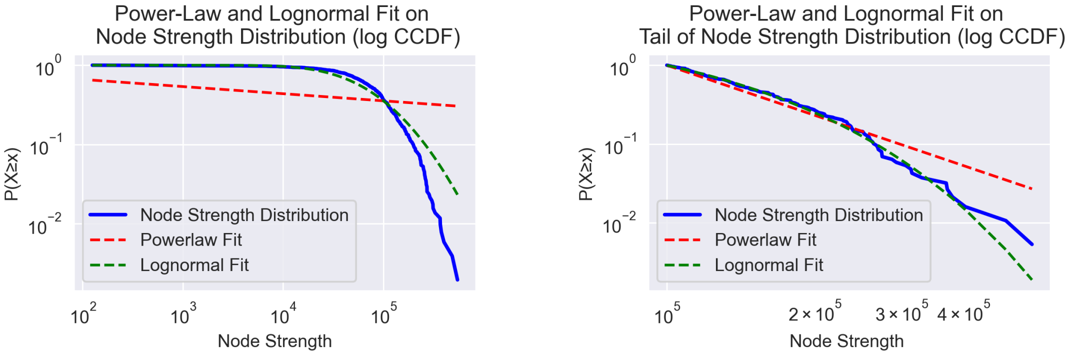


Figure 6: Plot of un-weighted in-degree versus node strength, including regression line

3.2.1 Test for Scale-Free Properties

In the context of the inflow of commuters to a region, the existence of a power-law distribution would provide evidence of a predictable self-organizing property, where nodes with higher commuting flows are more likely to receive further commuters. The likelihood ratio test between power-law and lognormal distributions has a p-value below 0.01 in favour of the lognormal fit, shown in Figures 7(a) and 7(b).

This analysis provides evidence that the un-weighted in-degree distribution and node strength distribution are not organized by a general power-law phenomenon. This further justifies an analysis of the factors specific to the network which may influence its structure. (Barabási & Albert, 2019).



(a) Power-Law/Lognormal Fit on CCDF of Node Strength Full Distribution

(b) Power-Law/Lognormal Fit on CCDF of Node Strength Tail Distribution (xmin = 10,000)

Figure 7: Power-Law and Lognormal Fit on Node Strength Distribution (CCDF)

4 Multiple Regression Quadradic Assignment Procedure Analysis

4.1 Methodology

This section uses a Multiple Regression Quadradic Assignment Procedure (MRQAP) analysis to determine the specific contributions of edge characteristics to the daily commuting flow between two regions (Lee, Lee, & Sohn, 2016). A standard ordinary least squares (OLS) procedure is inadequate for assessing the factors contributing to a node-to-node relationship, as OLS assumes independent and identically distributed observations (Lee et al., 2016). To circumvent this, a p-value and standard error of the regression coefficient are calculated by permutating the dependent variable matrix 1000 times – a Monte Carlo sampling method. The same permutations are used for rows and columns to preserve dependence among elements in the same row or column, while eliminating the relationship between the dependent and independent variables (Simpson, 2001). Each randomly permutated instance is considered as a null-hypothesis scenario, and coefficients are determined for each one. The two-sided p-value is significant at the α=0.05 level if the proportion of null coefficients as extreme or more extreme (on either side) to the observed coefficient is less than $\alpha/2 = 0.025$ (Farine, 2017; Dekker, Krackhardt, & Snijders, 2007; Simpson, 2001)

The dependent variable is the count of individuals flowing from an origin region to a destination region. The first independent variable is the mean commuting time between two regions, as reported in the commuting survey. Second, a binary variable captures whether the regions are in the same municipality, due to the propensity of commuters to work within the municipality in which they live (Schéele & Andersson, 2018). The final independent variable is the population density ratio (P^{od}_{ratio}), calculated for an origin o and destination d using:

$$P^{od}_{ratio} = \frac{(\frac{pop^o}{area^o})}{(\frac{pop^d}{area^d})} \tag{3}$$

A population density ratio below one is interpreted as a commuting flow from a less dense region to a denser one, and vice versa for a population density above 1. The population density ratio is designed to capture the relationship between travel demand and urban density, based on literature showing that jobs and are spatially concentrated in population centers (Guerra, Caudillo, Monkkonen, & Montejano, 2018).

4.2 Results

The mean-time coefficient results are presented in Figure 8. The observed coefficient in the regression is -15.47. The other observations in the distribution are the permutated null coefficients, with a minimum of -6.79, median of -0.12, and maximum of 7.00. All 1000 simulated coefficients are greater than the observed coefficient, therefore the p-value is 0.00, meeting the statistical significance criteria.

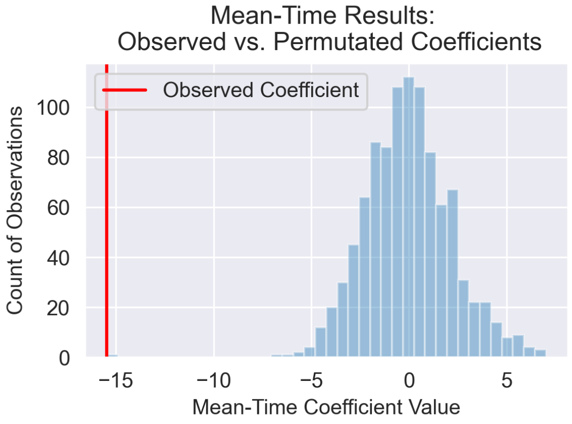


Figure 8: Mean-Time MRQAP Results, Vertical Line is Observed Coefficient

The dummy variable coefficient representing regions in the same municipality is 200.06. The minimum null scenario coefficient is -492.75, the median is 1.06, and the maximum is 458.30, as displayed in Figure 9. The proportion of values greater than or equal to the observed value is $110/1000 = 0.11$, failing the test of statistical significance at the $\alpha = 0.05$ level.

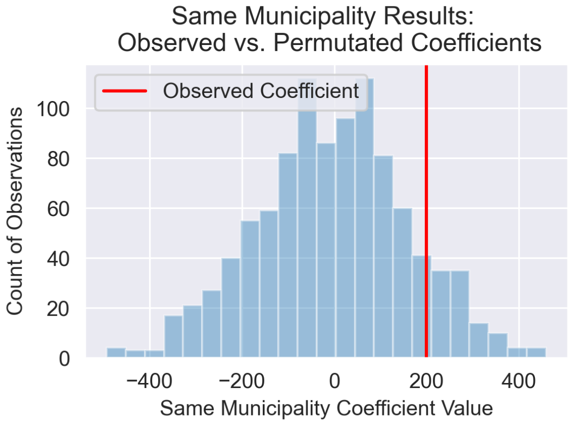


Figure 9: Within Same Municipality MRQAP Results, Vertical Line is Observed Coefficient

The results from the MRQAP on the population density ratio are presented in Figure 10. The observed coefficient from the commuting network is -1.36. The median coefficient of the null distribution is -0.40, with a minimum of -8.21 and maximum of 37.65. Of the 1000 generated coefficients, 319 are as low or lower than the population density ratio coefficient. Therefore the p-value of the observed coefficient is $319/1000 = 0.319$, failing the significance test at the α =0.05 level.

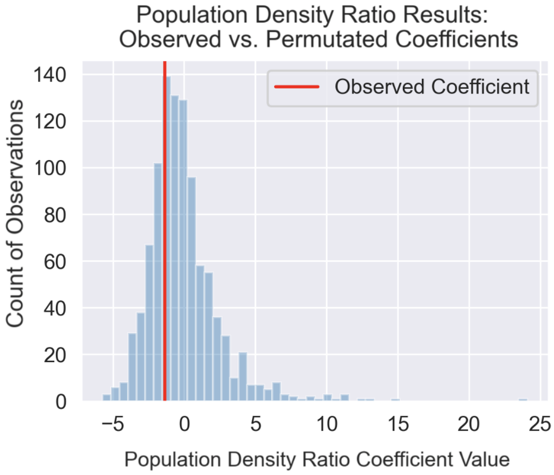


Figure 10: Population Density Ratio MRQAP Results, Vertical Line is Observed Coefficient

Indep. Var	**Min**	**Median**	**Max**	**SD**	**Observed Coef.**	**P**
Mean Time *	-6.79	-0.12	7.00	2.17	-15.47	0.00
Same Municipality	-492.75	1.06	458.30	163.15	200.06	0.11
Pop. Density Ratio	-8.21	-0.40	37.65	3.41	-1.36	0.319

Table 6: Summary of MRQAP Results – * Significant at $\alpha = 0.05$

5 Assessing Disease Risk with Betweenness Centrality

Betweenness centrality is a node-level measure of influence, defined by the extent to which a node intermediates flow over a network – a useful tool to determine regions at greatest risk for disease exposure from other commuting areas (Kivimäki, Lebichot, Saramäki, & Saerens, 2016; Tsiotas & Polyzos, 2015). The betweenness centrality of vertex i, denoted as $C_{\mathrm{B}}(i)$, is calculated as:

$$C_{\mathrm{B}}(i) = \frac{g_{jk}(i)}{g_{jk}}, \tag{4}$$

where $g_{jk}(i)$ is the number of shortest paths with endpoints j and k, that contain vertex i, and g_{jk} is the total number of shortest paths between j and k (Raghavan Unnithan, Kannan, & Jathavedan, 2014; Opsahl et al., 2010)

In an un-weighted betweenness centrality measure, path length is determined by the number of edges between nodes. In a weighted network, the sum of the weights, or 'cost', along a given path determines its length, known as Dijkstra's algorithm (Opsahl et al., 2010). The weighted portion of this analysis seeks to assess betweenness centrality using tie strength on a path, rather than a cost determined simply by the sum of weights. For this to conform to a shortest path analysis, the population flow of each edge must be inversed, so that strong population flow edges have a lower cost on the path (Opsahl et al., 2010).

A limitation of a simple weighted betweenness centrality is that it fails to incorporate the number of intermediary nodes on the shortest path, focusing only on the weights along a path. In the context of disease spread, this an important aspect to consider, as both tie strength *and* the number of intermediate nodes along a path can be conducive to transmission (Opsahl et al., 2010). The Opsahl method overcomes this limitation by incorporating both the number of intermediating nodes and the tie strength with a tuning parameter α, used as an exponent on the edge weight (Opsahl et al., 2010; Yustiawan, Maharani, & Gozali, 2015). The shortest path $d^{w\alpha}(i, j)$, with inversed edge weights and including the alpha parameter, is defined as:

$$d^{w\alpha}(i,j) = \min\left(\frac{1}{(w_{ih})^{\alpha}} + \ldots + \frac{1}{(w_{hj})^{\alpha}}\right) \tag{5}$$

The updated betweenness centrality becomes:

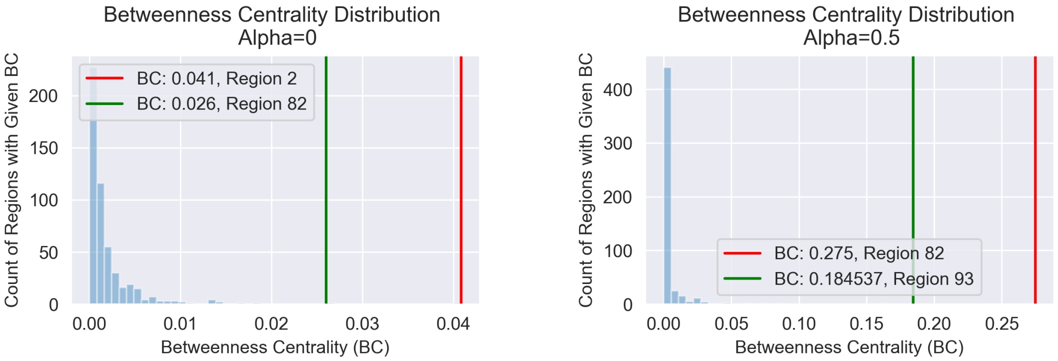
$$C_{\mathrm{B}}^{w\alpha}(i) = \frac{g_{jk}^{w\alpha}(i)}{g_{jk}^{w\alpha}} \tag{6}$$

An alpha parameter of 0 sets all edge weights to one, which is simply the unweighted betweenness centrality, based on the number of edges between endpoints. When $\alpha = 1$, Dijkstra's algorithm is used, by calculating the lowest sum of inversed population flow along a path. With an alpha between 0 and 1, the shortest path is determined by balancing the aggregate strength while favouring stronger paths with less diversity as the parameter approaches 0. An alpha parameter above 1 more heavily weights the commuting flow of the path, and tends to favor more intermediary nodes and weaker, diverse paths (Opsahl et al., 2010). Summary statistics for the betweenness centrality distributions calculated with each α parameter are presented in Table 7.

α	**0**	**0.5**	**1**	**1.5**
min	0	0	0	0
median	0.0009	0.0002	0.0001	0.0001
max	0.4	0.275	0.39	0.436
SD	0.003	0.018	0.026	0.032

Table 7: Betweenness Centrality Summary Statistics for all α Parameters

Figures 11(a) and 11(b) display the betweenness centrality distributions with α equal to 0 and 0.5. The two largest outliers in each distribution are highlighted with vertical lines. Only observing the topological structure of the network ($\alpha = 0$), nodes 2 and 82 are involved in 4.1% and 2.6% of all shortest paths, compared to an overall median centrality of 0.09%. With $\alpha = 0.5$, strong, short population flow paths are favoured, and regions 82 and 93 are identified as the greatest outliers. They are part of 27.5% and 18% of all shortest paths, respectively, while the median is 0.01%.

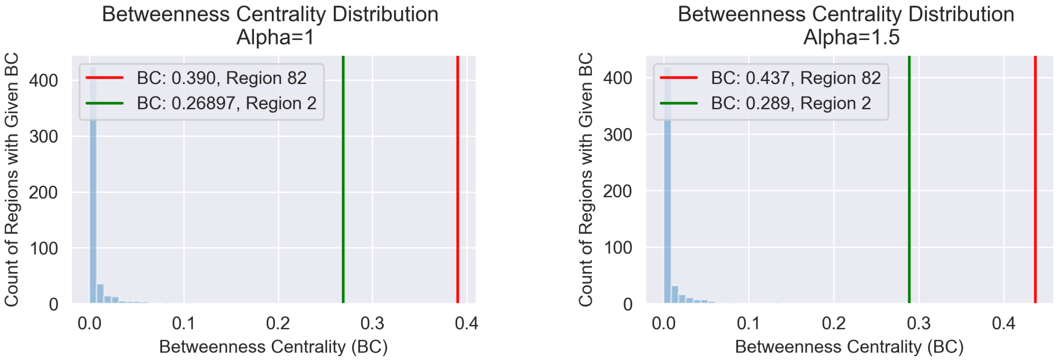


(a) Betweenness Centrality Distribution with $\alpha = 0$

(b) Betweenness Centrality Distribution with $\alpha = 0.5$

Figure 11: Betweenness Centrality Distribution with Opsahl Tuning Parameter $\alpha = 0$ and $\alpha = 0.5$

Using the path where inversed commuting flow defines length ($\alpha = 1$), nodes 82 and 2 have the highest centrality. They take part in 39% and 26% of all shortest paths, respectively, while the median centrality is 0.01%. Defining the shortest path in terms of tie strength, while favouring diverse intermediary nodes ($\alpha = 1.5$), regions 82 and 2 are of particular influence, taking part in 43.7% and 28.9% of all shortest paths, compared to a median of 0.01%.



(a) Betweenness Centrality Distribution with $\alpha = 1$

(b) Betweenness Centrality Distribution with $\alpha = 1.5$

Figure 12: Betweenness Centrality Distribution with Opsahl Tuning Parameter $\alpha = 1$ and $\alpha = 1.5$

From a topological perspective ($\alpha = 0$), this analysis has identified regions 82 and 2 as the most influential based on betweenness centrality. Where shorter, stronger ties are favored ($\alpha = 0.5$), regions 82 and 93 have the highest influence. With $\alpha = 1$, defining tie length as the sum of inverted commuting flow, regions 82 and 2 have the greatest role as intermediaries. Finally, with an $\alpha = 1.5$, which favors, weaker, diverse paths, 82 and 2 are the largest outliers in betweenness centrality. These three nodes, all present in the Sao Paulo municipality, are identified in Table 8 and Figure 13.

	α			
Region Ranking	**0**	**0.5**	**1**	**1.5**
1	2	82	82	82
2	82	93	2	2

Table 8: Ranking Regions with Highest 2 Betwenness Centrality Measures, Separated by α Value

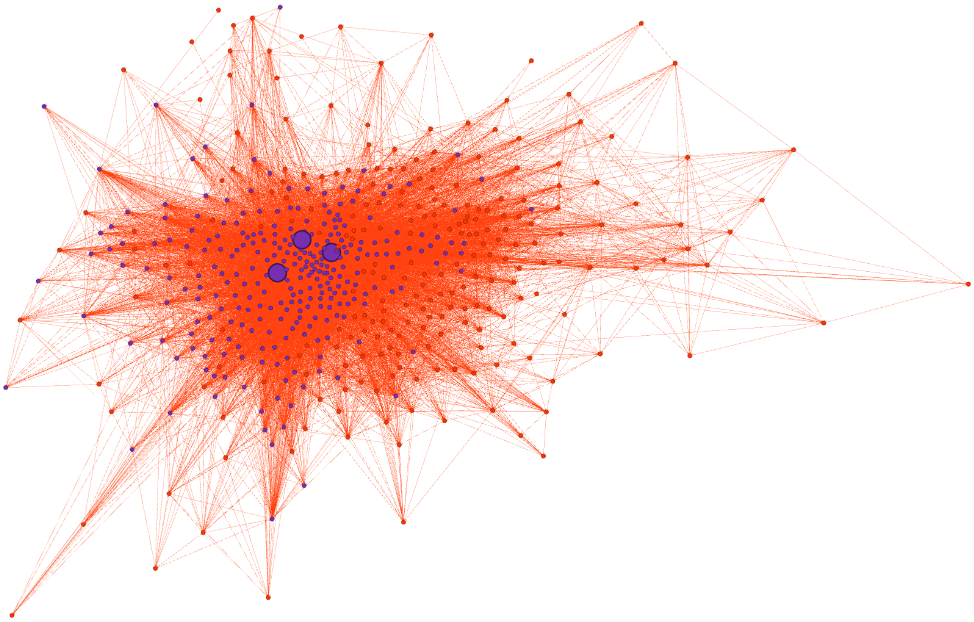


Figure 13: Regions with Highest Betweenness Centrality; **Left – 82**, with α=0, 0.5, 1, 1.5; **Top – 93** with α=0.5; **Right – 2** with α=0, 1, 1.5; Blue Nodes are Nearest Neighbors.

6 Conclusion

This commuting network analysis seeks to describe and understand the mechanisms behind commuting flow in the Metropolitan Area of Sao Paulo. In the context of a pandemic, understanding the characteristics between regions which influence daily commuting can help to identify regions most at risk to exposure. The in-degree distribution, representing the diversity of locations which flow to a given region, and the node strength, capturing the aggregate daily flow to a region, are not best described by a power-law. This is an important finding as it communicates that the mechanisms behind these primary commuting statistics do not follow a general phenomenon found in many large networks (Barabási & Albert, 2019).

The results from the MRQAP analysis illustrate that when controlling for regions in the same municipality and the population density ratio between two regions, the mean commuting time has a statistically significant negative association with greater commuting flow. While useful as controlling variables, the MRQAP analysis indicates that being in the same municipality and population density ratio variables have a statistically insignificant relationship ($\alpha = 0.05$) to the observed commuting flow edges between regions.

The betweenness centrality section analysis aims to identify the commuting areas most at risk for disease spread by determining the most influential nodes in the network. Using various adjustments of Opsahl's tuning parameter α, regions 82, 2, and 93 from the Sao Paulo municipality are the regions that disproportionately serve as intermediaries in the network. These 3 regions and their neighbors may serve as target areas for public health authorities to monitor disease evolution. Future research including information on COVID-19 transmission would do well to assess how different values for α in Opsahl et. al's betweenness centrality formula relates to the disease burden of a region.

References

Alstott, J., Bullmore, E., & Plenz, D. (2014). Powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE*, *9*(1). doi: 10.1371/journal.pone.0085777

Barabási, A. L., & Albert, R. (2019). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512.

Bloodgood, J. M., Hornsby, J. S., Rutherford, M., & McFarland, R. G. (2017). The role of network density and betweenness centrality in diffusing new venture legitimacy: an epidemiological approach. *International Entrepreneurship and Management Journal*, *13*(2), 525–552. Retrieved from `https://doi.org/10.1007/s11365-016-0412-9` doi: 10.1007/s11365-016-0412-9

Candido, D. S., Claro, I. M., de Jesus, J. G., Souza, W. M., Moreira, F. R., Dellicour, S., … Faria, N. R. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*, *369*(6508), 1255–1260. doi: 10.1126/SCIENCE.ABD2161

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2020). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*(January). Retrieved from `http://dx.doi.org/10.1038/s41586-020-2923-3` doi: 10.1038/s41586-020-2923-3

COMPANHIA DO METROPOLITANO DE SÃO PAULO (METRÔ-SP). (2017). Pesquisa Origem Destino 2017: A mobilidade urbana da Região Metopolitana de São Paulo em detalhes. , 136. Retrieved from `http://www.metro.sp.gov.br/pesquisa-od/`

Dekker, D., Krackhardt, D., & Snijders, T. A. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, *72*(4), 563–581. doi: 10.1007/s11336-007-9016-1

de Souza, W. M., Buss, L. F., Candido, D. d. S., Carrera, J. P., Li, S., Zarebski, A. E., ... Faria, N. R. (2020). Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behaviour*, *4*(8), 856–865. Retrieved from `http://dx.doi.org/10.1038/s41562-020-0928-4` doi: 10.1038/s41562-020-0928-4

Farine, D. R. (2017). A guide to null models for animal social network analysis. *Methods in Ecology and Evolution*, *8*(10), 1309–1320. Retrieved from `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12772` doi: https://doi.org/10.1111/2041-210X.12772

Guerra, E., Caudillo, C., Monkkonen, P., & Montejano, J. (2018). Urban form, transit supply, and travel behavior in Latin America: Evidence from Mexico's 100 largest urban areas. *Transport Policy*, *69*, 98–105. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0967070X17305930` doi: https://doi.org/10.1016/j.tranpol.2018.06.001

Hauge, S. H., & Meijerink, H. (2020). *Urbanization and preparedness for outbreaks with high-impact respiratory pathogens* (Tech. Rep. No. June). Norwegian Insititute of Public health.

Kivimäki, I., Lebichot, B., Saramäki, J., & Saerens, M. (2016). Two betweenness centrality measures based on Randomized Shortest Paths. *Scientific Reports*, *6*(February), 1–15. doi: 10.1038/srep19668

Lee, W. J., Lee, W. K., & Sohn, S. Y. (2016). Patent network analysis and quadratic assignment procedures to identify the convergence of robot technologies. *PLoS ONE*, *11*(10), 1–16. doi: 10.1371/journal.pone.0165091

Lewis, T. G. (2009). Random Networks. *Network Science*, 97–130. doi: 10.1002/9780470400791.ch4

Marques, G., Giulio, D., Bedran-martins, A. M. B., Vasconcellos, P., Costa, W., & Carmen, M. (2018). Mainstreaming climate adaptation in the megacity of São Paulo , Brazil. *Cities*, *72*(January 2017), 237–244. Retrieved from `https://doi.org/10.1016/j.cities.2017.09.001` doi: 10.1016/j.cities.2017.09.001

Mateos, G. (2016). *Degrees, Power Laws and Popularity Degree distributions Degree distributions* (Tech. Rep.). University of Rochester.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. Retrieved from `http://dx.doi.org/10.1016/j.socnet.2010.03.006` doi: 10.1016/j.socnet.2010.03.006

Peixoto, P. S., Marcondes, D., Peixoto, C., & Oliva, S. M. (2020). Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS ONE*, *15*(7 July), 1–23. doi: 10.1371/journal.pone.0235732

Raghavan Unnithan, S. K., Kannan, B., & Jathavedan, M. (2014). Betweenness Centrality in Some Classes of Graphs. *International Journal of Combinatorics*, *2014*, 241723. Retrieved from `https://doi.org/10.1155/2014/241723` doi: 10.1155/2014/241723

Ribeiro, H. V., Sunahara, A. S., Sutton, J., Perc, M., & Hanley, Q. S. (2020). City size and the spreading of COVID-19 in Brazil. *PLoS ONE*, *15*(9 September 2020), 1–12. doi: 10.1371/journal.pone.0239699

Schéele, S., & Andersson, G. (2018). Municipality attraction and commuter mobility in urban Sweden: An analysis based on longitudinal population data. *Urban Studies*, *55*(9), 1875–1903. Retrieved from `https://doi.org/10.1177/0042098017705829` doi: 10.1177/0042098017705829

Simpson, W. (2001). The Quadratic Assignment Procedure (QAP). *Stata Users Group, North American Stata Users' Group Meetings 2001*.

Tsiotas, D., & Polyzos, S. (2015). Introducing a new centrality measure from the transportation network analysis in Greece. *Annals of Operations Research*, *227*(1), 93–117. doi: 10.1007/s10479-013-1434-0

Yustiawan, Y., Maharani, W., & Gozali, A. A. (2015). Degree Centrality for Social Network with Opsahl Method. *Procedia Computer Science*, *59*, 419–426. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1877050915020888` doi: https://doi.org/10.1016/j.procs.2015.07.559