Question 1

Section 1: Measures

The three measures used to rank the stations (nodes) of the London Underground are Betweenness, Degree, and Eigenvector Centrality.

Betweenness Centrality for a node i is a measure of the number of shortest paths (g) between all node pairs in a graph that pass through node i; it is a measure of how important node i is in connecting other nodes in the graph

$$bet_i = \sum_{jk} \frac{g^i_{jk}}{g_{jk}} \tag{1}$$

As there may be more than one shortest path between nodes j and k, ${g^i_{jk}}/{g_{jk}}$ gets the fraction of these shortest paths that pass through node i (Bloch, Jackson, and Tebaldi 2019).

Degree Centrality is a measure of the number of immediate connections a node has. In a directed graph, the in degree and out degree are different. Degree centrality is useful for capturing the local importance of a node (ibid).

Eigenvector centrality is able to capture the global importance of a node by basing its score as a function of the importance of its neighbors (Bonacich 1987). It is a recursive measure where score is based on connections to high-scoring nodes.

To quantify changes in the network behavior, the two metrics used by Berche et al. (2009) to study transport network resilience are chosen.

1. ***Mean Inverse Shortest Path Length $\langle l^{-1} \rangle$***

$$\langle l^{-1} \rangle = \frac{1}{N(N-1)} \sum_{j \in V} \sum_{k \neq j \in V} \frac{1}{g_{jk}} \tag{2}$$

Equation (2) is from (Holme et al. 2002). g_{jk} is the geodesic between node j and k, and $N(N-1)$ is the number of node pairs in the graph. The Mean Shortest Path is an average of shortest paths between all node pairs in a graph; it measures the efficiency of flow in a network (ibid). However, as nodes are removed and the graph breaks up into disconnected subgraphs, the average geodesic length becomes infinity. The inverse shortest path is used instead to avoid this problem (Berche et al. 2009; Holme et al. 2002). As the average path length is also a measure of the flow of information through a network, it is not specific to transport network analysis; it is also used to study social networks, the internet, and biological networks (Asif et al. 2014).

2. ***Normalized Largest Connected Component $\langle S \rangle$***

$$\langle S \rangle = \frac{N_1}{N} \tag{3}$$

Equation (3) is from (Berche et al. 2009). A connected graph is one where all nodes are connected; you can travel through the graph to any node (Barabási 2016). If this is not the case, then the network is made up of multiple disconnected components. This metric shows how the network splits up as nodes are removed. N_1 is the number of nodes in the Largest Connected Component and N is the total number of nodes in the graph, and so as the network splits up, $\langle S \rangle$ gets smaller.

This metric is related to percolation theory (Barabási 2016) and useful in observing how the network breaks down (Berche et al. 2009). It can be used to analyze the vulnerability of different types of networks to attacks, and to understand the impact of random and targeted attacks on scale-free networks such as the internet (Barabási 2016).

Section 2: Analysis

Nodes are removed based on their centrality measures. In the first case (iterative), the centrality measure is recalculated after each node removal and nodes are ranked again before the next deletion. In the second case (All at Once – AAO) nodes are deleted based on the initial centrality ranking. For both cases, $\langle l^{-1} \rangle$ and $\langle S \rangle$ are calculated after each node deletion.

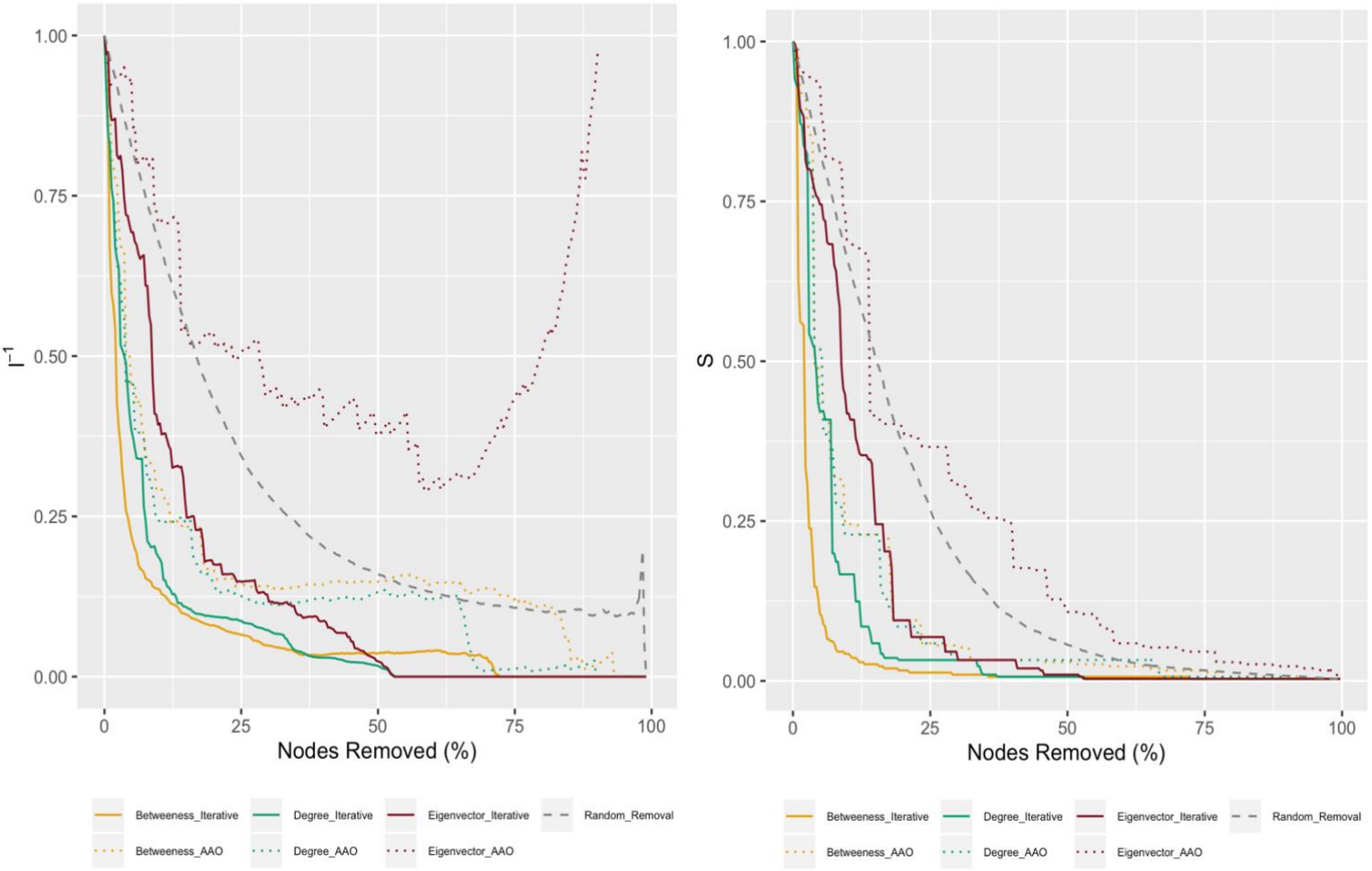

Figure 1: Change in (LEFT) Mean Inverse Shortest Path (l^{-1}) (normalized by value before any node is removed) and (RIGHT) Largest Connected Component (S) as nodes are removed based on different strategies

For all three centrality measures used, recalculating the measure in between node deletions has a bigger impact on the network structure compared to deletion based on initial node rankings. Holme et al. (2002) explain that recalculating centrality measures in between node removals is more harmful to the network and so the results are not surprising.

Random node removal[1] shows a more gradual decline in both $\langle l^{-1} \rangle$ and $\langle S \rangle$. It is clear that targeted attacks are more effective than random ones at breaking down the network.

Comparing $\langle l^{-1} \rangle$ and $\langle S \rangle$, we see that they provide similar results with the different centrality measures having the same impact relative to one another. In both cases, removal based on iterative betweenness centrality is most effective, closely followed by iterative degree centrality.

Section 3: Discussion

Scale-free networks are known to be especially vulnerable to targeted attacks but robust against random attacks (Barabási 2016; Callaway et al. 2000). Given that the degree distribution of stations in London's public transport network exhibit power law behavior (von Ferber et al. 2009), it makes sense for random attacks to have significantly less impact on it than targeted attacks. Targeted attacks based on betweenness and degree centrality appear to be more effective than those based on eigenvector centrality, as can be seen in both the $\langle l^{-1} \rangle$ and $\langle S \rangle$ graphs.

Comparing betweenness and degree, we can see that the difference between the iterative and AAO approaches is more pronounced in the former. Betweenness centrality is a global measure while degree is a local measure; studies have shown that iterative recalculation of global measures has a more pronounced effect on network breakdown than that based on local measures (Berche et al. 2009; Williams and Musolesi 2016).

While it could be assumed that degree centrality may not be a good indicator of node importance, given that it does not capture important aspects of a networks architecture (Bloch, Jackson, and Tebaldi 2019), the results show otherwise. Node removal based on degree was more effective than that based on eigenvector values and almost as effective as that based on betweenness values.

The results also show that node degree and betweenness do not change dramatically as nodes are removed, as can be seen by the small difference in effect between the iterative and AAO curves (Figure 1). However, node eigenvector centrality appears to change considerably as nodes are removed, and the breakdown of the network based on AAO eigenvector node removal is even less effective than random node removal. This could be due to the recursive nature of Eigenvector centrality, where important nodes are those that are connected to other important nodes (Bloch, Jackson, and Tebaldi 2019). Some of the highest ranking nodes at the start are therefore not hubs, but stations connected to hubs, meaning that their removal is not as effective as the removal of nodes with high betweenness or degree centrality.

[1] 100 iterations of random node removal were carried out and the results were averaged

The analysis conducted is quite simple and more complex evaluation metrics would consider the effect on total passenger travel time, accessibility, or unsatisfied demand (Jafino, Kwakkel, and Verbraeck 2020).

The analysis can be used by TfL to determine which stations to prioritize for extra security. The London Underground has been shown to be vulnerable to attacks on hubs but resilient to random attacks (Chopra et al. 2016), and so it makes sense to direct more investment to hubs. Stations that both rank highly and do not have neighboring stations where passenger flow can be redirected (ibid) should be prioritized for resource allocation. Bus routes should also be integrated with the underground network to ensure that there is redundancy at critical stations and on critical routes (Jin et al. 2014). This utilitarian approach may however raise concerns regarding the fairness of resource distribution (Jafino, Kwakkel, and Verbraeck 2020).

Question 2

Section 1

Gravity models are based on Newton's theory of gravitation, where the force of attraction between two objects is proportional to their masses and inversely proportional to the distance separating them. Equation (4), from Wilson (1971) shows the general formulation of a gravity model:

$$T_{ij} = KW_i^{\alpha} W_j^{\beta} f(c_{ij}) \quad (4)$$

W_i and W_j are the attraction components of the origin and destination, normally represented by total flow out of i and total flow into j, and $f(c_{ij})$ is a deterrent function that is inversely proportional to c_{ij}, the generalized cost of travel between i and j. In the case of modeling refugees fleeing their country we can use the production constrained model. This allows the total flows originating from each geographic area O_i to be constrained to the actual observed flows, as shown in Equation (5).

$$\sum_j T_{ij} = O_i \quad (5)$$

A Poisson regression model, suitable for estimating non-negative integers would be used (Flowerdew and Aitkin 1982). If we have refugee migration data on a regional level, then a multilevel spatial interaction model (Dennett and Wilson 2013) could be used to predict the flows of refugees into different regions of each country.

Firth et al. (2019) used real data on Syrian Refugee migration and found the following characteristics of the destination country to be statistically significant in determining where refugees ended up: distance, average income, unemployment rate, language similarity and probability of being granted asylum. Breitenecker et al. (2017) also include a variable for state fragility and one for the existing number of migrants at the destination. While distance is included in the impedance function $f(c_{ij})$ of our chosen model, the rest need to be incorporated. These variables could be standardized, and their weighted average would represent the attraction component (W_j) of the different destinations (Breitenecker et al. 2017). Alternatively, some of them could be added separately to the numerator in equation (4). He and Pooler (2003) did this and found that, not only did the predictive power of the model improve significantly, but the effect of distance impedance is replaced by the other variables.

The proposed model is an aggregated model, and so is not ideal to capture complex phenomena where actions are dependent on one another. Suleimenova, Bell, and Groen (2017) note that "push-pull characteristics can be insufficient to explain forced migration". ABM does not have this limitation as it can incorporate different variables at the level of the individual and allow for emergent behavior to take place through embedded migrant routes and autonomous interaction between migrants (ibid).

The traditional gravity model also does not account for spatial dependance in the flows (Lesage and Fischer 2014), meaning that flow characteristics of neighboring areas are assumed to be independent. This is not the case in reality, as Griffith and Jones (1980) have shown that

distance decay is not independent of the spatial structure of a study area, and that spatial autocorrelation is significant when determining flows between origins and destinations.

Given restrictions on the number of refugees being allowed into different countries, we would have to use a production-attraction constrained model (Wilson 1971). In this model, the total number of people arriving at a country D_j is constrained to the actual imposed limits, as shown in Equation (6). Equation (5) would also still apply.

$$\sum_i T_{ij} = D_j \quad (6)$$

A limitation with this model is that it assumes that flow demand into the receiving countries exceeds the imposed limits, and that the limits will be met. This is a reasonable assumption, as can be seen with the large demand and long waiting times in the asylum process of Syrian refugees.

Section 2

A production constrained model is used to assess the impact of a change in salaries and a change in transport costs. This is based on the assumption that the same number of people still need to commute for work, but their destinations will change depending on the change in relative attractiveness of the different Boroughs. This is a valid assumption if the decrease in salaries is part of wider changes in employment numbers by industry (If we were to assume that the number of jobs in each Borough would remain fixed, then a doubly constrained model would be used).

When the average salary in the City of London decreases by 30%, we see a significant decrease (more than 100,000 people) of total flow into the Borough (Figure 2). This flow is redistributed among the other Boroughs, with the increase being more significant in high-paying Boroughs, such as Camden, Kensington, and Westminster.

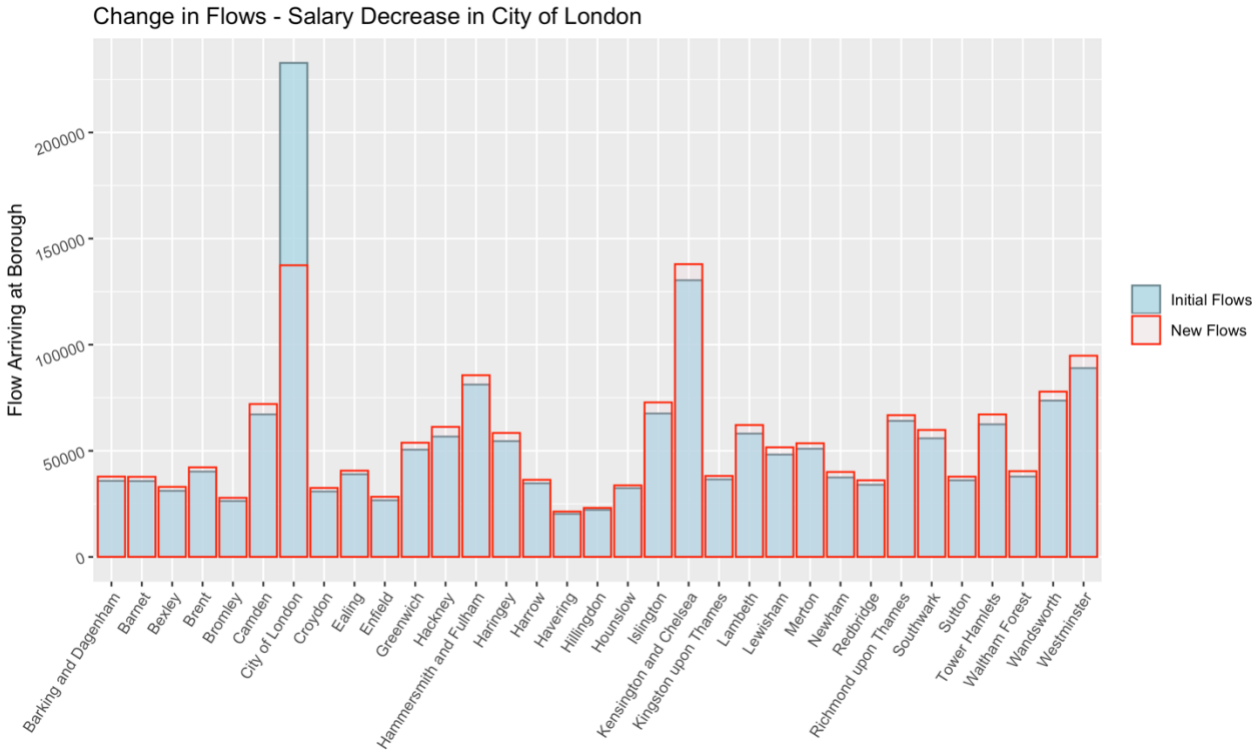

Figure 2: Results from a production constrained model after salary is reduced by 30% in the City of London

An increase in the cost of transport is modelled by doubling and tripling the original impedance factor (Beta), but the impact is different for each Borough (Figure 3) and does not correlate with salaries (Figure 5).

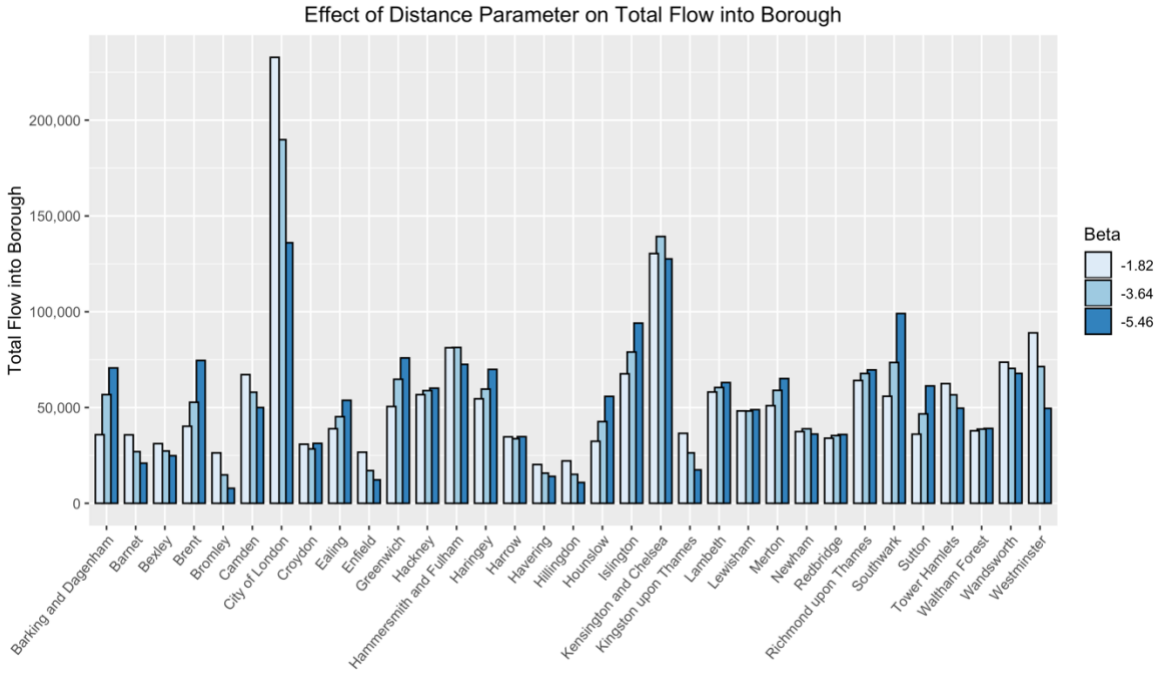

Figure 3: Modelled Flow Change Resulting From Changing Distance Impedance

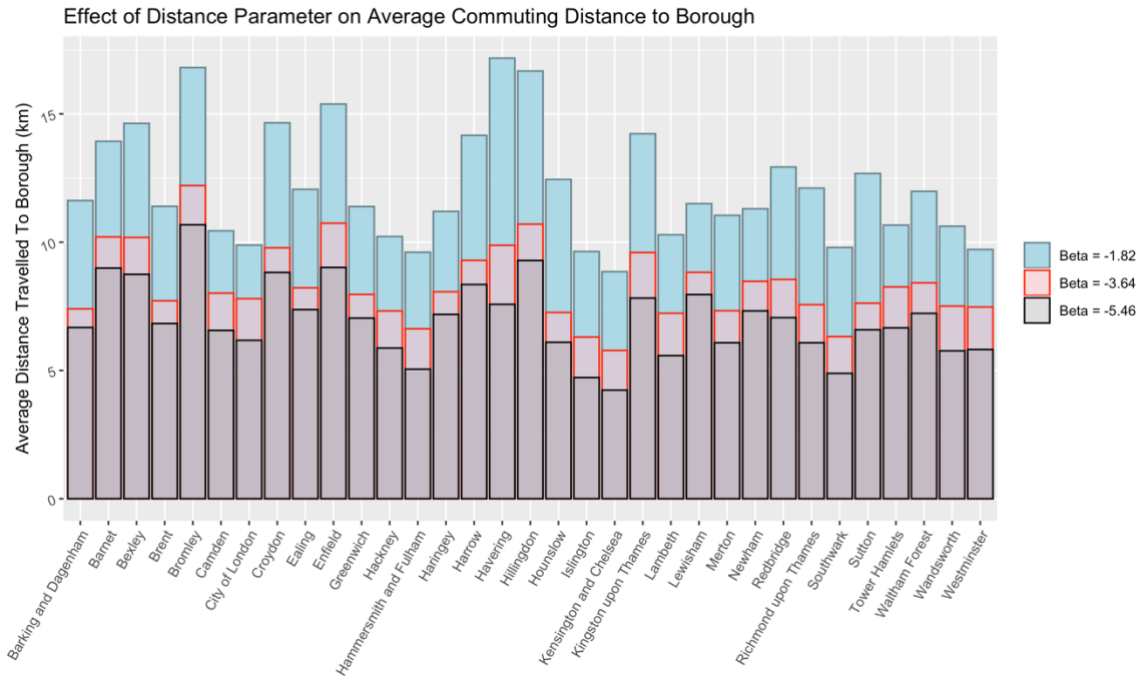

Figure 4: Average Distance Travelled to Borough at Different Distance Impedance Values

Looking at the average distance travelled to each Borough (Figure 4), we find that it decreases significantly when Beta is doubled, showing that people would travel less when journeys are more expensive. Tripling the Beta further decreases the average distance, but by a relatively smaller amount.

Doubling Beta has the least effect on average travel distance to Boroughs with high salaries such as Westminster, City of London and Tower Hamlets (Figure 4), but again there is little correlation between the salary and change in distance travelled to work (Figure 5).

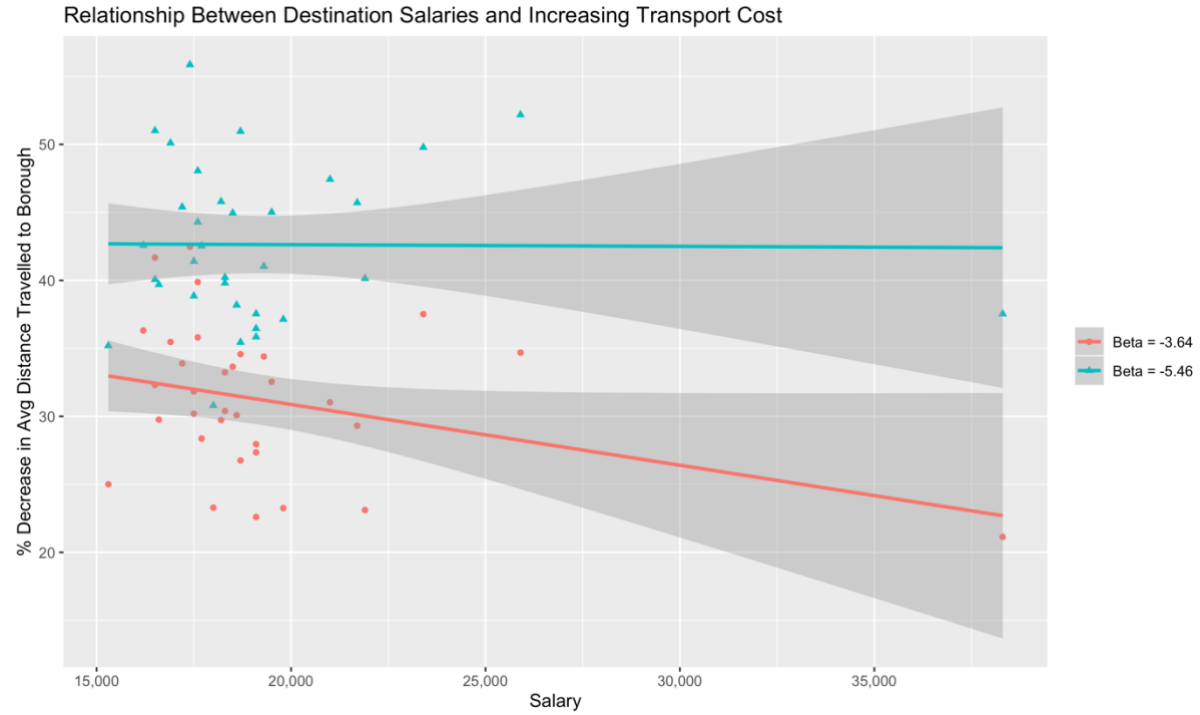

Figure 5: Comparing Salary to Reduction in Distance Travelled Due to Increase in Transport Cost. In both cases % decrease is calculated relative to original avg distance (Beta = -1.82)

If we were to focus on the City of London, we find that a 30% salary decrease almost halved the flows into it. An equivalent reduction in flows was achieved by quadrupling the distance decay parameter (Beta), indicating that a change in salary is more significant than one in transport cost.

Question 3

Section 1: CA vs. ABM.

Cellular Automata (CA) and Agent-Based Models (ABM) are bottom-up, disaggregated models that model at the individual level (Clarke 2014). In both cases, aggregate patterns "emerge" from the individual behavior, which itself requires calibration and validation (ibid).

CA are one of the simplest models where emergence can be shown. They study areas that can be represented on a grid and are preferable if the probability of state transition of the cells is known. A CA is made up of 1) a grid; 2) a definition of the neighborhood that effects each cell in the grid; 3) the initial condition of each cell; 4) rules that change cell condition based on condition of neighboring cells (Clarke 2014). The neighborhood of the cell is normally defined as the cells bordering it, but it can also be a larger catchment area, as done by White and Engelen (2016), who give cells within a specified radius a weight depending on their proximity to the origin cell.

ABMs model agents that move in space, and in doing so are able to show how the interaction between these agents affects the system (Clarke 2014). The decision of agents is based on "bounded rationality", where they act in their own interests. They are also able to learn from previous decisions (ibid). Agent decision-making is based on field work or domain knowledge, and one of the main challenges is to apply rules that accurately capture agent incentives and limitations in their-decision making (Bithell, Brasington, and Richards 2008).

Whereas CA results are normally spatial, ABM results also include aggregate statistics that show system behavior resulting from different rules and agent characteristics (Clarke 2014).

Section 2: ABM Description

To understand the interactions in the WolfSheepAgeing model, BehaviorSpace was used with the following parameters:

Table 1: Wolf-Sheep-Grass Model Parameters in BehaviorSpace

Index	***Parameter***	***Values***
1	wolf-gain-from-food	20 (default)
2	wolf-reproduce	5 (default)
3	Initial-number-of-wolves	10 35 60 85 110
4	Initial-number-of-sheep	60 90 120 150 180
5	sheep-gain-from-food	4 (default)
6	Grass-regrowth-time	20 40
7	Sheep-reproduce	4 (default)

In total, there were 50 parameter combinations, each run 20 times. In none of the combinations do we see an extinction of the sheep or wolves after 1000 steps, but there is significant variation in the final populations when the grass-regrowth-time is set at 20. At a grass-regrowth-time of 40, most of the combinations exhibit less variation in the final populations (Figure 6 - bottom).

We see that, by step 150[2], all combinations seem to have reached a steady state when the grass-regrowth-time is 40, with the number of wolves between 10 and 35 and the number of sheep between 150 and 180 (Figure 7 – bottom). The difference in the final populations of each animal is attributed to the initial populations in the model. When grass-regrowth-time is 20, the populations are still varying at step 150, with no indication of a steady state being reached. We see an increase in the sheep population followed by a spike in the wolf population which in turn leads to more sheep being hunted, and their numbers falling again.

[2] When recording values at each step (Figure 7), the number of steps is limited to 150 as exceeding it made the output too large for analysis in R. This is a limitation of the work

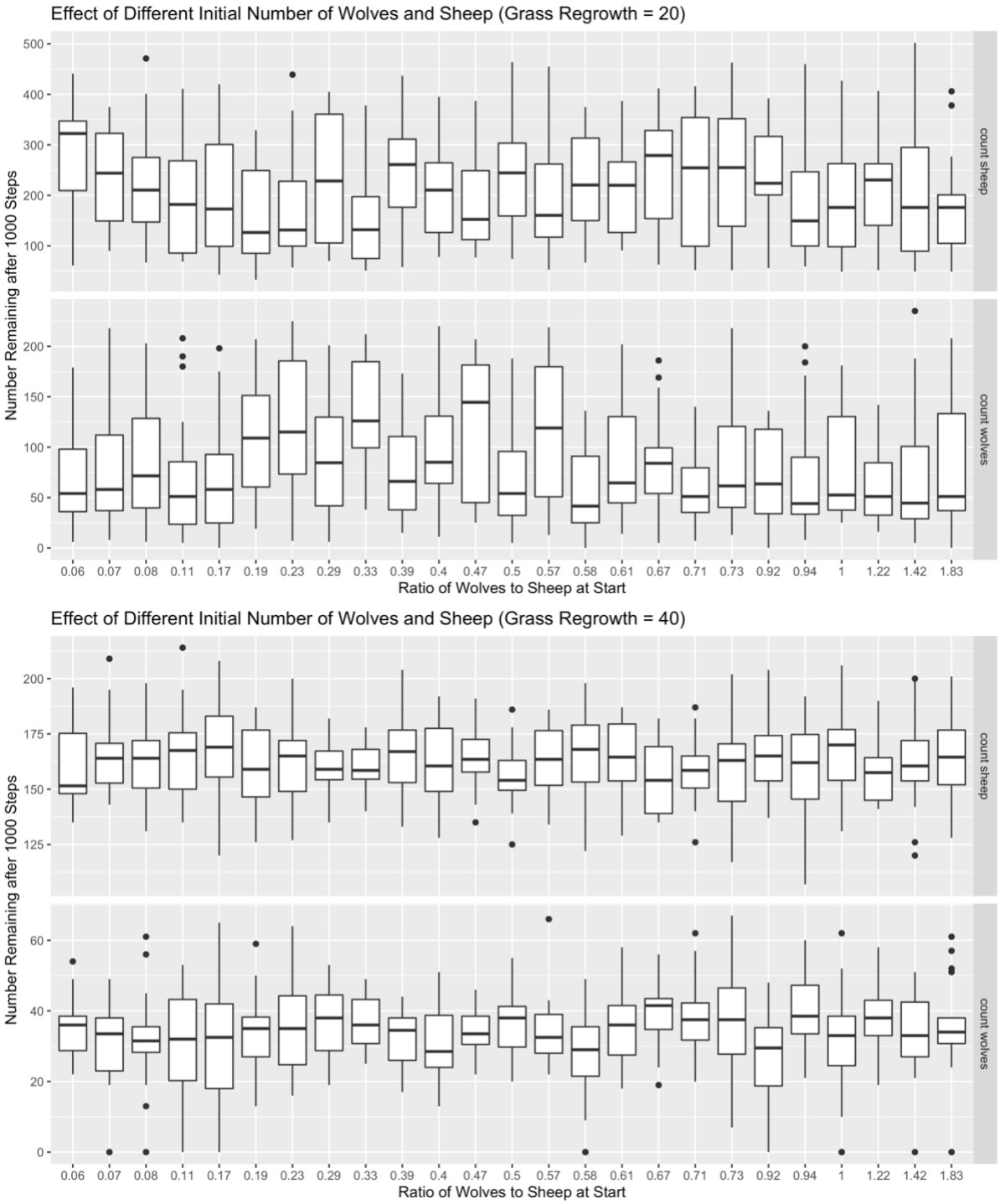

Figure 6: Simulation Results at Different Wolf/Sheep starting Ratios and grass-regrowth-time. TOP: Grass Regrowth = 20 and BOTTOM: Grass Regrowth = 40

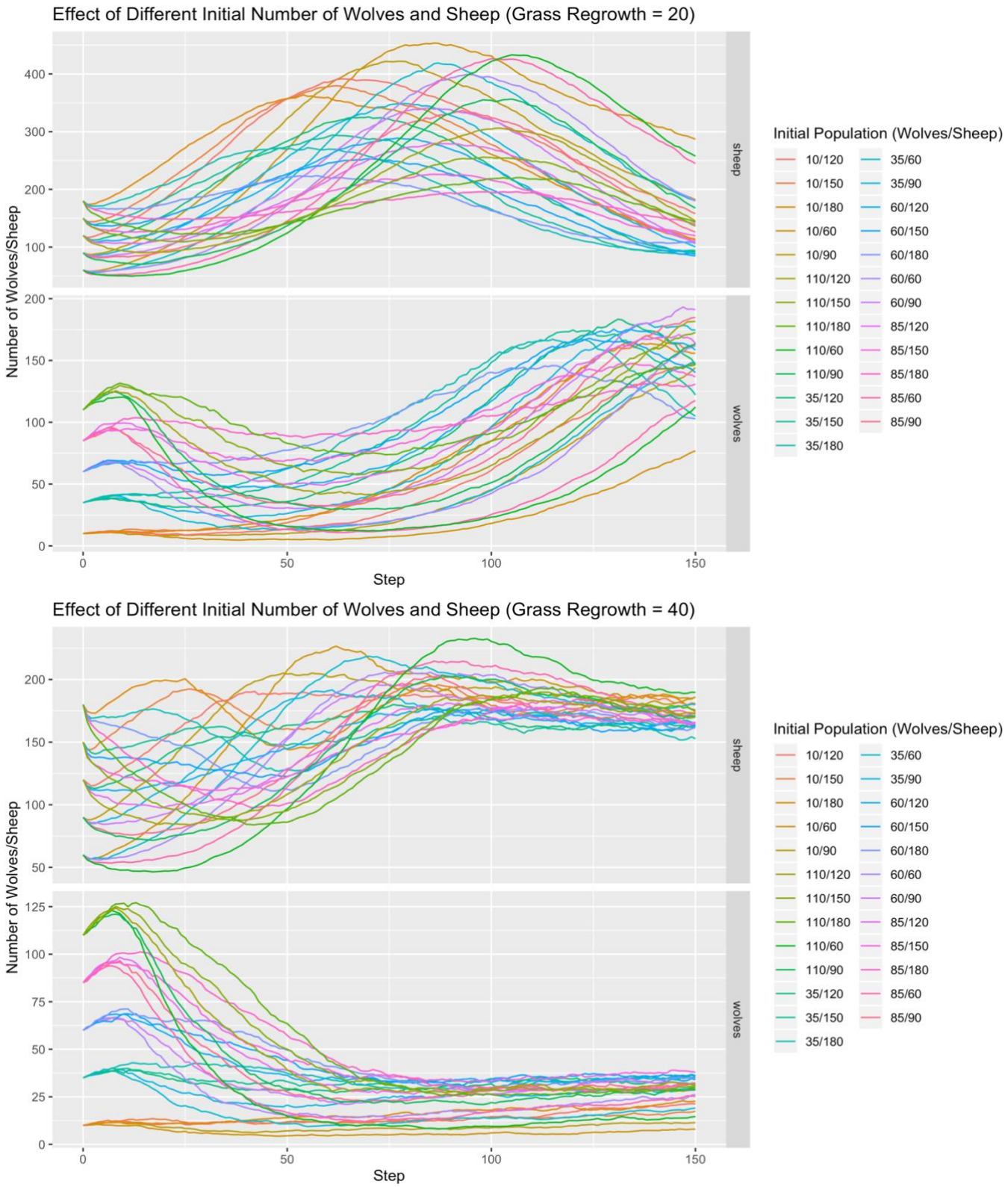

Figure 7: Number of Wolves/Sheep alive at different steps using different initial populations. TOP: Grass Regrowth = 20 and BOTTOM: Grass Regrowth = 40

Section 3: ABM Analysis

The higher grass-regrowth-time may be associated with less variation due to the sheep not having abundant amount of food; if they have less food then their numbers do not exhibit a pattern of increasing rapidly and then falling again when the grass is not sufficient to sustain the population increase. At grass-regrowth-time = 40, there appears to be a balance between the consumption needs of the wolves and the reproduction of the sheep, as the number of sheep is partially kept down by the lower availability of grass.

Figure 7 shows that the higher the initial ratio of wolves to sheep, the steeper the decline in the number of wolves, as wolves die out due to a lack of food. The wolf population stabilizes when the grass-regrowth-time is 40 but bounces back up when it is 20, as the number of sheep increase, and food becomes available to sustain a larger wolf population.

The number of runs is calculated based on the following equation:

$$n = \left(\frac{t*s}{ME}\right)^2 \quad (7)$$

Initial exploration showed that the parameter combination with the highest normalized standard deviation had a mean of ~178 and a standard deviation (s) of ~50 in the final number of sheep. Using a 10% margin of error (ME), we get n = 20. Each parameter combination was run 20 times. This is double the number of replications (10) specified by Olsen et al. (2018) to decrease outcome variation in this model to an acceptable level.

The limitations of such an analysis is that not all parameter combinations can be considered when trying to identify steady states and tipping points. Olsen et al. (2018) overcome this limitation by using a genetic algorithm to find the combinations that achieve a steady state; with that state being achieved if the standard deviation of predator and prey populations is consistently below a certain threshold in the last N steps of the model run.

Section 4: ABM Evaluation

The results show that the wolves-sheep-grass model is more sensitive to the abundance of grass than to the initial number of wolves and sheep. Nevertheless, initial figures have an effect on the number of wolves/sheep alive when the system reaches a steady state, a fact that could be used when monitoring wildlife populations, as was done by Sinclair et. al (1998) when modeling endangered marsupial numbers in Australia after their reintroduction to the wild.

As the populations are sensitive to the availability of grass, researchers could model the impact of climate change directly through varying that parameter in line with forecasts. The impact of increased CO_2 levels on water and soil quality and pest diseases has been researched using ABMs (Tubiello, Soussana, and Howden 2007). Integrating this research with predator/prey models would help better understand predator-prey dynamic within the larger ecosystem.

Bibliography

Asif, W., A. Qureshi, M. Iqbal, and M. Rajarajan. 2014. "On the Complexity of Average Path Length for Biological Networks and Patterns." *International Journal of Biomathematics* 7 (4). https://doi.org/10.1142/S1793524514500387.

Barabási, Albert-László. 2016. *Network Science*.

Berche, B., C. von Ferber, T. Holovatch, and Yu. Holovatch. 2009. "Resilience of Public Transport Networks against Attacks." *The European Physical Journal B* 71 (1): 125–37. https://doi.org/10.1140/epjb/e2009-00291-3.

Bithell, Mike, James Brasington, and Keith Richards. 2008. "Discrete-Element, Individual-Based and Agent-Based Models: Tools for Interdisciplinary Enquiry in Geography?" *Geoforum* 39 (March): 625–42. https://doi.org/10.1016/j.geoforum.2006.10.014.

Bloch, Francis, Matthew O. Jackson, and Pietro Tebaldi. 2019. "Centrality Measures in Networks." SSRN Scholarly Paper ID 2749124. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.2749124.

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92 (5): 1170–82. https://doi.org/10.1086/228631.

Breitenecker, Felix, Tamara Vobruba, Andreas Körner, and Niki Popper. 2017. "Pathways of Migrants and Refugees – a Simulation Approach." *SNE Simulation Notes Europe* 27 (December): 191–202. https://doi.org/10.11128/sne.27.tn.10394.

Callaway, Duncan S., M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. 2000. "Network Robustness and Fragility: Percolation on Random Graphs." *Physical Review Letters* 85 (25): 5468–71. https://doi.org/10.1103/PhysRevLett.85.5468.

Chopra, Shauhrat S., Trent Dillon, Melissa M. Bilec, and Vikas Khanna. 2016. "A Network-Based Framework for Assessing Infrastructure Resilience: A Case Study of the London Metro System." *Journal of The Royal Society Interface* 13 (118). https://doi.org/10.1098/rsif.2016.0113.

Clarke, Keith. 2014. "Cellular Automata and Agent-Based Models." In *Handbook of Regional Science*, 1217–33. https://doi.org/10.1007/978-3-642-23430-9_63.

Dennett, Adam, and Alan Wilson. 2013. "A Multilevel Spatial Interaction Modelling Framework for Estimating Interregional Migration in Europe:" *Environment and Planning A*, January. https://doi.org/10.1068/a45398.

Ferber, C. von, T. Holovatch, Yu. Holovatch, and V. Palchykov. 2009. "Public Transport Networks: Empirical Analysis and Modeling." *The European Physical Journal B* 68 (2): 261–75. https://doi.org/10.1140/epjb/e2009-00090-x.

Flowerdew, Robin, and Murray Aitkin. 1982. "A Method of Fitting the Gravity Model Based on the Poisson Distribution*." *Journal of Regional Science* 22 (2): 191–202. https://doi.org/10.1111/j.1467-9787.1982.tb00744.x.

Frith, Michael J., Miranda Simon, Toby Davies, Alex Braithwaite, and Shane D. Johnson. 2019. "Spatial Interaction and Security: A Review and Case Study of the Syrian Refugee Crisis." *Interdisciplinary Science Reviews* 44 (3–4): 328–41. https://doi.org/10.1080/03080188.2019.1670439.

Griffith, D. A., and K. G. Jones. 1980. "Explorations into the Relationship between Spatial Structure and Spatial Interaction:" *Environment and Planning A*, February. https://doi.org/10.1068/a120187.

He, J., and J. Pooler. 2003. "Modeling China's Province-to-Province Migration Flows Using Spatial Interaction Model with Additional Variables." *Geography Research Forum* 23 (January): 30–55.

Holme, Petter, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. 2002. “Attack Vulnerability of Complex Networks.” *Physical Review E* 65 (5). https://doi.org/10.1103/PhysRevE.65.056109.

Jafino, Bramka Arga, Jan Kwakkel, and Alexander Verbraeck. 2020. “Transport Network Criticality Metrics: A Comparative Analysis and a Guideline for Selection.” *Transport Reviews* 40 (2): 241–64. https://doi.org/10.1080/01441647.2019.1703843.

Jin, Jian Gang, Loon Ching Tang, Lijun Sun, and Der-Horng Lee. 2014. “Enhancing Metro Network Resilience via Localized Integration with Bus Services.” *Transportation Research Part E: Logistics and Transportation Review* 63 (March): 17–30. https://doi.org/10.1016/j.tre.2014.01.002.

Lesage, James, and Manfred Fischer. 2014. “Spatial Regression-Based Model Specifications for Exogenous and Endogenous Spatial Interaction.” *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.2420746.

Oldham, Stuart, Ben Fulcher, Linden Parkes, Aurina Arnatkeviciute, Chao Suo, and Alex Fornito. 2019. “Consistency and Differences between Centrality Measures across Distinct Classes of Networks.” *PLOS ONE* 14 (July): e0220061. https://doi.org/10.1371/journal.pone.0220061.

Olsen, Megan, Jessa Laspesa, and Theodore Taylor-D’Ambrosio. 2018. “On Genetic Algorithm Effectiveness For Finding Behaviors In Agent-Based Predator Prey Models.” In . https://doi.org/10.22360/summersim.2018.scsc.025.

Sinclair, A., R. Pech, Christopher Dickman, David Hik, P. Mahon, and A. Newsome. 1998. “Predicting Effects of Predation on Conservation of Endangered Prey.” *Conservation Biology - CONSERV BIOL* 12 (June): 564–75. https://doi.org/10.1046/j.1523-1739.1998.97030.x.

Suleimenova, Diana, David Bell, and Derek Groen. 2017. “A Generalized Simulation Development Approach for Predicting Refugee Destinations.” *Scientific Reports* 7 (October). https://doi.org/10.1038/s41598-017-13828-9.

Tubiello, Francesco N., Jean-François Soussana, and S. Mark Howden. 2007. “Crop and Pasture Response to Climate Change.” *Proceedings of the National Academy of Sciences* 104 (50): 19686–90. https://doi.org/10.1073/pnas.0701728104.

White, R., and G. Engelen. 2016. “Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns:” *Environment and Planning A*, November. https://doi.org/10.1068/a251175.

Williams, Matthew, and Mirco Musolesi. 2016. “Spatio-Temporal Networks: Reachability, Centrality and Robustness | Royal Society Open Science.” *The Royal Society* 3 (6). https://royalsocietypublishing.org/doi/full/10.1098/rsos.160196.

Wilson, A. G. 1971. “A Family of Spatial Interaction Models, and Associated Developments:” *Environment and Planning A*, March. https://doi.org/10.1068/a030001.