The London Underground's resiliency: comparing the effects of three node removal strategies on the Tube's global efficiency

Introduction

Accommodating approximately 5,000,000 daily riders (Transport for London, 2019), the Underground is crucial to London's stability. This report determines which node removal strategy most impact the Underground's efficiency and which station's removal causes the greatest disruption. For its purpose, the Underground consists of 11 lines, 306 stations, and 410 links, constituting an undirected network with links weighted by length, where stations represent nodes and links represent edges.

Node Removal Strategies

Degree centrality (DC), betweenness centrality (BC), and a centrality index of these metrics (CI) are considered. DC measures a node's links, whereas BC captures the number of geodesics, or shortest paths between node pairs, that pass through a node (Freeman, 1977; Freeman, 1979). Both measures exhibit shortcomings. For instance, DC assigns nodes of the same degree equal importance. Similarly, BC weights nodes by the geodesic volume passing through them, even though information—and Tube riders—do not consistently optimize journeys (Liu et al., 2016). For these reasons, CI is proposed. Results from the three removal methodologies are compared to determine which most affects network efficiency over two iterations.

First, a station is identified using a removal strategy. Next, a duplicate network omitting the station is generated. Finally, a second station is identified and removed, creating a third network. Networks are evaluated at all three stages.

Figure 1 illustrates the presence of high correlation among methodologies, indicating methodologies may advocate removal of the same stations. While this complicates strategy comparisons, removals of the same station emphasize its importance in the network.

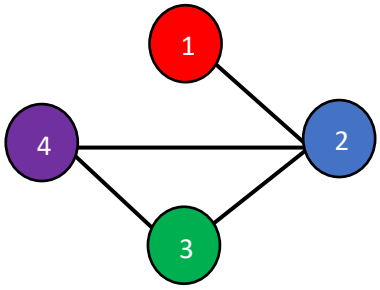
Figure 1: Correlation matrix of removal methodologies

	DC	***BC***	***CI***
DC	1.00	0.63	0.89
BC	0.63	1.00	0.91
CI	0.89	0.91	1.00

Degree Centrality

DC is calculated by summing stations' links (Opsahl et al., 2010). For an undirected network, Figure 2 illustrates that rows or columns of the adjacency matrix, A_{ij}, are summed to obtain DC (Borgatti, 2005). The red row and column show Station 1's only connection is with Station 2, as is indicated by the $A_{1,2}$ and $A_{2,1}$ entries; therefore, Station 1's degree is 1.

Figure 2: Calculating degree centrality for a Tube network with four stations



$$A_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad or \quad A_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

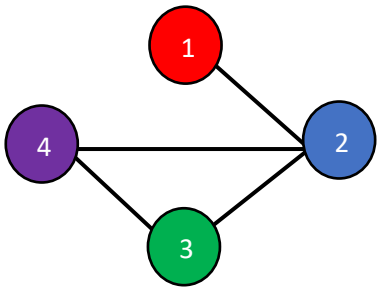
$$\begin{aligned} Station\ 1: \quad & 0 + 1 + 0 + 0 = 1 \\ Station\ 2: \quad & 1 + 0 + 1 + 1 = 3 \\ Station\ 3: \quad & 0 + 1 + 0 + 1 = 2 \\ Station\ 4: \quad & 0 + 1 + 1 + 0 = 2 \end{aligned}$$

Sequentially removing nodes by DC eliminates the best-connected stations, interrupts lines, and forces riders to recalculate journeys. It is suspected that these effects will affect the network's efficiency.

Betweenness Centrality

High BC denotes the 'bridges' or 'brokers' of a network (Abbasi et al., 2012). Figure 3 illustrates how Station 2's normalized BC, x_2, is obtained, where: $n_{s,t}^2$ is 1 if Station 2 occupies the s-t pair's geodesic and 0 otherwise; g_{st} is the number of pair s-t's geodesics; and n is the number of pairs (Meghanathan, 2016). With a BC of 1, Station 2 is the network's broker.

Figure 3: Calculating betweenness centrality for a Tube network with four stations



$$x_2 = \frac{1}{n}\sum_{st}\frac{n_{s,t}^2}{g_{s,t}} = \frac{1}{4}\left(\frac{n_{1,3}^2}{g_{1,3}} + \frac{n_{3,1}^2}{g_{3,1}} + \frac{n_{4,1}^2}{g_{4,1}} + \frac{n_{1,4}^2}{g_{1,4}}\right)$$

$$\text{where} \quad \frac{n_{1,3}^2}{g_{1,3}} = \frac{1}{1}, \quad \frac{n_{3,1}^2}{g_{3,1}} = \frac{1}{1}, \quad \frac{n_{4,1}^2}{g_{4,1}} = \frac{1}{1}, \quad \text{and} \quad \frac{n_{1,4}^2}{g_{1,4}} = \frac{1}{1}$$

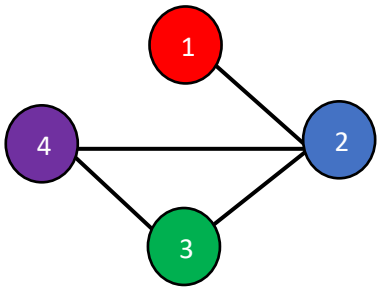
$$\therefore \quad \frac{1}{4}\left(\frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1}\right) = 1 = x_2$$

Sequentially removing nodes with the highest BC eliminates the most important 'brokers', interrupts optimal paths, and forces riders to recalculate journeys. It is suspected that these effects will affect the network's efficiency.

Centrality Index

CI utilizes a normalized index of DC and BC to determine removals. Figure 4 shows how Station 2's CI is calculated. Notably, DC measures are normalized before index calculations.

Figure 4: Calculating centrality index for a Tube network with four stations



$$DC = 1 + 0 + 1 + 1 = 4$$

$$Normalized\ DC = \frac{Degree - Min(Degree)}{Max(Degree) - Min(Degree)} = \frac{4-1}{4-1} = 1$$

$$BC = 1$$

$$Centrality\ Index = \frac{1+1}{2} = 1$$

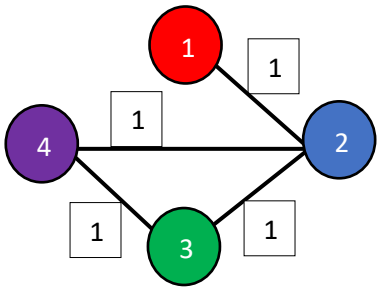
Use of CI should reduce drawbacks of using DC and BC by considering various global and local characteristics (Del Rio et al., 2009). For instance, CI should rarely indicate multiple removals in the same iteration, like DC.

Evaluation Methodologies

Since the Tube's resilience is of primary interest, removal methodologies should be evaluated using global efficiency, which measures "how efficiently [a network] exchanges information" (Latora & Marchiori, 2003, p. 249). Here, it will measure how efficiently riders navigate the Tube.

As Figure 5 illustrates, global efficiency is calculated by summing nodal pairs' inverted geodesics: $\sum_{i \neq j \in G} \frac{1}{d_{ij}}$. This sum is then divided by $N(N-1)$, where N is the network's number of nodes (Latora & Marchiori, 2001). With a uniform link weight of 1, the graph's global efficiency is one-sixth.

Figure 5: Calculating global efficiency for a Tube network with four stations



$$E(G)_{Global} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{i,j}}$$

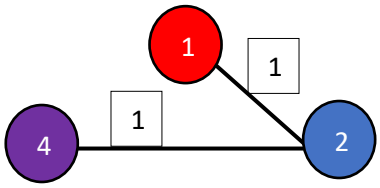
$$E(G)_{Global} = \frac{1}{4(4-1)} \left(\frac{1}{d_{1,3}} + \frac{1}{d_{1,4}} + \frac{1}{d_{3,1}} + \frac{1}{d_{4,1}} \right)$$

where $\frac{1}{d_{1,3}} = \frac{1}{2}$, $\frac{1}{d_{1,4}} = \frac{1}{2}$, $\frac{1}{d_{3,1}} = \frac{1}{2}$, and $\frac{1}{d_{4,1}} = \frac{1}{2}$

$$\therefore \ \frac{1}{4(4-1)} \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} \right) = \frac{1}{12} \times 2 = \frac{1}{6} = E(G)_{Global}$$

Figure 6 shows the effects of Station 3's removal. One would expect efficiency to decrease, but since Station 3 does not occupy the 1,4-4,1 geodesic, the network's efficiency remains unaffected. This illustrates that a removal *could* leave efficiency unaffected; however, this is unlikely for Underground station removals.

Figure 6: Calculating global efficiency for a Tube network with three stations



$$E(G)_{Global} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{i,j}}$$

$$E(G)_{Global} = \frac{1}{3(3-1)} \left(\frac{1}{d_{1,4}} + \frac{1}{d_{4,1}} \right)$$

$$\text{where } \frac{1}{d_{1,4}} = \frac{1}{2} \quad \text{and} \quad \frac{1}{d_{4,1}} = \frac{1}{2}$$

$$\therefore \quad \frac{1}{3(3-1)} \left(\frac{1}{2} + \frac{1}{2} \right) = \frac{1}{6} \times 1 = \frac{1}{6} = E(G)_{Global}$$

In addition to global efficiency, contextual characteristics are considered. Removed stations are documented to determine if correlations shown in Figure 1 emerge through removals. Furthermore, removals segment Tube lines, so the number of uninterrupted lines is tracked. Finally, the number of components or 'islands' resulting from removals indicates which method most segments the network (Esperet & Ochem, 2016).

Analysis

Analysis is conducted in R, using the `igraph` package to compute centralities and the `braingraph` package to calculate global efficiency. Only two iterations are conducted for each measure, since the third iteration of DC advocates removing four stations. This results from DC's inability to distinguish between the importance of nodes with the same degree. If included, results are incomparable.

Results

Table A: Evaluation metrics of the proposed methodologies

	Degree Centrality (DC)	Betweenness Centrality (BC)	Centrality Index (CI)
Initial Network			
Efficiency	100%	100%	100%
Lines Intact	11	11	11
Components	1	1	1
First Removal			
Station Removed	King's Cross St. Pancras	Green Park	Baker Street
Efficiency	89.79%	94.36%	88.90%
Lines Intact	5	8	6
Components	2	1	1
Second Removal			
Station Removed	Baker Street	King's Cross St. Pancras	Green Park
Efficiency	79.21%	82.52%	85.53%
Lines Intact	3	4	4
Components	3	2	1

Discussion

Table A illustrates that removing stations according to DC results in more interruptions than BC or CI. By removing King's Cross and Baker Street, the Tube's efficiency falls by more than 20% and 72% of lines are interrupted. Consequentially, the Tube consists of three components instead of one; northeast London is isolated, as is a segment from Great Portland Street to Euston Square.

While BC's and CI's removals affect efficiency less, they still result in disruptions. BC interrupts Westminster by removing Green Park and isolates northeast London by removing King's Cross, yielding two components. CI also disrupts Westminster through Green Park but segments five lines at Baker Street; however, it fails to create multiple components. By removing different stations, BC and CI both disrupt 63% of Tube lines.

As Figure 1 anticipated, the methodologies advocate removal of the same three stations; however, King's Cross's removal causes the greatest disruption. Depending on the methodology and iteration, removing King's Cross results in a(n) 11.84-10.21% efficiency reduction, while removing Baker Street and Green Park decrease efficiency by 11.10-10.58% and 5.64-3.37%, respectively. While Baker Street and King's Cross removals affect efficiency similarly, Baker Street does not disrupt as many lines. Furthermore, for a Baker Street removal to produce the Great Portland Street/Euston Square component, it must follow a King's Cross removal.

Conclusion

This analysis has produced two findings. First, removal of well-connected stations by DC disrupts the network more than BC or CI. Second, eliminating King's Cross causes more disruption than any other station removal. Considering these findings, we advocate better maintenance and protection of well-connected, high degree stations, particularly King's Cross. Furthermore, research into the practicality of centrality indexes formed from less-correlated measures should be conducted.

References

Abbasi, A., Hossain, L., and Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Infometrics*, 6(3), p. 402.

Borgatti, S.P. (2005). Centrality and network flow. *Social Networks*, 27(1), p. 62.

Del Rio, G., Koschützki, D., and Coello, G. (2009). How to identify essential genes from molecular networks?. *BMC Systems Biology*, 3(102), pp. 1-12.

Esperet, L. and Ochem, P. (2016). Islands in Graphs on Surfaces. *SIAM Journal of Discrete Math*, 301(1), pp. 206-219.

Freeman, L. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), pp. 35-41.

Freeman, L. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3), pp. 215-239.

Latora, V. and Marchiori, M. (2001). Efficient Behavior of Small-World Networks. *Physical Review Letters*, 87(19), pp. 1-2.

Latora, V. and Marchiori, M. (2003). Economic small-world behavior in weighted networks. *The European Physical Journal B*, 32(1), p. 249.

Liu, J., Xiong, Q., Shi, W., Shi, X., and Wang, K. (2016). Evaluating the importance of nodes in complex networks. *Physica A*, 452(1), p. 210.

Meghanathan, N. (2016). A computationally lightweight and localized centrality metric in lieu of betweenness centrality for complex network analysis. *Vietnam Journal of Computer Science*, 4(1), pp. 24-26.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), pp. 245-251.

Transport for London (2019). London Underground. *Transport for London*. Available at: https://tfl.gov.uk/corporate/about-tfl/what-we-do/london-underground [Accessed on 25 March, 2019].

The Barking Riverside development: forecasting its effects on Barking & Dagenham's commuter inflows

Introduction

The Barking Riverside development in Barking & Dagenham (B&D) will renovate the Barking Power Station's surrounding area. New schools, commercial space, community facilities, 10,800 residences, and an Overground station extension have been proposed (Latham, 2018). This report determines the appropriate spatial interaction model (SIM) for analyzing Barking Riverside's increase in B&D's commuter outflows and population, and the decrease in London's cost of transportation. Analysis also determines how these changes impact B&D's inflows.

The Family of SIMs

Before analysis of Barking Riverside's effects, the appropriate SIM must be identified. This report considers Wilson's renditions of the gravity model (1971).

The Unconstrained Model

The unconstrained model is commonly used when total flows is known (Fotheringham & O'Kelly, 1989). In Figure 1, T_{ij} is a matrix of flows, in this case commuters, between origins, i, and destinations, j; V_i is a vector capturing emissive characteristics of origins; W_j is a vector capturing attractive characteristics of destinations; and d_{ij} is a cost matrix of flows between origins and destinations (Oshan, 2016). Parameters k, μ, α, and β will be discussed later.

Figure 1: The unconstrained model's notation

$$T_{ij} = k\frac{V_i^{\mu}W_j^{\alpha}}{d_{ij}^{\beta}} = kV_i^{\mu}W_j^{\alpha}d_{ij}^{-\beta} \quad (\mathbf{1})$$

While this model conserves total flows, it fails to preserve outflows or inflows. For this reason, it is appropriate when the number of total flows is known, but not inflows or outflows (Fotheringham & O'Kelly, 1989). Ultimately, it is a residential location model, since it models flow accumulations at a location (Schirmer et al., 2014).

The Production-/Attraction-constrained Models

The production-constrained and attraction-constrained models differ from the unconstrained model. In addition to total flows, they conserve outflows or inflows, respectively (Fotheringham & O'Kelly, 1989). In Figures 3 and 4, T_{ij}, W_j, V_i, and d_{ij} have identical representations as in the unconstrained model. Furthermore, A_i ensures that origin i's modelled outflows equal its known outflows, O_i. Similarly, in Figure 4, B_j ensures that destination j's modelled inflows equal its known inflows, D_j (Oshan, 2016). Parameters μ, α, and β will be discussed later.

Figure 2: The production-constrained model's notation

$$T_{ij} = A_iO_iW_j^{\alpha}d_{ij}^{-\beta} \quad (\mathbf{2})$$

where $O_i = \sum_j T_{ij}$ and $A_i = \frac{1}{\sum_j W_j^{\alpha}d_{ij}^{-\beta}}$

Figure 3: Notation of the attraction-constrained model

$$T_{ij} = B_j D_j V_i^{\mu} d_{ij}^{-\beta} \quad (\mathbf{3})$$

where $D_j = \sum_i T_{ij}$ and $B_j = \frac{1}{\sum_i V_i^{\mu} d_{ij}^{-\beta}}$

Production- and attraction-constrained models conserve outflows and inflows, respectively; therefore, they are appropriate when these flows are known (Fotheringham & O'Kelly, 1989). Ultimately, these models determine flows accumulations at a location; therefore, they are residential location models (Schirmer et al., 2014).

The Doubly-constrained Model

The doubly-constrained model is a combination of the production- and attraction-constrained models, since it conserves total flows, inflows, and outflows. Subsequently, all variables in Figure 4 have identical representations to those of the previously specified models.

Figure 4: Notation of the doubly-constrained model

$$T_{ij} = A_i B_j O_i D_j c_{ij}^{-\beta} \quad (\mathbf{4})$$

where $O_i = \sum_j T_{ij}$, $D_j = \sum_i T_{ij}$, $A_i = \frac{1}{\sum_j B_j D_j c_{ij}^{-\beta}}$, and $B_j = \frac{1}{\sum_i A_i O_j c_{ij}^{-\beta}}$

The doubly-constrained model is applied when flows among origins and destinations are known (Fotheringham, 2001). Unlike other gravity model variations, the doubly-constrained model determines trip diffusion among origins and destinations; therefore, it is a transport model (Wilson, 1974).

Model Parameters

Calibration produces parameter estimates for forecasting system interactions. In this process, the natural logarithm of both sides of equations (1)-(4) are first taken. Next, linear regression analysis determines parameters' explanatory power (Fotheringham & O'Kelly, 1989). However, this method relies upon flawed assumptions, such that flows are continuous entities (Oshan, 2016).

To address these flaws, Poisson log-linear regression analysis is used (Flowerdew & Aitkin, 1982). This method assumes the Poisson distribution's mean, λ_{ij}, is equal to T_{ij} and related logarithmically to variables. After substituting λ_{ij} for T_{ij}, both sides of equations (1)-(4) are exponentiated and regression analysis is conducted (see Figure 5; Oshan, 2016).

Figure 5: Logged and exponentiated calibration formulae

***Unconstrained*:**

$$T_{ij} = \exp(k + \mu \ln V_i + \alpha \ln W_j - \beta \ln d_{ij}) \quad (\mathbf{5})$$

Production-constrained:

$$T_{ij} = \exp(k + \mu_i + \alpha \ln W_j - \beta \ln d_{ij}) \quad (\mathbf{6})$$

Attraction-constrained:

$$T_{ij} = \exp(k + \mu \ln V_i + \alpha_j - \beta \ln d_{ij}) \quad (\mathbf{7})$$

Doubly-constrained:

$$T_{ij} = \exp(k + \mu_i + \alpha_j - \beta \ln d_{ij}) \quad (\mathbf{8})$$

In equations (1)-(4), k conserves total flows, μ denotes origins' 'emmisiveness', α indicates destinations' attractiveness, and β captures generalized cost of travel. In equations (6) and (8), μ_i represents i's balancing factor, A_i, which conserves outflows. In (7) and (8), α_j represents j's balancing factor, B_j, which conserves inflows (Oshan, 2016).

While models' explanatory power is correlated with their number of parameters, these elements are negatively correlated with the conservation of observed information. For instance, the unconstrained model has four parameters but conserves total flows, while the doubly-constrained model preserves total flows, inflows, and outflows, but has one parameter (Oshan, 2016).

Analysis

Using the production-constrained model to constrain B&D's outflows and population, the impact of Barking Riverside's population and outflow increases can be determined. Borough boundary, commuter flow, population, and median salary data are analyzed (Office for National Statistics, 2016; 2001). Figure 6 shows summary statistics for the system.

Figure 6: Summary statistics for London's 33 boroughs and Barking & Dagenham

	Outflows	**Inflows**	**Population**	**Median Income (£)**
Mean	54,557.97	54,557.97	102,636.4	19,269.70
Median	51,659	30,744	99,000	18,300
Standard Deviation	17,547.57	67,653.4	30,510.06	4,035.35
Barking & Dagenham	37,223	18,313	56,000	16,200

Household size distribution inherently affects the population and commuter increase produced by new residences. Table A shows the U.K.'s household size distribution (Office for National Statistics, 2017), while Figure 7 illustrates commuter and population increase calculations, when 63% of B&D's population is working-age (London Borough of Barking & Dagenham, 2017). Notably, calculations are underpinned by two unrealistic assumptions: four-person and multi-family households do not exist. Ultimately, B&D's population and commuter estimates increase by 24,840 and 13,706 people, respectively.

Table A: Distribution of U.K. household sizes

Size	**Percent**
1	28
2	35
3	16
4+	21

Figure 7: Barking & Dagenham's increase in population and commuter outflows

***Population*:**

$$10{,}800\big(.28(1) + .35(2) + .16(3) + .21(4)\big) = 24{,}840$$

$$56{,}000 + 24{,}840 = 80{,}840$$

***Commuter outflows*:**

$$.63(80{,}840) \approx 50{,}929$$

Additionally, Barking Riverside's station extension will marginally reduce London's travel costs, as it is a terminal station. To model this, β is increased by 1% of its original, calibrated value: $-1.812009(1.01) = -1.84843$.

To determine Barking Riverside's impact, two models are conducted: one without the proposed effects ("Original Model"), and one with them ("Barking Riverside Model"). Subsequently, their results are compared.

Results

Table B: Comparison of predicted flows to and from Barking & Dagenham

	Original Model		Barking Riverside Model			
Borough	**to B&D**	**from B&D**	**to B&D**	**from B&D**	**to B&D Δ**	**from B&D Δ**
Barnet	597	433	554	1,196	-43	763
Bexley	3,199	2,367	2,805	3,775	-394	1,408
Brent	335	288	294	670	-41	382
Bromley	1,341	671	1,128	1,892	-213	1,221
Camden	221	708	268	815	47	107
City of London	15	3,346	15	32	0	-3,314
Croydon	616	388	520	1,040	-96	652
Ealing	346	244	304	611	-42	367
Enfield	913	622	829	1,331	-84	709
Greenwich	1,971	2,409	1,736	2,613	-235	204
Hackney	454	1,180	591	926	137	-254
Hammersmith and Fulham	124	523	160	386	36	-137
Haringey	574	867	566	872	-8	5
Harrow	317	272	230	318	-87	46
Havering	8,684	4,547	7,787	8,215	-897	3,668
Hillingdon	223	166	177	281	-46	115
Hounslow	189	183	172	286	-17	103
Islington	248	884	341	786	93	-98
Kensington and Chelsea	118	876	132	380	14	-496
Kingston upon Thames	151	268	136	196	-15	-72
Lambeth	500	608	498	1,072	-2	464
Lewisham	1,015	1,005	959	1,661	-56	656
Merton	262	361	224	425	-38	64
Newham	2,823	3,083	2,935	4,444	112	1,361
Redbridge	6,886	4,976	6,524	9,741	-362	4,765
Richmond upon Thames	193	386	167	292	-26	-94
Southwark	464	774	531	1,095	67	321
Sutton	243	287	194	381	-49	94
Tower Hamlets	477	1,668	678	1,263	201	-405
Waltham Forest	1,480	1,449	1,447	2,203	-33	754
Wandsworth	394	582	412	1,036	18	454
Westminster	121	801	169	697	48	-104
Totals	35,494	37,222	33,483	50,931	-2,011	13,709

Discussion

As Table B illustrates, Barking Riverside's effects on population, commuters, and transportation costs negatively impact B&D: commuter flows into the borough decrease by 2,011 (5.67%). Notably, neighboring boroughs' inflows generally decrease the most (see highlighted rows). The decrease in city-wide transportation cost, coupled with no increase in B&D's median salary, likely encourages

commuters from nearby boroughs previously working in B&D to seek work in boroughs that once had too costly of a commute. However, Barking Riverside's employment opportunities would likely increase B&D's median salary, implying that this model underestimates/overestimates B&D's inflows/outflows. Future research should ascertain whether a median salary increase reduces B&D's inflow loss or boosts commuters.

Overall, the model used is appropriate, as it constrains B&D's origin estimates. While, the model's R-squared value (0.2645) indicates a mediocre fit, the production-constrained model is the only one specified that constrains origin estimates *and* allows variation of destination estimates. Therefore, the production-constrained model is most appropriate for determining how B&D's increased origin estimates affect its commuter flows and should be extended in ascertaining the effects of a median salary increase.

Conclusion

While Wilson has popularized four SIMs for modelling residential location and transportation choices, the production-constrained model is most appropriate for determining Barking Riverside's effects on B&D's commuter flows. This report illustrates that, in the absence of a median salary increase, B&D's inflows decrease by 5.67 percent. Future research should determine Barking Riverside's projected median salary increase and its effects on B&D's commuter flows.

References

Flowerdew, R., and Aitkin, M. (1982). A Method of Fitting the Gravity Model Based on the Poisson Distribution. *Journal of Regional Science*, 22(2), 191-202.

Fotheringham, A. S. (2001). Spatial Interaction Models. In: *International Encyclopedia of the Social & Behavioral Sciences*. New Castle: University of New Castle, 14794-14800.

Fotheringham, A. S., and O'Kelly, M. E. (1989). Spatial Interaction Models: Formulations and Applications. London: Kluwer Academic Publishers.

Latham, L. (2018). The mighty Thames and the developments that are reclaiming the riverside. *Telegraph Media Group Limited*. Available at: https://www.telegraph.co.uk/property/buy/mighty-thames-developments-reclaiming-riverside/ [Accessed 6 April, 2019].

London Borough of Barking & Dagenham (2017). Population and demographic data. *London Borough of Barking & Dagenham*. Available at: https://www.lbbd.gov.uk/population-and-demographic-data [Accessed 7 April, 2019].

Office for National Statistics (2001). Table T103: Persons aged 16-74 in employment: Method of travel to work. *Office for National Statistics*.

Office for National Statistics (2016). Local Authority Districts (December 2015) Generalised Clipped Boundaries in Great Britain. *Office for National Statistics*. Available at: http://geoportal.statistics.gov.uk/datasets/8edafbe3276d4b56aec60991cbddda50_2 [Accessed 7 April, 2019].

Office for National Statistics (2017). Families and Households: 2017. *Office for National Statistics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2017 [Accessed 7 April, 2019].

Oshan, T. (2016). A primer for working with the Spatial Interaction modelling (SpInt) module in the python spatial analysis library (PySAL). *The Journal of ERSA*, 3(2), R11-R23.

Schirmer, P. M., van Eggermond M. A. B., and Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *The Journal of Transport and Land Use*, 7(2), 3-21.

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space*, 3(1), 1-32.

Wilson, A. G. (1974). Some New Forms of Spatial Interaction Model: A Review. *Transportation Research*, 9(2-3), 167-179.

Measles (agent-based) modelling: analyzing the relationships between vaccination rates, exposure rates, and the number of fatalities resulting from an outbreak

Introduction

Since the measles vaccine was discovered in 1963, incidences have rapidly declined (WHO, 2018); however, there has been a resurgence in recent years (Mundasad, 2018). This report uses an agent-based model (ABM) to analyze the relation among measles' fatality rate and its vaccination and exposure rates. The results of three scenarios will be compared to determine these relations.

Model Types

Before modelling, the ABM's use must be justified. Only cellular automata (CA) and ABMs are considered, but future research should consider other model forms.

Cellular Automata

CA model "discrete [spatiotemporal] dynamic systems based on local rules" (Miller, 2009, p. 409). They consist of cellular lattices, predefined cellular neighborhood structures, initial cell states, and rules governing how cells' states change. Cells possess defined states, which have the propensity to change every iteration. Predefined rules regarding cellular neighborhoods guide these changes. Cellular neighborhoods are spatially-defined lattice subsections, commonly surrounding cells (Clarke, 2014).

Agent-Based Models

ABMs consist of agents, managerial rules, learning procedures, agent action processes, and environments. Agents different types and granularities. Cells and agents possess defined states, which have the propensity to change every iteration. Informative research usually guides agent procedures, which range from movement to agent and environment interactions. These interactions change agents' and cells' states (Clarke, 2014).

"Close Cousins"

Couclelis states that "CA and ABM are…close cousins" (2019, p. 2); however, they have differences. For instance, ABMs' agents move about the environment and alter cells' and agents' states. CA's agents—lattice cells—are static and only interact with neighbors. Furthermore, ABMs model emergence resulting from agents' interactions, while CA model complexity resulting from cells' neighborhood-level interactions (Clarke, 2014). Essentially, CA's cells and ABMs' agents make decisions, respectively (Couclelis, 2019). Since people spread measles, this report uses an ABM to model their mobile nature.

The Model

Analysis uses a retrofit of the Netlogo, epiDEM Basic model (Wilensky, 1999; Yang & Wilensky, 2011; Wise, 2019). Modellers chose when and how often a contagion is introduced. Upon introduction, agents may be infected. These agents infect susceptible agents with which they share a cell. Agents have specified immunity probabilities and recovery probabilities. If infected, recovery probability recalculates every iteration. Initial number of agents, recovery probability, immunity probability, and number of agents exposed to the contagion are parameters.

The model exhibits micro-scale granularity, since agents represent individuals; therefore, iterations represent hours. The four parameters denote four state variables, so the state space is four-dimensional. Ultimately, the system has three steady states: all agents are infected, all agents are uninfected, or a certain ratio of infected to uninfected persists, *ad infinitum*.

Model Extensions

Model extensions tailor the ABM to measles. For instance, recovered agents gain immunity. Furthermore, recovery occurs between 17-22 days, or 408-528 iterations (WHO, 2018). Additionally, there is a 0.2% chance of dying following a 12-day incubation period, or 288 iterations (CDC, 2015). Chance of death and recovery time are initialized once per agent, when the agent is infected. Ultimately, these extensions supersede recovery probability as a state variable, since surviving agents recover and gain immunity.

Since recovery probability now takes a random integer between 407 and 529, and the additional agent states are fixed at similar ranges, the state space is three-dimensional. Furthermore, the new system has two steady states: some agents die, or all agents survive.

These additions are flawed in their oversimplicity. For instance, disease-complications are common and sometimes lifelong or terminal (CDC, 2015; Vaccine Knowledge Project, 2017; Wendorf et al., 2017). The model's temporal granularity does not permit modelling these cases.

The Three Scenarios

Three scenarios will be simulated, in which vaccination rate and exposure rate (number of exposed agents) vary. In Scenario 1, herd immunity is achieved (95%) and 100 agents are exposed (Fine, 1993), representing a town with many epidemiologists. In Scenario 2, vaccination rate is 50%, and fifty agents are exposed, depicting an Equatorial Guinean community, circa 2013 (BBC, 2013). Scenario 3 characterizes an Amish or Orthodox Jewish community where no one is vaccinated and ten agents are exposed (Belluz, 2019). In all scenarios, one-thousand agents are initialized.

In Scenarios 1 and 2, virus proliferation may not occur, since agents have a chance of being vaccinated; therefore, the minimum number of iterations could be zero. In Scenario 3, exposure always results in propagation. The minimum number of iterations could be 288, since all exposed agents *could* die in the shortest possible time, without infecting others; however, considering the community's low vaccination rate, this is unlikely.

Analysis

Fatality rate will be measured as the number of fatalities upon achieving steady state. Before analysis can be conducted, the minimum number of repetitions required for statistical significance, n, must be determined (see Figure 1).

Figure 1: Formula for calculation of minimum required runs

$$ME = \frac{t \cdot s}{\sqrt{n}} \quad \leftrightarrow \quad n = \left(\frac{t \cdot s}{ME}\right)^2$$

Given agents' chance of death, a 5% margin of error, ME, seems appropriate. Sample scenarios initialize one hundred agents. One hundred repetitions of one thousand iterations are conducted, generating standard deviation estimates, s. This number of repetitions implies 99 degrees of freedom. There cannot be negative fatalities, so a right-tailed t-score of 1.660391 is used.

Table A shows samples' parameters, Figure 1 depicts their distributions, and Figure 2 illustrates minimum run calculations.

Table A: Samples' parameter sets

	Scenario 1	Scenario 2	Scenario 3
Initial agent(s) exposed	10	5	1
Initial agents	100	100	100
Vaccination rate (%)	95	50	0
Repetitions	100	100	100
Time limit (iterations)	1000	1000	1000

Figure 1: Distribution of fatalities, by scenario sample

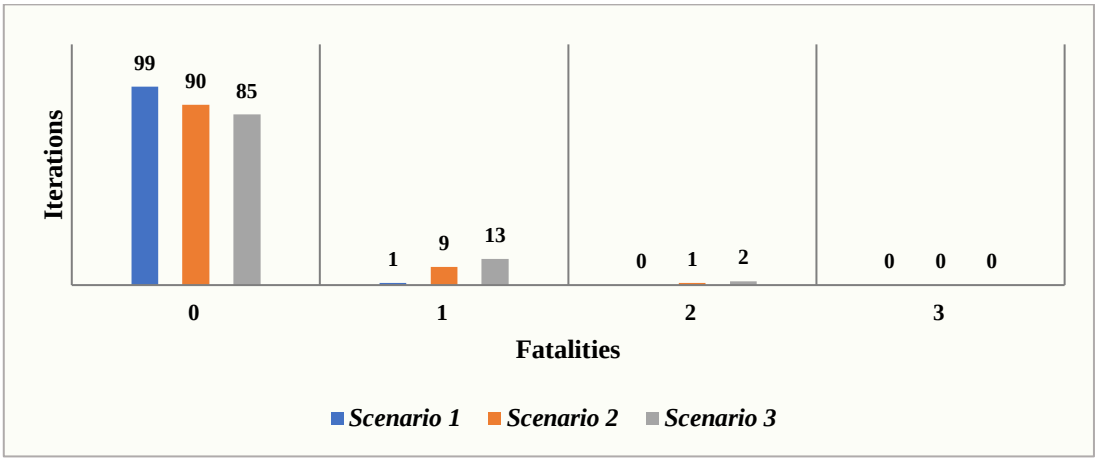


Figure 2: Calculations of minimum runs required for statistical significance

Scenario 1:

$$s \approx 0.10000$$

$$n = \left(\frac{t \cdot s}{ME}\right)^2 = \left(\frac{1.660391 \cdot 0.10000}{.05}\right)^2 \approx 11$$

Scenario 2:

$$s \approx 0.34510$$

$$n = \left(\frac{t \cdot s}{ME}\right)^2 = \left(\frac{1.660391 \cdot 0.34510}{.05}\right)^2 \approx 131$$

Scenario 3:

$$s \approx 0.42770$$

$$n = \left(\frac{t \cdot s}{ME}\right)^2 = \left(\frac{1.660391 \cdot 0.42770}{.05}\right)^2 \approx 201$$

Ultimately, 250 repetitions are conducted, for each scenario. Table B shows scenarios' parameter sets.

Table B: Scenarios' population parameter sets

	Scenario 1	Scenario 2	Scenario 3
Initial agent(s) exposed	100	50	10
Initial agents	1000	1000	1000
Vaccination rate (%)	95	50	0
Repetitions	250	250	250
Time limit (iterations)	1000	1000	1000

Results

Figure 4: Distribution of fatalities, by scenario: population

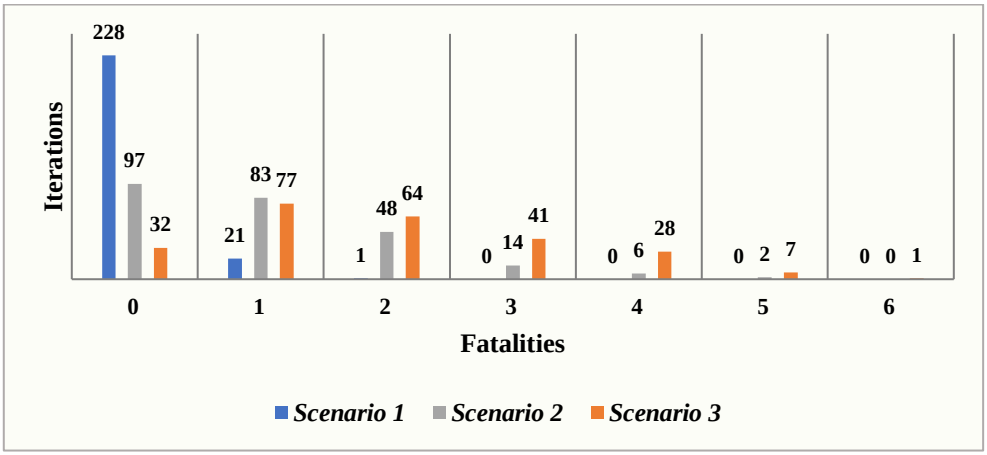


Table B: Summary statistics of fatalities

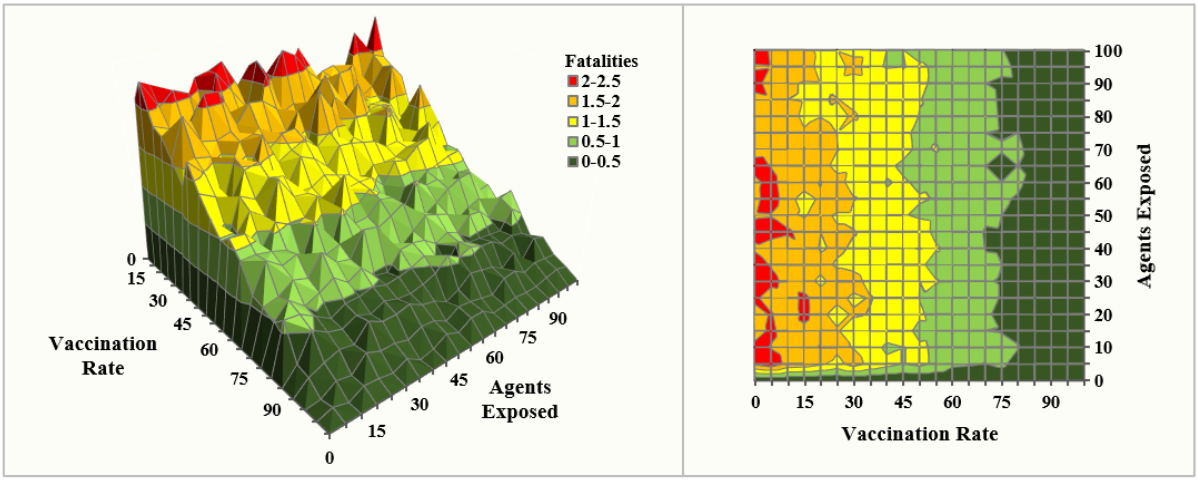
	Scenario 1		Scenario 2		Scenario 3	
	Fatalities	*Iterations*	*Fatalities*	*Iterations*	*Fatalities*	*Iterations*
Mean	0.09	682.48	1.02	546.85	1.92	546.36
Median	0	672	1	546	2	546
Mode	0	640	0	546	1	545
Std. Dev.	0.30	87.58	1.07	3.97	1.32	3.38
Maximum	2	909	5	563	6	559
Minimum	0	0	0	539	0	540
Range	2	909	5	24	6	19

Discussion

Figure 4 and Table B illustrate inverse correlations between vaccination rate and average number of fatalities, and between average run-time and vaccination. Essentially, less fatalities occur in highly vaccinated communities. Conversely, measles takes longer to propagate throughout a highly vaccinated community, because there are fewer carriers. Ultimately, the outbreak lasts longer but affects fewer individuals. Research supports these findings (McLean, 1998; BBC, 2019).

Furthermore, a parameter sweep of vaccination and exposure rates, supports the finding that vaccination rate and average number of fatalities are inversely correlated, regardless of the number of exposed agents (see Figure 5). However, this sweep uses a statistically insignificant number of repetitions (75) and parameter combinations that vary by five. Future research should reimplement this sweep, using a statistically significant number of repetitions and combinations that vary by one. It is expected that doing so would smooth the surface map, furthering the reliability and extendibility of this report's findings.

Figure 5: Parameter sweep surface map



Conclusion

Ultimately, this report shows that an ABM can be used to analyze relationships between a disease and its vaccination and exposure rates. In the case of measles, herd immunity drastically reduces fatality rates, even when one-tenth of the community is exposed. These results should inform future research and policymakers in their endeavor to prove that vaccinations are effective and important for combatting outbreaks.

References

BBC (2013). The growth of global immunisation. *BBC*. Available at: https://www.bbc.com/news/health-24519949 [Accessed 16 April, 2019].

BBC (2019). Measles cases triple globally in 2019, says UN. *BBC*. Available at: https://www.bbc.com/news/health-47940710 [Accessed 15 April, 2019].

Belluz, J. (2019). New York's Orthodox Jewish community is battling measles outbreaks. Vaccine deniers are to blame. *Vox Media*. Available at: https://www.vox.com/science-and-health/2018/11/9/18068036/measles-new-york-orthodox-jewish-community-vaccines [Accessed 16 April, 2019].

CDC (2015). Epidemiology and Prevention of Vaccine-Preventable Diseases. Washington: Public Health Foundation. Available at: https://www.cdc.gov/vaccines/pubs/pinkbook/meas.html#complications [Accessed 14 April, 2019].

Clarke, K. (2014). Cellular Automata and Agent-Based Models. In: *Handbook of Regional Science*. Berlin: Springer-Verlag, pp. 1220, 1225-1226.

Couclelis, H. (2019). Cellular automata and agent-based models: what next?. *ResearchGate*. Available at: https://www.researchgate.net/publication/265159398_Cellular_automata_and_agent-based_models_what_next [Accessed 14 April, 2019].

Fine, P. E. M. (1993). Herd Immunity: History, Theory, Practice. *Epidemiologic Reviews*, 15(2), p. 265.

McClean, A. R. (1998). Vaccines and their impact on the control of disease. *British Medical Bulletin*, 54(3), pp. 545-556.

Miller, H. J. (2009). Geocomputation. In: *The SAGE Handbook of Spatial Analysis*. London: SAGE, p. 409.

Mundasad, S. (2018). Measles resurgence 'due to vaccine hesitancy', WHO warns. *BBC*. Available at: https://www.bbc.com/news/health-46387167 [Accessed 15 April, 2019].

Vaccine Knowledge Project (2017). Measles. *Oxford Vaccine Group*. Available at: http://vk.ovg.ox.ac.uk/measles [Accessed 14 April, 2019].

Wendorf, K. A., Winter, K., Zipprich, J., Schechter, R., Hacker, J. K., Preas, C., Cherry, J. D., Glaser, C., and Harriman, K. (2017). Subacute Sclerosing Panencephalitis: The Devastating Measles Complication That Might Be More Common Than Previously Thought. *Clinical Infectious Diseases*, 65(2), pp. 226-232.

WHO (2018). Measles. *World Health Organization*. Available at: https://www.who.int/news-room/fact-sheets/detail/measles [Accessed 14 April, 2019].

Wilensky, U. (1999). Netlogo. *Center for Connected Learning and Computer-Based Modeling, Northwestern University*, Evanston, IL. Available at: http://ccl.northwestern.edu/netlogo/.

Yang, C., and Wilensky, U. (2011). NetLogo epiDEM Basic model. *Center for Connected Learning and Computer-Based Modeling, Northwestern University*, Evanston, IL. Available at: http://ccl.northwestern.edu/netlogo/models/epiDEMBasic.

Word counts

First report: 999

Second report: 993

Third report: 1,008

Total: 3,000