Urban Simulation Report

Student number: 23083053

1 Part 1: London's Underground Resilience

The first part of this report aims to address the resilience of the London Underground using network analysis.

1.1 Topological Network

1.1.1 Centrality Measures

Centrality measures are characteristics of nodes showing their *importance* in various aspects. In this section, we will identify nodes on the network with the highest centrality in the following measurements: degree centrality, closeness centrality, and betweenness centrality. The 3 central measures that I will cover in this report are: degree centrality, betweenness centrality, and closeness centrality. We will consider a network of n nodes, and the number of links between nodes i and j will be denoted as A_{ij}.

Degree centrality is the number of links that are connected to each node. Considering the underground as an undirected graph, the degree centrality k_i for node i is calculated as

$$k_i = \sum_j A_{ij}$$

In the context of the underground network, the degree corresponds to the number of lines that serve each station counting 1 for each direction. A high degree centrality indicates there are many lines that serve the station, thus identifies importance of the station as a transit hub that allows for transfer between multiple lines.

The stations with the highest degree centrality are shown in Table 1. These stations match the characteristics of being a large transfer station.

Table 1: Stations with the highest degree centrality. All stations tied at a degree of 6 are listed.

Rank	Station	Degree
1	Stratford	9
2	Bank and Monument	8
3	Baker Street	7
3	King's Cross St. Pancras	7

Rank	Station	Degree
5	Oxford Circus	6
5	Earl's Court	6
5	Canning Town	6
5	Waterloo	6
5	Liverpool Street	6
5	West Ham	6

Betweenness centrality is defined by the number of shortest paths that run through the node (or link). The betweenness centrality x_i can be calculated as

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}}$$

where

$$n_{st}^i = \begin{cases} 1 & (\text{if node } i \text{ is on geodesic from } s \text{ to } t) \\ 0 & (\text{otherwise}) \end{cases}$$

and g_{st} is the total number of geodesic paths from s to t. To normalised, the raw value should be divided by all possible combinations of shortest paths between the other $n-1$ nodes: $\frac{(n-1)(n-2)}{2}$. High betweenness centrality on the underground shows there are many passengers travelling through the station during their journeys, which are stations en route of connecting the suburban areas with the city centre, or stations on short-distance routes. When this station becomes inaccessible, a large amount of people will be affected.

The stations with the highest betweenness centrality are shown in Table 2.

Table 2: Stations with highest betweenness centrality.

Rank	Station	Betweenness Centrality
1	Bank and Monument	17,681
2	King's Cross St. Pancras	16,625
3	Stratford	14,563
4	Oxford Circus	13,548
5	Euston	13,181
6	Baker Street	12,130
7	Earl's Court	11,474
8	Shadwell	11,128
9	Waterloo	10,428
10	South Kensington	10,335

Closeness centrality is the inverse of the main geodesic distance l_i of one node to all the other $n-1$ nodes. Given the geodesic distance between nodes i and j as d_{ij}, the closeness centrality C_i is calculated as

$$C_i = \frac{1}{l_i} = \frac{n-1}{\sum_j d_{ij}}$$

A high closeness centrality in a rail network indicates the station is within a short distance from all the other stations, located at the physical centre of the network. Provided that the network spreads out radially from the city centre, the stations with highest closeness centrality are assumed to be located within the city centre, as illustrated by the high-ranked stations in London shown in Table 3. These stations are located at the West End area - which can be assumed as the central area of London.

Table 3: Stations with the highest closeness centrality.

Rank	Station	Closeness Centrality $(\times 10^{-5})$
1	Holborn	7.924
2	King's Cross St. Pancras	7.896
3	Tottenham Court Road	7.889
4	Oxford Circus	7.873
5	Leicester Square	7.835
6	Picadilly Circus	7.831
7	Charing Cross	7.830
8	Chancery Lane	7.823
9	Covent Garden	7.807
10	Embankment	7.799

1.1.2 Impact Measures

For impact measurement, we considered the delta centrality, a concept introduced by Latora and Marchiori (2007) that compares the *performance* of the whole network when a node is removed from the network in question. For a performance measure P of the network G, the delta centrality C_i^Δ of node i is defined as

$$C_i^\Delta = \frac{\Delta P}{P} = \frac{P(G) - P(G')}{P(G)}$$

where G' denotes the network where node i is removed. This can be applied to any performance measure P - we have considered the measures summarised in Table 4.

Table 4: Performance measures considered for delta centrality. The impact of node removal are measured through the change in these performance values.

Performance Measure P	Explanation
Inverse of Average Shortest Path Length	The average length of the shortest path between all pairs of nodes indicates the efficiency of the network, the smaller being more efficient. An inverse of this value indicates a better network with higher values.
Number of Cycles in Networks	Cycles in networks allow for supplementary routes in case of incidents (Katifori, Szöllősi and Magnasco, 2010). More cycles in the network indicate a more resilient network.

These measures can be applied to any transport network, especially the average shortest path length is directly addressing the puropose of transport systems of connecting different areas. The number of cycles also impact the resilience of transport networks, although may not be applicable for road networks that inherently have a large number of cycles. An applicable example of these measures outside the London underground is the destruction of the subway network in New York City after the September 11 attacks, which resulted in the closure of multiple stations and line segments (Wyatt, 2002; Paaswell, 2012).

1.1.3 Node Removal

The removal of nodes with high centrality and the impact on the performance of the network are considered. For each centrality measure, two strategies are used to remove 10 nodes from the network: a **non-sequential** strategy removes nodes in order that appears in the rank tables (Table 1; Table 2; Table 3), while a **sequential** strategy removes the node with the highest centrality when recalculated after the previous node removal step. The stations removed for each centrality measure are shown in Table 5.

1.2 Flows: Weighted Network

1.2.1 Centrality Measures

1.2.2 Impact Measures

1.2.3 Node Removal

2 Part 2: Spatial Interaction Models

In the second part, we will analyse the OD matrix of the London Underground. The variables will be denoted as follows:

Table 5: Stations removed for impact assessment for each centrality measure and strategy. For each scenario, the stations were removed one at a time from the top of the table to the bottom.

(a) Degree Centrality

Non-sequential	Sequential
Stratford	Stratford
Bank and Monument	
Baker Street	
King's Cross St. Pancras	
Oxford Circus	
Earl's Court	
Canning Town	
Waterloo	
Liverpool Street	

(b) Betweenness Centrality

Non-sequential	Sequential
Bank and Monument	Bank and Monument
King's Cross St. Pancras	
Stratford	
Oxford Circus	
Euston	
Baker Street	
Earl's Court	
Shadwell	
Waterloo	
South Kensington	

(c) Closeness Centrality

Non-sequential	Sequential
Holborn	Holborn
King's Cross St. Pancras	
Tottenham Court Road	
Oxford Circus	
Leicester Square	
Picadilly Circus	
Charing Cross	
Chancery Lane	
Covent Garden	
Embankment	

- T_{ij}: estimated flow from station i to j
- T: total flow within a network, thus $T = \sum_{i,j} T_{ij}$
- d_{ij}: distance between stations i and j - the cost function is a function of d_{ij} hence denoted as $f(d_{ij})$
- O_i, D_j: attractors (often population) at the origin and destination

2.1 Models and Calibration

2.1.1 Family of Spatial Interaction Models

The family of spatial interaction models are as follows:

The **unconstrained model** only constrains the model with matching the total flows with the observed value, written as

$$T_{ij} = KO_i^{\alpha} D_j^{\gamma} f(c_{ij})$$

where K is the constant

$$K = \frac{T}{\sum_i \sum_j O_i^{\alpha} D_j^{\gamma} f(d_{ij})}$$

so that the total number of journeys are constrained.

The **singly constrained model** constrains the total number of observations for each component at the origin or the destination. The origin constrained model fixes the total number of journeys at the origin, as:

$$T_{ij} = A_i O_i^{\alpha} D_j^{\gamma} f(d_{ij}) \tag{1}$$

The parameter A_i is determined so that the total at the origin is constrained, that is:

$$\sum_j T_{ij} = O_i \therefore A_i = \frac{1}{\sum_j D_j f(d_{ij})}$$

Similarly, the **destination constrained model** constrains the total at the destination.

$$T_{ij} = O_i^{\alpha} B_j D_j^{\gamma} f(c_{ij}) \quad \text{where} \quad \left(B_j = \frac{1}{\sum_i O_i f(c_{ij})} \right)$$

The **doubly constrained model** constrains both the total at the origin and destination.

$$T_{ij} = A_i O_i^{\alpha} B_j D_j^{\gamma} \,\text{exp}(-\beta d_{ij}) \tag{2}$$

2.1.2 Calibration of Cost Function

The dataset used for analysis have the following data for every origin-destination pair for the London Underground Stations.

- population of origin
- jobs at the destination
- distance between origin and destination
- flow from the origin to the destination

Using this dataset, we will use the doubly constrained model to estimate the best cost function and parameters by comparing with the observed flow. Constraining both the total journeys for both the origin and destination will enable the full utilisation of observed data, enabling a most accurate calibration of the cost function.

The doubly constrained model after a logarithm transformation of Equation 2 into a Poisson expression can be written as:

$$\text{ln}(T_{ij}) = \text{ln}\, A_i + \text{ln}\, O_i^{\alpha} + \text{ln}\, B_j + \text{ln}\, D_j^{\gamma} + \text{ln}\, f(d_{ij})$$

We will compare the negative exponential and inverse power relationships as the cost function $f(d_{ij})$ as follows:

$$f(d_{ij}) = \begin{cases} \text{exp}(-\beta d_{ij}) \text{ (Negative Exponential)} \\ d_{ij}^{-\beta} \text{ (Inverse Power)} \end{cases}$$

For each cost function, we have run the Poisson Regression to calculate the optimal parameter β.

Table 6: Comparison of results for the doubly constrained model using inverse power and negative exponential cost functions.

Cost Function	Parameter β	R^2 value
Negative exponential $f(d_{ij}) = \exp(-\beta d_{ij})$	$\beta = 1.543 \times 10^{-4}$	$R^2 = 0.4979$
Inverse Power $f(d_{ij}) = d_{ij}^{-\beta}$	$\beta = 9.096 \times 10^{-1}$	$R^2 = 0.4077$

From these results, the negative exponential model for the cost function has a better fit to the observed flows. When considering the scenarios, the cost function

$$f(d_{ij}) = \exp(-\beta d_{ij}) \quad (\beta = -1.543 \times 10^{-4}) \tag{3}$$

will be used.

2.2 Scenarios

The scenarios we will observe in this report are summarised in Table 7.

Table 7: The scenarios explored in this report

Scenario	Explanation
Scenario A	Jobs at Canary Wharf decrease by 50 %
Scenario B	Increase in cost of transport - considering 2 parameters

Since scenario A involves the change in the characteristics of the destination, the origin constrained model is used for the analysis to preserve the number of commuters starting their journeys in each area.

2.2.1 Calibration of γ

Since the origin-constrained model will be used for the analysis of the scenarios, the gamma variable introduced in Equation 1 must be calibrated before applying to the new scenarios. γ with the highest R-squared value for estimating the original flow will be used, resulting as follows:

$$\gamma = 7.556 \times 10^{-1} \quad (R^2 = 0.4680) \tag{4}$$

2.2.2 Scenario A

We will first decrease the number of jobs at Canary Wharf by 50%, from the original 58,772 to 29,386. Using the origin-constrained model, the procedure is as follows:

1. Calculate the flow based on the origin-constrained model $T_{ij} = A_i O_i D_j^{\gamma} \exp(-\beta d_{ij})$ to retrieve the optimal γ value for this dataset. The value for parameters β, γ is derived from Equation 3 and Equation 4.
2. Reduce the number of jobs at Canary Wharf (transform D_j into D'_j)
3. Recalibrate the adjusted A'_i parameter with the new distribution of using the relationship $A'_i = \frac{1}{\sum_j D_j'^{\gamma} f(d_{ij})}$
4. Calculate the new flows using the relationship $T'_{ij} = A'_i O_i D_j'^{\gamma} \exp(-\beta d_{ij})$

Using the origin constrained model, we will observe how the destination of commuters changed in reaction to the decrease in jobs in Canary Wharf.

We have observed a significant drop in the number of journeys terminating at Canary Wharf from 47,690 in the original simulation to 29,496 in scenario A, which is 61.9 % of the original amount. As observable from Figure 1, the decrease in the flows to Canary Wharf occured evenly among all origins, and has been redistributed into other destinations.

2.2.3 Scenario B

We will consider 2 scenarios for the change in the cost of transport. Given the original parameter as β, the parameters β_1, β_2 for the two scenarios B1 and B2 will be modified as shown in Equation 5 and Figure 2.

$$\begin{cases} \beta_1 = 2\beta \\ \beta_2 = 10\beta \end{cases} \qquad (\beta = -1.543 \times 10^{-4}) \tag{5}$$

The flows between origin-destination pairs for each scenario, plotted by distance of journeys, are shown in Figure 3. As the β increases, the longer distance journeys are disencouraged, and the distribution of the distance of journeys become negatively skewed.

2.2.4 Discussion

In this section, we will discuss the impact of each scenario on the distribution of flows in the underground network. The measurements summarised in Table 8 are considered to examine the changes that occured as a result of each scenario.

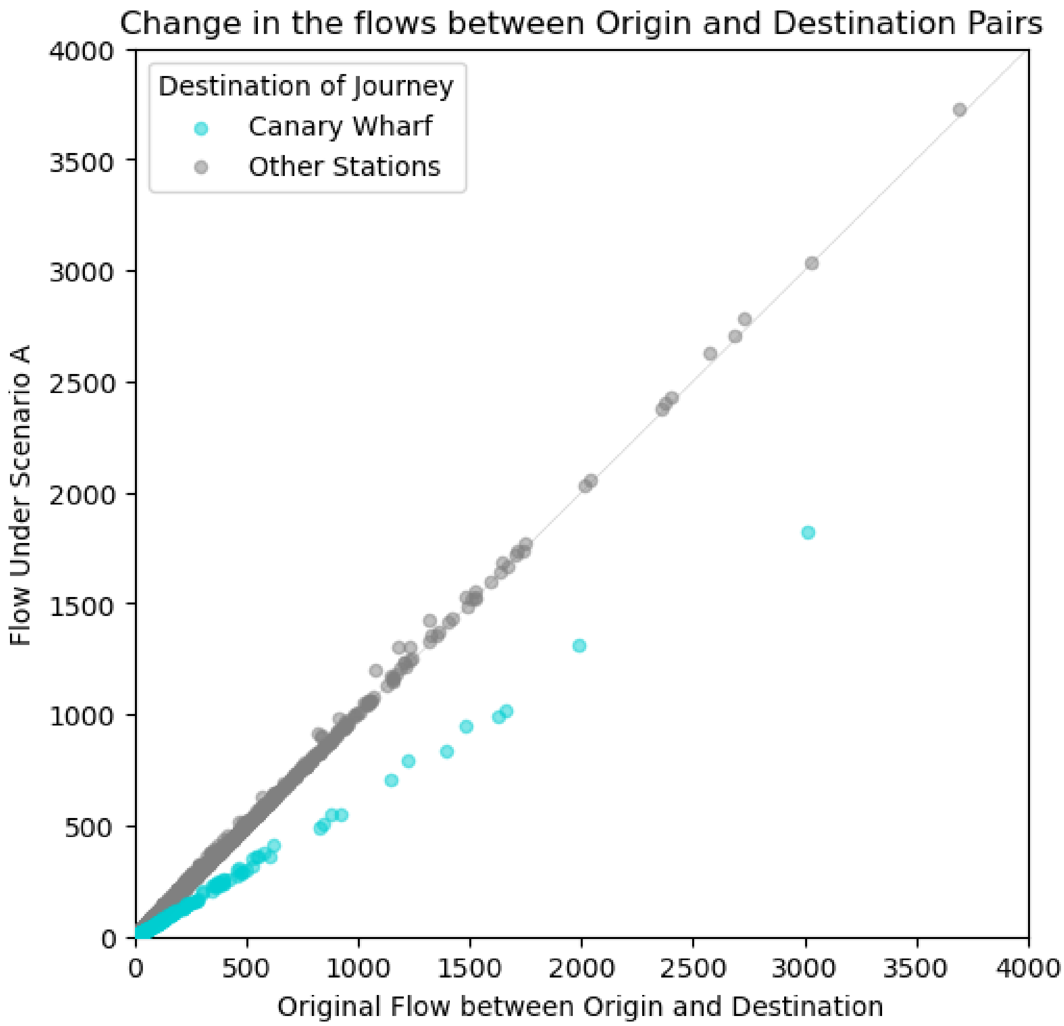

Figure 1: Comparison of the flows of OD pairs the original simulation and the flows under scenario A.

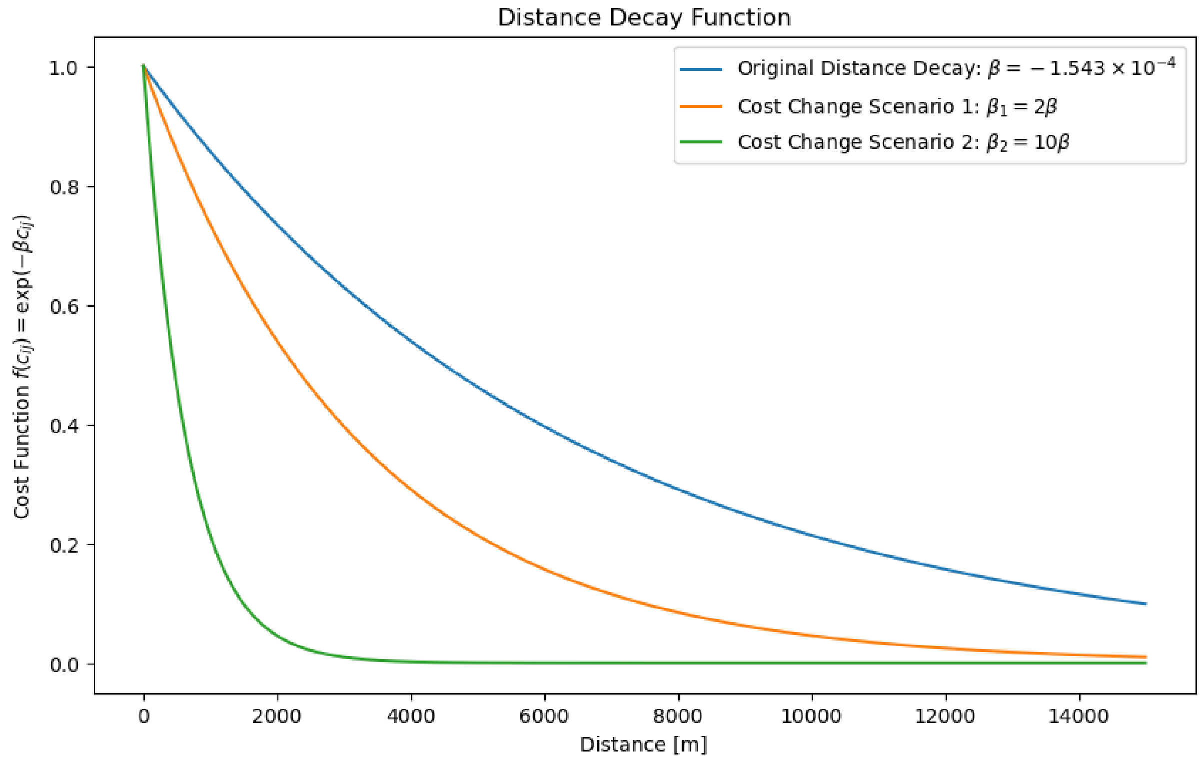

Figure 2: Distance decay function for the original scenario and the two modified scenarios. This indicates the cost for travelling a fixed distance will be equivalent to that of travelling twice (B1) or 10 times (B2) the distance in the original model.

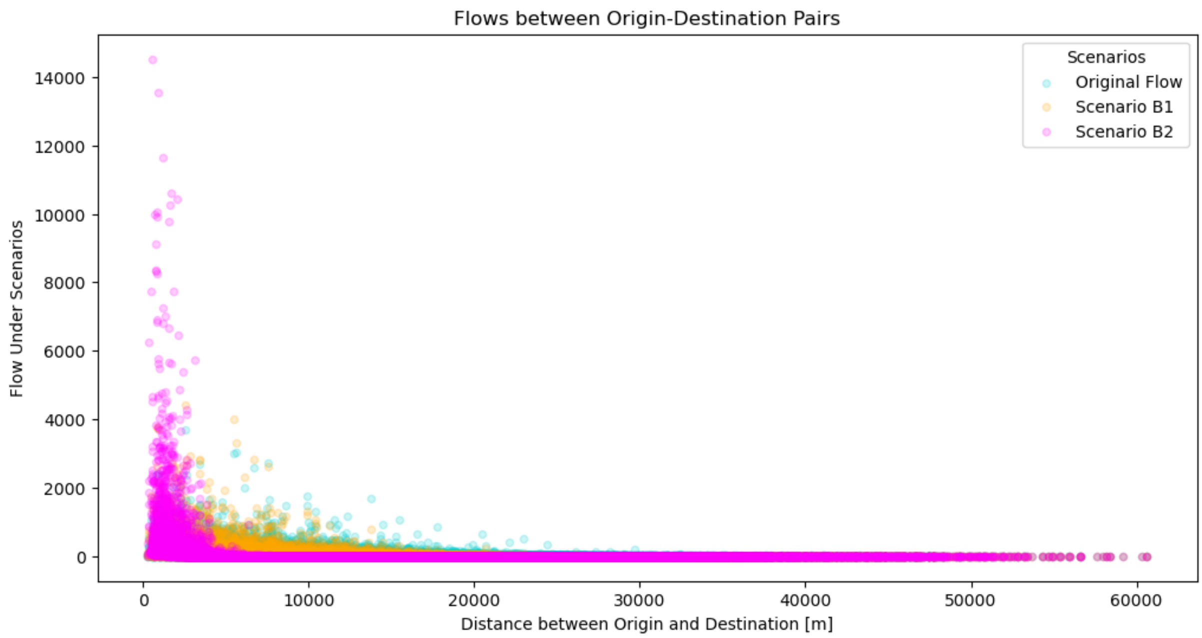

Figure 3: Comparison of the distance of flows for scenario B and the original prediction.

Table 8: Measurements consider to quantify the change in flows for each scenario. T'_{ij} is the number of journeys from station i to j under the scenario.

Measurement	Explanation
Quantity of change in destination T_{diff}	The number of flows that have different destination compared to the original scenario. A larger number of change in destination indicate larger impact.
Mean distance of journeys $\bar{\delta}^X$	The mean distance of all journeys simulated in each scenario. A larger change in the distance compared to the original simulation indicate larger impact.

For each scenario, the values for these measurements are calculated as shown in Table 9. The number of destinations changed and the change in the mean distance of journeys both indicate that scenario B2 had the largest impact on the number of flows, and scenario A being the least impactful.

Table 9: Value of measurements for each scenario.

Scenario	T_{diff}	$\bar{\delta}$ [m]
Scenario A	18,193 (1.2 % of total flows)	8,579
Scenario B1	346,503 (22.47 %)	6,030
Scenario B2	1,222,191 (79.25 %)	1,613
Original Simulation	Total flows: 1,542,283	8,583

The intervention to the travel cost (scenario B) directly impacts all OD pairs, while the reduction of jobs in one area (scenario A) only impacts journeys that end at the station affected. This different nature of the interventions dictate the magnitude of change it causes; thus scenario B has a larger impact on the flows on the network as a whole. Recalling the impact of scenario A on Canary Wharf was a 39 % decrease in the total inflows, this figure falls between scenarios B1 and B2 in terms of T_{diff}. Since $\bar{\delta}$ remains unchanged the impact on Canary Wharf is difficult to compare between scenarios A and B1, scenario B2 has the largest impact even if the focus is on Canary Wharf. To conclude, a drastic impact on the cost function has more impact throughout the network compared to a large change in the demand in a particular area. Impact of the actual changes in the underground fares that occured should be investigated to compare with the results of our model.

Word count: x words

GitHub repository (as hyperlink): Urban_Simulation_Report

References

Katifori, E., Szöllősi, G. J. and Magnasco, M. O. (2010) 'Damage and Fluctuations Induce Loops in Optimal Transport Networks', *Physical Review Letters*, 104(4), p. 048704. doi: 10.1103/PhysRevLett.104.048704.

Latora, V. and Marchiori, M. (2007) 'A measure of centrality based on network efficiency', *New Journal of Physics*, 9(6), p. 188. doi: 10.1088/1367-2630/9/6/188.

Paaswell, R. E. (2012) 'Approaches to Infrastructure Redevelopment, WTC, Lower NYC and the Region', pp. 215–226. doi: 10.1061/40717(148)18.

Wyatt, E. (2002) 'Subway Service to Resume On Routes Closed After 9/11', *The New York Times*.