Access graph: a novel graph representation of public transport networks for accessibility analysis

Tina Šfiligoj*[1], Aljoša Peperko[2], and Oded Cats[3]

[1]PhD student, Faculty of Maritime Studies and Transport, University of Ljubljana, Slovenia
[2]Associate professor, Faculty of Mechanical Engineering, University of Ljubljana and Institute of Mathematics, Physics and Mechanics, Slovenia
[3]Full professor, Department of Transport and Planning, Delft University of Technology, The Netherlands

Short summary

We study shortest-path-based accessibility by building a so-called access graph where two nodes share an edge if one is reachable from the other within a given time budget. This is achieved by obtaining generalised travel times between all pairs of nodes and directly connecting the pairs where the generalised travel time condition is satisfied. We then observe the time evolution of the access graph, increasing the time budget from zero to the maximum travel time in the network. Average degree and degree distributions of the newly proposed access graphs for 51 metro networks are analysed and two global measures are proposed as network-level indicators of accessibility. For all metro networks in our empirical analysis, a logistic-like growth of average degree with time budget is observed. We see a large potential of the introduced graph representation for in-depth studies of accessibility.
Keywords: public transport, accessibility, network science, metro network, access equity

1 Introduction

Accessibility is one of the main determinants of public transport (PT) use. The Hansen definition of accessibility as travel impedance between spatially dispersed opportunities for activity is at the core of defining accessibility indicators (Hansen, 1959; Geurs & Van Wee, 2004). In PT access studies, generalised travel time is typically comprised of in-vehicle and waiting times, together with transfer costs. Network-science-based indicators are increasingly used in investigating transport network structure and performance (Derrible & Kennedy, 2011; Ding et al., 2019; Shanmukhappa et al., 2019). Luo et al. (2019) suggested connecting network science and accessibility by calculating average shortest paths for each node to all other nodes, providing a Hansen-like indicator similar to closeness centrality.

Passengers' travels take place on (approximately) shortest paths. Past studies focused on either analyzing the relative importance of nodes using centrality indicators (Cats, 2017; Šfiligoj et al., 2025) or compared different networks using an aggregate metric such as the average shortest path from each node to all other nodes, known as network (in)efficiency (Dimitrov & Ceder, 2016; de Regt et al., 2019). We argue that aggregated values like these do not provide a sufficiently complete and insightful information when comparing accessibility across networks. To this end, we introduce a novel graph representation, called an access graph, where two nodes share an edge if they can be reached from one another within a given time budget. The distance matrix of generalised travel times is based on the unweighted and frequency-weighted P-space representation, and in-vehicle-time-weighted L-space representation. Importantly, in the resulting graph representation, while retaining the same set of nodes, the edges directly indicate accessibility between two nodes. Specifically, node degree in the access graph directly provides the number of reachable nodes within the time budget. Based on this representation, we study the temporal dependence of node degree distributions, focusing on average degree. In addition, the Gini coefficient of the degree distribution is used as an indicator of access equity (Gori et al., 2020).

2 Methodology

In the access graph $G_{\mathcal{A}}$, the nodes represent the stops of a PTN, and the edge set is obtained as follows. The generalised distance matrix $\mathcal{D}$ of the network is obtained by calculating distance matrices in three different PTN representations: $\mathcal{D}_{P^u}$ for the unweighted P-space representation, $\mathcal{D}_{P^f}$ for the frequency-weighted P-space representation, and $\mathcal{D}_{L^t}$ for the in-vehicle-time-weighted L-space representation. The generalised distance matrix is then calculated as:

$$\mathcal{D} = \mathcal{D}_{L^t} + w^{\text{wait}}\mathcal{D}_{P^f} + w^{\text{transfer}}(\mathcal{D}_{P^u} - 1), \tag{1}$$

where $w^{\text{wait}} = 2$ is the waiting time weight and $w^{\text{transfer}} = 5$ min is the transfer penalty. The values were determined from literature and reflect average passengers' valuation of time (Yap et al., 2024). The term with the unweighted P-space representation, $(\mathcal{D}_{P^u} - 1)$, counts the number of transfers. The generalised travel time between nodes i and j is denoted by d_{ij} and equals the corresponding element in the generalised distance matrix $[\mathcal{D}]_{ij}$.
In the next step, the time budget t_b is introduced, i.e. the maximum total travel time allowed. Then there exists an edge between nodes i and j if $d_{ij} < t_b$. The adjacency matrix A with entries a_{ij} of the access graph is thus obtained from the generalised distance matrix:

$$a_{ij} = \begin{cases} 1 & d_{ij} \leq t_b \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The edge set of the access graph thus depends on the time budget. $G_{\mathcal{A}}$ obtained in this process is an undirected unweighted graph. Although the standard graph representations are implicit in its construction, the edges connect nodes reachable within the time budget by following any shortest path. Therefore, node degree D of $G_{\mathcal{A}}$ is a direct measure of stop-level accessibility.
The evolution of the access graph is observed with increasing the time budget from 0 to $\max_{i,j} d_{ij}$. At $t_b = 0$ there are no edges in the access graph (i.e. $G_{\mathcal{A}}$ is an empty graph), and at $t_b = \max_{i,j} d_{ij}$, $\mathcal{A}$ is a complete graph. The average degree $\overline{D}$ in each case is 0 and $N-1$, respectively. To examine global accessibility indicators, the shape of the average degree growth with increasing t_b is studied. For almost all cities in our case study, a logistic-like convergence to the maximum degree is observed. Thus, there is a point where the rate of average degree growth starts to decrease. In a continuous limit, this would correspond to the point of inflection, where a function changes from convex to concave, or vice-versa, and where typically the second derivative is zero. At this point, from the access perspective, the point at which the maximum of the first derivative is obtained, is interesting. While the access graph describes cumulative accessibility, the first derivative approximation gives an indication of interval accessibility (i.e. number of nodes reachable in the time interval $[t_b(k), t_b(k+1)]$, where k is the time step counter). Here, we are working with numerical approximations and use difference quotients $\frac{\Delta \overline{D}}{\Delta t_b}$ as approximations of the derivatives. In data, non-smooth first-order difference quotients, together with multiple points of inflection are observed for several cities in the case study. Thus, to get a robust measure, the time budget at the global maximum of $\frac{\Delta \overline{D}}{\Delta t_b}$ is observed. Note that this corresponds to the time interval with the highest interval accessibility. The value of $\delta t_M = t_M/\tau_{max}$, where t_M is the time budget at this maximum, and τ_{max} is the maximum generalised travel time in the network, is proposed as a global indicator of PTN accessibility. In addition, the value of the average degree at t_M, $\overline{D}_M$, can be observed.
Alongside the average degree, we observe the t_b-dependent degree distributions of the access graph for access equity assessment. The Gini coefficient is defined as:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} |x_i - x_j|}{2n^2\overline{x}}, \tag{3}$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the distribution mean. The Gini coefficient of the degree distribution at t_M, G_M, is used as an indicator of access equity. The Gini index is chosen here rather than skewness, because the degree distributions in our empirical analysis were found to often exhibit a bimodal shape.

3 Results and discussion

The proposed methodology was applied to a dataset of metro networks of 51 cities worldwide (Vijlbrief et al., 2022a,b). For each metro network, access graphs were built for varying time budgets. t_b was varied in the interval $[0, \tau_{max}]$, where τ_{max} is the maximum generalised travel time in the network, with progressively increasing steps of two minutes. The maximum travel time varies for each city and increases approximately logarithmically with the size of the network N (Figure 1). Note that N is correlated with city size and population.

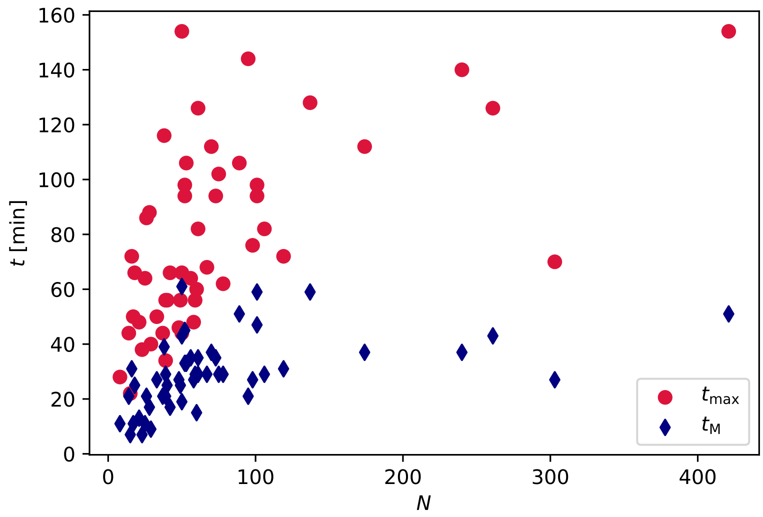

Figure 1: Scatter plots of the maximum generalised travel time t_{max} and time at fastest average degree growth t_M vs. number of nodes N. For each variable, one data point represents a single metro network.

The dependence of degree distributions and average degree on t_b results are shown in Figure (2). Each of the subplots represents a separate city and shows the time evolution of the access graph with the heatmaps representing node degree distributions. t_b is increased in steps of 2 minutes. In each subplot, the x-axis represents the time budget, and y-axis represents the degree. Heatmap cell colors represent the percentage of nodes in each degree bin. For a visual explanation of the results, see Figure 3 for a detailed explanation for an example of the Amsterdam metro network. The average degree of the access graph at t_b is shown with red markers. An S-shaped convergence to maximum degree is observed in almost all cases. Some networks exhibit a more complex behaviour (e.g. Athens, Los Angeles and Naples).

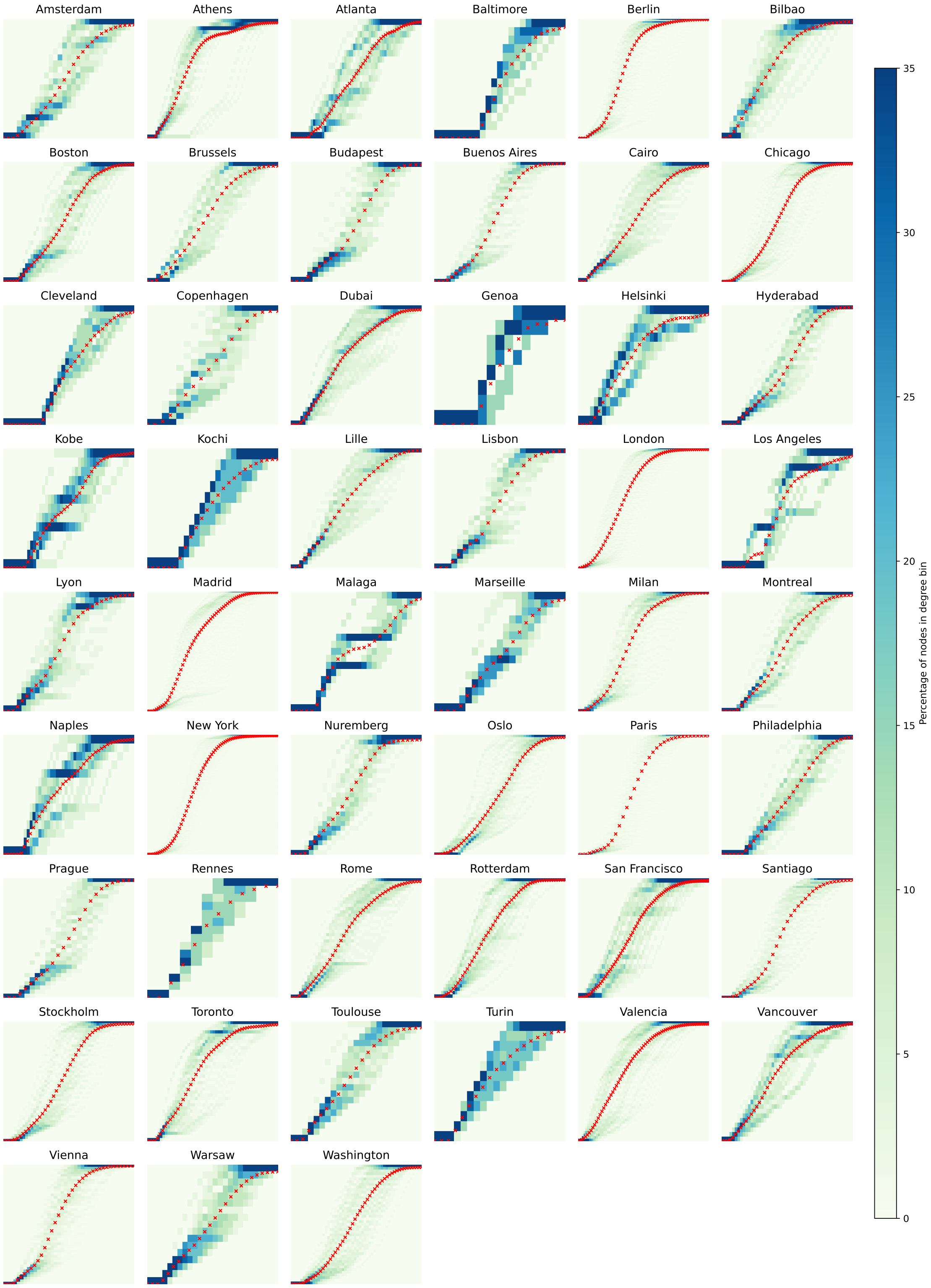

Figure 2: Degree distributions of access graphs with varying t_b. Each subplot shows the time evolution of the access graph with the heatmaps representing node degree distributions. t_b is increased in steps of two minutes. In each subplot, the x-axis represents the time budget, and y-axis represents the degree. Heatmap cell colors represent the percentage of nodes in each degree bin. Average degree of the access graph at t_b is shown with red markers.

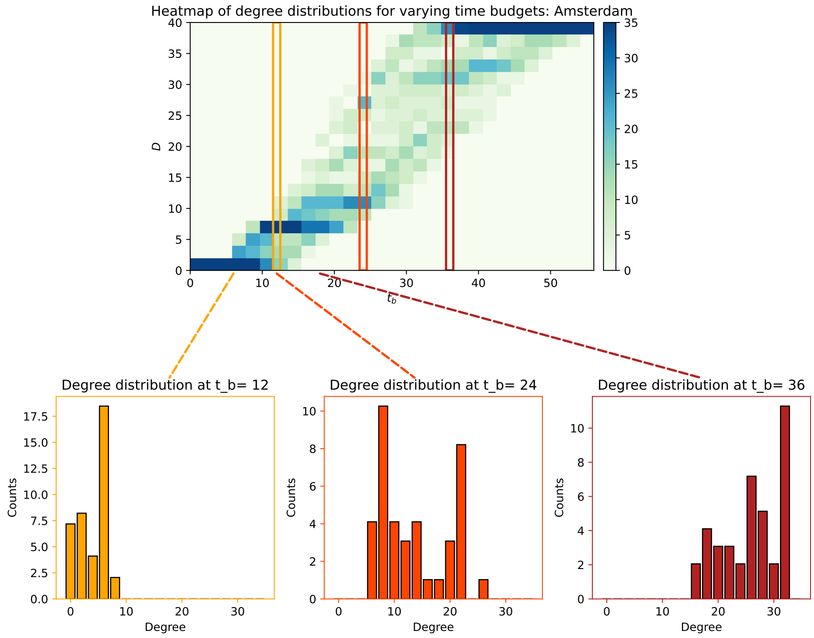

Figure 3: Explanation of the results on the example of the Amsterdam metro network. The upper plot shows the heatmap of degree distributions with varying time budgets. t_b is increased in steps of 2 minutes. The x-axis represents the time budget, and the y-axis represents the node degree. Heatmap cell color represents the percentage of nodes in each degree bin. Thus, each vertical column represents the color-coded histogram of the node degree distribution. For three values of the time budget, $t_b = \{6, 12, 18\}$, the heatmap visualisation is translated into the node degree histograms in the three bottom plots.

The most regular S-shaped behaviour is observed for the largest networks (New York, London and Paris, among others). In these cases, the value of t_M corresponds to the inflection point (i.e. the "turning point" of the S-curve). Since many of the other networks exhibit less smooth patterns, the t_b at the global maximum of the first difference quotient was taken as a more robust measure. Figure (4) shows the rate of growth of the average node degree, calculated in discrete steps as $y = \frac{\overline{D}(k+1)-\overline{D}(k)}{t_b(k+1)-t_b(k)}$, where k is the time step. In other words, the y-axis values represent the average change in the average degree of the access graph when the time budget is increased from the kth value $t_b(k)$ to the next, i.e. $(k+1)$th value. The value of t_b where y reaches the global maximum is taken as the value of t_M. For example, for the New York metro, the maximum change of $y = 8$ is observed at $t_b = 51$ min. This means that when the time budget is increased from 50 to 52 minutes, the average degree of the access graph increases by approximately 16, or 8 per minute, on average. The middle of the interval where this is achieved is taken as the maximum growth time accessibility indicator t_M, meaning $t_M = 51$ min for the New York metro network.

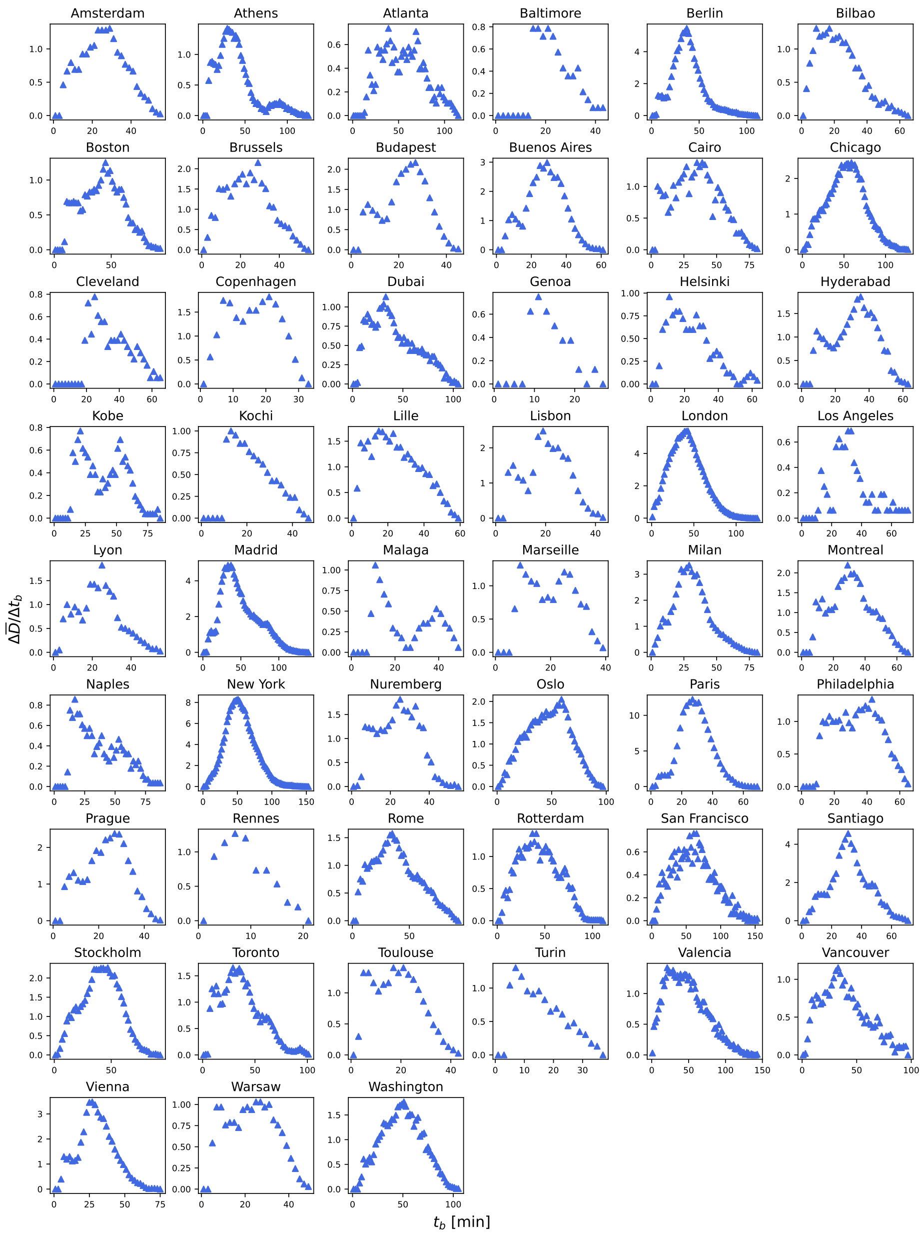

Figure 4: Rate of the average node degree growth, calculated in discrete steps as $y = \frac{\overline{D}(k+1)-\overline{D}(k)}{t_b(k+1)-t_b(k)}$, where k is the time step. The x-axis shows t_b in minutes, while the y-axis shows the average node degree growth (per minute). The value of t_b where y reaches the global maximum is taken as the value of t_M.

Similarly to t_{max}, t_M increases sublinearly with t_b (Figure 1), but converges much faster for large N. This indicates a generally better performance for larger networks. This is expected, as the individual time budget does not increase, or increases only marginally, with city size. Thus, both the absolute value of t_M and the relative value δt represent distinct accessibility indicators, the former reflecting the (approximately) universal passenger travel time budget, and the latter reflecting network performance properties. Note that low values of t_M and δt stand for higher accessibility.

The Gini coefficient G_M of the degree distribution of the access graph in $t_b = t_M$ is taken as an indicator of access equity. G can take values in the interval $[0, 1]$, where $G = 0$ means complete equality and $G = 1$ means the highest inequality. The relationship between G_M and N is shown in Figure (5). Again, we observe a roughly logarithmic behaviour with a group of outliers in the upper left corner. For the largest networks, with size $N \gtrapprox 150$, the value of the Gini coefficient is approximately constant, indicating that larger networks tend to offer more equal distributions of access, compared to the relation observed for small and moderately sized networks (i.e. the growth

of G_M stalls after this N).
The values of all variables are shown in Table (1). The Spearman correlation matrix for all variables is shown in Figure (6). Spearman correlation was chosen over Pearson due to non-linear relations between variables. Significant positive correlations ($r_S \approx 0.65$) are observed for the number of nodes N and maximum travel time t_{max}, and N and the time at fastest average degree growth t_M, which is expected. Similarly, a strong correlation between t_M and t_{max} is observed. Interestingly, the correlation of the absolute value of the time of fastest growth t_M and the value relative to maximum time $\delta t = t_M/t_{max}$ is relatively low ($r_S \approx 0.4$), suggesting the potential of using both values as accessibility indicators. Notable is the lack of correlation between δt and N, indicating a similar behaviour over all sizes of networks. On the other hand, the strong positive correlation between δt and the average degree at the time of fastest growth $\overline{D}_M$ points to a universal behaviour of average degree growth (observed to be logistic-like in most cases). Moderate positive correlation of the Gini index of the degree distribution at t_M, G_M, and N, G_M and t_{max} indicate that larger networks also tend to have larger disparities in access. Significant negative correlation between the Gini index and δt indicates that inequity decreases with δt, complementing the previous observation on the relation to N.

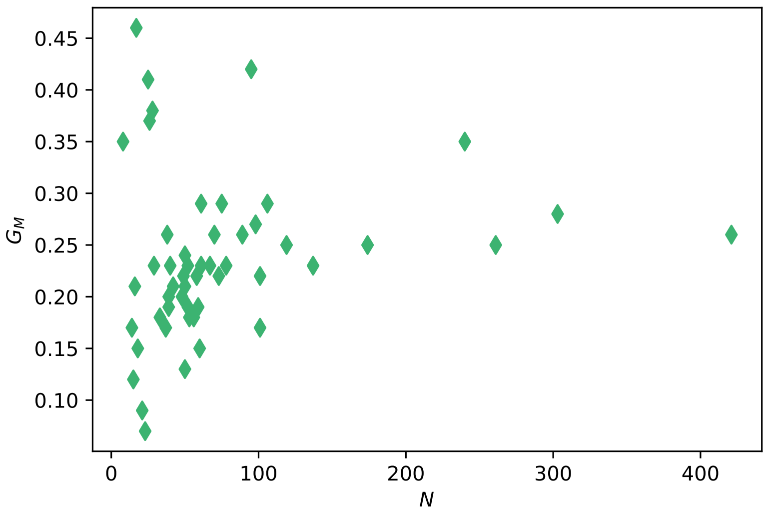

Figure 5: Scatter plot of the Gini coefficient G_M and network size N. For each variable, one data point represents a single metro network.

Table 1: Values of the studied variables for each city. N: number of nodes in the network; t_{max}: maximum generalised travel time; t_M: time of maximum average degree growth; $\delta t = \frac{t_M}{t_{max}}$; $\overline{D}_M$: average degree at t_M; G_M: Gini index at t_M.

City	N	t_{max}	t_M	δt	$\overline{D}_M$	G_M
Amsterdam	39	56	29	0.52	0.57	0.2
Athens	61	126	29	0.23	0.34	0.29
Atlanta	38	116	39	0.34	0.27	0.26
Baltimore	14	44	21	0.48	0.33	0.17
Berlin	174	112	37	0.33	0.48	0.25
Bilbao	42	66	17	0.26	0.34	0.21
Boston	52	94	45	0.48	0.51	0.23
Brussels	59	56	29	0.52	0.62	0.19
Budapest	48	46	27	0.59	0.6	0.2
Buenos Aires	78	62	29	0.47	0.49	0.23
Cairo	61	82	35	0.43	0.48	0.23
Chicago	137	128	59	0.46	0.57	0.23
Cleveland	18	66	25	0.38	0.17	0.15
Copenhagen	39	34	21	0.62	0.64	0.19
Dubai	53	106	33	0.31	0.39	0.18
Genoa	8	28	11	0.39	0.16	0.35
Helsinki	25	64	11	0.17	0.12	0.41
Hyderabad	56	64	35	0.55	0.55	0.18
Kobe	26	86	21	0.24	0.14	0.37
Kochi	21	48	13	0.27	0.09	0.09
Lille	60	60	15	0.25	0.26	0.15
Lisbon	50	44	19	0.43	0.38	0.21
London	261	126	43	0.34	0.53	0.25
Los Angeles	16	72	31	0.43	0.41	0.21
Lyon	40	56	25	0.45	0.51	0.23
Madrid	240	140	37	0.26	0.35	0.35
Malaga	17	50	11	0.22	0.06	0.46
Marseille	29	40	9	0.22	0.05	0.23
Milan	106	82	29	0.35	0.44	0.29
Montreal	67	68	29	0.43	0.41	0.23
Naples	28	88	17	0.19	0.11	0.38
New York	421	154	51	0.33	0.44	0.26
Nuremberg	49	56	25	0.45	0.48	0.22
Oslo	101	98	59	0.6	0.68	0.17
Paris	303	70	27	0.39	0.39	0.28
Philadelphia	50	66	43	0.65	0.67	0.13
Prague	58	48	27	0.56	0.57	0.22
Rennes	15	22	7	0.32	0.28	0.12
Rome	73	94	35	0.37	0.44	0.22
Rotterdam	70	112	37	0.33	0.37	0.26
San Francisco	50	154	61	0.4	0.49	0.24
Santiago	119	72	31	0.43	0.44	0.25
Stockholm	101	94	47	0.5	0.6	0.22
Toronto	75	102	29	0.28	0.35	0.29
Toulouse	37	44	21	0.48	0.55	0.17
Turin	23	38	7	0.18	0.09	0.07
Valencia	95	144	21	0.15	0.16	0.42
Vancouver	52	98	33	0.34	0.4	0.19
Vienna	98	76	27	0.36	0.38	0.27
Warsaw	33	50	27	0.54	0.57	0.18
Washington	89	106	51	0.48	0.51	0.26

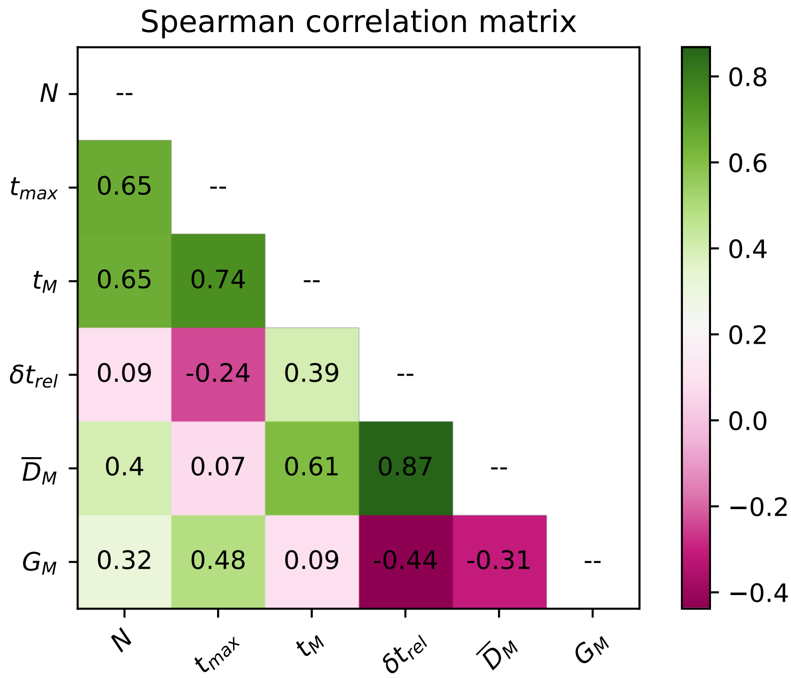

Figure 6: Spearman correlation matrix for accessibility variables.

4 Conclusions

We proposed a novel graph representation of public transport networks: the access graph $G_{\mathcal{A}}$. The edges in $G_{\mathcal{A}}$ are based on shortest paths, as determined from the classic L- and P-space representations, and directly connect nodes, reachable from one another within a given time budget. Average degree and degree distributions of the access graphs for 51 metro networks were examined and two global measures were proposed as accessibility indicators: the time budget of the fastest growth of average degree, t_M, and the relative value t_M/t_{max}. The Gini coefficient of the degree distribution at $t_b = t_M$ was used as an access equity indicator.
We see a number of venues for future research. First, we expect a detailed analysis of the degree distributions and average degree growth with a focus on understanding differences in behaviour among networks to offer new insights into accessibility. Second, a more detailed analysis of the access graph introducing other indicators and examining the behaviour from several perspectives will provide a more comprehensive equity analysis. Most importantly, we believe that the introduced access graph offers numerous opportunities for advancing network science and PT accessibility research. Among the most promising ones we see: *i*) detailed analyses of the t_b-varying connectedness and topology of $G_{\mathcal{A}}$, *ii*) studying the impacts of disturbances on accessibility via simulations, and *iii*) introducing land use data into the model.

References

Cats, O. (2017). Topological evolution of a metropolitan rail transport network: The case of stockholm. *Journal of Transport Geography*, *62*, 172–183.

de Regt, R., von Ferber, C., Holovatch, Y., & Lebovka, M. (2019). Public transportation in great britain viewed as a complex network. *Transportmetrica A: Transport Science*, *15*(2), 722–748.

Derrible, S., & Kennedy, C. (2011). Applications of graph theory and network science to transit network design. *Transport reviews*, *31*(4), 495–519.

Dimitrov, S. D., & Ceder, A. A. (2016). A method of examining the structure and topological properties of public-transport networks. *Physica A: Statistical Mechanics and its Applications*, *451*, 373–387.

Ding, R., Ujang, N., Hamid, H. B., Manan, M. S. A., Li, R., Albadareen, S. S. M., … Wu, J. (2019). Application of complex networks theory in urban traffic network researches. *Networks and Spatial Economics*, *19*, 1281–1317.

Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, *12*(2), 127–140.

Gori, S., Mannini, L., & Petrelli, M. (2020). Equity measures for the identification of public transport needs. *Case Studies on Transport Policy*, *8*(3), 745–757.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, *25*(2), 73–76.

Luo, D., Cats, O., van Lint, H., & Currie, G. (2019). Integrating network science and public transport accessibility analysis for comparative assessment. *Journal of Transport Geography*, *80*, 102505.

Šfiligoj, T., Peperko, A., Bajec, P., & Cats, O. (2025). Node importance corresponds to passenger demand in public transport networks. *Physica A: Statistical Mechanics and its Applications*, 130354.

Shanmukhappa, T., Ho, I. W.-H., Chi, K. T., & Leung, K. K. (2019). Recent development in public transport network analysis from the complex network perspective. *IEEE Circuits and Systems Magazine*, *19*(4), 39–65.

Vijlbrief, S., Cats, O., Krishnakumari, P., van Cranenburgh, S., & Massobrio, R. (2022a). *A curated data set of l-space representations for 51 metro networks worldwide.* 4TU.ResearchData. Retrieved from `https://data.4tu.nl/articles/dataset/A_curated_data_set_of_L-space_representations_for_51_metro_networks_worldwide/21316824/1` doi: 10.4121/21316824.v1

Vijlbrief, S., Cats, O., Krishnakumari, P., van Cranenburgh, S., & Massobrio, R. (2022b). *A curated data set of p-space representations for 51 metro networks worldwide.* 4TU.ResearchData. Retrieved from `https://data.4tu.nl/articles/dataset/A_curated_data_set_of_P-space_representations_for_51_metro_networks_worldwide/21316950/2` doi: 10.4121/21316950.v2

Yap, M., Wong, H., & Cats, O. (2024). Passenger valuation of interchanges in urban public transport. *Journal of Public Transportation*, *26*, 100089.