Towards Reducing School Segregation by Intervening on Transportation Networks

Dimitris Michailidis, Mayesha Tasnim, Sennay Ghebreab, and Fernando P. Santos

Civic AI Lab, Socially Intelligent Artificial Systems
Informatics Institute, University of Amsterdam
{d.michailidis, m.tasnim, s.ghebreab, f.p.santos}@uva.nl

Abstract. Urban segregation is a complex phenomenon associated with different forms of social inequality. Segregation is reflected in parents' school preferences, especially in context of free school choice modes. Studies have shown that parents consider both distance and demographic composition when selecting schools for their children, potentially exacerbating levels of residential segregation. This raises the question of how intervening on transit networks — thereby affecting school accessibility to citizens belonging to different groups — can alleviate spatial segregation. In this work-in-progress paper, we propose a new agent-based model to explore this question. Conducting experiments in synthetic and real-life scenarios, we show that improving access to schools via transport network interventions can lead to a reduction in school segregation over time. The mathematical framework we propose provides the basis to simulate, in the future, how the dynamics of citizens preferences, school capacity and public transportation availability might contribute to patterns of residential segregation.

Keywords: Transportation Networks · One-sided Matching · Agent-based Simulations · Dynamic Preferences

1 Introduction

Urban segregation is a complex phenomenon that reverberates across multiple socio-economic contexts — from social mobility to educational opportunities. In the context of education, centralized school admissions systems such as *Deferred Acceptance* and *Random Serial Dictatorship* have been popularized across the world for their simplicity and fairness in student allocation [8, 2]. However, school segregation can emerge in such preference-based systems, reflecting (or even amplifying) existing residential segregation patterns [6]. There is evidence that parents do not send their children to schools in their residential neighborhoods; if they did, schools would be less segregated than how they currently are [12].

Although parents might prefer schools outside their neighborhoods, distance and commuting time are important factors for attending a school [6]. With the exception of high-income households, most do not tend to move house and thus

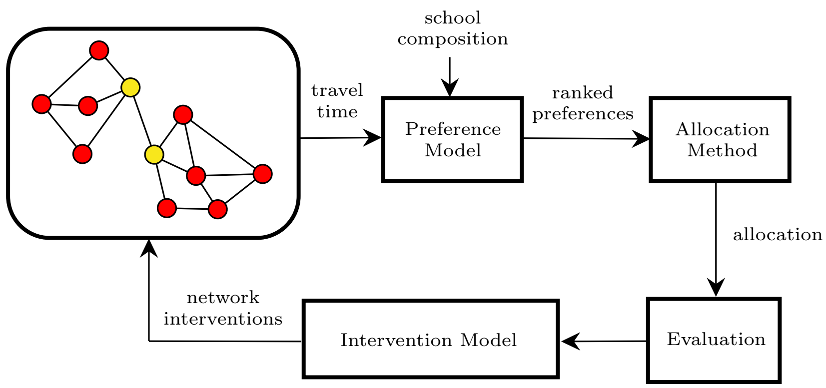

Fig. 1. Proposed agent-based model to study the impact of transport network interventions on school segregation. We consider an environment where citizens, schools, and a transportation graph are distributed in space (Section 2.1). At each round, agents A generate preferences for schools F, using a preference model (section 2.2). Agents are assigned to schools via an allocation method (Section 2.3), which is evaluated on segregation (Section 2.4). An intervention model creates edge-based interventions to the transportation network, aiming to improve segregation (Section 2.5).

their choice is limited by their location [3]. Intervening on public transportation networks can thereby affect segregation, by allowing citizens from different societal groups to attend a wider set of schools. This raises a natural question: *Can transportation networks be designed, or extended, to efficiently reduce school segregation?*

Here we resort to agent-based modeling (ABM) to explore the previous question. Prior studies focused on the complexity of residential and school segregation via ABMs [15, 6], and preference models based on both school composition and distance have been explored [6, 14]. However, these works do not study the effect of strategically increasing accessibility to specific schools. Graph-based interventions have been utilized before to reduce accessibility inequality [10], but not to tackle school segregation. We assess whether graph-based transportation interventions can be used to reduce disparities in group composition within schools, under a centralized admission system.

We test transport network intervention strategies based on greedy optimization of classic graph centrality measures such as *closeness*, *betweenness*, and *degree* centrality. We conduct experiments in a synthetic and a real-life environment in the city of Amsterdam and show that targeted interventions can lead to a significant reduction in segregation over time.

2 Methods

2.1 Environment: Citizens, Transportation, and Schools

We model the environment as an undirected graph $\mathbb{G} = (V, E)$, where $V = \{v_1, ..., v_{n_v}\}$ are nodes, one for each census tract in the city, and $E = \{e_{i,j}\}, i, j \in V, i \neq j$ are edges that represent transportation connections between nodes. For the sake of simplicity, the edges are unweighted, but the model can be used with weighted edges too (e.g., representing transportation times). We define the shortest path between i and j as $t_{i,j},\ i, j \in V$.

We define a set of N agents (citizens), $A = \{a_1, ..., a_N\}$. An agent is characterized by its residence node $v_a \in V$. Each agent is belongs to a group $g \in G$, defined based on characteristics such as ethnicity, income, or other socio-economic status. Finally, each agent has a homophily attribute, $h_i \in [0, 1]$, defining a preference for an optimal fraction of agents from the same group attending a school [6, 11]. Note that agents are abstract entities that represent students in a city.

We define schools $f \in F$, which are located in nodes $v_f \in V$. Each school is associated with a capacity (maximum number of allowed agents) $s_f \in [0, N]$ and a group composition (fraction of assigned agents from each group) $c_{g,f} \in [0, 1],\ g \in G$. Note that $\sum_g c_{g,f} = 1,\ \forall f \in F$.

2.2 Preference Model

At every round, each agent $a_i \in A$ creates a preference list $P_i \subseteq F$, over schools. Each school appears once. The preference list is based on a utility function $U_{if}, f \in F$, and schools are sorted in descending order. We adopt the widely used Cobb-Douglas utility function, based on a function of school composition $C : c_{g,f} \to \mathbb{R}$ and travel time from the agent's residence to the school $t_{i,f}$ [6, 14]

$$U_{i,f} = c_{g,f}^{\alpha}\ t_{i,f}^{(1-\alpha)}, \tag{1}$$

where g denotes the group that agent a_i belongs to and $0 \leq \alpha \leq 1$ is a parameter that controls the weight of the group composition over the travel time. Travel time is normalized by the maximum value and is calculated as follows [6]:

$$t'_{i,f} = \begin{cases} \frac{t_{max,i} - t_{i,f}}{t_{maxi} - t_{min,i}}, & \text{if } t_{i,f} \leq t_{max,i} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

For the school composition, we use a single-peaked utility function, that is maximized when the number of agents of the same group in a school $x_{g,f}$ is equal to the homophily attribute h_i [6, 14]. Values above h_i incur a constant penalty M:

$$C(x_{g,f}, h_i, M) = \begin{cases} \frac{x_{g,f}}{h_i}, & \text{if } x_{g,f} \leq h_i \\ M + \frac{(1-x_{g,f})(1-M)}{1-h_i}, & \text{if } x_{g,f} > h_i \end{cases} \tag{3}$$

M controls the level of dissatisfaction when the fraction of similar agents exceeds the optimal h_i. With this formulation interventions in the transportation network are performed to reduce the travel time $t_{i,f}$ of agents to school, with the goal of increasing utility towards more segregated schools.

2.3 Allocation Method

Once the preference lists P have been generated at each simulation round for all $a_i \in A$, they are then provided as input to an allocation method R. R is defined as a function $R : P \rightarrow F$ which takes as input a preference list p_i for agent a_i and capacity s_f for all $f \in F$ and assigns a school $f_i \in p_i$. Random Serial Dictatorship (RSD) is a popular mechanism for one-sided matching between schools and students [2]. In RSD a lottery number is first uniformly drawn for each student. The students are then serially allocated to the top-preferred school with remaining capacity in increasing order of the lottery. For our simulations we implement RSD and perform allocations at every round; schools have, overall, capacity to allocate all students, i.e., $\sum s_f \geq N$. Additionally, for each student the preference model from Section 2.2 provides a ranking for all schools, and RSD can allocate all students. The allocation result is then aggregated for evaluation.

2.4 Allocation Evaluation

After each simulation round, the allocation of agents to schools is evaluated on segregation. To measure segregation, we use the Dissimilarity Index (DI), a measure that captures the differences in the proportions of agents from two groups assigned to a school [7]. DI has been widely used in assessing segregation, as it takes into account the total number of agents from each group, making it suitable to use even when one group is a minority [1]. DI is defined as follows:

$$DI = \frac{1}{2} \sum_{f=1}^{|F|} \left| \frac{g_{1,f}}{G_1} - \frac{g_{2,f}}{G_2} \right|, \quad DI \in [0, 1] \tag{4}$$

Where $g_{j,f}$ is the number of agents of group j in school f; G_j is the number of agents in group j. Segregation is minimum (maximum) when $DI = 0$ ($DI = 1$).

2.5 Intervention Model

We explore the impact that intervening on public transport networks has on school choices. By improving transportation, we aim to elevate the rank of schools composed of majority groups in the preference lists of minority groups, increasing their accessibility to popular (yet distant) schools. Transport interventions are performed in the form of graph augmentations, by creating a new edge set $E' : \mathbb{G}, B \rightarrow \mathbb{G}'$ to the spatial graph, under a budget B [10]. It follows that $\mathbb{G}' = (V, E \cup E')$. Interventions can be seen as a proxy to the creation/expansion of

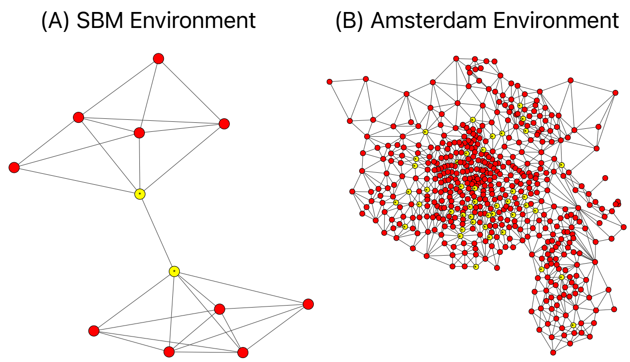

Fig. 2. We study synthetic (left image) and real (right image) environments. Nodes represent neighborhoods and yellow nodes (marked with *) indicate nodes with schools.

public transportation lines in a real city, such as bus, metro or tram. We constrain the total number of interventions to a budget B, reflecting resource limitations.

The goal of interventions is to find the best set of edges E' to add to the graph, such that total segregation is reduced. Segregation depends on the allocation method (section 2.3), which has a random element to it. Therefore, optimizing directly for the dissimilarity index is not possible. We look for targeted interventions that increase accessibility to certain schools for certain groups, aiming to affect the agent's preferences in such a way that segregation is reduced.

We test two classes of greedy interventions: 1) **Centrality** and 2) **Group-based Centrality Optimization**. We identify the schools that have the lowest network centrality measure (*closeness*, *betweenness* or *degree*) [4] with respect to any group and then add the intervention that leads to the maximum increase in that node's corresponding 1) centrality or 2) group-based centrality.

3 Experimental Setup

We perform experiments on two graph environments: a real-life city environment based on Amsterdam neighborhoods, demographic and transportation data; and a synthethic environment based on the stochastic block model (SBM) [9], which allows us to have full control over the level of modularity and segregation in a hypothetical city. For more details please refer to Appendix B and Fig. 2.

4 Preliminary Results

In Figure 3, we present the 95% confidence interval of the Dissimilarity Index on each simulation round, over a total of 50 rounds. Our preliminary experiments

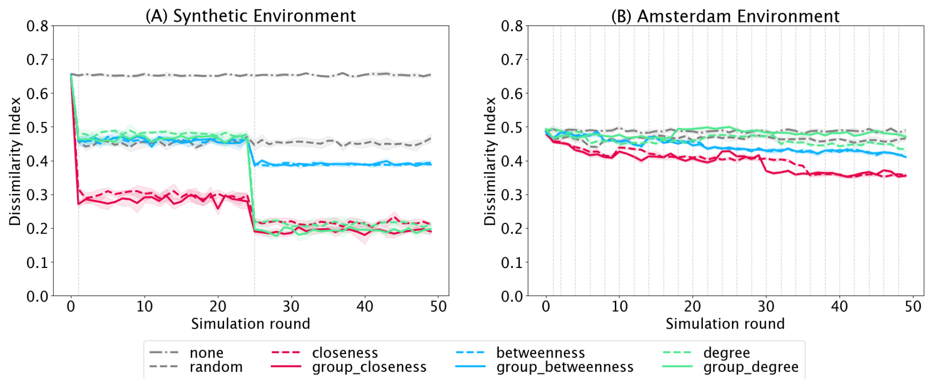

Fig. 3. We show that targeted interventions in the network can significantly decrease segregation over time. Strategies based on closeness perform best over other centrality measures. Vertical dashed lines indicate rounds with graph interventions.

show that, under the settings outlined above, all targeted network intervention strategies proposed in Section 2.5 lead to a significant reduction of segregation over time, when compared to a no-intervetion scenario (none) or random interventions. Specifically, we observe that greedy interventions aimed at increasing the closeness of the least-accessible nodes lead to the highest reduction of segregation over time. We also observe that degree-based interventions can have similar effects to closeness, but only in small networks, like SBM. This is because, when the number of nodes is low, increasing a the degree of a node also increases its closeness to other nodes. A betweenness-based strategy reduces segregation and outperforms degree-based ones in a bigger environment, like that of Amsterdam. Finally, there are seemingly not big differences between centrality and group-based centrality strategies, but depending on the budget, group-based closeness can outperform its classic counterpart.

5 Conclusion and Future Work

In this work-in-progress paper, we used an agent-based simulation model to study the impact of transport network interventions on school segregation, under the prevalence of a centralized school choice algorithm. We have demonstrated in both a synthetic and a real-life environment that, by affecting citizens preferences for particular schools, targeted transportation interventions can ultimately reduce school segregation over time. In the future, we plan to further experiment with the parameters of the preference model, to assess the sensitivity of network interventions to different types of agent school preferences. We plan to further experiment with group-based interventions, aiming at identifying the contexts where they become more efficient than centrality-based interventions.

Acknowledgements This research was supported by the Innovation Center for AI (ICAI, The Netherlands) and the City of Amsterdam.

References

1. Abbasi, S., Ko, J., Min, J.: Measuring destination-based segregation through mobility patterns: Application of transport card data. Journal of Transport Geography **92**, 103025 (Apr 2021). https://doi.org/10.1016/j.jtrangeo.2021.103025
2. Abdulkadiroğlu, A., Sönmez, T.: Random serial dictatorship and the core from random endowments in house allocation problems. Econometrica **66**(3), 689–701 (1998)
3. Boterman, W.R.: Socio-spatial strategies of school selection in a free parental choice context. Transactions of the Institute of British Geographers **46**(4), 882–899 (2021). https://doi.org/10.1111/tran.12454
4. Chen, D., Lü, L., Shang, M.S., Zhang, Y.C., Zhou, T.: Identifying influential nodes in complex networks. Physica a: Statistical mechanics and its applications **391**(4), 1777–1787 (2012)
5. Crescenzi, P., D'angelo, G., Severini, L., Velaj, Y.: Greedily Improving Our Own Closeness Centrality in a Network. ACM Transactions on Knowledge Discovery from Data **11**(1), 9:1–9:32 (Jul 2016). https://doi.org/10.1145/2953882
6. Dignum, E., Athieniti, E., Boterman, W., Flache, A., Lees, M.: Mechanisms for increased school segregation relative to residential segregation: a model-based analysis. Computers, Environment and Urban Systems **93**, 101772 (Apr 2022). https://doi.org/10.1016/j.compenvurbsys.2022.101772
7. Duncan, O.D., Duncan, B.: A Methodological Analysis of Segregation Indexes. American Sociological Review **20**(2), 210–217 (1955). https://doi.org/10.2307/2088328, publisher: [American Sociological Association, Sage Publications, Inc.]
8. Erdil, A., Ergin, H.: What's the matter with tie-breaking? improving efficiency in school choice. American Economic Review **98**(3), 669–689 (2008)
9. Holland, P.W., Laskey, K.B., Leinhardt, S.: Stochastic blockmodels: First steps. Social Networks **5**(2), 109–137 (Jun 1983). https://doi.org/10.1016/0378-8733(83)90021-7
10. Ramachandran, G.S., Brugere, I., Varshney, L.R., Xiong, C.: GAEA: Graph Augmentation for Equitable Access via Reinforcement Learning. arXiv:2012.03900 [cs] (Apr 2021), arXiv: 2012.03900
11. Schelling, T.C.: Dynamic models of segregation. The Journal of Mathematical Sociology **1**(2), 143–186 (Jul 1971). https://doi.org/10.1080/0022250X.1971.9989794, publisher: Routledge _eprint: https://doi.org/10.1080/0022250X.1971.9989794
12. Sissing, S., Boterman, W.R.: Maintaining the legitimacy of school choice in the segregated schooling environment of Amsterdam. Comparative Education **59**(1), 118–135 (Jan 2023). https://doi.org/10.1080/03050068.2022.2094580
13. Sousa, S., Nicosia, V.: Quantifying ethnic segregation in cities through random walks. Nature Communications **13**(1), 5809 (Oct 2022). https://doi.org/10.1038/s41467-022-33344-3, number: 1 Publisher: Nature Publishing Group
14. Stoica, V.I., Flache, A.: From Schelling to Schools: A Comparison of a Model of Residential Segregation with a Model of School Segregation. Journal of Artificial Societies and Social Simulation **17**(1), 5 (2014)

15. Zuccotti, C.V., Lorenz, J., Paolillo, R., Rodríguez Sánchez, A., Serka, S.: Exploring the dynamics of neighbourhood ethnic segregation with agent-based modelling: an empirical application to Bradford, UK. Journal of Ethnic and Migration Studies **49**(2), 554–575 (Jan 2023). https://doi.org/10.1080/1369183X.2022.2100554

Appendix

A Intervention Methods

Section 2.5 introduces the design choice to test two classes of greedy algorithms in the intervention model of the ABM. The algorithms and their usage as intervention methods are discussed below:

A.1 Greedy Centrality Optimization

Making a school more accessible is a non-trivial optimization problem, especially for large graphs [5]. We use a greedy algorithm to approximate the optimal set of interventions to apply to the graph with respect to accessibility. This translates to increasing a school node centrality $\mathbb{C}$ with respect to the other nodes. We evaluate strategies based on the classic graph measures of *closeness* ($\mathbb{C}_C$), *betweenness* ($\mathbb{C}_B$), and *degree* ($\mathbb{C}_D$) centrality.

At every intervention step, we find the school that has the lowest centrality measure with respect to any group and then add the intervention that leads to the maximum increase in this node's centrality. The process is described in Algorithm 1.

Algorithm 1 Greedy Centrality Optimization

Input $\mathbb{G} = (V, E)$

$E' \leftarrow \{\}$
for $b = 1, 2, ...B$ **do**
  $v_{g_{min}} \leftarrow argmin\{\mathbb{C}(v, g) \mid v \in V,\ g \in G\}$
  $\mathbb{C}_{max} = 0$
  $e_{max} \leftarrow null$
  for $u \in V,\ u \neq V$ **do**
    $e \leftarrow (u, v)$
    Compute $\mathbb{C}(v_{g_{min}}, E \cup E' \cup e)$
    if $\mathbb{C}(v_{g_{min}}, E \cup E' \cup e) > \mathbb{C}_{max}$ **then**
      $\mathbb{C}_{max} = \mathbb{C}(v_{g_{min}}, E \cup E' \cup e)$
      $e_{max} \leftarrow e$
    end if
  end for
  $E' \leftarrow E' \cup e_{max}$
end for

Output $\mathbb{G}' = (V, E \cup E')$

A.2 Group-based Centrality

Classic centrality measures fail to capture group dynamics in a graph. In segregated environments like cities, central areas can exhibit high closeness centrality, despite having low accessibility to specific groups. Examples of this phenomenon include cities where low-income households concentrate in the outskirts, while high-income households are situated closer to the center. To account for this disparity in measurement, we introduce group-based extensions of the classic centrality measures $\mathbb{C}^g,\ g \in G$, that take into account the distribution of groups within nodes. These are namely group-based closeness $\mathbb{C}^g_C$, betweenness $\mathbb{C}^g_B$ and degree $\mathbb{C}^g_D$. Let $D_g,\ g \in G$ be the distribution of group g on all nodes V in the network such that $\sum_g D_g = 1$.

Group-based Closeness Centrality Group-based closeness $\mathbb{C}^g_C$ of a node $v \in V$ is defined as the reciprocal of the sum of travel times from all other nodes u, weighted by the fraction of agents of group g in u, $p(g|u)$.

$$\mathbb{C}^g_C(v) = \sum_u \frac{1}{t(u,v)\ p(g|u)} \tag{5}$$

Where $t(u,v)$ is the travel time between nodes u and v.

Group-based Betweenness Centrality Group-based betweenness $\mathbb{C}^g_B$ of a node $v \in V$ is defined as the number of shortest paths σ from all nodes $o \in V$ to all nodes $d \in V, o \neq d$, that pass through v, weighted by the fraction of agents of group g in d. $p(g|d)$.

$$\mathbb{C}^g_B(v) = \sum_{o \neq v \neq d} \frac{\sigma_{t_{o,d}(v)}}{\sigma_{t_{o,d}}} p(g|d) \tag{6}$$

Group-based Degree Centrality Group-based degree $\mathbb{C}^g_D$ of a node $v \in V$ is defined as the total number of edges connected to a node $E_v = e_{u,v}$, $u \in V, u \neq v$, weighted by the fraction of agents of group g in u, $p(g|u)$.

$$\mathbb{C}^g_D(v) = \sum_{u \in V, u \neq v, e_{u,v} \in E} p(g|u) \tag{7}$$

Optimizing for group-based centrality measures leads to interventions that target schools where specific groups are underrepresented, instead of arbitrarily increasing the centrality of a school.

B Simulation Environments

We perform experiments on two graph environments, a synthetic stochastic block model (SBM) [9] and a real city environment based on Amsterdam, Netherlands.

SBM Environment The SBM graph is specifically generated to form clusters of communities, where nodes are densely connected with other nodes in their community and scarcely connected with nodes outside of it. We generated an SBM graph of $n_v = 12$ nodes and $n_e = 27$ edges; nodes clustered in 2 communities, which represent the majority group of their respective nodes. The parameters are chosen speficially to create a highly segregated graph, in which we aim to study the impact of the proposed intervention strategies. In-community edge probability is set to 0.7 and out-community probability is set to 0.01. In Figure 2 (A) we show the realization of the SBM graph we used for the simulation.

Further, we generated a population of $N = 1000$ agents and sampled both their residence node and their group membership, from a total of 2 groups. Group samples are chosen in such a way that each group, within their respective community has a majority of ≥ 0.8 and outside of their community a minority of ≤ 0.2. Since agents do not start at random nodes and there is no moving action in the model, we assume that the optimal fraction of similar agents is equal to the fraction of the majority group of each node. Formally, the homophily parameter of an agent i in a node v is set to $h_{i,v} = \max\{c_{g,v}\},\ g \in G$, where $c_{g,v}$ is the composition of group g in node v.

Finally, we place two schools on the graph, located in the two most connected nodes of the SBM graph. The initial group composition of each school is set to be equal to the group composition of the node it is located in.

Amsterdam Environment To model the real-life environment of Amsterdam, we create a graph where census tracts are converted to nodes, which are connected with their neighboring tracts via an unweighted edge. This graph structure has recently been used to quantify segregation because it provides a scale-free and generalizable method [13]. In total, the graph consists of $n_v = 517$ nodes and $n_e = 1611$ edges. In Figure 2 (B) we show the graph used for the Amsterdam experiments.

Similar to SBM, we generate a population of $N = 7000$ agents. However, in this environment, agents are generated to represent the real-life population of Amsterdam and are split in groups of western (W) and non-western (NW) ethnic background. More details on the population can be found in Table B. Here the homophily parameter is set in the same way as in the SBM environment.

We use the publicly available Amsterdam secondary school dataset provided by DUO[1] which contains 47 secondary schools and their locations. We combine this information with the admissions dataset collected by OSVO[2] which provides

[1] Education Executive Agency: http://duo.nl

[2] The association of school boards in Amsterdam: https://www.verenigingosvo.nl/

the capacities for each school based on the admission results of the previous year.

Table 1. Parameters used in running the experiments.

	SBM	**Amsterdam**
Groups	g0, g2	W, NW
Total Population	1000	7000
Group Populations	524, 476	4547, 2453
Group Populations (%)	52%, 48%	65%, 35%
No. of Nodes	12	517
No. of Edges	27	1611
α, M	0.2, 0.6	0.2, 0.6
Budget (B)	1	1
Simulation Rounds	50	50
Allocation Rounds	5	5
Interventions	2	25

B.1 Simulation Parameters

For the experiments shown in this work, we follow the setup of Dignum et al. and set the relative weight of the composition in the preference model to $\alpha = 0.2$ and the constant $M = 0.6$ for both environments. All experiments are run over 50 simulation rounds, with 5 random serial dictatorship allocations at each environment. We perform 2 intervention rounds in the SBM environment with a budget of $B = 1$ each, while in Amsterdam, we perform a total of 25 intervention rounds, also with $B = 1$. Parameters including total number of intervention rounds and budget B are determined beforehand. Other parameters and their values used in the experimental studies are listed in Table 1.

At every simulation step of the agent-based model agents submit preferences and are allocated to schools. However, interventions are applied to the transport network in intervals. We evaluate the performance of the intervention strategies against a *null* baseline, where no interventions are being done, and against a *random* baseline, where interventions are performed randomly.