Transport Network Design Based on Origin/Destination Clustering During the COVID-19 Pandemic Use Case

Matthieu Guillot[1], Angelo Furno[1], El-Houssaine Aghezzaf[2,3], and Nour-Eddin El Faouzi[1]

[1] Univ Gustave Eiffel, Univ Lyon, ENTPE, LICIT, F-69518, Lyon, France.
[2] Department of Industrial Systems Engineering and Product Design
Faculty of Engineering and Architecture, Ghent University, Belgium
[3] Industrial Systems Engineering (ISyE), Flanders Make.
matthieu.guillot@univ-eiffel.fr

Abstract. During the COVID-19 health crisis, the demand for public transport has decreased, and even today has not recovered its pre-crisis level. In this paper, we propose a network design methodology on a bus network, which, starting from existing bus stops and a (decreasing) demand on them, constructs a set of lines with interesting properties. On the one hand, the lines created do not promote any geographical areas such as city centers, and ensure a certain equity from a geographical point of view. On the other hand, for any given trip, a user can take only one bus line, which limits the contamination locations. This methodology, based on origin/destination clustering, is decomposed into several independent steps that we explain separately. Then the methodology is tested on real data from the region of Lyon, in France.

Keywords: Network Design, Clustering, Covid-19

1 Introduction

In a public transport network, one of the challenges is to match supply with demand. Demand is defined as a set of trips that users want to take, as well as the number of users who want to take them. Supply, on the other hand, is the set of infrastructure, vehicles, and human resources that are in place to meet the demand. The match between supply and demand is an important element: if supply exceeds demand, then unnecessary costs are incurred, which is detrimental especially for a public operator. If, on the other hand, demand exceeds supply, then some trips will be difficult (delays, congestion, etc.), or even impossible, which causes problems if we want everyone to feel free to travel.

During the health crisis related to the COVID-19 pandemic, the demand for public transport was strongly impacted. During the lockdowns, of course, all mobilities have been very strongly reduced reaching a level of almost zero for public transport [2]. However, even nowadays, when a relative "back to normal" is observed, the use of public transport has still not recovered its pre-crisis level [4][6][14][11]. This phenomenon can be explained in several ways: on the one hand, teleworking, which increased during and after the successive lockdowns, may have changed the mobility habits of workers. Thus, if a public transport subscription was profitable for 5 days of work on site, it is less so if one teleworks two or three days a week. On the other hand, the fear of contamination in public transport may have pushed some users to individual modes (car, bicycle...). The increase in the number of kilometers of traffic jams in the major cities since January 2022, as well as the increase in cycling, corroborate this hypothesis. Thus, public transport operators are seeing their revenues decrease: in the Lyon region, it is several hundred million euros since the beginning of the crisis, and in the Île-De-France region, this figure reaches one billion euros in lost revenue.

As a consequence of the change in demand for public transport, the transport supply has to be adapted in order to satisfy a sufficient quality of service while being financially sustainable. When the change in demand is small, adaptation of frequencies, timetables and vehicle schedules may be sufficient. For example, D'Acierno et al. propose rescheduling strategies to cope with budget cuts, which can be adapted to a declining demand. Kang et al. focus on rescheduling the last trains in case of delays during the day, which results in a slight variation in demand during the day [9]. Concerning a periodic variation in demand related to vacations, Gkiotsalitis et al. focused on changes in schedules for leisure activities during these periods [8].

However, when demand is overly affected, simply changing the frequency of public transport does not sufficient for effective adaptation. It is then necessary to completely rebuild the lines: this is called network design or network redesign.

LeBlanc defines transport network design as the problem of finding the optimal frequencies of transit lines. Lee et al. also consider the possibility that the transit demand could change over time [10]. In this paper, the authors assume that the total demand (cars and transit network) is fixed, but the proportion of demand for transit network can vary. More recently Cipriani, Gori and Petrelli have proposed a case study of network design for the city of Rome, with multi-modal properties and complex road network topology [3]. Lo et al. consider network design under demand uncertainty for ferries [12]. Ukkusuri et al. focus instead on multi-stage network design in order to accommodate demand fluctuations and demand uncertainty [15].

In this article, we focus on a network design methodology on a bus network. As we have seen, the demand for public transport has been strongly impacted by the health crisis, but the evolution of this demand is uncertain. It is not yet known whether the changes that have occurred will last over time, become more important, or whether a return to a pre-crisis level will ever be achieved. Thus it is advisable to choose a mode that is easily adaptable to possible future changes. This is why we have chosen to focus on bus networks, because it is relatively easy to change bus lines or to use or not to use bus stops, which is more difficult with heavy infrastructures such as tramways, subways or trains.

In order to lie down on existing infrastructure, we use existing bus stops. From these existing bus stops and a known demand on them, we propose a methodology to create a new bus network. We ensure that (i) for each trip, it is necessary to take only one bus line and (ii) no geographical zone is *a priori* given a preferential treatment, guaranteeing a certain equity within the network we consider. This methodology is composed of several independent blocks, which can be modified without questioning the global path. We present here a version with simple assumptions, but it is possible to make each block more complex in order to improve the model in a modular way. The paper is organized as follows: in a first part we present the theoretical model. We present each block independently while explaining the main thread of the methodology. Then we apply the methodology to the bus network of Lyon to give and analyze results with real data.

2 Methodology

In this part, we theoretically present the methodology that from a set of bus stops and a demand on them, constructs a set of lines with interesting properties that we will analyze later in the paper. This methodology, based on clustering of origin/destination pairs, is decomposed into several blocks that follow each other but are independent. Figure 1 presents a diagram representing the different steps of the methodology. Let us now formalize this methodology.

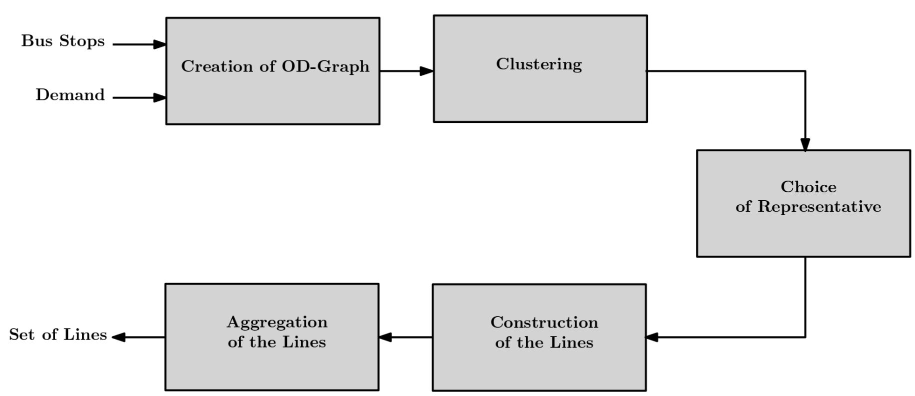

Fig. 1. Outline of the methodology

Let $G = (V, E)$ be a graph representing the bus network: V represents the $|V| = n$ bus stops, and E the $|E| = m$ road network axes that connect these bus stops. We assume that we have a demand on the bus stops, i.e. an origin/destination matrix OD such that for any $v_1, v_2 \in V$, $OD(v_1, v_2)$ represents the number of people wanting to go from stop v_1 to v_2. This demand is considered to be time-independent, but it may result from aggregating dynamic data.

We also assume that we know the minimal distance between each pair of vertices of G, i.e. that for any $v_1, v_2 \in V$, we know the value of the shortest path $d(v_1, v_2)$.

2.1 Creation of the OD-Graph

The first step is to create an auxiliary graph G', whose vertices represent the different origin/destination pairs of G (OD pairs for short) that are interesting, *i.e.* which origin and destination are not *too close*. Moreover, we want to link two vertices of G' if and only if the sum of the distance between the corresponding origins and destinations is not *too big*. More formally, we define a threshold d_{walk} which represents the distance we accept to walk, and we define a graph $G' = (V', E')$ with:

$$V' = \{v' = (o, d) \in V | d(o, d) > d_{walk}\} \tag{1}$$

$$E' = \{(v'_1 = (o_1, d_1), v'_2 = (o_2, d_2)) \in V' | d(o_1, o_2) + d(d_1, d_2) < d_{walk}\} \tag{2}$$

For instance, d_{walk} can be based on the distance that an average pedestrian can walk in a given time (15, 20min...). Thus the vertices of G' represent the OD pairs that we do not allow to walk.

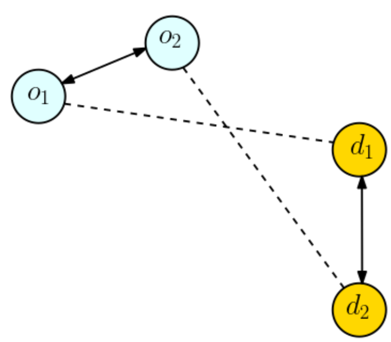

Fig. 2. Distance between two ODs (o_1, d_1) and (o_2, d_2)

Moreover, two OD pairs (v'_1, v'_2) are linked in G' if and only if doing the trip v'_2 instead of v'_1 induces an increase of the distance of maximum d_{walk}. Figure 2 illustrate the distance between two ODs. So G' represents the graph of *close* OD of G.

By abusing the notations, we call $v \in V'$ the path (o, d) such that $v = (o, d)$.

2.2 Clustering on OD-Graph

The second step is to apply a clustering algorithm to G'. There are several known clustering algorithms.

The spectral clustering is a clustering algorithm in which one can choose the number of clusters as input parameter [16]. It is based on the Laplacian matrix of the graph on which it is applied, and evaluates the differences between the eigenvalues of this matrix in order to create clusters which can be related to the "almost connected components" of the graph. Thus clusters have both a high connectivity within them, and a low connectivity between clusters. On the one hand, the advantage of this algorithm is that it is very fast and gives clusters that have good properties. One of the big issues, on the other hand, is to decide the number of clusters. There are statistical methods to decide a "natural" number of clusters [13], but in our case, the number of clusters will be strongly related to the number of output lines. Thus, the fact that the number of clusters is an input parameter of the clustering algorithm allows us to size the network.

The Louvain method does not take the number of clusters as an input parameter. This algorithm computes a set of clusters on the graph, and optimizes the modularity [5]. The modularity is a metric that relates to both the connectivity within the cluster, but also to the connectivity between clusters. Thus a cluster having a high modularity will have a strong connectivity between the vertices composing it, but a very weak one between the nodes of the cluster and those of other clusters.

For the following, let us assume that we apply a spectral clustering algorithm to G'. We define a number $c \in \mathbb{Z}^+$ of clusters. We thus obtain a set of clusters $\mathcal{C} = \{C_1, C_2, ..., C_c\}$.

Thus, given the definition of G', $\mathcal{C}$ corresponds to a set of clusters grouping similar ODs. Indeed, two nodes of G' are linked if and only if their respective ODs can easily swap. The next step is to cleverly choose a representative of each cluster.

2.3 Choice of Representative

In this third step, we will consider each cluster C_i separately. For each $i \in [\![1, c]\!]$, we want to choose a *good* representative of C_i. A representative $r(C_i) \in C_i$ must have some properties.

First, it must be *central*: for each vertex $v \in C_i$, the distance between v and $r(C_i)$ must not be too large. Here we define the distance D as the minimum number of edges connecting two vertices. Thus, if $r(C_i)$ is central, then making the path $r(C_i)$ instead of $v \in C_i$ will not induce a too big increase of the distance to walk.

On the other hand, we would like to choose $r(C_i)$ taking into account the demand. Since $v \in V'$ corresponds to an OD, there is a demand associated to v. Thus, for any $v' = (o, d) \in V'$, we define $OD(v') = OD(o, d)$. To choose the representative of C_i, we would like to promote vertices with a large demand.

We define the weighted distance D^{C_i} on the vertices of C_i as follows. For all $v \in V'$:

$$D^{C_i} = \min_{v \in C_i} \sum_{u \in C_i} D(u,v) OD(u) \tag{3}$$

Where $D(u,v)$ represents the minimum number of edges in G' between $u \in V'$ and $v \in V'$. In particular, we can compute D for any $(u,v) \in V'^2$ by computing a shortest path matrix by setting the cost of any edge to 1.

Finally, for any $i \in [\![1,c]\!]$, we define the representative $r(C_i)$ of C_i as the vertex of C_i minimizing the weighted distance:

$$r(C_i) = \arg\min_{v \in C_i} D^{C_i} \tag{4}$$

In the following, the representative of each cluster will be used as a starting point to create the lines of the bus network.

Note that this choice of representative is not the only one possible. We could also take less complex metrics but faster to compute, such as centrality in terms of degrees, i.e. take as representative of C_i the vertex with the highest degree. As indicated in the introduction, it is possible to take different assumptions at each step of this methodology.

2.4 Construction of the Lines and Aggregation

Now that we have defined for each cluster C_i a central representative that takes into account the demand, we will build bus lines from these representatives. For each $i \in [\![1,c]\!]$, we will define a line l_{C_i} whose terminus will be the origin and the destination associated to $r(C_i)$.

Let $i \in [\![1,c]\!]$, we can define l_{C_i} as the shortest path in G (in the initial graph of bus stops). The advantage of taking the shortest path is that in this case, by definition, l_{C_i} will be the line that connects these terminals the fastest. However, one can think of other choices because if one does not choose the shortest path, then one passes through potentially more bus stops, so the network serves a larger number of bus stops. Thus for $r(C_i) = (o_{C_i}, d_{C_i})$, we define l_{C_i} as:

$$l_{C_i} = (o_{C_i}, .., d_{C_i}) = SP_G(o_{C_i}, d_{C_i}) \tag{5}$$

where $SP_G(o_{C_i}, d_{C_i})$ is the shortest path in G between o_{C_i} and d_{C_i}.

Thus we create one line per representative. However, it is likely that some lines are redundant. Indeed, it is possible that the lines created in this way include a large number of bus stops in common. Figure 3 shows an example where such a configuration occurs. Cluster A contains ODs, and cluster B contains ODs. Thus, when we compute the representative of each cluster, and then the rows associated with the representatives, the line resulting from cluster A is contained in the line resulting from cluster B.

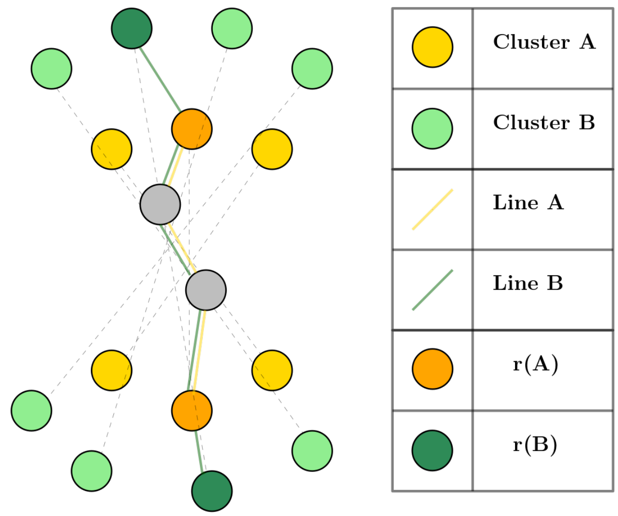

Fig. 3. Example of situation where two lines have common bus stops

We define $p_{com}(l_1, l_2)$ as the percentage of bus stops in common between l_1 and l_2.

$$p_{com}(l_1, l_2) = \frac{|l_1 \cap l_2|}{\min(|l_1|, |l_2|)} \tag{6}$$

Thus, if as in Figure 3, line A is completely included in line B, then $p_{com}(A, B) = 1$. If on the contrary there is no stop in common between two lines A' and B', then $p_{com}(A', B') = 0$.

To avoid the kind of situation where two lines have a lot of bus stops in common, we define a threshold $p_{aggreg} \in \mathbb{R}^+$ which represents the percentage of joint bus stops above which we will aggregate two lines. We apply the algorithm 1. At the end of this algorithm, we ensure that the percentage of joint bus stops between two lines of the network is at most p_{aggreg}.

Algorithm 1 Aggregate lines

Require: a set of lines $\mathcal{L} = l_1, .., l_c$
Ensure: a subset $\mathcal{L}' \subseteq \mathcal{L}$ such that $p_{com}(l, l') < p_{aggreg}$ for all pair $(l, l') \in \mathcal{L}'$

```
continue ← True
L_curr ← L
while continue do
  continue ← FALSE
  for l ∈ L_curr do
    for l' ∈ L_curr do
      if l ≠ l' AND p_com(l, l') > p_aggreg then
        if |l| < |l'| then
          Delete l from L_curr
        else
          Delete l' from L_curr
        end if
        continue ← TRUE
      end if
    end for
  end for
end while
return L_curr
```

Thus, by applying the methodology just described, we construct a set of bus lines. By construction, this network has some interesting properties that we will describe in the next section.

3 Analysis of the Network

The methodology we used to build a bus network is based on groupings of ODs. The principle is quite natural: it is about grouping ODs that we would be likely to swap as users of the network. Indeed, by defining d_{walk}, we define both the distance we accept to walk, and thus the minimal length of the ODs we consider. Moreover, this threshold represents the maximum distance that one will walk if one chooses $v \in V'$ rather than $v \in V$ if $(v, v') \in E'$. Once the similar ODs are grouped, we choose a central representative of each cluster, and we build a line from the OD corresponding to the representative. Finally, we aggregate the lines that have a high percentage of bus stops in common.

The bus network resulting from this methodology has some interesting properties. On the one hand, any user can be advised to take only one bus line, and this network guarantees a certain equity geographically.

Let $(o, d) \in V^2$ be a trip desired by a user of the network. We can very simply advise him a trip. Indeed, let $d(o, d) \leq d_{walk}$ and in this case the trip is supposed to be feasible by walking. We suppose that $d(o, d) > d_{walk}$. In this case, there exists $v' \in V'$ such that $v' = (o, d)$ by definition of V' (equation 1). Once we have applied the clustering algorithm on G', there exists $C_i \in \mathcal{C}$ such that $v' \in C_i$. Once the cluster containing v' is known, it is then sufficient to advise the line l_{C_i} created from $r(C_i)$. Figure 4 summarizes the assignment of a line to an OD.

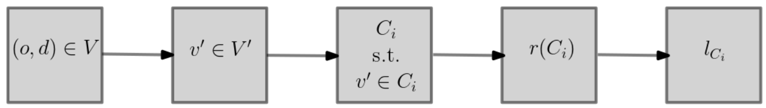

Fig. 4. Line to assign according to the OD wanted

Thus it is possible to use only one bus line for any trip one wants to make. In times of pandemic, this has several advantages. On the one hand, taking only one line means that there are no transfers, which limits the risks of contamination and the mixing of users. On the other hand, since there is only one line used by OD, it is easy to know the minimum capacity that a bus line must support. It is therefore easy to size the transport supply since it is easy to define the frequency of buses on each line in order to satisfy the demand.

Moreover, not only can we recommend only one bus line for each user's trip, but we can quantify the extra walking distance that the user will have to walk compared to the existing network. Indeed, for $v' = (o,d) \in V' \cap C_i$ a path (contained in the cluster C_i, it is enough to compute $D(v', r(C_i))$ (recall that $D(u,v)$ is the minimal distance, in terms of number of edges, between $u \in V'$ and $v \in V'$). The distance to walk in addition will be bounded by $D(v', r(C_i)) \times d_{walk}$. We thus have a performance guarantee on the access time and the egress time of our network.

Finally, to build the OD graph V', the demand has not been taken into account. In fact, the demand only affects the choice of the representative of each cluster. Thus we do not promote certain ODs *a priori* in our model. In most of the current public transport networks, the periphery/center city trips are often promoted. This is due to the fact that it is usually necessary to make connections, and that these are more efficient in terms of travel time if they are made in the city center. In our case, however, since no connections are necessary, these types of trips are not promoted over others. Thus the network resulting from our methodology performs efficiently for suburb-to-suburb trips, which are currently difficult, guaranteeing a certain equity in geographical terms.

In order to test our methodology on a concrete example to evaluate the resulting network, we applied it with real data on the region of Lyon, France.

4 Case Study: bus network of Lyon

We are interested in the bus network of Lyon. We will follow the methodology described in the previous section until we obtain a network of bus lines. We will follow the different steps while discussing the sensitivity to the input parameters.

For the topology of the network, we use the data of the transit network of Lyon (TCL) (see [7]). Moreover, for the demand, the OD has been built iteratively with the LICIT traffic simulation open-source tool SYMUVIA [1], combined with information from loop detectors in Lyon's urban area. SYMUVIA has been parametrized with different origin/destination inputs until the output traffic matches the real traffic data collected with the loop detectors. So the OD has been built to match real cases.

The network of Lyon has 468 bus stops. We show the bus stops and the corresponding road network on Figure 5

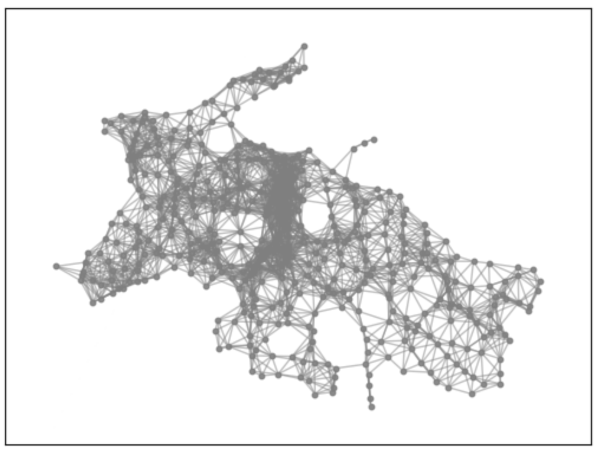

Fig. 5. Bus network for the urban area of Lyon, France

This network corresponds to G in the denomination of the methodology. The first step is to create G'. For this, we choose d_{walk} and we fix it by taking into account the average walking speed, and a maximum time that we accept to walk. Knowing that an average pedestrian walks at $3km.h^{-1}$, and that we fix a maximum walking time of 30min, we fix $d_{walk} = 1500m$.

Then, we apply a clustering algorithm. Here we choose the spectral clustering algorithm since it is one of those that allows to fix the number of clusters c as input parameter. We can choose several c values. We show in Figure 6, in spite of the fact that with such a number of clusters it is not possible to obtain good illustrations, an example with $c = 32$, $c = 50$ and $c = 150$

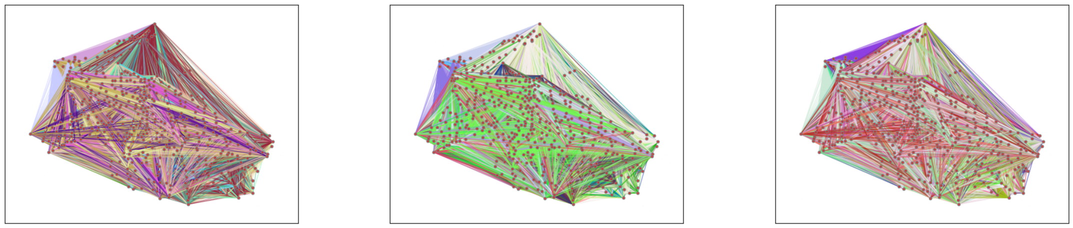

Fig. 6. Three examples of clustering of G' with (from left to right) 32, 50 and 150 clusters

Once the clusters are created, the next step is to choose the representatives of each cluster. For the same values of c, we show in Figure 8 the representatives of each cluster by connecting the corresponding origin and destination for each cluster.

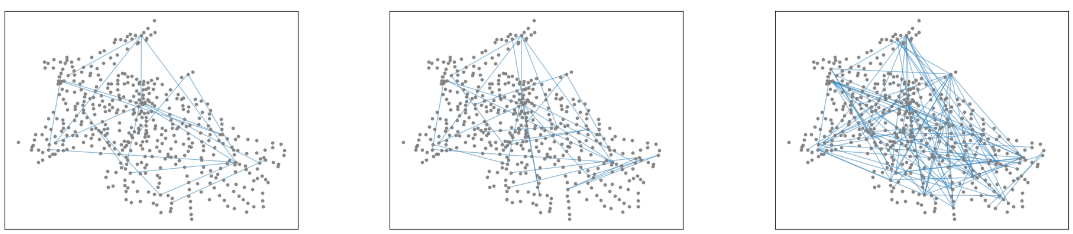

Fig. 7. Representatives of the clusters for (from left to right) 32, 50 and 150 clusters

We continue to apply the methodology described in the previous section by computing the shortest path between the origin and destination of each representative in G. We obtain the following network design for the same values of c.

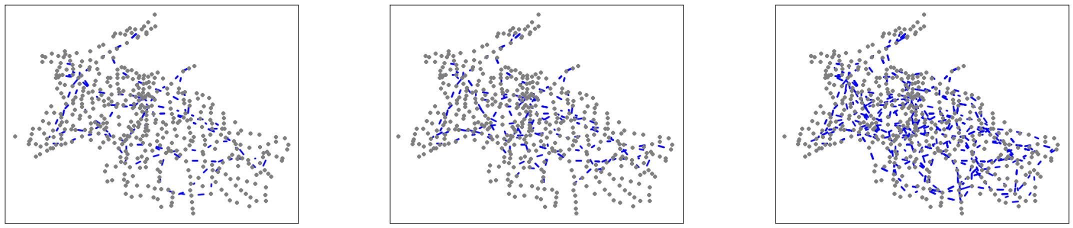

Fig. 8. Set of lines for (from left to right) 32, 50 and 150 clusters

For $c = 150$, we apply the aggregation procedure to reduce the number of lines and to ensure that the percentage of common bus stops between the lines is not too high. Figure 9 presents the lines at the end of the procedure for different values of p_{aggreg}, from $p_{aggreg} = 0.4$ to $p_{aggreg} = 0.7$.

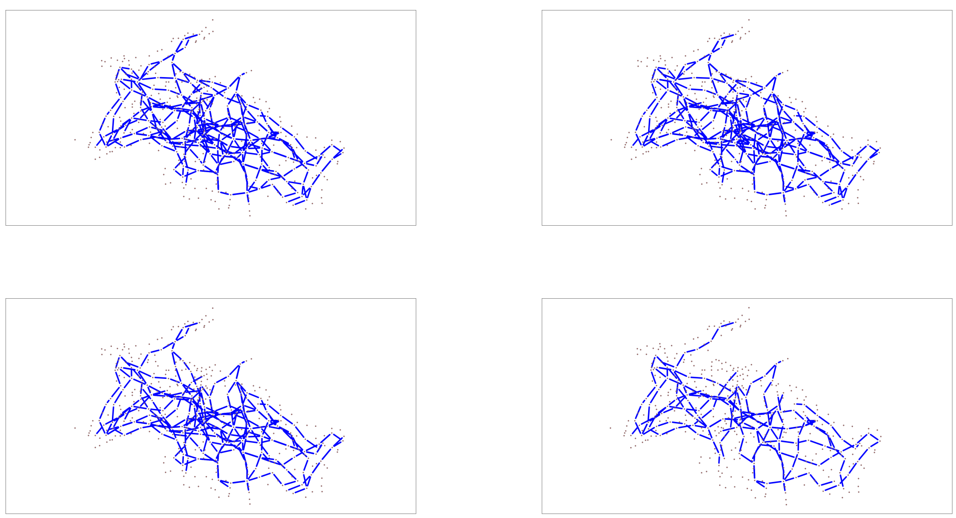

Fig. 9. Procedure of aggregation for p_{aggreg} from 0.4 to 0.7

Thus we applied our methodology to create a network of bus lines in the Lyon area with real data.

5 Conclusion

In this paper, we focus on a network design methodology based on origin/destination clustering. We considered a bus network, and after creating an auxiliary graph containing the origin/destination pairs of the bus network, we applied a clustering algorithm allowing to group similar ODs. We then chose a good representative for each cluster, and created a bus line from each of these representatives, before grouping the lines with a high percentage of common stops.

This methodology results in a network design with interesting properties, especially in pandemic times. On the one hand, the network allows to use only one bus line for any trip, which allows to lower the risk of contamination and also to be able to size the transport supply easily. On the other hand, the final network guarantees a certain equity between the geographical areas of the considered urban area.

In future work (to be completed for the conference), we now want to evaluate the network resulting from our methodology by confronting it with real traffic conditions. A comparison of travel times for some interesting DOs between the current network and our proposed network will be done in order to evaluate the network. The network will be tested with dynamic demand and bus speeds based on the bimodal (bus and car) MFD model. Even if this part is very important to validate the global approach, we think that the methodology itself is an interesting contribution, as well as the case study on the Lyon area.

References

1. Symuvia.
2. Priyabrata Chowdhury, Sanjoy Kumar Paul, Shahriar Kaisar, and Md Abdul Moktadir. Covid-19 pandemic related supply chain studies: A systematic review. *Transportation Research Part E: Logistics and Transportation Review*, page 102271, 2021.
3. Ernesto Cipriani, Stefano Gori, and Marco Petrelli. Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies*, 20(1):3–14, 2012.
4. Collectif Transformation Mobilité Ile de France. Analyse comparée du trafic francilien et international pendant la crise de la covid-19, 2021.
5. Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. Generalized louvain method for community detection in large networks. In *2011 11th international conference on intelligent systems design and applications*, pages 88–93. IEEE, 2011.
6. Laura Di Renzo, Paola Gualtieri, Francesca Pivari, Laura Soldati, Alda Attinà, Giulia Cinelli, Claudia Leggeri, Giovanna Caparello, Luigi Barrea, Francesco Scerbo, et al. Eating habits and lifestyle changes during covid-19 lockdown: an italian survey. *Journal of translational medicine*, 18:1–15, 2020.
7. Dinsic. Réseau urbain tcl sytral.
8. Konstantinos Gkiotsalitis and Antony Stathopoulos. Demand-responsive public transportation re-scheduling for adjusting to the joint leisure activity demand. *International Journal of Transportation Science and Technology*, 5(2):68–82, 2016.
9. Liujiang Kang, Jianjun Wu, Huijun Sun, Xiaoning Zhu, and Bo Wang. A practical model for last train rescheduling with train delay in urban railway transit networks. *Omega*, 50:29–42, 2015.
10. Young-Jae Lee and Vukan R Vuchic. Transit network design with variable demand. *Journal of Transportation Engineering*, 131(1):1–10, 2005.
11. Luyu Liu, Harvey J Miller, and Jonathan Scheff. The impacts of covid-19 pandemic on public transit demand in the united states. *Plos one*, 15(11):e0242476, 2020.
12. Hong K Lo, Kun An, and Wei-hua Lin. Ferry service network design under demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 59:48–70, 2013.
13. Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.
14. Alejandro Tirachini and Oded Cats. Covid-19 and public transportation: Current assessment, prospects, and research needs. *Journal of Public Transportation*, 22(1):1–21, 2020.
15. Satish V Ukkusuri and Gopal Patil. Multi-period transportation network design under demand uncertainty. *Transportation Research Part B: Methodological*, 43(6):625–642, 2009.
16. Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.