

Passenger flow connectivity in collective transportation line networks

Eva Barrena^{1,†}, Alicia De-Los-Santos², Gilbert Laporte¹ and Juan A. Mesa²

¹ *Canada Research Chair in Distribution Management, HEC Montreal, Canada.*

² *Departamento de Matemática Aplicada II, Universidad de Sevilla, Spain.*

Abstract. In this work we analyze the connectivity, in terms of number of transfers, of collective transportation line networks. In contrast to other existing works in the literature, where connectivity is analyzed from a topological point of view, we analyze it from a passengers flow perspective. This enables us to provide a real picture of the connectivity of a transportation system. We accomplish this task by means of hypergraphs, as well as their linearization and measures coming from Complex Network Theory.

Keywords: Collective Transportation Line Networks, Connectivity measures, Hypergraphs

† **Corresponding author:** eva.barrena@cirreht.ca

Received: November 30th, 2013

Published: December 31th, 2013

1. Introduction

In the last two decades, graph theory representation has been widely applied and several coefficients and measures on them have been introduced. Watts and Strogatz [12] introduced two measures which are now commonly used: average path length and clustering coefficient. From these measures, the authors were able to identify complex systems. Later, Latora and Marchiori [8] defined new measures which play a similar role to those defined in [12], but that can be easily applied to transportation networks. Thus, in the last paper, the local and global efficiency were adapted so that they could be applied to transportation networks, instead of the average path length and clustering coefficient, respectively. In particular, the local and global efficiency concepts have been applied to the Boston subway [9]. Indexes to evaluate the robustness of a railway network against interruptions in the normal functioning of its links (both

accidental interruptions and intentional attacks) have been introduced in [6]. However, sometimes the graph structure alone is not sufficient to adequately represent real situations or it is difficult to apply a measure in order to analyze characteristics of determined networks. So, Sen et al. [11] computed several measures on a real railway system carrying out many approximations.

In this work, we are interested in analyzing the transfer system of collective transportation line networks (CTLN). Due to its complexity, the transport planning process has traditionally been decomposed into a succession of stages, namely, line planning, timetabling, resource scheduling, etc. ([10], [5], [4] and [7]). Three layers can be distinguished from the network perspective: the infrastructure network, the line network, and the passenger system. Recently, Barrena et al. [1] introduced several measures on the line network by means of hypergraph theory ([2, 3]). Hypergraphs are the natural extension of graphs and allow us to describe and apply different concepts which cannot be used by graphs. Barrena et al. [1] introduced relationships between structures of graphs and hypergraphs. On each structure, they defined new connectivity measures expressing how easy or hard it is to transfer from one line to another. In this paper we are concerned with the passenger system level, where data on the travel patterns are introduced. From this perspective, the connectivity measures are better to evaluate the difficulty of transferring between lines.

The remainder of the paper is organized as follows. In Section 2.. 1, we formally describe the different representations of a CTLN topology. In Section 2.. 2 we describe the travel patterns in order to use them in Section 3. to adapt the topological connectivity measures to a passenger system level.

2. Previous definitions and demand patterns

We assume the existence of a CTLN. As in [1], we will represent a CTLN G by its set of lines, that is, $G = \{L_1, \dots, L_\ell\}$, where $L_i = \{s_1^i, \dots, s_{k_i}^i\}$ is the i -th line given by its ordered set of stations, so that s_j^i and s_{j+1}^i are directly linked, for all $i = 1, \dots, \ell$, $j = 1, \dots, k_i - 1$. If $s_{k_i}^i$ and s_1^i are also linked, then L_i is a circular line.

2.. 1 Topology of the CTLN. Previous definitions

Depending on the level of information about the network, a CTLN $G = \{L_1, \dots, L_\ell\}$ can be described by several graph structures: a hypergraph, its associated linear graph and a multigraph, as follows.

- Hypergraph

Let $\mathbb{H} = (V(\mathbb{H}), E(\mathbb{H}))$ be the hypergraph associated to G , where the node set $V(\mathbb{H}) = \{s_1, \dots, s_k\}$ contains the stations of G , and the hyperedge set, $E(\mathbb{H}) = \{L_1, \dots, L_\ell\}$ contains the network lines so that each hyperedge L_i consists of a subset of $V(\mathbb{H})$ which are the stations where vehicles from line L_i stop. Note that, as opposed to standard graphs, the elements in $E(\mathbb{H})$ are not necessarily pairs of elements of $V(\mathbb{H})$, but sets of elements. From now on, we call this hypergraph \mathbb{H} the *transit hypergraph*.

On this structure, distance $d_{\mathbb{H}}(s_i, s_j)$ on the elements of $V(\mathbb{H})$ is the length of the shortest ordinary (s_i, s_j) -chain. So, all nodes belonging to the same hyperedge are one unit of distance apart. More precisely, $d_{\mathbb{H}}(s_i, s_j)$ is the minimum number of different lines one needs in order to travel from station s_i to station s_j . For the sake of readability we will identify a station by its subindex whenever this creates no confusion.

- Linear graph

Let $\mathcal{L}(\mathbb{H}) = (V(\mathcal{L}(\mathbb{H})), E(\mathcal{L}(\mathbb{H})))$ be the linear graph associated to hypergraph \mathbb{H} . Its node set $V(\mathcal{L}(\mathbb{H})) = \{L_1, \dots, L_\ell\}$, represents the network lines (hyperedges of \mathbb{H}) and its edge set $E(\mathcal{L}(\mathbb{H}))$ is the set of transfer edges connecting lines with intersections between them. These transfer edges are denoted by e_{pq} . Observe that each hyperedge in \mathbb{H} corresponds to a node in $\mathcal{L}(\mathbb{H})$, and two nodes in $\mathcal{L}(\mathbb{H})$ are linked if and only if the corresponding hyperedges in \mathbb{H} have a non-empty intersection. For the sake of readability we will identify a line by its subindex whenever this creates no confusion.

In this graph, the concept of distance is the usual topological distance in graphs. Concretely, the distance $d_{\mathcal{L}(\mathbb{H})}(L_i, L_j)$ from node L_i to L_j is the number of edges of the shortest path between L_i and L_j . From the point of view of transfers, this distance indicates the number of transfers one needs to make when traveling from one line to other different line in the CTLN.

- Multigraph

Note that the linear graph $\mathcal{L}(\mathbb{H})$ is assumed to be a simple or strict graph, where multiple edges between nodes are not allowed. Recall that an edge e_{ij} in this graph means that the lines L_i and L_j have a non-empty intersection, that is, these lines have at least one common station in \mathbb{H} . However, it does not indicate the number of common stations between the lines, which is interesting in order to measure the how easy it is to transfer between lines. For some purposes, it can become helpful to

consider the linear graph as a multigraph $\mathcal{L}^M(\mathbb{H})$, i.e., a graph in which multiple edges are permitted. In this last case, the number of edges connecting two lines in $\mathcal{L}(\mathbb{H})$ will be equal to the number of transfer stations between them in \mathbb{H} . In the rest of the paper $\mathcal{L}^M(\mathbb{H})$ will be referred to as the *linear multigraph* of G .

As in the linear graph, distance defined on the multigraph is the topological distance.

2.. 2 Passenger system level. Demand patterns

In this work we analyze the functioning of a CTLN with respect to the number of transfers from a passengers flow perspective. To this end, we consider the passenger flow between stations, as well as between the lines of the network by means of origin-destination matrices.

Let $OD \in \mathcal{M}_{k \times k}$ be the origin-destination demand matrix, whose elements $OD(i, j)$, $i, j \in \{1 \dots k\}$ represent the number of passengers traveling from station s_i to station s_j and the diagonal elements are equal zero since there is no demand within a station. Regarding travels between lines, let $LOD \in \mathcal{M}_{\ell \times \ell}$ be the origin-destination lines demand matrix where its elements $LOD(p, q)$, $p, q \in \{1 \dots \ell\}$ represent the number of passengers traveling from line L_p to line L_q , and its diagonal elements $LOD(p, p)$ represent the number of passengers traveling within line L_p . Note that matrices OD and LOD are not necessarily symmetric.

The flow structure is as follows. In terms of hypergraphs, the flow f_{ij} between stations s_i and s_j is the sum of $OD(i, j)$ and $OD(j, i)$, whereas, in its associated linear graph, the flow f_{pq}^L , between lines L_p and L_q , $p \neq q$, is the sum of $LOD(p, q)$ and $LOD(q, p)$. In the hypergraph \mathbb{H} , the total flow N , is computed as the sum of all flows f_{ij} , $i, j \in \{1 \dots k\}$ and in the linear graph the total flow N^L is obtained as the sum of all flows f_{pq}^L , $p \neq q$, since the flow inside lines is not considered. Note that the total flow N can also be expressed by terms of linear graphs, i. e., $N = N^L + \sum_p f_{pp}^L$.

3. Passenger system level. Flow-based connectivity measures

An interesting measure to evaluate the connectivity of a collective transportation line network is the *characteristic path length*. In the next section we show how to adapt the topological measure *characteristic path length* to a passenger system level.

3.. 1 Characteristic path flow-weighted length

In a previous work [1], the characteristic path length is defined on the transit hypergraph, its associated linear graph and the multi-linear graph. From a topological point of view, this measure indicates how easy it is for passengers to transfer in the CTLN. However, if the mobility patterns are introduced, this measure has several limitations and should be extended.

The following example depicts two different CTLN having the same characteristic path length in the topological level.

Example 3..1 Consider a simple three-lines CTLN and its hypergraph representation, which is formed by three hyperedges, L_1, L_2 and L_3 , all having the same number of stations. Figure 1 illustrates two hypergraphs representations \mathbb{H}_1 and \mathbb{H}_2 of CTLN with these characteristics.

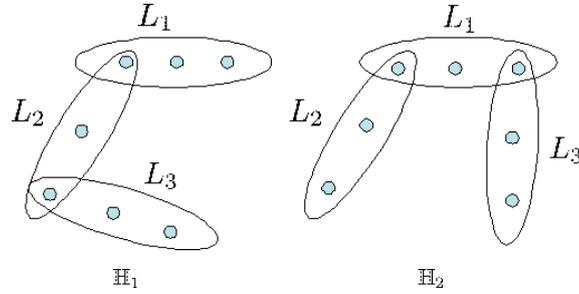


Figure 1: A sample illustration for two hypergraph \mathbb{H}_1 and \mathbb{H}_2 representation of a CTLN.

The corresponding linear graphs $\mathcal{L}(\mathbb{H}_1)$ and $\mathcal{L}(\mathbb{H}_2)$ associated to the networks above defined are represented in Figure 2.

Then, the characteristic path length [1] on hypergraphs \mathbb{H}_1 and \mathbb{H}_2 are both equal to 1,76, and on the linear graphs [1] $\mathcal{L}(\mathbb{H}_1)$ and $\mathcal{L}(\mathbb{H}_2)$ are both equal to 1,67.

However, the networks represented in the example may not be equal if demand patterns are included. In order to distinguish between the cases presented in this example, we now describe the characteristic path flow-weighted length on the topological structures described in Section 2.. 1.

3.. 1.1 Characteristic path flow-weighted length on the transit hypergraph

Over the hypergraph level of information, the characteristic path length will provide an average measure of how the topological configuration of the CTLN affects to the passengers who transfer between the stations of a CTLN.

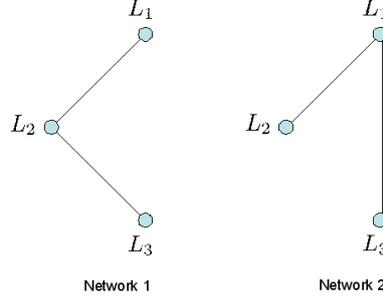


Figure 2: A sample illustration for two linear graphs $\mathcal{L}(\mathbb{H}_1)$ and $\mathcal{L}(\mathbb{H}_2)$ representation of a CTLN.

Definition 3..1 We define the characteristic path flow-weighted length of the transit hypergraph graph \mathbb{H} , with $|V(\mathbb{H})| > 1$, as the average flow weighted distance in \mathbb{H} , i.e.,

$$\mathfrak{L}_f(\mathbb{H}) = \frac{2}{|V(\mathbb{H})|(|V(\mathbb{H})| - 1)} \sum_{s_i < s_j} d_{\mathbb{H}}(s_i, s_j) \frac{f_{ij}}{N},$$

So, the higher the number of passengers who transfer between stations s_i and s_j , the higher the weight we assign to the distance between these two stations. In this way, when the characteristic path length is minimized, the lines with more passengers transferring between them will tend to have better connections.

In order to illustrate this measure, we introduce travel patterns in the Example 3.1 and obtain Example 3.3.

Example 3..2 Let us consider the CTLN represented in Example 3..1, and the following flow pattern f_{ij} , $i, j \in \{1, \dots, k\}$, over their stations $s_i, s_j \in V(\mathbb{H})$:

$$f_{ij} = \begin{cases} 1 & \text{if } s_i \in L_1, s_j \in L_2 \\ 2 & \text{if } s_i \in L_1, s_j \in L_3 \\ 3 & \text{if } s_i \in L_2, s_j \in L_3 \end{cases}$$

If s_i and s_j belong to the same hyperedge or some of them is a intersection node, $f_{ij} = 1$. The characteristic path flow-weighted length for \mathbb{H}_1 and \mathbb{H}_2 are different when demand patterns are included, and equal to $\mathfrak{L}_f(\mathbb{H}_1) = 0,09$ and $\mathfrak{L}_f(\mathbb{H}_2) = 0,11$, respectively. The characteristic path flow-weighted length is larger for hypergraph \mathbb{H}_2 than for \mathbb{H}_1 because the maximum flow is given between stations L_2 and L_3 and hypergraph \mathbb{H}_1 has a direct connection between these two lines whereas \mathbb{H}_2 does not.

The next lemma proves that the characteristic path flow-weighted length above defined, is a natural extension of $\mathfrak{L}_f(\mathbb{H})$ defined in the paper [1].

Lemma 3.1 $\mathfrak{L}_f(\mathbb{H})$ is an extension of $\mathfrak{L}(\mathbb{H})$, which yields proportional results if the number of passengers between each pair of stations $s_i, s_j, i \neq j$, is the same, that is, all the elements of matrix OD , except its diagonal elements, are the same.

Proof. Trivially, if we consider a constant number α of passengers for each OD pair (i.e. $f_{ij} = 2\alpha$), the following expression is obtained:

$$\mathfrak{L}_f(\mathbb{H}) = \frac{2}{|V(\mathbb{H})|(|V(\mathbb{H})| - 1)} \sum_{i < j} d_{\mathbb{H}}(s_i, s_j) \frac{2\alpha}{N} = \frac{2\alpha}{N} \mathfrak{L}(\mathbb{H}).$$

The following proposition states that the characteristic path length of the transit hypergraph, which gives a measure of the connectivity of a CTLN, satisfies the same two basic properties of this type of measures: it lies within a predefined range and satisfies a monotonicity property.

Proposition 3.1 Consider a CTLN G , and let \mathbb{H} be its associated transit hypergraph. Let OD be a non-empty demand matrix between stations. The characteristic path flow-weighted length on \mathbb{H} satisfies the following two properties:

1. $\frac{1}{N} \leq \mathfrak{L}(\mathbb{H}) \leq \frac{1}{3N}(\ell + 2)$.
2. Let G' be a CTLN obtained when adding one new link joining two lines of G (with at least one passenger), and let \mathbb{H}' be the associated hypergraph. Then $\mathfrak{L}_f(\mathbb{H}) \geq \mathfrak{L}_f(\mathbb{H}')$.

3.. 1.2 Characteristic path flow-weighted length on the linear graph

The following definition is the natural extension of the characteristic path length defined in [1]. The characteristic path length in $\mathcal{L}(\mathbb{H})$ gives information about how the number of transfers (the topology of the network) affects to the passengers.

Definition 3.2 We define the characteristic path flow-weighted length of the linear graph $\mathcal{L}(\mathbb{H})$ with $|V(\mathcal{L}(\mathbb{H}))| > 1$ as the average flow weighted distance in $\mathcal{L}(\mathbb{H})$, i.e.,

$$\mathfrak{L}_f(\mathcal{L}(\mathbb{H})) = \frac{2}{|V(\mathcal{L}(\mathbb{H}))|(|V(\mathcal{L}(\mathbb{H}))| - 1)} \sum_{i < j} d_{\mathcal{L}(\mathbb{H})}(L_i, L_j) \frac{f_{ij}^L}{NL},$$

where f_{ij}^L/N^L is then the proportion of passengers transferring from line L_i to line L_j over all passengers who transfer.

So, the higher the number of passengers who transfer between lines L_i and L_j , the higher the weight we assign to the distance between these two lines. In this way, when the flow characteristic path length is minimized, lines with more passengers transferring between them will have better connections. Another possibility is to consider in the denominator, instead of the number N^L of passengers who transfer, the total number of passenger N in the network. Let us introduce travel patterns between lines in the Example 3.1:

Example 3.3 We consider a simple case in which the flow between lines are: $f_{12}^L = 1$, $f_{13}^L = 2$ and $f_{23}^L = 3$. The line flow characteristic path length is at each case, $\mathfrak{L}_f(\mathcal{L}(\mathbb{H}_1)) = 0,44$ and $\mathfrak{L}_f(\mathcal{L}(\mathbb{H}_2)) = 0,5$. As expected, these measure is larger for the \mathbb{H}_2 than for \mathbb{H}_1 since the higher flow is given between lines without direct connection in \mathbb{H}_2 and directly connected in \mathbb{H}_1 .

The next lemma proves that the characteristic path flow-weighted length above defined, is a natural extension of $\mathfrak{L}_f(\mathcal{L}(\mathbb{H}))$ defined in the paper [1].

Lemma 3.2 $\mathfrak{L}_f(\mathbb{H})$ is an extension of $\mathfrak{L}(\mathbb{H})$, which obtain proportional results if the number of passengers between each pair of lines $L_i, L_j, i \neq j$, is the same, that is, all the elements of matrix LOD , except its diagonal elements, are the same.

Proof. Trivially, if we consider a constant number α of passengers for each OD pair (i.e. $f_{ij}^L = 2\alpha$), the following expression is obtained:

$$\mathfrak{L}_f(\mathbb{H}) = \frac{2}{|V(\mathbb{H})|(|V(\mathbb{H})| - 1)} \sum_{i < j} d_{\mathbb{H}}(s_i, s_j) \frac{2\alpha}{N^L} = \frac{2\alpha}{N^L} \mathfrak{L}(\mathbb{H}).$$

The next proposition proves two properties which, together with the trivial invariance to scale changes noted in the previous remark, allow us to use the characteristic path length as a connectivity measure for collective transportation line networks. These two properties are: staying within a predefined range of variation and monotonicity.

Proposition 3.2 Consider a CTLN G , and let $\mathcal{L}(\mathbb{H})$ be its associated linear graph. Let LOD be a non-empty demand matrix between stations. The characteristic path flow-weighted length $\mathfrak{L}_f(\mathcal{L}(\mathbb{H}))$ of the linear graph satisfies the following properties:

1. $\frac{1}{N^T} \leq \mathfrak{L}_f(\mathcal{L}(\mathbb{H})) \leq \frac{\ell-1}{T}$.

2. $\mathfrak{L}_f(\mathcal{L}(\mathbb{H}))$ is monotone decreasing in the sense that, if G' is obtained when adding a new link to G connecting two lines (with at least one passenger), we have that $\mathfrak{L}_f(\mathcal{L}(\mathbb{H})') \leq \mathfrak{L}_f(\mathcal{L}(\mathbb{H}))$, where $\mathcal{L}_f(\mathbb{H})'$ is the linear graph of G' .

Similar definitions and properties hold for the linear multigraph $\mathcal{L}^M(\mathbb{H})$.

Acknowledgements

This research work was partially supported Ministerio de Economía y Competitividad (Spain)/FEDER under grant MTM2012-37048, by Junta de Andalucía (Spain)/FEDER under excellence projects P09-TEP-5022 and P10-FQM-5849 and by the Canadian Natural Sciences and Engineering Research Council under grant 39682-10.

References

- [1] E. Barrena, A. De-Los-Santos, J. A. Mesa, and F. Perea. Analyzing connectivity in collective transportation line networks by means of hypergraphs. *European Physical Journal Special Topics*, 215(1):93–108, 2013.
- [2] C. Berge. *Graphes et Hypergraphes*. Elsevier Science, Paris, 1972.
- [3] C. Berge. *Hypergraphs: combinatorics of finite sets*. North-Holland mathematical library. North-Holland, Amsterdam, 1989.
- [4] A. Ceder and N. H. M. Wilson. Bus network design. *Transportation Research Part B: Methodological*, 20(4):331–344, 1986.
- [5] J.F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32(4):380–404, 1998.
- [6] A. De-Los-Santos, G. Laporte, J.A. Mesa, and F. Perea. Evaluating passenger robustness in a rail transit network. *Transportation Research Part C: Emerging Technologies*, 20:34–46, 2012.
- [7] V. Guihaire and J.K. Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273, 2008.
- [8] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87(19), 2001.

- [9] V. Latora and M. Marchiori. Is the Boston subway a small-world network? *Physica A*, 314(1):109–113, 2002.
- [10] C. Liebchen and R. Möhring. The modeling power of the periodic event scheduling problem: railway timetables and beyond. In F. Geraets, L. Kroon, A. Schoebel, D. Wagner, and C. D. Zaroliagis, editors, *Algorithmic Methods for Railway Optimization*, volume 4359 of *Lecture Notes in Computer Science*, pages 3–40. Springer, Berlin, Heidelberg, 2007.
- [11] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. Small-world properties of the indian railway network. *Physical Review E*, 67, 036106(3):1–5, 2003.
- [12] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.