

Document downloaded from:

<http://hdl.handle.net/10251/58847>

This paper must be cited as:

Barrena, E.; De-Los-Santos, A.; Mesa López-Colmenar, JA.; Perea Rojas Marcos, F. (2013). Analyzing connectivity in collective transportation line networks by means of hypergraphs. *European Physical Journal - Special Topics*. 215(1):93-108. doi:10.1140/epjst/e2013-01717-3.



The final publication is available at

<http://dx.doi.org/10.1140/epjst/e2013-01717-3>

Copyright EDP Sciences: EPJ; Springer

Additional Information

# Analyzing Connectivity in Collective Transportation Line Networks by means of Hypergraphs

Eva Barrena<sup>1,a</sup>, Alicia De-Los-Santos<sup>1,b</sup>, Juan Antonio Mesa<sup>1,c</sup>, and Federico Perea<sup>2,d</sup>

<sup>1</sup> Departamento de Matemática Aplicada II. Universidad de Sevilla (Spain).

<sup>2</sup> Departamento de Estadística e Investigación Operativa Aplicadas y Calidad. Universitat Politècnica de València (Spain)

**Abstract.** In this paper we will show how hypergraphs and some measures related to them can help in extracting information about Collective Transportation Line Networks. We will also prove that these measures satisfy certain properties that validate their use to compare the connectivity of different networks.

## 1 Introduction

Collective Transportation Line Networks (CTLN) have been classified as complex systems. In order to understand the organization and functioning of these systems, in the last two decades graph theory representation has been applied and several coefficients and measures have been introduced. Some of the features studied to characterize networks representing complex systems are the average path length and the clustering coefficient (see [1]). These two measures were modified by Latora and Marchiori [2] in order to fit them 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, these concepts have been applied to the Boston subway [3]. The related concept of vulnerability has been applied (see [4]) to other subway networks along the world. More recently, 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 [5].

CTLN planning contains several intertwined problems: network design, line planning, timetabling, resource scheduling, etc., with several agents (public organizations, private companies, users) involved. Three layers can be distinguished from the network perspective: the infrastructure network, the line network, and the passenger system. In this work, we are concerned with the line network level, where one is usually interested in analyzing the connectivity of its lines, i.e., how easy/hard it is to

---

<sup>a</sup> e-mail: [ebarrena@us.es](mailto:ebarrena@us.es)

<sup>b</sup> e-mail: [aliciasantos@us.es](mailto:aliciasantos@us.es)

<sup>c</sup> e-mail: [jmesa@us.es](mailto:jmesa@us.es)

<sup>d</sup> e-mail: [perea@eio.upv.es](mailto:perea@eio.upv.es)

transfer from one line to another. CTLN riders are reluctant to transferring from one line to another because of inconvenience and extra travel time. It can be noted that, in urban CTLN, the proportion of users who transfer twice is very low and nearly nobody transfers three times, in these cases they prefer to use a different mode of transportation.

Since graphs are not enough to study the transfer system, in order to analyze the connectivity of CTLN, in Section 2 we represent them as hypergraphs and their corresponding linear graphs. Some related measures will be presented and analyzed in Section 3. For the sake of readability, all technical proofs have been omitted here and are presented in the technical report [6] instead. Four metro systems have been chosen for comparing the values of these measures in Section 4.

## 2 Representation of line networks by means of hypergraphs and their associated linear graphs

In this paper we are interested in analyzing the functioning of a collective transportation line network (CTLN) with respect to the number of transfers. For this purpose, several graph representations are used. We will represent a collective transportation line network  $G$  by its set of lines, that is,  $G = \{L_1, \dots, L_\ell\}$ , where  $L_i = \{s_1^i, \dots, s_{k_i}^i\}$  are the  $i$ -th line and its 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,  $L_i$  is a circular line. In the upper graph of Figure 1 we show an example of a collective transportation line network.

Let us define three different structures associated to the CTLN: the transit hypergraph and its associated linear graph and multigraph. A collective transportation line network can be represented by a hypergraph (see [7, 8]) as follows.

**Definition 1** Let  $G = \{L_1, \dots, L_\ell\}$  be a CTLN. We define  $\mathbb{H} = (V(\mathbb{H}), E(\mathbb{H}))$  as its associated hypergraph, where the node set  $V(\mathbb{H}) = \{s_1, \dots, s_k\}$  is formed by the stations of  $G$  and the hyperedge set  $E(\mathbb{H}) = \{L_1, \dots, L_\ell\}$  consists of the network lines so that each hyperedge  $L_i$  consists of a subset of  $V(\mathbb{H})$ : the stations/stops where  $L_i$  stops. From now on, we call this hypergraph a transit hypergraph.

Note that elements in  $E(\mathbb{H})$  do not need to be pairs of elements of  $V(\mathbb{H})$ , as opposed to standard graphs.

Figure 1 bottom left graph shows the transit hypergraph of the Vienna metro network where, for the sake of readability, only transfer stations ( $T_i$ ) have been depicted. On every hyperedge, the number of stations constituting the corresponding metro line is indicated. It is interesting to note how the graphical representation of a relatively complex network can be highly simplified by means of its corresponding hypergraph. In the following,  $H$  refers to a general hypergraph and  $\mathbb{H}$  to the transit hypergraph.

The following definition introduces the second network structure we are to study: the concept of *linear graph* (see [7, 8]) associated with the transit hypergraph.

**Definition 2** Let  $G = \{L_1, \dots, L_\ell\}$  be a CTLN, and let  $\mathbb{H}$  be its associated hypergraph. The linear graph of  $\mathbb{H}$  is denoted by  $\mathcal{L}(\mathbb{H}) = (V(\mathcal{L}(\mathbb{H})), E(\mathcal{L}(\mathbb{H})))$ , where:

- $V(\mathcal{L}(\mathbb{H})) = \{L_1, \dots, L_\ell\}$  representing the network lines (hyperedges of  $\mathbb{H}$ ).
- $E(\mathcal{L}(\mathbb{H}))$  is the set of transfer edges connecting lines with intersections between them. These transfer edges are denoted by  $e_{ij}$ .

Observe that each hyperedge in  $\mathbb{H}$  corresponds with 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 non-empty intersection.

Note that the linear graph is assumed to be a simple or strict graph, where multiple edges between nodes are not permitted. So, if lines  $L_i$  and  $L_j$  have common stations then there exists a unique edge  $e_{ij}$ , independently of the number of common stations. For some purposes, it becomes interesting 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 multiple 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  $G$ .

Figure 1 bottom center and bottom right graphs show the linear graph and linear multigraph of the Vienna metro network. Note that each node represents a line and there is an edge (or a multiple edge) between two lines if and only if they share one station (or more than one transfer stations).

### 3 Connectivity measures at each level of abstraction

In this section we introduce several measures that will evaluate the connectivity of a collective transportation line network in terms of its transfers: the *characteristic path length*, the *clustering coefficient*, the *local efficiency* and the *global efficiency*.

*Remark 1* All these measures are invariant to scale changes in the measurements of physical (spatial) distances when they are defined topologically. In the rest of this section this property will be complemented with others which validate the use of these measures in order to compare the connectivity of different networks.

#### 3.1 Characteristic path length

A measure of the separation between pairs of nodes in a graph is given by the average length of the shortest paths, also known as characteristic path length, defined as the mean of the length of the shortest path over all pairs of nodes, (see [1, 9]). We will extend this definition so that this index can be applied to the transit hypergraph and its associated linear graph, and we will see that such new definitions give valuable information about the connectivity of a CTLN.

##### 3.1.1 Characteristic path length on graph $\mathcal{L}(\mathbb{H})$ and multigraph $\mathcal{L}^M(\mathbb{H})$

We begin with the linear graph. Over this level of abstraction, the characteristic path length will give an average measure of how easy it is for passengers to transfer between the lines of a CTLN. Before defining this first measure, we need to introduce the distance on this linear graph that will be used.

**Definition 3**  $d_{\mathcal{L}(\mathbb{H})}(L_i, L_j)$  is the topological distance between the nodes  $L_i, L_j$  of the graph  $\mathcal{L}(\mathbb{H})$ , i.e., the number of edges that a shortest path between  $L_i$  and  $L_j$  contains. In other words,  $d_{\mathcal{L}(\mathbb{H})}(L_i, L_j)$  is the number of transfers one needs for traveling from line  $L_i$  to line  $L_j$ .

We trivially have that  $d_{\mathcal{L}(\mathbb{H})}(L_i, L_j) = 1$  if and only if  $L_i \cap L_j \neq \emptyset$ , that is, if there is a transfer station from line  $L_i$  to line  $L_j$ , and  $d_{\mathcal{L}(\mathbb{H})}(L_i, L_j) > 1$  otherwise. Besides,  $d_{\mathcal{L}(\mathbb{H})} \leq \ell - 1$  because we are assuming a connected CTLN.

**Definition 4** We define the characteristic path length of the linear graph  $\mathcal{L}(\mathbb{H})$  with  $|V(\mathcal{L}(\mathbb{H}))| > 1$  as the average distance in  $\mathcal{L}(\mathbb{H})$ , i.e.,

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

Therefore the characteristic path length in  $\mathcal{L}(\mathbb{H})$  represents the average number of transfers one needs for traveling between lines in the CTLN. Note as well that this measure makes no sense when the network consists of one line only.

The next proposition proves two properties that, 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. Such two properties are: staying within a predefined range of variation and monotonicity.

**Proposition 1** Consider a CTLN  $G$ , and let  $\mathcal{L}(\mathbb{H})$  be its associated linear graph. Let  $\mathfrak{L}(\mathcal{L}(\mathbb{H}))$  be the characteristic path length of the linear graph. We have that:

1.  $1 \leq \mathfrak{L}(\mathcal{L}(\mathbb{H})) \leq \ell - 1$ . The nearer to 1, the more interconnected the lines are.  $\mathfrak{L}(\mathcal{L}(\mathbb{H})) = 1$  means that for all pairs of lines of  $C$  there is a transfer station that directly connects them.  $\mathfrak{L}(\mathcal{L}(\mathbb{H})) = \ell - 1$  can only be achieved when the  $G$  consists of two lines that are connected (note that in this case  $\ell - 1 = 1$ ).
2.  $\mathfrak{L}(\mathcal{L}(\mathbb{H}))$  is monotone decreasing in the sense that, if  $G'$  is obtained when adding a new link to  $G$  connecting two lines, we have that  $\mathfrak{L}(\mathcal{L}(\mathbb{H})') \leq \mathfrak{L}(\mathcal{L}(\mathbb{H}))$ , where  $\mathcal{L}(\mathbb{H})'$  is the linear graph of  $G'$ . Moreover,  $\mathfrak{L}(\mathcal{L}(\mathbb{H})') < \mathfrak{L}(\mathcal{L}(\mathbb{H}))$  if and only if the new link connects two lines that were not directly connected in  $G$ .

**Proof.** See [6]. □

The reader may note that the same properties proven in the previous section hold for the characteristic path length on the linear multigraph  $\mathcal{L}^M(\mathbb{H})$ .

### 3.1.2 Characteristic path length on transit hypergraph $\mathbb{H}$

We now continue with the transit hypergraph. Over this level of abstraction, the characteristic path length will give an average measure of how easy it is for passengers to transfer between the stations of a CTLN. Before defining this first measure, we need to introduce the distance on this transit hypergraph that will be used.

**Definition 5**  $d_{\mathbb{H}}(s_i, s_j)$  is the length of the shortest ordinary  $(s_i, s_j)$ -chain. All the nodes incident to the same hyperedge (stations belonging to the same line) lie distance 1 apart. So,  $d_{\mathbb{H}}(s_i, s_j)$  is the minimum number of different lines one needs in order to travel from station  $s_i$  to station  $s_j$ . Another interpretation for  $d_{\mathbb{H}}(s_i, s_j)$  is the number of transfers one needs in the trip from  $s_i$  to  $s_j$  plus one.

Note that in our collective transportation line network setting  $d_{\mathbb{H}}(s_i, s_j)$  is well defined since we are considering connected networks only. From this distance, the characteristic path length on the transit hypergraph, denoted by  $\mathfrak{L}(\mathbb{H})$ , can be analogously defined as the characteristic path length on the linear graph, see Definition 4.

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 2** Consider a CTLN  $G$ , and let  $\mathbb{H}$  be its associated transit hypergraph. We have that the characteristic path length on  $\mathbb{H}$  satisfies the following two properties:

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

**Proof.** See [6]. □

### 3.2 Clustering coefficient

Transitivity is an important concept in social network analysis. In a social system, the probability of two friends having common friends is very high. The usual way of measuring the transitivity of a network  $G$  is by means of the fraction of transitive triples, i.e. the fraction of connected triples of nodes which also form triangles of interactions (see [10–12]). This is a classic measure used in social sciences to indicate how much a network is clustered locally. Note that the number of possible triples of nodes in  $G$  is  $\binom{n}{3}$ . Another indicator to measure the local degree of clustering known as *clustering coefficient*  $C$  is defined in [1] and it is adapted to  $\mathcal{L}(\mathbb{H})$ ,  $\mathcal{L}^M(\mathbb{H})$  and  $\mathbb{H}$  in the following sections.

#### 3.2.1 Clustering coefficient on linear graphs $\mathcal{L}(\mathbb{H})$ and $\mathcal{L}^M(\mathbb{H})$

The clustering coefficient in  $\mathcal{L}(\mathbb{H})$  will measure the degree/existence of connectivity between its nodes. Since nodes in  $\mathcal{L}(\mathbb{H})$  are lines in the CTLN and edges represent the connections between pairs of lines, the clustering coefficient  $C$  will be an indicator of the robustness/interconnectivity of the line configuration. More precisely, the clustering coefficient at this linear graph will take into account the number of transfers one needs to do for traveling from one line to another. Thus, at this level of abstraction, the number of stations of each line is not considered, and so, neither are the travels between pairs of stations of the same line. These aspects will be taken into consideration at the transit hypergraph later on.

**Definition 6** Let  $G$  be a CTLN and let  $\mathcal{L}(\mathbb{H})$  be its associated linear graph. We consider the clustering coefficient  $C$  on the linear graph  $\mathcal{L}(\mathbb{H})$  as in [1]. So, for each node  $i \in V(\mathcal{L}(\mathbb{H}))$ , the subgraph  $\mathcal{L}_i(\mathbb{H})$  formed by all first neighbors of  $i$  is considered. In such subgraph, node  $i$  and all edges incident to  $i$  are eliminated. If node  $i$  has  $k_i$  neighbors, then  $\mathcal{L}_i(\mathbb{H})$  will have  $k_i$  nodes and at most  $k_i(k_i - 1)/2$  edges.  $C_i(\mathcal{L}_i(\mathbb{H}))$  is the fraction of these edges that actually exist and  $C(\mathcal{L}(\mathbb{H}))$  is the average of  $C_i$ , calculated over all nodes:

$$C(\mathcal{L}(\mathbb{H})) = \frac{1}{|V(\mathcal{L}(\mathbb{H}))|} \sum_{i \in V(\mathcal{L}(\mathbb{H}))} C_i(\mathcal{L}_i(\mathbb{H})),$$

where

$$C_i(\mathcal{L}_i(\mathbb{H})) = \frac{\text{number of edges in } \mathcal{L}_i(\mathbb{H})}{k_i(k_i - 1)/2}.$$

We consider that if  $|\mathcal{L}_i(\mathbb{H})| = 1$ , then  $C_i(\mathcal{L}_i(\mathbb{H})) = 0$ .

As stated before, the clustering coefficient in  $\mathcal{L}(\mathbb{H})$  measures the degree/existence of connectivity between its nodes, but it does not take into account the number of transfer nodes in  $\mathbb{H}$ . In order to consider, not only if one can transfer between two lines, but also the number of possibilities to transfer (i.e., number of transfer nodes in  $\mathbb{H}$ ), one should consider the following definition of the clustering coefficient  $C$  in  $\mathcal{L}^M(\mathbb{H})$ .

**Definition 7** Let  $\mathcal{L}_i^M(\mathbb{H})$  be the neighbor multigraph associated to  $i \in V(\mathcal{L}^M(\mathbb{H}))$  and let  $U^{max}$  be a threshold that represents the maximum number of transfer nodes that can exist between two lines. If node  $i$  has  $k_i$  neighbors, then  $\mathcal{L}_i^M(\mathbb{H})$  will have at most  $U^{max} \frac{k_i(k_i-1)}{2}$  multiedges.  $C_i(\mathcal{L}_i^M(\mathbb{H}))$  is the fraction of these edges that actually exist and the clustering coefficient  $C(\mathcal{L}^M(\mathbb{H}))$  on the linear multigraph  $\mathcal{L}^M(\mathbb{H})$  is the average of  $C_i(\mathcal{L}_i^M(\mathbb{H}))$ , calculated over all nodes:

$$C(\mathcal{L}^M(\mathbb{H})) = \frac{1}{|V(\mathcal{L}^M(\mathbb{H}))|} \sum_{i \in V(\mathcal{L}^M(\mathbb{H}))} C_i(\mathcal{L}_i^M(\mathbb{H})),$$

where

$$C_i(\mathcal{L}_i^M(\mathbb{H})) = \frac{\text{number of edges in } \mathcal{L}_i^M(\mathbb{H})}{U^{max} k_i(k_i - 1)/2}.$$

### 3.2.2 Clustering coefficient on transit hypergraph $\mathbb{H}$

The global clustering coefficient on hypergraphs is defined in [13]. In this work, we analyze the clustering coefficient from a local point of view, which, to the best of our knowledge, has not been considered for hypergraphs in the literature. The natural extension of the local clustering coefficient for graphs [1] to the case of hypergraphs would be to consider hyperedges instead of edges. However, with this extension, the higher the number of hyperedges, the higher the value of  $C$ , which in the case of CTLN would mean that, for two given networks with the same number of nodes, the one with more lines would yield a better clustering coefficient than the network with less lines. This is not the aspect we would like to measure in this section, but the level of connectivity between stations, i.e., a measure of the number of transfers one needs to do for traveling between two stations. So, stations in the same line would be very well connected, followed by the stations separated by one transfer, then by those separated by two transfers and so on.

For that reason, in order to define the local clustering coefficient on hypergraphs, we refer to the primal graph of a hypergraph and will make calculations on it using the terminology of hypergraphs. Some previous definitions related to hypergraphs are necessary.

**Definition 8** For a hypergraph  $H$  and a set  $X \subseteq V(H)$ , the subhypergraph induced by  $X$  is the hypergraph  $H[X] = (X, \{e \cap X : e \in E(H)\})$ .

The primal graph, also called the Gaifman graph (see [14]) of a hypergraph is the graph with the same nodes as the hypergraph, and edges between all pairs of nodes contained in the same hyperedge.

In [15], a graph representing direct connections in the Indian Railway Network (IRN) is defined, and the usual local clustering coefficient is calculated over it, requiring for that a high number of calculations. It is easy to see that this graph structure coincides with the primal graph. We will prove that using hypergraph structures simplifies calculations. We now introduce the notion of clustering coefficient in the primal graph following the usual local clustering definition [1].

### 3.2.3 Primal graph. Clustering coefficient on $G_{\mathbb{H}}$

**Definition 9** Let  $G_{\mathbb{H}}$  be the primal graph of the transit hypergraph  $\mathbb{H}$ . We will consider the clustering coefficient of  $G_{\mathbb{H}}$  defined as in [1] and denoted by  $C(G_{\mathbb{H}})$ ,  $G_{\mathbb{H}_i}$

being the subgraph formed by all first neighbors of  $i$ . Mathematically,

$$C(G_{\mathbb{H}}) = \frac{1}{|V(G_{\mathbb{H}})|} \sum_{i \in V(G_{\mathbb{H}})} C(G_{\mathbb{H}_i}),$$

where

$$C(G_{\mathbb{H}_i}) = \frac{\text{number of edges in } G_{\mathbb{H}_i}}{k_i(k_i - 1)/2},$$

$k_i$  being the number of nodes of  $G_{\mathbb{H}_i}$ .

Two drawbacks of this coefficient in CTLNs are:

- $C(G_{\mathbb{H}_i})$  equals one when  $i$  is a nontransfer station (belongs to just one hyperedge),
- $C(G_{\mathbb{H}_i}) \in [0, 1]$  for transfer stations (belongs to more than one hyperedge).

In other words, transfer nodes are somehow penalized by this measure.

### 3.2.4 Computing the clustering coefficient in the primal graph using the terminology of hypergraphs

Computing the clustering coefficient in the primal graph requires a high computational effort. First of all, the primal graph must be calculated by adding edges between each pair of nodes belonging to the same hyperedge. Secondly, this high increment of the number of edges at the primal graph makes its application to complex networks too hard. For this reason, in [15] the clustering coefficient is calculated for only some stations of the network instead of the complete IRN. One of the disadvantages of the primal graph is that the information about hyperedges is lost (in CTLN this is the line structure). This information about the line structure is relevant and it even enables the simplification of calculations for computing the clustering coefficient. Therefore, we will propose a different methodology to calculate the clustering coefficient based on hypergraph properties.

*Remark 2* Note that at the hypergraph level, every pair of nodes in a same hyperedge are related and this relation is represented by an edge in the primal graph. Then, at the primal graph, the number of edges between nodes that belong to the same hyperedge in the original transit hypergraph is equal to the number of pairwise combinations between them.

To calculate this number of pairwise combinations the number of pairwise combinations between elements of a hyperedge is defined.

**Definition 10** Let  $H$  be a hypergraph. For each node  $s_i \in H$ , we consider the associated neighbor subhypergraph  $H_i$ . Let  $L_j^i \in E(H_i)$  be a hyperedge of  $H_i$ , we define  $Comb(L_j^i)$  as the number of possible pairwise combinations between elements of  $L_j^i$ , i.e.

$$Comb(L_j^i) = \frac{|L_j^i| \cdot (|L_j^i| - 1)}{2}.$$

The clustering coefficient of the primal graph of the hypergraph can be more easily calculated by means of the hypergraph terminology, as will be shown in Theorem 1, where some simplifications and combinatorics help to group the same type of pairs. This simplifies the counting process and reduces the computational complexity.

Since  $V(G_H) = V(H)$ , we refer to the original hypergraph when dealing with nodes and their membership to a hyperedge. So, intersection nodes are the ones belonging to more than one hyperedge.



**Theorem 1** Let  $s_i \in V(H) = V(G_H)$  be a node of  $H$  and let  $H_i$  be the neighbor subhypergraph associated to  $s_i$ . Denote by  $L_j^i$  a hyperedge of  $H_i$ , for  $j \in I = \{1, \dots, |E(H_i)|\}$ . Then the clustering coefficient of the primal graph can be calculated using the hypergraph terminology, distinguishing between intersection and non-intersection nodes, as follows:

$$C(G_H) = \frac{1}{|V(G_H)|} \sum_{i \in V(G_H)} C(G_{H_i}),$$

where

- (a) If  $s_i \in V(H)$  is a non-intersection node, then  $C(G_{H_i}) = 1$ .  
(b) If  $s_i \in V(H)$  is an intersection node, then the following is applied:

$$C(G_{H_i}) = \frac{\sum_{k=1}^{|I|} (-1)^{k-1} \text{Comb}(\bigcap_{L_j^i \in E(H_i), j \in J \subseteq I, |J|=k} L_j^i)}{\text{Comb}(\bigcup_{j \in I} L_j^i)}.$$

**Proof.** See [6]. □

### 3.3 Local and global efficiency

The clustering coefficient suffers from several limitations, as it is shown in [3, 16]: it is ill-defined in several cases, and it only works in the topological network, where the only information retained is about the existence or the absence of a link, and nothing is known about the link's length. Some new measures are proposed in [2, 3, 16–18], by introducing the *efficiency*  $E$ , which is valid both for weighted and unweighted networks. They argued that the global efficiency plays the role of the inverse of the characteristic path length and the local efficiency is similar to the clustering coefficient. The description of the network in terms of its efficiency extends the small-world analysis to disconnected networks and to real systems that are better represented as weighted networks.

#### 3.3.1 Local and global efficiency on linear graph $\mathcal{L}(\mathbb{H})$

In this section we define both the global efficiency and the local efficiency measures for the linear graph level, which measure how efficiently passengers can move from one line to another from a global and a local point of view, respectively.

**Definition 11** We define the global efficiency indicator of the linear graph  $\mathcal{L}(\mathbb{H})$  as the average of the inverse of the distances in  $\mathcal{L}(\mathbb{H})$ , i.e.,

$$E_{glob}(\mathcal{L}(\mathbb{H})) = \frac{2}{|V(\mathcal{L}(\mathbb{H}))|(|V(\mathcal{L}(\mathbb{H}))| - 1)} \sum_{i < j} \frac{1}{d_{\mathcal{L}(\mathbb{H})}(L_i, L_j)}.$$

Once more, this indicator is defined if  $|V(\mathcal{L}(\mathbb{H}))| > 1$ , which implies that the corresponding CTLN consists of at least two lines.

The following proposition proves that this measure can be applied to evaluate the connectivity of CTLNs, because it satisfies the two basic properties we also proved for the characteristic path length: it belongs to a predefined range, and it is monotone (as well as the general invariance to scale changes). These two properties allow us to use this index to compare different networks.

**Proposition 3** Consider a CTLN  $G$ , and let  $\mathcal{L}(\mathbb{H})$  be its associated linear graph. We have that:

1.  $\frac{1}{\ell-1} \leq E_{glob}(\mathcal{L}(\mathbb{H})) \leq 1$ . The nearer to 1, the more interconnected the lines are.  $E_{glob}(\mathcal{L}(\mathbb{H})) = 1$  means that for all pairs of lines of  $G$  there is a transfer station that directly connects them.  $E_{glob}(\mathcal{L}(\mathbb{H})) = \frac{1}{\ell-1}$  if and only if  $\ell = 2$ .
2.  $E_{glob}(\mathcal{L}(\mathbb{H}))$  is monotone increasing in the sense that, if  $G'$  is obtained when adding a new link that connects two lines in  $G$ , we have that  $E_{glob}(\mathcal{L}(\mathbb{H})') \geq E_{glob}(\mathcal{L}(\mathbb{H}))$ , where  $\mathcal{L}(\mathbb{H})'$  is the linear graph of  $G'$ . Moreover,  $E_{glob}(\mathcal{L}(\mathbb{H})') > E_{glob}(\mathcal{L}(\mathbb{H}))$  if and only if the new link connects two lines that were not directly connected in  $G$ .

**Proof.** This proof can be done analogously as the proof to Proposition 1.  $\square$

We now define a new measure: the local efficiency indicator, which evaluates how efficiently passengers can move between lines from a local point of view.

**Definition 12** We define the local efficiency indicator as the average global efficiency of the subgraph  $\mathcal{L}_i(\mathbb{H}) = (V_i = V(\mathcal{L}_i(\mathbb{H})), E_i = E(\mathcal{L}_i(\mathbb{H})))$  formed by all first neighbors of  $L_i$  in  $\mathcal{L}(\mathbb{H})$ .

$$E_{loc}(\mathcal{L}(\mathbb{H})) = \frac{1}{|V(\mathcal{L}(\mathbb{H}))|} \sum_{i \in V} E_{glob}(\mathcal{L}_i(\mathbb{H})),$$

where

$$E_{glob}(\mathcal{L}_i(\mathbb{H})) = \frac{2}{|V_i|(|V_i| - 1)} \sum_{j < k} \frac{1}{d_{\mathcal{L}(\mathbb{H})}(L_j, L_k)}.$$

The reader may note that the same properties proven for the global efficiency indicator hold for the local efficiency indicator, that is:  $\frac{1}{\ell-1} \leq E_{loc}(\mathcal{L}(\mathbb{H})) \leq 1$ , it is invariant against scale changes, and it is monotone increasing. It is interesting to note that our intervals improve those that have traditionally been stated (in [16–18] the global and local efficiency indicators are shown to be in the interval [0,1]).

### 3.3.2 Local and global efficiency on transit hypergraph $\mathbb{H}$

We now adapt the same efficiency indicators as before for the transit hypergraph  $\mathbb{H}$ . In this case, the global efficiency and the local efficiency will measure how efficiently passengers can move between stations. We define the *global efficiency indicator* of the transit hypergraph  $\mathbb{H}$  as the average of the inverse of the distances  $d_{\mathbb{H}}(s_i, s_j)$  analogously as we did in Definition 11. Again, this indicator can be used to compare networks due to the properties proven in the following proposition:

**Proposition 4** Consider a CTLN  $G$ , and let  $\mathbb{H}$  be its associated hypergraph. Let  $E_{glob}(\mathbb{H})$  be the global efficiency of the transit hypergraph  $\mathbb{H}$ . We have that:

1.  $\frac{1}{\ell} < E_{glob}(\mathbb{H}) < 1$ . The nearer to 1, the more interconnected the stations are.  $E_{glob}(\mathbb{H}) = 1$  would mean that all pairs of stations of  $G$  belong to the same line, which is not possible because we are considering CTLN with at least two lines.
2.  $E_{glob}(\mathbb{H})$  is monotone increasing in the sense that, if  $G'$  is obtained when adding a new link to  $G$ , we have that  $E_{glob}(\mathbb{H}') \geq E_{glob}(\mathbb{H})$ , where  $\mathbb{H}'$  is the transit hypergraph of  $G'$ . Moreover,  $E_{glob}(\mathbb{H}') > E_{glob}(\mathbb{H})$  if and only if the new link connects two lines that were not directly connected in  $G$ .

**Proof.** See [6].  $\square$

Just like we did for the primal graph, we can define the local efficiency measure for the transit hypergraph  $\mathbb{H}$ , which will be denoted by  $E_{loc}(\mathbb{H})$ . The reader may note that  $E_{loc}(\mathbb{H})$  satisfies the same properties, that is, it belongs to  $(1/\ell, 1)$  and it is monotone increasing.

Table 1: Results of the studied measures for network Vienna, Los Angeles, Montreal and Moscow.

Networks	Levels	$C$	$\mathfrak{L}$	$E_{loc}$	$E_{glob}$
Vienna	Hypergraph	0.9413	1.9067	0.9678	0.5971
	Linear graph	0.8666	1.2	0.9333	0.9
	Multigraph	0.5833	–	–	–
Los Angeles	Hypergraph	0.9698	2.2113	0.9744	0.5621
	Linear graph	0.5333	1.5	0.6	0.7833
	Multigraph	0.3999	–	–	–
Montreal	Hypergraph	0.9258	1.7291	0.9598	0.6767
	Linear graph	0.5833	1.3333	0.5833	0.8333
	Multigraph	0.2916	–	–	–
Moscow	Hypergraph	0.9054	2.0050	0.9463	0.5506
	Linear graph	0.7866	1.3636	0.8919	0.8181
	Multigraph	0.4975	–	–	–

## 4 Numerical example

In this section, the applicability of the measures presented in this paper will be illustrated on several real CTLNs. In order to obtain comparable results, we focus on the same type of CTLN. More precisely, since some aspects of metro systems have been researched in the last years [3,4] we will consider four real world metro networks.

We have chosen the Vienna, Moscow, Montreal and Los Angeles metro networks since they constitute a good spectrum of several types of metro networks from size and connectivity points of view.

- Vienna metro network (see Figure 1) is a medium size network (88 stations) with, a priori, well connected lines. With respect to its structure, it can be classified as a cartwheel configuration. The percentage of transfer stations with respect to the total number of stations equals 0.1136.
- Los Angeles metro network (see Figure 2) is a medium size network (71 stations) with, a priori, low connected lines. With respect to its structure, it can be classified as a tree configuration. The percentage of transfer stations with respect to the total number of stations equals 0.0563.
- Montreal metro network (see Figure 3) is a medium size network (68 stations) with, a priori, rather well connected lines. With respect to its structure, it can be classified as a grid configuration. The percentage of transfer stations with respect to the total number of stations equals 0.0589.
- Moscow metro network (see Figure 4) is a large size network (147 stations) with, a priori, well connected lines. With respect to its structure, it can be classified as a cartwheel configuration. The percentage of transfer stations with respect to the total number of stations equals 0.1701.

For metro networks, the threshold  $U^{max}$  (needed to calculate the clustering coefficient  $C(\mathcal{L}^M(\mathbb{H}))$  for the linear multigraph) can be fixed to two, since two connected lines seldom share more than two transfer stations. It should be mentioned that, if two lines share several contiguous stations, which happens when two or more lines have a common track (as it occurs in Los Angeles network), then we only consider the first and the last stations of the track as transfer stations between these lines.

In Table 1 and Table 2, it can be observed that the behaviour of the measures is not same at both the linear and the hypergraph levels, in the sense that they do not lead

Table 2: Different orderings of the metro networks Vienna (Vn), Los Angeles (LA), Montreal (Mn) and Moscow(Ms), depending on measure and level of abstraction (based on values of Table 1).

Levels	$C$	$\mathcal{L}$	$E_{loc}$	$E_{glob}$
$\mathbb{H}$	LA-Vn-Mn-Ms	Mn-Vn-Ms-LA	LA-Vn-Mn-Ms	Mn-Vn-LA-Ms
$\mathcal{L}(\mathbb{H})$	Vn-Ms-Mn-LA	Vn-Mn-Ms-LA	Vn-Ms-LA-Mn	Vn-Mn-Ms-LA
$\mathcal{L}^M(\mathbb{H})$	Vn-Ms-LA-Mn	–	–	–

to the same ordering of the networks based on the values of each coefficient. That is, according for example to the clustering coefficient, at the hypergraph level Los Angeles achieves the best value, followed by Vienna, Montreal and Moscow, whereas at the linear level is Vienna the one that reaches the best value of this coefficient, followed by Moscow, Montreal and Los Angeles.

In order to better understand the behavior of the measures at each level, we will analyze the particularities of the networks and their relation with the values obtained in Table 1 for each level.

Concerning the clustering coefficient at the hypergraph level, as stated above, Vienna's is lower than Los Angeles', which is due to the fact that the percentage of non-transfers stations in Vienna is lower than in Los Angeles. Note that, when considering the primal graph or the graph of direct connections [15], non-transfer nodes lead to  $C_i = 1$  (see Theorem 1), which increases the average  $C$ . In other words, for two networks with the same number of stations, the higher number of transfer stations the lower the clustering coefficient. So, transfer stations are somehow penalizing the value of the clustering coefficient. This analysis of the transfer system should be complemented to have a more complete vision of the reality and also in order to avoid extreme cases, such as the example of a network with all its lines disconnected and whose clustering coefficient would be equal to one (one is the upper bound of the clustering coefficient).

The linear level becomes then necessary in order to complete the analysis of the CTLN so that transfer stations do not penalize. Note that at this level, stations are not taken into consideration but just lines and their connections. A network with all its lines disconnected has a clustering coefficient equal to zero at the linear graph level, since the corresponding linear graph has no edges. So, depending on the level of abstraction taken into consideration, the maximum or the minimum threshold for the clustering coefficient is achieved.

At the linear graph it can be observed that, for example, Vienna and Los Angeles have the same number of nodes, since both are networks consisting of five lines. However, the linear and multiple linear clustering coefficients are higher for Vienna than for Los Angeles at this level. This is due to the fact that Vienna's linear graph is nearly a complete graph (see Figure 1), in the sense that most lines are pairwise connected. So, transferring between two lines is normally allowed by means of just one transfer.

Note that lines in the Moscow network are well connected. This fact is reflected in the average degree of the nodes in the linear graph (close to 7) and in the high values of the clustering coefficient at the linear and multiple graph. However, at the hypergraph level, the value of this measure is the lowest of the four networks, due to the fact that the percentage of non-transfer nodes is the highest amongst them.

Recall that, at the multigraph level, each edge between two nodes (lines) represents a transfer station between the lines. As it can be observed in Figure 3, any pair of lines of Montreal network share no more than one transfer station, so there are no multiple

edges at its corresponding multigraph. This leads to the fact that the value of the clustering coefficient at the linear level is exactly twice the clustering coefficient at the multigraph ( $U^{max}$  equals two). Another consequence of the absence of multiedges at Montreal multigraph is that, if we compare it with Los Angeles network, although the clustering coefficient at the linear graph is larger for Montreal than for Los Angeles, the opposite occurs considering the clustering coefficient at the multigraph since Los Angeles network does have multiedges at the multigraph.

As stated by [2] and commented in Section 3.3, it can be observed that the clustering coefficient and the local efficiency play the same role, as it can be observed at the hypergraph level, where the order induced by the local efficiency is the same as the one induced by the clustering coefficient: Los Angeles achieves the best value, followed by Vienna, Montreal and Moscow. Also, the global efficiency and the inverse of the characteristic path length have the same behavior (see [2] and Section 3.3), as it can be seen in the order of the networks induced by these concepts: Vienna, Montreal, Moscow and Los Angeles.

The characteristic path length represents the average number of transfers one needs for traveling between nodes, understanding nodes as lines at the linear level and as stations at the hypergraph level. Analogously as commented for the clustering coefficient, the characteristic path length does not necessarily induce the same order at both levels. For example, the average number of transfers one needs to travel between lines is smaller for Vienna than for Montreal and the opposite happens for transferring between stations. Again, this responds to the fact that Vienna's linear graph is nearly a complete graph, which makes it the network that achieves the best values of the measures at the linear level (see Table 2). The number of transfers one needs for transferring between lines at Vienna's metro network is then the lowest of the studied ones since most of its lines are directly connected. However, if one observes in deep the Vienna's metro network (Figure 1), line 6 is the one with the highest number of stations and it has no direct connection with line 2 neither with line 3. This absence of connections between lines with high number of stations elevates the average of the number of transfers one needs to do for transferring between stations. On the contrary, the line in Montreal's metro network (Figure 3) with the highest number of stations is line 1, followed by line 3. These two lines are directly connected, which decreases the average number of transfers one needs to do for transferring between stations in the whole network.

## 5 Conclusions

In this work, we have represented a collective transportation line network (CTLN) by means of hypergraphs and their corresponding linear graphs. In order to analyze the functioning of a CTLN, with respect to the number of transfers, several measures have been considered.

Two different abstraction levels and their respective coefficients regarding line routes are presented in this work. In the examples, each level corresponds to different interpretations of reality: the linear graph level measures connectivity between lines, whereas the hypergraph level measures connectivity between stations.

We have also seen that hypergraphs significantly reduce the computation of the clustering coefficient with respect to other methods found in the literature.

## Acknowledgements

This work was partially supported by Ministerio de Educación, Ciencia e Innovación (Spain)/FEDER under project MTM2009-14243 and by Junta de Andalucía

(Spain)/FEDER under excellence projects P09-TEP-5022 and FQM-5849. Special thanks are due to two anonymous referees for their valuable comments and advice.

## References

1. Watts, D.J. and Strogatz, S.H., *Nature* **393**, (1998) 440-6684
2. Latora, V. and Marchiori, M., *Physical Review Letters* **87**, (2001) 198701-19
3. Latora, V. and Marchiori, M., *Physica A* **314**, (2002) 109-113-1
4. Criado, R. and Hernández-Bermejo, B. and Romance, M., *International Journal of Bifurcation and Chaos* **17**, (2007) 2289-2301-7
5. De-Los-Santos, A. and Laporte, G. and Mesa, J.A. and Perea, F., *Transportation Research Part C: Emerging Technologies* **20**, (2012) 34-46
6. Barrena, E. and De-Los-Santos, A. and Mesa, J.A. and Perea, F., Technical proofs of paper "Analyzing Connectivity in Collective Transportation Line Networks by means of Hypergraphs" , (2012)
7. Berge, C., *Graphs and Hypergraphs* (Elsevier Science Ltd, 1985)
8. Berge, C., *Hypergraphs: combinatorics of finite sets* (North Holland)
9. Watts, D.J. , *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton 1999) 262
10. Newman, M.E.J., Technical report, Santa Fe Institute, (2001)
11. Newman, M.E.J., *Physical Review E* **64**, (2001) 016131-1
12. Newman, M.E.J., *Physical Review E* **64**, (2001) 016132
13. Estrada, E. and Rodríguez-Velázquez, J.A., *Physica A: Statistical Mechanics and its Applications* **364**, (2006) 581-594
14. Dechter, R., *Constraint Processing* (Morgan Kaufmann, 2003) 450
15. Sen, P. and Dasgupta, S. and Chatterjee, A. and Sreeram, P.A. and Mukherjee, G. and Manna, S.S., *Physical Review E* **67**, (2003)
16. Crucitti, P. and Latora, V. and Marchiori, M. and Rapisarda, A., *Physica A* **320**, (2002) 642-665 cond-mat/0205601
17. Latora, V. and Marchiori, M., *Chaos Solitons Fract.* **20**, (2004) 69-75-1
18. Latora, V. and Marchiori, M., *The European Physical Journal B* **32**, (2002) 249-263-2

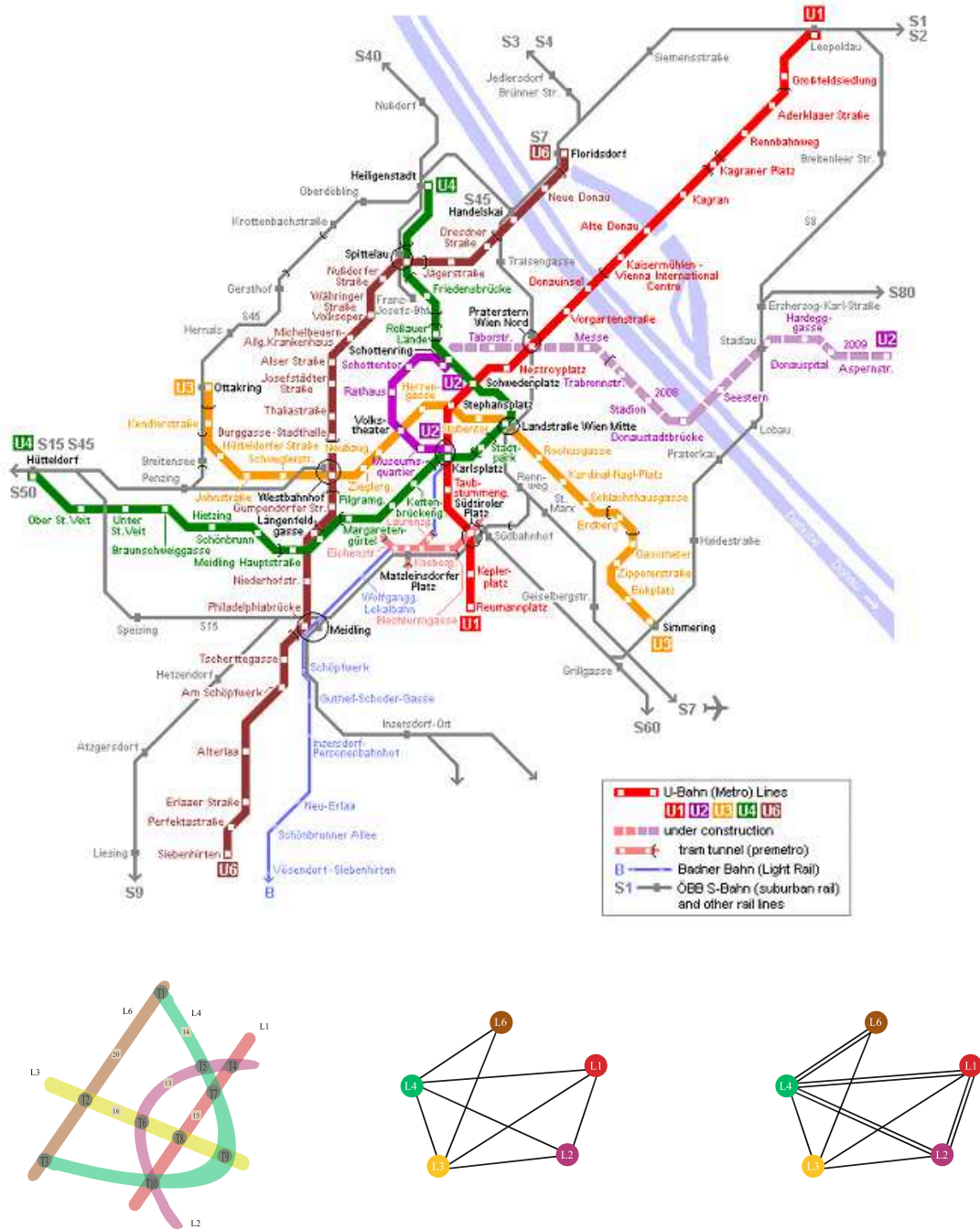


Fig. 1: Vienna metro network. Top graph is the complete network, consisting of 88 stations, out of which 10 are transfer stations (9 double and 1 triple), and 5 lines in 2012. Source: Urban Rail 2012. Bottom graphs are, from left to right, the corresponding transit hypergraph, the linear graph, and the linear multigraph.

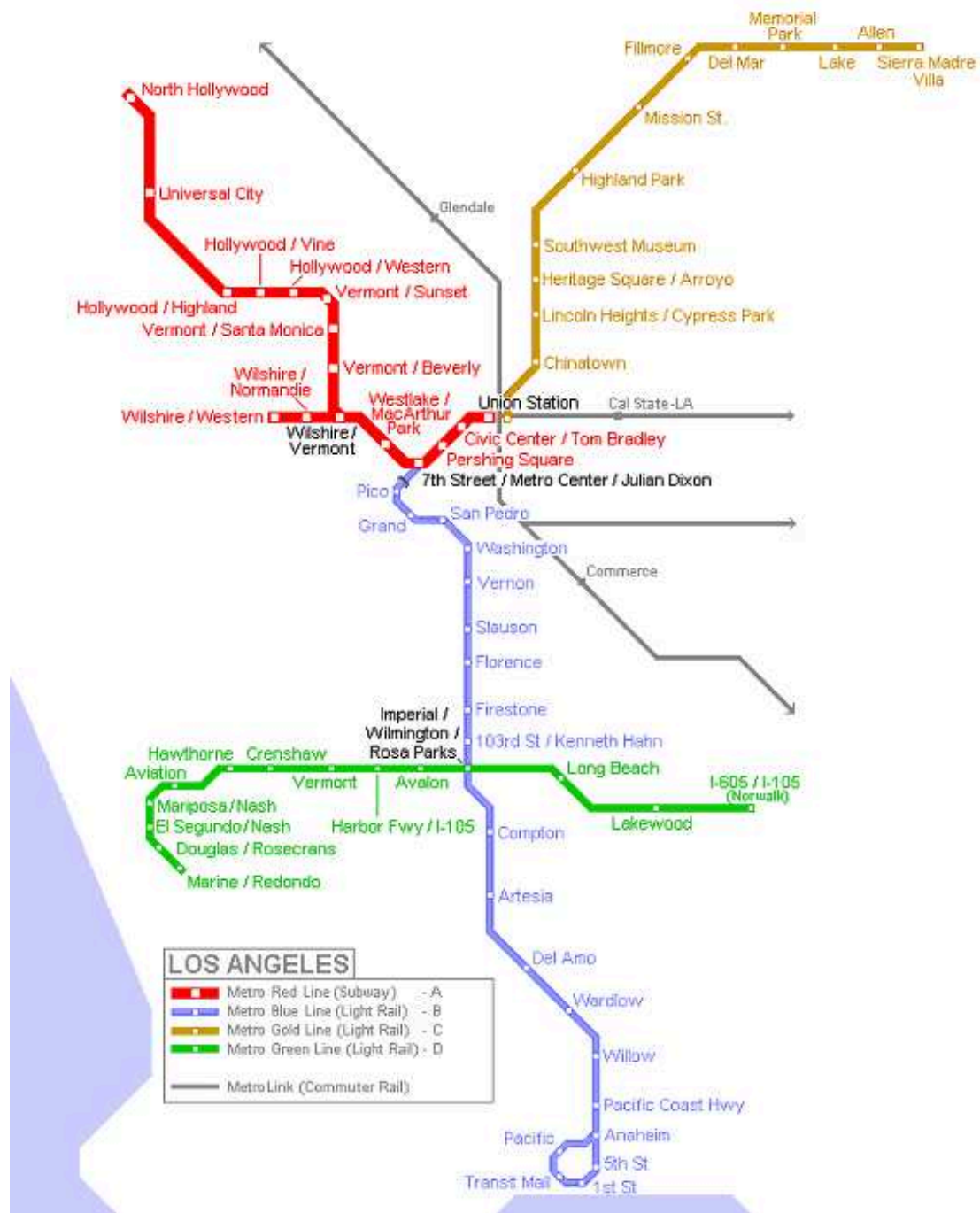


Fig. 2: Los Angeles network, consisting of 5 lines, 71 stations out of which 4 are transfer stations (2 double and 2 triple), in 2004. Source: Urban Rail. The reader should note that the red line is considered as two different lines.





Fig. 3: Montreal network, consisting of 4 lines, 68 stations out of which 4 are transfer stations (3 double and 1 triple), in 2012. Source: Urban Rail



Fig. 4: Moscow network, consisting of 11 lines, 147 stations out of which 25 are transfer stations (18 double, 6 triple and 1 quadruple), in 2006. Source: Urban Rail