COMPARED ANALYSIS OF SUBWAY NETWORKS SUPPORTED BY INFORMATION THEORY

Josep Barberillo and Joan Saldaña

Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona, Girona

INTRODUCTION

The structural properties of the subway network are crucial in effective transportation in cities. This study presents an informational perspective of navigation in four different subway networks: New York, Paris, Barcelona (including tramway lines) and Moscow. We want to address our study to investigate what makes it complicated to navigate in these kinds of networks. Furthermore, we want to carry out a comparison between them and their intrinsic geographical and evolutional constraints. This poster is focused on a navigational evaluation of subway networks through the definition of a methodological approach based on a set of indicators which are defined in the references section.

ACCESS AND HIDE INFORMATION

To characterize the ease or difficulty of navigation in subway networks, we use the "Search Information" (S) [1]. Without prior knowledge, the information needed for locating a given exit from a node of degree k, is $\log_2(k)$ yes/no questions to guess the correct link. For each path p(s,t) from *s* to *t* the probability to follow it is:

$$P[path(s,t)] = \frac{1}{k_s} \cdot \prod_{j \in path(s,t)} \frac{1}{k_j - 1} \qquad (1)$$

The total informational value of knowing any one of the degenerate paths between *s* and *t* is therefore:

$$S(s \to t) = -\log_2 \sum_{\{path(s,t)\}} P[path(s,t)] \qquad (2)$$

To quantify how difficult it is to find one vertex starting from an arbitrary node in the network in [1] it is defined the "Hide Information" (H) as:

$$H_t = \frac{1}{N-1}\sum_{s \neq t} S(s,t) \;\neq\; A_t = \frac{1}{N-1}\sum_{s \neq t} S(t,s) \qquad (3)$$

In the same way the quantity A_t is a measure of how good the access to the network is from node t (average Access Information needed to reach any other node).

RANDOMIZATION PROCEDURE

Randomized versions of the subway networks were constructed by randomly reshuffling links maintaining the same degree as in the original network [3]. The iterative algorithm consists of first randomly selecting a pair of edges {A,B} and {C,D}. The two edges are then rewired in such a way that A becomes connected to D, while C connects to B. This step is aborted if one or both of these new links already exists (preventing the appearance of multiple edges connecting the same pair of nodes). A repeated application of this step leads to a randomized version of the original subway network keeping it globally connected.

GRAPH REPRESENTATION — SPACE L & SPACE P

The ideas of Space L and Space P are proposed in general terms in [2]. The first topology (Space L) consists of nodes representing subway stops, and a link between two nodes exists if they are consecutive stops on one subway line. In Space P, the nodes are the same as in the previous topology but here an edge between nodes means an underground line links them (see Fig.1).

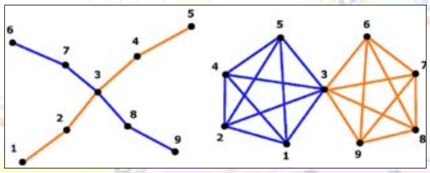

Figure 1. Explanation of Space L (left) and Space P (right) [4].

SUBWAY NETWORKS

We have based our study on different networks with different sizes because it permits us to establish behavior patterns as a result of these specific characteristics (common geographical constraints, etc).

	SPACE L			
	N	M	<k>	<A>
New York	496	605	2.44	16.43
Paris	300	353	2.35	10.69
Barcelona	203	237	2.33	8.99
Moscow	183	215	2.35	7.34

Table 1. Network indicators: N (number of nodes), M (number of links), <k> (average degree), <A> (average Access Information).

RESULTS

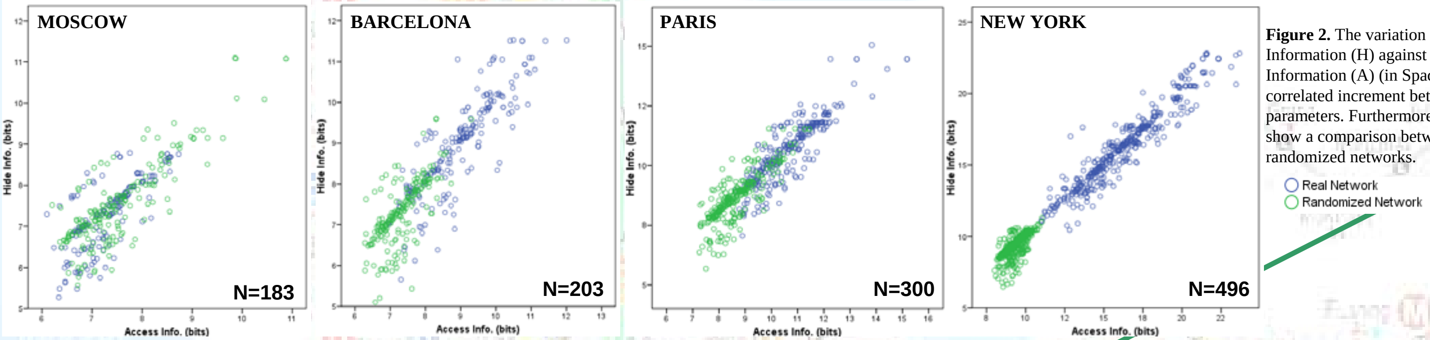

Figure 2. The variation of the Hide Information (H) against the Access Information (A) (in Space L) indicates a correlated increment between these two parameters. Furthermore, these graphics show a comparison between real and randomized networks.

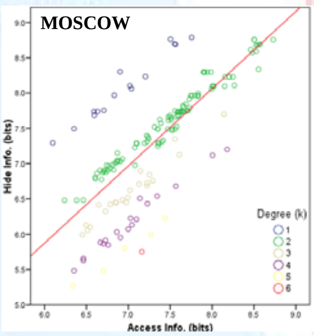

Figure 3. The Hide Information average value as a function of degree (k). Circles are colored according to the node degree.

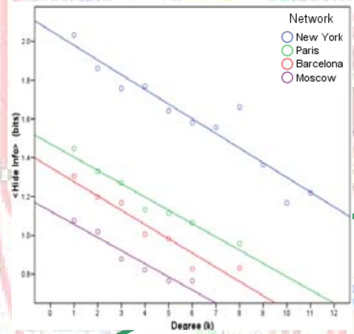

Figure 4. The Hide Information average value (<H>) as a function of degree (k) for four chosen cities (Space L).

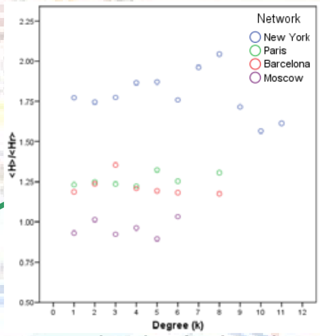

Figure 5. It shows the real-random Hide ratio $\langle H(k)\rangle / \langle H_r(k)\rangle$. The degree of a node plays a minor role for Hide Information (H) (Space L).

BARCELONA INTEGRATED NETWORK

In order to show the connectivity among railroad tracks in the integrated subway network of Barcelona, we analyzed the Space P selecting one node of each railroad track (Fig 6).

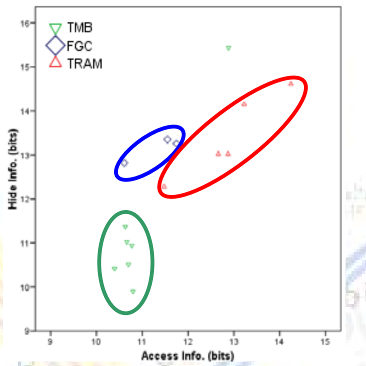

Figure 6. Plot of the dependence between Hide Information (H) and Access Information (A) for each railroad track considered (Space P) as a function of the operating company (TMB, TRAM or FGC).

CONCLUSIONS

- The observation of a universally large S relative to S_r in all subway networks means that the ability to obtain information is more important in these real networks. This value increases with the size of the network (N) (Fig.2).
- The topological differences between real and randomized networks increases with the size of the network (N). These differences are caused by the geographical constraints (embedding the network in a two-dimensional (2D) space).
- The randomized version of the subway network does not show sensible differences with the size of the network (N). Topological features are similar for randomized networks independent of their size.
- Basic levels of Hide Information (H) which are needed to find a specific station are fixed depending on the size of the network (N). From this residual value, the Hide Information (H) depends only on the degree of the station (k). The same behavior was observed in every subway network analyzed (Fig. 4).
- Having good levels of accessibility depends on the good position of the station inside the subway system.

CONCLUSIONS

- Poor levels of overall communication between railroad tracks are showed in Barcelona Integrated Network (Fig 6).
- Depending on which company is providing the service, the different railway lines present huge variations in the navigability levels (there is a strong modularity in the service).
- TMB lines have good levels of accessibility and are easily reachable (low Hide Information value).
- It is important to plan the integrated rail network systems in an effective manner to maintain good cooperation between the different parts of the subway networks because of the massive investment needed to build them.

References

[1] M. Roswall, A.Trusina et al. *Networks and Cities: An information Perspective*, Phys. Rev. Lett. **94**, 028701 (2005).
[2] S. Dasgupta, A. Chatterjee et al. *Small-world properties of the Indian railway nework*, Phys. Rev. E **67**, 036106 (2003).
[3] S. Maslov and K. Sneppen. *Specifity in topology of protein networks*, Science **296**, 910-913 (2002).
[4] J. Sienkiewicz and J. Holyst. *Statistical analysis of 22 public transport networks in Poland*, Phys. Rev. E **72**, 046127 (2005).

4th UniNet Workshop **"New Directions in Network Modelling"**
9th June - 12th June 2008, Université René Descartes, Paris, France