Does Road Network Topology Affect Real Estate Pricing? The Naples Case Study

Arianna Nocente[1], Jarir Salame Younis[1], Marco Cozzolino[1] and Giulio Rossetti[2]

[1] University of Pisa, Italy
arcente@gmail.com, jarir.s.y@gmail.com, marcocozzolino0@gmail.com
[2] KDD Lab, ISTI-CNR, Italy
giulio.rossetti@isti.cnr.it

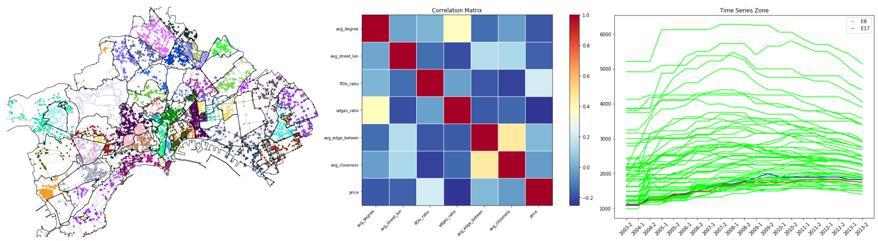
1 Introduction

Nowadays, the estimation of house selling prices represent a hot and challenging task.The most widely used methods to address it rely on standard machine learning techniques, namely Artificial Neural Network and Hedonic. The former approach performs well with incomplete or unknown data [1], while the latter assumes that the selling price can be seen as a set of attributes and that the buyer tends to maximize its utility function [2, 3]. The main challenge of such a task lies in identifying those parameters that influence selling value since, often, they are interdependent or even hidden.

Using the road network of an Italian city, Naples (Figure 1(a)) – along with the dynamics of its points of interest –, our research aims to understand the relation between it and the real estate market prices. The road network was built identifying as edges the city roads and as nodes their intersections. Anomalies in the original data have been corrected through a cross-reference with the walk network. From a preliminary structural analysis emerges that the network shares some characteristics with subcritical random networks (e.g., the degree distribution) while at the same time, it does not exhibit the small-world property[4]. The degree distribution is close to the Gaussian distribution; planarity imposes severe constraints on the degree of a node and on its distribution, which is generally peaked around its average value[5].

We partitioned the network into sub-networks identifying the territorial boundaries of the 65 land registry areas[1]. For economic analysis, the network was enriched with main information about points of interests (POIs) which might influence the purchase choices of the housing stock made by population. The geospatial coordinates of such POIs have been identified and assigned to the nodes of the network which were nearest to them. A different approach has been used for educational and health institution POIs since the capacity and consequently the size of the universities and the hospital are not negligible. To take into account their relevance they have been mapped to areas and the nodes within them have been assigned to the same POI. The categories to which a POI belongs to and the number of nodes identified by them in the original network are, respectively: Public schools (70 nodes), Universities (21 nodes), Public and private hospitals (186), Parks (34 nodes). The POI nodes are uniformly distributed among the

[1] The coordinates of these area boundaries were obtained through Geopoi, a cartographic visualization software for territorial navigation service.



(a) Naples Road Network (b) Correlations (c) House Sales Estimates

Fig. 1: (a) Naples Road Network (with nodes subdivided by cadastral areas); (b) feature/pricing correlations; (c) pricing trends per area.

identified subgraphs with the exception of the cadastral areas corresponding to industrial area and the city centre.

2 Preliminar Results

To investigate the correlations between estimates of sales prices, points of interest and structure of the network, each subgraph was characterized by the following network measures: (i) *Average degree*: average degree of nodes in the area; (ii) *Average street length*: the average value of the street length; (iii) *Average edge betweenness*: represent the importance of the streets and it is related to the number of commercial activities [6]; (iv) *Average closeness centrality*: related to the distance from any other node in the network; (v) *POIs Ratio*: number of POIs divided by the population of the area; (vi) *Edges Ratio*: number of streets divided by the population; this measure shows how well the street network serves citizens. The number of POIs was normalized using the expected population of the area to take into account the number of persons who would benefit from the services offered by them. The expected population of each cadastral area was estimated using the population and the borders of neighbourhoods assuming an equal distribution of the inhabitants inside neighbourhoods.

From our analysis, the Kendall and Pearson correlations among the identified features and the average cluster pricing appear close to zero (Figure 1(b)). Such a negative result seems to highlights how other features than the road network structural ones can explain the laws of the real estate market for Naples.

Indeed, one limit of the actual analysis lies in the absence of road network historical data. However, if such data were available, it would have been possible to train a predictive regressor that, taking into account how the city connectivity has evolved as well as when new POIs appeared, could have shown interesting predictive performances.

In absence of such information, we conducted a punctual survey by comparing the evolution of the price for the 65 Naples cadastral areas. To such extent, we leveraged real estate quotations data[2] over the decade from 2002 to 2013, detailed by semesters;

[2] Agenzia delle Entrate - Banca dati delle quotazioni immobiliare (OMI)

this data has been used to generate and compare the time series of each cadastral area (Figure 1(c)). For each area we computed the price linear trend; the trend coefficient shows how prices change over time. In particular, we focused our attention on the two areas where two important POIs (a hospital and a university) have been introduced during the observed period.

The average trend coefficient across all areas is 19.24. However, when considering the two selected areas the coefficient changes considerably reaching 35.84 in the new hospital area (Figure 1(c) red) and 41.42 in the new university one (Figure 1(c) blue). Such results underlying how for those areas the sales prices tend to grow twice as fast as in other areas of the city. Indeed, the evolution of the road network through the introduction of relevant POIs causes a strong evolution in prices. Moreover, such a result confirms what observed by the static analysis conducted so far: the prices in the cadastral areas where new POIs appear to change in an evident way, undergoing a rise, never become the highest prices in absolute terms, as they are the result of a historical evolution of more than ten years.

Dynamic analysis can be used as a tool to support forecasts combining the historical trend of prices with the evolution of the road network with a particular focus on the evolution of points of interest. As future work, we plan to collect data on the evolution of the Naples road network so to better formalize a dynamic graph theory framework able to explain the real estate market in terms of topology dynamics. In particular, integrating dynamic road network analysis with economic and cultural information (hospitals, schools, population) we are certain that a novel class of approaches for pricing forecasting can be devised. Moreover, such a framework could be used as a guide to identifying those areas for which urban re-qualification is needed or, in an entrepreneurial context, to choosing where to open a new POI to maximize the effects on sales prices.

Acknowledgment

This work is partially supported by the European Community's H2020 Program under the funding scheme "INFRAIA-1-2014-2015: Research Infrastructures" grant agreement 654024, http://www.sobigdata.eu, "SoBigData".

References

1. Shaw, J.:Neural network resource guide. AI Expert. 8(2), (1992)
2. Griliches, Z.: Price Indexes and Quality Changes: Studies in New Methods of Measurement. Harvard University Press (1971)
3. Rosen S.: Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economics. 82, (1974)
4. Katzav, E. and Biham, O. and Hartmann, A. K.:Distribution of shortest path lengths in subcritical Erdős-Rényi networks. Phys. Rev. E. 98, (2018)
5. Barthlemy, M.: Spatial networks. Phys. Rep. 499(13), 1 – 101 (2011)
6. Porta, S. and Strano, E. and Iacoviello, V. and Messora, R. and Latora, V. and Cardillo, A. and Wang, F and Scellato, S.: Street centrality and densities of retail and services in Bologna, Italy. Environment and Planning B: Planning and design. 36(3). (2009)