

Data Extraction from Social Networks for Urban Analyses

Combining Space Syntax and user-generated data analysis on Lisbon

Stefano Fiorito¹, Francesco Orsi², Francisco Manuel Serdoura³, Victor Ferreira⁴

CIAUD, Faculty of Architecture, Technical University of Lisbon, Rua Sá Nogueira, Pólo Universitário, Alto da Ajuda, 1349-055 Lisboa, Portugal

¹<http://home.fu.utl.pt/~fiorito/>, ²<http://home.fu.utl.pt/~orsi/>, ³<http://home.fu.utl.pt/~fs>

⁴<http://home.fu.utl.pt/~victor/>

¹fiorito@fa.utl.pt, ²orsi@fa.utl.pt, ³fs@fa.utl.pt, ⁴victor@fa.utl.pt

Abstract. *The present work constitutes the first stage of an ongoing research on the interaction between morphological characteristics of the urban fabric and the amount of social activity in such spaces. In order to analyze such correlation, the current research links together two different fields of studies: Space Syntax on one side, for the morphological analysis of the spatial configuration of the urban fabric, and a remote sensing study about social activity in the analyzed urban context, on the other. Data extracted from location-based online Social Networks databases (e.g. Foursquare) are employed in order to perform such survey. The resulting methodology constitutes an early attempt to set a novel approach to the study of the relationships between the morphological and configurational characteristics of urban systems and actual human dynamics in urban contexts.*

Keywords. *Space Syntax; urban morphology; remote sensing; social networks; urban dynamics.*

INTRODUCTION

During the last years online Social Networks (OSNs) increasingly grew in popularity becoming a major way of communicating over the internet. The advent of location-based online Social Networks (LBOSNs), e.g. Foursquare, where users voluntarily share their current location via the global positioning system (GPS) embedded in their mobile phones, opens the door to novel detailed analyses on patterns of human movement through urban environments, ultimately allowing for the study of the role of spatial configurations and geographic variables in influencing flows of movement and social dynamics in the urban context.

Ultimately this work aims at developing a feedback system for evaluating the correspondence between the data gathered by the analysis of human activity extracted through location-based online Social Networks and the results returned by Space Syntax analyses. Thus the validity of this methodology is tested for mapping – or forecasting – patterns and trends of human dynamics in urban contexts. The potential of such methodology lays in the foundation of a cross-referenced feedback system based on both the real-world mapping of actual human behavior in urban context and the Space Syntax analysis. In the proposed work we attempt to deve-

lop such comparative study and to achieve some early conclusions how the two entities – spatial configuration and social patterns – relate and influence each other. This paper presents an early sketch of such methodology.

THEORETICAL FRAMEWORK

The first part of the research is constituted by a morphologic and configurational study of the city of Lisbon, Portugal. Such study has been conducted following the Space Syntax methodology, as presented by Bill Hillier (1989). Space Syntax is a theory and a set of methods for the analysis of urban morphological configurations: it has been used to treat spatial configuration as a fundamental variable in a multitude of studies about social functions, cultural significance and behavioral implications of urban layouts, in contemporary and historical contexts. Space syntax has also been used to explore, predict and evaluate the likely effects of design alternatives. It has been extensively applied in the fields of architecture, urban design, planning and transportation as a tool to analyze urban fabrics in terms of their spatial configuration characteristics, obtaining information about different indicators developed by Space Syntax itself – such as Integration, Choice and Connectivity – that could be later treated numerically for mathematical and statistical calculus.

Thus Space Syntax aims to describe how spatial patterns in the city influence the way people tend to flow through the city. It can also help to understand how a specific urban layout can lead to social and economic benefits such as safer public spaces, more successful shops and higher property values.

Based on mathematical analysis of street networks and their spatial configurations, the Space Syntax methodology allows for a clear, quantitative and objective comparison between the configurational values returned by each element in the urban grid and also for a flawless cross-reference with other kinds of statistical information gathered about the same place. Due to this reason, for each element of the urban grid, it is possible to put in direct correlation Space Syntax results stored in a specific da-

tabase with other kinds of data gathered from other sources.

In the present study we decided to make use of Space Syntax in order to be able to formulate hypothesis about the correlation between street network layouts and actual amount of human activity within a given urban context. To do so, we aimed at compare the results returned by Space Syntax with the data retrieved through location-based online Social Networks in order to seek for congruence relations – or possibly incongruence or fuzzy relations.

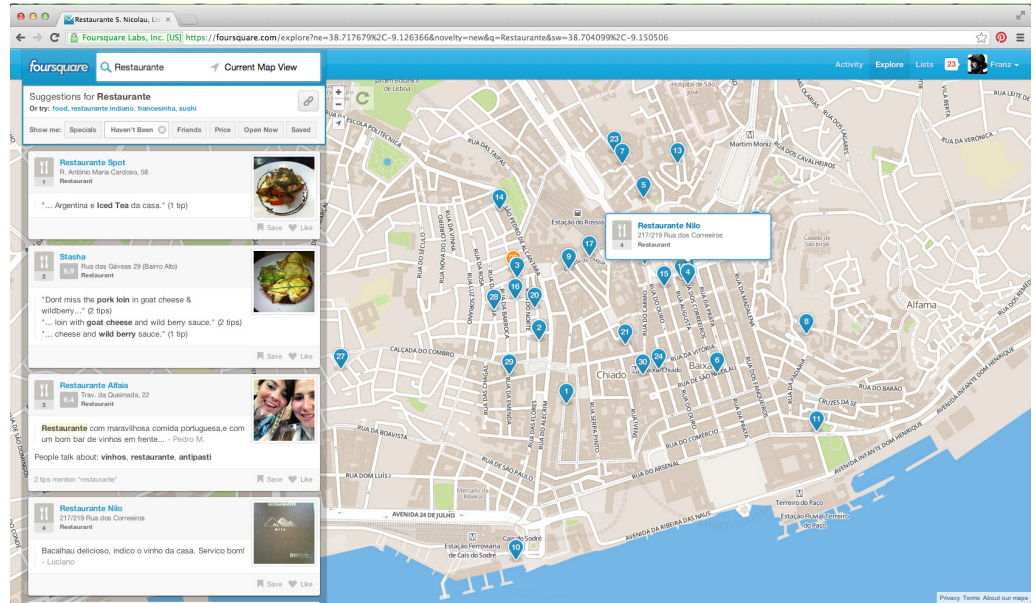
Foursquare is a location-based social networking website for mobile devices, such as smartphones (Figure 1). Users “check-in” at venues using a mobile website, text messaging or a device-specific application by selecting from a list of venues that the application locates nearby. Location is based on GPS hardware in the mobile device or network location provided by the application.

A major hypothesis at the base of this study is that the geo-referenced information resulting from “check-ins” created by individual users throughout the town can constitute a valid sample – although very narrow and partially biased – of the patterns of social activity and interaction throughout the town. Certainly the chosen sample suffers from being bounded to the specific behaviors of a very narrow subset of smartphone equipped users (willing to share their geo-referenced information on Foursquare) that doesn’t necessarily document the behavior of the all urban population. Moreover we must take into account that the kinds of information that users are willing to transmit over the social networks does not describe actual human behavior, but declared human behavior. Therefore one can say that behavior in social networks is fundamentally flawed and highly biased.

Although being aware of the fundamental issues that an analysis based on such foundations can encounter, we defend nevertheless the validity of the present approach, believing that valid emergent patterns of more general behaviors can be extracted from such partially flawed sample and also knowing that the insights we are gaining about actual social

Figure 1

A screenshot of the Foursquare web interface – the Baixa area of Lisbon. It is possible to run queries about venue category (such as Restaurants, Bars, Hotels, etc.) as well as subcategories (such as restaurant typology: Portuguese, Italian, Asian Cuisine, etc.). All this data is also augmented by reviews and pictures provided by the users.



behaviors and trends due to such approach proved to be highly valuable in order to understand the dynamics that come into play to form the social fabric of the city.

OBJECTIVES

The main objective of the present research is to develop a methodological framework for verifying the actual correlation between urban morphology indicators and the amount of social and commercial activity present in a specific place.

This operation is performed by comparing the results returned by the morphological and configurational analysis performed by means of Space Syntax method with the results of quantitative and qualitative analysis on the user-generated data provided by the Foursquare database. Automating the process of retrieving data from social networks databases and refining them through an automatized routine is a side condition – although very fundamental – for the correct definition of a viable methodological approach to such analysis.

A major objective for the present work would be combining the results of Space Syntax analysis and the gathering of geo-located data retrieved from Foursquare in order to throw light on how and why human activity in urban areas arises in correlation with certain morphological configurations of the urban system.

What is interesting in such combination of data gathered from different sources is the possibility of crossing information based on purely spatial analysis with other kinds of information provided by the detailed data about land use, type of activities, popularity of specific places on the grid returned by the Foursquare database.

DESCRIPTION

The methodology employed in the present research is constituted by a set of well-defined subsequential steps, capable of being replicated for other contexts and case studies when needed.

The first stage consisted in a preliminary survey on the case study area through remote sensing us-

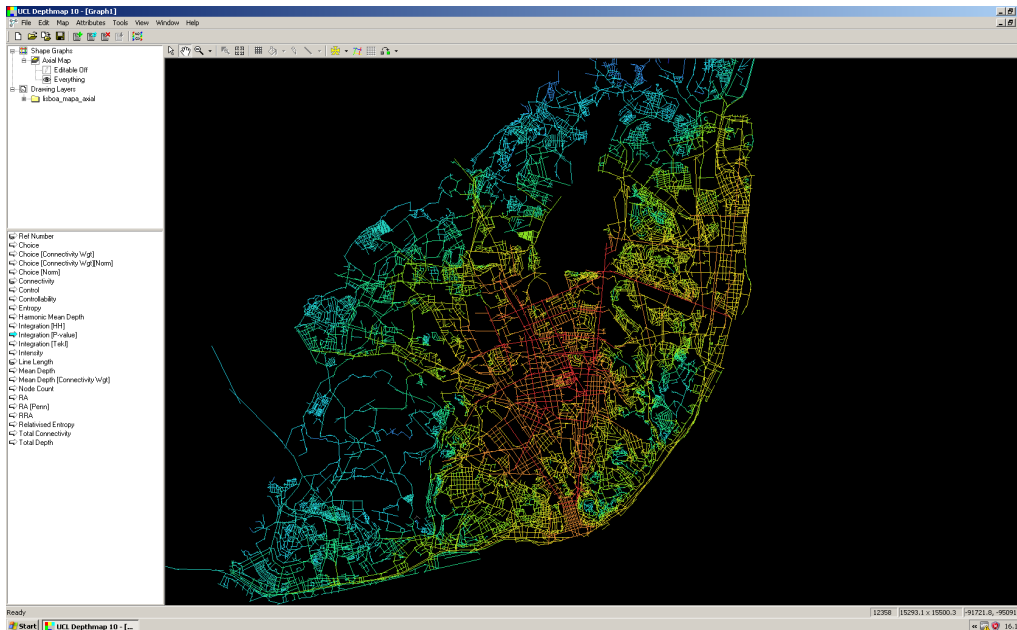


Figure 2
A screenshot of the spatial configuration analysis performed by Depthmap software – the analysis of Integration value for Lisbon city area. Given a specific axial map Depthmap is able to compute the numerical values about the related indicator for each segment in the map. The warmer the color, then more integrated the segment in the overall system.

ing location-based online Social Networks data (e.g.Foursquare). At this stage we checked the amount and quality of available data regarding the chosen case-study area in order to be able to decide whether the data available for the chosen area was enough to justify such approach over a more traditional in-field research. This step allows for a reasonable delimitation of the case-study area and for a definition of the variables considered capable of returning consistent results.

Then a set of procedures for automating the process of retrieving data from Foursquare API has been developed and the information gathered from this process has been stored in a local database. In order to do so we created a custom-made query system base on Apigee console for automating the retrieving of large amounts of data about the selected case study. Alongside this operation, a methodology for refining such data has also been defined: here we made use of Google Refine web application to clean the gathered data and make comparative analysis

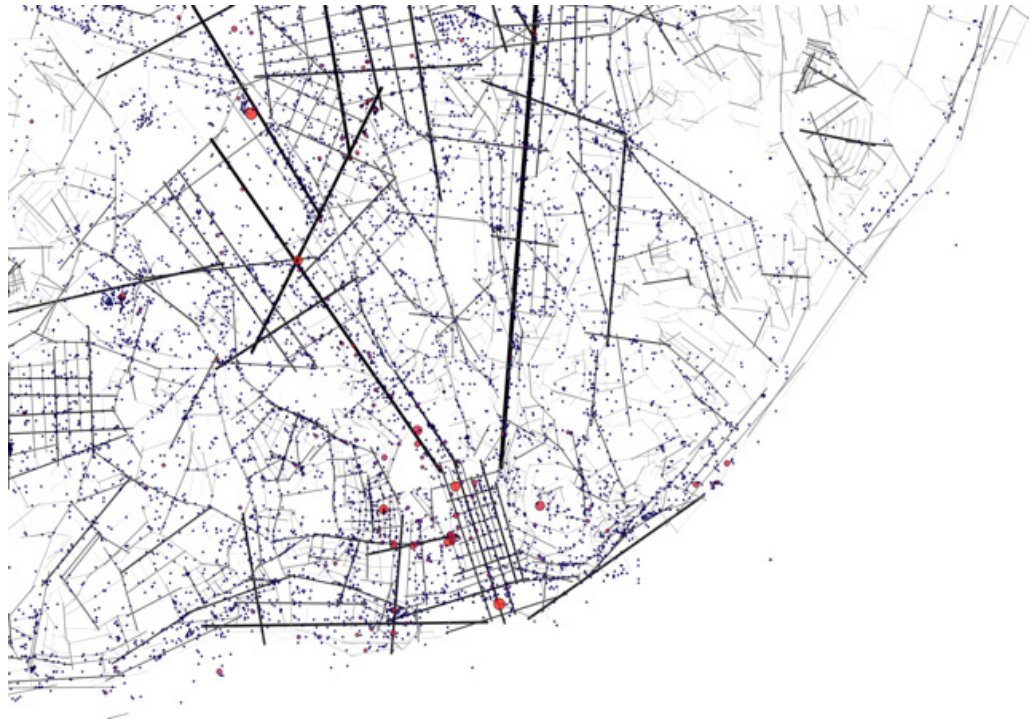
between this data and the Space Syntax data possible in a later stage of the research.

Such tasks – together with the task of implementing an intuitive user interface for visualizing the research results – are to be considered as fundamental parts of the research.

The third step of the work is then constituted by the morphological analysis of the chosen case-study area following the Space Syntax routine. Firstly, we created an extensive axial map of the whole municipality of Lisbon (taking into account 20 segment crossings outside of the delimited borders of the case study area in order to minimize the error in calculus with Depthmap on a later stage) (Figure 2). In order to accomplish this tasks both Depthmap (the original and fundamental piece of software developed by Space Syntax labs to compute calculus on the basis of provided axial maps) and GIS software were used.

The last step of the research is constituted by the comparison between the results of the morpho-

Figure 3
Map of Lisbon – Baixa area.
The axial map showing the connectivity (using the thickness and gray gradient); the coloured dots (blue to red gradient) are showing the amount of check-ins at the selected location and the dimension of the radius of each dot is showing the amount of users that checked-in there.



logical analyses (performed with GIS software) and the results of the remote sensing survey (performed by the analysis of user-generated data of Foursquare). In order to intuitively visualize the relations between the results of the morphological analysis and the geo-located information retrieved from location-based online Social Networks, the two types of information are superimposed on a single map displaying the results of such analyses on different layers (Figure 3). Such visual comparison allows for a quick and intuitive mapping of the different kinds of information displayed making possible to cross-reference the results of the morphological analysis with the actual data about human behavior in the urban system.

The process of gathering user-generated data provided by Foursquare database finally gives us the opportunity to dig into vast repertoire of infor-

mation about land uses, commercial activities and the relative popularity of such activities that can be referenced with the spatial configuration analysis performed by Space Syntax. This allows us to gain interesting insights about the correlation between spatial configuration and social patterns. What is interesting here is that commercial activities can be filtered by typology (e.g. Italian restaurant, Asian food market, etc.) thus allowing a more deep understanding of the way how certain types of activities tend to gather together over time and under what spatial conditions.

In order to intuitively map the different kinds of activities on the grid, we adopted a strategy for allowing quick visualization that employs a colour code for identifying each type of activity and the relative popularity of the same activity (Figure 4).

The user interface for allowing such visualiza-

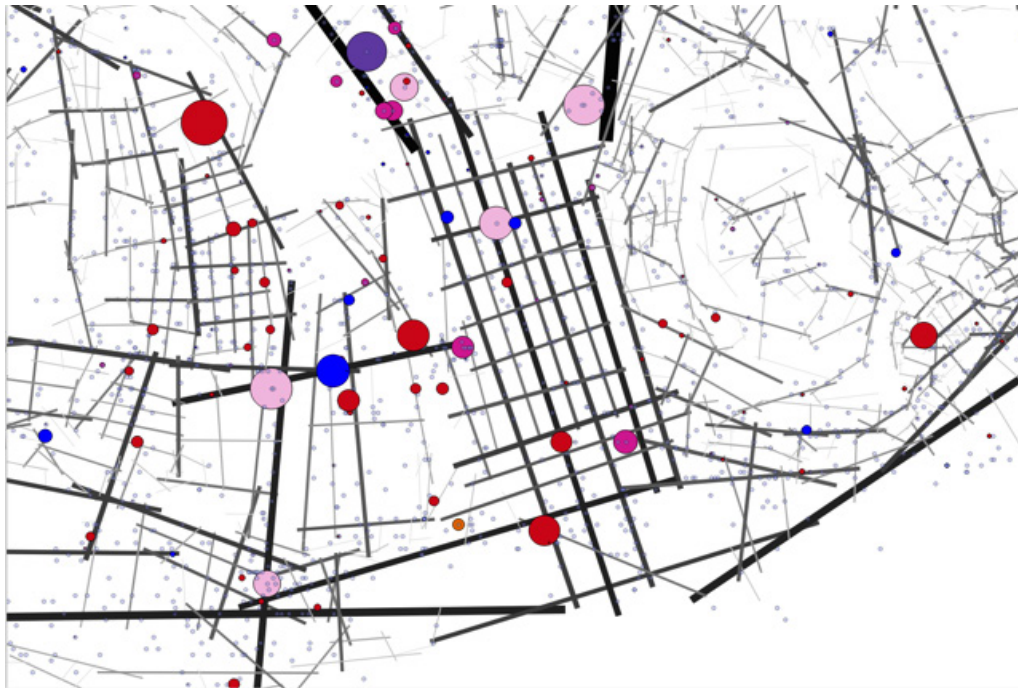


Figure 4
The colour of the dot is telling us to what category the dot belongs to, the size of the radius of the circle is expressing the amount of users doing checkin there.

tion is constituted by a web application capable to showing in real-time the possible correlations between spatial configurations and Foursquare data in a intuitive and dynamic way.

A full integration of the two sets of data – one from Space Syntax indicators' values, the other from Foursquare information – is currently under development, the rationale being that once the two sets of data are merge together a whole new type of mathematical and statistical operations with them will be possible.

At the present stage of the research we are working on normalizing the results returned by Space Syntax and Foursquare in order to allow a flawless analysis and comparison of data from both sources. To do so we are able to correlate the geo-referenced location of each check-in point with the road or street segment with closest relative distance. In this way we can assign check-in point to a road

segment: this again allows for the integration of the information gathered by Space Syntax with the ones returned by Foursquare analyses.

FUTURE WORK

The implementation of a methodology for cross-referencing the numerical results returned by Depthmap software with the data retrieved from location-based online Social Networks is currently under development. This will greatly enhance the potential of a detailed comparison of the results returned by the morphological analysis and the ones returned by the remote sensing process. Such comparison can constitute the basis for a novel urban analysis methodology that hopefully could be able of returning interesting insight about the relation between spatial configuration and amount of social networking activity.

Another interesting research direction would be employing the Place methodology (Stähle et al., 2006) for enhancing the capacity of the system to model actual relation between spatial configurations, geographic variables and amount of social activity in a chosen study area.

In fact Place Syntax, as developed by the research group of Spatial Analysis and Design at the Royal Institute of Technology in Stockholm, builds up on Space Syntax theory by adding to the traditional Space Syntax methodology reflections about geographic accessibility analysis performed in GIS – thus combining geographically oriented accessibility research with geometrically oriented research in architecture – making the comparison between Foursquare check-in geo-located points and spatial configuration analysis easier.

Finally another step towards the completion of the work would be achieved by adding analyses about geo-referenced data retrieved by Twitter database to the ones returned by Foursquare. Twitter in fact is an online social networking service and micro-blogging service that enables its users to send and read text-based messages of up to 140 characters, known as “tweets”. Since Twitter has released a

mobile version which can be installed on any mobile device such as smart phones or tablets, its “tweets” can be geo-referenced by the use of the GPS device embedded in the mobile device itself on any given position of Earth’s surface. Tweets, not being bound to the localization of a specific commercial activity or point of interest in the town, in fact can be sent from everywhere in the town’s area.

This would complement the strictly commercial Foursquare data that are only referred to points of interests, such as shops, restaurants and other socially relevant location with the addition of data in the form of “tweets” that map the movement in urban space in a more comprehensive way.

REFERENCES

Hillier, B, 1989, *The Social Logic of Space*, Cambridge University Press, Cambridge.

Stähle, A, Marcus, L, Karlström, A, 2006, “Place Syntax – Accessibility with axial lines, *Environment and Planning B: Planning and Design*.

[1] Foursquare, <http://www.foursquare.com> (last accessed: February, 2013)

