BUILDING KNOWLEDGE LAYERS AND NETWORKS FROM URBAN DIGITAL INFORMATION

Joaquín Borrego-Díaz[79]
Juan Galán-Páez[80]
Jaime de Miguel-Rodríguez[81]

Key words: Urban data, Knowledge Engineering, Collective Intelligence, Formal Concept Analysis

Abstract

The understanding and management of complex digital information on cities need the use of tools providing experts with new insights about the knowledge hidden within this great amount of data. In this paper a methodology to provide such a kind of knowledge is presented. This methodology is based on Formal Concept Analysis and allows visualizing abstract concepts that can be interpreted (and hence discovered) by city researchers.

1. Introduction

Nowadays Urban (eco)systems produce data that enables to manage, compare and share huge amounts of information about the city. According to its nature, urban data can be grouped in different sets of information layers, which as other information repositories and systems, are prone to problematic issues such as noise, inconsistencies, ambiguity, etc.

Citizens can access many of these layers through www, for their consult, enrichment, etc., and obtain useful information about their urban environment. However, in order to build understanding, it is well known that humans process information by means of (bounded) rational strategies and skills. One of the most important methods at stake in these events is qualitative reasoning (as well as the ability of extracting qualitative judgements from data). The simulation of these processes (a task of Artificial Intelligence) is essential for understanding how human minds behave upon available inputs. Consequently, this knowledge can be used to find more

[79] Departamento de Ciencias de la Computación e Inteligencia Artificial, E.T.S. Ingeniería Informática-Universidad de Sevilla, Avda. Reina Mercedes s.n. 41012-Sevilla Email: jborrego@us.es

[80] Departamento de Ciencias de la Computación e Inteligencia Artificial, E.T.S. Ingeniería Informática-Universidad de Sevilla, Avda. Reina Mercedes s.n. 41012-Sevilla Email: juangalan@us.es

[81] Departamento de Ciencias de la Computación e Inteligencia Artificial, E.T.S. Ingeniería Informática-Universidad de Sevilla, Avda. Reina Mercedes s.n. 41012-Sevilla Email: demiguel.jaime@gmail.com

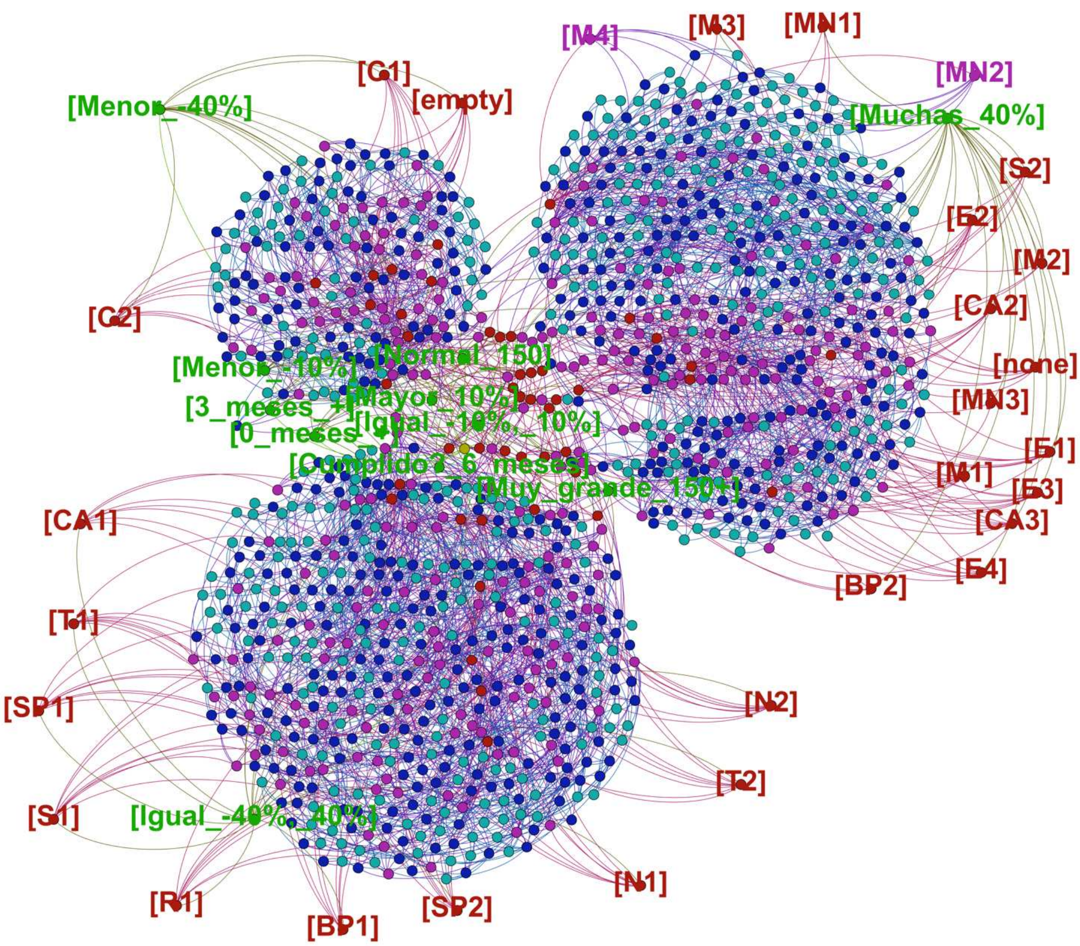
coherent strategies to present the information, and in addition, it can be targeted to attempt predictions over the interaction between users and data.

1.1 Aim of the paper

Every project should consider two middle/long-term goals. The first one is to understand special features of the urban facet under study in order to explain results. The second one is to obtain results respecting urban policies (Vaccari et al., 2009).
In this paper the authors propose a methodology to extract qualitative knowledge from urban digital information. The methodology is based on the scalable use of Formal Concept Analysis (FCA) tools and methods in order to retrieve abstract ontologies from data, according to a number of criteria. FCA provides abstract semantic structures associated to qualitative data, which can be (partially) interpreted, discovering new concepts and semantic relationships among them.

The results of its application consist of different semantic structures (knowledge networks) from data. The analysis and reasoning with this (complex) network (and with the associated Knowledge Bases) can be useful to obtain insight knowledge about: the city considered as a complex system, the evolution of the features represented by urban database, the semantic comparison between different urban dynamics, etc.

Figure 1 **– Concept Lattice built from digital real state data**



The methodology is illustrated by applying it on historical records of the real state market data in a city (Seville), which have been extracted from real state B2C and C2C WWW portals. The result is a broad view on housing mobility within the city by means of the discovery of new abstract concepts associated with it as well as an historical survey tool to explain some phenomena associated with this particular market. It also shows how the semantic structure differs between different Spanish cities, suggesting that there exist specific features about the dynamics of this market, strongly influenced by buyers and sellers. Preliminary results show a semantic structure, which relates spatial features with real state attributes (see Fig. 1). The concept lattice is represented as a complex network where spatial attributes are highlighted.

2. Urban Informational Ecosystem

The increasing availability of digital information about cities and their dynamics represents an enormous challenge for urban planners, data scientists and, in general, for the scientific community involved in the development of smart cities. Some problems they face are, among others, the human understanding of the potential knowledge that this information can provide, as well as the extraction and processing of inputs are particularly difficult to tackle due to the big amount of data. In fact, contemporary cities are complex systems with complex information (eco)systems. And the computational processing of this information is a hard task. However, in successful cases, it allows the understanding of features that are difficult to estimate otherwise. The recollection, processing and use of urban data applied to endeavours such as the design of new applications, urban services, urban planning and the analysis of the social dimension of the cities, is an emergent research & development line which is beneficed greatly from the current initiatives of open data.

There are two ways of considering digital information in the case of cities. On one hand, it is the view of the city as a *Data city*, that is, a place where data collection became the basis of the Geo-CyberInfrastructure (GCI). On the other hand there exists the vision of the city as an interactive system, which considers social networks (real and virtual) and their local interactions (between users, included in specific environments). In both cases it is essential to analyse the GCI, in which information of diverse nature is processed. The metabolic nature of digital information in the city, due to the spectacular penetration of social networks, is considered essential in today's cities.

The natural evolution of the GCI in the Smart Cities is directed towards a constellation of services for collection, access, display and analysis (which would allow the modelling and simulation of phenomena). However, nowadays, few cities offer such an opportunity, i.e., few cities have a GCI advanced and flexible enough to produce that kind of knowledge. Therefore,

we should use classic data and web services (accessibility variable), and data collected from the WWW. Both options provide information of varying quality. Clearly, these limitations do not offer a perfect digital image of the city, but the available information allows us to analyse some aspects of urban behaviour. As formerly mentioned, the image has been enhanced greatly thanks to open data initiatives in cities, which allow hatching of R + D initiatives around them.
From the point of view of Smart Cities, the quality of the available information is not laid out in such a weak manner: in pre-Smart Cities the information is available for consumption, while in Smart Cities this is available for the use and production of new data (applications, user communities, etc.), but in general we should not assume such features.

There exist four types of information flows to be considered in order to analyse how digital information can be used in simulation. Those four information flows types (Aranda-Corral et al., 2011) are: U2U (users to users), I2U (institutions to users), U2I (users to institutions) and I2I (institutions to institutions).

2.1 Institutions to Institutions (I2I)

The flow of information between institutions (e.g. cadastral changes, commercial tax collection areas, etc.) is often inaccessible to researchers and citizens, although open data flow initiatives make this flow visible and accessible for its study. An interesting feature of I2I is the need to solve the interoperability problem: How to (in a semantically sound manner) integrate several and heterogeneous structured data. Semantic technology can be applied in this case.

2.2 Institutions to users (citizens), I2U

This source is of vital importance when it comes to the understanding of certain processes that affect the liability, dynamics and evolution of urban environments. It provides historical analysis of urban phenomena that possess documentation. Moreover, these entities do not need to be solely formal institutions; social media also provide such information.

2.3 Users to institutions (U2I)

This type of flow is growing dramatically due to the adaptation of Web 2.0 technologies and philosophy at the Urban Informatics (Foth, 2009). Crowdsourcing (and crowdfunding) are novel philosophies that originate new forms of both, to collect information, and to help with urban features and dynamics.

2.4 Users to users (citizens), U2U

The flow U2U is essentially different, due to the heterogeneity of sources and interactions. It stresses the information generated by the mass use of mobile (devices) phones. They enable citizens to gather, as agents of the MAS, digital information and often leave fingerprints (digital footprints) on the Web 2.0. The quality of this information may even surpass I2U. In fact, P2P networks are promising urban information resources in Smart-cities, and their analysis can be helpful for smart planning activities. U2U flow can be used to add new behaviours to citizens-agents.

From a GCI context, metabolic information flows comprise several digital practices. When it considers production/consumption of data, the data is from both U2I and I2U. Social dynamics have to be considered as U2U networks. Lastly, institutions feed the system by means of I2U and I2I information, which hopefully is open data.

3. Formal Concept Analysis

The convergence of the Social Web towards the Knowledge Web depends on how we deal with semantic techniques of knowledge organization as tagging. A very useful tool to manage this convergence process is Formal Concept Analysis (FCA). According to R. Wille (Ganter et al., 1997), FCA mathematizes the philosophical understanding of a concept as a unit of thoughts composed of two parts: the extent and the intent. The extent covers all objects belonging to this concept, while the intent comprises all common attributes valid for all the objects under consideration. With this idea as starting point, FCA is a branch of applied mathematics whose goal is the discovery, extraction and organization of knowledge from qualitative data, by means of concept networks (hierarchies).

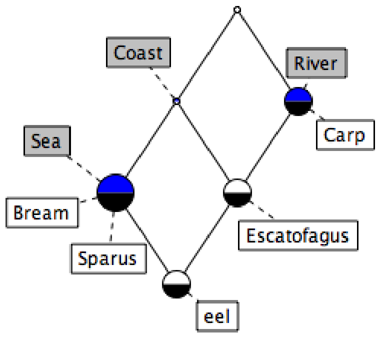
This theory provides methods to reason with the knowledge extracted. FCA can be considered as an approach to data mining with a semantic component (formal concepts) and a computational-logic component (association rules and automated reasoning). In this section, we succinctly present basic FCA elements.

The basic data structure is the *formal context* M = (O, A, I), which consists of two sets, O (objects) and A (attributes) and a relation between them. Finite contexts can be represented by a 0, 1 values table. See Fig. 2 (left) for an example of formal context about fish species with three attributes. Given an object set X (or an attribute set Y) the intension (or the extension of Y respectively) of X as X' = {a $\in$ A | oIa for every o $\in$ X} (respectively Y' = { o $\in$ O | oIa for every a $\in$ Y }).

The FCA main goal is the computation of every concept within the context. A (formal) concept is, following the former idea, a pair (X, Y) such that X'=Y and Y'=X. For example, the concept lattice with every concept from formal context about fish species is shown in Fig. 2 (right). In this concept hierarchy (concept lattice) each node represents a concept, and it can be considered as a bidirectional hierarchy. The extension can be read in top to bottom direction (from the most generic concept to the most specific) and analogously the intension in bottom to top direction. For instance the most specific concept (extension) in Fig. 2 (right) is ({*eel*}, { *Coast, Sea, River*}) which means euryhaline fish. The aim of theory is to discover new concepts hidden within the data.

Figure 2 **– Formal Context and associated Concept Lattice**

	River	Coast	Sea
Carp	X		
Escatofagus	X	X	
Bream		X	X
Sparus		X	X
eel	X	X	X



3.1 FCA and digital footprints

FCA can be used to organise knowledge and extract new concepts from rear data, with a number of advantages. On the one hand, digital information describing features of cities hide human behaviours and beliefs (in qualitative form). The extraction of knowledge from digital footprints aids to understand how citizens live and work within the city. To sum up, there exist hidden ontologies useful for (bounded) reasoning on cities and their structure.

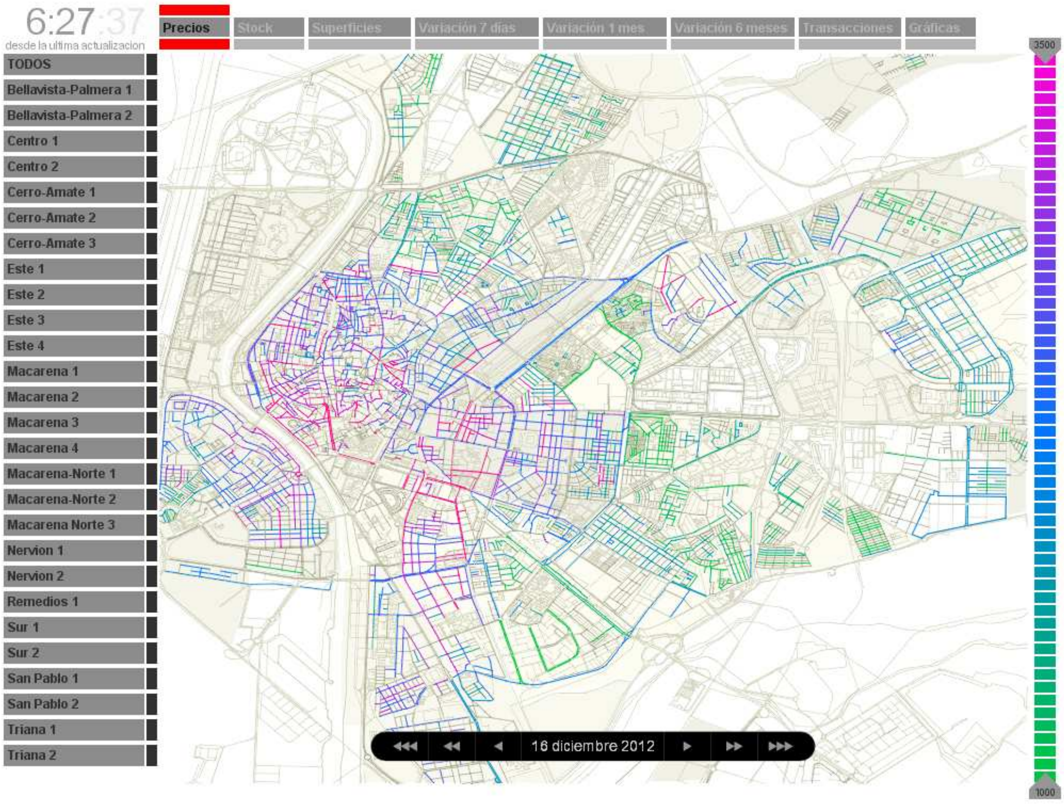
4. A case study: housing markets

Housing market values form a data set that plays a key role in the understanding of urban processes. They provide us not only with a meaningful vision of the socio-economic distribution of our cities, but also with more profound insights into the urban realm. For instance, when focusing on a certain neighbourhood, the data can be used to measure sale-price variation indexes upon which substantial knowledge can be gained regarding issues like social integration and diversity of urban areas. Moreover, urban processes of re-adjustment (i.e. gentrification) are often connected to changes in the economic value of real estate properties and vice versa. Therefore, the crossing of other layers of urban information such as infrastructural transformations, including transportation network, health, education and commercial centres and so on, can put together an analytical setup from which both the social and economic impact of these transformations can be evaluated. In addition, the reverse approach brings forward an interesting scenario; by monitoring and detecting abnormal trends within the values of the housing market, new opportunities raise to react upon urban processes that otherwise would remain hidden to the planner's eye.

4.1 SelfCity

City planning departments tend to be large and rigid institutions with highly bureaucratic mechanisms of operation, and therefore slow to absorb advances in the outside world. In contrast, SelfCity platform enables the incorporation of real time urban information in a meaningful way, striving to make dynamic data both efficient and fundamental to city management and decision-making. In order to achieve these targets, SelfCity involves a complex articulation of visualization techniques, crossing of planning data, clustering algorithms and most importantly, monitoring strategies in tune with the agency's needs. The right setup of these different apparatuses will determine to what extent will the information served be assimilated and thus, the chances of reaction upon data. Also, in terms of impact, the success of the research targets would mean a remarkable step forward in urban planning practices %at a national level in Spain; if not the first, at least one of the first significant implementations of real time urban information management platforms (see Fig. 3).

Figure 3 **– SelfCity system**



Finally, the system also provides an excellent opportunity for cooperation between citizens, public agencies and private corporations in the generation, sharing and management of meaningful urban data. In other words, housing sales-adverts are published by individual citizens within the framework of private corporations but are of interest of public agencies for the better management of the city; a better management that finally reverts in the common's best for all the parties at stake. In this context, the methodology will contribute to create awareness regarding the importance of data sharing and the positive value of synergetic attitudes involving individuals as well as private and public entities.

Figure 4 **– Streets selected in Seville**

4.2 Contexts from digital footprints

Contexts are built from the SelfCity information flow, as i.e. temporal and spatial information on homes for sale: Objects are homes for sale and attributes represent a qualitative description of the homes at a given moment; description of the item (i.e. price, dimensions, environment, etc.) and description of the item evolution (i.e. price changes, environment evolution, etc.). From this context, the concept lattice is computed. The object set is a Collection of 6000 (approx.) for sale homes in the city of Seville. The global concept lattice with all the info aggregated is depicted in Fig. 1. Also subareas of Seville can be considered (i.e. streets, zones, etc.) for a more detailed analysis. For example, in Fig. 4 two loosely related streets from Seville are selected, and they are compared by means of the associated concept lattices (Fig. 5). A more precise analysis of conceptual differences can be provided by means of logical reasoning with association rules associated to the lattice. Also, it is possible to consider how qualitative patterns are reproduced in the city.

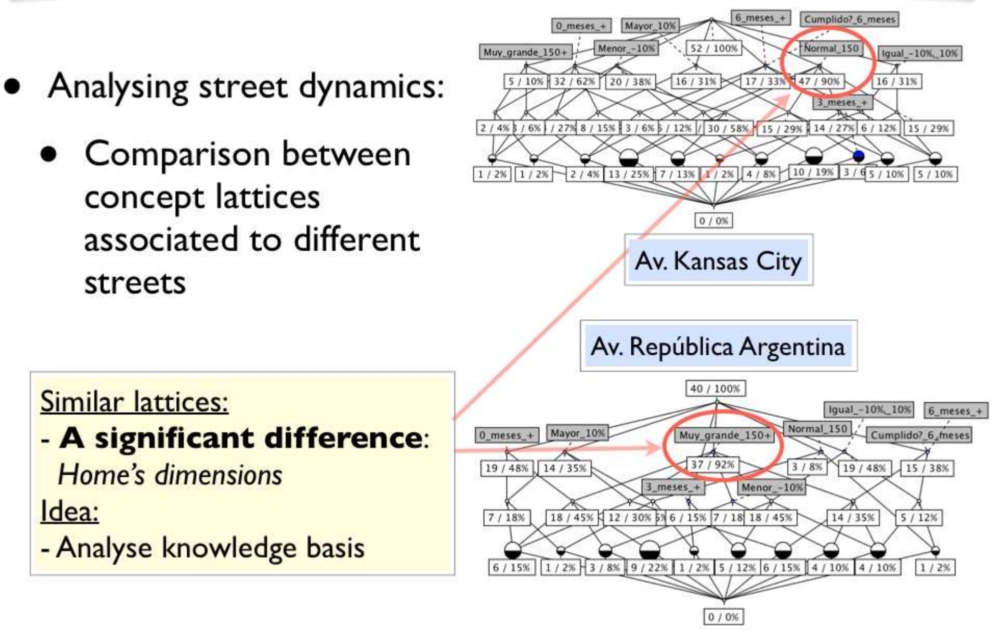
5. Conclusions and Future Work

Currently, SelfCity manages only digital information. Such information can be visualized, but a sound comprehension of its dynamics and conceptual structure requires a semantic processing. The understanding of the full dataset needs a semantic interpretation of the concepts involved in this complex system.

The methodology above described, which was applied in this case, provides a rich semantic structure (Fig. 1 shows the concept lattice). The study of this structure allows:

- To discover new concepts, that are relevant for housing markets
- To compare conceptual structures of different cities, as well as their dynamics
- To compute concept lattices for different districts and compare them by means of FCA-based similarity metrics
- To compute concept lattices for different dates and analyse their evolution
- To obtain knowledge bases useful to reason on the housing market

Figure 5 **– Comparing two streets**



Bibliography

VACCARI, A., **LIU**, L., **BIDERMAN**, A., **PEREIRA**, F.C., **OLIVERINHA**, J., **GERBER**, A., and **RATTI**, C. A holistic framework for the study of urban traces and the profiling of urban processes and dynamics, in Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems (ITSC2009), IEEE Press 273-278. Accessible in http://senseable.mit.edu/papers/pdf/2009_Vaccari_et_al_Holistic_framework_ITSC.pdf

ARANDA-CORRAL, Gonzalo, **BLANCO-ESCUDERO** Alejandro, **BORREGO-DÍAZ**, Joaquín, **GOMAR-ACOSTA** Manuel, *Data-in-the-cloud City,* in Proceedings of the 7th International Conference on Virtual Cities and Territories, 2011 pp. 565-571

FOTH, Marcus (Ed.) Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City. Information Science Reference, IGI Global, Hershey, Pa. 2009. ISBN: 978-1-60566-152-0

Ganter, B. and **Wille**, R. Formal Concept Analysis - Mathematical Foundations, Springer, Berlin, Heidelberg, 1999. ISBN: 978-3-540-62771-5