



## Social Media Analysis about City

---

Domenico Monaco

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 25, 2019

# Social Media Analysis about City

PhD proposal

Domenico Monaco

d.monaco2@campus.unimib.it

April 25, 2019

## 1 Introduction

People, Technologies and Places are in every moments more closely linked ([1]). Today the vast majority of people is living in a cities ([2]) and the world is characterized by presence of millions mobile connected devices<sup>1</sup>.

There are many challenges to be resolved related to increasing urbanization and theses can not ignore the new technologies. In particular all human's life aspects are progressively digitized into Social Media Data: created and used in Urban context that we can extract to understand how people use and perceive the city through their Social Media content ([3]).

This research proposal provide a general overview of the social and economic dimensions of the urban economy through the Urban Informatics, Social Media Analysis and Machine Learning techniques.

## 2 Context

Urban Informatics is wide set of methodologies, techniques and applications of Information technologies applied to urban aspects, as Batty define it in [3] "*Urban Informatics is loosely defined as the application of computers to the functioning of cities*".

One of activities of the Urban Informatics is related to data collection, analysis and using of geolocated informations into differents applications. But this in most case is in purely statistical way, such as considering only punctual information on points density: number of photos in a place, number of check-ins and/or which tags have been mostly used.

But this is not enough because you get a rigid and partial picture of the city.

**Geolocated Social Media Data** Increasing amount of connected device and the diffusion of using of Social Media feeds huge flow of data about city, its using, its space, its qualities, its issues ([3]).

People, voluntarily or involuntarily, tell the city through social media. They express opinions, comment and tell their lives through social media. So, they potentially offer a glimpse of the city through their eyes: essentially they tell us how they perceive it, how they live it, and use it or how imagine it.

I refers to this type of information about city as *city-related informations* why these are not only geolocalized, but can refer to the city, to its spatially identifiable place or it can express spatial facts in differents ways from geolocation.

---

<sup>1</sup>The internet of things: mapping the value beyond hype, McKinsey Global Institute, June 2015. [www.mckinsey.com](http://www.mckinsey.com)

**Tales of the Places** There are many social media related to places of interest and geolocalized activities, they offer a pool of content and descriptions of places without precedent. People often not only comment on the quality of a service, but contextualizing it by telling, for example, why they were there, how they got there or where they were going, or how they used a place habitually. They offer us a clear and detailed picture of how they perceive services, spaces and places in the city and they often do so with unstructured information.

The challenge is to extract information about places from largely unstructured data in order to create a sort of litmus paper of places by starting from the bottom, from the information that the users themselves give us, going beyond the "what" and "where", but rather by extracting complex information and spatially contextualizing it.

## 3 Project

The objective of this study is to provide a general overview of the social and economic dimensions of the urban economy, using a technical in-depth analysis by Data Analysts and Machine Learning, as well as the Clustering of data and the processing of linked and natural data.

It will be informative for the description of the city, its causes, its spaces; bringing out the city from the point of view of who is through current Social Media by connecting complex, unstructured information with more complex information to.

### 3.1 Research activities

At the current state, the starting point is: *What activities could we make to improve the description of the city, its spaces, its dynamics, from the point of view of citizens through Social Media?*

**Object of photo** In several searches, geolocated photographs are already analyzed ([4]), but in a rather superficial way, considering almost only structured information such as timestamp or geolocation information. More could be achieved by interpreting the content of the photo through the recognition of objects, understanding if it is, for example, monuments, parks and dishes. This could provide information on the interest of that place or how it is used.

**Spatial discourses** The idea is to bring out *Spatial Discourses* from the reviews, a bit like the "sentiment", such as the *spatial/geographical context* that the text expresses.

*Spatial Discourses* are one or more spatial informations or relations extracted from a text, in theory is extractable from any text, but we take advantage by using the reviews of a social media like Trip Advisor, Yelp or Google Place puts us in first condition having an initial Reference Geolocation, we already know from the beginning what area a reference is referring to, secondly if it is possible to extract additional spatial information, this information would be related to the initially extracted position. Or more we can have some location to verify what we extracted from a text.

In addition it is possible that there are more information of this type in the form of text in social media dedicated to places rather than on a social media dedicated to other.

**Information integration** Another possible application derives from the fact that often geolocalized POIs on the various Social Media may not be the same between one and the other, this implies the need to merge between the different platforms in order to create a full image of the city, but this implies the need to characterize and/or identify the places with a unique and shared taxonomy, also in optics to integrate our data to Governative Open Data Datasets<sup>1</sup>.

Often the meta-information on POIs is not sufficient to characterize a place in a comprehensive way: they may be missing or of poor quality. It is often possible to identify a POI through a number

---

<sup>1</sup><https://www.dati.gov.it/>

of related information such as the name, a photo, the website etc. that are simple to interpret for a human, more complex for an algorithm.

It would be interesting to use all this related complex information to improve the characterization of places where necessary.

## 3.2 Techniques

*What are the techniques that will potentially be used to integrate geolocalized complex informations such as photos and text?*

In this context it is essential to use a mix of techniques such as the Spatial Data Mining and Clustering, Natural Language Processing, Linked Data and Data Visualization.

Spatial Data Mining is important part of analysis because it is one of the most popular techniques to bring out spatial relationships on geolocalized data. But this does not tell us everything, but expresses only one type of relationship, as a concentration of a certain type of activity on the territory.

They do not express how they are used or for what reason, in this case the techniques of "Natural Language Processing" and "Image Recognition" come to help extracting information from unstructured data, which represent the vast majority of data on social media.

With regards to the extraction of information and the organization of information, it is certainly useful to use Linked Data which offers a method of mapping the previously extracted concepts automatically.

Finally, visualization, an important component essential to representation, understanding and communication of concepts. Without a doubt, various visualization techniques will be useful both to the understanding itself and to the communication of potentially complex and heterogeneous spatial concepts.

Mixing together these techniques to extract spatial detail from data can offer an improving the understanding of city.

## 3.3 Sources

Previously we talked about city-related information created by users through Social Media. So, *What are this social media with where we found city-related information?*

**Social Media** Many Social Media allow the sharing of geolocated information, but some have a stronger orientation to geolocation. And it is on these that I want to focus.

The first to be mentioned are certainly the most classic GIS such as Google Maps and HERE, which represent widely used and reliable Web Geographic Information System that offer a general overview of the POIs in the cities. In some cases even with some rich related information like photos, descriptions and reviews. These are a good starting point.

Another possible category is represented by the Social Media for sport activities. Sport is an important theme in the urban context, for this reason it is useful to consider some social media like Garmin Connect or Strava that offer an accurate view of the most frequented areas for sport activities.

Facebook is not specifically oriented to geolocation, but with its functionalities of Check-in and participation in geolocated Events can give an alternative view of places such as those related to big events.

How not to mention Instagram, although not specifically oriented to geolocation, still offers the possibility of extracting alternative information related to the use of the city through the huge amount of photography often also the apparently most insignificant.

Finally, perhaps the most interesting for this research, we find those oriented to tourism, which I personally find extremely interesting for a series of reasons related to their way of being used and the

strong tendency to do-it-yourself in the modern tourism. Among these, we find Yelp and TripAdvisor that offer a detailed mapping of the commercial activities related to tourism with related reviews sometimes exhaustive and really accurate that even on the context of use of a place or part of city.

## 4 Applications

This research can be useful to all city stakeholders, users and actors of city in different applications like: Society, Governance & Government, Economics & Finance, Inequality, Exclusion, Environment & Sustainability, Service, Quality of Life ([3]), Education, Safety and Transportation, Urban Planning and urban Regeneration.

Can be used to improve planning, verify or observing the activities to predict or monitor.

**Walkability** For example, the Walkability is an interesting concept used to evaluate the quality of an area for pedestrian use through a set of factors usually related to the perception that each of us has of a place such as it must be useful, safe, comfortable, and interesting in a same time.

These assessments they have often been made through manual surveys or other repetitive activities. And they are all personal and relative. As quercia2015digital expose in his research [4], this activities can be improved with Social Media data. In same, way but in more deep, I want to link information to Walkability concept by Social Media analysis, through alternative information such as place's reviews, generally unconsidered in a researches for this scope.

**Partecipatory Urban Planning** Another application is to provide a contribute to new partecipatory Urban Planning like Milano 2030, a series of public events around key topics about urban regeneration, centrality and public space at the support to the P.G.T. (Piano di Governo del Territorio <sup>1</sup>) which aims to involve different actors of the city in its evolution over the next 10 years, regarding mobility, environment, housing, sociality, inclusion and enhancement of public space and in general relating to urban planning and urban regeneration that will affect Milan in the coming years.

**Application Questions that we can answer** With regard to previous applications, as example, we can formulate some questions that we can try to answer through this research project.

– How much are separated periphery and center? What are the objects of separation or merging of urban areas?

– Are shopping mall, like City Life in Milan, the new squares? What is the effect of developing traditional squares in opposition of modern shopping mall?

– Are food activities the boost of pedestrian movement in a city? And the big events?

– A park, is really used for jogging or is an area considered too dangerous? Are there photos taken in this park of people in sportswear? Picnic photos?... and dogs?

– For example the Affori metro station is being used in large part by people coming from outside Milan to reach the center, there are sufficient parking spaces?

**ICSCU 2019, Jan 21-22, 2019** 21st International Conference on Smart Cities and Urban. <https://waset.org/conference/2019/01/london/ICSCU/home>

**BDASE 2018, Dec 17-20** International Workshop on Big Data Analytics for Sustainable Environments will be held in conjunction with the 11th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2018) <https://hud-cs-research.github.io/bdase2018/>

---

<sup>1</sup>[http://www.comune.milano.it/wps/portal/ist/it/servizi/territorio/pgt\\_citta\\_metropolitana](http://www.comune.milano.it/wps/portal/ist/it/servizi/territorio/pgt_citta_metropolitana)

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Context</b>	<b>1</b>
<b>3</b>	<b>Project</b>	<b>2</b>
3.1	Research activities . . . . .	2
3.2	Techniques . . . . .	3
3.3	Sources . . . . .	3
<b>4</b>	<b>Applications</b>	<b>4</b>

# References

- [1] Franco Zambonelli. Engineering self-organizing urban superorganisms. *Engineering Applications of Artificial Intelligence*, 41:325–332, 2015.
- [2] Guido Martinotti. *Sei lezioni sulla città*. Feltrinelli, 2017. ISBN 9788858830130.
- [3] Michael Batty. Urban informatics and big data. *ESRC (The UK Economic and Social Research Council) Cities Expert Group*, 2013.
- [4] Daniele Quercia, Luca Maria Aiello, Rossano Schifanella, and Adam Davies. The digital life of walkable streets. In *Proceedings of the 24th International Conference on World Wide Web*, pages 875–884. International World Wide Web Conferences Steering Committee, 2015.