

## MAKING SENSE OF BIG DATA – PRESENTATION OF THE “MARSEILLE SPEAKS, IN & OUT, ON TWITTER” PROJECT

AGATA LUDZIS-TODOROV

Culture and Communication Department, University of Social Sciences  
and Humanities, Chodakowska Street 19/31, Warsaw, Poland

[agata.todorov@gmail.com](mailto:agata.todorov@gmail.com)

LUDMILA GIRARDI

Nijmegen School of Management, Radboud University,  
Thomas van Aquinostraat 5, Nijmegen, The Netherlands

E-mail Address: [ludgirardi@gmail.com](mailto:ludgirardi@gmail.com)



### ABSTRACT

The present article seeks to present the conceptual and procedural findings of a brief sample research being part of a wider project called “How do cities speak about themselves on Twitter?”, guided during the Digital Methods Summer School at the University of Amsterdam in 2014. Our research project, called “Marseille speaks in & out”, contains an exploratory analysis of Twitter data and uses a cartography approach as an empirical method towards mapping comparisons. The objective of this methodology is to overlay the city’s scales in terms of different perspectives of the collected data: languages, locations and places in Marseille at an assumed period of time.

**Keywords:** Twitter, city studies, digital methods, mapping, Marseille.

### INTRODUCTION - THE INTERNET AS A RELATIONAL SPACE

We understand the internet as a vast space of topological metrics where actors can relate from places all over the world, turning it a social space pertinent on a Global scale (Lévy, 2003; 2008). The relations on the internet produce other types of spaces and they form “places of reticular metrics” (Beaude, 2012). The terrain (the physical space) and the territories, or cities’ administrative limits, are no longer essential to understand a society, considering this approach. We can therefore identify where cities relate to each other in this “new kind” of place which has other boundaries, limits and hierarchies. In this sense, we assume that the internet rapport a societal problem: living within this almost unlimited connectivity, cities’ spaces are now “hybrid” (Latour, 1993), where territories are more dense and more capable of embracing the complementarity of this common global space.

Networks have the same significance as territories for studying a place under a spatial dimension, even more in the context of urban and city studies since they have the tradition to rank and to compare hierarchies and centralities, postulating the relevance of studying topological spaces' metrics. Because of the emergence of this new "huge scale", Lévy (2008) understands places to be now comparable to others places in the world in terms of distance - to define places' proximities in other metrics than the Euclidean system - and in terms of scale - to define the limits or regions by the phenomena's substance. In the context of the present study, we ask the question which are the links that connect people and, therefore, the places where a city is being designed by smartphone users' relations beyond borders.

Within this framework, a city is part of a "complete society" (Lévy, 2003), by which we tried to identify what were the connections "within" and "without" of the French city of Marseille, considering places' mentions in one of the main social media on the internet, Twitter, at a specified space-time period (June, 2014). We use the application DMI-TCAT, through which researchers from the European Observation Network, Territorial Development and Cohesion (ESPON) had already extracted a huge dataset of 5 European Cities (Amsterdam, Edinburgh, Marseille, Bologna, Brussels) to study the topics associated with these cities and the specific location of users speaking about the cities and their places. We chose Marseille and looked for its network of places at a local scale to associate it with a global perspective of the city, considering the wider frontiers of a place and its global pertinence. Marseille can be considered a smart city because its inhabitants and visitors are "smart users" tweeting their location and experiences, but also creating a myriad of links, networking places inside and outside the city limits, and making sense of the city through the media.

### MAKING SENSE OF BIG DATA

As a huge, vast and unlimited space, the internet addresses empirical problems due to the amount of data available. Social media users exhibit and archive physical experiences to places to identity markers and employ the "spatial self" as a way to communicate where they are/were, or what they are/were doing, as well as who they are (Schwartz, & Halegoua, 2014). This type of check-in selection demonstrates the way social media users coordinate and incorporate the places displayed in their profile with other presentations of the self. The authors encourage social media researchers "to think about place and space as strategically chosen markers of identity" (Schwartz, & Halegoua, 2014). How can one filter all that information to grasp the details?

On Twitter, for example, hashtags (#) are essential to organize information, making it very relevant for researchers to consider when analysing digital data, because they aggregate content in a semantic way, therefore it is possible to study the profile of individual social media users based on their physical activity. Also, "content curation" is a trend evident in places and cities studies because it serves to filter information (Wilken, 2014). The profession of information cura-

tor is naturally becoming increasingly necessary to organize the flood of data we face every day, a work source of social evidence that can be applied to research and knowledge production.

The curator role is a challenge for communicators: this abundance of information is likely to continue growing, and dealing with it is a great challenge. Google, for example, can only scan 5% of the information available on the Internet, but it is the emerging tools, such as customized search engines, that reach other audiences and diversify the information availability. There are three online curatorial models: the curator as strainer, as an agent and as a curatorial platform device, according to Beiguelman (2011): the first is a conservative and individualistic model (“I am what I link”), the second has the role to create a point of tension that causes others to produce favourable mechanisms to tools development and the third considers information as “things are as you link”. This is an information program of curated feeds being fed by a so claimed “distributed intelligence” across the network, which is based on the knowledge society. The distributed intelligence is not measured by the amount of likes and followers, contrary to what may be the most common way people think (Beiguelman, 2011).

## METHODOLOGY AND RESULTS

We tried to answer the following research questions: What Marseille places do twitter users speak about? And from where do users speak about Marseille on Twitter, in what languages? We worked with two different data collections extracted in a short period related to the main ESPON project, from 23rd to 29th of June 2014. Firstly, we extracted, all the hashtags from the full dataset available, which considers all mentions of Marseille in all places and all European languages, to identify which one was the most frequent, as well as to select among them the occurrence of places. Secondly, we extracted the hashtags from a geolocation database which comprises all tweets generated within Marseille limits. Briefly, the methodological procedures were as follows:

1. Hashtags analysis (digital curation):
  - a. #marseillemlife: top hashtag;
  - b. #places: Marseille’s correlation with places in different scales considering the frequency;
2. Full datasets filtering:
  - a. Keyword database: organized by user specified language;
  - b. Geolocation database: organized by user location and classified in different scales (filters sequence: exclusives registers, categories of places).

We conducted an exploratory analysis of Twitter data using a Cartographic approach as an empirical method, specifically to organize a database towards mapping comparison. According to Lévy (2008), the visual resonance of maps with each other is a cartographic means that considers the unity of the world. The objective with this empirical method was to overlay scales in terms of types of languages, location and places’ correlation of Marseille’s mentions on

Twitter. Additionally, we followed the analysis with more specific questions: What are the places correlated to Marseille in Twitter conversations? How strong are their connections, considering the frequency of mentions? What are the places and languages speaking about Marseille on a global and on a local scale? What are the scales of users tweeting from Marseille according to their specified location?

Firstly, we analysed the top hashtag #marseillemlife, which is both a username on Twitter and a local blog on Tumblr that curates geotagged Instagram pictures, co-created by many authors. The whole creation of the blog and the hashtag is based on intermedia transfer and confronts the multi-level use of links to design and multiply the content. Places are the common point of every post and users activity in the blog and in using # because they are all geotagged feed, so we can identify all the places in the city and in the surroundings areas that are being socially constructed in the web. The tweets' texts containing the #marseillemlife indicate a close relationship between the users and the local environment. The visualized discourse shows numerous references to characteristic places in the city, like "Cassis", a coast city frequented by locals and tourists. The users' discourses on #marseillemlife mainly reproduce places of leisure that may be touristic areas. It is the good life in the city being talked about: "love", "summer", "sun", both in English and in French: "amour", "soleil", "vie".

Additionally, we proceed to the analysis of all hashtags containing the name of the places in Marseille mentioned in tweets, which turn out to be specific places inside the city at an inter-urban scale, including the hashtags for each quartier, but the most frequent mentions are for the more central ones. Also, the hashtags emphasize Marseille's regional location in the Provence region (Provence, Vitrolles, & Pennes-Mirabeau), addressing an indication of the places' connectivity. There are other French cities with hashtags, especially those close to Marseille in terms of cultural relations and physical proximity.

Considering this overall analysis, we built four maps to compare them with one another, two focusing on the local scale and the other two considering the global scale. We built them in the CartoDB platform, which is an open source solution that includes Geographical Information System to store and manipulate data. From the functional point of view, the tool is simple and accessible, it offers a user-friendly and comprehensive online editor, but also provides more interfaces that enable a developer to go a little further, without having to specialize in manipulation of geographic information. The outcome maps are available online, as their links follow with general comments:

Map 1. The sense of the local (#marseillemlife) <http://cdb.io/1rqrCxn>

- a. The relevance of communication in places of circulation: port, airport, roads;
- b. The relevance of communication clustered in the city center (the port is top location to tweet) and leisure places like beaches (mainly Cassis) indicate a strong link to tourism;

Map 2. Marseille world places relations: #places by frequency of mentions <http://cdb.io/1q29u9e>

- a. Brazil's outstanding situation is visible probably due to the World Cup, a global debate at the moment of data extraction;
- b. Although there are not many mentions in the USA/Canada, it seems to have a relevant connection with Marseille;
- c. Physical proximity matters a lot (Western and Central Europe), also the substance of historical background, considering the connection with North African countries which were French colonies, especially Algeria which has more mentions to Marseille than the Capital Paris. Algeria and Marseille have also important port relations.

Map 3. World speaking about Marseille (by user-specified language) <http://cdb.io/1q29zdg>

- a. On every continent there were users speaking about Marseille mainly in their own territory-based language except for the globally overspread pattern of English speakers;
- b. Globally, there is also the use of a place-based language around the world (French in this case), maybe France-based people traveling the world;
- c. In the local-regional scale, besides French and English speakers, there are only Dutch users tweeting.

Map 4. Scales of user(s)-specified location <http://cdb.io/1q29AxD>

- a. The tweets are not only localized within Marseille's city limits, but mainly within Provence, a local scale from a global perspective;
- b. Looking closer, it's possible to differentiate French users from the other users. There's a concentration of more outsiders than the locals in the airport and touristic areas, especially in "La Ciota", where there's a predominance of world users;
- c. In the city center there is a miscellany of users, which weakens in the cities' suburbs where there is a predominance of local users;
- d. After filtering the geodatabase, local users' conversations differ from the world topic conversations (people mainly talking about the World Cup) because they don't have a specific topic and use common words in French.

## CONCLUSIONS

Hashtags give the sense of the local as well as the other ways of filtered digital content. Marseille has hashtags for each of the city's zones and the people report to specific places in Marseille, not the city, therefore it is easier to identify in the hashtags the users and the conversations concerning local themes, more than global topics and subjects. This data is more appropriate for event occurrences. Because of this, it is mandatory to select a precise range or time period that mostly explains the issue, for content analysis to bring more relevant findings. To differentiate the locals from the outsiders, the location and the language combined give a fair indication. It is not easy to identify a user identity based on this data, so it would be better to filter the data in the sense of using the most

precise chunks of it and working with clean samples, taking off the “noises” of big data.

The geotagged content production concentrates in the city centre and the circulation places, also affirming the importance of fixed internet access. This concentration also affirms that availability of time does matter(s): to produce content, one needs time to do it, so people generate tweets in spare time e.g. during their trips. Citing Anna Maj’s expression, we can call such travellers “high-tech travellers” (Maj, 2010).

In isolation, the maps do not explain great things, but combined they brought a fair understanding of what the city of Marseille is on Twitter, locally and globally. Mapping twitter data can be useful for communication plans and city management. This sample project intends to bring highlight to the fact that nowadays one must consider the global scale of cities to understand the local arrangements for social and spatial planning. Although in very general terms, we demonstrate that web data can help re-scaling cities’ limits, through a worldwide significance appointed in a single social media. Marseille connects itself to the world through Twitter using different substances and specific parts of its space, redefining internal frontiers and reinforcing boundaries.

## REFERENCES

- Beaude, B. (2012). *Internet, changer l’espace, changer la société. Les logiques contemporaines de synchronisation*. France: FYP éditions.
- Beiguelman, G. (2011). Curadoria de informação. Conferece *Encontros com o Futuro*, by Commais, at ECA-USP, 21/06/2011. Retrived from <http://pt.slideshare.net/gbeiguelman/curadoria-informacao>.
- Latour, B. (1993). *We have never been modern*. Harvard University Press.
- Lévy, J. Quels espaces pour la société-monde? *Le Monde.fr*. Retrived from [http://www.lemonde.fr/savoirs-et-connaissances/article/2003/06/25/jacques-levy-quels-espaces-pour-la-societe-monde\\_325363\\_3328.html](http://www.lemonde.fr/savoirs-et-connaissances/article/2003/06/25/jacques-levy-quels-espaces-pour-la-societe-monde_325363_3328.html)].
- Lévy, J. (2008). *L’invention du Monde: une Géographie de la mondialisation*. Paris: Sciences Po Les Presses.
- Maj, A. (2010). Digital Memories of High-Tech Tourists and Travelling Media. *Twittering and Globalhood. Popular Culture*, 3-4, 29-30.
- Schwartz, R., Haleboua, G.R. (2014). The spatial self: Location-based identity performance on social media. *New Media & Society*. DOI: 10.1177/1461444814531364. Retrived from <http://nms.sagepub.com/content/early/2014/04/09/1461444814531364.abstract>
- Wilken, R. (2014). Twitter and Geographical Location. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter & Society*, (pp155-168). New York: Peter Lang.