

Audio paper 6 – MEDIATING SUSTAINABLE CITIES: quantitative analysis of social media data

Welcome to the series of audio papers: Mediating Sustainable Cities!

I am Paola Monachesi, researcher at Utrecht University. In the previous audio paper, I have discussed how to identify specific communities in Twitter and how to extract their data. In this one, I will focus on a quantitative analysis of the data collected. Special attention will be dedicated to how skilled creative migrants and elderly contribute to the smart city discourse through their communication.

I propose an innovative methodology that combines a language analysis of the (geo-tagged) Twitter data with a quantitative investigation of the communication patterns through social network analysis. While the language analysis reveals the interests and activities of the users, the social network analysis provides a geographical visualization of their communicative interactions. This methodology can make a relevant contribution to the use of Twitter geodata in the analysis of communication for urban studies research.

Jingle

What kind of language analysis can be carried out on the basis of the collected data?

The language analysis I have proposed relies on computational methods and focusses on an analysis of hashtags to investigate the topics of discussion of the users. It relies on word frequency. In this case the frequency of the hashtags employed by the users.

The reason why I have decided to use hashtags instead of normal words is that hashtags are added by users to their tweets. They are a way to add metadata to shared content and highlight in this way the debate that is being carried out on the platform.

Hashtags make tweets more visible through the search functionality and it becomes possible to communicate with a specific community around the hashtag topics. Through hashtags people can follow topics that interest them in an easy way. A frequency analysis based on hashtags reveals the topics that social media users want to highlight and provide us with information about what interests them.

In the case of the creative migrants, I have analysed the four groups of creatives present in our sample. The categories are 'writers', 'new media arts', 'fine arts', and 'architects and designers'. The aim was to investigate whether there are differences in the interests of the various categories of creative skilled migrants. A visualization of hashtag's frequency has been provided through word clouds and can be visualized in the published paper. You might remember that in audio paper 2 and 3, I discussed the results and implications of this analysis for a human-centred smart city.

Jingle

The language analysis can be connected to social network analysis in order to provide insights on the geographical distribution of the various patterns of communication possible in Twitter. We analyse the followers, reply and mention networks of our users.

I will focus first on how social network analysis works in the case of creative migrants and then I will discuss what we can learn from these visualizations.

In the case of creative migrants, I have examined the nature of the relationships between users. I have quantified and categorized the users' connections in four dimensions: creative vs. non-creative and local vs. global users. The latter, that is, local and global users, have been calculated on the basis of the geo-location provided. In this way, it is possible to investigate with whom creative skilled migrants interact and where these users are located.

For each user, we have extracted his network and then aggregated the networks of all users. The results are visualized by means of three maps representing the locations of the users. The maps visualize the followers, reply and mention network of the users. They were created by gathering the latitude and longitude of the users' locations and mapping them onto a world map using the software Gephi. They can be visualized in the published paper linked to this audio paper.

So, now that we have these maps, what do they tell us about creative migrants?

Well, we can figure out whether creative migrants are connected to more non-creative users than creative ones, for example. In this way, one can investigate whether they use the platform for professional or for private use. It is also possible to assess whether creative migrants connect more within the Netherlands or internationally. A visualization of followers, mention and reply networks also allows to analyse the level of engagement of these users within their network.

I think you can easily see what the advantages of integrating language analysis with social network analysis are.

This innovative approach makes possible to discover the interests of the community through language analysis while social network analysis shows the way discourse emerges through the local and global communication networks of the users.

Social media play an important function in this context since they give rise to new spaces. They contribute to the creation of a digital urban imaginary that complements the physical one, as discussed in audio paper 1. Social media foster the development of networks of innovation that rely on cities as relevant hubs, instead of the nation-state.

Jingle

One might wonder whether we can use this methodology to analyze other groups of users.

The methodology proposed here adapts nicely also to the other community considered in the previous audio paper, that is the elderly. This provides evidence for its validity.

Recall that three age groups have been identified, in the case of the elderly. I have carried out a frequency analysis of hashtags and words employed by these three age groups. The word frequency analysis allows for an investigation of language use that can reveal differences in terms of behavior and interests. This is especially the case if hashtags are employed since they highlight the topics being addressed, as I have mentioned previously.

In the case of the elderly, I have extended the methodology by combining the hashtag frequency analysis with the geo-location information available in Twitter. This approach allows to carry out interesting investigations on the spatial distribution of users that are particularly relevant in the case of urban sustainability. It provides an example of how digital sustainability and environmental sustainability are related. The data provided by the users make possible to assess whether certain age

groups have an interest for environmental sustainability in connection with the place they live and can play a role in stimulating an alternative discourse.

Another question that might arise here is how we can we exploit the geo-location information in this analysis.

As you know, Twitter profiles might include geolocation information. If it is present, it becomes possible to assess the distribution of the living places of the various users and to visualize it through maps. The maps can be created with the software Tableau that exploits geographical coordinates to represent the locations in maps.

By analysing the locations that are mentioned in hashtags two possibilities arise:

first, we can investigate the interest of users for specific places and second, we can analyze whether there is a correspondence between the places of residence and the locations users mention in their hashtags.

While a quantitative analysis can reveal patters in this respect, a qualitative analysis might provide deeper insights into the data revealing in which context these locations are employed, for example whether they are employed in relation to professional or private life.

I will tell you more about ways to combine quantitative and qualitative analysis in the next audio paper. For now, it is interesting to see what quantitative analysis alone might reveal about our data.

Our analysis shows that location hashtags are used mainly by people living in the place that is being mentioned or close to it. It also reveals that, elderly above 67, make the most use of location hashtags, especially locations within The Netherlands. They seem to show an interest for the community they live in.

Furthermore, a frequency analysis of hashtags related to nature and to sustainable development reveals that the elderly understand sustainable development, mainly as environmental sustainability. This is different from the other age groups.

If properly involved they can act as social innovators in broadcasting an alternative discourse on sustainable development, as discussed in more details in audio paper 3.

In this respect, it is important to stress the importance of social network analysis within this methodology. We can discover whether the message of the elderly gets broadcasted worldwide through social network analysis. It reveals their communication patterns within and outside of the Netherlands and makes possible to assess their impact in broadcasting their messages to different locations. The software Gephi is used to visualize the network of the followers and followees of our users.

Jingle

I want to conclude by saying few words about the relevance of the methodology I have proposed.

The innovative methodology I have proposed relies on language and media but can have broad applications in urban studies: the results of the data analysis allow for an understanding of the storytelling behind the smart city discourse. This is important knowledge for urban planners and policy makers as is knowledge based on spontaneous data.

The hashtags employed by the creative migrants highlight social projects, issues related to environment, sustainability, circular economy as well as a strong interest for technology and data in relation to common goods, being in line with the claims of grassroots movements. Similarly, the hashtags employed by the elderly reveal an interest in nature as well as in environmental sustainability.

Urban policies must acknowledge that technology and innovation should be employed to tackle social challenges and should not only support a privatized smart city driven by corporations that often neglect social issues. The communities I have investigated are interesting because they provide a good example of the way digital and environmental sustainability interact. The data produced by these groups, if properly analysed, can support an alternative human-centred smart city discourse.

However, in certain cases, a quantitative analysis, like the one I have discussed in this audio paper, might only confirm what we already know, and might not be very informative. In order to discover new knowledge, we might need more than that. More about the methodology in the next audio paper.

Don't miss it!