Cyber Cities:

The Digital Basis of Social Buzz in Smart Cities[1]

Abstract

'Big' urban data are increasingly becoming accessible for scientific research and policy use. They may enhance the intelligence that is needed for smart cities. The present paper aims to highlight and demonstrate the rich potential of information based on digital technology in modern cities. As a case-study example of the power of social media data as a support tool in smart cities, we consider Twitter data in the municipality of Amsterdam. We use machine learning techniques to identify temporal patterns that we then relate back to their spatial dimension, effectively connecting the digital with physical aspect of cities. We also show that analysis of geo-referenced tweets can shed significant light on physical aspects of the city and on the spatial distribution of urban functions.

Keywords: social buzz; digital technology; Twitter; ICT; '*New Urban World*'; smart city; internet geography

1.   Cities as Cyberplaces for Social Capital

Cities are the heart of a complex body called 'the space-economy', while interconnected (infrastructural, knowledge, social, trade or business) networks make up the blood circulation of this body. With the radical transformation of our space-economy into an 'urban planet' (coined the '*New Urban World'* by Kourtit 2013), the urban economy has become a prominent feature of the 21st century. The fate of our economies seems to rest in particular in the hands of large cities, as these cities outperform all other human settlement patterns in terms of productivity and wealth creation (Glaeser 2012). Such cities have turned into complex and efficient intra-urban networks, while also being connected to the outside – sometimes global – world through multi-layer inter-urban networks (see Neal 2013). The '*New Urban World'* has thus become a complex evolutionary organism.

The complexity of the emerging global urban system is further rising as a result of the structural increase in the world population in the decades to come. Cities will become the necessary havens to accommodate the rising volume of people. The abovementioned long-term megatrend in population movement towards the city will intensify large-scale urbanization, and will mirror both the natural population increase and large migration movements in various parts of the world. Consequently, urban challenges may become the most complex and critical factors for sustainable development in the future. Despite (sometimes) negative perceptions of cities, it should be noted that in general cities are meant to be the 'home of man'. Nevertheless, they will have to meet strict sustainability conditions in a dynamic – often global – environment. It is widely accepted that cities are able to create many positive external benefits, so that from an economic perspective there will be a structural tendency for an increasing influx of people into urban areas. Socio-demographic changes (e.g., ageing), migration and mobility, entrepreneurial dynamics, sustainability and efficiency of transport and energy systems, ICT (and other advanced technologies) and increasing returns to scale in urban agglomerations are the driving forces behind the continuation of urban settlement patterns in our modern society.

Cities are clearly no longer 'isolated islands', and have to be competitive to acquire a strong position in an open world (see Kourtit et al. 2011). This new trend has prompted the popularity of the concept of a 'smart city', a city that is able to exploit its indigenous assets (knowledge, technology, entrepreneurship, accessibility, sustainability, culture etc.) to gain a relatively strong (socio-)economic position in an open spatial system (see Caragliu et al. 2011; Kourtit et al. 2013). An important vehicle for smart city initiatives is formed in the design, access and use of appropriate ICT technologies and facilities on a fit-for-purpose basis. This awareness has led to the development of a varying nomenclature for ICT-instigated smart cities, such as: computer city, plug-in city, cyber city, fiber city, software city, wiki city, wired city, digital city etc. (see also Grosveld 2002).

The urban system is indeed a multidimensional dynamic organism with many characteristics of a complex economic, social and cultural nature. Cities are evolving systems driven by a multiplicity of actors and stakeholders. It goes without saying that cities have many 'faces'; they are not uniform or identical. The management and governance of these modern, complex and ever-rising urban agglomerations calls for effective and focused decision-making tools. There is a general awareness that traditional urban planning tools (models, stakeholder analysis, and consultation methods) are no longer able to cope with the multiplicity of challenges faced by cities in an open world, driven by digital technologies.

It is increasingly recognized that advanced ICT use in an urban setting may drastically change the organization of smart cities. ICT may, for example, stimulate the competitive position of business firms in an urban economy (e.g., through outsourcing, call centers, B2B services, B2C services, enhanced logistics), improve the level of public services in the city (e.g., standardized procurement services, e-

democracy, efficient delivery of community services), or intensify the potential of social capital in the city (e.g., better access to collective initiatives, appropriate information on educational or health services, encouragement of voluntary community activities). A review of the various opportunities created by ICT services in the context of urban social capital can be found in Cohen-Blankshtain and Nijkamp (2013). It is clear that ICT adds a new dimension to social capital in the city, as introduced amongst others by Bourdieu (1986). ICT offers a low-cost, high-frequency and user-friendly addition to local social capital, for instance, through digital opportunities such as Facebook, Twitter or text messages. The use of ICT reinforces the impact and position of smart cities, not only from efficiency, but also from a social cohesion perspective (cf. Sadler 2005). Social media are likely to change the functioning of urban spaces, as a permanent flow of interactive information will change the behaviour of citizens and visitors into dynamic urban consumerism (see also Pentland 2009). Modern cities tend to be increasingly controlled and influenced by – top-down and bottom-up organized – complex data platforms. Access to and smart use of such data platforms – sometimes called 'reality mining' – will most likely become critical success factors for our urban future.

The trend towards new geo-science applications (e.g., geodesign, geo-imaging) ties in with the current discussion on intelligent and smart cities, in which the use of digital technology in cyber space (through the Internet, web, etc.) lays the foundation for modern urban planning and analysis. Essentially, a smart city emerges when urban intelligence is added to information. This information is created and stored in 'electronic warehouses' (e.g. Internet, GPS, sensors, etc.) and offers a new map of the structure and interactions in urban life. This new approach is often also coined internet geography or cyber geography (see Batty 2012, Leamer and Storper 2001; Malecki 2002; Storper and Venables 2004; Tranos 2013). This information incorporated in these – often real-time – databases is increasingly used in social media and location-based services. For example, Foursquare and Facebook are able to inform others on the spatial imprint of daily urban activity behaviour of residents or visitors, and opens new ways for digital geography and even psycho-geography (in relation to spatial mental maps). In the public domain this type of information can also be used for geocentric incident management and urban or regional crowd management (see also Boyd and Crawford 2012).

Recent years have witnessed a rapid increase in the development and application of advanced geodesign methods for effective urban planning and policy (see Flaxman 2010). The need for up-to-date innovative and operational planning tools originates from: (i) the multi-faceted nature of urban policy goals; (ii) the great potential offered by the massive presence of 'big' urban data; and (iii) the emergence of complex multi-stakeholder choice and planning challenges. Especially the use of digital data and interactive communication tools offers new perspectives for original urban planning tools in our 'urban century'.

As argued by Nijkamp and Kourtit (2012), density, proximity, accessibility and connectivity are the cornerstones of a modern, ICT-connected city. Modern ICT-led cities are able to meet a series of strategic objectives (cf. Hollands 2008):

- benefits from networked infrastructure in terms of enhanced economic and political efficiency and more opportunities for social, cultural and urban development;
- exploration of new opportunities offered by a business-led urban development;
- utilization of urban social capital through the inclusion of urban residents and visitors in public services;
- strategic reliance on high-tech potentials of the city (including creative industries) for long-run urban development;
- creation of a new socio-economic potential through the development of social connectivity and trust principles and initiatives in the city;
- development of new strategic initiatives and plans for social and ecological sustainability as a solid and promising bases for a smart city.

This paper presents a novel spatial analysis of one of these new sources of urban data from digital feeds. In particular, it analyses a set of Twitter messages geo-referenced within the municipality of Amsterdam. The contribution is two-fold: at a specific level, it offers insight into the spatial structure, dynamics and the distribution of population in a global city, adopting a distinct (spatial-)data-driven approach; at a more conceptual level, it presents a concise application, almost a pilot study, of how elements of the smart-city discourse can be readily applied to better understand the functioning of an urban area. This is realized by combining not only different data sources of recent appearance, but also techniques from the field of statistical and machine learning in an exemplification of how urban studies can benefit from this seemingly foreign domains of expertise.

This paper is organized as follows. After this introductory section, we will highlight in Section 2 the role of modern cities as social network machines producing 'tonnes' of socio-spatial interaction information, including tweets. Next, Section 3 will be devoted to a description of the Twitter database used to study spatially detailed information patterns in the city of Amsterdam. In Section 4, a series of statistical and visual findings will be presented based on the Twitter data analysis for Amsterdam. And finally, Section 5 will draw some policy lessons on the use of 'big' digital data for smart city management.

2.   Cities as Social Network Machines

As argued in the introductory section, we live in the '*New Urban World*', which is a new epoch in the history of mankind. The *'New Urban World'* refers to the persistent rise in the share of a nation's population that lives in urban areas, be it in a geographically concentrated form (e.g. cities) or a deconcentrated but functionally connected form (e.g. metropolitan areas, poly-nuclear spatial patterns, sprawl areas). The '*New Urban World*' does not display a homogeneous and stable settlement pattern, but rather a spiky and dynamic landscape. This megatrend offers various great opportunities for urban development, but puts, at the same time, enormous pressure on our urban areas by inducing also negative externalities, such as pollution, congestion, security issues and social degradation. In addition, cities – especially in Europe – have an abundance of historical-cultural heritage that needs to be protected and exploited. A smart city has to integrate various dimensions that shape the future of the city through proactive intelligent data management. Such a city has to face the challenging task to govern a complex and open spatial system as a set of (internally and externally) connected intelligent subsystems (see also Komninos 2002).

A prominent feature of all cities – as dynamic and interconnected living ecosystems – is that they house human beings, be it residents or visitors. Their social behaviour forms altogether a fuzzy set of social interactions, sometimes weak (or even negligible) and sometimes strong (positive or negative). The type of interactions depends on various functions ranging from joint consumption (e.g., visiting a theatre or a football match) or joint production (e.g., working together in an office) to strong social interdependence (e.g., a protest demonstration or a meeting in a bar). Thus, social communication is not one-dimensional, but multifunctional and takes place in different layers and functional hierarchies. Modern high-tech cities tend to become multi-tasking smart engines of social and cognitive proximity (see also Boschma 2005).

Social networks – and in a more general sense – social capital (see e.g. Bourdieu 1986) – has received a great deal of attention in social science research over the past decades. Social networks are functionally organized ways of social – often interpersonal – interaction (see Hanneman and Riddle 2005). It should be added that social relational structures have also a spatial constellation, the social space which refers to the geographical constellation that shapes or facilitates interactions as interdependences among individuals or social groups in a given geographical environment. Thus, geographical referencing is the dual side of social interactions (see also Abbott 1997; Bottero and Crossley 2011; Pattison and Robins 2004).

Social network analysis does not refer to a 'wonderland of no spatial dimensions', but is tied to concrete and systematic geographical pattern analysis, including the presence of nodes and ties (edges or links), as well as neighbourhoods. Thus, social networks represent interrelational and functionally determined systems mapping out the behaviour of social actors.

4

An intriguing question is now whether a similarity among social actors (e.g., gender, age, education, social class, profession, language, religion) favours socially interactive behaviour and, in general, social communication. The reverse question is whether social distance (or socio-psychological heterogeneity) discourages interaction among people (see Fleming and Petty 2000; Hipp and Perrin 2009; Massey 1981). Clearly, interaction may refer to the intensity or frequency of communication ('strong and weak ties') or to the nature and contents of communication among individuals ('functional ties'). In both cases, geographical distance may play an important role and even induce hierarchical interaction patterns (see Zipf 1949; Mouw and Entwistle 2006; Verdery et al. 2012; Modica et al. 2013). This observation is in agreement with the first law of Waldo Tobler (1970) in geography: "everything is related to everything else, but near things are more related than distant things" (see also Nijkamp 2013). The role of geography is also prominently present in the 'death of distance' hypothesis, in which the assumption is advocated that in a digital world geography hardly matters (see for a critical review Tranos and Nijkamp 2013).

In recent years, some researchers have pointed at an interesting paradoxical development: when both physical and virtual mobility increases (e.g., thanks to the internet), the sense of neighbourhoods and places we know is intensified (see e.g. Iannides and Zabel 2008; Krysan and Bader 2009; Blacksher and Lovasi 2011), most likely as a response to the need for self-identity in an open global world. From this perspective, digital information may be helpful in the search for social spaces where self-identity (including friendship, social entertainment etc.) can be realized (see e.g. Eagle et al. 2009; Raento et al. 2009).

As mentioned above, our paper will address the issue how ICT – and more specifically, digital Twitter information shared by group members (where groups are defined as social entities that have access to the same type of information) – will affect group behaviour in space and time. This calls for the analysis of large digital data systems on social network behavior, taking into consideration the fact that cities turn into smart spatial data and information engines that map out on-line or real-time spatial interactions in physical and digital urban space. As a consequence, digital spatiality tends to shape the future morphology and interaction patterns in urban agglomerations. In a smart city environment, open digital information sources (e.g., intelligent location-based services) allow for a reliable identification, understanding, contextualization and timely response to events, especially when there is a common willingness of all urban actors concerned to offer intelligent solutions for vital and sustainable cities (see also Abdoullaev, 2011).

Smart cities are thus in general characterized by an intensive use of digital technology, including social media. Not only the socio-economic acceptance pattern of social media, but also the spatial distribution of the use of social media in smart cities is an important research challenge. In the present study we will focus on the analysis of Twitter data in the city of Amsterdam. Twitter data are popular vehicles for

transmitting short messages to a wide audience, are easy to use and to receive, and have a high penetration rate in many countries and cities. The use of this digital technology symbolizes the image of a smart city.

In the current research project, we will address the question of functional-cultural delineation of urban areas from a bottom-up perspective of micro-data from social media, acquired at the spatial scale (6-digit postal code) in the city of Amsterdam. In this way, Twitter data can be used to depict the social composition and diversity of tweets in a geographically referenced smart city, e.g. in terms of language, user categories, user frequency of tweets, etc. In the next section, we will describe the data base used for our investigation.

3. The Digital Database for Amsterdam

 Much data from information machines comes from social network use. A social network is an organized social structure made up of individuals (or organizations) called 'nodes', which are tied (connected) by one or more specific types of bonds or interdependency, such as friendship or a common interest. Social media create new communication channels to collect, generate, share, circulate, and exploit information, as well to generate a different type of value-added information (personal comments and insights).

Social networks are relatively recent phenomena: Facebook was launched in 2004, Twitter in 2006, FourSquare in 2009. The diffusion has been so rapid and viral that the number of users of some of these websites is larger than most nations in the world. With the advent of social media, the entire logic of communication has changed. While the institutional channels still remain in place, a series of autonomous and unstructured channels has emerged. Usually, the first source of information of a specific event is formed by 'eyewitnesses', who share images or texts almost immediately through websites and social networks. This process can be much faster than any other form of communication. This information is a prominent source for both the public and the media. The latter implements ways to capture and organize bottom-up sources and comments to create reports, as well as to ensure a real-time information feed. For example, twitter messages for the emergency services embody three vital emergency functions: assessment (What is happening?), coordination (What is needed?), and response (What must be done?). This provides an additional source to support Situational Awareness for emergency organizations. The large amount of data pumped into the information stream from the public, the media, and government sources leads also to questions regarding its accuracy, reliability, and privacy. In our study on smart cities we address in particular the potential of Twitter.

Twitter is an online social networking and micro-blogging service that enables its users to send and read text-based messages of up to 140 characters, known as 'tweets'. The structure of a Twitter message usually includes:

- User information
    - Name and nickname
    - User ID
    - Time zone
    - Number of followers
    - Number of following
    - Account information (e.g. language and home location)
- Message text
    - Message ID
    - Message text
    - Links to external sources, hashtags (#) and references to other user (@)
    - Source (service through which the tweet is generated)
    - Status (e.g. retweet)
- Message metadata
   - Time stamp
    - Location (if applicable)
    -

Through this structure, a tweet can contain the following elements of information:

- Users' comments or reports concerning current events, which they consider important or interesting;
- Direct reference to the particular topic or group of posts by a hashtag "#" e.g. #eurocup. The creation of hashtags is community-driven;
- Reference to the particular user with the '@' sign followed by a username, e.g. @Rijkwaterstaat.
- Links to external websites or pictures using a hyperlink.

Twitter offers four different types of Application Programming Interfaces *(*API): (1) Twitter for websites with ready-to-use website components, e.g. 'tweet' and 'follow' buttons; (2) REST API for handling

posting messages to Twitter; (3) Search API with the possibility of querying back in time up to one week, with a limited number of queries types (access via HTTP GET); and (4) Streaming API which is the most suitable for downloading big amounts of data, and recommended for data mining (access via HTTP POST). A twitter database contains basically two options: messages with and without geo-location. For our case study on Amsterdam, we have used three months of data (October 2012 – December 2012) which contained about 200,000 geo-located tweets of Greater Amsterdam. Messages are filtered based on their location: only those with attached geographic coordinates, located within the Greater Amsterdam area. The Twitter data feed can be biased by artificial activity of bots and astroturfing. Such bias has been filtered out. The hourly distribution of tweets over an entire day can be seen in Figure 1. The dashed line represents the average day, while coloured straight lines show the aggregate for every day of the week.
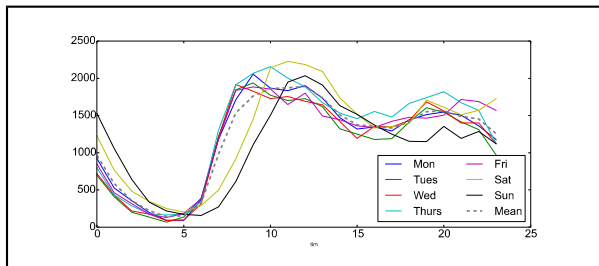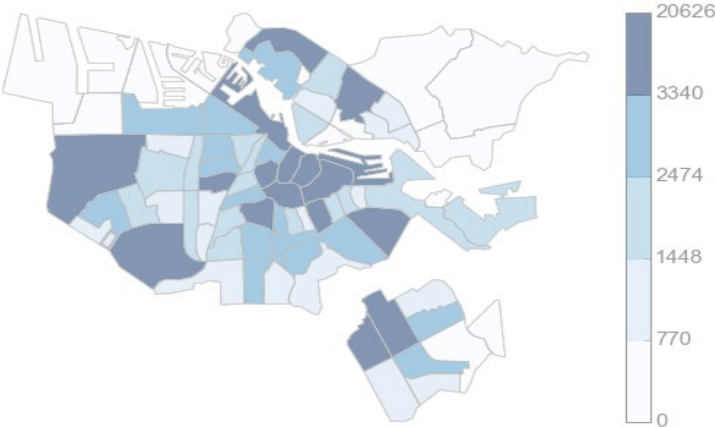

Figure 1. Tweet volume aggregated by hour of the day (0h-24h) by day of the week in Amsterdam

The main pattern is clear and intuitive: early hours in the morning show low activity. Activity starts picking up around 6am, and continues increasing until the mid-morning when it reaches the day peak; the afternoon displays a small valley in the volume of activity that is overcome towards the evening when work hours are over. Besides this general trend, there is also an interesting duality between weekdays and weekends. The time when the morning activity begins to pick up is clearly delayed for Saturday and, particularly, for Sunday, while it is very homogeneous and earlier for the rest of the week. The highest day peak is reached on Saturdays around 11am. Interestingly, Sundays show the largest drop between the morning peak and the evening trends, although it is just second to Saturday in terms of morning activity; tweets on Sunday appear to decrease to the lowest level of the week in the evening hours. In the next section we outline an analysis of Twitter data representing smart functions for the city of Amsterdam.

4. Visualization and Analysis of Social Buzz in a Digital City – Twitter Feeds in Amsterdam

8

In order to identify the structure and to explore regularities in the Twitter activity of Amsterdam, we aggregate tweets by hour and to a postal code 4 level. This allows us to tackle the temporal and spatial dimensions at a level at which we do not loose much detail but also at which we can still count with enough data to see variation (using lower levels of detail yields many "empty bins", points in time at which activity is zero in a particular area). The spatial distribution of the aggregate volume is shown in Figure 2.

Figure 2. Tweet volume aggregated to a postal Code-4 level

The map in Figure 2 contains 81 polygons distributed in the region of the Amsterdam municipality. The main challenge is to extract patterns out of the time trajectories of each postal code to investigate the space-time distribution of tweets. In a literal sense, this implies obtaining lines or trajectories, such as those in Figure 2 for every single polygon. This is shown in Figure 3, but only for illustrative purposes, as it is very hard to extract any pattern visually due to the large number of lines.
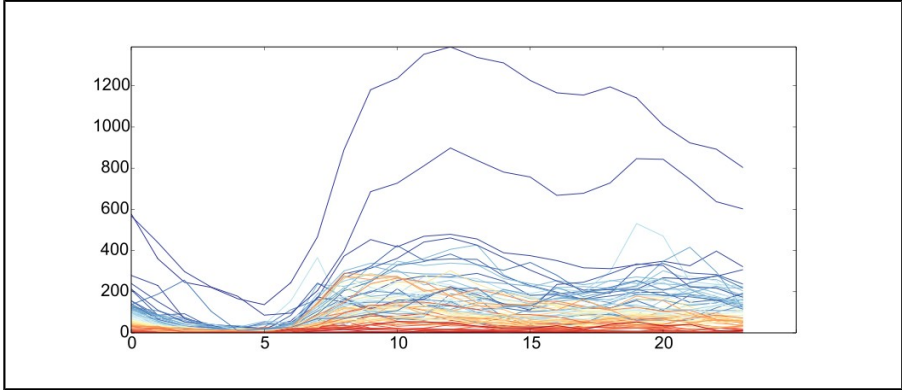
Figure 3. Hourly tweet volume by postal code-4 level

The method we propose is to use a clustering algorithm to group similar trajectories of postal codes by their shape. In order to avoid "level distortions" and to focus really on the shape of the trajectory, we use standardized trajectories for clustering. This means that, instead of using raw volumes of tweets per hour, we calculate the percentage that each hour represents, as a share of the total volume of that postal code. Clustering of time series is a well-known problem in pattern recognition. Following the taxonomy established in Warren Liao (2005), we use an approach that employs a general-purpose clustering algorithm on the raw-data (as opposed to on features extracted from the data or on model outputs). Robustness of this approach is tested by using two very distinct techniques to cluster the trajectories: the widely-known K-means algorithm, and the somewhat more advanced self-organizing map (SOM, Kohonen 2001).

After experimenting with different number of solutions and alternative parameters in both algorithms, we find that the outputs look very similar[2]. The number of postal codes within each solution is surprisingly very close to each other, and so is the overall shape of the trajectories in each cluster. The robustness of these results is particularly reaffirming, given the very different paradigm that each algorithm uses. Given this similarities, we will use results only from the K-means algorithm in the remainder of the analysis, since it is arguably a more popular technique.

Figure 4 shows the output from a solution with five clusters, which we find particularly interesting. After experimenting with higher and lower number of clusters in both the SOM and K-means, we decide to use five for the rest of our study. Including more additional clusters does not seem to add particularly different cases, while smaller solutions substract some groups which we find interesting to focus on.

There are three main clusters containing the vast majority of areas ($\approx$ 93% of the sample) and two clusters with outliers. Almost half of the observations are labelled in a group with, roughly, two peaks and two valleys (#1, in lime colour). Hours of low volume in this case are late at night (1am to 6am) and, albeit with more activity, the afternoon (around 3pm). Rush hour in this group is divided between mid-morning, with the day peak around 10-11am, and the evening where it picks up from the afternoon. The second largest cluster (#0, in red) displays a similar profile, but the morning peak is not as pronounced and more activity seems to occur in the evening. The third most relevant cluster (#3, in blue) has a markedly different profile: it is characterized by only one peak that occurs towards the end of the morning and, after that, a decay in the intensity of tweets spans throughout the entire day, unlike in the previous two cases, without any second comeback

---

[2] Results are not shown, but available from the authors upon request.

in the evening.

The remaining two clusters only contain a handful of areas and thus capture very outlying trajectories. The one with more than one area (#2, in green) is characterized by a significant portion of its activity concentrated in the morning peak and a low profile for the rest of the day, while cluster #4, in purple, with only one area (a peripheral postal code with very little volume of tweets) has a profile that appears random throughout the day, but shoots up at the very end of the day (after 9 pm), indicating a spree of activity towards the end of the day.
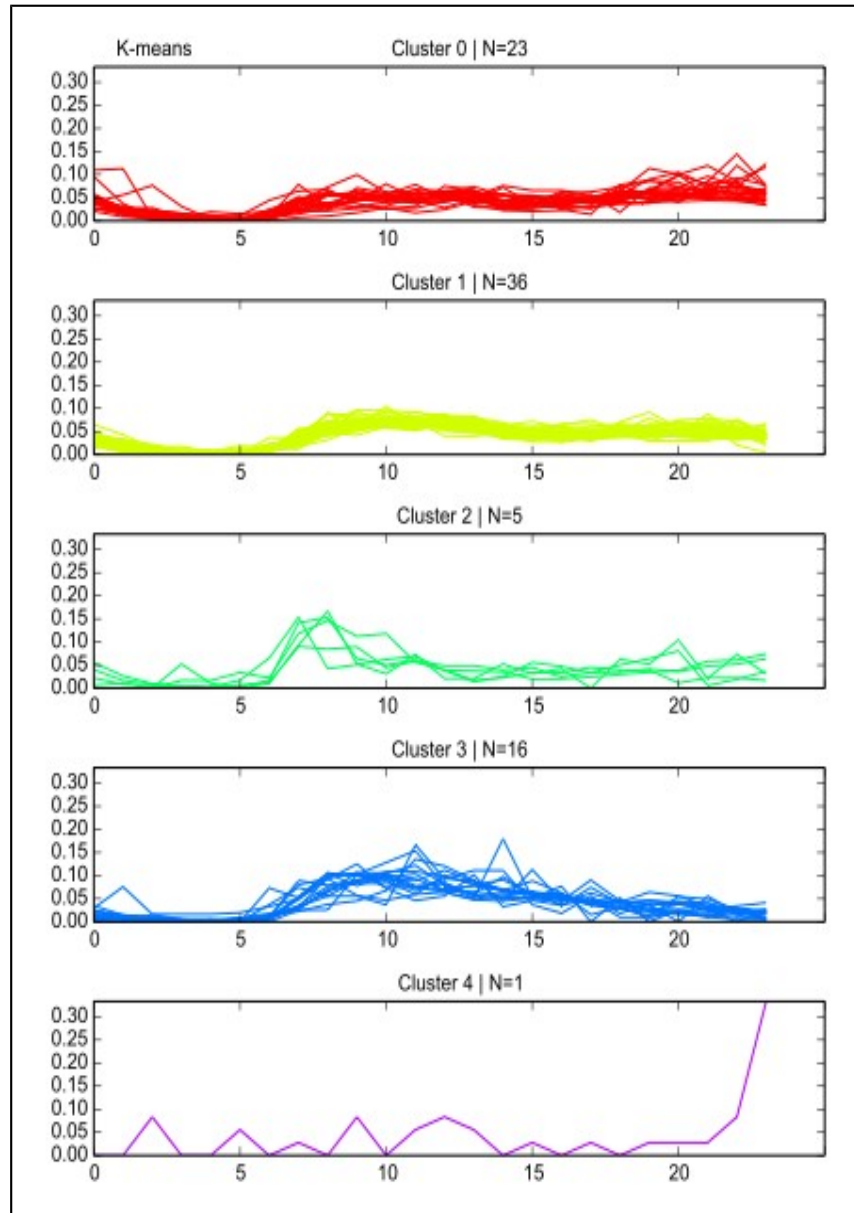
Figure 4 . Postal code-4 trajectories by cluster

Combining the temporal dimension with its spatial distribution, Figure 5 shows the geographic location of the clusters and a time plot of the average trajectory in each of them.
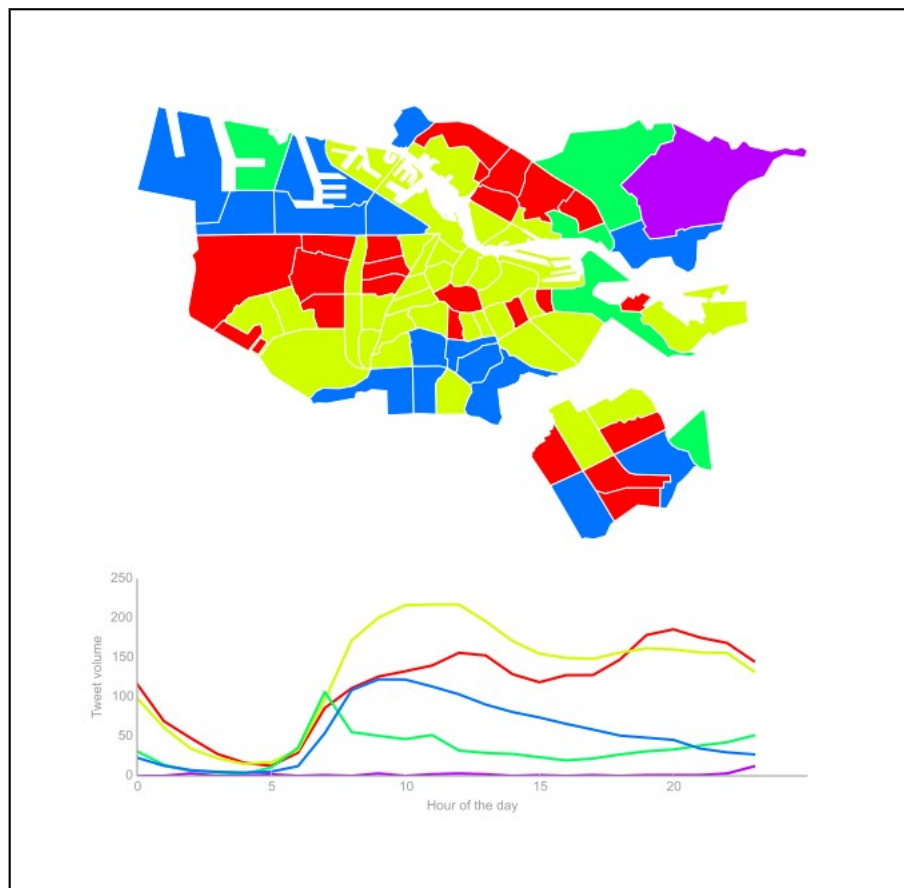
Figure 5. Space-Time distribution of clusters

The colouring scheme is consistent across views, so that a visual comparison can be quickly established. The lower panel of the figure displays the average volume of tweets by hour for the postal codes in each cluster. It is important to reinforce that, although the graph is based on the raw volume, the clustering was performed, as explained before, based on standardized changes. Clustering on standardized values allows us to group by similar shapes, ignoring the level effect, while representing the output of the clustering with actual volumes allows us to get a sense not only of the shape of each group of areas, but also of the overall level of the cluster in the general picture. From a functional standpoint, the clusters can be interpreted as representations of very different types of areas in a city.

Lime: Mixed uses in the case of the one in lime are highly frequented areas from which people tweet throughout the day and, particularly, during the morning, when the day peak is. Still, the volume of tweets sent from these areas remains high well into the evening. This is typically associated with areas where a range of urban uses, such as residential, commercial or office space, are intertwined and people are attracted, and hence end up tweeting, not because of only one reason but because of several of them (e.g. residence, job, amenities). Perhaps not surprisingly, most of these areas are located in the center of the city as well as in more peripheral neighbourhoods which, as it is well known for the Netherlands, maintain a high degree of land-use mix.

Red: Residential areas are classified in red show a similar trajectory, with a high total number of tweets and two marked peaks in the day, be it that in this case, the most intense time of the day is concentrated towards the evening. This may be indicative of areas which people particularly frequent once they have left work, either because of these are their places of residence or because they offer leisure activities such as bars and cafes, as is the case of the touristic. Leidseplein in Amsterdam, the most central area in red, which is full of tourist amenities and remains lively well into the late evening. Red polygons in the West and North of the city are classical examples of

residential areas and also fit this profile of Twitter activity, particularly high at the end of the day, albeit that in this case the tweets originate most likely from homes.

Blue: Work space, i.e. parts of the city with a high concentration of employment and a limited degree of mixed use, are captured in the blue cluster. This is reflected in the evolution of Twitter activity over the day. In effect, these areas are characterized by only one peak in the morning that starts a bit later than in the lime and red places, and a slow but constant decay from there throughout the rest of the day. This is typical of work places, where people are attracted only for one reason and which fulfil only the work function. This lack of range of activities concentrates activity in only one part of the day, and leaves the rest of it with a lower number of tweets. Spatially, they are located in the periphery of the city and, to an eye familiar to the geography of Amsterdam, blue areas map out nicely the largest employment centers of the city, e.g. the docks and the World Trade Center, among others.

Green and purple: These are outliers referring to five areas classified into green and one into purple. Green areas have a marked peak very early in the morning, a sudden drop and a bit of activity picking up towards the end of the day. These are also not the most popular places: on average, the volume of activity they attract is not small, as compared to the previous groups. This profile could indicate remote residential areas where people can only tweet at the beginning and the end of the day because they are somewhere else (working) during the rest of the day. The only area categorized as purple is located in the north-east part of the city and comprises mostly countryside, very sparsely populated. With a very low number of tweets, even a small peak in one hour (as it is the case at the end of the day, where there are ten tweets sent) can distort the entire profile of the area and this change might be picked up by the clustering algorithm as something very distinct. This case is a good example of the relevance of performing data quality checks throughout the entire analysis This a caveat for the researcher to always be cautious when interpreting results from this type of approaches with data from social media.

Finally, in order to verify some of our hypotheses about the function and use of the areas delineated by the five clusters, Figure 6 shows the respective shares in each cluster for residential, office, industrial and commercial space. Very much according to our previous description, clusters 0 and 1 (red and lime) have a high presence of residence and commercial space, and a medium level of office activities. As mentioned, areas in lime align more with a highly mixed use of space and, hence, the proportion of offices is also higher,. Industrial use is very low in both clusters.

Established as work destination by the Twitter trajectories, blue areas are markedly defined by their industrial and office space, medium levels of commercial and low shares of residential use. Green areas contain a mix of residence and industrial uses, with very low space devoted to other functions. Conversely, the outlier in purple contains a high share of residential and industrial use and virtually no space devoted to office or commercial.
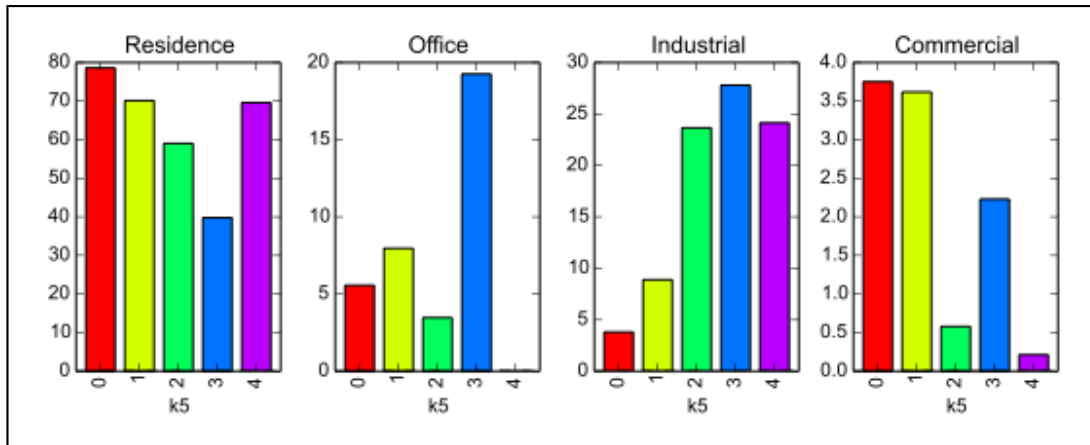
Figure 6 . Average proportion of land-use by cluster

5. Policy Lessons for Smart Cities

This paper has focussed the attention on the use of digital information – in this case, tweets – in the city of Amsterdam. As argued before, smart cities are characterized by an intelligent use of e-data, and Twitter is one of these information technologies. Its rapid transmission potential and wide acceptance makes Twitter an appropriate social capital tool in a smart city.

Such social media are still in their infancy, but they tend to have increasingly a radical impact on the social fabric of the city. First, they have a drastic impact on communication behaviour, especially because of their fast and pervasive nature, so that communication patterns in urban areas are changing substantially. Next, even though the cyber space seems to act as a disjoint dimension of physical space, in reality there is a close connection between these two spaces. Patterns of tweet may resemble and mirror land use, diversity and proximity patterns in cities.

Consequently, urban policy should not regard social media use as a communication technology per se, that is disconnected from real urban space. Instead, current land use patterns and uses (e.g., entertainment centers) prompt the use of social media as a modern communication tool, whereas on the other hand social media may exert a great impact on mobility, business and leisure patterns of urban residents and visitors. It seems thus plausible that the role of digital technology in urban dynamics and morphology will increase substantially in the future. Complementarily, regardless of direct effects this technology has on

the actual patterns of behaviour, these new data sources offer an opportunity to more closely monitor, study and, ultimately, understand the idiosyncrasies and needs of cities. In this direction, the present study offers a window to look into more ambitious possibilities. As an example, it does not take much invention to envision a real-time system based on the analysis displayed here but that is constantly updated, offering the urban planner and policy maker a close connection to the city.

In conclusion, digital information has in the past years become a source of novel research. The hundreds of millions of tweets sent out on a daily basis offers an unprecedented research potential with great policy implications, not only at a societal level but also at a local community level. These information sources are also extremely interesting for analysing cultural diversity in cities on the basis of the language of tweets. The density and time pattern of tweets offers also important strategic information on the human geography (consumers and producers) in modern cities. For geoscience applications in urban planning (e.g., crowd management, spatial segregation trends, socio-spatial profile analysis, traffic management etc.), the use of digital information including cell phone data, tweets, GPS information and other micro-based digital sources may open new avenues for smart cities.

References

Abdoullaev, A. (2011), A Smart World: A Development Model for Intelligent Cities, Keynote Address 11th IEEE International Conference on Computer and and Information Technology (CIT-2011) (http://www.cs.ucv.ac.cv/CIT2011).

Abbott, A. (1997), Of Time and Space: the Contemporary Relevance of the Chicago School. *Social Forces* 75, 1149–82.

Batty, M. (2012) Building a Science of Cities, *Cities*, 29, S9–S16.

Beinat, E., Sitko, I., Hawelka, B., Kazakopoulos, P., Stanislavska, K. and Steenbruggen, J. (2012). Social Media Data Analysis for Crisis Management in the Netherlands. Report for the Dutch Ministry of Infrastructure and Environment. 1 November 2012.

Blacksher, E., and Lovasi, G. (2011). Place-focused Physical Activity Research, Human Agency, and Social Justice in Public Health: Taking Agency seriously in Studies of the Built Environment. *Health & Place* 2011, doi:10.1016/j.healthplace.2011.018.019.

Boschma, R. (2005) Proximity and Innovation: A Critical Assessment, *Regional Studies*, 39, 61-74.

Bottero, W. and Crossley, N. (2011). Worlds, Fields and Networks: Becker, Bourdieu and the Structures of Social Structures. *Cultural Sociology* 5 (1), 99-119.

Bourdieu P (1986) The Forms of Capital. In: Richardson J.G. (ed.), *Handbook of Theory and Research for the Sociology of Education*. New York, Westport, CT and London: Greenwich Press, 241–58.

Boyd, D., and Crawford, K. (2012), Critical Questions for Big Data, *Information Communication & Society*, 15 (15), 662-679.

Caragliu, A., Del Bo, C., and Nijkamp, P. (2011). Smart Cities in Europe. *Journal of Urban Technology* 18 (2), 65-82.

Eagle, N., Pentland, A., and Lazer, D. (2009). Inferring Friendship Network Structure by Using Mobile Phone Data. *Proceedings of the National Academy of Sciences* 106, 15274-15278.

Fleming, M.A., and Petty, R.E. (2000). Identity and Persuasion: An Elaboration Likelihood Approach. In: D.J. Terry and M.A. Hogg (eds.), *Attitudes, Behavior, and Social Context: The Role of Norms and Group Membership.* Mahwah, NJ: Lawrence Erlbaum, 171-199

Flaxman, M. (2010): Fundamentals of Geodesign. In: Buhmann/Pietsch/Kretzler (Eds.): *Digital Landscape Architecture* 2010, Wichmann Verlag, VDE Verlag GmbH, Berlin and Offenbach, 28-41.

Glaeser, E. (2012). *Triumph of the City*. New York: Penguin Press.

Grosveld, H., (2002). *The Leading Cities of the World and their Competitive Advantages. The Perception of 'Citymakers'*, Ph.D. Dissertation. Amsterdam: University of Amsterdam.

Hanneman, R.A., and Riddle, M. (2005). *Introduction to Social Network Methods*. Riverside, CA: University of California.

Hipp, J.R., and Perrin, A.J. (2009). The Simultaneous Effect of Social Distance and Physical Distance on the Formation of Neighborhood Ties. *City & Community* 8 (1), 5-25.

Hollands, R. G (2008). Will the Real Smart City Please Stand Up?. *City* 12 (3), 303–320.

Ioannides, Y., and Zabel, J.E. (2008). Interactions, Neighborhood Selection and Housing Demand. *Journal of Urban Economics* 63(1), 229-252.

Kohonen, T. (2001). Self-organizing Maps. Berlin: Springer, 3rd. edition.

Komninos N. (2002), *Intelligent Cities: Innovation, Knowledge Systems and Digital Spaces,* London and New York, Routledge.

Kourtit, K. (2013). *The New Urban World.* Amsterdam: VU University.

Kourtit, K. Deakin, M. , Caragliu, A., De Bo, C., Nijkamp, P., Lombardi, P., and Giordano, S. (2013), An Advanced Trpile Helix Network Framework for Smart Cities Performance, *Smart Cities* (M. Deakin, ed.), Routledge, London, 196-216.

Kourtit, K., Nijkamp, P., Lowik, S., Vught, F. van, and Vulto, P. (2011). From Islands of Innovation to Creative Hotspots. *Regional Science Policy and Practice* 3 (3), 145–161.

Krysan, M., and Bader, M. (2009). Racial Blind Spots: Black-White-Latino Differences in Community Knowledge. *Social Problems* 56 (4), 677-701.

Malecki, E.J. (2002), The Economic Geography of the Internet's Infrastructure, *Economic Geography*, 78, 399–424.

Massey, D.S. (1981). Social Class and Ethnic Segregation: A Reconsideration of Methods and Conclusions. *American Sociological Review* 46, 641-50.

Modica, M., Reggiani, A. and Nijkamp, P. (2013). Are Gibrat and Zipf Monozygotic or Heterozygotic Twins? A Comparative Analysis of Means and Variances in Complex Urban Systems, paper European Regional Science Association Conference, Palermo, August, 2013.

Mouw, T., and Entwisle, B. (2006). Residential Segregation and Interracial Friendship in Schools. *American Journal of Sociology* 112 (2), 394-441

Leamer, E.E., and Storper, M. (2001), The Economic Geography of the Internet Age, *Journal of International Business Studies*, 32 (4), 641-666.

Neal, Z. (2013). *The Connected City: How Networks are Shaping the Modern Metropolis*. New York: Routledge.

Nijkamp, P. (2013). The Universal Law of Gravitation and the Death of Distance. *Romanian Journal of Regional Science* (forthcoming).

Nijkamp, P. and Cohen-Blankshtain, G. (2013). The Importance of ICT for Cities: e-Governance and Cyber Perceptions, *Metropolitan Regions* (J. Klaesson, B. Johansson and C. Karlsson, eds), Springer, Berlin, 295-308.

Nijkamp, P., and Kourtit, K. (2012), The 'New Urban Europe': Global Challenges and Local Responses in the Urban Century. *European Planning Studies*, 21 (3), 291-315,

Pattison, P., and Robins, G.L. (2004). Building Models for Social Space: Neighbourhood Based Models for Social Networks and Affiliation Structures. *Mathematiques des Science Humaines* 168: 11-29

Pentland, A. (2009). Reality Mining of Mobile Communications, in: S. Dutta and I. Mia (eds.), *Mobility in a Networked World.* Paris: World Economic Forum/INSEAD, 75-80.

Raento, M., Oulasvirta, A., and Eagle, N. (2009). Smartphones: An Emerging Tool for Social Scientists. *Sociological Methods and Research* 37, 426-454.

Sadler, S. (2005). *Archigram: Architecture without Architecture*. Cambridge: MIT Press.

Storper, M. and Venables, A.J. (2004), Buzz: Face-to-Face Contact and the Urban Economy, *Journal of Economic Geography*, 4, 351-370.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography* 46(2), 234-240.

Tranos, E. (2013). *The Geography of the Internet: Cities, Regions and Internet Infrastructure in Europe*. Cheltenham: Edward Elgar.

Tranos, E., and Nijkamp P. (2013). The Death of Distance Revisited: Cyberplace, Physical and Relational Proximities. *Journal of Regional Science.* Online first publication available at: http://onlinelibrary.wiley.com/doi/10.1111/jors.12021/abstract

Verdery, A. M., Entwisle, B., Faust, K., & Rindfuss, R. R. (2012). Social and Spatial Networks: Kinship Distance and Dwelling Unit Proximity in Rural Thailand. *Social Networks,* 34, 112 - 127.

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11),1857–1874.

Zipf, G. (1949). *Human Behaviour and the Principle of Least Effort.* Reading MA: Addison-Wesley.