Innovation – The European Journal of Social Science Research
Vol. 25, No. 2, June 2012, 229–246

Smart cities in perspective – a comparative European study by means of self-organizing maps

Karima Kourtit*, Peter Nijkamp and Daniel Arribas

Department of Spatial Economics, VU University, Amsterdam, The Netherlands

(*Received 13 October 2010; final version received 10 October 2011*)

Cities form the heart of a dynamic society. In an open space-economy cities have to mobilize all of their resources to remain attractive and competitive. Smart cities depend on creative and knowledge resources to maximize their innovation potential. This study offers a comparative analysis of nine European smart cities on the basis of an extensive database covering two time periods. After conducting a principal component analysis, a new approach, based on a self-organizing map analysis, is adopted to position the various cities under consideration according to their selected "smartness" performance indicators.

Keywords: self-organizing maps; smart cities; creative cities; cities' performance; urbanization; location; concentration; proximity externalities; geographic knowledge spillovers

A creative urban world

The year 2007 marked an important development in the history of human settlements: for the first time, the share of total population living in cities exceeded 50%. Urbanization has become a major global trend, with ever increasing degrees of urbanization – reaching 70% and more – in various European and Asian countries.

The current developments essentially reflect the third revolution in urbanization in our world. The first revolution took place in antiquity during the rural to urban development period, when the earliest cities were formed. The second radical transformation in urbanization emerged from the Industrial Revolution, with the rise of massive industrial and population concentrations as a result of scale and agglomeration advantages, and the third settlement revolution took place in the post-World War II period, when cities were not only expanding in size, but also adopted a pivotal role in the spatial development of industrializing countries (including urban sprawl, new town development, edge cities, and networked and satellite cities). In this third phase, the functional position of cities also changed dramatically: cities were no longer rather passive human settlements, but generated an indigenous strength and outreach thanks to their creative and innovative potential. This urban transformation has prompted much original research, instigated, inter alia, by the melting pot hypothesis (see Jacobs 1961), the creative class hypothesis (see Florida 2002, Landry 2003), or the islands of innovation hypothesis (see Cooke 1998, Kourtit *et al.* 2011). Knowledge orientation and cultural diversity have become key factors for modern urban developments.

*Corresponding author. Email: kkourtit@vu.nl

ISSN 1351-1610 print/ISSN 1469-8412 online

http://dx.doi.org/10.1080/13511610.2012.660330
http://www.tandfonline.com

A modern city is not characterized by a "sense of place" but by a "place of senses". With their historical strength, creative and supra-infrastructures and important mobilized socio-economic resources and activities functioning as open platforms for a new and open future, modern cities shape and display a spectacular new urban cultural space and urban design, as well as improvements in lifestyle and liveability and economic viability from different perspectives (across the interfaces of economy, technology, society and culture). All this is supported by smart and innovative governance and elite lifestyles in the city. The socio-economic pluriformity (e.g. "cultural diversity") of cities aims to determine the various impulses that stimulate many different creative actors to become engaged in the city. This approach can significantly contribute to the entire space-economy by enhancing the economic and innovative urban environment of a city.

The present study addresses the issue of creative or smart cities from a broad multidimensional perspective. It seeks to provide an operational analytical framework for characterizing and identifying cities in Europe that have a high creativity potential based on a set of pre-selected performance indicators that play a critical role in creativity performance. The database for our comparative study is derived from the Urban Audit database in Europe. This multidimensional database is then analyzed by using two methodological frameworks, viz. principal component analysis and self-organizing map analysis. The latter method is a recently developed technique that – as one of the family of computational neural network models – has great scope for mapping out the relative differences in creativity potential of the cities under investigation.

The creative or smart cities movement

Our world is moving towards a new settlement pattern, in which urbanization (sometimes in a concentrated form – like Mexico City or Shanghai; sometimes dispersed with satellites and edge cities – like Los Angeles or Paris) becomes the dominant feature. In the "new urban world", metropolitan areas and cities are increasingly functioning as seedbeds for creativeness, innovation, entrepreneurship and spatial competitiveness. They are characterized by product heterogeneity and behave according to the laws of monopolistic competition in economics (see Frenken *et al.* 2007). Modern cities try to provide the highest possible quality or image in terms of culture, arts, sports, innovativeness, entrepreneurship, financial markets, sustainability etc. (Glaeser *et al.*, 2001) Density and proximity are the key features of modern cities (or, in general, urban areas). The past centuries have been characterized by a persistent trend towards urbanization. Some 200 years ago, less than 20% of the world population lived in cities, whereas nowadays the degree of urbanization is moving towards 80%. Worldwide, not only has the number of cities increased rapidly but also the size of cities, with a tendency towards megacities, which are large urban conglomerations with global power and considerable local/regional autonomy (Sassen 1991). Some people wonder whether this trend towards "more and bigger" cities might eventually come to a halt. However, from an economic perspective there is no valid argument that would convincingly demonstrate that there is a "natural limit" to city size or urbanization. It is plausible to argue that cities will continue to gain importance – in size and numbers – as long as the agglomeration benefits exceed the dark sides of agglomerations.

When Barbara Ward (1976) made a passionate plea for a positive view of modern cities as "the home of man", she meant to say that cities are the natural habitat for the human species in the post-industrial period, provided cities can offer favorable living and working conditions as a result of density externalities. Proximity is – despite the emergence of the ICT sector – still a major driving force for urban rise and expansion. Proximity is a frequently used concept in geography, but it has different connotations. First, there is *physical proximity* in terms of a short straight-line distance or a short distance based on the a transport network. In fact, what matters in interaction is the time or efficiency in bridging such a distance. Then there is *geographic proximity* is either a physical or a time concept, or both. However, in a social space there is also social proximity, i.e. a perceived small distance as a result of impacts from social relationships, common habits and interests, etc. (see, e.g., Gertler 2003). Clearly, both these concepts may be intertwined in an urban area: cities can be seen as agglomerations of economic activities based on the advantages of both kinds of proximity.

Proximity and agglomeration are two sides of the same coin. In the history of urban economics much attention has been paid to density and proximity externalities (à la Hoover, Isard), where often a distinction was made between scale, localization and urbanization economies. The density externalities perspective takes for granted that urban size has no limits, as long as the economies of density are greater than the diseconomies. According to the density externalities framework, cities offer important socio-economic and cultural advantages that are far higher than any other settlement pattern. In particular, in our modern age, cities offer spatial advantages related to knowledge spillover effects and an abundant availability of knowledge workers in the labor market (Acs *et al.* 2002). The spatial concentration of activities, involving spatial and social proximity, increases the opportunities for interaction and knowledge transfer, and the resulting spillover effects reduce the cost of obtaining and processing knowledge. In addition, knowledge workers preferably need to interact with each other in agglomerated environments to reduce interaction costs, and they are more productive in such environments (Florida 2002). Following this argumentation, cities are the cradle of new and innovative industries. Companies in the early stages of the product and company life cycle – when dealing with manifold uncertainty – prefer locations where new and specialized knowledge is abundantly available for free (see, e.g., Audretsch 1998; Camagni 1991; Cohen and Paul 2005). Cities offer an enormously rich potential for a wide array of business opportunities.

Clearly, the spatial extent of knowledge spillovers is limited, owing to various kinds of geographic conditions, for example, a wide daily activity system where people can meet easily and where people change jobs in the course of their careers, or smaller areas such as quarters in a central business district or university premises where people can see each other often by chance (e.g. Rosenthal and Strange 2001). The need for spatial proximity in order to benefit from knowledge spillovers seems, however, at odds with the impacts of the recent telecommunication revolution, as the costs of electronic communication have drastically declined, while advanced information and communications technology (ICT) allows for long-distance video-conferencing, data-mining, virtual design, computer-assisted decision making, etc. ICT offers an unlimited spectrum of virtual communication opportunities. But does it affect urban size?

To understand this paradoxical situation in the geography of knowledge spillovers, we need to look into the type of knowledge concerned (Howells 2002).

On the one hand, there is *codified knowledge* (partly just information) that can easily circulate electronically over large distances, for example, prices determined at a stock exchange and statistical data. On the other hand, there is *tacit knowledge* and its context, and these are critical in innovation processes. This kind of knowledge is vague and difficult to codify, and, accordingly, spreads mainly through face-to-face contacts of the persons involved. Tacit knowledge is transferred through observation, interactive participation, and practice. In addition, there is *contextual knowledge*, which is achieved through long-term and interactive learning, often in relatively open (unstructured) processes (Bolisani and Scarso 2000). All such density externalities present in a modern city offer a very powerful tool to enable cities to survive and grow and ultimately become hubs in a space-economy. Cities are essentially learning machines.

One of the most prolific authors who has advocated the learning region as a paradigm is Florida (1995). Earlier seminal work underlying the learning regions paradigm was done by Aydalot (1986), Camagni (1991), Maillat (1991) and others, while the paradigm was fertilized from different angles in regional studies, such as those on (regional) innovation systems, technology complexes (including knowledge spillover phenomena), post-Fordism and clusters, and those on technology policy, local and regional institutions, and community action (see, e.g., Benner 2003; Morgan 2002; Ratti *et al.* 1997; Cooke 1998; Maskell and Malmberg 1999; Gertler and Wolfe 2002). The learning regions approach has an advantage over other approaches in that it explicitly addresses the quality of creative policy-making and of other institutional conditions in the regional economy and society. In particular, it is a regional development concept in which the emphasis is on improving the individual and collective learning processes of the regional actors involved through open and flexible networks (OECD 2001). This concept does not imply, however, that the learning is exclusively taking place between regional partners. In fact, regional actors (e.g. policy institutes and companies) learn through both regional (local) and global networks.

Cities are turning into the geographical hubs (virtual and real) of a modern networked space-economy. They are the source of progress and global orientation, and hence deserve the full attention of economists, geographers, planners, sociologists, political scientists and urban architects. Thus, cities – and, more generally, metropolitan areas – will continue to be engines of economic growth, creativity and innovativeness. Clearly, R&D expenditures and investments in education and knowledge will be essential in this context, as these elements are the key ingredients for the increase of productivity at local and regional levels. This calls for pro-active and open-minded governance structures, with all actors involved, in order to maximize the socio-economic and ecological performance of cities, and to cope with negative externalities and historically grown path dependencies.

Smart cities in Europe

Smart cities is a policy concept in Europe designed to mobilize all knowledge centers into innovation hubs in order to strengthen the socio-economic progress in EU Member States. Smart cities have a high productivity as they have a relatively high share of highly educated people, knowledge-intensive jobs, output-oriented planning systems, creative activities and sustainability-oriented initiatives (see, e.g., Holland 2008; Komninos 2008; Landry 2003; Lee *et al.* 2006; Torres *et al.* 2005;

Paskaleva-Shapira 2008). To qualify as a smart city, it is necessary to comply with various quantitative indicators that can provide an informed picture of the performance of the cities under consideration. Such indicators should be measurable, comparable, transferable and consistent over all relevant cities. An important source of information on European cities is the so-called Urban Audit data set (EUROSTAT), which contains data on indicators in the following fields: demography, social aspects, economic aspects, civic involvement, training and education, environment, ICT, travel and transport, information society, and culture and recreation.

This information is systematically available for many medium-sized to large cities in Europe, and covers most European countries. The increasing European and global competition leads to the need to map out the innovative potential of these cities, if they want to play a significant role on the European or global scene. To that end, a benchmarking analysis of their relative position is necessary, in order to identify the strengths and weaknesses of these cities.

It should be noted that, in addition to the above conceptualization of smart cities, there are two complementary approaches on smart cities in Europe, each of them having a different scope and coverage. The first study is concerned with the analysis of "Smart Cities; Ranking of European Medium-Sized Cities" (2007)[1] and offers a comparative perspective on the performance of medium-sized cities in Europe, on the basis of six angles: competitiveness, social and human capital, civic participation, transport and ICT, natural resources, and quality of life. This study concludes that, in general, northwestern European cities are performing rather well.

Another study is called SmartCities[2] and serves a network of nine cities and academic partners with the aim of developing and delivering better e-services to cities and businesses in the North Sea Region. This is mainly an indicator study on the use of ICT for the development of e-Gov services (see Deakin 2010). We will use the latter set of nine cities for our comparative benchmark study, for the simple reason that detailed information on each of these smart partner cities is available.

Database and methodology

In our empirical work, we have collected a data set with a wealth of information on factors related to the characteristics of smart cities' performance from the Urban Audit data set and complementary data sources. This smart cities group belongs to a subset of regions bordering the North Sea and comprises, among others, European cities belonging to the SCRAN network (SmartCities (inter) Regional Academic Network). Table 1 shows the nine cities in our sample.

The list of smart cities indicators covers the various fields of the SCRAN cities' urban creativity performances for two periods. The data were collected over three years at two points in time: the first period is from 1999 to 2002, and the second period is from 2003 to 2006. Hence, the time period of the available data is from

Table 1. The set of nine cities belonging to the smart cities network (names in Urban Audit).

Bremerhaven	Kristiansand	Kortrijk
Edinburgh	Lillesand	Osterholz
Karlstad	Groningen	Norfolk

1999–2002 to 2003–2006, and this allows us to apply the analysis for both periods. Table 2 gives an overview of the variables from the list of the general data.

In order to compare the urban characteristics, and to identify "*smartness*" indicators that play a critical role in the creativity performance of the smart cities in a comparative way, it was pertinent to focus on the performance of a broad range of "*smartness*" indicators rather than on only one aspect of smart city development and performance. Clearly, there is no common definition of the "smart cities" concept, as we observe in reality multiple and different meanings. Therefore, for our study we decided as a starting point in the selection of our variables to focus on various literature sources (Cohen and Levinthal 1990; Coe *et al*. 2001; Graham and Marvin 2001; Florida 2002; Berry and Glaeser 2005; Poelhekke 2006; Southampton City Council 2006; Abreu *et al*. 2008; Nijkamp 2008; Glaeser and Berry 2006; Shapiro 2008; Caragliu *et al*., 2011) in order to identify critical "*smartness*" characteristics and indicators for the next operational step, viz. the use of self-organizing map (SOM) analysis.

In order to avoid the creation of an overwhelming amount of data, as well as to obtain a better insight into the evolution of smart cities over the period considered and to get some idea of the most crucial characteristics of the different regions in our subsequent statistical analysis, the long list of indicators was systematized and summarized by means of a principal component analysis (PCA). After the application of a PCA, we decided to select only the most relevant variables with the highest loading factors – in regard to *advanced business and socio-cultural attractiveness* (ADBA), *presence of a broad (public and private) labour force and public facilities* (PBLFPF) and *presence and use of sophisticated e-services* (PUSS) of smart cities, in order to explore the environmental sustainability of smart cities' development and performance, based on the PCA results. Table 3 shows the selected data on characteristics and factors from such an analysis, which will be used again at the visualization stage of the SOM approach below.

A SOM approach to visualize smart cities performance

In modern quantitative statistical analysis in the social sciences the concept of SOM analysis has become increasingly popular. The SOM[3] is a special kind of unsupervised computational neural network (Fischer, 2001) that combines both data projection (reduction of the number of attributes or dimensions of the data vectors) and quantization or clustering (reduction of the number of input vectors) of the input space without loss of useful information while preserving topological relationships in the output space. In a less technical way, we can say the SOM is a method that takes multidimensional data and compresses the information contained in them to present it in an understandable way for the human brain. The output of a SOM can be thought of as a spatial representation of the statistical relations between the observations; in this map, the axes are not north–south or east–west but measures of statistical similarity, which is expressed in the distance between observations. However, examples in the social sciences are still limited (e.g. Arribas-Bel *et al*. 2011). It is usually employed for two main purposes: (1) to visualize complex data sets by reducing their dimensionality; and (2) to perform cluster analysis in order to group similar observations into exclusive sets. Because of the former purpose, it is sometimes compared with other methods such as principal components or factor analysis; and, because of the latter, is also associated

Table 2. List of general data.

Indicators	Proxies	Years
GDP	Gross domestic product of city	1999–2002/ 2003–2006
Population	Total resident population	1999–2002/ 2003–2006
	Total economically active population	1999–2002/ 2003–2006
Employment	Proportion of employment in agriculture and fishery	1999–2002/ 2003–2006
	Proportion of employment in mining, manufacturing and energy	1999–2002/ 2003–2006
	Proportion of employment in commercial services	1999–2002/ 2003–2006
	Proportion of employment in construction	1999–2002/ 2003–2006
	Proportion of employment in trade, hotels and restaurants	1999–2002/ 2003–2006
	Proportion of employment in transport and communication	1999–2002/ 2003–2006
	Proportion of employment in financial intermediation and business activities	1999–2002/ 2003–2006
	Proportion of employment in public administration, health and education	1999–2002/ 2003–2006
	Proportion of employment in culture and entertainment industry	1999–2002/ 2003–2006
Human capital (level of education)	Proportion of population aged 15–64 with secondary level education living in Urban Audit cities (%)	1999–2002/ 2003–2006
	Proportion of population aged 15–64 with some college education living in Urban Audit cities (%)	1999–2002/ 2003–2006
Infrastructure	Length of public transport network (km)	1999–2002/ 2003–2006
Density	Population density in Urban Audit cities	1999–2002/ 2003–2006
Business	All companies	1999–2002/ 2003–2006
	New business registered in reference year	1999–2002/ 2003–2006
Local government	Annual expenditure of the municipal authority per resident	1999–2002/ 2003–2006
	Proportion of municipal authority income from transfers from nationa, regional, provincial and state government	1999–2002/ 2003–2006
	Debt of municipal authority per resident	1999–2002/ 2003–2006
Income	Median or average disposable annual household income	1999–2002/ 2003–2006
Unemployment	Number of unemployed	1999–2002/ 2003–2006

(*Continued*)

Table 2 (*Continued*)

Indicators	Proxies	Years
	Unemployment rate	1999–2002/ 2003–2006
Tourism and cultural heritage	Number of museums	1999–2002/ 2003–2006
	Annual visitors to museums per resident	1999–2002/ 2003–2006
	The number of theaters	1999–2002/ 2003–2006
	Annual attendance of theaters per resident	1999–2002/ 2003–2006
	The number of public libraries	1999–2002/ 2003–2006
	Total book loans and other media per resident	1999–2002/ 2003–2006
	Proportion of employment in culture and entertainment industry	1999–2002/ 2003–2006
	Tourist overnight stays in registered accommodation in Urban Audit cities (number of nights per year)	1999–2002/ 2003–2006
Leisure and recreation	Total land area (km^2) according to cadastral register	1999–2002/ 2003–2006
	Proportion of the area in recreational sports and leisure use	1999–2002/ 2003–2006
	Proportion of the area in green space	1999–2002/ 2003–2006

with techniques like the *k*-means or hierarchical clustering. It was first developed at the beginning of the 1980s (Kohonen, 1982) with the purpose of explaining the spatial organization of the brain's functions (Kohonen and Honkela, 2007), but the range of fields and applications for which the SOM has been used in the last years has grown exponentially.[4]

Although there have been several variants and modifications, depending on the kind of data and specific purposes for which it is used, we will consider here the basic algorithm, which performs a "non-linear, ordered, smooth, mapping of high dimensional input data manifolds onto the elements of a regular, low-dimensional array" (Kohonen, 2001, p. 106). Given the characteristics of the data and the purpose of this study, we believe the SOM is worth exploring as an empirical method to analyze and identify urban spatial structures.

Before detailing the actual procedure, it is helpful to clarify a couple of concepts that will be used throughout the explanation. The first is that of *input space* (also called signal space), which refers to the set of input data we employ to feed the algorithm; typically, the observations are multidimensional, and are thus expressed by using a vector for each of them. The second concept is that of output space (trained network, network or SOM), which defines the low-dimensional universe in which the algorithm represents the input data. It usually (although not necessarily) has two dimensions, and is composed of a set of elements called neurons (or nodes) that are interconnected, hence the network idea. What the algorithm does is to map the input space onto the output one, keeping all the relevant information and

Table 3. List of selected data (based on principal component analysis).

Context	Indicators	Code	Proxies	Years
ADBASOM	Employment	EC2011I	Proportion of employment in financial intermediation and business activities	1999–2002/ 2003–2006
ADBASOM	ICT	IT3001V	Number of local units manufacturing ICT products	1999–2002/ 2003–2006
ADBASOM	Population	EC1001I	Total economically active population	1999–2002/ 2003–2006
ADBASOM	Population	DE1001I	Total resident population	1999–2002/ 2003–2006
PBLFPF	e-Government	IT2003I	Number of administrative forms available for download from official web site	1999–2002/ 2003–2006
PBLFPF	Local government	CI2014I	Debt of municipal authority per resident	1999–2002/ 2003–2006
PBLFPF	Local government	CI2006I	Annual expenditure of the municipal authority per resident	1999–2002/ 2003–2006
PBLFPF	Infrastructure	TT1066V	Length of public transport network (km)	1999–2002/ 2003–2006
PBLFPF	Human capital	TE2028V	Proportion of population aged 15–64 with secondary level education living in Urban Audit cities (%)	1999–2002/ 2003–2006
PUSS	e-Government	IT2004I	Number of administrative forms which can be submitted electronically	1999–2002/ 2003–2006
PUSS	ICT	IT1005I	Percentage of households with Internet access at home	1999–2002/ 2003–2006

ordering observations in a way such that topological closeness in the output space implies statistical similarity in the input space.

The input space is composed of n-dimensional vectors that we want to visualize/cluster in a low (usually two) dimensional environment (output space). We can express the input vector t as:

$$x = [\xi_1(t), \xi_2(t), \cdots, \xi_n(t)]^T \in \Re^n \tag{1}$$

where ξ_i (t) represents the value for each dimension. There are no specific requirements for the data, before they become part of the signal space; however, as in other statistical procedures, depending on the distribution of the dimensions and their scales (if they represent different variables), it may be useful to normalize them, so that they all range within the same bounds and/or take a logarithmic transformation to avoid skewness. The output space is an array of x by y neurons (nodes) topologically connected, following a kind of geometrical rule (the most common topologies being squares and hexagons). Each of the nodes is assigned a parametric real vector, which we call "model" and is expressed as follows:

$$m_i = [\mu_{in},\ \mu_{in},\ \cdots \mu_{in}]^T \in \Re^n \tag{2}$$

The actual values of m_i may be chosen randomly, or assigned with any kind of previous knowledge in order to speed up the computing process. We may also define as d(x, m_i) any distance metric between two vectors x and m_i. The most widely used metric is the Euclidean distance, although other specifications are valid as well. What we are now looking for is a topologically ordered mapping of the signal space onto the network. That is done by the SOM in an iterative process called training, in which each signal vector is sequentially presented to the output space. The best matching unit (BMU) for x is defined as:

$$c = arg\ \min_i \{d(x,\ m_i)\} \tag{3}$$

When c is found, the neuron m_i is activated and the adaptive process starts, such that the BMU and its topological neighbors are modified by the following scheme:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \tag{4}$$

where t and $t + 1$ represent, respectively, the initial state and the final state after the signal has activated the neuron respectively; h_{ci} (t) is called a *neighborhood function*, and expresses how the BMU and its neighbors are modified when activated by a signal; usually, the linear or Gaussian versions are used. This process is repeated over many cycles before the training is finished. The neighborhood function depends on several parameters relevant for this stage: the distance between the BMU and the modified neuron (so the further away the neuron is, the smaller the adjustment); a *learning rate* $\alpha(t)$ that defines the magnitude of the adjustment, and gradually decreases as the training cycles advance; and the *neighborhood radius*, which decides which of the surrounding neurons of the BMU are also modified by the neighborhood function, and also decreases over the training stage. Altogether these make possible a topological preservation of the distances between input vectors in the output space.

It is important to stress that this learning process just explained is at the heart of the method and is used because the neighborhood function does not perfectly adjust the BMU and its neighbors, but only partially (this is controlled by the learning rate), and that this process is repeated a fairly large number of times, which leads to the arranging and self-organization of the input information onto the output space. It is also of interest to remark that, unlike other clustering or data reduction methods, the SOM does not compare directly the input vectors directly, but only makes comparisons from signal to neuron.

Once the training stage is completed, the network is ready to be used. There are two main (complementary) approaches that we will consider in order to explore the information provided by a trained SOM: first, we look at the network created, and then compare it with the original input vectors. The former gives an idea about the general trends and relationships of the data set, while the latter is intended to give information about individual vectors, and how they interact with each other.

First, the network itself is a useful tool of analysis. It has been said (Kohonen, 2001, p. 160) that a SOM tries to represent $p(x)$ as if it were a probability surface. A useful way to extract information is by visualizing the values of one of the dimensions (*plane*) on the network. This procedure allows us to find out how different values of different dimensions map together and interact. By construction, the SOM preserves the topology, but it tries to fill in all the available space, thus distorting actual distances from the elements. In order to visualize such distances and to identify clusters, Kraaijveld *et al*. (1992) propose the so-called *U-Matrix*, in which the average distance from each neuron to its neighbors is mapped on a grayscale. Here, a cluster border might be seen as an area where distances are large, although the decision of what is an actual cluster, and what is not is always a subjective one that the researcher must make.

The second option we will explore here is to link the original data on the network using the BMUs for each vector. This technique allows us to find out where the input vectors are mapped according to the SOM, and which ones are close, based on the data used. We will now address the empirical issues of smart cities.

Smart cities performance results

In this section we present the results from the SOM analysis applied to the above-mentioned data set on smart cities. Since we are dealing with space–time data, our approach here is geared to exploit both these dimensions in order to gain better insight into the spatial aspect, as well as its evolution over the period considered. There are two parts to this approach: in the first, we focus on each of the two time periods in order to understand the situations at the beginning and the end of the time span; and in the second, we analyze the evolution of the cities by means of trajectories, and give an interpretation. They are linked by the study of the trained SOM, for which we will refer to the principal components analysis from above.

The bulk of the analysis is based on the results of a SOM performed using standardized data for all the cities in the sample and for the two periods of time. This provides an output space based on all the available data (see Table 2) that allows us to map both periods onto the same network, and thus study the evolution of the cities in the period concerned. Also, since our aim is to analyze patterns and let the data speak for themselves, we decided to use a topology with enough neurons ($15 \times 15 = 225$), so that cities can be mapped without needing to be assigned to the same neuron, unless the similarity between them is very high. The 225 neurons refer to the number of neurons (hexagons) in the SOM that are used. It is a lattice of 15 by 15, which results in a total of 225 hexagons. The point is to note that the total number of hexagons is much larger than the number of observations that are used, thus allowing for true self-organization. In order to gain more confidence in the SOM chosen, we selected the best run out of 100, following the criterion of the minimum average *q*-error. Table 4 serves to specify more technical details; it basically sets up the specification to run the algorithm. The table shows all the technical parameters one needs to know to reproduce the application that is undertaken.

Static results

The first step we take is to consider separately the results for each point in time considered. Figure 1 displays the trained network, onto which the cities have been mapped according to their BMUs. Figure 1(a) shows the city values in the first period, while (b) shows those in the second. Let us recall from the previous section the interpretation: a short distance between two BMUs in the SOM implies two highly similar cities, particularly in the present context, where the network contains

Table 4. Technical parameters in the SOM analysis.

	Initialization	
Topology		Hexagonal
x, y		15, 15
Neighbourhood function		Gaussian
	First part	*Second part*
Number of cycles	10,000	100,000
a	0.04	0.03
Initial radius	10	7

many more neurons than original observations (225 > 18); conversely, observations far apart from each other will tend to imply very different cities.

In the first period (see Figure 1a), cities are fairly separated from each other, and tend to locate along the borders of the network; in fact, all but Lillesand have been mapped by the algorithm to edge neurons. Kortrijk is in the upper-left corner, and along the bottom border (left) we find Edinburgh and Norfolk; Osterholz and Bremerhaven are on the top border, and Groningen is located on the right border.

If we consider the most recent period (see Figure 1b), the picture is quite different: most cities now map into central neurons, and only four of them are located at the borders. In this case, cities are more similar and those located in the corners may be considered to be *outliers.*

The next step involves the component planes, and links the cities with the actual variables employed in the analysis, in an attempt to give an idea of the characteristics of the different regions of the SOM. In this context, we use information from the PCA from above to consider only the most relevant variables. Figure 2 displays the distribution of the 11 variables with the highest loading factor, that is, those aspects of the smart cities that most define each and differentiate them from the rest.

Figure 2 shows the component planes of the 11 most relevant variables, which are based on the result of a SOM performed using standardized data for all the cities in the sample, and for the two periods of time. This provides an output space based on all the available data that allows us to map both periods onto the same network, and thus study the evolution of the cities in the period considered. The data from both periods are pooled together to run the algorithm, and to obtain a statistical space based on all the information available (that is from both periods). In this context,

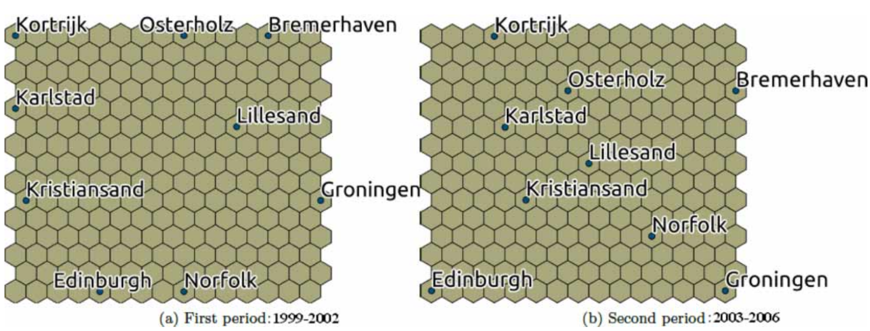

Figure 1. BMUs by time period, 1999–2002 and 2003–2006.

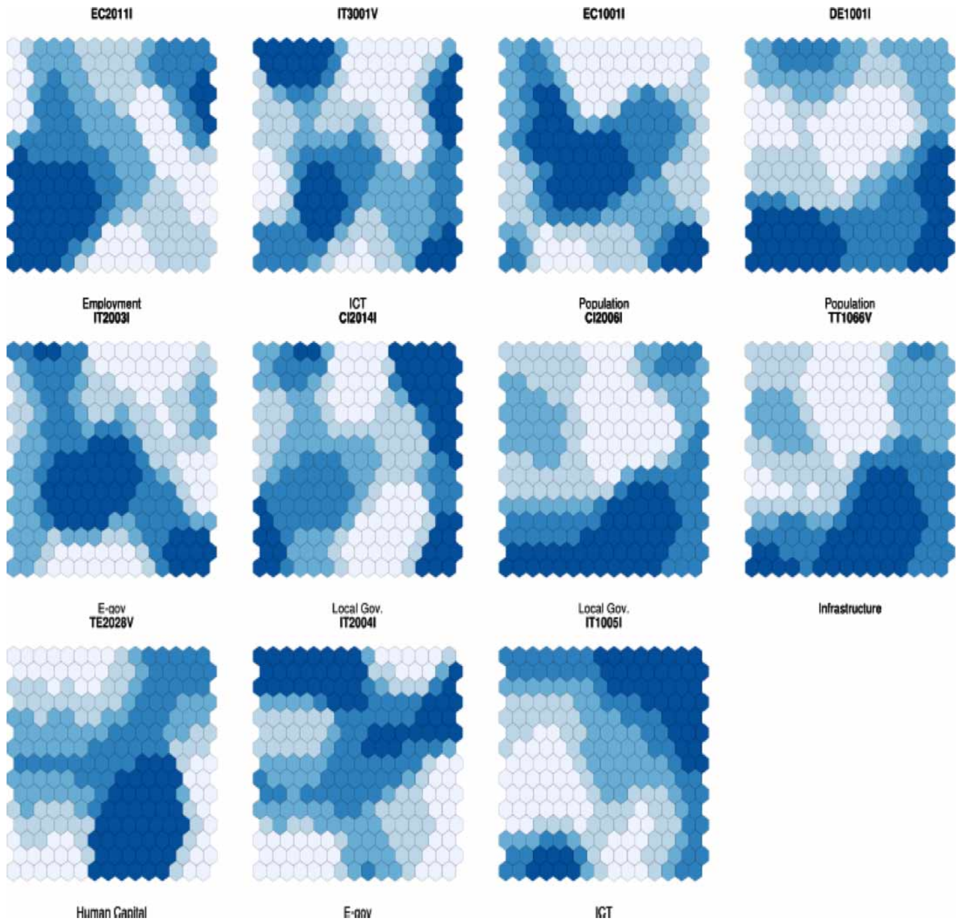

Figure 2. Component planes of the 11 most relevant variables.

every city enters twice as two different observations, one with the values of the first period and the other with those from the second one.

It is worth noting in Figure 2 there is no clear pattern that applies to all the variables. There are some, such as the annual expenditure of the municipal authority per resident (CI2006I) and the length of the public transport network (TT1066V) that show similar distributions but, overall, every variable shows a different pattern. This is a sign of the complexity of the interrelations between the various components of the *smartness* of a city, and highlights the need to consider different variables instead of trying to summarize them into one indicator, because it is likely to be missing relevant information. These graphics represent in a colored gradient of five quantiles the regions of the SOM where values are low (white) and high (dark blue; color in online version only). As an example of interpretation, we refer to the case of the annual expenditure of the municipal authority per resident (CI2006I): cities like Edinburgh show high values in both periods, and thus locate into a dark zone, while others like Kortrijk display low values, and thus may be found in the areas with lighter colors.

Finally, a note of caution is necessary in this context, since the analysis of individual variables, because of the mechanisms of the SOM algorithm, may be inaccurate in certain cases; SOM looks for global patterns where all variables are considered that work best; for particular cases, it is preferable to refer to the original data. The SOM analysis presents in a simple way the general patterns of association for the evolution of the cities over time, based on all the variables used.

Trajectories of "travelling cities"

Having looked now at how the cities are mapped into the SOM in the first and the second periods in time, the next obvious step is to consider the transitions that have occurred over these periods, and thus analyze the dynamics of the process. In comparison with the previous section, in which each period is analyzed separately, this section bridges both periods and looks at how cities have evolved and changed. The tool employed in this context is known as *trajectories*, and has already been suggested by Kohonen (2001). In this work, we follow a very similar approach to the one in Skupin and Hagelman (2005), where they analyze a set of 254 US counties over three points in time with a SOM based on the observations of each county in each year.

The main idea of a trajectory is fairly simple: to map the change over the statistical space of an observation (a city in this case); in other words, to analyze the evolution of the cities over the period analyzed. One way to do that is by plotting both the location of the city in the SOM in the first period (origin) and that in the second period (destination) and by drawing a directed line or arrow to visualize the evolution over the period of interest. By means of the trajectories it is possible to link each city back to the different component planes in order to see the behavior of particular variables, and to find out how the values have changed. The trajectories' excel shows in a transparent and visual way how the different cities have become more or less similar over the period, which ones have experienced some sort of "convergence" and which ones have "diverged" and become outliers. Since there is only one SOM that encompasses both periods, one can assign the observations of one city in each period to the same map, and look at where each of them is located (see Figure 3).

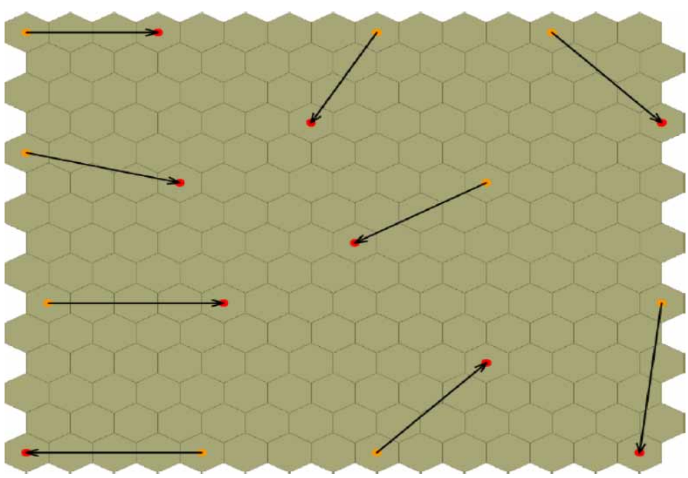

Figure 3. Trajectories of smart cites over time.

Figure 3 shows in orange the origins of the nine cities in our data set, and in red their destinations, and indicates the trajectories by drawing arrows between both (color in online version only). The first interesting aspect to note is that all trajectories are short, implying that the cities do not "travel" more than five or six neurons. This is indicative of no dramatic changes, and thus of only a slight evolution.

It is also noticeable that most cities experience a shift to the center of the SOM space, which in turn implies that the majority of them end up closer (i.e. more similar) by the end of the period than at the beginning. This is what makes the cities in the second period (see Figure 1b) appear more clustered than in the first round (this is what is called "club convergence" in the spatial disparities literature). At the same time, we may also observe that there are a few outliers that do not follow this trend, and, conversely, end up further apart from the rest. These are Edinburgh, Groningen and, to some extent, Bremerhaven.

The interpretation of these changes is that, over the period concerned, these three cities have become statistically separated from the remaining set of cities, while the other group has become more alike in a sort of convergence process. The outlier cities (Edinburgh, Groningen and Bremerhaven) have moved to places within the SOM that are located remarkably further from the rest of the cities, which, at the same time, have experienced movements from the first to the second period that have made them end up clustered around the same region. These movements across the SOM are always based on all the variables used, which is precisely why we employ a technique like the SOM, in order to reduce data sets of higher dimensions to two dimensions.

Bearing in mind what was noted in the previous subsection about the analysis of the component planes, we can also think about the implications of these "travels" of the cities over the statistical space with respect to their characteristics. One way to do that is to analyze by means of the component planes what the values are in the areas where most destinations are located (see Figure 2). As an example, we may note the case of the number of administrative forms available for download from official websites (variable "IT2003I"). A similar example, the total economically active population (variable "EC1001I"), shows that most of the cities have moved to a dark area, which implies higher levels for the values of the variables. The central/center-left area, where many of the cities are mapped in the second period, displays the highest values of the SOM; since the origins are all in areas with lighter coloring, we can see how the majority of the cities in our sample have increased their values of this "smartness" variable.

Conclusion and recommendation

Cities are operating in a dynamic and competitive world. They have to exploit their indigenous strength so as to be more distinct ("centrifugal"), but they also operate in the same global environment and have to learn from each other ("centripetal"). From this perspective, a space–time analysis of the relative position of cities is very interesting. In this paper we have considered a few European cities over a period of time that ranges from 1999 to 2006, and analyzed them using a SOM approach, examining both the static of each point in time and the dynamics involved to evolve from one period to another. The most relevant insight that can be drawn from our

SOM analysis is the idea that most cities have converged in the period analyzed, becoming more similar over time. At the same time, we found that a few of them became outliers. The main advantage of this study compared with others that apply more traditional techniques is the ability to bring together all the information from several variables in two points in time and visualize it in an intuitive way that can allow for pattern extraction and substantive knowledge creation.

Acknowledgment

The authors wish to thank Chiara del Bo and Andrea Caragliu for their assistance in the data collection.

Notes

1. Published by the Centre of Regional Science, Vienna University of Technology, Vienna, October 2007 (www.smart-cities.eu).
2. For more details on the latter network, we refer to www.smartcities.info
3. This section briefly explains the general idea and functioning of the basic SOM algorithm. For a complete and rigorous treatment of the SOM methods, the reader is referred to Kohonen (2001).
4. For a bibliography of papers using the SOM algorithm, see Oja *et al*. (2003).

References

Abreu, M., Grinevich, V., Kitson, M., and Savona, M., 2008. *Absorptive capacity and regional patterns of innovation*. Research Report DIUS RR-08-11. Cambridge, MA: MIT.
Acs, Z.J., de Groot, H.L.F., and Nijkamp P., eds., 2002. *The emergence of the knowledge economy, a regional perspective*. Berlin: Springer.
Arribas-Bel, D., Nijkamp, P., and Scholten, H.J., 2011. Multi-dimensional urban sprawl in Europe: a self-organizing map approach. *Computers, environment and urban systems*, 35 (4), 263–275.
Audretsch, D.B., 1998. Agglomeration and the location of innovative activity. *Oxford review of economic policy*, 14 (2), 18–29.
Aydalot, P., 1986. *Milieux innovateurs en Europe*. Paris: Gremi.
Benner, C., 2003. Learning communities in a learning region: the soft infrastructure of cross-firm learning networks. *Environment and planning A*, 35, 1809–1830.
Berry, C.R. and Glaeser, E.L., 2005. The divergence of human capital levels across cities. *Papers in regional science*, 84 (3), 407–444.
Bolisani, E. and Scarso, E., 2000. Electronic communication and knowledge transfer. *International journal of technology management*, 20, 116–133.
Camagni, R., ed., 1991. *Innovation networks: spatial perspectives*. London: Pinter.
Caragliu, A., Del Bo, C., and Nijkamp, P., 2011. Smart cities in Europe. *The journal of urban technology*, 18 (2), 65–82.
Coe, A., Paquet, G., and Roy, J., 2001. E-governance and smart communities: a social learning challenge *Social science computer review*, 19 (1), 80–93.
Cohen, W. and Levinthal, D., 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly*, 35 (1), 128–152.
Cohen, J. and Paul, C.J.M., 2005. Agglomeration economies and industry location decisions. *Regional science and urban economics*, 35, 215–237.
Cooke, P., 1998. Introduction: origins of the concept. *In*: J-J. Braczyk, P. Cooke, and M. Heidenreich, eds. *Regional innovation systems. The role of governances in a globalized world*. London: UCL Press, 2–25.
Deakin, M., 2010. SCRAN: the smartcities (inter) regional academic network supporting the development of a trans-national comparator for the standardisation of e-government

services. *In*: C. Reddick, ed. *Comparative e-government: an examination of e-government across countries*. Berlin: Springer, 425–446.
Fischer, M.M., 2001. Computational neural networks: tools for spatial data analysis. *In*: M. M. Fischer and Y. Leung, eds. *Geocomputational modelling: techniques and applications*. Heidelberg: Springer, 79–102.
Florida, R., 1995. Towards the learning region. *Futures*, 27 (5), 527–536.
Florida, R., 2002. *The rise of the creative class*. New York: Basic Books.
Frenken, K., Oort F. van, and Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. *Regional studies*, 41 (5), 685–697.
Gertler, M., 2003. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of economic geography*, 3, 75–99.
Gertler, M. and Wolfe, D.A., eds., 2002. *Innovation and social learning: institutional adaptation in an era of technological change*. Basingstoke: Palgrave Macmillan.
Glaeser, E.L. and Berry, C.R., 2006. *Why are smart places getting smarter?* Cambridge, MA: Taubman Centre Policy Brief 2006-2, Taubman Centre.
Glaeser, E., Kolko, J., and Saiz, A., 2001. Consumer city. *Journal of economic geography*, 1 (1), 27.
Graham, S. and Marvin, S., 1996. *Telecommunications and the city: electronic spaces, urban place*. London: Routledge.
Holland, G., 2008. Will the real smart city please stand up?, *Cities*, 12 (3), 303–320.
Howells, J.R.L., 2002. Tacit knowledge, innovation and economic geography. *Urban studies*, 39, 871–884.
Jacobs, J., 1961. *The death and life of great American cities*. New York: Random House.
Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43 (1), 59–69.
Kohonen, T., 2001. *Self-organizing Maps*. 3rd edn. Berlin: Springer.
Kohonen, T. and Honkela, T., 2007. Kohonen network. *Scholarpedia*, 2 (1), 1568.
Komninos, N., 2008. *Intelligent cities and globalisation of innovation networks*. London: Taylor & Francis.
Kourtit, K., Nijkamp, P., Lowik, S., van Vught, F., and Vulto, P., 2011. From islands of innovation to creative hotspots. *Regional science policy and practice*, 3 (3), 145–161.
Kraaijveld, M.A., Mao, J., and Jain, A.K., 1992. *A non-linear projection method based on Kohonen's topology preserving maps*. ICPR-92: Proceedings of the 11th IAPR International Conference on Pattern Recognition, 2.
Landry, C., 2003. *The creative city*. London: Earthscan.
Lee, L., Tan, X., and Trimmi, S., 2006. Current practices of leading e-government countries. *Communications of the ACM*, 48 (10), 100–104.
Maillat, D., 1991. Local dynamism, milieu and innovative enterprises. *In*: J. Brotchie, M. Batty, P. Hall, and P. Newton, eds. *Cities of the 21st century*. Harlow, UK: Longman, 265–274.
Maskell, P. and Malmberg, A., 1999. Localised learning and industrial competitiveness. *Cambridge journal of economics*, 25, 167–185.
Morgon, K., 2002. The new regeneration narrative: local development in the multi-level polity. *Local economy*, 17 (3), 71–98.
Nijkamp, P., 2008. *E pluribus unum*. Research Memorandum, Faculty of Economics, VU University, Amsterdam.
OECD, 2001. *Cities and regions in the new learning economy*. Paris: OECD.
Oja, M., Kaski, S., and Kohonen, T., 2003. Bibliography of self-organizing map (SOM) papers: 1998–2001 addendum. *Neural computing surveys*, 3 (1), 1–156.
Paskaleva-Shapira, K., 2008. Assessing local e-governance in Europe. *International journal of electronic governance research*, 4 (4), 17–36.
Poelhekke, S., 2006. *Do amenities and diversity encourage city growth? a link through skilled labor*. Economics Working Papers ECO2006/10. San Domenico di Fiesole: European University Institute.
Ratti, R., Bramanti, A., and Gordon, R., eds., 1997. *The dynamics of innovative regions – the GREMI Approach*. Aldershot, UK: Ashgate.
Rosenthal, S.S. and Strange, W.C., 2001. The determinants of agglomeration. *Journal of urban economics*, 50, 191–229.

Sassen, S., 1991. *The global city*. Princeton, NJ: Princeton University Press.
Shapiro, J.M., 2008. Smart cities: quality of life, productivity, and the growth effects of human capital. *The review of economics and statistics*, 88 (2), 324–335.
Skupin, A. and Hagelman, R., 2005. Visualizing demographic trajectories with self-organizing maps. *GeoInformatica*, 9 (2), 159–179.
Southampton City Council, 2006. *Southampton on-line*. Available from: http://www.southampton.gov.uk/thecouncil/thecouncil/youand-council/smartcities/ [Accessed March 13, 2009].
Torres, L., Vicente, P., and Basilio, A., 2005. e-Government developments on delivering public services among EU cities. *Government information quarterly*, 22, 217–238.
Ward, B., 1976. *The home of man*. New York: Norton.