Going Beyond the Euclidean Setting in the Statistical Analysis of Human Movement in Urban Landscape

Anna Calissano[1], Paola Sturla[2], Paola Pucci[3], Valeria Fedeli[3], Simone Vantini[4]

[1]Politecnico di Milano, MOX Dept. of Mathematics, Milano/Italy · anna.calissano@polimi.it
[2]Harvard University GSD, Massachusetts/USA
[3]Politecnico di Milano, Dept. of Architecture and Urban Studies, Milano/Italy
[4]Politecnico di Milano, MOX Dept. of Mathematics, Milano/Italy

Abstract: The statistical analysis of human movement within the built environment has always been framed in a Euclidean setting. Due to the intrinsic complexity of the urban landscape and how humans interact with it, reductionism is an element that should be taken carefully into account. Extending the mathematical tools to studying complex data and trying to avoid reductionism in the mathematical representation of the problem is the first step to conduct a meaningful and less reductive quantitative study. Object-Oriented Data Analysis (OODA) is a novel statistical approach aiming at going beyond the classic (and sometimes reductive) Euclidean framework. OODA aims at describing the natural complexity of a type of datum such as graphs, text, trajectories, defining a natural geometrical framework for the datum.

Keywords: Non-Euclidean statistics, Object-Oriented Data Analysis, human movement

1 Introduction

Human Movement in urban studies refers to the analysis of how individuals move in a built environment (KEYFITZ 1973). In our era, data concerning movement can be gathered from a variety of sources, and these data could reveal exciting insight into the phenomenon of movement of individuals in environments. These data are complex both in terms of their nature (they can often have the form of a graph, a trajectory, a text, an image, a video), and in terms of the phenomenon they describe and in terms of the mathematical tools required to analyse them. However, that new eclectic source of digital data has a complex intrinsic nature. In our opinion, rather than "big", these types of data should be referred to as "complex." The majority of traditional statistical tools are too reductive to embrace such complexity. In this paper, we are discussing the possibility and the potential gain in studying spatial phenomena with complex data in a non-Euclidean setting. We suggest the usage of non-Euclidean statistics, also referred to as Object-Oriented Data Analysis, as a possible technique to avoid reducing a complex phenomenon such as the human movement in an urban landscape. Non-Euclidean geometry has been a successful tool for describing reality in other disciplines like physics. We are showing how a "non-spatial" mathematical setting could reveal exciting and unexpected insights into a spatial phenomenon. This paper aims to discuss if the measurement of the urban landscape and human movement should always be performed using a Euclidean setting, showing the pro and cons of both Euclidean and non-Euclidean settings through a simulation exercise. As for relativity theory, landscape theory should ask if the measurements are performed in the most suitable space.

Journal of Digital Landscape Architecture, 5-2020, pp. 150-157. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-690-1, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537690016.

2 Human Movement

Human Individual Movement was defined as "how individual humans move within a network or a system" (KEYFITZ 1973). Understanding how individuals move in a specific environment is an intrinsically valuable research question, potentially revealing essential information for planners, policymakers, urban designers, sociologists etc. (TOOLE et al. 2012, WANG et al. 2012). Human individual movement can be studied at a different scale in space, according to the selected scale of the system. Focusing on the human-scale (GIAMPIERI et al. 2017), human movement can be framed in the concept of "Public Life" by GEHL (2011), referring to citizens' "daily interactions with others within the built environment". The human movement in a square, a park, or a pedestrian cross could suggest to planners and policymakers' insights on the usage of the public space (GEHL & SVARRE 2013, RIVA et al. 2019). The analysis of the human movement in urban space can be conducted in a different methodological way. The problem can be tackled using a qualitative, quantitative, or mixed approach (LORD et al. 2011, GONZALEZ et al. 2008, GOETZ et al. 2009). It is not the aim of this work to open a discussion about whether it is better to use one approach or the other to dig into the complexity of this problem (see GOETZ et al. 2009 for a discussion). Among all the developed quantitative approached to understand human movement, we focus on statistics. The reasons are mainly two: the increasingly available data sources concerning human movement, and the traditional role of statistics, as a mathematical field design to study complex phenomena, not easily described by deterministic laws.

3 Statistical Analysis of Human Movement

From 2011, the word big data started spreading in both industry and research environment without a shared clear definition of the concept (BOYD & CRAWFORD 2012). Such data flow gives rise to a new era of digital positivism (FUCHS 2017), and consequential post-positivism movements. Scholars, including MARRON & ALONSO (2014), have started interpreting this stream of data not only in terms of velocity, variety, and volume but in terms of complexity. Complex data are the one scholars are gathering to study human movement and perception in the urban landscape such as text, biometric data, trajectories and graphs (e. g., CALISSANO et al. 2019, ROSS 2019). So, we should look at these types of data in this peculiar applicative context. The analysis of complex data poses serious methodological challenges concerning analytical methods for different applications. Complexity resides not only in the intrinsic nature of the data but also in the methodological complexity required to mine those data and in the analysis of the relation between the complex data and the complex system the data are gathered from. When dealing with complex system analysis, reductionism could be a common mistake (MANSON & O'SULLIVAN 2016). There is a thin line between the holistic attempt in describing a complex system (through complex data), the emergence phenomenon, and the reductionism of the analysis conducted. In analysing a complex system, scholars could default to a methodological reductionism, which is the scientific attempt to explain in terms of ever-smaller entities. Methodological reductionism has a deep root in the statistical field and it is strongly related to the concept of sufficience. If we are interested in describing a specific phenomenon with certain summarising statistics, the estimation of these statistics is going to be exhaustive if we have represented the phenomenon with a sufficient amount of information. For example, if we are interested in how inhabitants move around the square, sampling one position for each individual every day is an under-representation of the phenomenon, as well as measuring the microsecond movement with sensors on the legs of each individual are an over-representation of the phenomenon. The choice of the representation of a problem is a key and first step in the quantitative analysis because it frames the problem and the analysis in "space". The sufficient representation should not only be related to the granularity of the information we gathered but also with the type of geometrical embedding we chose, and/or the summarized statistics we used. The question we would like to address is: Which is the sufficient geometry to use when studying a complex phenomenon as the human experience of a landscape? Which are dimensions involved when we depart from the 3D Euclidean space? What is our agency, as an interdisciplinary team of statisticians and designers, in studying such geometry of reality?

3.1 Geometry and Complex Data

The notion of how to measure space is thousands of years old. First introduced by the Egyptians to measure field dimensions, the theoretical development of the geometric theory as a set of axioms and theorems was formalized by Euclide in the 4th century BC. In thousands of years, scholars used Euclidean geometry as a "comfortable" mathematical framework to measure reality such as space and shape (NAGEL 1961). Alternative geometries (rather than Euclidean geometries) such as Hyperbolic Geometry, Elliptic Geometry, Projective Geometry, have only been developed as recently as the 19th century. As in the analysis of human movement, the Euclidean space has been considered as the unquestionable natural framework to study space phenomena, the same happened in other disciplines such as physics (e. g., Newton considered geometry as a simple tool to study physics). However, choosing a suitable geometry is crucial, and it strongly influences the results of an analysis. An example is Einstein's relativity theory, which is based on the assumption that space is a hyperbolic space, and all the computations are performed according to this assumption. As shown by this somewhat relevant example, non-Euclidean geometries can be adopted to measure space with successful outcomes. In the statistical field, we refer to the problem of framing the correct geometry where complex data should be analysed as Object-Oriented Data Analysis (OODA, WANG & MARRON 2007). OODA approach questions the Euclidean setting as a "give for granted setting" and opens a discussion about which geometrical framework should be used. This approach enables the analysis of a variety of complex statistical objects, such as curves, shapes, images, trees, networks, and text.

3.2 Methodology

In this work, we introduce the OODA approach in the specific setting of the analysis of the population of graphs. See MARRON & ALONSO (2014) for a broader overview. Graphs are a mathematical representation of relational phenomenon such as transport infrastructure, social activities, chemical reaction etc. A graph is made of a set of nodes and a set of edges representing relations among nodes (e. g., in a subway network, nodes are the stations and edges are the train ways). Relational representation is also common in landscape analysis. Within the design disciplines, such kind of representation is called diagrammatic, and has been widely applied both in architecture, landscape architecture and urban planning over the past 50 years (e. g. BATTY 2004). One example of such diagrammatic representation is the one obtained through the Space Syntax method (e. g. HILLIER et al. 1987). When dealing with a population of graphs, or diagrams, every graph could have a different number of nodes or different node labels, so a geometrical framework where graphs are comparable is required. The framework proposed here is Graph Space (JAIN et al. 2009, CALISSANO et al. 2020), a space where every element is an equivalent class of graphs obtained by permuting the nodes. This space is not an Euclidean space; neither is it locally comparable to it (i. e., it is not a manifold). In the next section, we are going to show the analysis of a set of human movement graphs in Piazza Leonardo da Vinci (Milan). The analysis consists of clustering similar usage of the square using Graph Space and not the actual geographical space to perform the analysis. In order to perform cluster analysis, a concept of distance between objects should be defined in the Graph Space as details explained in CALISSANO et al. (2020). Once the pairwise distance is computed, closer graphs are grouped with a hierarchical clustering procedure.

4 Case Study

Leonardo da Vinci Square is a relevant testing ground for the scope of this research. The square, adjacent to the main entrance of Politecnico di Milano was redesigned in 2018 to foster pedestrian circulation by eliminating parking lots and vehicular paths. The resulting public realm supports spontaneous gatherings, and pedestrian and bicycle crossings. Because of its geometry, it allows for installing fixed and temporary features like benches, kiosks, pavilions, and stages to support cultural events. The existing grasslands have been extended and reshaped to fit the spontaneous pedestrian desire lines defined by PoliMi students crossing them. Finally, the new lighting design expands the usability of the plaza by enabling evening activities and safety.

Fig. 1:
Piazza Leonardo da Vinci, Milano, 2011. Note the pedestrian desire lines in the grasslands (Image: Google Earth Pro)

Fig. 2:
Piazza Leonardo da Vinci, Milano, 2020 (Image: Google Earth Pro)

One year after completion, the new layout has shifted the practices of use of the plaza. The use of space varies throughout the day. In the early morning, students, professors and scholars from the two universities located in the area (Politecnico and Università Statale) cross the space from the bike-sharing stalls and public transportation stops (subway, tramways, and buses) to the universities main gates. At the same time, the users of the two primary public services on site, the BESTA neurology hospital and the Italian National Institute for Tumors hospital, cross the space in the same directions. Around 11 AM, residents walk around for leisure. During the lunch break, people buy street food at the numerous kiosks on site and eat seating on the benches or the grass. In the evening, space is quieter but still used because of a theater and the library of the PoliMi School of Architecture, which is open until midnight. Figure 1 and 2 show the view of the square before and after the redesign. The pedestrian-bikes shared surface, together with the lack of hard paths built in the grasslands, makes Leonardo da Vince Square a porous and flexible space, allowing for spontaneous use of the space, avoiding impositions by design. That is the reason why the case is relevant to this research.

4.1 Data and Analysis

Following the square description usage previously described, three different types of square usage are simulated: (1) simple square crossing represented as a graph with two nodes (the entry point and the exit point of the square) and one link representing the trajectory; (2) seating in the square, represented as a graph with three nodes (the entry point, the seat, and the exit point) linked with three trajectories; (3) Kiosk, represented as a graph with four nodes (the entry point, the seat, the Kiosk, and the exit point) linked with four trajectories.

In Figure 3, we show the entrances and the kiosk. All the seats are parametrized according to both the real branches designed in the square and the steps in front of the university, where students frequently appropriate as seating. In Figure 3, all the generated trajectories are shown on the map.

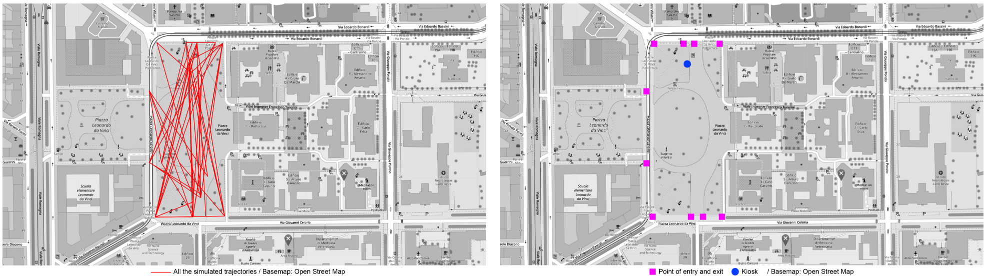

Fig. 3: The simulated trajectories in Piazza Leonardo da Vinci. Each trajectory is simulated as a graph with an entry and an exit point, and eventually a seating point and a break at the Kiosk. Entry points, exit points, and kiosks are also shown (Image: OpenStreetMap).

To analyse them, we extrapolate geo-referenceed data and embed them in Graph Space, where we compute the distance between graphs. Figure 4 shows three graphs disassociated from the geographical space. In a real dataset, geolocalised data are sampled. There are several strategies to build graphs from trajectories. Here we interpret: (1) as graph nodes both the first and last measurement of every trajectory (i. e. the enter and the exit points) and all the points where the users stop for more than a certain time; (2) as graph edges the presence or absence of the trajectory between nodes. Different encodings are possible, as well as different attributes choice (e. g. the velocity of the trajectory, the stopping time).

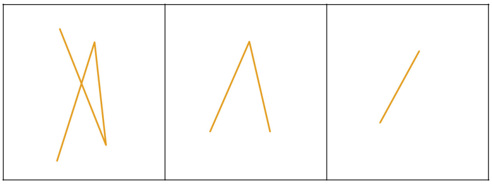

Fig. 4:
Conceptual representation of the three observations in Graph Space

4.2 Results Interpretation

As shown in Figure 5, trajectories are correctly clustered using Hierarchical Clustering methodology (ward linkage). The cluster of single square crossing (top), a cluster of people sitting in the square (central), and the cluster of people enjoying a break at the kiosk (bottom) The results show how the methodology can cluster the simulated data, thereby demonstrating the potential of the Graph Space approach. Also, the methodology can deal with unconventional paths. For example, in the top cluster in Figure 5, we can see two paths that degenerate to a single point, where the pedestrian enters and exits the square at the same point without even crossing it. From the point of view of the designer, the not Euclidean graph could suggest unexpected, emergent uses of space, which can inform design in terms of paths, pedestrian shared surfaces and urban furniture location. The multidimensional nature of the Graph, not literally linked to a Euclidean 2D geometry, provides the designer with insights without forcing a layout, enabling interpretation and critical thinking in the workflow. The OODA produces a sort of diagram of the space, stressing on the relational aspects of the plaza inhabitation.

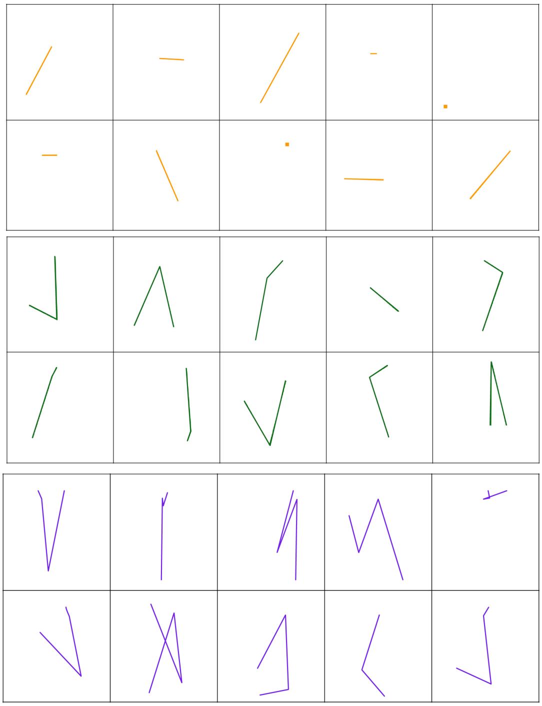

Fig. 5:
All the square graph trajectories are shown,
divided in clusters

Cluster 1:
trajectories of simple square cross

Cluster 2:
trajectories of people crossing and seating in the square

Cluster 3:
more complex trajectories of people seating in the square and going to the kiosk

5 Conclusion

This paper aims to discuss if the measurement of the urban landscape and human movement should always be performed using a Euclidean setting. We introduced human movement and we discussed how the phenomenon could be studied with "complex" data more than big data. Human movement data could be modelled as graphs, trajectories or images, which are complex. They are complex in terms of the mathematical tools required to analyse them. We then introduce the statistical theory known as Object-Oriented Data analysis. This theory aims at understanding the natural geometrical framework where complex data should be embedded. As an example, we focused on the graph representation of trajectories describing human movement in a square in Milan: Piazza Leonardo da Vinci. To show the potential of this framework, we simulated some simple trajectories of pedestrian crossing the square, seating in the square, and having a break at the kiosk in the square. Each trajectory is then mathematically represented as a graph and embedded into the not Euclidean Graph Space. Cluster analysis is performed, showing how the framework is able to cluster the simulated data, recognizing the three different types of usage. The simulation serves as a simple example of how the not Euclidean setting can show exciting results about a geographical dataset. The presented framework could be easily extended to another type of data generated by users interacting with the landscape. It could be also easily enriched considering other variables such as the time evolution of the square usage. The resulting Graphs, or diagrams, provide the designer with insights on potentially unexpected practices of use of the plaza, without generating an Euclidean 3D geometry, enabling interpretation as part of the design workflow.

References

BATTY, M. (2004), A new theory of space syntax. Centre for Advanced Spatial Analysis UCL.

BOYD, D. & CRAWFORD, K. (2012), Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15.5 (2012), 662-679.

CALISSANO, A., ARENA, M. & VANTINI, S. (2019), Monitoring rare categories in sentiment and opinion analysis: A Milan mega event on Twitter platform. Statistical Methods & Applications, 1-26.

CALISSANO, A., FERAGEN, A. & VANTINI, S. (2020), Graph Space: Geodesic Principal Components for a Population of Network-valued Data, MOX Report 2020.

FUCHS, C. (2017), From digital positivism and administrative big data analytics towards critical digital and social media research. European Journal of Communication, 32 (1), 37-49.

GEHL, J. (2011), Life between buildings: using public space. Island Press, USA.

GEHL, J. & SVARRE, B. (2013), How to study public life. Island Press, USA.

GIAMPIERI M. A., WILLIAMS, S., XU, W., GUNC, H., LANGSTON C. M. & HARVEY D. C. (2017), Benchmark: using sensors to study public space. Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings, 17, 22.

GOETZ, A. R., VOWLES, T. M. & TIERNEY, S. (2009), Bridging the qualitative-quantitative divide in transport geography. The Professional Geographer, 61 (3), 323-335.

GONZALEZ, M. C., HIDALGO C. A. & BARABASI, A. L. (2008), Understanding individual human mobility patterns. Nature, 453 (7196), 779.
HILLIER, B., HANSON, J. & GRAHAM, H. (1987), Ideas are in things – an application of the space syntax method to discovering house genotypes. Environ Plan b, 14 (4), 363-385.
JAIN, B. & OBERMAYER, K. (2009), Structure spaces. Journal of Machine Learning Research, 10, 2667-2714.
KEYFITZ, N. (1973), Individual mobility in a stationary population. Population Studies, 27 (2), 335-352.
NAGEL, E. (1961), The Structure of Science. American Journal of Physics.
LORD, S., DESPRES, C. & RAMADIER, T. (2011), When mobility makes sense: A qualitative and longitudinal study of the daily mobility of the elderly. Journal of Environmental Psychology, 31 (1), 52-61.
MANSON, D. & O'SULLIVAN, D. (2006), Complexity theory in the study of space and place. Environment and Planning A, 38 (4), 677-692.
MARRON, J. S. & ALONSO, A. M. (2014), Overview of object-oriented data analysis. Biometrical Journal, 56 (5), 732-753.
OPENSTREETMAP (2015), Retrieved from https://planet.openstreetmap.org.
RIVA, P., STURLA, P., CALISSANO, A. & VANTINI, S. (2019), Landscape Perception through Complex Data: Exploring George Hargreaves's Queen Elizabeth Olympic Park in London. Journal of Digital Landscape Architecture, 4-2019, 42-56.
ROSS, I. (2019), Finding 'A Space for Being' at Salone Del Mobile in Milan. Google Keyword. https://blog.google/technology/design/a-space-for-being-salone-del-mobile-milan.
TOOLE, L., ULM, M., GONZALEZ, D. & BAUER, M. C. (2012), Inferring land use from mobile phone activity. Proceedings of the ACM SIGKDD international workshop on urban computing, 1-8.
WANG, H. & MARRON, J. S. (2007), Object-oriented data analysis: Sets of trees. The Annals of Statistics, 35 (5), 1849-1873.
WANG, P., HUNTER, T., BAYEN, A., SCHECHTNER, K. & GONZALEZ, M. C. (2012), Understanding road usage patterns in urban areas. Scientific Reports, 2, 1001.