090

Modelling dependency networks to inform data structures in BIM and smart cities

Kinda Al Sayed
Bartlett School of Architecture, UCL
k.sayed@ucl.ac.uk

Mark Bew
UK Government Department of Business, Innovation and Skills
mark.bew@ecstrategies.co.uk

Alan Penn
Bartlett School of Architecture, UCL
a.penn@ucl.ac.uk

Dan Palmer
BSI Group
dan.palmer@bsigroup.com

Tim Broyd
Civil, Environ & Geomatic Eng, UCL
tim.broyd@ucl.ac.uk

Abstract

The pervasive deployment of "smart city" and "smart building" projects in cities world-wide is driving innovation on many fronts including; technology, telematics, engineering and entrepreneurship. This paper focuses on the technical and engineering perspectives of BIM and smart cities, by extending building and urban morphology studies as to respond to the challenges posed by Big Data, and smart infrastructure. The proposed framework incorporates theoretical and modelling descriptions to verify how network-based models can act as the backbone skeletal representation of both building and urban complexity, and yet relate to environmental performance and smart infrastructure. The paper provides some empirical basis to support data information models through building dependency networks as to represent the relationships between different existing and smart infrastructure components. These dependency networks are thought to inform decisions on how to represent building and urban data sets in response to different social and environmental performance requirements, feeding that into void and solid descriptions of data maturity models. It is concluded that network-based models are fundamental to comprehend and represent the complexity of cities and inform urban design and public policy practices, in the design and operation phases of infrastructure projects.

Keywords

Space syntax, BIM, smart cities, future cities, urban planning, policy.

1. Introduction

Research on the science of buildings and cities is witnessing an upsurge, mainly triggered by the pervasive use of smart technology applications and connected devices. The effective use of these technologies in built assets is expected to cast significant impact on the Digital Economy in Great Britain, and in many other countries. These new trends, enabled by the Internet of Things (IoT) gave rise to Big Data. Urban and architectural modelling theories need yet to respond to these trends, in order to provide an adaptive basis for understanding and making interventions in technology, telematics, and in modelling information flows and data sharing interactions.

There is a vast amount of data that are made available through technology. Yet, there is no comprehensive regulatory framework by which different types of data can be grouped and organised in response to performance requirements. On a building scale, Building Information modelling (BIM) schemes often account for the solid built "atomic" elements and their associated supply and operational infrastructure. Where there are "abstract" void descriptions (Jeong and Ban, 2011a, 2011b), they need to be organised and systemised to relate to social, cognitive and behavioural performance criteria (Schultz and Bhatt, 2011). One could argue that, with a structured and semantic data (LOD) the best possible "fidelity" of any output would be proportional to the lowest quality data. There is therefore a need for structuring information about the built form in such a way as to improve on delivering performance indicators. A proposition for a network description of built spaces that is perhaps associated or complemented by a shape description might hence be sensible in this context. A combined spatial and shape descriptions of the built environment that are compatible with and complementary to energy and lighting performance requirements would enable forecasting user behaviour and comfort during the design stage. The key issue is to really outline the set of performance requirements for buildings, hence find the reduced set of variables and parameters that are essential for analysing and forecasting the performance of built assets. There is also a need to identify a priority structure for different performance criteria depending on what is essential for a building to function and what would improve the comfort of the built environment.

This task is perhaps easier on an urban scale, since cities –unlike buildings- settle into universal patterns and trends (Bettencourt et al, 2007), and embody a well-defined generic function; that is to optimise journeys from all origins to all destinations (Hillier, 1996). The complexity aspect of cities might rather be in the multiplicity of actors and factors and how they contribute individually and collectively to the emergence and planning of built form (Allen, 1997). A prioritised structure model of urban information was proposed in (Al_Sayed, 2014c) for aiding urban design. It was suggested there that spatial accessibility of street networks needs to be prioritised to predict other spatially dependent variables (i.e. Building height, street width, block density, land use distributions and property values), moving forwards to predicting energy and lighting performance of urban form.

The notion of "Smart" was often perceived as an additional technical layer that is overlaid on top of an existing urban infrastructure, to enable new forms of operation and use to urban space. This layer is equipped by telematics to enable new business models in response to specific city problems, such as the rise of unemployment in certain sectors, the decline in local retail markets, gentrification and others. There are some issues, however, that raise concern with this type of adaptation of technology. The fact that the smart layer is not integral to existing urban infrastructure, might implicate that this layer will be secluded from the internal dynamics of the urban system on the long term, and will eventually become incompatible with its urban context and alien to its nature. This will all be at a higher cost rather than greater value. Hence, there is a need to make smart urban systems more integral to the existing infrastructure, perhaps learning from dependencies between different urban variables in order to enable a sustainable open market of analytics.

In response to these challenges, this paper aims to outline key aspects of the nature of building and urban dependencies, in an effort to reach a theoretical framework and representational scheme that would respond to the majority of performance requirements of buildings and cities. We propose establishing a network-based approach as the backbone for smart infrastructure in building and urban contexts. An integrative network-based model can be used to enhance the design and test and simulate the operation of smart infrastructure. At the fundamental level, and as a step forwards to find that integrative model there is a need to understand the network of relationships that characterize; how the physical infrastructure and its configurations relate to different types of performance criteria and how a smart infrastructure corresponds to performance requirements.

To answer this question, this paper addresses first the need to extend the theoretical and modelling frameworks of building and urban morphology to respond to the challenges posed by Big Data. The paper also addresses how space syntax might be used to incorporate social performance indicators in BIM and smart city frameworks, through establishing a relationship between the configurations of built form and social structures (Hillier and Hanson, 1984; Hillier, 1996). Together with environmental performance, space syntax is set to close the loop between post-occupancy operation phase of the building life cycle and the design phase, through the use of evidence-based empirical models to inform architectural and urban design. The paper then reviews previous research on building and urban analytics that incorporated configurational variables. Following that, a methodology for visualising dependency networks is introduced, and used to reveal the relationships between building and urban spatial components and other performance criteria. The final section reflects on all that by making some propositions on how to integrate frameworks and incorporate empirical models of dependency networks as to inform data information models and public policies.

2. The need to extend existing frameworks to respond to Big Data

Decades ago, theoretical frameworks and predictive models were mostly dependent on the need to predict more variables by using very little data. Space Syntax was one of these theoretical and modelling frameworks, to reason about the relationship between physical space and society; between the built environment and human behaviour. In Space Syntax, urban systems were often represented as networks of streets, where accessibility of the street network infrastructure was considered as the main predictor for other urban variables such as; pedestrian, and vehicular movement, retail land use, property prices, indices of socioeconomics. In the meanwhile, the positivist approach in geography was more focused on the utilisation of agent-based models in urban simulations (Batty, 2013). Long term transformations and challenges were traditionally targeted in both the analytical and modelling approaches. This trend in built environment research has recently diverted towards thinking and modelling short-term challenges induced by the era of Big Data and the Internet of Things (IoT).

The need for predictive models that can predict behaviour based on a measurable set of variables might be seen as no longer valid considering the number of open data sources that are becoming available to communities and stakeholders. In the advent of real-time information flows enabled through sensors and social networking websites, it is often argued that the availability of these types of data along with the more traditional data sources will make it easier to forecast the operational performance in the built environment more accurately and on a temporal basis. Researchers at CASA[1] for example, are making the argument that we no longer need to estimate pedestrian flows in an urban setting because we can infer it from Oyster card data[2]. In the light of such arguments, there is a need to reconsider the need for the orthodox explanatory descriptions of urban space, by accounting for the temporal dimension, and the dynamics underlying human behaviour in buildings and cities.

3. Space Syntax use in design, construction and operation of assets

Construction contributes with 90BN to the UK economy (6.7% of the national GDP) (2013). About 10% of the UK population works in the construction sector. Construction 2025 set targets of 33% lower costs, 50% faster delivery, 50% lower emissions, 50% improvement in exports[4]. In the current government strategy, and due to funding restrictions, the delivery of BIM level'2, level'3 and level'4 strategies needed to be separated (Bew & Underwood, 2009). Ideally, this separation should have been avoided, particularly in what concerns the link to human behaviour in built assets and the

[1] UCL Centre for Advanced Spatial Analysis
[2] Source: http://simulacra.blogs.casa.ucl.ac.uk/2012/05/pulse-of-the-city-reboot/

cultural aspects of smart buildings and smart cities, but the need to effectively communicate with the 3M people involved in the industry a managed migration was seen as essential.

To complement the vision for data analytics, there is a need to attend to the value of space syntax as a configurational performance measure for social behaviour, and also as an indicator for economic performance. Through improving the performance of space and optimising access and vision, it might be argued that space syntax can have a major role during the operational phase of a building.

The model Mark Bew developed for BIM[3], considers a data structure that provides a comprehensive account for the issue of interoperability throughout the design, construction, and operation stages (figure 1). It takes into consideration the technical, commercial, and cultural aspects of data sharing infrastructure. We will be focusing mostly on the design and operation of built assets, seeking explanatory descriptions of performance and operation requirements to suggest that these requirements could be incorporated in the design stage to improve on the performance of building throughout the whole life-cycle, or else better adapt smart systems to fit with existing infrastructure and improve its functioning performance and durability.

The Level 3 programme was built on this delivery and operational process, and introduced the concept of measurement. This opened the potential for consistent and continual review of asset performance with respect to designed or desired state. In the original proposition made by Bew, the design and operation of buildings needed to account for dependency analytics, in order to better outline the performance requirements of an infrastructure. From these requirements stems the relationship between existing building infrastructure and the smart systems that improve its function. It is usually argued that performance requirements might be mined from smart building and smart city projects. However, in order not to be limited to current descriptions of smart infrastructure projects, there is a need to go beyond the smart layer to reveal intelligent relational descriptions in the physical built environment, and perhaps expose how the smart layers might be integrated with building and urban infrastructure to improve their overall performance. More importantly, there is a need to find a universal spatial representation and description that allows for handling the layout complexity without having to review details of building and urban layouts. In this context, an adapted form of space syntax description might be used as a universal description for layout structures (figure 2), but need to be also complemented with a description of solid surfaces that envelope spaces. This is mainly to do with the impact of data fidelity when dealing with more than one performance criteria. For example, if we have a complete data model of the built environment, where BIM values are filled in along with Space Syntax values (figure 3), the values for each component will fall into the same attribute and entity positions with different provenance. The analytics devised to measure performance will need to be adapted to provide a tolerance of error value so the user can interpret potential uses of data analytics.

For the purpose of building a universal and integrative framework that brings together BIM and smart cities, there is a need to outline an extended network-based representation, accounting for the dependencies between different layout attributes and the temporal, operational and economic dimensions. There is also a need to define what Space Syntax means for the engineering and design phase in BIM, and how it might fulfil certain types of social and operational performance assessment, and where it fits in a data maturity model of the built environment (figure 3).

Space Syntax can be seen as a universal framework for describing and representing buildings and cities. For smart cities schemes, space syntax can offer an integrative network-based model that accounts for the dependencies between urban variables. The space syntax description of urban form can entwine the layers of street networks, block density, building height, land use, street width, and can offer good proxies for estimating and forecasting human behaviour, vehicular movement, socioeconomic data, along with some aspects of the demand for energy, water and power.

[3] See Digital Built Britain plan; https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/410096/bis-15-155-digital-built-britain-level-3-strategy.pdf [accessed 23 April 2015]

For BIM models and tools, space syntax can perhaps offer the inverted void description of buildings and cities, something that needs to be incorporated in current BIM models (Jeong and Ban, 2011a, 2011b). It can be suggested that space syntax can complement the current BIM schemes by accounting for the operational, social and economic value of space.

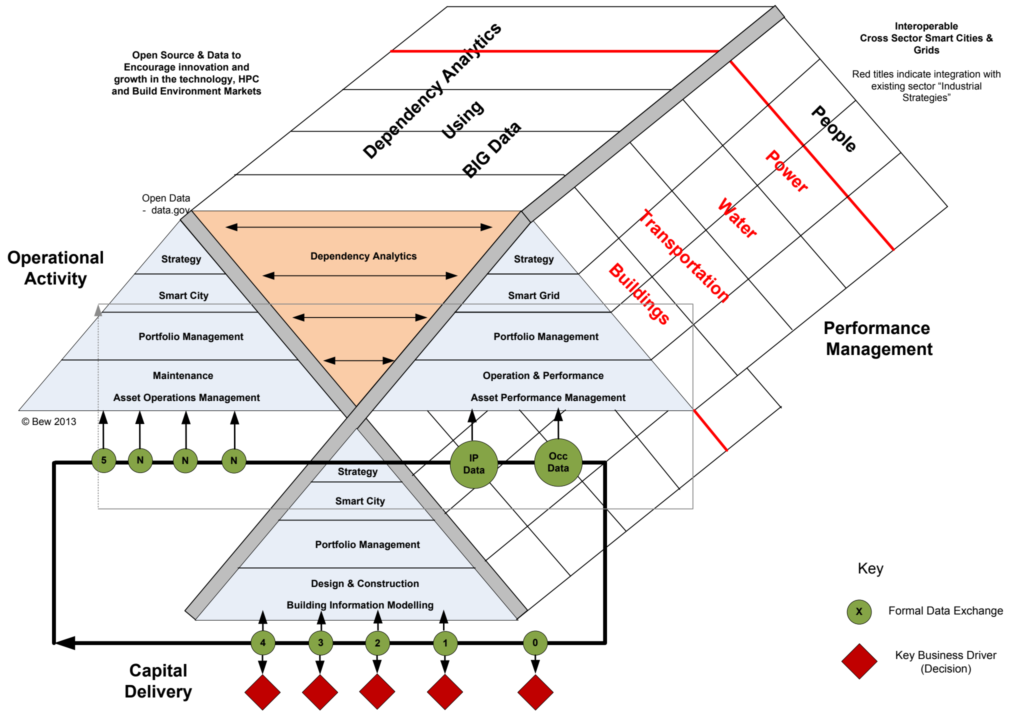


Figure 1 BIM level'3 and dependency analytics

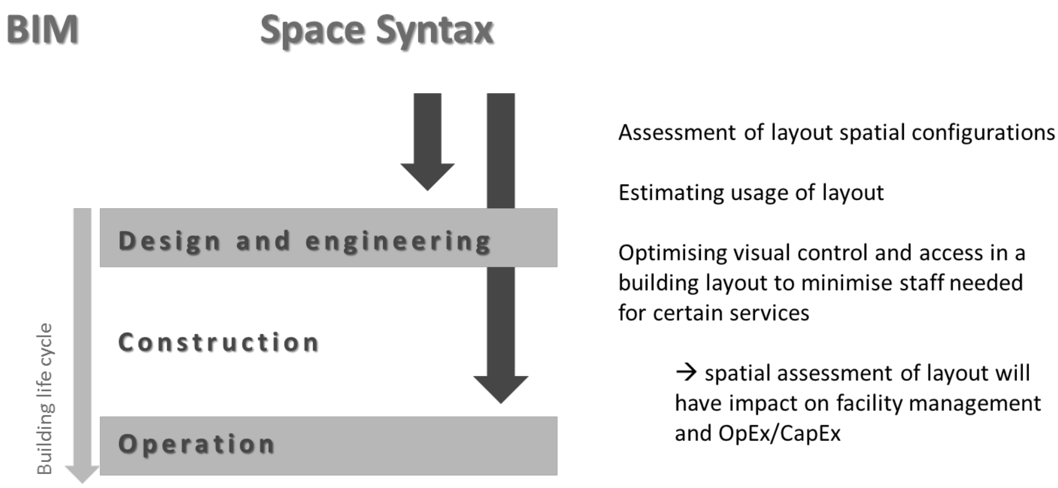


Figure 2 The relationship between BIM and Space Syntax

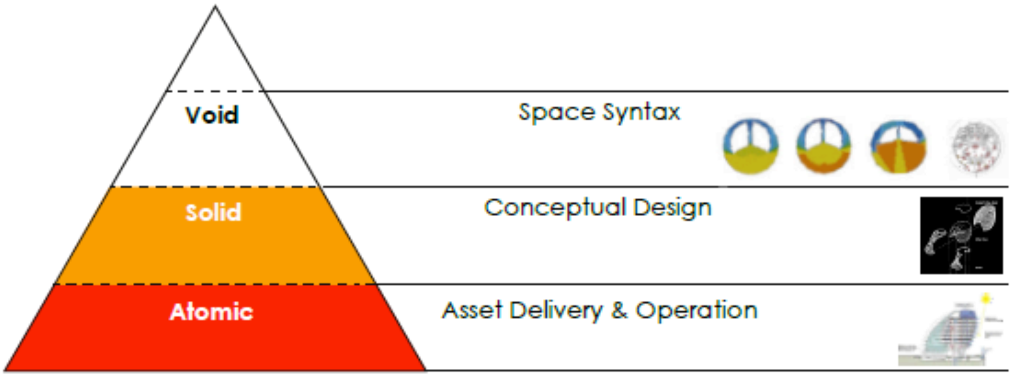


Figure 3 A data maturity model for Building Information Modelling; accounting for void and solid descriptions of the built environment (Source of GLA data: Foster and Partners, 2007).

4. Research on dependency analytics in buildings and cities

In this paper, it is suggested that standardisation models of smart cities and BIM would require a representational scheme and a modelling framework that incorporates spatial relationships and shape proportions, which are indicative of a larger set of performance criteria. An argument for that is made on both the building and urban scales.

Dependency analytics in buildings

There are multiple performance criteria in buildings; some are intended and some are a by-product of their size, shape and configurations (Al_Sayed, 2014b). It is possible perhaps to describe buildings as organised complex systems, but this description is restricted and incorporates limited dynamics; in that the dynamics are mostly affiliated with the way the smart grid and supply networks operate, and with human occupation and behaviour in facilities, and perhaps with changes on furniture and temporary structures. This is less the case with the physical structure of building; unless the building incorporates dynamic components in its structure.

It is perhaps useful to start from the implicit dependencies in the void descriptions of buildings, and move further to explain how the shape, configurations and size of spaces in buildings might have many implications on different performance criteria; such as sensed social behaviour (Sailer and Penn, 2007), social media (Conroy Dalton et al., 2013), Behavioural psychology, wayfinding and cognition (Kuliga et al., 2013; Orellana and Al_Sayed, 2013), morphological and typological parameters (Shayesteh and Steadman, 2005; Steadman, 2014), and energy performance (Steadman et al., 1991; Batty et al., 2008; Salat, 2009). It is then important to acknowledge dependencies between the atomic void and solid elements of buildings and different utility networks that supply buildings with water, gas and electricity.

Interdependencies between shapes and configurations in buildings can be described discursively, through relating the network structure of spaces in a building to the shape proportions and size (Al_Sayed, 2014b; 2014d). These basic dependencies might have many implications on the social and energy performance of buildings, hence the need to distinguish between core dependencies that characterise other more specific dependencies. An understanding of these fundamental compositions and performance criteria of void descriptions is much needed to complement the solid descriptions of buildings.

Dependency analytics in urban settings

Finding a standardised framework for smart cities would very much depend on the standard expression of the problem of cities; that is how the different components, and variables that make cities relate to each other in space and time, as well as jurisdiction issues that relate to data sharing with regards to both open and private data. The latter types of issues are beyond the scope of our paper, hence our focus on data analytics on both the building and urban scales.

Essentially, any proposition for a standardised model for smart cities needs to be supported by evidence on universality in urban systems; that is how the distribution of certain variables and their dependencies settle into universal trends (Batty et al, 2008; Bettencourt et al, 2007; Bettencourt, 2013). Aspects of universality were often traced in network-based models of cities, where cities are reduced into a set of spatial elements that are connected through network relationships between adjacent elements. Dual network representations of streets (space syntax), and primal network representation of streets (famous in transport planning), are often found to bear dependencies with other urban variables. These dependencies were traditionally attributed to the relationship between spatial accessibility of road infrastructure and pedestrian movement (Hillier et al, 1987; Hillier et al, 1993; Hillier, 1996b), as well as vehicular movement (Turner, 2012), adding also the distribution of retail land uses and street width in demand-supply models (Penn et al, 1998; Banister et al, 1998). In that, there was emphasis on the averaged trends that mark the overall performance of an urban system at its synchronic state, calculating the configurations of all origins and destinations in the street network at one frame in time. In Hillier's theory of "movement economies", the relationship between street accessibility and pedestrian and vehicular movement is fundamental, for it is this relationship that triggers retail investment (Hillier, 1996b). It is argued that where there are higher rates of accessibility, there are more people walking, hence stockholders choose these locations to invest in retail and maximise trade transactions. Furthermore, street accessibility relates to patterns of change in land use and land values (Desyllas, 1996; Hossain, 1999; Shin et al, 2007; Foltête and Piombini, 2007; Ortiz-Chao and Hillier, 2007; Enström and Netzell, 2008; Chiaradia et al, 2009; Marcus, 2007; Matthews and Turnbull, 2007; Al-Ghatam, 2009; Porta et al, 2009). Other studies have also established a relationship between street accessibility and carbon emissions (Croxford et al, 1996), as well as how carbon emissions of buildings change with their mass and configurational location within an urban setting (Park et al, 2013).

Street accessibility might also act as proxy indicators of latent variables such as poverty and social inequalities (Vaughan and Penn, 2006), and crime (Al_Sayed and Hanna, 2013). These relational models could be represented as dependencies between complex networks when looking at the totality of existing and superimposed urban layers. Street centrality -for example- might correspond to centrality measures in the social network of Twitter (Deni and Al_Sayed, 2015). Networks of visibility in streets could form the basis for considering networks of pervasive technologies (Gen. Schieck et al, 2013), and HCI (Dalton et al, 2010).

Recent work by Al_Sayed (2014a) has also revealed dependencies between the configurations of street networks and land uses, building density and height, street width and property values. These dependencies might be subject to time delay. In a dynamic loop of transactions, accessibility and land values contribute to the characterisation of block density and street width on the short term whilst triggering high-rise development and controlling the overall distribution of industrial and retail land uses on the long term. These relationships are thought to characterise the overall trends that cities converge to in the process of urbanisation, and might explain why planning restrictions hinder the development of urban regions (Albouy and Ehrlich, 2011). This might be explained in how property development in a certain region tends to begin by investing in the best sites available in terms of physical location and access to amenities, accounting also for other geotechnical characteristics. It is suggested therefore, that the costs of regulation outweigh the benefits (Cheshire and Sheppard, 2002; Glaeser et al, 2005). An account of the dependencies between different urban variables, both in static and dynamic terms, is therefore vital for appropriating an accurate information modelling description, to reduce construction costs and improve on the operational performance of infrastructure projects.

From a risk management perspective, the network structure of roads could be related to gas and electrical power networks, as well as water supply networks, by identifying critical locations in a Spatial network with graph theory (Demšar et al, 2008), and understanding how sewer networks and storm drainage operate (Rodriguez et al, 2003).

There has also been some extensive research into automating urban design systems (Gil and Duarte, 2008; Duarte et al, 2012; Beirão et al, 2012; Al_Sayed, 2014). Some of that research incorporated conceptualisations of how a City Information Modelling might look like, whilst also using space syntax modelling tools as to evaluate performance of urban systems. There is a real value in that, since research in space syntax often accounted for empirical definitions of the relationship between street configurations, formal dimensions of urban space and spatial economic (functional) variables. With such framework, there is a universal ground for modelling and forecasting visible layers of information in cities.

To follow on with this line of research, there is a need to appropriate a framework to incorporate other types of data including artificial lighting (Choi et al, 2007), and semantic information that might –for example- describe the material specifications of road infrastructure and street furniture. Interdependencies between utility networks and road networks are essential for analysing the impact of major disruptions failures and attacks on certain components in an urban infrastructure and the cascading effects that follow (Pederson et al, 2006). Network Theory might also be applied to quantify the redundancy and structural robustness of water distribution systems (Yazdani and Jeffrey, 2012). A sensible interoperability model that represents urban systems as networks is therefore essential to capture redundancies in the system and simulate its performance. However, in any interoperability model, there is a need to give priority to the road network, since it is perhaps the only type of network that operates in an off-line city. This is considering that street spaces are prospectively the kind of spaces that people escape to, in case of emergencies.

5. Using partial correlations for causal inference in spatial datasets

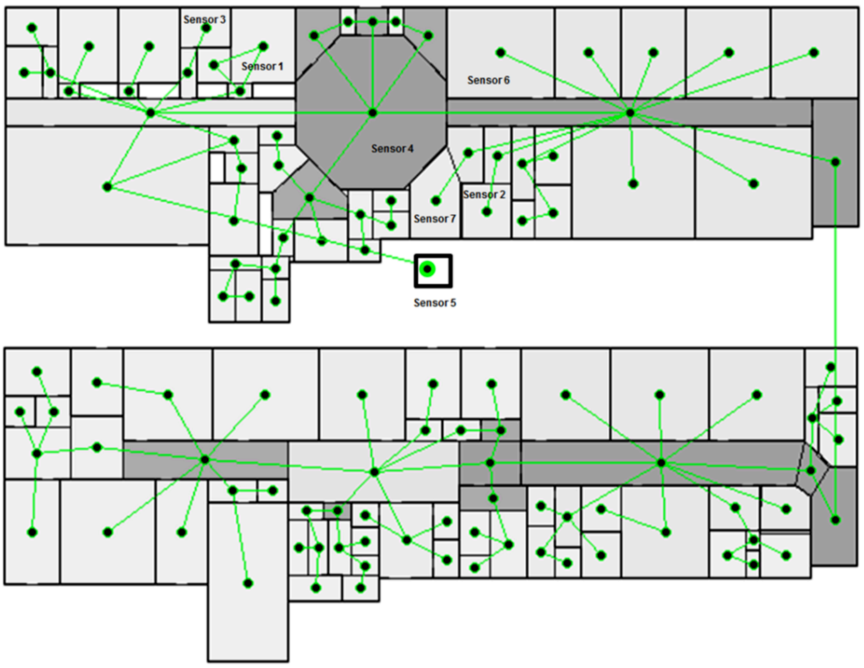
Previous sections have discussed how dependence between pairs of variables was investigated separately in the literature. For the purpose of representing relationships between larger groups of variables in the built environment, there needs to be a methodological intervention that explains the sequence and structure of interactions between performance variables and the affordances of the physical infrastructure. To reveal networks of dependencies between different data sets in buildings and cities, a methodological framework was adapted from biomedical research (De La Fuente et al., 2004)[4] to outline the relationships between different spatial components in architectural and urban layouts. The Pearson[5] product moment correlation coefficient was used in measuring associations between continuous random variables. For this purpose, a partial correlation coefficient was used to reveal dependencies and identify independence between built environment data sets. A partial correlation coefficient[6] quantifies the correlation between two variables (e.g. temperature x and humidity y) when conditioning on one z or several other variables ($z_1, z_2, z_3, \ldots z_i$)[7]. If a correlation between two variables yields a zero partial correlation (or a correlation not significantly different from zero), the algorithm removes that edge (representing a relationship between two variables) from the correlation network. The recursive application of this algorithm on all possible edges results in a network that represents putative direct interactions. In this study, we propose to use a 0 to 2nd –order correlation coefficient to interpret relationships between spatial components in buildings and cities. The application of partial correlation coefficients to represent dependencies between spatial

[4] Please refer to this paper for further details about the algorithms. The associated software were used to calculate the Pearson coefficients.
[5] As an alternative, Spearman rank correlation could be used for this analysis since it does not depend on normality and linearity of interactions, thus can be useful for a variable like Choice (Betweenness Centrality) in street networks which follows an exponential distribution.
[6] See Appendix 1
[7] A partial correlation coefficient between $rxy.z$ is the correlation between the parts of x and y that are uncorrelated with z. To obtain these parts of x and y, they are both regressed on z. The residuals of the regression are then the parts of x and y that are uncorrelated with z.

variables in the built environment can reveal some interesting patterns that might help understanding different types of social, configurational, functional and environmental performance and link it to existing and smart infrastructure.

Revealing dependency networks in buildings

This section will demonstrate the possibility of applying graph theoretic models of dependency networks to represent relationships between building data sets (configurations and room size), and environmental datasets that are collected from 7 sensors reporting a set of environmental qualities[8] of a 6th form school building[9]. In the context of buildings, social performance variables can be inferred from building configurations using convex representations of space (Hillier and Hanson, 1984). The topological connections between different convex spaces[10] might be represented by an adjacency graph (figure 4). Spaces with high connections might have more accessibility and afford higher likelihood of people moving through them to reach others. Hypothetically, the accessibility of a convex space along with its physical area might have implications on the sensed environmental and comfort qualities of the environment.



Higher values of integration

Figure 4 A topological network description[11] of a 6th Form school building (source of data: Mark Bew, EC Strategies), with the locations of sensors identified. The grayscale colouring is showing higher levels of centrality closeness (darker colours) in the network configurations of space.

Through applying the Pearson product coefficient (De La Fuente et al., 2004), it was possible to visualise an undirected dependency network that represents the relationships between lighting, area of convex spaces and spatial integration of building configurations, noise, pressure, humidity, VOC, and relative temperature of interior to exterior. The relationships visualised using the energy model of Kamada Kawai (separate components) in figure 5 show that temperature, pressure and noise are strongly related. Integration, humidity and VOC form another cluster; where humidity seems to bear

[8] Environmental performance is calculated based on average sensed values during normal workday operational hours.
[9] The data belongs to Mark Bew, EC Strategies.
[10] Fewest and fattest spaces in a layout
[11] The topological network was visualised using DepthmapX, UCL.

a strong connection to noise. The analysis yields negative correlations between the physical area of building spaces, and integration, pressure, temperature, and VOC. The analysis also yields that light bears significant positive correlations with integration and noise, and less significant with pressure and temperature. It is not clear whether these performance criteria do actually relate to each other in reality. Due to the limited number of observations and issues with accuracy of the data being generated at present, the results of this approach should be seen as an initial estimate of the real underlying network, enabling us to develop new hypotheses for interactions between configurations, physical characteristics of building components, and environmental performance.

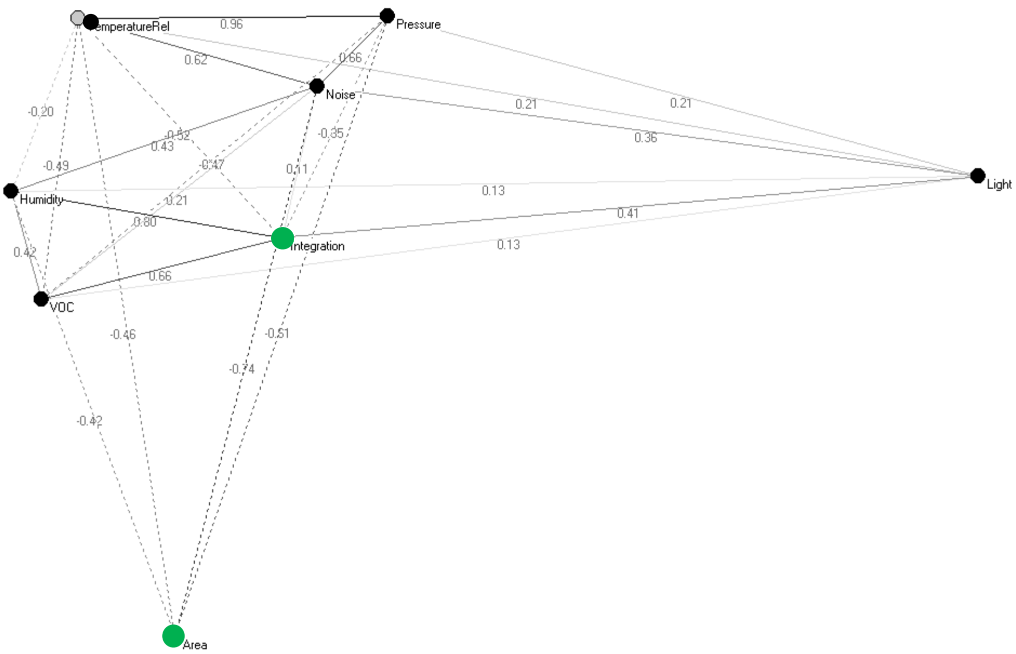


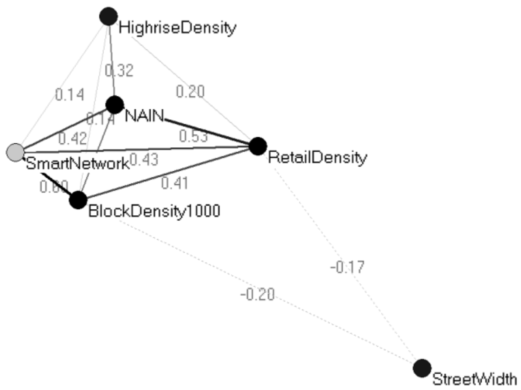
Figure 5 Dependency network[12] of the school building dataset (see figure 4), revealing relationships between eight variables. Darker edges indicate higher values for zero-order partial correlation coefficient between each two variables. The green coloured nodes represent the physical and configurational variables of rooms, the rest of the nodes represent environmental variables.

Revealing dependency networks in cities

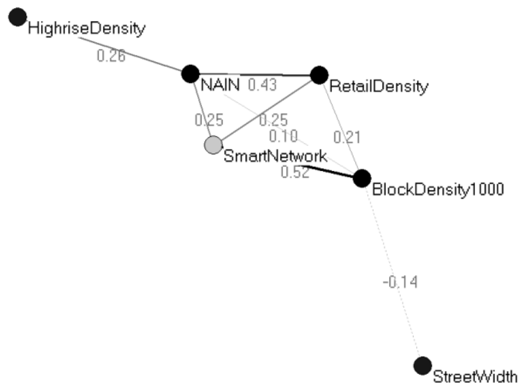
This section will demonstrate the possibility of applying graph theoretic models of dependency networks to represent relationships between urban data sets. For that we analysed street network data from Barcelona[13] and deduced measures for network accessibility NAIN[14] (highest accessibility values per 500 metre grid square), relating it to street width (ordinal number of street width correlated with highest accessibility value), and then relating accessibility and street width to; density of blocks, density of high buildings, density of retail land use, and density of sensors in the smart municipal network[15] of Barcelona all measured against a 500metre grid square area with the exception of building density which is calculated per 1000metres square).

[12] The dependency network was visualised using PAJEK software programme (De Nooy et al., 2005).
[13] Ajuntament De Barcelona [online] Available from <http://w20.bcn.cat> [Date accessed:2 September 2011]
[14] Normalised Angular Integration (NAIN); Centrality Closeness that is normalised to relate to the network size and the cost of segregation (Hillier et al, 2012), see appendix 2 for algorithm
[15] Source: Municipal wireless networks in Spain http://www.muniwireless.com/2010/04/24/municipal-wireless-networks-in-spain/ [Date accessed:1 April 2015]

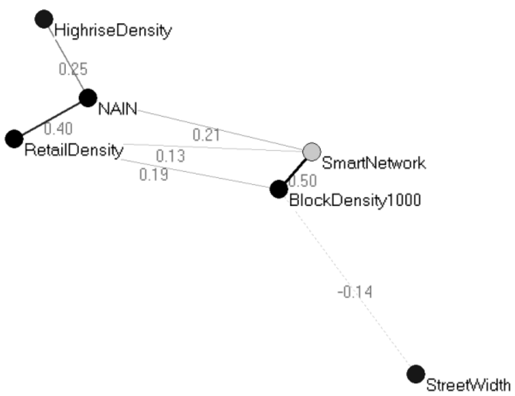
The Pearson product coefficient was again used to visualise dependency networks in urban data. To weigh the network connections against the values of coefficients, the Kamada Kawai energy model (separate components) was used for the 0rders 0-1-2 (figure 6). The different orders reveal how weak relationships disappear as to reveal the degree of interaction between the different variables. The relationship between street accessibility, building density, smart network density, and retail density forms a clique in orders 0-1. When applying the 2nd order algorithm, the relationship between street accessibility and block density disappear, whilst other relationships hold. The density of high-rise buildings conserves a relatively strong relationship with street accessibility. The remaining variable; street width, seems to bear a negative correlation with building density.



Order -0



Order-1



Order-2

Figure 6 Dependency network[16] of spatial data from Barcelona – 1438 observations (see figure 4), revealing relationships between six variables calculated using Pearson correlation coefficient of the orders 0-2. Darker edges indicate higher values for partial correlation coefficient between each two variables. The grey coloured nodes represent the density of the municipal smart network of Barcelona, the rest of the nodes represent variables that are characteristic of the existing urban infrastructure (building density (within 1000metres), and street accessibility, street width, retail density, and high-rise density (all maximum values are calculated for 500metres grid squares)).

6. Some propositions on structuring datasets for built assets

For the purpose of constructing a standardised framework to represent the built environment, it is argued that we need to account for the dependencies explained and revealed in the previous sections to arrive at a data structure that differentiates primary from secondary datasets. Primary datasets are those that are proven to be good predictors of other datasets (e.g. street networks). In making these distinctions, we build a hierarchy structure of datasets arriving perhaps at semantic

[16] The dependency network was visualised using PAJEK software programme (De Nooy et al., 2005).

descriptions of data. This hypothetical structure would help create an empirical description of different categories of data. The relational models will be needed to complement columns of existing row data with columns for estimating missing and incomplete data. Some models might also act as proxy indicators of latent variables.

A categorisation scheme must also distinguish between core data (those that exist across different cities, and building typologies) and subject data-matching smart services that are designed to respond to specific problems that cities or buildings suffer from. This can only be done by examining a wide range of case studies to generalise our observations about dependency networks and build an empirical basis for structure data information models.

7. Conclusion

This paper has introduced a theoretical framework on how to address the use of social performance analysis (using space syntax) in data maturity models. The paper has also demonstrated some propositions on how to empirically represent dependencies between different building and urban data sets by adapting a novel Pearson Correlation method -used previously in biomedical research (De La Fuente et al., 2004)– and applying it in the context of buildings and cities. Using this method, it was possible to derive dependency network representations from partial correlation coefficients.

There are nontrivial benefits for dependency network representations in the context of smart buildings and smart cities; some are to do with testing the degree of fitness between artificial smart systems and existing infrastructure, whilst others are to do with outlining redundancies, disruptions, and cascading effects in urban systems as well as in building systems. The paper concluded with some propositions on how to approach standardisation of datasets, both by attending to the dependencies between different variables and systems, and by surveying existing and sensor networks in different building typologies, and different smart city projects. There is perhaps a hierarchy of categorisation and representation that need to be addressed in any proposition for BIM and smart city standards. This hierarchy will depend on which data sets cities share in common, and the kind of datasets that are specific to certain city problems shaping the identity of smart city projects. The hierarchy will also depend on the relational models that characterise how different datasets relate to each other within buildings and on a city scale. Whilst it is difficult to establish universality across building typologies, it is perhaps more viable to identify universal models that characterise cities. Yet, it is important to admit that urban complexity is a very different matter to building complexity; the pace of change in these systems is very different and the functional descriptions are of different character. In explaining the nature of urban complexity, there is a need to emphasise that the whole is more than the parts; that a city as a whole is more than aggregations of building blocks, and that there is an emergent aspect to them. On a building scale, complexity might also have some underlying universal principles; in how spatial structures relate to shape and size of spaces, and in how a combined description of shapes and configurations bears a relationship to energy consumption, carbon emissions, lighting, and noise.

At this stage, it is important to raise some caveats with regards to the interpretation of our findings, considering the small data set we had for buildings and the variance in environmental performance measures that have much to do with the operation of buildings alongside many other factors. It is also important to recognise that, whilst partial correlation coefficients do not necessarily indicate causality, their ability to exclude weak correlations legitimises their use as indicators for causal inference, hence their use makes it possible to rule out primary from secondary datasets. There is a need to emphasise here that “spurious” correlation models of building and urban relationships must not be explained as matters of causality (Simon, 1957), since many different causal relationships can be mapped onto a correlation. Therefore, the application of zero-order[17] to 2nd order correlation networks in the context of buildings and cities need to be cautiously interpreted. Pearson correlations might fail in some occasions to correctly identify a system of significant relationships

[17] The order of the partial correlation coefficient is determined by the number of variables it is conditioned on. The zero-order For example, *rxy.z* is a first-order partial correlation coefficient, because it is conditioned solely on one variable (*z*).

between different variables, and might on other occasions coincidentally show unrealistic correlations between variables that don't have any shared performance requirements. Despite these deficiencies, the method can be used to develop, with reasonable degree of confidence, plausible hypotheses of interactions between physical, configurational and performance variables, whilst also revealing correspondence between existing and smart infrastructure. The use of dependency networks will therefore be very helpful in building an empirical basis for building and urban information models, and in structuring performance data to enable better predictions about design and operation throughout the lifecycle of infrastructure projects.

References

Allen, P. M. (1997). *Cities and Regions as Self-Organizing Systems: Models of Complexity*, Taylor and Francis.

Al_Sayed, K. and Hanna, S. (2013). 'How city spaces afford opportunities for riots', In *Proceedings of the 9th International Space Syntax Symposium*, Edited by Y O Kim, H T Park, K W Seo, Seoul, Korea.

Al-Sayed, K. (2014a) 'Urban Morphogenesis: How form-function complexity coupled temporal changes on street configurations in Manhattan and Barcelona over the past centuries'. PhD thesis, UCL (University College London).

Al_Sayed, K. (2014b) 'On the evolution of thoughts, shapes and spatial structure in architectural design'. In *Design Computing and Cognition DCC'12*. J.S. Gero (ed), pp. 393-411. © Springer 2014.

Al_Sayed, K. (2014c) 'Thinking systems in urban design: A prioritised structure model'. In Explorations in Urban Design. M. Carmona (ed), (Farnham: Ashgate, 2014), pp. 169–181.

Al_Sayed K. (2014d). 'How designs evolve', The Journal of Space Syntax. Special Issue on Models and Diagrams in Architectural Design, In D. Koch and P. Miranda Carranza (ed), 5(1): 68-90.

Batty, M. (2013). *The New Science of Cities*. MIT Press.

Batty, M., Axhausen, K., Fosca, G., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., and Portugali, Y. (2012).'Smart cities of the future'. *European Physical Journal-Special Topics*, *214*(1), 481.

Batty, M., Carvalho, R., Hudson-Smith, A., Milton, R., Smith, D., and Steadman, P. (2008). 'Scaling and allometry in the building geometries of Greater London'. *The European Physical Journal B-Condensed Matter and Complex Systems*,*63*(3), 303-314.

Beirão, J. N., Montenegro, N., and Arrobas, P. (2012). 'City Information Modelling: parametric urban models including design support data'. Available at: http://utl. academia.edu/JoseBeir%C3%A3o/Papers/1829358/City_Information_Modelling_parametric_urban_models_including_design_support_data [Accessed August 3, 2012].

Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007). 'Growth, innovation, scaling, and the pace of life in cities'. *Proceedings of the national academy of sciences*, *104*(17), 7301-7306.

Bettencourt, L. M. (2013). 'The origins of scaling in cities'. *science*, *340*(6139), 1438-1441.

Bew, M. and Underwood, J. (2009). 'Delivering BIM to the UK Market'. *Handbook of Research on Building Information Modeling and Construction Informatics: Concepts and Technologies, Hershey, PA: IGI Global*, 30-64.

Choi, A. S., Jang, S. J., Park, B. C., Kim, Y. O., and Kim, Y. S. (2007).'Rational-design process and evaluation of street-lighting design for apartment complexes. *Building and environment'*, 42(8), 3001-3013.

Conroy Dalton, R., Kuliga, S. F. and Hoelscher, C. (2013) 'POE 2.0: POE 2.0: exploring the potential of social media for capturing unsolicited post occupancy evaluations'. *Intelligent Buildings International*. 5(3), 162-180.

Croxford, B., Penn, A., and Hillier, B. (1996). 'Spatial distribution of urban pollution: civilizing urban traffic'. *Science of the total environment*, *189*, 3-9.

Dalton, S. N., Marshall, P., and Dalton, R. C. (2010). 'Measuring environments for public displays: A space syntax approach'. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3841-3846). ACM.

De La Fuente, A., Bing, N., Hoeschele, I., and Mendes, P. (2004). 'Discovery of meaningful associations in genomic data using partial correlation coefficients'. *Bioinformatics*, 20(18), 3565-3574.

Demšar, U., Špatenková, O., and Virrantaus, K. (2008). 'Identifying critical locations in a spatial network with graph theory'. *Transactions in GIS*, *12*(1), 61-82.

Deni, S. and Al_Sayed K. (2015). 'Virtual communities and patterns of social interactions in 'Tech City''. AESOP Planning and Complexity. 13th meeting: Complexity and digitalization of cities – Challenges for urban planning and design, Tampere, Finland

De Nooy, W., Mrvar, A., and Batagelj, V. (2005) 'Exploratory Social Network Analysis with Pajek', Structural Analysis in the Social Sciences 27, Cambridge University Press.

Department for Business, Innovation & Skills (2015) Digital Built Britain, Level 3 Building Information Modelling - Strategic Plan. [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/410096/bis-15-155-digital-built-britain-level-3-strategy.pdf, [Accessed: 23rd April 2015]

Duarte, J.P. et al. (2012). 'City Induction: formulating, generating, and evaluating urban plans'. In *Digital Urban Modelling and Simulation. CCIS Series Communications in Computer and Information Science Series*. Springer-Verlag.

Gil, J. and Duarte, J.P., (2008). 'Towards an Urban Design Evaluation Framework'. In *Architecture in Computro-26th eCAADe Conference Proceedings*, Antwerpen. pp. 257–264.

Gen Schieck, A. F., Al_Sayed K., Kostopoulou, E., Behrens, M., and Motta, W. (2013). 'Networked architectural interfaces: Exploring the effect of spatial configuration on urban screen placement', In *Proceedings of the 9th International Space Syntax Symposium*, Edited by Y O Kim, H T Park, K W Seo, Seoul, Korea.

Hillier, B. and Hanson, J. (1984). *The social logic of space*. Cambridge: Cambridge university press.

Jeong, S. K. and Ban, Y. U. (2011a). 'Computational algorithms to evaluate design solutions using Space Syntax'. *Computer-Aided Design*, 43(6), 664-676.

Jeong, S. K. and Ban, Y. U. (2011b). 'Developing a topological information extraction model for space syntax analysis'. *Building and Environment*, 46(12), 2442-2453.

Kuliga, S., Conroy Dalton, R. and Hölscher, C. (2013) 'Aesthetic and Emotional Appraisal of the Seattle Public Library and its relation to spatial configuration'. In: 9th International Symposium, Seoul, South Korea

Orellana, N. and Al_Sayed, K. (2013). 'On spatial wayfinding: Agent and human navigation in virtual and real worlds', In *Proceedings of the 9th International Space Syntax Symposium*, Edited by Y O Kim, H T Park, K W Seo, Seoul, Korea.

Park, H. T., Kim, M. H., and Kim, Y. O. (2013) 'Beyond compact city: Spatial configuration model for carbon emission reduction'. , In *Proceedings of the 9th International Space Syntax Symposium*, Edited by Y O Kim, H T Park, K W Seo, Seoul, Korea.

Pederson, P., Dudenhoeffer, D., Hartley, S., and Permann, M. (2006). 'Critical infrastructure interdependency modelling: a survey of US and international research'. *Idaho National Laboratory*, 1-20.

Penn, A., and Croxford, B. (1998). 'Fingerprinting urban kerbside carbon monoxide concentrations: interaction between street grid configuration, vehicle flows and local wind effects'. *International journal of vehicle design*, *20*(1), 60-70.

Rodriguez, F., Andrieu, H., and Creutin, J. D. (2003). 'Surface runoff in urban catchments: morphological identification of unit hydrographs from urban databanks'. *Journal of Hydrology*, *283*(1), 146-168.

Sailer, K., and Penn, A. (2007). 'The performance of space–exploring social and spatial phenomena of interaction patterns in an organisation'. In: *Proceedings of the Architecture and Phenomenology Conference*. Faculty of Architecture and Town Planning, The Technion, Israel Institute of Technology: Haifa, Israel.

Salat, S. (2009). 'Energy loads, CO2 emissions and building stocks: morphologies, typologies, energy systems and behaviour'. *Building Research & Information*, *37*(5-6), 598-609.

Schultz, C., and Bhatt, M. (2011). 'Toward accessing spatial structure from building information models'. In *UDMS 2011: 28th Urban Data Management Symposium, Delft, The Netherlands, September 28-30, 2011*. Urban Data Management Society; OTB Research Institute for the Built Environment; Delft University of Technology.

Schultz, C., Bhatt, M. and Mora, R. (2013) 'Mindyourspace - a tool for evidence-based qualitative analyses of user experience and navigation behavior in the built environment'. In *edra44providence - 44th Environmental Design Research Association Conference*.

Shayesteh, H. and Steadman, P. (2005). 'Typo-morphological approach to housing transformation in Tehran'. In *Proceedings of the 5th Space Syntax Symposium*, TU Delft.

Simon, H. A., (1957). *Models of man; social and rational*. Wiley. New York.

Steadman P., Brown F. and Rickaby P. (1991). 'Studies in the morphology of the English building stock'. *Environment and Planning B: Planning and Design*. 18(1) 85 – 98

Steadman, P. (2014). *Building Types and Built Forms*. Troubador Publishing Ltd.

Turner, A. and Penn, A. (2002). 'Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment'. *Environment and Planning B: Planning and Design*, 29(4), 473-490.

Varoudis T. (2012) "DepthmapX - Multi-Platform Spatial Network Analysis Software", The Bartlett School of Architecture, UCL, http://varoudis.github.io/depthmapX/

Appendix 1

A partial correlation can be calculated to any pre-defined order. Partial correlation coefficients can be used to distinguish between causal type of correlations and correlations between two variables that originate *via* intermediate variables (sequential pathways) or those that embed direct relationship to other variables (common causes). The following three Equations (1,2, and 3) can be used to calculate partial correlation coefficients of orders 0–2. Similar type of equations can also be used to calculate higher order partial correlation coefficients.

zeroth-order correlation $$r_{xy} = \frac{cov(xy)}{\sqrt{var(x)var(y)}} \quad (1)$$

first-order correlation $$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (2)$$

second-order correlation $$r_{xy.zq} = \frac{r_{xy.z} - r_{xq.z}r_{yq.z}}{\sqrt{(1 - r_{xq.z}^2)(1 - r_{yq.z}^2)}} \quad (3)$$

Appendix 2

In order to enable cross-scale comparisons between different parts of a city or between different cities, Hillier *et. al.* (2012) suggested a normalisation procedure for angular weighted graph distance. The normalisation was based on a hypothetical relationship between the tendency of an urban system to optimise travel distance from all origins to all destinations and the cost of segregation that is an effect of the system size. Normalised angular Integration $NAIN_\theta$ for a graph *G'* with the size *n* is defined in Hillier *et. al.* as follows;

$$NAIN_\theta(x) = \frac{(n+2)^{1.2}}{\left(\sum_{i=1}^{n} d_\theta(x,i)\right)}$$

where d_θ is the length of a geodesic (shortest path) between vertix x and i.