Smart Infrastructure and Construction
ISBN 978-0-7277-6127-9

doi:10.1680/tfitsi.61279.729

A dependency network description of building information models

K. Al Sayed[*1], M. Bew[2] , and A. Penn[1]

[1] University College London, London, UK
[2] The Department for Business, Innovation & Skills (BIS), London, UK
[*] Corresponding Author

ABSTRACT. The pervasive deployment of "smart building" projects 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, by extending building 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 building 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 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 buildings and inform architectural design and public policy practices, in the design and operation phases of infrastructure projects..

1 INTRODUCTION

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 & Ban, 2011), they need to be organised and systemised to relate to social, cognitive and behavioural performance criteria (Schultz & 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 paper aims to outline key aspects of the nature of building dependencies, in an effort to build a dependency network description of the variables that make up their complexity. At essence, we plan to model 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.

With this objective, this paper addresses how network-based descriptions of built environment might be incorporated in BIM frameworks, through establishing a relationship between the configurations of built form and social structures, as well as environmental performance. For that, a methodology for visualising dependency networks is introduced, along with some propositions on how to integrate frameworks and incorporate empirical models of dependency networks as to inform data information models and public policies.

2 A DATA MATURITY MODEL FOR BIM

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. 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 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 the social performance embedded in the description of building layouts, e.g. how room spaces connect through adjacency relationships, and how these adjacencies influence social interactions on the long term. It might be argued that any improvements made on the performance of building layouts, would have positive impact on the social and economic performance of buildings during the operational phase.

In the model proposed by Mark Bew[1] for BIM level 3 strategy, the design and operation of buildings needed to account for dependency analytics, in order to better outline the performance requirements of infrastructure. From these requirements stems the relationship between existing building infrastructure and the smart systems that are designed to improve its function. It is usually argued that performance requirements might be mined from smart building 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 infrastructure to improve its overall performance. In this context, a network description of building layouts might be used, but need to also have a complementary 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 the spatial attributes of each room space (e.g. network configurations (space syntax), shape proportions), the values for each component will fall into the same attribute and entity positions with different provenance, to make a data maturity model of the built environment (figure 1). 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 social performance indicators, 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. For BIM models and tools, a network-based description of building space that accounts for the shape parameters of each room might perhaps offer the inverted void description of buildings.

[1] 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]

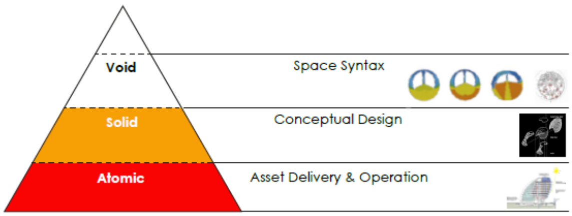

Figure 1 A data maturity model for Building Information Modelling; accounting for void and solid descriptions of the built environment (Source of GLA data: Foster and Partners, 2007).

3 RESEARCH ON 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, 2014a). 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 & Penn, 2007), social media (Conroy Dalton et al., 2013), Behavioural psychology, wayfinding and cognition (Kuliga et al., 2013; Orellana & Al_Sayed, 2013), morphological and typological parameters (Shayesteh & 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, 2014a; 2014b). 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.

4 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, a methodological framework was adapted from biomedical research (De La Fuente et al., 2004)[2] to outline the relationships between different spatial components in architectural layouts. The Pearson product[3] 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[4] quantifies the correlation between two variables (e.g. temperature x and humidity y) when conditioning on one z or several other varia-

[2] Please refer to this paper for further details about the algorithms. The associated software was used to calculate the Pearson coefficients.

[3] 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.

[4] See Appendix

bles (z1, z2, z3, … zi)[5]. 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 (a second-order UDG approximation graph). In this study, we propose to use a 0 to 2nd –order correlation coefficient to interpret relationships between spatial components in buildings. The application of partial correlation coefficients to represent dependencies between spatial 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.

4.1 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[6] of a 6th form school building[7]. In the context of buildings, social performance variables can be inferred from building configurations using convex representations[8] of space (Hillier & Hanson, 1984). The topological connections between different convex spaces might be represented by an adjacency graph (figure 2). 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.

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 (table 1). The relationships were visualised using the energy model of Kamada Kawai (separate components) in figure 3, revealing that temperature, pressure and noise are strongly related. Integration, humidity and VOC form another cluster; where humidity seems to bear 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.

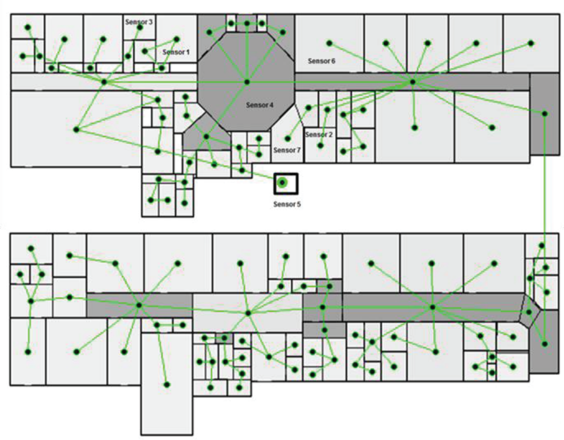

Figure 2 A topological network description[9] of a 6th Form school building, with the locations of sensors identified. Darker colours indicate higher levels of centrality closeness.

[5] 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*.

[6] Environmental performance is calculated based on average sensed values during normal workday operational hours.

[7] The data belongs to Mark Bew, EC Strategies.

[8] Fewest and fattest spaces in a layout

[9] The topological network was visualized using DepthmapX, UCL.

Table 1 zeroth[10] order Pearson correlation matrix for school data

	R_Temp	VOC	Light	Noise	Humidity	Pressure	Integ	Area
R_Temp	1.00	-0.49	0.21	0.62	-0.20	0.96	-0.52	-0.46
VOC	-0.49	1.00	0.13	0.21	0.42	-0.47	0.66	0.07
Light	0.21	0.13	1.00	0.36	0.13	0.21	0.41	0.00
Noise	0.62	0.21	0.36	1.00	0.43	0.66	0.11	-0.74
Humidity	-0.20	0.42	0.13	0.43	1.00	0.00	0.80	-0.42
Pressure	0.96	-0.47	0.21	0.66	0.00	1.00	-0.35	-0.61
Integ	-0.52	0.66	0.41	0.11	0.80	-0.35	1.00	-0.03
Area	-0.46	0.07	0.00	-0.74	-0.42	-0.61	-0.03	1.00

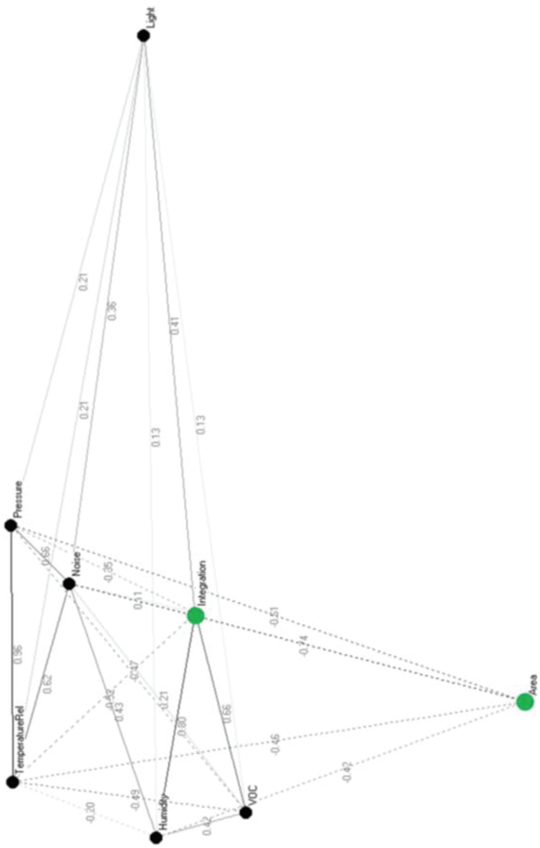

Figure 3 Dependency network[11] of the school building dataset (see figure 2), revealing relationships between eight variables. Darker edges indicate higher values for zero-order partial correlation coefficient between each two variables. The green coloured nodes represent the physical and configurational variables of rooms, the rest of the nodes represent environmental variables.

[10] 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*).

[11] The dependency network was visualized using PAJEK software (De Nooy et al., 2005).

5 CONCLUSION

This paper 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 a method on how to empirically represent dependencies between different building data sets by adapting a novel Pearson Correlation technique -used previously in biomedical research (De La Fuente et al., 2004)– and exploring its application in the context of buildings. 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; 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 building systems. 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 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 to 2nd order correlation networks in the context of buildings need to be cautiously interpreted. Pearson correlations might fail in some occasions to correctly identify a system of significant relationships 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 information models, and in structuring performance data to enable better predictions about design and operation of buildings.

REFERENCES

Al_Sayed, K. 2014a. 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. 2014b. 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.

Al_Sayed, K., Bew, M., Penn, A., et al. 2015. Modelling dependency networks to inform data structures in BIM and smart cities. In *Proceedings of the 10th International Space Syntax Symposium*. Edited by K. Karimi et al, London, UK.

Bew, M. & 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.

Conroy Dalton, R., Kuliga, S. F. & 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.

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

De Nooy, W., Mrvar, A., & 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]

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

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

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

Orellana, N. & 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.

Sailer, K., & 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., & 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. & 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. & 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. & 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.

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

APPENDIX

A partial correlation can be calculated to any pre-defined order (table 2). 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.

Table 2. Different orders for the partial correlations.

0th order correlation	$r_{xy} = \frac{cov(xy)}{\sqrt{var(x)var(y)}}$	(1)
1st order correlation	$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$	(2)
2nd order correlation	$r_{xy.zq} = \frac{r_{xy.z} - r_{xq.z}r_{yq.z}}{\sqrt{(1-r_{xq.z}^2)(1-r_{yq.z}^2)}}$	(3)