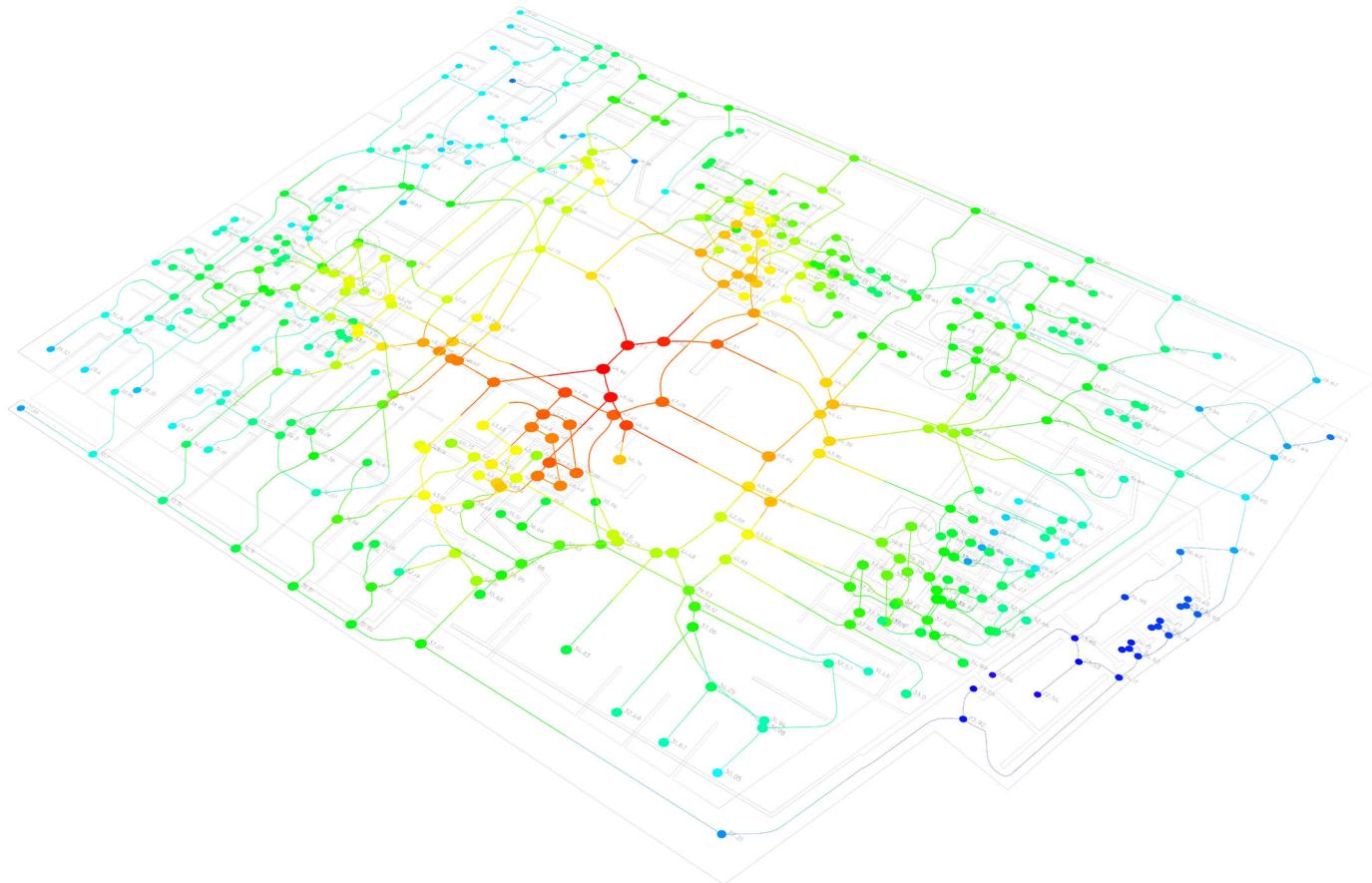


Digital intuition – Autonomous classifiers for spatial analysis and empirical design

Christian Derix, Prarthana Jagannath
Woods Bagot | Global Studio

Pages: 190-215



The Journal of Space Syntax

ISSN: 2044-7507 | Year: 2014 | volume: 5 | issue: 2 | Online Publication Date: 29 December 2014

Digital intuition – Autonomous classifiers for spatial analysis and empirical design

Christian Derix, Prarthana Jagannath

Woods Bagot | Global Studio

Unsupervised network models have found application as data mining tools in spatial analysis and as search tools in spatial design. Building on previous work at the urban scale, ‘self-organising feature maps’ (SOMs) are tested as mechanisms to classify heterogeneous spatial attributes of layouts and extract representative data profiles. These profiles have been presented in the paper as potentially new spatial typologies, which are implicit within the morphology of a layout and are representative of the occupational affordances of that layout. Contrary to traditional typologies, these typologies are derived rather than assigned. The methods explored in this paper establish a way in which spatial typologies can be autonomously recognised through associations between spatial attributes of a layout. Spatial typologies which are representative of occupant perception may be useful in incorporating user-centric approaches more systematically into the design process and could give rise to new key-performance indicators.

Keywords:
Spatial typologies, artificial neural networks, self-organising maps, associative models, spatial characterisation.

Although the models were developed and are described using test cases established in the context of spatial resilience and infrastructure security, they are presented with the intention of generalisation and with an eye on their application to spatial analysis and design outside the remit of security. The paper aims to focus on the development and structure of these specific models which are later generalised beyond the security perspective for application in other design sectors. The models are intended to act as heuristic underlays for the design process, maintaining light-weight interactivity and real-time feedback. This is largely to introduce analysis earlier in the design process and enable an iterative design-analysis workflow. The paper also speculates on the future of design methodologies beyond parametric procedures.

1. Introduction

Increasingly, user-centric objectives based on cognitive descriptions are suggested in building and urban design, for instance in best practice guidelines published by the Commission for Architecture and the Built Environment (CABE). These guides dealt with ‘best practice’ design objectives and mechanisms, which were based on qualitative and perceptive properties of space. In ByDesign, four of the seven ‘Objectives for Urban Design’ regard spatial properties as support for user perception: continuity and enclosure, ease of movement, legibility and quality of the public realm (CABE, 2000). ByDesign also details different ‘aspects of development form’, which underpins the cognitive nature of urban patterns through spatial

properties. Similarly, professional building sectors, such as workplace design, are discovering users and their behaviours as a potential source for ‘best practice’ (US GSA, 2006). Hardly any of the user-centric development criteria are backed up by measures or quantitative analysis methods.

Through a series of academic studies and professional projects, the authors have been principally testing feature classification models so as to explore the associations between the configurational properties and occupational patterns in buildings and urban environments. The intrinsic agenda behind these studies questions the correlation between the use of space and measures of spatial cognition. In a pertinent

enumeration of this agenda, the paper explores the potential of using unsupervised network models to extract spatial typologies from building configurations and their occupational patterns. The resulting typologies are generated without necessarily referring to the configuration's pre-assigned functional type. Such an emergent typology is defined by a weighted combination of the processed configurational and occupational properties of the space.

The most obvious application of these typologies is to use them to analyse the performance of a layout against relevant criteria. The mapping mechanism would discretise layouts by classifying them into profiles which would make the configurations more amenable to analysis and manipulation in the design process. An enhanced application of the same would be to reveal potential affordances of use, based on which spatial configurations can be classified as belonging to a particular 'use affordance type'.

In general, the paper aims to test the efficacy of unsupervised network models, specifically self-organised maps, as models to cluster spatial data in order to present an alternative discretised understanding of layout configurations. In the following, section 2 discusses the definition of spatial typologies in the context of unsupervised network models, and section 3 establishes some precedents for the use of unsupervised network models in spatial design. A specific model is presented as a part of the developments made for an EU-funded FP7 research project, Resilient Infrastructure and Building Security (RIBS). Section 4 outlines the data collected for this model, its structure and outputs in terms of spatial risk definitions. A brief description in section 5 outlines the generalisation of the model beyond a spatial security perspective in order to establish applicability to other sectors of architectural and urban design. The precedents presented discuss

architectural and urban design projects that extend and utilise emergent typology definitions to associate design interventions with existing context. They also provide some theoretical context to the concept of emergent spatial typologies and outline some precedence for the use of network models in architectural design.

As a tool for analysis, these methods can be used to query the performance of a layout while addressing issues relevant to the data combinations being used to generate spatial profiles. In design, real-time feedback can be obtained from the models in order to alter spatial configurations and close gaps in their performance early in the design process.

2. Clustering and spatial typologies

Comparable sets of data samples can be reduced to clusters or categories based on their similarities. These categories can be used as handles through which the data can be manipulated. During this reduction of dimensionality, some information is lost in favour of simplification. Patterns inherent in the data are extracted and used as typical representations of the categories or clusters. Through the clustering process, the data samples can then be placed in a comparative map by virtue of their contrasts and convergences. Data samples that are similar to each other will be placed close to each other and dissimilar objects will be placed farther apart. A user is then able to navigate this map using the associations between data samples.

Although other clustering models were tested for the following work, the main models developed and described are variations of a self-organising map (SOM) (Kohonen, 1995). In general terms, a SOM is a network model that consists of layers of components called nodes or neurons. A SOM basically aims to associate particular regions of its map with similar input data, thereby being

able to reduce high-dimensional data to lower dimensions for the purpose of visualisation and operation. Similar to other network models a SOM has to ‘learn’ in order to successfully map an input dataset. Where supervised network models need to be taught how to recognise a good solution, SOMs being unsupervised learning models can learn unlabelled input data. If the input data is not initially labelled, the resultant mapping would need to be processed semantically for visualisation.

Each node is assigned a weight vector of the same length as an input data sample. For the purpose of this paper, data samples consist of an array of disparate spatial attributes associated with a location in a layout. Weights are initially set at random or sampled from the data set in order to minimise the learning time. The input data samples are fed to the map over numerous iterations to enable learning. The output nodes of a SOM are linked to each other based on the topological structure of the map. The model utilises a neighbourhood function to distribute information across its regions. This aspect of a SOM allows it to preserve topological structures within an input data set despite the reduction of dimensionality.

The competitive learning mechanism of a SOM ensures that the output node that matches the input data sample best is a winner. The weights of the winning node and its neighbours are adjusted every iteration to match the input data sample, creating strong associations between that region of nodes and a particular type of data sample. With each iteration, the learning rate of the map is reduced, such that in the beginning large adjustments can be made, whereas towards the end adjustments are slight and the map is more rigid.

SOMs with n output nodes typically cluster data samples into n clusters. The structure of a network model can be referred to by its dimensionality (for example, a one-dimensional or two-dimensional

SOM). This reference to dimensionality describes the topology of connections between output nodes, with one-dimensional SOMs having a single line of output nodes and two-dimensional SOMs having a grid of output-nodes. The terms, however, do not relate to the dimensionality data being processed by the network model.

A one-dimensional SOM is usually built such that input samples mapped to a particular node are more similar to input samples mapped to nodes on either side of the winning node, as compared to other nodes in the map. Topological relationships between the output nodes of a SOM preserve equally complex topological structures present within the data. As a result, two-dimensional SOMs, being more topologically complex than a one-dimensional SOM, would map more complex data patterns. Although this forms the structure of a basic SOM, further complexity can be added to the model by, varying the dimensionality of the output nodes, by varying their structure, or by making the model adaptable over time. For the specialised integrated clustering model described in section 4.3, a basic one-dimensional SOM (4.3.1) and a clipped two-dimensional SOM (4.3.2) are used. The models highlight different structural patterns within the same dataset as a consequence of their distinctive structures.

In the following sections, the data samples being used are spatial in nature. Spatial data samples, for the purpose of this paper, can be points in a layout or convex spaces in a layout. In a more general sense, they can be any object with a set of attributes defining a location (points, rooms, floors, building layouts or urban layouts) and a set of attributes defining some configurational or occupational properties. When such data is clustered using a SOM, each cluster consisting of one or more data samples is represented by an output node. The cluster itself is defined by the associated weight vector associated with the

output node. The weight vector is considered as a typical set of values, representative of all the data samples that are members of that cluster. The paper suggests the use of these typical samples as new emergent spatial typologies, which have been extracted from a set of spatial data samples through a clustering process. Data samples can be reduced to their representative spatial typologies for simplification and further manipulation.

In a broader sense, however, all data samples can be considered as members of every category but to varying degrees. While only the strongest memberships are visualised, understanding the outputs as a gradation of memberships could allow for the definition of ‘fuzzy’ rather than fixed archetypes.

3. Precedents / Experiential typologies

In this section, precedent implementations of neural networks in architectural design are examined in order to establish an understanding of experiential typologies. Richard Coyne introduced the concept of connectionism as an associative differential engine to architectural research (Coyne, 1990) using David Rumelhart's Parallel Distributed Processing model (PDP) as a tool for demonstration of abstract residential plans. The PDP is a type of information processing model

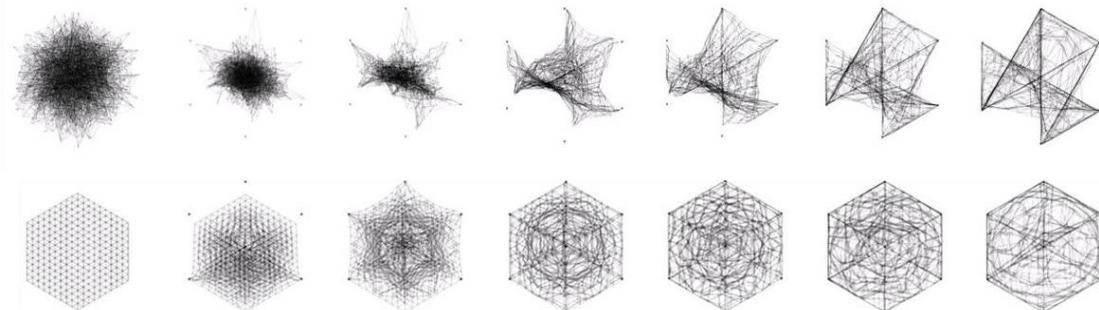
inspired by the manner in which information is processed in the brain. Representations of the input data are stored implicitly by being distributed in the connections between the units of the model. Learning occurs as the matrix of connections is adjusted to match pre-defined target states from the input data. This type of learning is known as supervised learning wherein new classes of associations do not emerge. Subsequently, Ivan Petrovic and Igor Svetel used the PDP as a template to generate three-dimensional residential layouts based on semantic associations (Petrovic and Svetel, 1993).

At first, studies of human movement behaviour triggered mapping exercises into non-standard cognitive maps of spatial dimensions. These deviated from standard metric plans, and were comparable to Kevin Lynch's 'mental maps' (Lynch, 1960). It became clear that people navigated space by associating spatial properties to events in time and adjusted their behaviour to account for what they had learned in a space previously, resulting in seemingly subjective associative maps. The neuroscientist Teuvo Kohonen developed a mechanism that would be able to map disparate features via associations into perceptive maps (Kohonen, 1995). His self-organising feature map (SOM) was adopted in 2000 to explore a

Figure 1:

Self-Organising Space (SOS) model.

Early 1999 tests where the same cubic input vectors are classified by the SOM with diverging final clusters, depending on the initial distribution of neurons.



spatial machine that could build autonomous representations of space, and was called 'self-organising space' or SOS (Figure 1) (Derix and Thum, 2000). The SOM was distinguished from contemporary unsupervised network models by its ability to represent features with continuous real numbers as opposed to binary values. This is particularly useful to represent gradients of intensities in architectural design. Additionally, the feedback process of the map can be easily visualised, making the logic of the model more transparent to the designer.

In the SOS, the SOM was used to reduce high-dimensional data to three-dimensional spatial outputs. The SOS mapped associations of spatial data such as vertices and edges of buildings and proximities to edges over time, into volumetric network clusters. They were visualised as spatial boundaries by generating an implicit surface using the marching cubes algorithm. The network was 'released' within a three-dimensional urban site model and the nodes had to select their input samples autonomously within a perceptive reach. Like the human movement behavioural maps, the model would experience the site and build associations between locations and spatial features over time, visualising the results through spatial expressions (Figure 2) (Coates, Derix et al., 2001).

The SOS model represented a prototype at the Centre for Evolutionary Computing in Architecture (CECA, University of East London), that served as a research foundation for many following student projects (and eventually the Aedas R&D spatial classification model). A multiple network approach was tested with Tim Ireland in 2002, where each room in an accommodation schedule would be dissolved into a network rather than a geometric boundary or a simple graph representational node. This allowed each room to build differentiated associations with other rooms simultaneously via their many neuronal nodes, resulting in organic building layout studies (Ireland and Derix, 2003). In 2004 Derix proposed to use the SOS as an input mechanism to a more standard SOM in order to analyse the perceived spatial structures of the SOS into classes of empirical spatial typologies (Figure 3) (Derix, 2004; Benoudjite, Derix and Coates, 2004).

Sean Hanna at UCL developed a method for the classification of architectural plans from large data sets, using 'spectral graphs' to reduce complex matrices to manipulable minimal graphs (of eigenvectors) (Hanna, 2006). This method was incorporated by John Harding at UEL in 2008 to build an 'artificial curator' model for exhibition planning (Harding and Derix, 2010) (Figure 4).

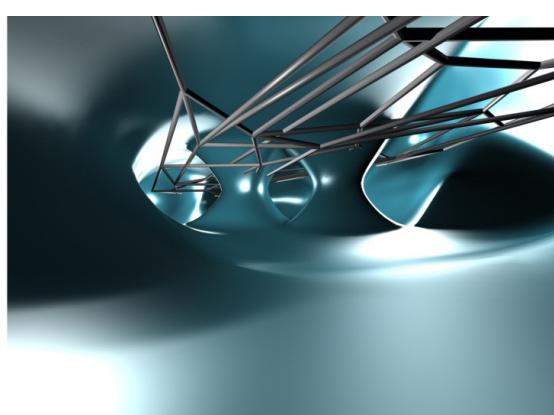
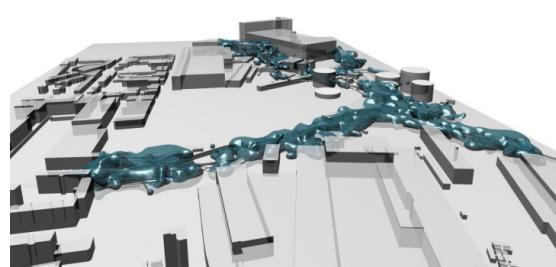


Figure 2:

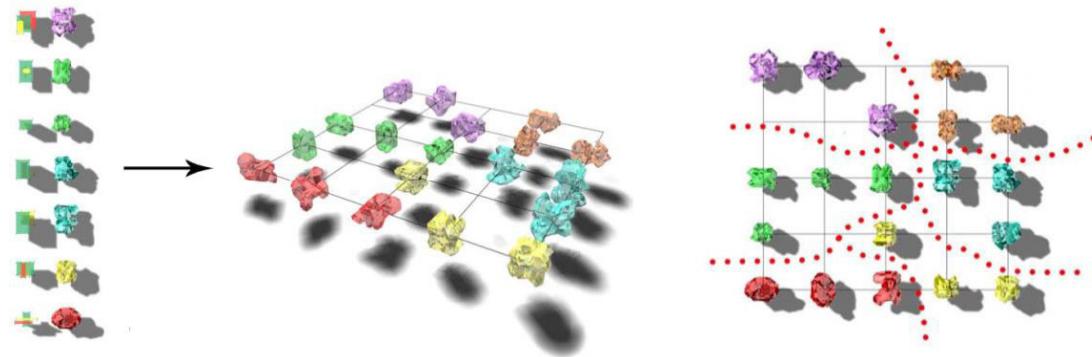
**Self-Organising Space
(SOS) model.**

Left: Accumulation of classified spaces on site (King's Cross, London) showing the path followed by the SOS along the canal. Right: view into the self-generated typology, resulting from cluster wrapping with an implicit surface algorithm.

Figure 3:

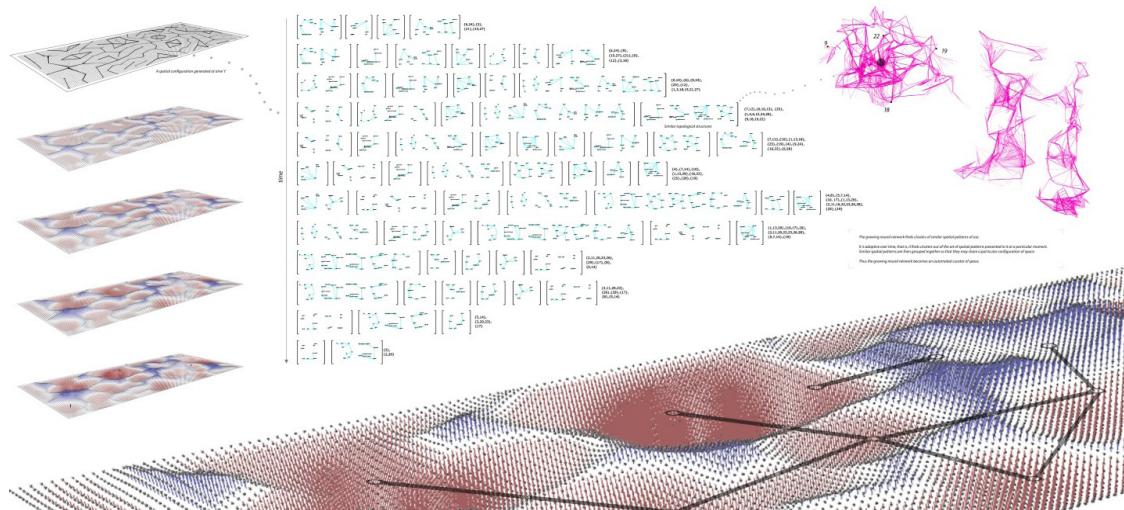
Classifying the resulting instances from the SOS model into a two-dimensional map to reveal similarities between spatial organisations from a site.

(From (Derix, 2004) and (Benoudjitt and Derix, 2004)).

**Figure 4:**

The artificial curator: the image shows at top centre the series of organised spectral graphs derived from the SOM layouts for each exhibition.

The top left shows all spectral graphs organised over time into a layout, to produce (bottom) a potential exhibition layout that is based on the features of the exhibits and their differences/ similarities over time (Harding and Derix, 2010).



The 'artificial curator' was a hierarchical self-organisation model built on three stages of classifications that synthesised,

- the strength of the SOM to learn generic similarities between features (of exhibits);
- the reduction of the resulting spatial graphs representing associative clusters of exhibits into spectral graphs;
- the comparison across exhibits to generate exhibition layouts (Figure 4).

The model allowed the layout of exhibits in a single exhibition space over time depending on the visitor's experiences of the 'sub-features' in the space. Twelve exhibitions were laid out in sequence by the associations of the qualities of the artefacts and their locations, in the hope that visitors could intuitively navigate exhibition spaces by their experience of previous exhibits. The 'artificial curator' represents the most complete spatial classification model to date that integrates analysis with design generation based

on experiential associations. Although its strength lay in the instrumentalisation of associative spatial typologies, the choice of input data and feature space were not as relevant to the architectural profession.

4. Spatial Resilience and Infrastructure Security

The spatial classification system presented in this paper was developed as a part of the EU-funded FP7 project, Resilient Infrastructure and Building Security (RIBS). Aedas R&D was one of a seven member consortium undertaking work to develop effective, integrated security measures aimed at protecting building infrastructures from terror attacks.

As a part of the consortium, the group focused on spatial resilience and infrastructure analysis, in an attempt to analyse a layout for vulnerabilities

based on emergent risk profiles. These risk profiles are generated from the spatial attributes inherent in the layout.

4.1. Case studies

Three case study buildings were used to test the methods outlined in this paper. The methods were developed through implementation on an operational but undisclosed building and generalised over two historic case study buildings. This stage of generalisation was specific to the project and therefore still under the realm of spatial security. The latter two cases were chosen to establish a variation in size, partitioning and circulation structures among the buildings. All the case studies are banking environments that have public as well as private spaces. The case studies are as follows:

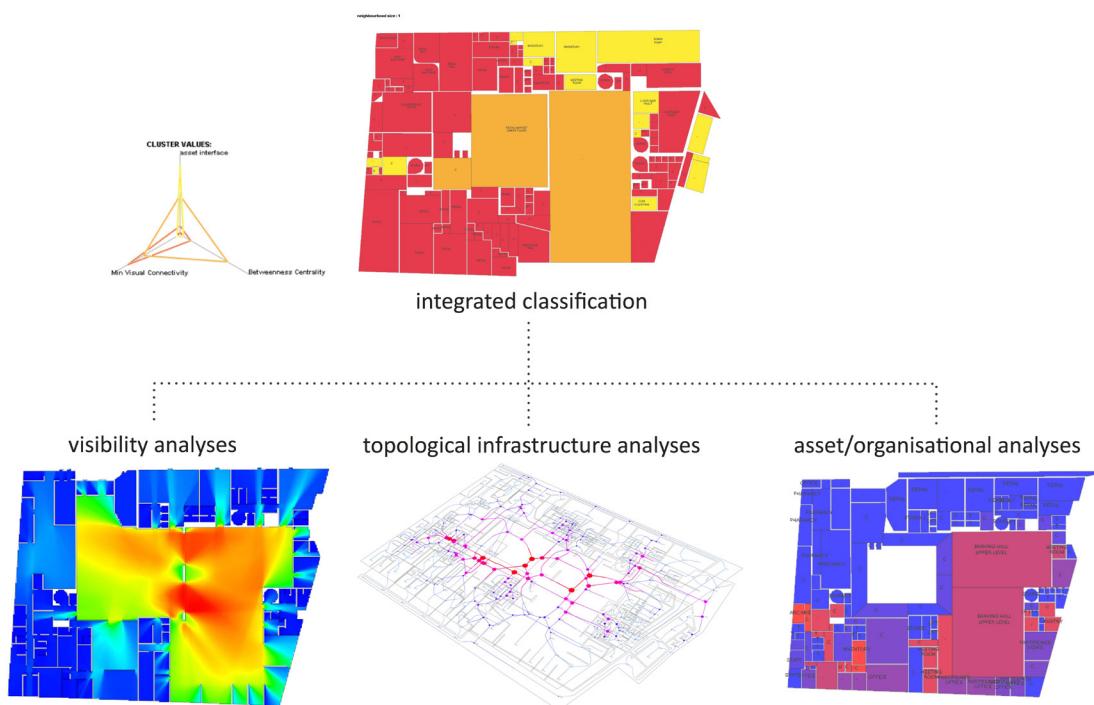


Figure 5:

Integration of three strands of development in the RIBS project.

- Undisclosed European Financial Institution.
- Länssparbanken Falun, designed by Hultman and Holmer in 1970-73, is a medium sized building with an orthogonal grid spanning over two floors. The layout is partly open plan and partly consists of closed rooms.
- The National Farmers' Bank in Owatonna, designed by Louis Sullivan in 1907-08, is a small to medium sized building with an orthogonal grid. It has closed partitioned spaces and clear definition of rooms and functions.

4.2. Input data

The tests outlined in later sections define a single data sample as a location within a building layout. For the spatial security implementation, a single data sample is equivalent to a room-like partition. The case study layouts are divided into a set of two-dimensional partitions (section 4.2.1). Although these are not used as data for clustering directly, they form containers for other data types. In the following examples, the partitions are furnished with visibility, accessibility and asset based attributes. The specifics of these attributes are outlined below (section 4.2.2). They are specific

to this project and can be altered or extended for future work.

4.2.1. Partitioning of the layout

The case study layouts, being continuous regions, needed to be discretised in order to generate containers to which different spatial values could be attached. The discretisation method aimed to provide a common representation of the layout that could be shared between collaborators of the project. Collaborators could then populate the common representation with values relevant to their individual strands. For this reason, the discretisation also needed to be such that the distribution of values across the layout was meaningful for the different data types included.

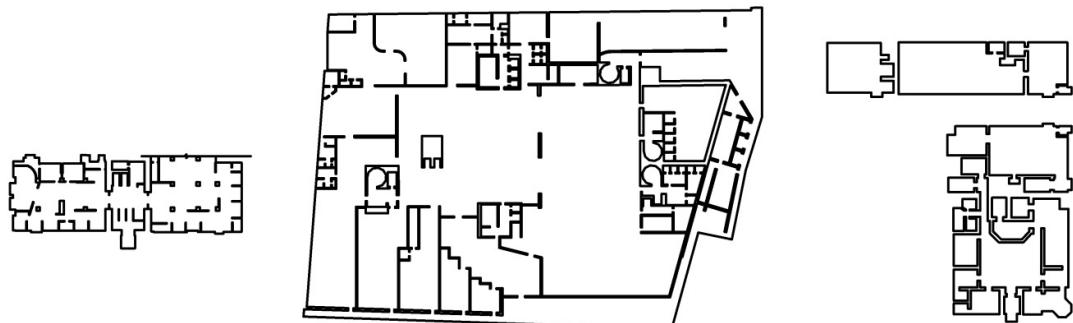
The case study layouts were divided into convex partitions formed by mapping the fewest and flattest discrete convex spaces available (Hillier and Hanson, 1984). This ensured that closed rooms in layouts could be captured as they were, whereas large open plan spaces could be divided for a finer sampling. Consequently, values could be assigned to room-like partitions and the layout could be evaluated as a collection of groups of rooms.

Finally, the partitions were then connected using a simple topological connectivity graph

Figure 6:

Case study layouts at 1:1200.

Undisclosed (left),
Laensspar Banken Falun
(centre), National Farm-
ers' Bank (right).



based on the permeability in the building (Deleuran and Derix, 2013). Each convex partition forms a node in this graph, where each edge represents a physically accessible connection between two such nodes/partitions.

4.2.2. Spatial attributes

Based on precedent research of the Computational Design and Research group at Aedas R&D, different strands of development were carried out for the RIBS project within the group. The data generated from these strands were used as inputs for the clustering model, thereby integrating their outputs in a single comparable platform. Three strands of development were pursued within the group for the RIBS project: Visual Risk analysis (Izaki and Derix, 2013), Topological Infrastructure Resilience (Deleuran and Derix, 2013), and the mapping of asset locations. The relevance and treatment of these values shall be outlined further in this section.

Various assigned and derived values from the strands were collated for each partition. Particular instances of potential spatial risk are defined by selecting combinations of these values. The advantage of using a clustering system for this process lies in being able to compare seemingly disparate values within a single framework. However, while these values are used to define potential spatial situations of risk, they are not considered as definitions of risk in themselves.

4.2.2.1. Visual Risk: Visibility analysis

Visibility values generated in this strand were derived from the geometry of an isovist drawn at a particular location in the layout (Izaki and Derix, 2013) (Figure 8). These values are generated on a dense grid as shown in the figure and later reduced to a single value per partition (this value may be an average, minimum or maximum value within the partition, depending on the nature of the metric). The measures relate to visual levels of perceived

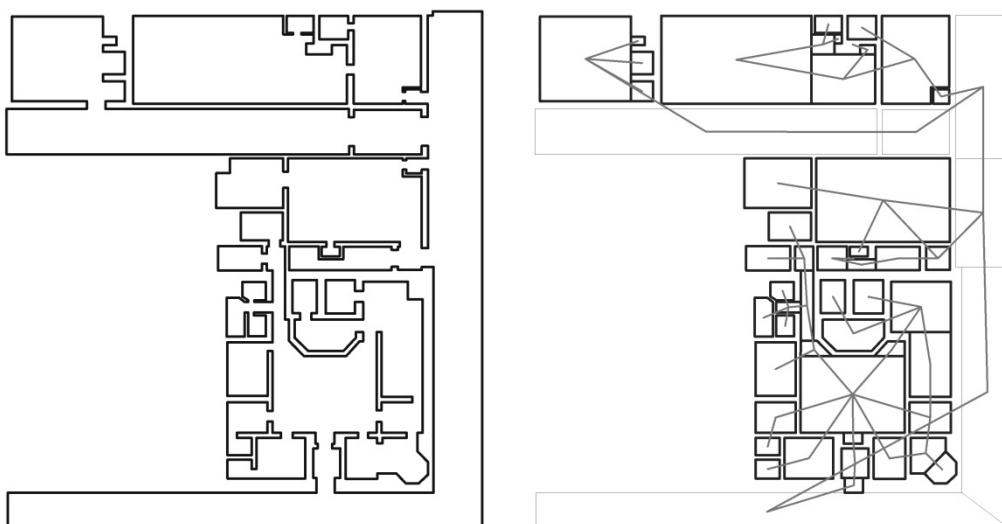
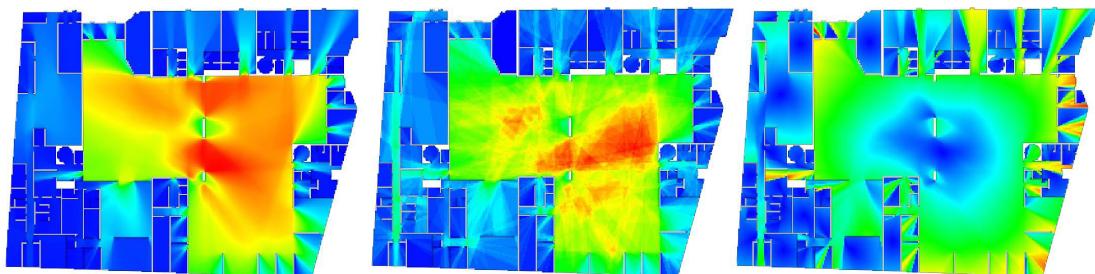


Figure 7:

Convex partitions and access graph for the ground floor of the National Farmers' Bank.

Figure 8:**Visual fields measures.**

Visual connectivity (left), number of permeable edges (centre), visual drift (right) for Länssparbanken Falun.



risk, either as opportunities for an intruder or as vulnerabilities for an occupant. The main spatial values taken from this strand were,

- Visual connectivity: For this development, visual connectivity is defined by the isovist area at a certain location and is used to represent the visual exposure of a location (Benedict, 1979).
- Minimum visual connectivity is given by the area of the smallest isovist amongst isovists at all the locations within the partition. This location can be considered as the least visually exposed or most hidden location in the partition.
- Maximum visual connectivity is defined by the area of the largest isovist in a partition. It is used to represent the maximum visual exposure of a partition.
- Maximum permeable edges in partition: The number of permeable edges of an isovist is defined by the number of occluded edges in its geometry. For the purpose of this research, we consider this to be a measure of visual choice at a given location.
- Maximum visual drift: The metric for drift is comparable to 'drift magnitude' and represents how central a location is within its visible area. This value describes the strength of visual pull drawing an occupant towards the centre of a space (Ostwald and Dawes, 2013).

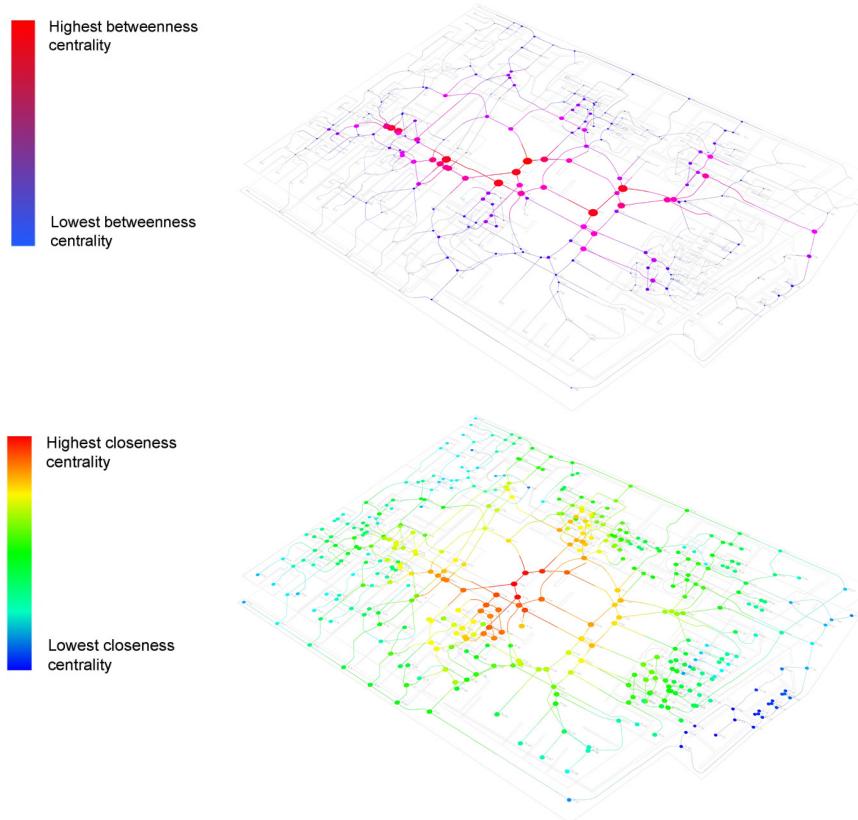
4.2.2.2. Network Resilience: Topological infrastructure analysis

Degree, closeness centrality and betweenness centrality measures were generated as a part of the topological infrastructure analysis strand (Figure 9, 10). They were obtained by conducting graph analysis on a pruned medial axes graph, constructed from the boundary domain of the layout (Deleuran and Derix, 2013). They are equivalent to measures found in graph theory and network analysis. These measures describe circulation properties and are representative of how significant a location is with respect to other points in the layout. They also highlight the resilience of a layout based on availability of otherwise redundant routes that may be used if a prominent link in the network is broken. In the context of spatial security, this could happen when a portion of the building is damaged under attack.

Degree is defined by the number of partitions/nodes a partition/node is connected to.

Closeness centrality is used to represent how isolated or central a partition/node is to the overall layout.

Betweenness centrality is used to represent the most traversed partitions/nodes in the layout. It can indicate which locations are most likely to be used often and therefore indicate strong



Figures 9 and 10:

Betweenness centrality measures (left) and closeness centrality measures (right) for Länssparbanken Falun.

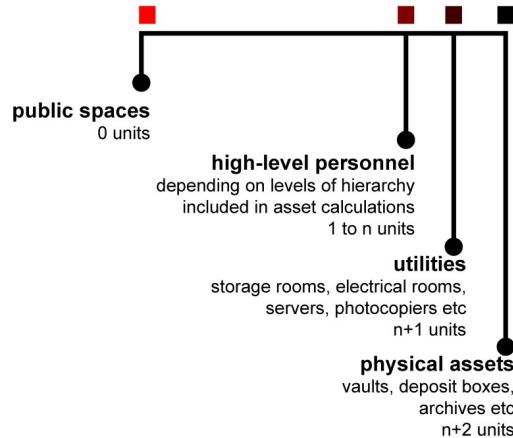
links within the circulation network that may be vulnerable to attack.

4.2.2.3. Asset analysis

Asset values were assigned to partitions as a measure of their vulnerability to attack in terms of having a higher degree of attractiveness to the intruder (Figure 11). The lowest asset values were attributed to public, circulation spaces. Nodes that represent generic functional and service spaces were attributed asset values derived from the importance of functions or the physical assets they accommodate. Utilities including storage spaces, electrical equipment, server rooms, photocopiers, and printers, were given higher asset values as compared to public and circulation spaces.

Physical assets including monetary storage, and archives, were given the highest asset values as they were the most secured assets in the case study layouts. Where available, offices of high-ranking persons in the organisational hierarchy were also given high asset values. This was done specifically for the cases being studied, as there was a threat of intrusion attributed to the offices of high-level personnel. The offices were given lower asset values than the physical assets, because unlike the physical assets they were not secured locations.

A new metric, called an 'asset interface' was set up to represent proximities between high value assets and low value assets (Figure 12). In a threat scenario these interfaces could be indicative of

Figure 11:**Asset value allocations.**

assets that are vulnerable against intruders by being easily accessible from public / circulation spaces. The value was derived from a linear compression of the depth (number of partitions one would have to traverse) between a high value asset and its neighbours. These values were also set up to indicate if assets were distributed or aggregated in a layout.

4.2.2.4 Proximity to entrance

This measure represents the proximity of a partition to the closest entrance. It is defined by the depth between the partition and the closest entrance calculated on the connectivity graph between the partitions.

4.3 Specialised model: Integrated clustering model

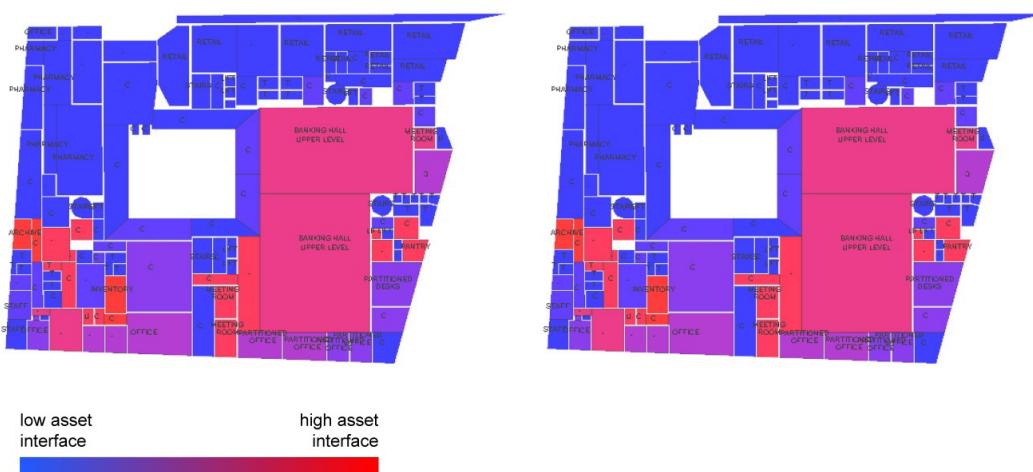
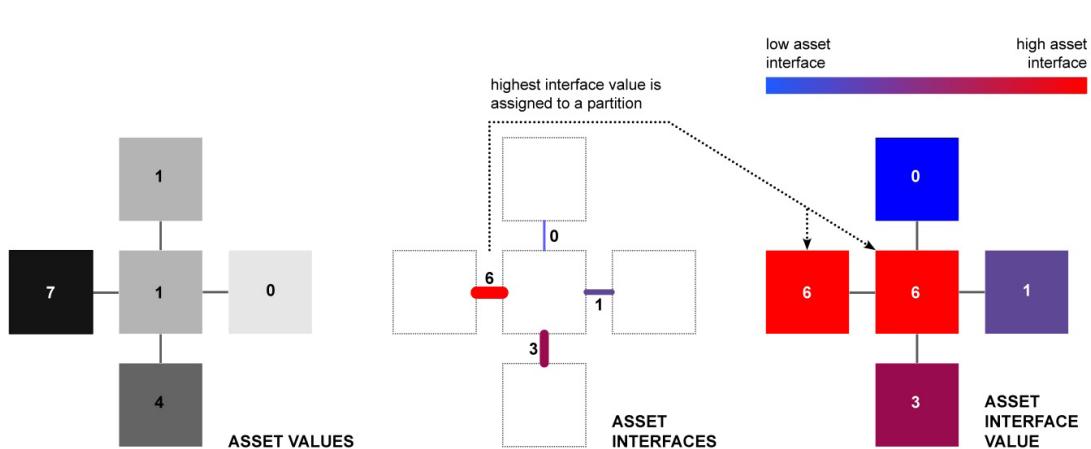
A spatial clustering model was developed as a part of the RIBS project. As previously mentioned, the model was used to integrate the outputs of development work undertaken by the team in the different strands of the project. This was done by defining spatial risk profiles as combinations of values produced from the different models (as described above).

The specific model built for the project consists of a main one-dimensional unsupervised SOM that is used to cluster partitions in the layout based on the input values selected. The clustering model is supplemented by two additional visualisations – a force-directed connectivity graph and a clipped two-dimensional SOM (Mayer et al., 2005). These visualisation models will be explained further in coming sections.

In the interface of the model, a set of attributes can be selected from a list of inputs. The combination selected is then visualised in the clusterer and the two supplementary models. Since the three models are visualising different aspects of the same layout, they are accessible from each other. Therefore, if a partition is selected in one particular visualisation, it is also highlighted in the other visualisations. This allows the user to select a partition and find its neighbouring partitions in a supplementary visualisation to be able to see where it is located in the layout. This is to allow a partition to be queried consistently across the different visualisations.

4.3.1 One-dimensional SOM

The integrated spatial clustering model utilised a basic one-dimensional self-organised map (Kohonen, 1995) to cluster partitions into a pre-defined number of spatial types. Selected values for each partition are normalised and fed to a one-dimensional self-organised map. The resulting outputs place each partition within a particular cluster. All partitions belonging to a particular cluster can be said to have similar values for the selected parameters. The cluster weights define the spatial typology representing its cluster members. Since the SOM used is one-dimensional, the output nodes are structured in a single line, with each node connected to two nodes, one on either side of itself. The first and last nodes are not connected to each other, because of which the



neighbourhood does not wrap around the edge of the map. According to this model, partitions mapped to a node are most similar to other partitions mapped to the same node. Second to this, they are more similar to partitions mapped to the next node than they are to partitions mapped to a node two steps away.

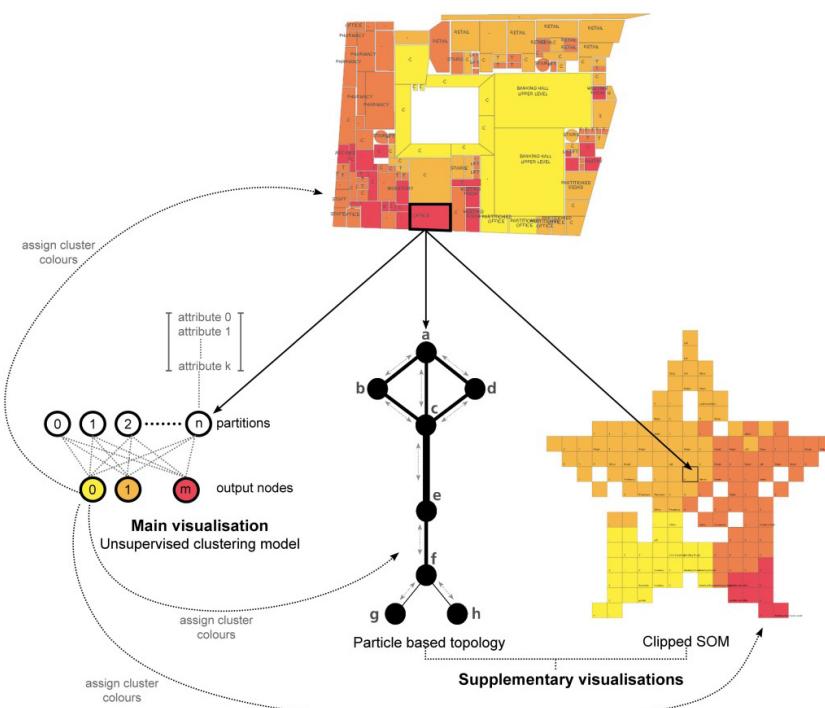
Clusters extracted by a SOM represent data profiles rather than discrete representative values that can be sorted linearly. Relationship between output clusters are determined by the topology

of the network. In a one-dimensional SOM, similarities between neighbouring cluster indices exist without providing exact reasoning for their order. Consequently, cluster definitions in this basic model are not constrained by a fixed range and are therefore not sorted or normalisable. This makes it difficult to compare outputs between iterations quantitatively. For example, the first cluster generated from the selection of data types - minimum visual connectivity, betweenness centrality and proximity to entrances - will not be

Figure 14:

Diagram showing the role of a partition in the different visualisations of the integrated clustering model.

The same partition from the layout is an input data for the main visualisation, a particle in the particle based topology visualisation and a cell in the clipped SOM. Cluster colours generated by the main visualisation are overlaid onto the supplementary visualisations and the main layout.



directly comparable to the first cluster generated for the same selection for a different layout. It will also not be directly comparable to output clusters generated from different data selections such as asset interfaces, maximum visual drift and minimum visual connectivity. However, in this research, visual inferences are made between different outputs that are not directly or quantitatively comparable.

Several clustering and associative models were tested on the data collected and derived from the case studies. For the purpose of this research, a basic one-dimensional SOM was used as it exhibited lesser maximum deviation per cluster as compared to other clustering models tested for the same task. For a simple SOM, the number of clusters needs to be determined beforehand. Testing with four clusters showed most stability during the clustering iterations for all

layouts. It also had the least maximum deviation within each cluster from the mean. While in some iterations, clustering over five clusters showed lower maximum deviation per cluster, the results fluctuated over a large range of values, the maximum of which was higher in most cases than results derived from four clusters.

4.3.2 Supplementary visualisations

The following section describes two supplementary models, used alongside the main clustering model in order to highlight alternative aspects of the data. The force-directed topology graph focuses on visualising the topology of the layout. The model helps to make visual correlations between the various clusters and their topological locations. For example, one can observe the distribution of a cluster to see if it tends to occur largely at end nodes, on major junctions or through nodes etc.

Such patterns could indicate correlations between particular data combinations and the topology of the layout.

The second supplementary model is a boundary clipped two-dimensional SOM. It adds a layer of detail to the main unsupervised clustering model. Through this model, it is possible to read the relative similarities between individual pairs of partitions instead of just being able to say that a set of partitions belong to the same cluster. This is effective for querying partitions within a cluster or partitions at the periphery between clusters.

4.3.2.1 Force-directed topology graph

In this alternative mode of visualisation, the partitions of the layout are represented by particles in a particle-spring system. A spring represents the connection between two partitions if they are one topological step away from each other on the connectivity graph (Figure 16). All the springs are given equal rest lengths and all the particles have mutual repulsion to each other. This mode of representation is similar to a connectivity graph that is force directed, except the length of the edges is not defined by the distance between the centres of the connected partitions. The edge lengths start at unit length and are stretched out by the forces on their end nodes. Over time, the system resolves itself to reveal a topology of the layout, stretching out those edges that suffer the most opposing forces. These edges coincide with most central connections in the layout and are rendered relatively thicker than other connections (Figure 16).

The layout was visualised as a force-directed graph in three dimensions as opposed to a planar connectivity graph to make the complex topology of multiple floor levels visually more accessible. It also visualised directly, although simplistically, the centrality of the graph connections. When partitions are selected, the three-dimensional

force-directed graph reveals correlations such as high asset value locations to public interfaces and their relative topological depth. Since the graph organises itself in three dimensions, being able to rotate the graph in an interactive viewer clarifies the structure better. For the purposes of this paper, images of the three-dimensional force-directed graph have been presented. Although this projection may cause some of the nodes to appear larger than others, they are all of the same size in the model.

The particles are then rendered based on the clustering outputs in the integrated clustering model explained in 4.4.1. This highlights the general locations of partitions of a certain risk characteristic on the graph. The Traer physics library for Java was used to generate this model. The visualisation method was meant to augment the main clustering visualisation by giving the clustering information a more visually accessible topological context.

4.3.2.2 Two-dimensional clipped SOM

The two-dimensional clipped SOM (Mayer et al, 2005) looks at the similarity between the partitions quite like the clustering model in 4.3.1. The difference between the resulting visualisations is that in this SOM, similar partitions occur as neighbouring cells in a two-dimensional grid of cells whose topology is restricted by a pre-defined bounding polygon (imported in a CAD format). Each cell in the two-dimensional SOM represents a partition in the layout. Referring to the general description of a SOM in section 2, each cell in a two-dimensional clipped SOM is an output node.

The model is used to assign every partition in the test layout to a cell in the two-dimensional grid, such that partitions in adjacent cells are most similar to each in terms of the attributes selected. The two-dimensional nature of the mapping adds a level of detail to the basic one-dimensional

Figure 15:

Force-directed graph for two different basic layouts.

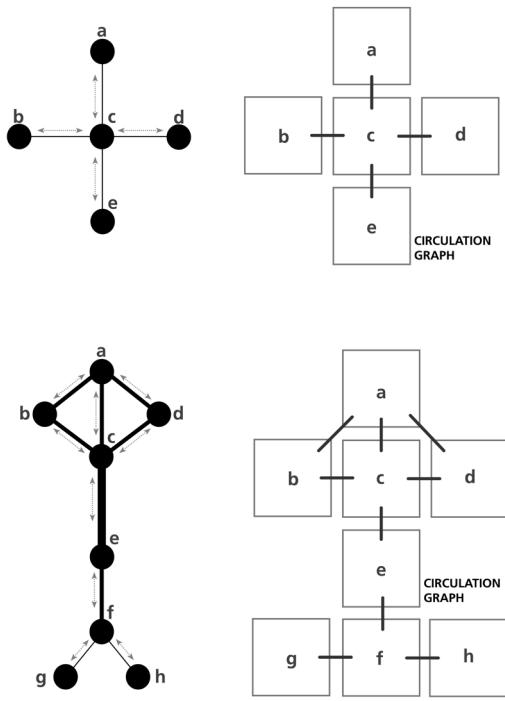


Figure 16:

Three-dimensional force-directed graph for the layout of the National Farmer's Bank.

Each partition in the layout is a node in the graph and its colour is given by the cluster it belongs to, as given by the main unsupervised clustering model. Thick edges are comparable to more central connections in the layout.



clustering model (section 4.3.1) by showing how similar or dissimilar partitions within a particular cluster are. It also shows if two clusters are close to each other on the grid (and are therefore similar) or if particular partitions in different clusters share similarities. The distinguishing feature of this map is that a bounding polygon is used to clip the structure of a regular two-dimensional grid of output nodes. The clipping of the grid with a bounding polygon alters the topology of the network. Altering the topology of the output nodes increases neighbourhood distances within the map, allowing it to represent more distinct dissimilarities between partitions than previously. It also creates corners in the grid topology, such that partitions that are assigned to the cells in the corners are most dissimilar to other cells in the map (Figure 17).

Bounding polygons with sharp corners (a regular star-shaped polygon) have been used to

clip the topology of the two-dimensional SOM. Such polygons create corners in the topology, increasing the distances between cells and allowing most dissimilar partitions to be assigned to these corners. Such topologies can display a greater range of dissimilarities present in the input data as compared to the range they could have accommodated before the clipping. Partitions assigned to the corners of the clipped grid, being most dissimilar to other partitions in the layout, can be representative of the bounds of input dataset. Although any polygon with ‘n’ number of corners would suit this purpose, in Figure 17 and in section 4.4.2, regular star-shaped polygons were sufficient for these tests.

Apart from visualising similarities and dissimilarities between partitions in a layout, the model also shows similarities and dissimilarities between and within clusters produced in the one-dimensional clustering model. Figure 17 shows

a two-dimensional clipped SOM that has been used to map partitions in the Laensspar Banken layout for three spatial attributes – proximity to an entrance, minimum visual connectivity and betweenness centrality. In the figure, high valued assets are not necessarily located near each other and are distributed in groups over the map. This indicates that the partitions containing high value assets do not necessarily share similar spatial attributes. Meeting rooms, pantries, archives and inventories share similar spatial profiles as circulation spaces and toilets. Similarly, parts of the pharmacy, retail market and slot machines share similar spatial profiles. When iteratively analysing design options, such a map is useful to understand the alignment between functional allocation of spaces and existing spatial profiles.

This particular model does not create cluster divisions or cluster centres. It is essentially a continuous two-dimensional adjacency mapping.

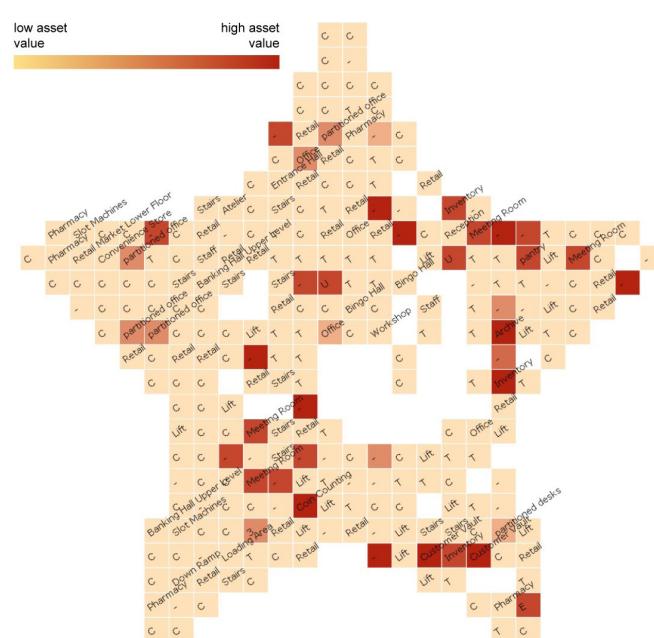


Figure 17:

*Two-dimensional
clipped SOM of
Länssparbanken
Falun with clusters
rendered later*

Cluster values from the main one-dimensional unsupervised clustering model can be overlaid on this mapping to visually verify the separation of clusters outputs.

4.4 Outputs

The specific model outlined above was applied to spatial data in the context of examining analysing security in banking environments. Consequently, specific combinations of data types were utilised to look at potential situations of risk. The case study layouts were clustered based on these specific combinations of attributes. The pre-defined combination of data, could thereby allow a user to query one or more than one layout for levels of the risk defined by that combination. Below are descriptions of some such combinations that were tested. This section aims at describing the combination of values and the potential situation of risk they might represent.

The output clusters are represented by a radar gram, in which the number of radials is equivalent to the number of data types selected. The data is normalised for comparison before processing, resulting in the radar gram having a range of 0 to 1 in all cases.

4.4.1. Exposure of asset interfaces

The combination of asset interface value, betweenness centrality and minimum visual connectivity were used to cluster partitions based on a value that could be identified as 'exposure of asset interfaces'. In such a case, a cluster representing partitions that have a high asset interface value and are at the same time easily accessible (betweenness centrality) and have low passive supervision levels (minimum visual connectivity) could be considered to be at risk.

In the Laensspar Banken Falun layout (Figure 18), cluster 1 contains partitions that have high asset interface values but are not well integrated in

the layout and have low supervision values. These partitions may be considered as highly vulnerable partitions as they are close to high value assets but may not be passively well supervised. Partitions in cluster 3 are centrally located, well supervised and have mid-level asset interface values. These partitions may be considered to be vulnerable, but less so than cluster 1 partitions because although they lie on asset interfaces, they are likely to have high through traffic and be passively well supervised. A large portion of the layout falls under cluster 4 consisting of partitions that have low asset interface values and are poorly integrated while having low supervision levels. Though these partitions can be hidden niches for intruders, they do not have easy access to high value targets.

In the National Farmers' Bank layout (Figure 19), partitions in cluster 3 and 4 have high asset interface values. Partitions in cluster 3 have higher betweenness centrality and minimum visual connectivity values; and are consequently likely to have better passive supervision. On the other hand, partitions in cluster 4 have lower betweenness centrality and minimum visual connectivity values; and can, consequently, be more vulnerable than partitions in cluster 3 as a result of likely less passive supervision.

Between the two layouts, Laensspar Banken has a larger percentage of the overall area of the layout in mid to high vulnerability as compared to the National Farmers' Bank. This is largely because very large publicly accessible spaces like the banking halls lie on asset interfaces in the Laensspar Banken layout.

4.4.2 Proximity of hidden spaces to an entrance

The combination of attributes like proximity to entrance, betweenness centrality and minimum visual connectivity can highlight spaces that are not well integrated visually and topologically, while

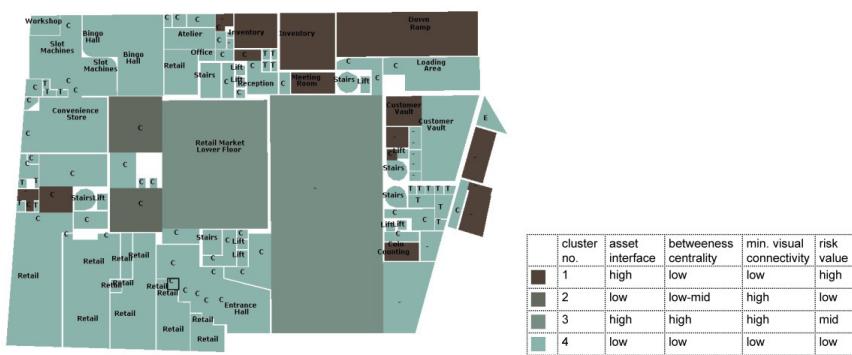


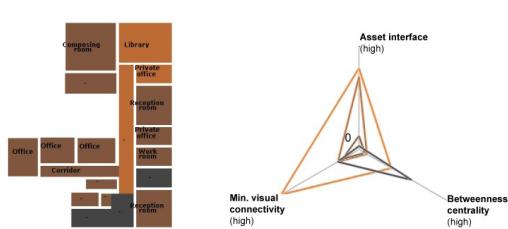
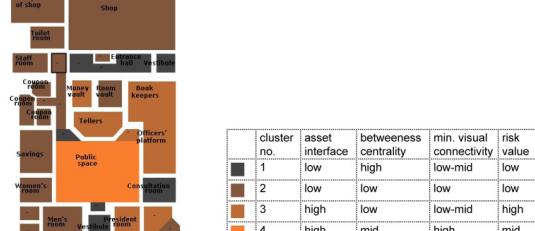
Figure 18:

One-dimensional SOM clustering to define 'exposure of asset interfaces' in Laensspar Banken Falun.



Figure 19:

One-dimensional SOM clustering to define 'exposure of asset interfaces' in the National Farmers' Bank.



being close to an external entrance. The clustering of partitions based on these attributes in the layout below generates the following distribution.

In the Laensspar Banken Falun layout (Figure 20), we see that a large portion of the layout is light orange and it has partitions that are proximal to an entrance while having low minimum visual connectivity and low betweenness centrality. These partitions might exhibit the spatial risk of being hidden spaces that are proximal to an external entrance. In the particle topology graph (Figure 21), we see that all of the red cluster partitions lie at major junction points whereas most of the dark orange cluster partitions lie at the end of branches in the tree. In the topologically clipped

SOM (Figure 22), we see that the light orange and yellow clusters are quite disparate. On the right the map shows the distribution of assets based on similarity. The meeting rooms and inventories share similar spatial properties and are dissimilar to some of the partitioned office spaces which share similar spatial properties. The two customer vaults are not very similar and the archive has the most dissimilar spatial properties as compared to the rest.

4.4.3 Clustering all attributes

One can also cluster the partitions simultaneously based on all the collated spatial attributes, where there is no pre-defined notion of what the

Figure 20:

One-dimensional SOM clustering to define ‘proximity of hidden spaces to an entrance’ in Laensspar Banken Falun.



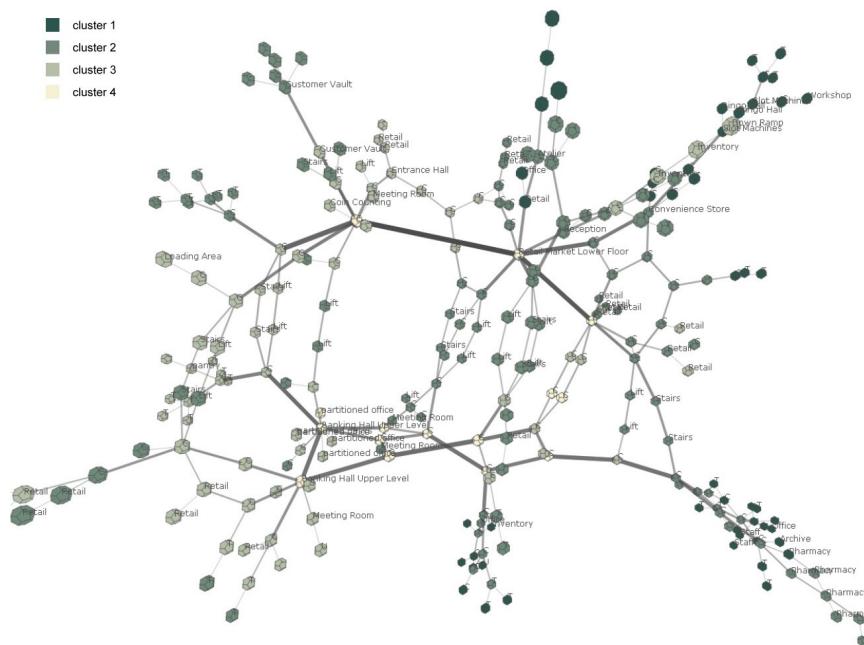


Figure 21:

One-dimensional SOM cluster locations in the particle based topology of Laenssparr Banken Falun.

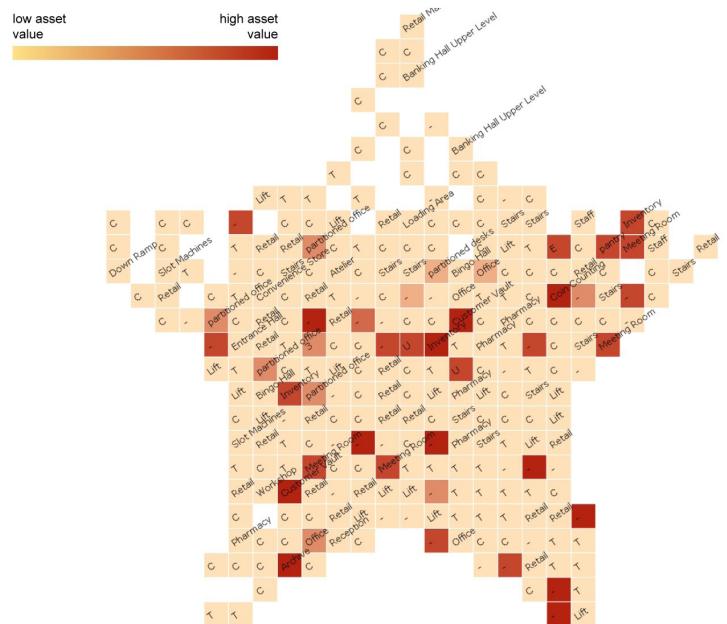


Figure 22:

Distribution of assets in the two-dimensional clipped SOM.

Figure 23:

One-dimensional SOM clustering of nine parameters in Laensspar Banken Falun.



combination of values imply synchronously. The clusters formed as result (Figure 23) contain partitions whose values conform to the cluster values in some aspects and deviate in some aspects. Consequently all members of every cluster have similar values for a portion of the selected properties. These attributes are then used to define the spatial risk of the cluster. The advantage of clustering several/all attributes at once is to allow unexpected associations to be visualised in the layout. Although this method throws up interesting associations when compared to the pre-specified combinations, the implications of each cluster differ between layouts, making

the method more specific to a layout instead of enabling the comparison of layouts.

In the following example, an unconstrained clustering of all available attributes divides the layout into distinct spatial types. All the major circulation areas are categorised together in cluster 4 – having large visual exposure and higher centrality measures. Most of the larger retail partitions in the light orange category have higher centrality measures than those in the dark orange category. They are also slightly closer to the entrances. The partitions in dark red are distinguished by their proximity to an asset interface.

4.5 Conclusions

The Resilient Infrastructure and Building Security (RIBS) project, being a project with divergent strands of research, was an appropriate platform to test the use of clustering systems as a means of integrating several disparate data types in a single model. The model tried to not only display these values together but compare them as well. The benefit of testing a clustering system or any other relevant data-mining system over simply visualising the data types alongside each other lies in the opportunity to cull trends from the data. Additionally, high-dimensional data, that would otherwise reach a limit in legibility with traditional visualisation methods, can be reduced with these clustering techniques.

Some of the outputs described in the paper contain only three dimensions. They were useful for testing the method such that results could be visually verifiable. However, the model is also able to deal with more complex inputs, like those in section 4.4.3. Layouts furnished with several layers of spatial data were successfully reduced to archetypes that were easier to handle and compare between layouts. From a design perspective, highlighting similarities and generating categories may be useful when making functional allocations to particular regions or locations based on the suitability of that location's spatial attributes.

As a model choice, the SOM proved to be a lightweight and transparent model to work with while being easily extendable for future work. The system in its basic form, however is limited by the need to pre-define the number of clusters into which to divide the data. A trial and error method was used to find the appropriate number of clusters for the relevant data combinations. This parameter of the model will have to be re-tested as new data are added to the input collection. This aspect would limit the generalisation of the model over different data types. It can be addressed,

however, by extending the basic system into a more adaptable variation.

In a previous iteration of the model, the clusters were sorted in order to assign a linear gradation in the output values. The intention was to test if the outputs could be structured so as to directly be read as increasing/decreasing levels of risk. This would also make them more comparable to outputs generated from other layouts. Imposing such external constraints, however, overwhelmed the inherent structure of the result. On the contrary, results across layouts were comparable if they were clustered within the same instance of the SOM. In as much as the clustering system aids visualisation of complex, high-dimensional data, much work is needed in making the outputs more comprehensible in practice.

5. Generalised model

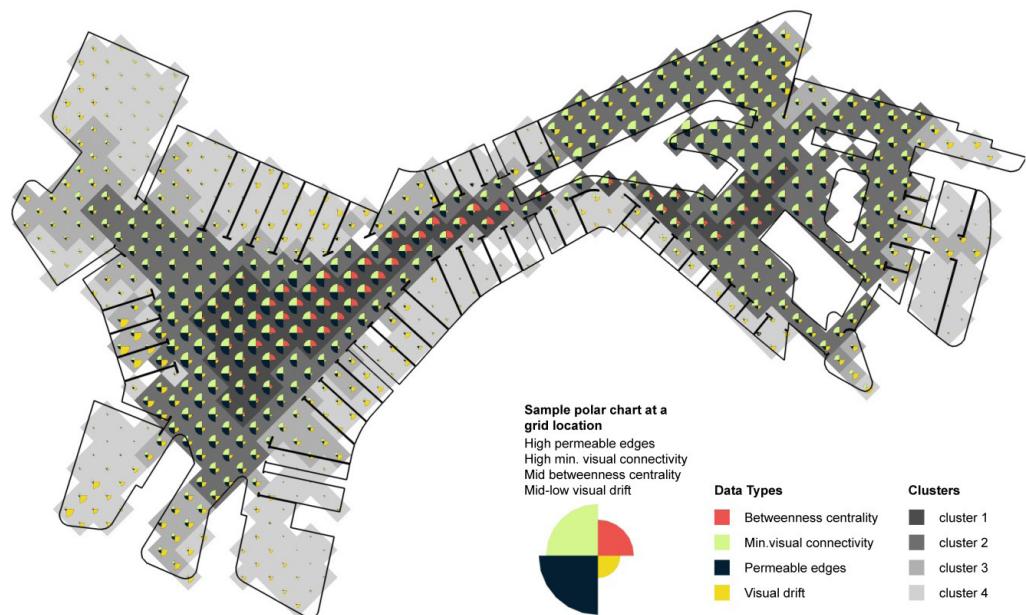
The integrated clustering model was generalised from a security perspective towards application in other sectors. For this purpose, the model was extended to include other numerical data inputs. Additionally, the interface was simplified and made sector-neutral. The general aim of the model was to profile a layout based on any of the selected input spatial attributes with the user determining the significance of the data selections themselves.

The layouts tested under RIBS were reduced to manually determined partitions and spatial values were reduced across these partitions (depending on the nature of the attribute, either an average value, minimum or maximum value was taken). By this process, the inherent trends in spatial values across the open plan spaces of the layouts were camouflaged by the pre-determined divisions of the convex spaces. In the generalised extension, an alternative to the convex partitions was sought. Another reason for this was that, since the partitions needed to be generally produced, it increased the time required for preparing the

Figure 24:

The classification visualisation consists of rendering polar charts at each location on the grid.

Each polar chart shows the distribution of values at that location (see sample polar chart). The locations are additionally separated into four clusters in the background. The clusters are formed based on similarities between the distributions of values found at different locations.



data. A grid of sample points, which was culled by the boundary of the layout, was instead used to locate the values. Spatial values were distributed across these sample points, which were then classified. The resulting visualisation highlighted more relevant partitioning in the open plan areas of the layout than was previously achieved. There was also an allowance for the refinement of the model to be adjusted by changing the spacing of the grid.

The convex partitions, if needed, could be used to compile values within a particular area of interest. A graph representing selected spatial values is rendered at every sample grid point (Figure 24). The model was based on visualisations like Minard's mapping of meat consumption in butchers' shops across Paris (Friendly 1982) and Christenson's (2010) isovist grids. A visual impression of the spatial classifications can be understood directly from the changing nature

of the graphs across the layout. The primary extension of this generalisation exercise is to be found in visualising the spatial values as a grid of distributed graphs and then displaying the clusters in the background.

6. Outlook

From an analytical point of view, it would be useful to extend the current set of building data to include a larger number of layouts, from a variety of functional and formal typologies. Varying the dataset could generate interesting correlations between the traditional typologies of these layouts and emergent ones that would result from a clustering process. Additionally, the datasets could also be augmented with more observational and qualitative data to arrive at more semantic descriptions of configurations. Such non-formal, implicit associative maps can be used to incorporate a form of commonality-based search

into design workflows. Considering the ability of these models to capture non-parametrically defined spatial descriptions, their malleability and lack of type specificity can be applied towards alternatively representing conceptual architectural designs.

Extending Harding and Derix's initial work, types of experiences can be classified across buildings. This should allow architects and urban planners to approach design by planning for experiences. In contemporary complex buildings such as railway stations or airports, building typologies are becoming hybridised to cater for a mix of experiences and services. Those new typologies are no longer rigorously classifiable as a sector, construction or formal category but revolve around sequences of use. This sequence evokes associations between new hybrid spatial typologies and the experiences they are meant to create. Hence, building typologies in the professional sense might give way to spatial experience archetypes.

References

- Benedikt, M. L. (1979), 'To take hold of space: Isovists and isovist fields'. In: *Environment and planning. B, Planning & Design*, Vol. 6 (1), p.47-65.
- Benoudjitt, A., Derix, C. and Coates, P. (2004), 'Human perception and space classification: The perceptive network'. In: *Proceedings of Generative Arts*, Milan.
- Christenson, M. (2010), 'Registering visual permeability in architecture: Isovists and occlusion maps in AutoLISP'. In: *Environment and planning. B, Planning & Design*, Vol. 37 (6), p.1128-1136.
- Cilliers, P. (1998), *Complexity and Postmodernism*, London: Routledge.
- Coates, P., Appels, T., Simon, C. and Derix, C. (2001), 'Dust plates & blobs'. In: *Proceedings of the Generative Arts conference*, Milan.
- Commission for Architecture & The Built Environment (2000), *By Design: Urban design in the Planning System: Towards Better Practice*, London: Department for the Environment, Transport and the Regions (DETR).
- Coyne, R. and Postmus, A.G. (1990), 'Spatial applications of neural networks in computer-aided design'. In: *Artificial Intelligence in Engineering*, Vol. 5 (1), p.9-22.
- Deleuran, A. and Derix, C. (2013), 'Topological infrastructure analysis of the built environment'. In: *11th EAEA Envisioning Architecture: Design, Evaluation, Communication Conference*, Milan.
- Derix, C. and Thum, R. (2000), 'Self-organising space'. In: *Proceedings of Generative Arts*, Milan.
- Derix, C. (2004), 'Building a synthetic cognizer'. In: *Proceedings of Design Computation Cognition and Cognition conference*, Boston MA..
- Derix, C. (2007), 'Approximating phenomenological space'. In: *Proceedings of Intelligent Computing in Engineering and Architecture*, Lecture Notes in Artificial Intelligence, Heidelberg: Springer Heidelberg, p.136-146.
- Derix, C. and Jagannath, P. (2014), 'Near futures: Associative archetypes'. In: *Empathic Space: The Computation of Human-Centric Architecture, Architectural Design*, Vol. 85 (2), New York and London: Wiley and Sons.

About the authors:

Christian Derix (christian.w.derix@gmail.com) is director of the *Design Futures Lab (DFL)* at WoodsBagot architects. Previously, he founded and directed the Computational Design Research group (*CDR*) of Aedas architects in London (UK) since 2004. *CDR* and now *DFL* develops computational simulations for generative and analytical design processes with an emphasis on spatial configurations and human occupation. Derix studied architecture and computation in Italy and the UK and has taught the subject at various European universities since 2001, including University of East London, University College London, Milan Polytechnic, Technical University Vienna and as visiting professor at Technical University Munich. Currently he is associate professor at IE University Madrid and visiting professor at the University of Sheffield.

Prarthana Jagannath has been a designer/researcher at WoodsBagot since June 2014. Before this, she was a part of the Computational Design and Research group at Aedas R&D since February 2011. She holds a bachelor's degree in architecture from the University of Mumbai and an *M.Sc. in Adaptive Architecture and Computation (AAC)* from the Bartlett Graduate Studies, UCL. For her masters' dissertation she studied non-parametric representations of geometry. At Aedas R&D and WoodsBagot, Prarthana has worked on several spatial analysis and computational design projects in the transport, master planning and retail sectors.

- Coates, P., Derix, C., Lau, T., Parvin, T. and Pussepp, R. (2005), 'Topological approximations for spatial representations'. In: *Proceedings of Generative Arts Conference*, Milan.
- Fritzke, B. (1995), 'A growing neural gas network learns topologies'. In: Tesauro, G., Touretzky, D. S. and Leen, T. K. (eds.), *Advances in Neural Information Processing Systems*, Vol. 7, p.625-632, Cambridge MA: MIT Press.
- Gero, J. (1990), 'Design prototypes: A knowledge representation schema for design'. In: *AI magazine*, Vol. 11 (4, Winter), p.26-36.
- Hanna, S. (2006), 'Automated representation of style by feature space archetypes: Distinguishing spatial styles from generative rules'. In: *International Journal for Architectural Computing*, Vol. 5 (1), p.2-23.
- Harding, J. and Derix, C. (2010), 'Associative spatial networks in architectural design: Artificial cognition of space using neural networks with spectral graph theory'. In: *Proceedings of Design Computation & Cognition*, Stuttgart.
- Hillier, B. and Leaman, A. (1974), 'How is design possible?'. In: *Journal for Architectural Research and Teaching*, Vol. 3 (1), RIBA, p.4-11.
- Ireland, T. and Derix, C. (2003), 'An analysis of the poly-dimensionality of living: An experiment in the application of 3-dimensional self-organising maps to evolve form'. In: *Proceedings of eCAADe*, Graz.
- Izaki, A. and Derix, C. (2013), 'Visibility polygon traversal algorithm'. In: *11th EAEA Envisioning Architecture: Design, Evaluation, Communication Conference*, Milan.
- Kohonen, T. (1995), 'Self-organizing maps'. In: *Springer Series in Information Sciences*, Vol. 30, Berlin: Springer Berlin.
- Langely, P., Derix, C. and Coates, P. (2007), 'Meta-cognitive mappings: Growing neural networks for generative urbanism'. In: *Proceedings of Generative Arts*, Milan.
- Laskari, A., Hanna, S. and Derix, C. (2008), 'Urban identity through quantifiable spatial attributes'. In: *Design Computing and Cognition '08*, p.615-634, Springer Netherlands.
- Luhmann, N. (1972), 'The autopoiesis of social systems'. In: Geyer, F. and van der Zouwen, J. (eds.), *Sociocybernetic Paradoxes*, London: Sage.
- Lynch, K. (1960), *The Image of the City*, Cambridge MA: MIT Press.
- Maturana, H. R. and Varela, F. J. (1987), *The Tree of Knowledge: The Biological Roots of Human Understanding*, New Science Library/Shambhala Publications.
- Mayer, R., Merkl, D. and Rauber, A. (2005), 'Mnemonic SOMs: Recognizable shapes for self-organizing maps'. In: *Proceedings of the Fifth International Workshop on Self-Organizing Maps*.
- Ostwald, M. J. and Dawes, M. (2013), 'Prospect-refuge patterns in Frank Lloyd Wright's Prairie houses: Using isovist fields to examine the evidence'. In: *The Journal of Space Syntax*, Vol. 4(1), p.136-159.
- Petrovic, I. and Svetel, I. (1993), 'A sketch of a distributed architectural design system'. In: *Proceedings of eCAADe*, Eindhoven.
- Simon, H. and Newell, A. (1958), 'Heuristic problem solving: The next advance in operations research'. In: *Operations Research*, Vol. 6 (1).
- Friendly, M. (2000), 'Re-visions of Minard'. In: *Statistical Computing & Statistical Graphics Newsletter*, Vol. 11 (1), p.1.
- US General Services Administration (2006), 'Workplace matters', Washington DC: Public Buildings Services.