Partitioning indoor space using visibility graphs: Investigating user behaviour in office spaces

Petros KOUTSOLAMPROS *1
Kerstin SAILER *1
Tasos VAROUDIS *1

(*1 Space Syntax Laboratory, Bartlett School of Architecture,
University College London, London, UK)

Abstract. An abstract representation of interior space is the foundation for any spatial analysis of human activity in such environments. It must capture high level concepts such as rooms, areas and corridors, but also allow for the discrete appearance of human behaviour (for example two people will not walk through the same corridor in the same way). Within the field of Space Syntax three such representations have been proposed, axial lines, convex spaces and visibility graphs. However none of these representations are both unambiguous and allow for aggregating results. Axial lines are reductions of the space into longest lines of sight and convex spaces are ”the largest and fattest convex spaces” possible. While both are meaningful abstractions, they are ambiguous and depend on the person creating them. Visibility Graphs on the other hand provide a uniform unit of analysis by dividing the space using a lattice grid into cells of equal size and connecting the cells if they are intervisible. This representation however does not allow for a meaningful aggregation of spatial human behaviour data, given its very precise nature. We propose a new representation, one which clusters adjacent cells of the visibility graph based on different metrics and thus provides both aggregatable areas and a robust method of creation. We explore how these various metrics and properties of the visibility graph create different types of clusters and specifically examine connectivity and Visual Mean Depth on various types of spaces, from simple shapes, to complex multi-floor buildings. Finally, we demonstrate how this aids the analysis of human activity in indoor spaces by focusing on a large sample of observed activity in office spaces. We argue that this new representation provides a robust but also meaningful foundation for the analysis of indoor space.

Keywords. office space, workspace, user behaviour, space syntax

Introduction

Evidence-based design is the practice of observing and recording human behaviour in order to strategically inform design decisions. Studies show that this practice is becoming more and more popular in the field of architecture. Some studies claim that up to 80% of their respondents identified the need to capture the effects of their designs [5] while others that architects wanted to know more about new tools that enable them to inform their designs [13]. This interest has been especially prevalent in the design of workspaces. Seeking to reduce space (and thus cost) and increase productivity, companies have started looking into ways of collecting information about human behaviour in their workplace and using that information to re-design or optimise the office configuration.

In the search for patterns in human behaviour as they relate to the spatial configuration of office spaces various metrics have been developed. Initially the developed metrics were simple, such as the proximity between workers [1] and the existence of barriers [7], but eventually more complex ones appeared, such as zone and path overlap [11][10]. These metrics have shown promise in some cases but they have only been examined in a handful of spaces, tend to be unique to each study and do not thus reveal generic patterns that are true for any workspace [15]. The field of Space Syntax has examined more complex methodologies, but those too have mainly been tested across small samples [8]. In cases where the methods were tested in larger samples the results tended to be inconclusive, or work with different subsets of cases [12][16]. Therefore, the absence of generic patterns can not only be attributed to a lack of data, but also to the way of the methods themselves were constructed.

One of the problematic areas in the existing methodologies is the representation chosen to allow for the comparison of human activity to the configuration of the space. A few have been proposed, especially within the field of Space Syntax but they vary in how much detail they capture and how well they are suited for the appearance of human behaviour which occurs in distinct locations. The classic Space-Syntax representation of longest lines of sight [9] (axial-lines) focuses on movement rather than occupancy of space and is very abstract thus providing very little of the geometric intricacies of space. Convex spaces on the other hand provide slightly more detail and focus on occupancy, but they can not be re-created reliably [14]. The representation suggested by Turner et al. [18] known as Visibility Graph Analysis (VGA) captures much more detail but, because it is based on a lattice grid, it is inflexible when trying to identify larger patterns of human behaviour.

In this paper we suggest a spatial representation based on VGA that retains the amount of detail captured but also allows us to examine these larger patterns. We construct this representation by grouping continuous parts of the lattice grid into larger areas using the existing VGA metrics. These areas serve as the main unit of analysis and allow for a compromise between the amount of geometric detail captured and flexibility when capturing human behaviour.

We evaluate the results on a large dataset that contains information about human behaviour and configuration of office spaces. Our aim is to reliably identify whether areas that have the same properties attract the same behaviour, for example whether areas more central to the building configuration attract more interaction or movement. We test across the dataset for generic patterns but we also examine each case separately with the aim of understanding whether this methodology can be used to predict activity for any single given site.

The next section will review the existing methods and representations and provide details on their strengths and weaknesses. It will be followed by two sections describing the dataset used for the evaluation, and the suggested representation and methodology in detail. We will explain how this representation is generated as well as which aspects of it are variable and can thus be tuned to better approach the problem at hand. The final three sections describe the statistical testing framework used and its outcome, discuss the results and their general implications in the field of spatial analysis in the workplace and lay down future plans for the research team.

1. LITERATURE REVIEW

Human activity in space may be recorded in many ways, through sensors [4], video [2] or participant observations [20]. In all these cases, the output captured is expressed spatially as points or traces on a plan, depending on the method. In order to compare this output to the configuration of a space, a common framework is needed that allows both the human activity and the properties of the space to be quantified using the same general unit of analysis. Office spaces are very complex with many more spatial parameters affecting human activity than the ones examined in early research, [7][1] highlighting the need for a more general framework, a reliable and fine-grained representation of indoor space.

Within the field of Space Syntax three such representations have been proposed, axial lines, convex spaces and visibility graphs.

Axial lines were suggested by Hillier and Hanson in The Social Logic of Space [9] as the longest lines of sight that go through every space. These lines may be generated through an all-line analysis of the space, which connects the vertices of the walls of the plan between them and eliminates all but the minimum required to reach every space in the building. Given their one-dimensional nature they can very reliably represent linear spaces, such as corridors, as a single unit, but can not unambiguously represent square-like spaces such as the large open-plan rooms found in office spaces. Axial lines were also made to represent movement rather than occupancy, thus although they can be used for indoor spaces, they are more suitable for urban scales for example to represent streets.

Hillier and Hanson identified that limitation and suggested a complimentary approach, what they called ”convex spaces”. These are ”the fewest and largest possible convex spaces” that can be carved out of a space. The process of constructing such spaces has been shown not be repeatable reliably by Peponis et al. [14] who examined the various existing approaches and highlighted the ambiguity in the process, evident in the example seen in Figure 1.

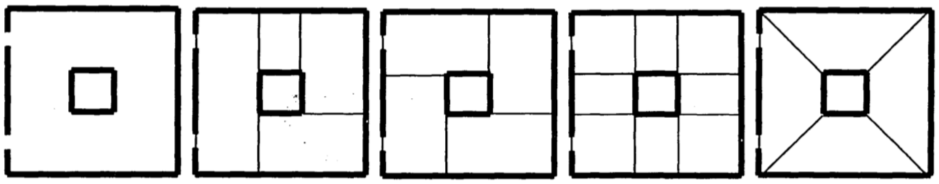

Figure 1. Sample plan and its possible minimum partitions by Peponis et al. [14]

While the lack of detail in both these representations can be thought of as a weakness, it may also be seen as a strength. The fact that they abstract so much of the geometric detail allows them to act as general units of analysis. In the case of axial-lines, one may compare the properties of a line (connections to other lines, depth) to the number of people found in its proximity. Convex spaces are also useful in that regard because they allow for comparing the properties of these spaces to the number of people within them.

Visibility graph analysis was developed as an alternative by Turner [18]. This method suggested splitting a space using a lattice grid and connecting the cells if they are inter-visible (through isovists as described by Benedict [3]), forming thus a dense graph. The underlying lattice grid provides a uniform unit of analysis (a cell) which also allows for a more detailed representation of the spatial configuration. On the other hand, it also creates a requirement to choose between the amount of detail and its ability to act as a general unit of analysis. If the lattice grid is created with small cells then it can capture more of the geometric detail of the space, but the distinct locations of human behaviour will be distributed to a small amount of cells that are in close proximity. This can be seen at the left in figure 2 where the observations of people moving (blue dots) can only be assigned to the cell that is closest to their centre leaving the adjacent ones (pink) empty, although in terms of spatial configuration they are likely very similar to the red ones. If a large cell size is used, then more observations of human behaviour can be attributed to the same cell, but not as much detail can be captured. A large cell size also prevents the representation from capturing continuous space effectively. If for example the cells are larger than the door openings then the space on either side of the door will be inferred to be two distinct spaces as seen on the right in figure 2.

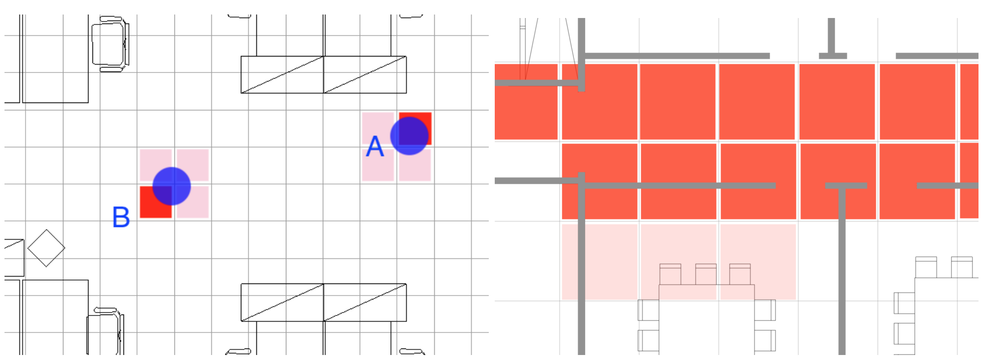

Figure 2. VGA Cell size issues. Left: Observations of people moving (blue dots), stepping on specific squares in the VGA lattice grid (red), while adjacent ones (pink), with similar values are ignored (cell size = 0.45m). Right: If the cell size is large enough (1.5m) then the corridor and the meeting rooms are viewed as different spaces

In previously published research we utilised VGA to identify patterns of human behaviour in the sample of office spaces also used in this study. We initially tried ”smoothing” the human activity across the lattice grid using a gaussian kernel function [12]. This allowed cumulative effects to appear (see figure 3) i.e. all cells were affected by the observations of human activity but the effect was greater for the cells closer to that activity. In a second study [16], we examined the effect of human activity without this smoothing, by matching activity to the lattice grid cells directly, as seen in figure 4. In both these cases the statistical tests returned insignificant results, stemming from confounding factors, but also by the lack of a proper method to generalise the presence of human behaviour.

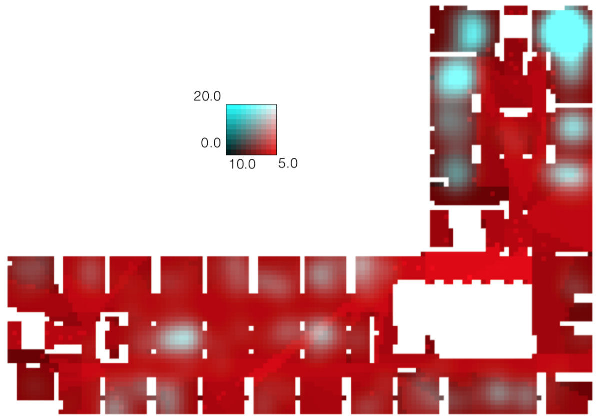

Figure 3. Visual Mean Depth Interaction density from Koutsolampros et al. [12]

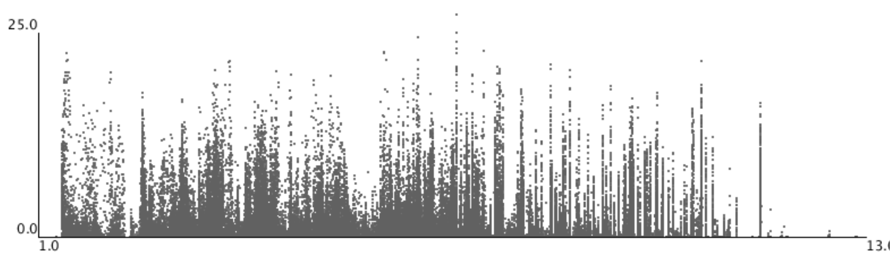

Figure 4. Visual Mean Depth (x-axis) and Interaction density (y-axis) for 27 case studies combined from Sailer et al. [16]

The small-scale decisions that happen during the design process specifically require practical spatial units that can also allow for meaningful generalisations of human behaviour. The above mentioned lattice grid provides units that are too small, making any insights impractical when a designer is planning for whole rooms such as meeting rooms, kitchens or workspaces, that contain many more cells. Hillier and Grajewski [8] used larger units (whole floors and buildings) in a study about configuration and human behaviour for offices in the UK, Scandinavia and the US, thus their results would only be useful in larger strategic decisions but not in the case of the many small companies that typically occupy a few rooms or floors within a building.

2. DATA

To evaluate our representation we examine an office-space dataset provided by Spacelab, an architectural office and consultancy in London, UK. The sample contains 34 different cases (sites), from 29 companies across the UK, compiled from 2012 to 2017. The companies examined vary in size (50 to 2700 desks) and come from different industries, such as Media, Advertising, Technology, Legal and Finance.

There are two types of data for each case: observation data collected by participant observation [20] and visibility graph analysis. The observation data is collected usually over a period of five days, every one hour for eight hours, and it contains information of where people sit, stand, walk and interact as points on a plan. Visibility graph analysis has been carried out with various versions of Depthmap [17] and depthmapX [19] that were available when each case was collected. For every site we examined two VGA metrics: Connectivity and Visual Mean Depth.

3. METHODOLOGY

Given that we consider the existing general units of analysis inadequate to capture both enough detail of the spatial configuration and sufficiently generalise human behaviour, we developed a solution that lies between the existing representations. Relying on the existing metrics of the VGA we group the various cells into different continuous areas, creating partitions of space similar to the convex-space representation. With this new representation we may now aggregate the observations of activity by the number of times it occurs within that specific group of cells, instead of matching each activity to a specific cell. This alleviates the problem of the activities happening in discrete cells and allows for general units of analysis comparison (areas) that more accurately describe the space than axial lines or convex spaces.

The process for generating these areas follows four steps:

- construct a visibility graph and select a metric i.e. Connectivity;
- split the distribution of the metric into two parts according to a specific rule i.e. by the median (this creates areas of high and low of the specific metric);
- run a modified blob-detection algorithm that assigns areas to the cells depending on whether their adjacent cells belong to the same part of the distribution (low/high);
- merge areas that are linked but are on different floors.

We initially tested the first three steps of this process across a small sample of simple-shaped plans with the results visible in figure 5 below. For each of the simple plans (first row) the connectivity of the various spaces is calculated (second row) and split according to the median (third row). The modified blob-detection algorithm is executed with the split pixels as input, resulting in the different areas seen in the last row.

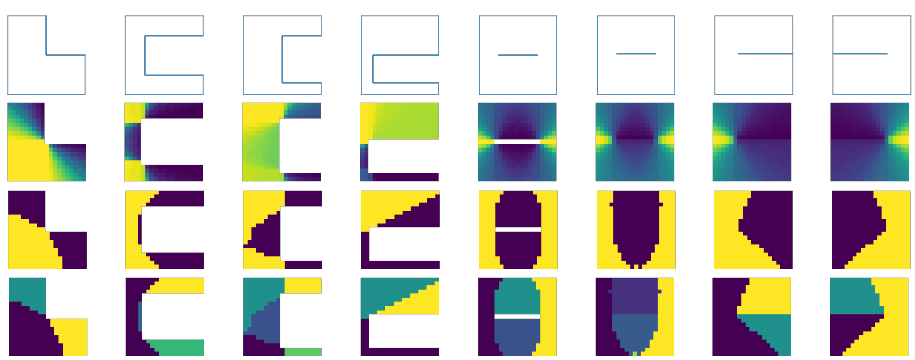

Figure 5. Per row: Plans, Connectivity metric, split by median and final areas

We examined two visibility-graph metrics, Connectivity and Visual Mean Depth as well as their combination. Connectivity of a cell is the amount of visible cells from that cell, and it can be thought of as a proxy for the size of visible space at each point. Visual Mean Depth on the other hand is the average number of turns required to reach every other cell in the building. We also combine Connectivity and Visual Mean Depth into a single representation to achieve a more nuanced separation of spaces. Connectivity alone does not allow us to differentiate open-plan spaces from a long corridor and Visual Mean Depth focuses mostly on their centrality. With this combination we can identify those spaces as different types independently of office characteristics.

The cells are grouped by splitting the distribution of each metric and assigning the grouping to each cell. This creates semantically continuous spaces, such as areas with high or low visibility (high / low connectivity) or areas that are deep or shallow (high / low Visual Mean Depth). We split the distribution in various ways to test whether these splits affect the generation of the areas: by the mid-range, the mean and the median. The distribution for the two metrics for all cases can be seen in figure 6 with the three splitting points.

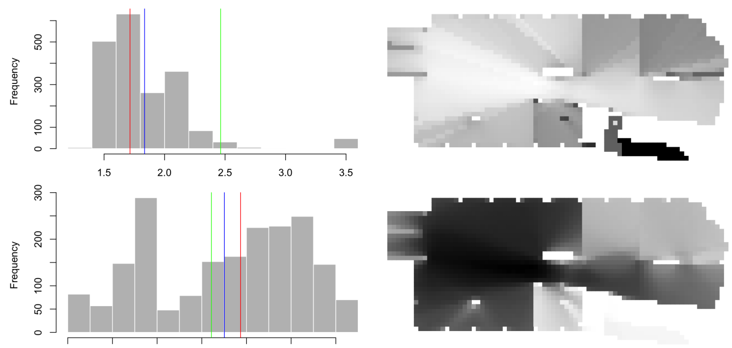

Figure 6. Distribution histograms (left) and values on plan (right, darker is higher) for Visual Mean Depth (top) and Connectivity (bottom). The different splits are shown in the histograms, green for Mid-Range, red for Median and blue for Mean

The distribution of Visual Mean Depth tends to be heavily skewed to the left and thus a mid-range split tends to create areas that are uneven in size, with the very high values in one group and the rest in another as seen in the first row of figure 7. The median provides a better split with half the values on the high area and half in the low, which in effect creates many more groups of similar value (figure 7, second row). We also tested the mean of the distribution as a compromise between the two (7, last row).

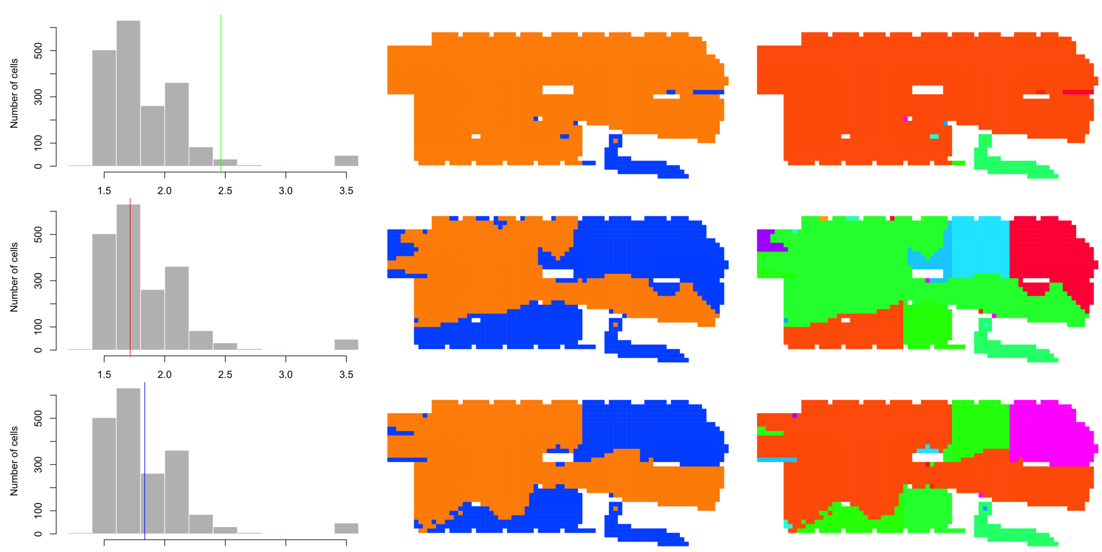

Figure 7. Visual Mean Depth histogram (left column), splits (middle column) and final areas (right column) for the three splits: Mid-Range (top row), Median (middle row) and Mean (bottom row)

The final stage of the process is an algorithm commonly used for blob-detection as described by A. Greensted [6]. In the typical use case it is employed as a computer-vision algorithm to identify and detect "blobs": areas of similar colour in images. Figure 8 shows a part of the process. Through this process the cells of a lattice grid are classified according to the colours of their adjacent neighbours. In our method the types that were generated from the above mentioned split (high/low Connectivity, VMD) are used as the colours of cells. Thus, if the neighbours of a cell are found to be of low connectivity then that cell is classified with low connectivity as well.

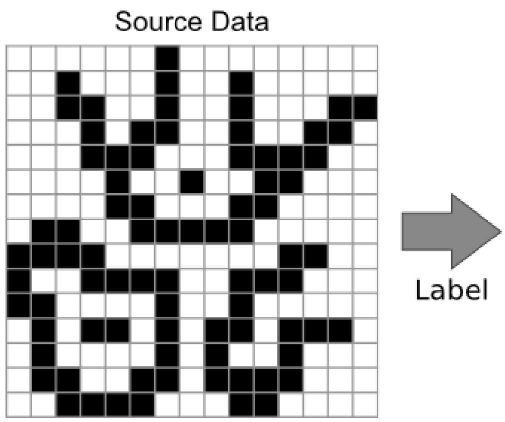

Figure 8. A part of the process of detecting "blobs": continuous areas of the same type by A. Greensted [6]

Various alterations have been introduced to this algorithm for it to work in this specific domain. In contrast to the typical use-case our lattice grid is not completely filled as in the case of a camera image, thus, the exterior space is considered a type in itself. Also in contrast to the typical use-case is the nature of the adjacency of the cell. While in blob-detection two cells are adjacent if there's no gap between them, in our case the true adjacencies are provided by the visibility graph. While the centres of two cells can be at a distance of one cell (i.e. the cells have no gap between them) it is possible they are not actually inter-visible. In cases where the cells are larger than the widths of the walls it is very likely that two adjacent cells will end up in the same split but with a wall between them. In this case the blob-detection algorithm will assign them to two different areas given that perceptually these are actually different spaces. The last alteration of the algorithm is to allow it to merge areas that are linked between floors, if they belong in the same split group, are thus considered part of the same continuous space.

The examined output metric is the density of people observed within a specific area over one hour. We examine two types of activities, movement and interaction. We define movement as the number of people observed moving or standing during that hour, and interaction as the number of people seen interacting. Given that the size of the areas differs we calculate the density by dividing the numbers of people carrying out each activity with the size of that area in m^2. We also divide by the total number of rounds of each observation, therefore the units of the density are people per m^2 and round.

4. STATISTICAL RESULTS

We tested against nine scenarios: all the possible combinations between the three metrics (Connectivity, Visual Mean Depth and their combination) and the three splits (Mid-Range, Median, Mean). In cases where the split only results in two types of areas (high / low) we examine with a t-test how different the areas with high values are from the low, in relation to the number of people they attract. In cases where the split results in more types (high Connectivity - low VMD / high Connectivity - high VMD etc.) we test whether the different types differ with an ANOVA. The tests are all independent given that we are examining non-paired populations of different sizes.

We test across the aforementioned dataset by grouping all the areas found in all sites along with their respective densities of people. This way we examine whether patterns can be found that are generic and valid for all types of spaces. We also apply this methodology on a case-by-case basis in order to understand whether it is reliable for an architect to use these tests within a specific project.

As we can see from the histogram below (figure 9), the distribution of the density in each area is heavily skewed to the right. This is a standard effect encountered when working with count data and can be countered by taking the logarithm of the number in question.

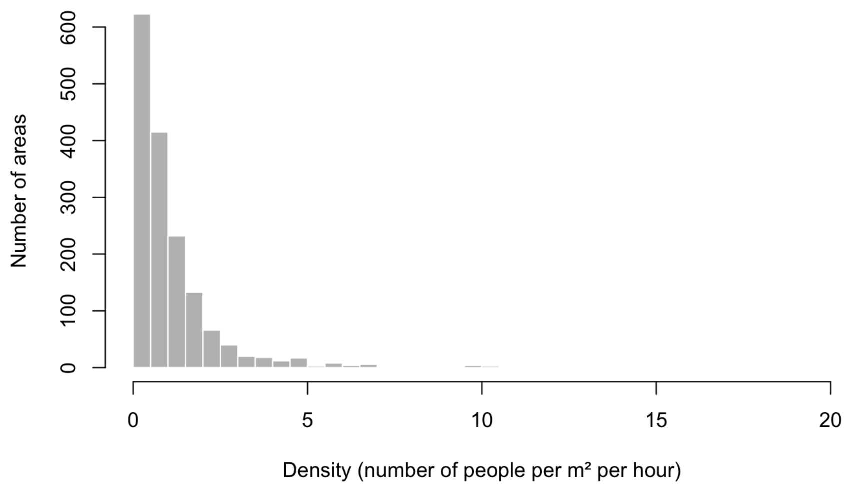

Figure 9. Numbers of areas split by the combination of the means of Connectivity and Visual Mean Depth, grouped by the density of people they attract.

The overall results of the tests for the density of people moving can be seen in table 1 below. Independently of the split used all the tests are significant. Also independently of the split we find that the order of the types in relation to the amount of people moving remains stable. This means that when the split is with Connectivity, areas with low values attract fewer people, specifically we observe around half the density. When the split is with Visual Mean Depth, areas with low values attract more people, around one and a half the density seen in areas with high values. The results for the combined metric are also relatively consistent, with the lowest density of people seen in areas with low connectivity and high mean depth, followed by low connectivity and low mean depth. The last two groups (high connectivity - low mean depth, high connectivity - high mean depth) seem to be interchangeable depending on the splitting method, but the split by the mean seems to be more reliable as it has a slightly higher R^2 value and more significant cases when those are examined one by one (see table 2).

	Connectivity Mid-Range	Connectivity Median	Connectivity Mean	VMD Mid-Range	VMD Median	VMD Mean	Connectivity-VMD Mid-Range	Connectivity-VMD Median	Connectivity-VMD Mean
p-value	0.000	0.000	0.000	0.000	0.006	0.009	0.000	0.000	0.000
t or R²	-6.874	-6.770	-6.565	3.987	2.745	2.636	0.022	0.021	0.025
Ordered types	L, H	L, H	L, H	H, L	H, L	H, L	LH, LL, HH, HL	LH, LL, HH, HL	LH, LL, HL, HH
Means	0.26, 0.46	0.24, 0.39	0.23, 0.41	0.22, 0.30	0.27, 0.34	0.29, 0.36	0.27, 0.28, 0.31, 0.45	0.25, 0.30, 0.31, 0.43	0.24, 0.29, 0.37, 0.43

Table 1. T-tests and ANOVAs for overall movement (L for Low, H for High, Means in number of people per m^2 per hour)

The following table (table 2) shows the p-values of the tests on a case-by-case basis, ordered by the number of people counted in each case. While a few cases show significant results, none of the splits works reliably for all. Most of the cases that follow the general rule seem to be the larger cases pointing perhaps to a requirement for the project to have a critical mass to accurately extract such results. The scenario that best fits most cases is Connectivity split by the mean, followed by the various cases split by Visual Mean Depth. We also ordered (but not show) the table by the average mean depth of each site, average connectivity and amount of teams created, but they did not seem to reveal useful patterns.

	Number of people	Connectivity Mid-Range	Connectivity Median	Connectivity Mean	VMD Mid-Range	VMD Median	VMD Mean	Connectivity-VMD Mid-Range	Connectivity-VMD Median	Connectivity-VMD Mean	Significant tests
All	**91310**	**0.000**	**0.000**	**0.000**	**0.000**	**0.006**	**0.009**	**0.000**	**0.000**	**0.000**	**9/9**
67	**137**							**0.922**	**0.629**	**0.887**	**0/3**
28	315				0.466	0.498	0.489	0.994	0.992	0.948	0/6
17	357	0.015	0.036	0.014	0.043	0.335	0.404	0.167	0.682	0.148	4/9
32	379		0.088	0.372	0.024	0.514		0.024	0.551	0.776	2/7
65	453		0.135	0.629	0.807			0.916	0.848	0.827	0/6
54	504	0.005	0.006	0.002	0.074	0.375	0.137	0.051	0.058	0.076	3/9
85	571		0.479	0.943		0.042	0.075		0.434	0.053	1/6
25	604	0.166	0.625	0.578	0.003	0.875	0.554	0.323	0.255	0.264	1/9
47	733	0.418	0.202	0.225	0.156	0.140	0.284	0.832	0.142	0.607	0/9
66	735	0.466	0.386	0.191	0.886	0.219	0.525	0.369	0.322	0.386	0/9
23	808	0.074	0.495	0.933	0.526	0.754	0.862	0.331	0.995	0.903	0/9
57	834	0.063	0.145	0.063	0.573	0.870	0.994	0.219	0.591	0.592	0/9
46	887	0.333	0.085	0.069	0.802	0.717	0.821	0.601	0.681	0.487	0/9
44	920	0.729	0.401	0.710	0.585		0.324	0.990	0.469	0.658	0/8
63	925		0.647	0.731	0.210	0.065	0.185	0.582	0.323	0.505	0/8
24	1001	0.688	0.156	0.080		0.506		0.007	0.049	0.000	3/7
27	1010	0.456	0.189	0.515	0.334	0.312	0.390	0.766	0.609	0.062	0/9
71	1480	0.851	0.371	0.717	0.681	0.534	0.494	0.783	0.060	0.120	0/9
29	1521	0.129	0.515	0.079	0.024	0.018	0.019	0.059	0.063	0.052	3/9
11	1635	0.003	0.579	0.012	0.003	0.004	0.001	0.014	0.006	0.003	8/9
37	1688	0.007	0.213	0.031	0.005	0.018	0.002	0.002	0.001	0.000	8/9
80	2023	0.892	0.550	0.041	0.255	0.756	0.034	0.329	0.904	0.035	3/9
73	2049		0.000	0.023		0.044	0.385	0.792	0.005	0.001	5/7
18	2197	0.105	0.727	0.394	0.977	0.027	0.089	0.269	0.039	0.201	2/9
45	2568	0.917	0.675	0.709	0.182	0.001	0.014	0.720	0.209	0.843	2/9
61	2871	0.191	0.037	0.014	0.230	0.259	0.429	0.343	0.027	0.030	4/9
49	3279	0.471	0.738	0.692	0.286	0.389	0.107	0.412	0.285	0.440	0/9
50	3309	0.553	0.553	0.194	0.092	0.051	0.030	0.688	0.662	0.034	2/9
51	4888		0.277		0.264	0.776	0.036	0.062	0.348	0.006	2/7
36	5748	0.000	0.000	0.000	0.004	0.033	0.025	0.000	0.000	0.000	9/9
75	9074	0.002	0.044	0.038	0.120	0.502	0.710	0.000	0.030	0.045	6/9
58	9844	0.686	0.057	0.130	0.000	0.180	0.311	0.085	0.931	0.988	1/9
68	10579	0.007	0.113	0.114	0.051	0.085	0.273	0.182	0.023	0.175	2/9
84	15384	0.010	0.003	0.021	0.295	0.088	0.232	0.168	0.801	0.774	3/9
Significant		**8/26**	**7/32**	**10/31**	**8/30**	**8/31**	**8/30**	**6/33**	**9/34**	**10/34**	**74/281**

Table 2. T-tests and ANOVAs for movement per site (p-values); empty cells are cases where the amount of groups were insufficient to run the test

The same two tables follow for interaction. For this behaviour the general results (table 3) are not as consistent, with only two scenarios appearing significant: Connectivity split by the median and the combined metric split by the means. In the former scenario low connectivity areas were found to attract around 80% fewer people interacting; in the latter, the fewer people interacting were found in low connectivity, low Visual Mean Depth areas. These are areas quite central to the building but without a lot of visible space. The two intermediate types of areas (low connectivity-high Visual Mean Depth and high connectivity - low Visual Mean Depth) were found to attract about the same amount of people interacting. The greatest densities of people interacting were found to be in areas of high connectivity and high Visual Mean Depth, areas that are large but not very central to the building. The combined metric by mid-range also provides significant results but, as mentioned previously, the areas created are greatly imbalanced when Visual Mean Depth is considered, therefore it is only shown here for completeness.

	Connectivity Mid-Range	Connectivity Median	Connectivity Mean	VMD Mid-Range	VMD Median	VMD Mean	Connectivity-VMD Mid-Range	Connectivity-VMD Median	Connectivity-VMD Mean
p-value	0.197	0.017	0.351	0.909	0.425	0.055	0.043	0.365	0.000
t or R²	-1.295	-2.412	-0.935	0.114	-0.800	-1.927	0.007	0.002	0.013
Ordered types	L, H	L, H	L, H	H, L	L, H	L, H	HH, LL, LH, HL	LL, LH, HL, HH	LL, LH, HL, HH
Means	0.64, 0.73	0.58, 0.71	0.58, 0.64	0.62, 0.63	0.61, 0.66	0.66, 0.79	0.55, 0.64, 0.73, 0.85	0.60, 0.65, 0.68, 0.68	0.53, 0.66, 0.67, 0.80

Table 3. T-tests and ANOVAs for overall interaction (L for Low, H for High, Means in number of people per m^2 per hour)

In the case-by-case table below (table 4) it is evident that while we can detect significant generic patterns in the overall data, when we examine each site, the results are not as clear. While more people were counted interacting, in contrast to movement they don't appear to prefer specific spaces. There also doesn't appear to be a relationship between the size of the study and whether it shows significant results.

	Number of people	Connectivity Mid-Range	Connectivity Median	Connectivity Mean	VMD Mid-Range	VMD Median	VMD Mean	Connectivity-VMD Mid-Range	Connectivity-VMD Median	Connectivity-VMD Mean	Significant tests
All	**199062**	**0.197**	**0.017**	**0.351**	**0.909**	**0.425**	**0.055**	**0.043**	**0.365**	**0.000**	**3/9**
67	**422**								**0.907**	**0.720**	**0/2**
17	785	0.741	0.548	0.486	0.817	0.141	0.417	0.995	0.576	0.671	0/9
28	877							0.771	0.367	0.973	0/3
54	912	0.347	0.245	0.286	0.010	0.311	0.540	0.136	0.351	0.598	1/9
46	1089	0.181	0.488	0.042	0.213	0.726	0.950	0.818	0.773	0.434	1/9
25	1119	0.172	0.547	0.340	0.433	0.052	0.591	0.760	0.473	0.199	0/9
32	1121		0.237	0.322	0.873	0.650		0.526	0.164	0.208	0/7
65	1327		0.591	0.193	0.276		0.567	0.907	0.951	0.987	0/7
85	1771		0.392	0.378		0.436	0.965	0.963	0.315	0.297	0/7
44	1815	0.250	0.634	0.590	0.258		0.603	0.377	0.255	0.924	0/8
47	1819	0.365	0.772	0.712	0.559	0.178	0.714	0.486	0.965	0.896	0/9
66	1846	0.784	0.431	0.748	0.252	0.016	0.690	0.470	0.703	0.582	1/9
57	2181	0.000	0.135	0.000	0.232	0.000	0.000	0.469	0.009	0.236	5/9
63	2411		0.392	0.291	0.534	0.574	0.525	0.388	0.724	0.340	0/8
23	2450	0.431	0.369	0.147	0.370	0.020	0.009	0.768	0.035	0.002	4/9
73	2564		0.184	0.961	0.961	0.005	0.580	0.011	0.003	0.001	4/8
24	2848	0.527	0.398	0.340		0.313		0.065	0.093	0.162	0/7
11	2934	0.504	0.028	0.002	0.406	0.007	0.007	0.907	0.016	0.023	6/9
27	2965	0.617	0.911	0.773	0.629	0.480	0.265	0.831	0.508	0.361	0/9
29	3244	0.450	0.105	0.241	0.567	0.464	0.887	0.202	0.161	0.629	0/9
37	3566	0.015	0.456	0.083	0.643	0.285	0.242	0.415	0.142	0.149	1/9
80	3692	0.658	0.385	0.676	0.013	0.813	0.003	0.020	0.375	0.018	4/9
18	4255	0.115	0.242	0.498	0.729	0.403	0.809	0.840	0.470	0.909	0/9
45	4928	0.317	0.350	0.811	0.012	0.641	0.367	0.071	0.394	0.662	1/9
71	5181	0.372	0.745	0.687	0.111	0.514	0.571	0.269	0.017	0.394	1/9
61	6905	0.582	0.533	0.989	0.506	0.870	0.961	0.732	0.161	0.275	0/9
50	7151	0.930	0.930	0.888	0.088	0.194	0.871	0.821	0.767	0.984	0/9
49	7523	0.821	0.124	0.796	0.867	0.916	0.654	0.972	0.340	0.531	0/9
36	9367	0.925	0.132	0.343	0.558	0.044	0.145	0.092	0.007	0.141	2/9
51	10067		0.890			0.033	0.245	0.918	0.723	0.228	1/6
75	18090	0.002	0.247	0.216	0.806	0.053	0.555	0.051	0.892	0.313	1/9
58	19726	0.679	0.448	0.815	0.111	0.260	0.002	0.870	0.669	0.230	1/9
68	21544	0.008	0.355	0.889	0.847	0.626	0.977	0.136	0.214	0.474	1/9
84	40567	0.169	0.435	0.295	0.545	0.525	0.935	0.159	0.002	0.000	2/9
Significant		**4/26**	**1/32**	**3/31**	**3/29**	**7/30**	**5/30**	**2/33**	**7/34**	**5/34**	**37/279**

Table 4. T-tests and ANOVAs for interaction per site (p-values); empty cells are cases where the amount of groups were insufficient to run the test

5. DISCUSSION

The statistical tests show that the suggested representation is useful in the search for patterns of human behaviour in spatial configurations. The larger unit of analysis (area) avoids the problem of observed behaviour happening at very specific locations as seen in our previous research [16] by aggregating those observations and thus reducing the noise in the sample. We also manage to capture sufficient detail to differentiate between various spatial configurations, i.e. large and small spaces, central and non-central ones, as well as spaces that have combinations of those properties.

This method specifically allowed us to identify how predictable the two examined behaviours are. Movement proved to be the more predictable of the two, even in disproportionate splits of the sample (i.e. with the mid-range). Interaction on the other hand seems much more scattered perhaps as a result of the influence of other parameters. The placement of various functions (kitchens, workspaces) attracts movement but not necessarily interaction which may happen anywhere. Interaction is also influenced by the social network of the office, thus making spatial factors less relevant.

The results about movement contradict previous findings in a smaller sample (seven offices) by Hillier and Grajewski [8], that movement increases in more segregated areas. Instead, we found that movement happened in central spaces, especially those that were highly visible. This contradiction, as well as the overall findings in both the studied behaviours highlight the value of sampling from a large dataset. It provides a critical mass of data that allows us to identify generic patterns that are otherwise unattainable with small studies. It also allows smoothing out the various idiosyncrasies of specific studies, such as the locations of the various areas and distortions in the data. But most of all, it provides a basis on which we can identify which of these idiosyncrasies are actually pervasive in the sample and can be modelled and studied later on.

The large differences of means we find in the movement point to interesting possibilities for a designer. While not true for each individual case, it can be expected that areas with less visibility (low Connectivity) and those that are deeper in a building (high Visual Mean Depth) will attract fewer people moving. The combined metric provides some more nuance, in that we can expect most of the movement in a newly designed building to be found in highly visible, and central spaces. These results can be used as evidence to strategically inform future design decisions for example allowing a designer to specifically configure a space so as to attract more movement and interaction.

6. CONCLUSION AND FUTURE WORK

In this study we examined the existing spatial representations of indoor space as vehicles to understand the patterns of human behaviour. We found that the existing representations can either not accurately represent indoor spaces, are ambiguous in their construction or do not provide a general enough unit of analysis. We proposed a new representation based on Visibility Graph Analysis that addresses this problem by creating larger units of analysis and allowing us thus to meaningfully compare spatial configuration and human behaviour. The unit proposed is an aggregation of the cells of the VGA lattice grid into similar and continuous areas that may then be used to aggregate the discrete observations of human presence.

We initially tested this representation on simple-shaped spaces to understand its output, but eventually evaluated it against a sample from a large dataset of office spaces that we used in previous studies. The results allowed us to identify generic patterns, true for the general set of office spaces, for two behaviours that of people moving, and people interacting. Movement was found to be consistently concentrated in places that are more visible or those that are more central to the building. Interaction on the other hand was only found to be predictable when visibility and centrality were examined together, with the most interactive areas being the ones with high visibility but which are not very central.

While our evaluation focused on office spaces the results of the analysis may be carried out to other types of indoor spaces. Given that office spaces are moderately programmed, we would expect this to perform worse for more strongly programmed spaces such as hospitals. The best fit might be loosely programmed spaces where the human behaviour is governed mainly by the configuration, such as museums and small-scale urban spaces such as public squares.

In the future we plan to more closely examine the cases one by one and identify the confounding factors and idiosyncrasies that make some of our results insignificant. We will incorporate other spatial information such as the locations of functions and desks as well as non-spatial data such as the office social network and other organisational parameters. We also plan to test the multitude of different metrics that have been created within the field of Space Syntax through the years, including, sets of topological, isovist, metric and angular metrics that capture local and global properties of the space. Finally, in search for more elegant solutions we will compare the above mentioned to ones that are agnostic to existing metrics but select continuous areas in random or uniform ways.

References

[1] Allen, Thomas John and Fustfeld, Alan R. "Research Laboratory Architecture and the Structuring of Communications". In: *R&D Management* 5.2 (1975), pp. 153–164.

[2] Backhouse, Alan and Drew, Peter. "The Design Implications of Social Interaction in a Workplace Setting". In: *Environment and Planning B: Planning and Design* 19 (1992), pp. 573–573.

[3] Benedikt, Michael L. "To Take Hold of Space: Isovists and Isovist Fields". In: 6.1 (1979), pp. 47–65. ISSN: 0265-8135. DOI: 10.1068/b060047.

[4] Cattuto, Ciro et al. "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks". In: *PLoS ONE* 5.7 (July 15, 2010). Ed. by Neylon, Cameron, e11596. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0011596.

[5] EBD. *The Knowledge Problem As Architects and Designers, What Do We Know about People?* Aug. 22, 2015.

[6] Greensted, Andrew. *Blob Detection - The Lab Book Pages*. Nov. 18, 2009.

[7] Hatch, Mary Jo. "Physical Barriers, Task Characteristics, and Interaction Activity in Research and Development Firms". In: 32.3 (Sept. 1987), p. 387. ISSN: 0001-8392. DOI: 10.2307/2392911.

[8] Hillier, Bill and Grajewski, Tadeusz. "The Application of Space Syntax to Work Environments inside Buildings: Second Phase: Towards a Predictive Model". In: *Unit for Architectural Studies*. London: The Bartlett School of Architecture and Planning, University College London, 1990.

[9] Hillier, Bill and Hanson, Julienne. *The Social Logic of Space*. Cambridge University Press (CUP), 1984. ISBN: 978-0-511-59723-7.

[10] Kabo, Felichism et al. "Shared Paths to the Lab A Sociospatial Network Analysis of Collaboration". In: *Environment and Behavior* 47.1 (2015), pp. 57–84.

[11] Kabo, Felichism et al. "Zone Overlap and Collaboration in Academic Biomedicine: A Functional Proximity Approach to Socio-Spatial Network Analysis". In: *Ross School of Business Paper* 1184 (2013).

[12] Koutsolampros, Petros et al. "Spatial Databases: Generating New Insights on Office Design and Human Behaviours in the Workplace". In: *Proceedings of the 10th International Space Syntax Symposium*. 10th International Space Syntax Symposium. London, 2015.

[13] Outram, Christine. *Part Two: What Starbucks Gets That Architects Don't. Or What Architects and Others Think of the Architecture Profession*. July 5, 2015.

[14] Peponis, J et al. "On the Description of Shape and Spatial Configuration inside Buildings: Convex Partitions and Their Local Properties". In: 24.5 (1997), pp. 761–781. ISSN: 0265-8135. DOI: 10.1068/b240761.

[15] Sailer, Kerstin and Penn, Alan. "Spatiality and Transpatiality in Workplace Environments". In: Royal Institute of Technology (KTH), 2009.

[16] Sailer, Kerstin et al. "Measuring Interaction in Workplaces". In: *Architecture and Interaction*. Ed. by Dalton, Nicholas Sheep et al. 2016.

[17] Turner, Alasdair. "Depthmap: A Program to Perform Visibility Graph Analysis". In: *Proceedings of the 3rd International Symposium on Space Syntax*. Vol. 31. 2001, pp. 31–12.

[18] Turner, Alasdair et al. “From Isovists to Visibility Graphs: A Methodology for the Analysis of Architectural Space”. In: *Environment and Planning B: Planning and Design* 28.1 (Feb. 2001), pp. 103–121. ISSN: 0265-8135, 1472-3417. DOI: 10.1068/b2684.

[19] Varoudis, Tasos. *depthmapX–Multi-Platform Spatial Network Analyses Software*. 2012.

[20] Vaughan, Laura. *Space Syntax Observation Manual (2001 Unpublished Revised Edition)*. 2001.