Nexus Network Journal (2021) 23:187–208
https://doi.org/10.1007/s00004-020-00530-z

NEXUS NETWORK JOURNAL
Architecture and Mathematics

RESEARCH

Comparing and Clustering Residential Layouts Using a Novel Measure of Grating Difference

Ran Xiao[1]

Accepted: 29 September 2020 / Published online: 15 October 2020

Abstract

Clustering is widely used as a knowledge discovery method in scientific studies but is not often used in architectural research. This paper applies clustering to a dataset of 129 residential layouts, which were collected from contemporary architectural practices, to reveal underlying design patterns. To achieve this, this paper introduces a novel measure for the topological properties of layouts: 'grating difference measure'. It was benchmarked against an alternative that measures geometrical properties and the advantages are explained. The grating difference measure indicates the extent of design differences, which is used in the clustering method to obtain the distance between datapoints. The results from clustering were grouped into design schematics and qualitatively assessed, showing a convincing separation of characteristics. The method demonstrated in this paper may be used to reveal topological patterns in datasets of existing designs for both academic and practical purposes.

Keywords Grating representation · Topology · Mappings · Design analysis · Residential architecture · Algorithms

Introduction

Clustering is widely used as a knowledge discovery method, to organise objects in a dataset into groups, thereby revealing underlying patterns. In architectural design, this term was first referred to by Carter and Whitehead (1975) in the clustering of rooms according to their functional connections. Their research, however, was primarily concerned with the spatial clustering of architectural programmes. Recently, clustering has been used in the categorisation of architectural designs generated by algorithms. Rodrigues et al. (2017) and Yousif and Yan (2019) addressed the problem of presenting large quantities of computer-generated layout

✉ Ran Xiao
rx206@cam.ac.uk

[1] Department of Architecture, University of Cambridge, Cambridge CB2 1PX, UK

options to human users. They used geometry-based methods to measure similarities between layouts, then applied a clustering method to categorise the generated design options into groups, which made them easier to assess and apply. Deep learning techniques have also been used in architectural studies to cluster datapoints. For instance, algorithms to recognise architectural styles or authorship were trained with datasets that consisted of prelabelled photographs of buildings, demonstrating varying degrees of success (Obeso et al. 2017; Yoshimura et al. 2019). In Yoshimura et al.'s research, an additional clustering step measured similarity between each architect's work and identified clusters of architects whose works are stylistically similar. However, these methods are limited by the visual information that can be captured by a camera.

In contrast, this paper uses a clustering method to uncover topological patterns in layout designs. It gathers data from plan drawings collected from contemporary architectural practice and relies on a topological representation, grating, which is intrinsically linked to spatial organisation. Categorising designs in this way may be useful in both academic and practical settings. For example, researchers may be interested in the topological similarities and differences of designs that belong to a certain typology; using clustering to group similar designs may point to certain shared causes. In a professional setting, the idea of extracting patterns from designs produced in practice and reapplying them is similar to the idea advocated in Alexander et al.'s *A Pattern Language* (Alexander et al. 1977). An architect may want to compile a best practice design guide using existing designs, a format found in both professional publications (Levitt and McCafferty 2019) and architects' offices. Clustering layouts systematically allows a designer to edit out repetitions of similar layouts, and to order and index such guides in a maximally coherent manner.

To achieve this aim, a novel grating difference measure has been developed. It allows for comparisons to be made between layouts via their gratings. This measure is benchmarked against a theoretical model based on a geometrical representation and is subsequently used in clustering 129 one-bedroom flat layouts. The outcome of this clustering is interpreted from an expert's view, thereby demonstrating the extraction of patterns from existing designs for future applications.

A Brief History of Grating-Based Methods

Grating is defined formally as a plane set of orthogonal grids bounded within a rectangular frame (Newman 1964). Such grids were applied in architectural research to describe rectilinear plans, the earliest examples of which can be found in research by Eastman (1970) and March (1972). Grating is a dimensionless layout, describing the topological relationships of its components without specifying their dimensions. During this period of research on grating, various schemes of representation were developed, including the original 'grating' (Eastman 1970), 'polyominoes' (Frew 1974) and 'wall representation' (Flemming 1978). Although later schemes adopted by different researchers may vary, it is possible to identify these grating-based approaches by the central idea of separating topological relationships from dimensional information.

From this early research on grating grew an interest in systematic studies of all possible grating layouts (Steadman 1983). 'Rectangular arrangements' are a subset of grating layouts, restricted to those that are composed of rectangles only. Researchers devised algorithms that can exhaustively generate all possible rectangular arrangements given a fixed number of rectangles (Earl 1977; Flemming 1978; Bloch and Krishnamurti 1978). Many computer programmes followed, with the common goal of generating all possible rectangular arrangements, as each of the generated arrangements can then be dimensionally optimised according adjacency and size criteria. However, these efforts were limited by the combinatorial explosion in possible arrangements (Steadman 1983).

Concurrently, graph theory was applied to study small architectural plans (Levin 1964) and graphs were introduced to represent spatial connectivity in plans. The link from plan graphs to rectangular arrangements was established by Grason with a view to automating design (Grason 1970). In Grason's dual graph representation, a connectivity graph is embedded in a grating plan formed of predefined rectangular modules; adjacency requirements were first translated into a graph and subsequently used to exhaustively generate rectangular arrangements that meet those requirements.

Research into grating-based methods had a strong emphasis on design generation. Of these approaches, some created tree-like structures in which every rectangular arrangement is obtained by manipulating a simpler one (Earl 1977; Flemming 1978). However, in such approaches each arrangement is treated as a discrete category and the inherent relationships between categories were not explored. In contrast, graph-based methods were applied to understanding spatial connectivity patterns in archaeological studies, and to reveal sociological and anthropological structures. Comparison methods for plans quickly developed using graphs (Ostwald 2011). This was most notably so in Hillier and Hanson (1984), which heralded the well-known space-syntax approach. The divergence of research interests between grating-based and graph-based approaches is curious and points to a lost opportunity in grating-based approaches. The remainder of this paper explores this opportunity.

Comparing Layouts Using a Grating-Based Representation

Polyomino-Based Gratings

Polyominoes, introduced by Frew (1974), are formed by square cells joined together along their edges. The simplest polyomino is a square. Rectilinear plans may be represented using polyominoes as a grating-based layout (Steadman 1983). This method captures the dimensionless polyomino types (e.g. rectangular, L-shape, T-shape, etc.) of internal spaces as well as adjacency relationships. This paper adopts polyominoes from amongst other alternatives as the representation scheme. This is because it lends itself to the sampling method used in this paper, which will be explained in detail in a later section. Hereafter in this paper 'gratings' refers to 'polyomino-based gratings'.

A grating that represents the topology of a layout can be easily derived from an existing design. For example, Fig. 1 shows a flat plan and the grating derived from it. It also shows the polyominoes that represent each room in isolation. These room polyominoes are grouped by type, in this case one rectangular and one L-shape. Although the number of cells in the room polyominoes do not hold any meaning in isolation, as they are dimensionless, the types do convey information about the topological shape of the rooms.

Gratings are always presented in such a way that the rows and columns of the cells are parallel to cartesian axes. It is worth noting that they do not contain information about the orientations in the 'construction world' (Mitchell 1990). Isometries obtained from rotating a grating are treated as identical to the original, as rotation does not affect the grating's topological structure. To achieve a higher degree of generality, this paper also treats isometries obtained from reflecting a grating as identical to the original. It is possible, however, to choose to differentiate isometries when orientation and handedness are important goals. This can be done by adding referential instruments to relate the 'design world' to the 'construction world' (Mitchell 1990).

For computational purposes, dummy cells are used to 'patch' a non-rectangular grating into the shape of a rectangle, thereby ensuring that all columns have the same number of cells and all rows have the same number of cells. Figure 2 shows the grating in Fig. 1 as a patched rectangle with a dummy cell at the top right corner.

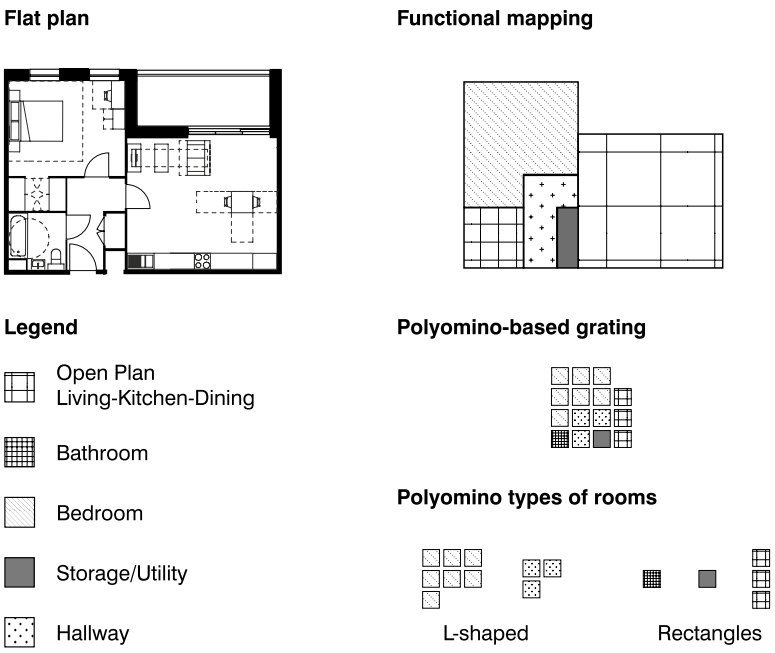

Fig. 1 Plan layout of a residential flat (top left), the functional mapping (top right), the polyomino-based grating (middle right) derived from the layout and polyominos that represent the rooms (bottom right) arranged by type

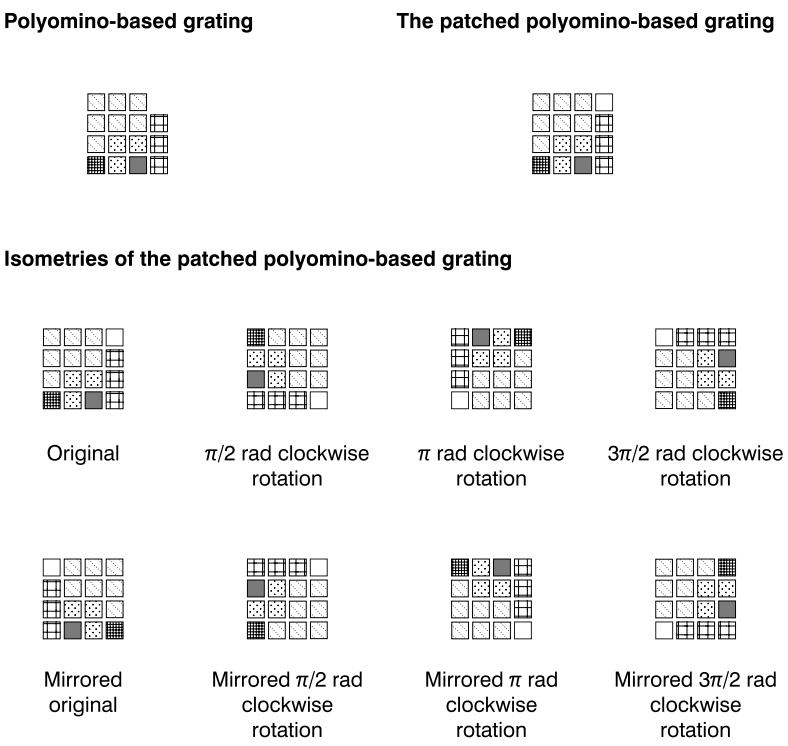

Fig. 2 A polyomino-based grating (top left), its patched rectangular version (top right), and the isometries obtained from reflection and rotations of the patched version (below)

It also shows all its isometries obtained from rotation and reflection. Using the above representation scheme, the next paragraphs will introduce a novel method of comparing gratings.

Difference Measure for Polyomino-Based Gratings

Grating research in the past focused primarily on generating new designs and much of the effort was dedicated to ensuring that the generative algorithms did not produce duplicates. This focus on new designs perhaps inhibited the realisation of the potential of grating for analysing existing layouts. Intuitively, one may consider two layouts as similar when their room adjacencies are similar. Such topological insight can be found in their gratings as well. To systematically analyse the topological similarities, or to the same effect differences, a grating difference measure is introduced. This measure lays the foundation for comparing layouts via their gratings.

A naïve method of measuring the difference between two gratings A and B might compare each cell of each row one at a time. If the cell in A is of a different function to the corresponding cell in B, it is added to a count of different cells:

$c_{diff}(A,B)$. Patching gratings with empty cells to form rectangles ensures all rows are equal in length.

When all cells have been compared, $c_{diff}(A,B)$ is divided by the total number of cells $c_{total}(A,B)$ to obtain a difference measure (represented as d_N):

$$d_N(A,B) = \frac{c_{diff}(A,B)}{c_{total}(A,B)}.$$

Of course, this naïve method does not consider the isometries of gratings. To compare A and B fully, one must consider each of these isometries. As shown in Fig. 2, each grating has four rotations—0, $\pi/2$, π and $3\pi/2 rad$ clockwise—and a mirror inversion set—which also has four rotations. One only needs to compare A with all isometries of B using the naïve method to cover all possible outcomes of d_N. If rotations of B are named B_r, B_{rr}, B_{rrr} at $\pi/2$, π and $3\pi/2 rad$ clockwise respectively, and the mirror inversion of B named B', then the difference measure $d(A,B)$ is defined as the minimum value of all d_N:

$$d(A,B) = \min\left\{\begin{array}{l} d_N(A,B), d_N\left(A,B_r\right), d_N\left(A,B_{rr}\right), d_N\left(A,B_{rrr}\right), \\ d_N\left(A,B'\right), d_N\left(A,B'_r\right), d_N\left(A,B'_{rr}\right), d_N\left(A,B'_{rrr}\right) \end{array}\right\}.$$

A worked example is shown in Fig. 3.

However, the naïve method would not be applicable to A and B if they had different numbers of rows or columns because the cells in A would not neatly correspond to those in B. As gratings are used to encode existing designs, such a situation is likely to happen when comparing layouts via grating. Transformations are thus performed on the gratings so that they have the same number of rows and columns before being compared. The exact transformation method is explained in the following example.

Grating C in Fig. 4 has two rows and five columns. To shorten the expression, one may say that C has the shape 2×5. Meanwhile, D has the shape 3×4. Taking the largest numbers of rows and columns from both, one obtains the smallest common shape of these two gratings: 3×5. A transformation can then be executed which will duplicate rows and/or columns of each layout to achieve this smallest common shape. Consequently, C needs to expand by one row and D by one column. This results in gratings which are expansions of the originals.

For C, there are two possible ways to reach shape 3×5: the first is to duplicate row 1 and the second is to duplicate row 2. The expanded gratings obtained are named $C_{D\varepsilon x}$, $x \in \{1,2\}$. Similarly, for D, one can obtain expanded gratings $D_{C\varepsilon y}, y \in \{1,2,3,4\}$ by duplicating each of the four columns. However, an arbitrary result would be produced if one chooses a specific pair of $C_{D\varepsilon x}, D_{C\varepsilon y}$ to be compared. Instead, all $C_{D\varepsilon x}$ and $D_{C\varepsilon y}$ have the same shape, so one may compare all possible combinations of $C_{D\varepsilon x}$ and $D_{C\varepsilon y}$ directly using the naïve method. Then, from the set of $\left\{d_N\left(C_{D\varepsilon x}, D_{C\varepsilon y}\right)\right\}$ containing results of all comparisons, the minimum value is taken as the difference measure $d_\varepsilon(C,D)$:

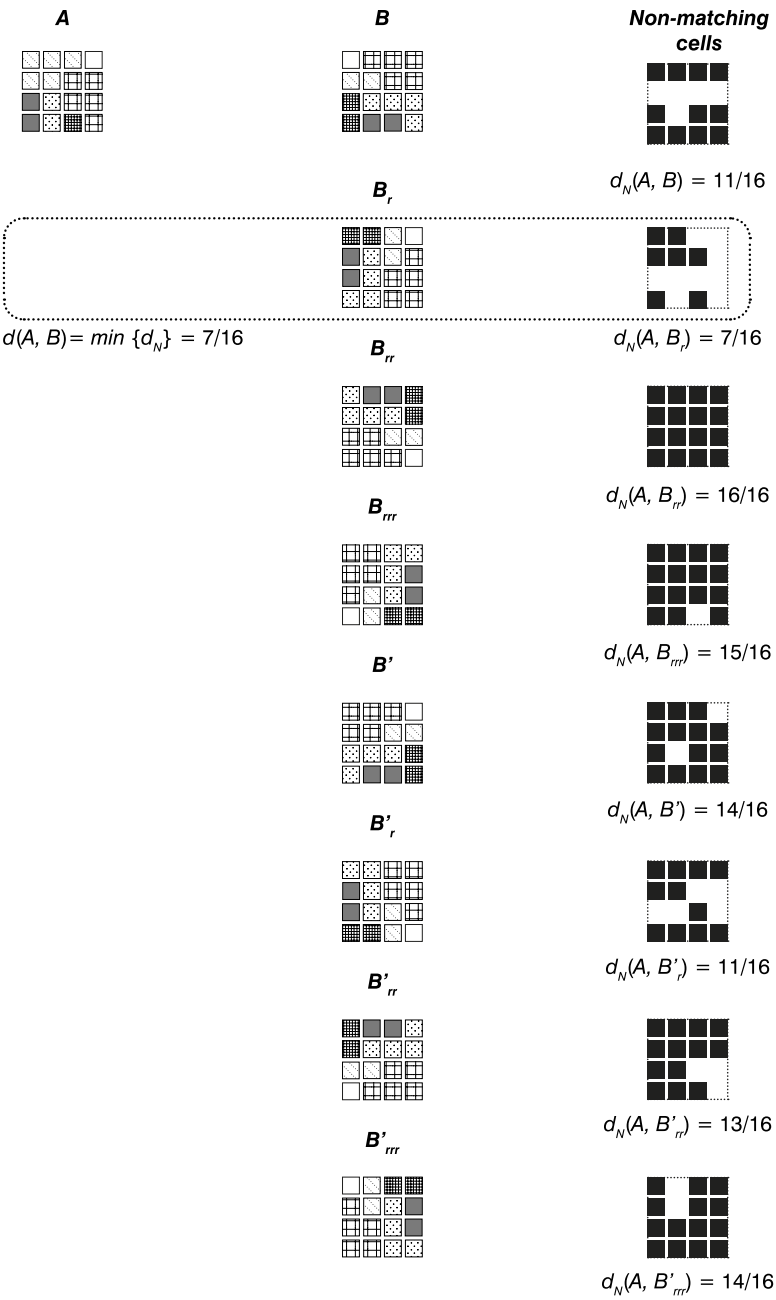

Fig. 3 A naïve comparison of grating *A* (left) and grating *B* includes all grating *B*'s isometries (middle) with each comparison result showing non-matching cells (right); the second row (highlighted) is the minimum value amongst all results

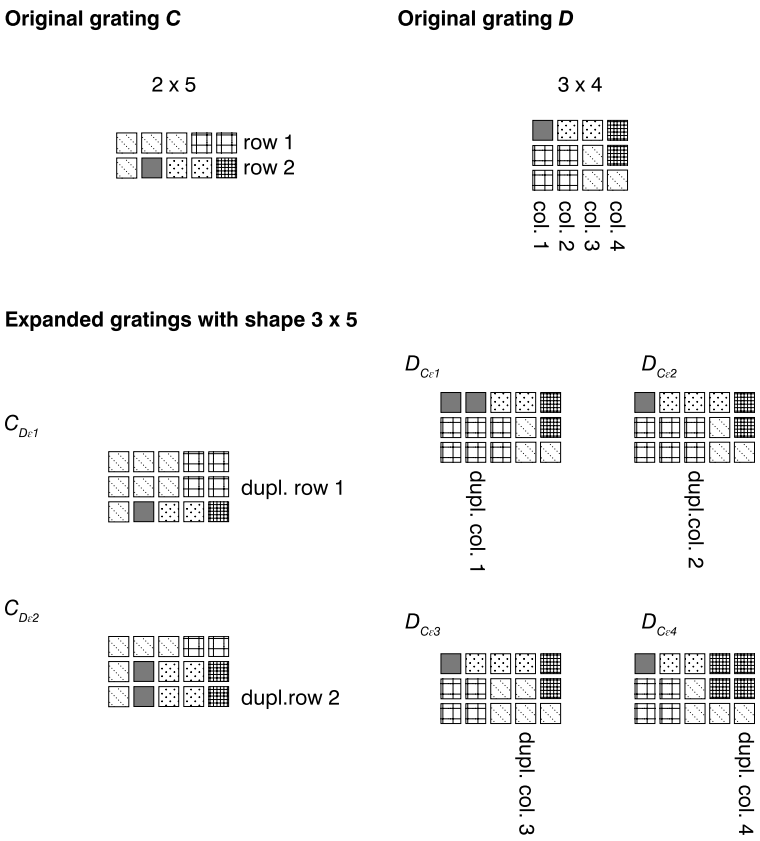

Fig. 4 Gratings C (top left) and D (top right), and their expanded gratings (below)

$$d_\varepsilon(C,D) = \min\left\{d_N\left(C_{D\varepsilon x}, D_{C\varepsilon y}\right), x \in \{1,2\}, y \in \{1,2,3,4\}\right\}.$$

The comparison of C and D is not yet complete as a full comparison will need to include all isometries; Fig. 4 only shows the gratings that are compared in order to obtain value $d_\varepsilon(C,D)$. In order to calculate $d(C,D)$, one must compare all expanded layouts obtained from the grating pairs formed by C and each of the isometries of D. This can be written thusly:

$$d(C,D) = \min\left\{\begin{array}{l} d_\varepsilon(C,D), d_\varepsilon\left(C,D_r\right), d_\varepsilon\left(C,D_{rr}\right), d_\varepsilon\left(C,D_{rrr}\right), \\ d_\varepsilon\left(C,D'\right), d_\varepsilon\left(C,D'_r\right), d_\varepsilon\left(C,D'_{rr}\right), d_\varepsilon\left(C,D'_{rrr}\right) \end{array}\right\}.$$

The complete comparison details are omitted here for brevity. The final calculated result using the method described above is $d(C,D) = \frac{2}{15} \approx 13.3\%$.

It should be noted that for C and D, duplicating a single row or column is enough to reach the smallest common shape. However, when multiple duplicated

rows or columns are needed, the duplications can be of the same row or column. An example is shown in Fig. 5. If C had to be expanded into a shape 4×5 to be compared with a certain grating G that has the shape 4×4, then there would have to be three expanded gratings: $C_{G\epsilon x}$.

It might be concluded from the naïve comparison that the smaller the difference measure, the smaller the difference between two gratings. If the difference between two gratings M and N is formally defined as the percentage of cells whose functional label must be changed in order to transform M into N, then the naïve comparison, $d_N(M,N) = \frac{c_{diff}(M,N)}{c_{total}(M,N)}$ displays the case where M and N have the same shape. However, in the expanded case, one needs to examine the expanded gratings closely to explain how this measure remains consistent.

A concept of equivalent grating is offered here. An equivalent grating describes the same polyomino types and adjacency relations as its original. Take the example in Fig. 5. The expanded gratings $C_{G\epsilon x}$ have the same adjacencies between rooms as in C. Although the number of cells in the polyominoes has changed in the transformation, the types (e.g. rectangular, L-shape, etc.) have stayed the same. As grating is a dimensionless descriptor, the number of cells does not matter. Therefore, comparing any two gratings M and N can be done through comparing their equivalent gratings M_N and N_M which have the same shape. The minimum value is taken so that the difference is always the smallest transformation of M to N via the smallest shape of M_N and N_M, or vice versa. From this definition, one may conclude that for any two gratings, M and N, their difference measure is $d(M,N) \in [0,1]$. The smaller the difference measure, the more similar two gratings are.

Grating Difference Measure in Relation to Layout Difference

The previous section provided a comprehensive algorithm to calculate the difference between any two gratings. This novel grating difference measure reflects topological differences between layouts, from which the gratings are derived. This section

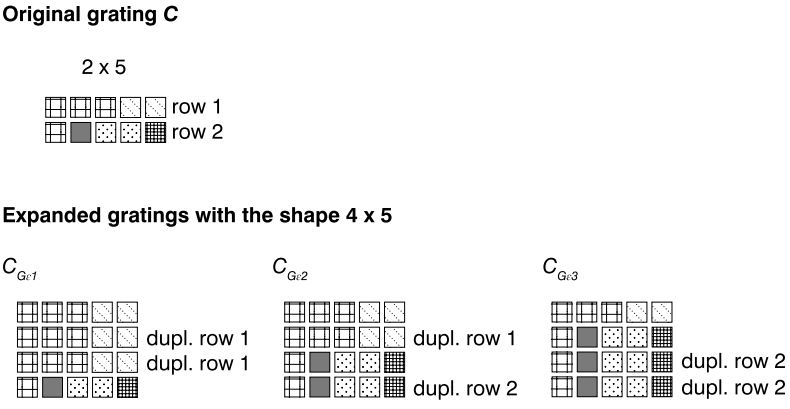

Fig. 5 Grating C (top) and its expanded gratings with two additional rows (below)

extends the use of the method to reflect layout difference in a general sense. However, a grating is a dimensionless representation, it contains no dimensional information as to the width and height of each cell. Therefore, the difference measure disregards dimensional differences that exist in the original layouts. To understand how this lost information impacts on the proposed extension of its use, this section introduces a benchmark measure and compares the results obtained from these two measures. The following worked example introduces this benchmark measure.

As shown in Fig. 6, floor plans are used to derive gratings C and D that were analysed in the previous section. The relative locations of rooms are present in the gratings, e.g. the bathroom and bedroom are adjacent and located on one side, the living space is located on the other side, etc. The grating difference measure is also reflected in the difference in the polyomino types of the rooms, e.g. C has an

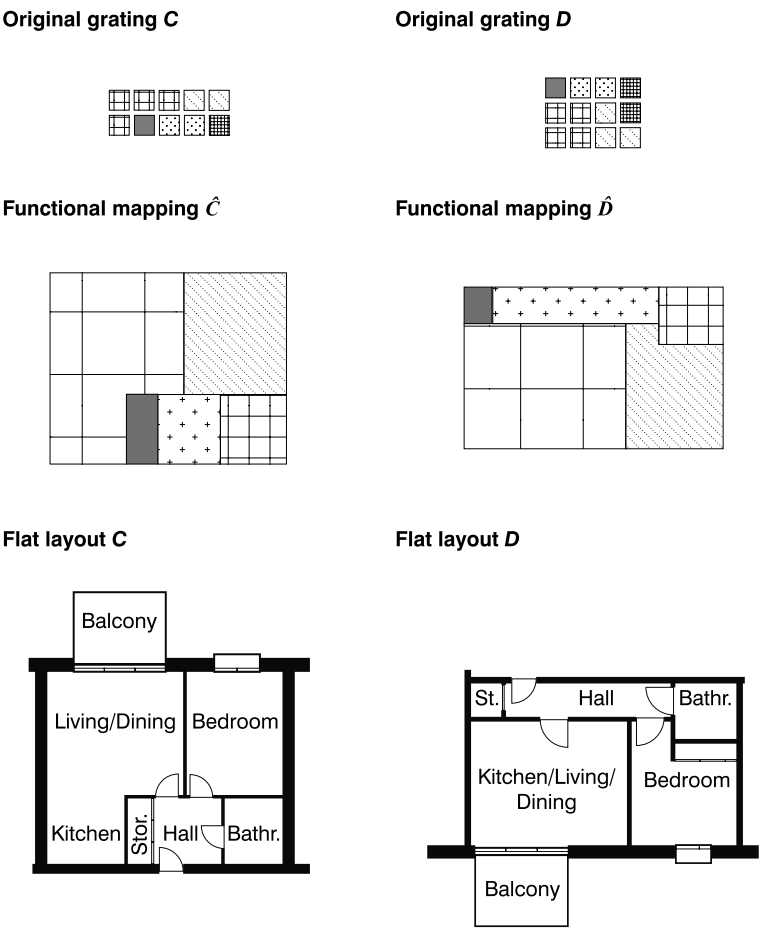

Fig. 6 Gratings C and D (top), the functional mapping $\hat{C}$ and $\hat{D}$ (middle), and the flat layouts C and D from which they derived (below)

L-shaped kitchen-living-dining space whereas D has a rectangular one. The grating difference measure obtained is 13.3%, which reflects these differences. What this measure fails to capture is the proportional relationships and the overall scaling of the layout, both of which are determined by dimensional information. As uniform scaling can be determined by simple metrics, such as the overall area, it can be easily compensated for. But proportional relations cannot be determined by a single metric and therefore require further investigation. It is important to understand the effect of disregarding this information when comparing layouts by their gratings. To do so, another difference measure of the layouts can be introduced.

In this new measure, the 'mapped layout difference' d_m, proportional relationships are retained and can be compared. It is derived directly from the functional mappings $\hat{C}$ and $\hat{D}$ in Fig. 6. Put simply, to calculate $d_m\left(\hat{C}, \hat{D}\right)$ is to measure the extent of changes one must make in order to transform $\hat{C}$ into $\hat{D}$. The first step is to scale $\hat{C}$ along horizontal and vertical axes to produce $\hat{C}_{\hat{D}}$, so that the overall rectangular bounding box $\boxed{C}$ of $\hat{C}$ will completely overlap with $\boxed{D}$ of $\hat{D}$. In this process uniform scaling will not be accounted for, as previously mentioned. However, when the original aspect ratios of bounding boxes are not the same, scaling will be non-uniform along the two axes. To account for this non-uniform transformation, an aspect difference $d_{m-aspect}\left(\boxed{C}, \boxed{D}\right)$ is defined as the angular difference between the diagonal vectors $\overrightarrow{\boxed{C}}$ and $\overrightarrow{\boxed{D}}$ from the lower left vertex to upper right vertex of the bounding boxes:

$$d_{m-aspect}\left(\boxed{C}, \boxed{D}\right) = \arccos\left(\frac{\overrightarrow{\boxed{C}} \cdot \overrightarrow{\boxed{D}}}{\left|\overrightarrow{\boxed{C}}\right| \cdot \left|\overrightarrow{\boxed{D}}\right|}\right) \Big/ \frac{\pi}{2}.$$

When $\boxed{C}$ and $\boxed{D}$ have the same aspect ratio, $d_{m-aspect}\left(\boxed{C}, \boxed{D}\right) = 0$. The more different the aspect ratios are, the larger this measure will be. Furthermore, because the vectors are always in the first quadrant, one may conclude that $d_{m-aspect}\left(\boxed{C}, \boxed{D}\right) \in [0, 1]$):

$$\arccos\left(\frac{\overrightarrow{\boxed{C}} \cdot \overrightarrow{\boxed{D}}}{\left|\overrightarrow{\boxed{C}}\right| \cdot \left|\overrightarrow{\boxed{D}}\right|}\right) \Big/ \frac{\pi}{2} = \arccos\left(\frac{\overrightarrow{\boxed{D}} \cdot \overrightarrow{\boxed{C}}}{\left|\overrightarrow{\boxed{D}}\right| \cdot \left|\overrightarrow{\boxed{C}}\right|}\right) \Big/ \frac{\pi}{2},$$

$$d_{m-aspect}\left(\boxed{C}, \boxed{D}\right) = d_{m-aspect}\left(\boxed{D}, \boxed{C}\right)$$

Once $\hat{D}$ is transformed into $\hat{D}_{\hat{C}}$, which has the same overall rectangular bounding box of $\hat{D}$, the functional difference $d_{m-func}\left(\hat{C}, \hat{D}_{\hat{C}}\right)$ is calculated as the non-overlapping functional area divided by the overall area of the bounding box. The

same procedures for comparing isometries as were used in calculating grating difference still apply here.

Because the transformation is proportional in both horizontal and vertical axes, one also obtains:

$$d_{m-func}\left(\hat{C}_{\hat{D}}, \hat{D}\right) = d_{m-func}\left(\hat{D}_{\hat{C}}, \hat{C}\right).$$

Finally, d_{m-func} and $d_{m-aspect}$ are combined into a single function that calculate the overall mapped layout difference:

$$d_m\left(\hat{C}, \hat{D}\right) = \sqrt{d^2_{m-aspect}\left(\boxed{C}, \boxed{D}\right) + d^2_{m-func}\left(\hat{C}_{\hat{D}}, \hat{D}\right)}.$$

This comparison method is summarised in Fig. 7. From the above definition, one can also conclude that:

$$d_m\left(\hat{C}, \hat{D}\right) = d_m\left(\hat{D}, \hat{C}\right).$$

It is difficult to define what 'ground truth' is when considering differences in layout designs. As the consideration may vary under different circumstances, it is a fundamentally subjective matter whose criteria must be decided by the architects involved in each case. Previous research has used crowd-sourced expert opinion to establish a standard of layout difference for computer-generated designs (Rodrigues et al. 2017). A perceptual approach such as this is limited by sampling issues.

The mapped layout difference introduced previously offers an alternative benchmark based on a theoretical model. This model, of course, has its limitations. For example, the combination of $d_{m-aspect}$ and d_{m-func} is non-trivial, as the weighting of each aspect of the difference is unknown. Despite its limitations, d_m offers a base case, inclusive of dimensional information and available for grating difference measure d to be examined against. In the next section, such an examination is conducted using a dataset containing functionally similar flat layouts.

Applying Difference Measures in a Dataset of Residential Flats

The dataset has been collected partly from an industry partner and partly from British architectural publications. It contains 129 one-bedroom flat layouts produced by 16 architects' offices. Each layout is sampled using a grid and manually labelled to obtain the functional mapping. This functional mapping grid is then used to derive gratings by eliminating repeating rows and columns. In data collection, the grid spacing was set at 600 mm, a dimension that is small enough to capture all functional zones in the layout. Figure 8 shows a summary of the data collection method. A small percentage of layouts is non-orthogonal. In these situations, the layouts were approximated to orthogonal layouts by eye.

Every possible layout pair is drawn from the database and compared using their derived functional mappings and gratings, giving difference values d_m and d

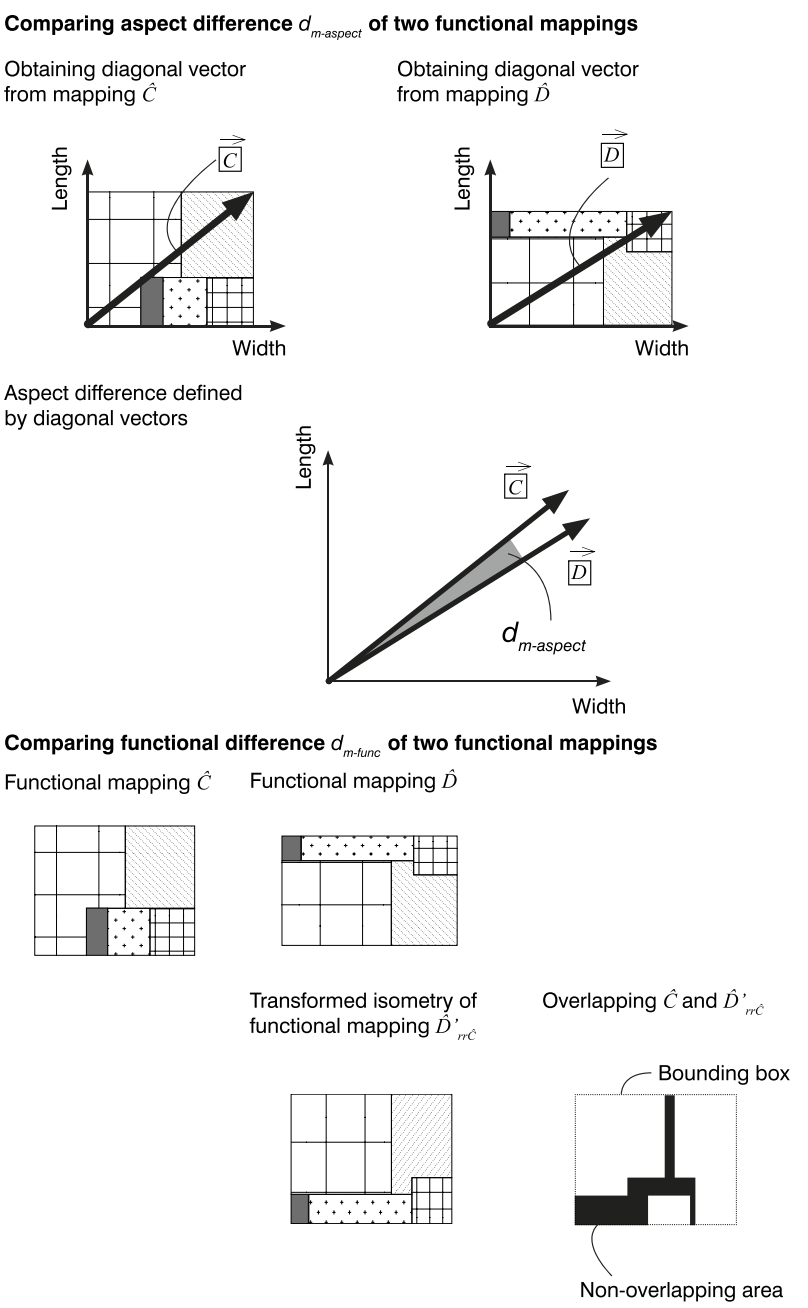

Fig. 7 Overview of the functional mapping comparison method: obtaining aspect difference (top) and obtaining functional difference (below)

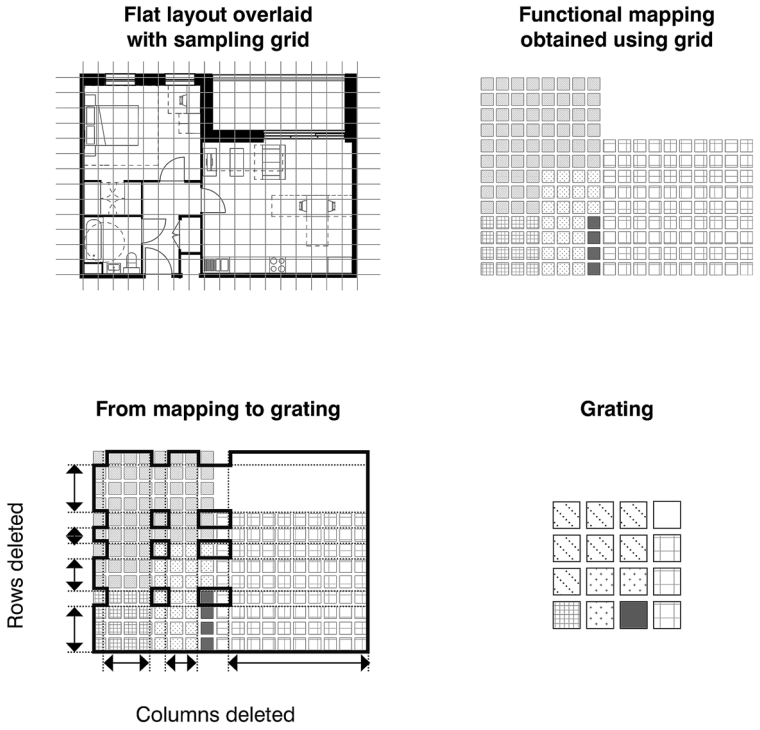

Fig. 8 Summary of data collection method: flat layout overlaid with a grid (top left), functional mapping obtained from sampling using the grid (top right), procedure of deleting repeating rows and columns from the mapping (below left), and the grating obtained from the procedure (below right)

respectively. Figure 9 illustrates all 8256 pairwise comparisons on a scatter plot, which shows a clear correlation between d_m and d. Of course, this correlation does not come as a surprise, as grating is derived from functional mapping and contains part of its information. The graph also shows a regression line, which is calculated using a linear model:

$$d_m = ad^2 + bd + c.$$

The square of the grating difference d^2 is included in this regression model because the functional mapping difference d_m, which includes the dimensional information, can increase in both horizontal and vertical axes. The strength of the correlation suggests that the grating difference measure can be used as a surrogate measure of layout difference. That is, grating difference reflects layout design difference. Because the grating difference measure can achieve a higher level of generality and drastically reduce calculation complexity by disregarding dimensional information, there is additional benefit of using it in some analytical tasks. However, the residual graph below the main graph shows a noticeable spread as d increases, pointing to heteroscedasticity in the data. This means that, although the correlation is significant

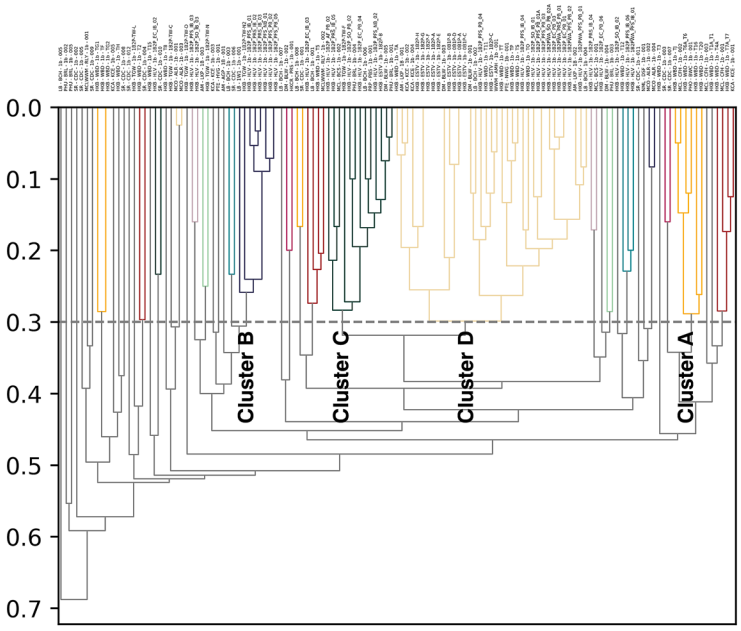

Fig. 9 A scatter plot with an OLS regression line (top) and residual value analysis derived therefrom (below); the y-axis shows mapped layout difference and residual values, respectively; both x-axes show grating difference

between d_m and d, predicting the value of d_m given d using the linear model is likely to give inconsistent errors. One can attribute this heteroscedasticity to dimensional variations that are disregarded but correlate to grating difference in some ways.

It is worth pointing out that this research does not guarantee that the same method would yield valid results in any dataset, as the data gathering process used in this paper was not designed for that purpose. Instead, it offers an examination which demonstrates that the two difference measures introduced in this paper are coherent in the typology of one-bedroom flat layouts, even though their relations cannot be reduced to a simple linear model. It also implies that similar datasets which are gathered to study a certain typology may be able to utilise the clustering method, explained in the next section, for the same purpose.

Clustering Residential Layouts Using Grating

Based on the grating difference measure introduced and benchmarked in previous sections, clustering architectural layouts can be performed. Using a grating difference measure as a distance, a clustering experiment was performed on the dataset introduced in the previous section. A hierarchical clustering function implemented in a SciPy module in Python was used (Virtanen et al. 2020). This method is called 'Unweighted Pair Group Method with Averaging', or the UPGMA algorithm (Sokal and Michener 1958). By establishing how similar two layouts are,

the algorithm joins layouts together into branches; branches are subsequently joined depending on their distances. The result is a dendrogram that links all the layouts in the dataset into a tree-like diagram. However, the dendrogram does not decide how many clusters there are in the dataset. A distance to 'cut' the dendrogram into clusters must be manually selected. Figure 10 shows the dendrogram produced from the clustering algorithm where the cutting line is set at 0.3, from which the clusters of layouts are obtained. 46 layouts are not joined with any other after the cut at 0.3, meaning they have not been assigned a cluster using the cutting line; these are linked with a grey line. The rest of the layouts are joined at least with one other layout.

Even though there are 129 flat layouts in the dataset, the combinatorial design possibilities for one-bedroom flats is vast and it is interesting that so many clusters emerge. Clear design similarities can be identified within the clusters with the

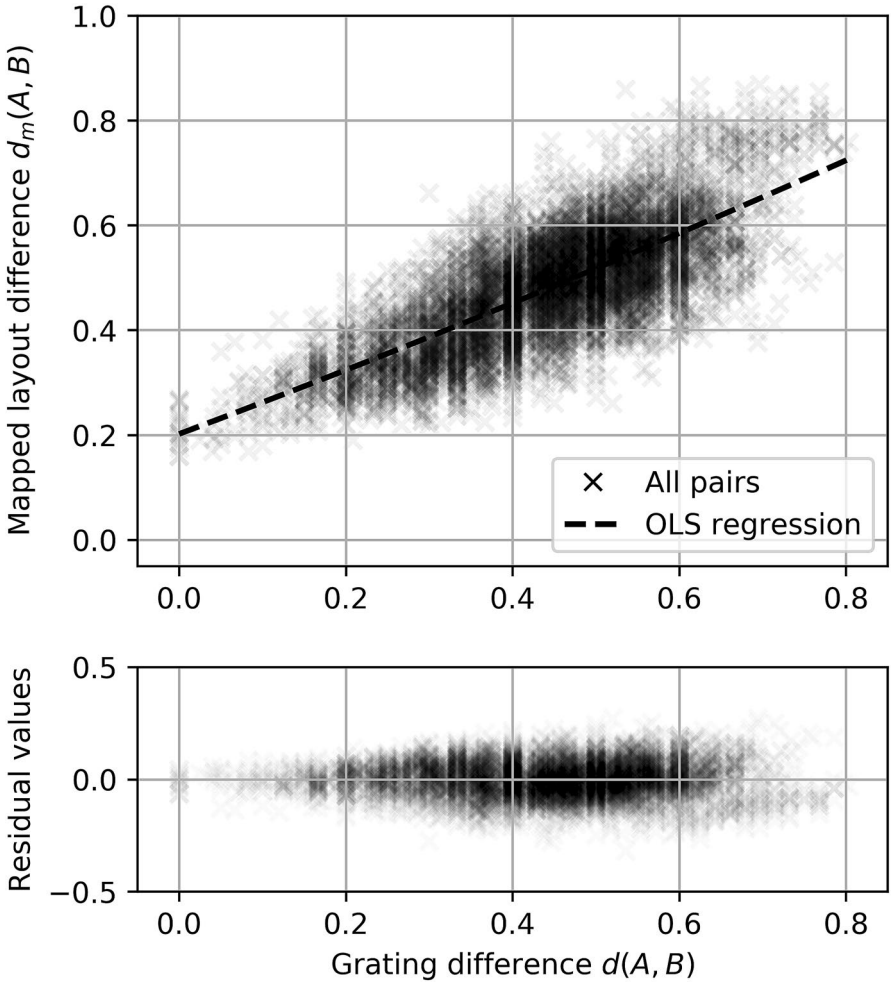

Fig. 10 Dendrogram of clusters, cutting line and four clusters selected for analysis

human eye and significant divergence is also seen between the clusters. Four main clusters—which each contain more than four layouts—are shown in Fig. 11a, b. For each cluster, the original plans are shown as thumbnails composed of functional

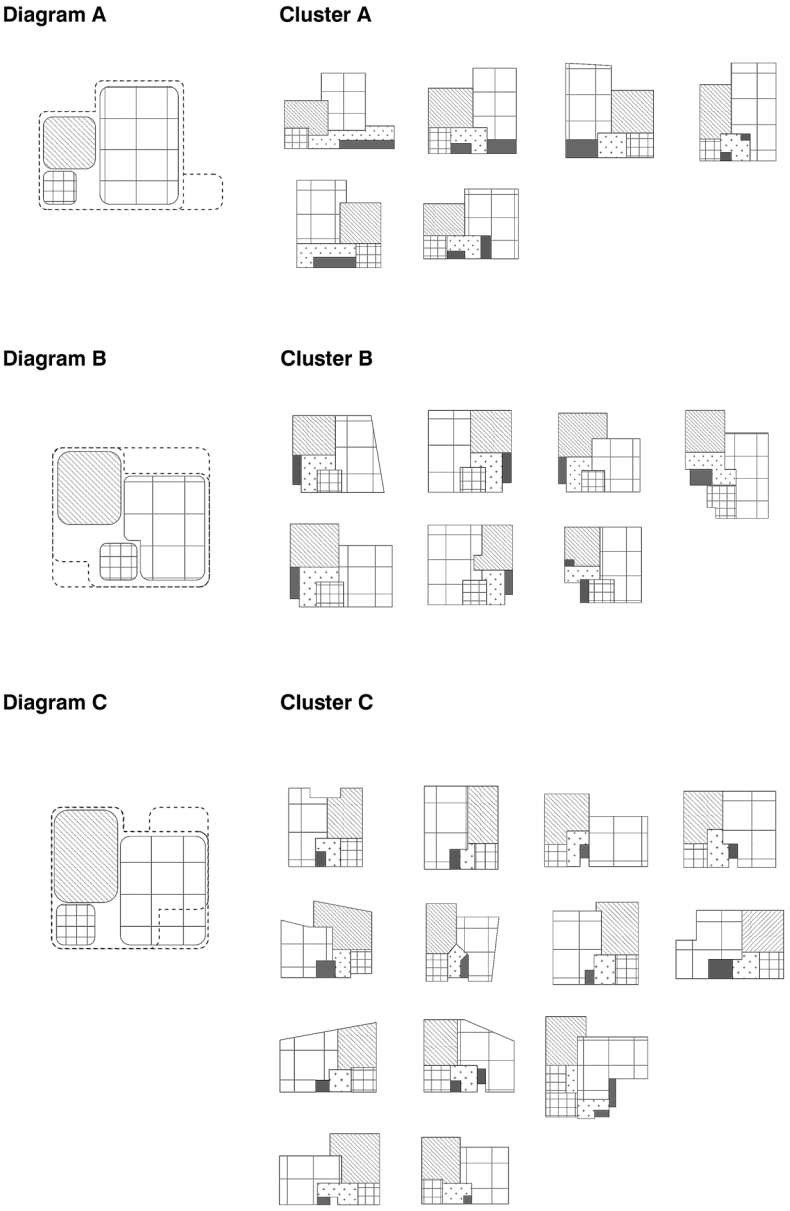

Fig. 11 **a** Layout design clusters identified using the grating difference measure: diagrams summarising the design idea of each cluster (left) and clusters of flat layouts represented by thumbnails of functional mappings (right). **b** Layout design clusters identified using the grating difference measure: diagrams summarising the design idea of each cluster (left) and clusters of flat layouts represented by thumbnails of functional mappings (right) (continued)

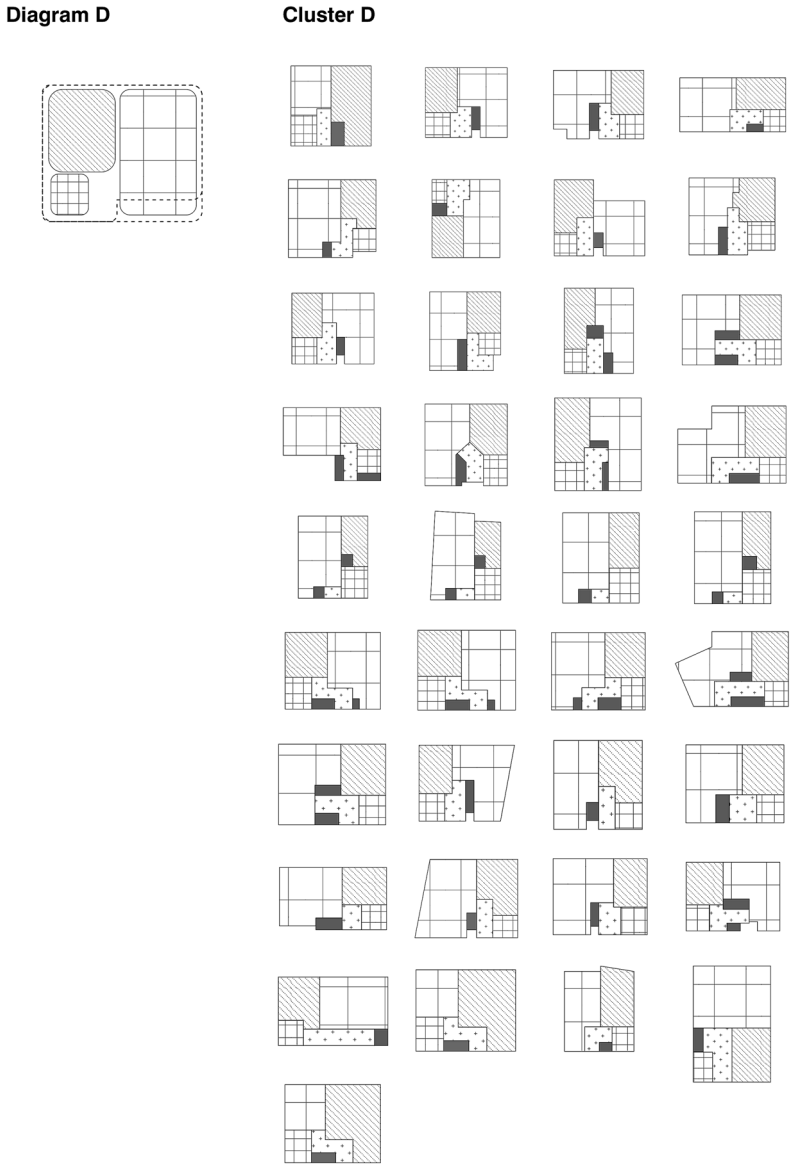

Fig. 11 (continued)

zones for the purposes of clarity. A separate diagram is offered as an expert's interpretive summary of the design ideas presented by each cluster. The diagrams identify the relative locations of three important functional spaces: the living-kitchen-dining space (LKD), bedroom and bathroom. The dotted outline shows variations in the shape of the footprints.

- Cluster A identifies a type of one-bedroom design that has a mostly L-shaped footprint. The bathroom and the living-kitchen-dining space are located on two sides of the plan. The living space protrudes from the main part of the plan, while the bedroom opens to an inset balcony.
- Cluster B identifies a type of design where the entrance is placed at the farthest corner away from both the LKD and the bedroom. It is also characterised by the bathroom sharing a partition wall with the kitchen area.
- Cluster C identifies another type of mainly L-shaped footprints but where, unlike type A, the bedroom protrudes from the main part of the plan whilst the living room opens to an inset balcony. The bathroom and the LKD are located on opposite sides of the plan.
- Cluster D identifies a design in which LKD and the bedroom share a straight external wall, and the bathroom and LKD are located on opposite sides of the plan.

The result of this clustering demonstrates a notable robustness in the grating difference measure. The clusters are not influenced by the overall footprint, as the variations in each cluster demonstrate. Only one layout in the clusters shown contradicts the expert's summary (this outlier layout in cluster D is highlighted with an asterisk and appears closer to cluster C). This error could be a side effect of the clustering algorithm. As each grating is assessed globally before assigning to a group, an expert may come to different conclusions when only the certain highlighted characteristics are assessed at a local level. In any case, the dendrogram does show a close link between clusters C and D. If the cutting line had been chosen at a lower point on the 'tree', clusters C and D would have been joined. Overall, clustering the layouts using the grating difference measure showed a convincing result that can be easily interpreted into design ideas.

Conclusion

While the dataset used in this paper is restricted to one architectural typology, the grating difference measure and clustering methods introduced are of direct relevance to architectural practice. Whereas traditionally, architects learn from instances of exemplary designs, the methods introduced here may be used to summarise generic patterns from a whole body of exemplary designs. Such patterns can be then critically assessed before being applied to future projects. This echoes Alexander et al.'s (1977) idea that patterns in architecture arises as responses to reoccurring challenges arise in our environment, and that one should not be afraid to use these patterns. Conversely, as new challenges emerge, designers should not be afraid of developing new patterns.

Beyond this practical dimension, this paper furthers the understanding of grating representation by extending its use as an analytical tool, providing a helpful level of abstraction that compliments current methods. The flat layouts analysed belong to a typology that is highly homogenous in spatial connectivity. However, in such typologies a high level of diversity of design can still exist.

The structure of diverse 'embeddings' (Steadman 1983), which share almost identical connectivity graphs, would not have been revealed if graph-based methods were used; nor would visibility analysis, introduced as a supplement to graph-based methods (Turner et al. 2001), offer much insight. This is because not all typologies are public spaces in which visibility plays an important role in wayfinding and social interaction. In contrast, grating as a topology-based representation has been proven to be highly effective.

This paper also showed that gratings are not discrete categories but are instead related to each other. This departs from previous grating-based analyses. For example, in a study of 74 small Cambridgeshire house plans, Dickens categorised plans according to their gratings and connectivity graphs, and statistically analysed the categories against historical periods (Dickens 1977). However, there has been no attempt to link the categories, even though some are conceptually more similar than others. In contrast, this paper's introduction of a grating difference measure enabled linking between designs. In future studies, it may be used to reveal structures of topological similarities, which may suggest underlying practical, intellectual, societal or historical influences.

Clustering in this paper differs from that in precedents in two aspects. Firstly, as a knowledge discovery technique, it has demonstrated effectiveness in analysing real-world designs from multiple architects; this contrasts with alternative approaches, which were only demonstrated using highly homogenous computer-generated layouts produced by a single programme. Secondly, the introduction of a grating difference measure has addressed a key issue identified by Rodrigues et al. (2017), that geometry-based measures can only assess overall footprint shape without considering internal spatial organisation as well. It is also notable that gratings' relations to each other are defined from an internal measure—the grating difference measure—rather than from an external metric, such as the Mean Depth or the Total Depth in space syntax theory. Therefore, the clustering result presented in this paper and the design patterns that emerged from it refer to a kind of relative structure, rather than a universal theoretical structure. The diversity of flat layout design is structured into generic schemes that may change depending on the dataset and the levels of generalisation. In the absence of a commonly accepted universal theory on design, building relative structures of design offers a pragmatic alternative.

Finally, apart from analytical tasks, the grating difference measure also points to future applications in machine learning. Although no direct comparison is made to current pixel-based model-free deep-learning research in architecture, grating as a coherent architectural formalism has the advantage of interpretability. The difference measure introduced in this paper serves as a complementary model to model-free approaches, which have certain theoretical limitations (Pearl 2018).

Acknowledgements This research benefited from design data acquired from Hawkins\Brown Architects LLP.

References

Alexander, Christopher, Sara Ishikawa, and Murray Silverstein. 1977.*A pattern language: towns, buildings, construction*. New York: Oxford University Press.

Bloch, CJ. and Ramesh Krishnamurti. 1978. The counting of rectangular dissections. *Environment and Planning B: Planning and Design* 5(2): 207-214.

Carter, David J and B Whitehead. 1975. The use of cluster analysis in multi-storey layout planning. *Building Science* 10(4):287-296.

Dickens, P. 1977. An analysis of historical house-plans: a study at the structural level (micro) in *Spatial Archaeology*, ed. D L Clarke. 33-45. London: Academic Press

Earl, Christopher F. 1977. A note on the generation of rectangular dissections. *Environment and Planning B: Planning and Design* 4(2):241-246.

Eastman, Charles M. 1970. Representations for space planning. *Communications of the ACM* 13(4):242-250.

Flemming, Ulrich. 1978. Wall representations of rectangular dissections and their use in automated space allocation. *Environment and Planning B: Planning and Design* 5(2):215-232.

Frew, RS. 1974. *Towards a Theory of Systems Architecture*. PhD thesis. University of Waterloo.

Grason, John. 1970. A dual linear graph representation for space-filling location problems of the floor plan type. In *Emerging methods in environmental design and planning*, ed. Gary T. Moore, 170-178. Cambridge MA: MIT Press.

Hillier, Bill and Julienne Hanson. 1984. *The social logic of space*. Cambridge: Cambridge University Press.

Levin, PH. 1964. The use of graphs to decide the optimum layout of buildings. *Architects' Journal* 140:809-182.

Levitt, David and Jo McCafferty. 2019. *The housing design handbook.* Abingdon: Routledge.

March, Lionel. 1972. *The Architecture of Form*. Cambridge: Cambridge University Press.

Mitchell, William J. 1990. *The logic of architecture: Design, Computation, and Cognition.* Cambridge MA: MIT Press.

Newman, Maxwell HA. 1964. *Elements of the topology of plane sets of points*. Cambridge: Cambridge University Press.

Obeso, Abraham Montoya, Mireya S. García Vázquez, Alejandro A. Ramírez Acosta, and Jenny Benois-Pineau. 2017. Connoisseur: classification of styles of Mexican architectural heritage with deep learning and visual attention prediction. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, 1–7*. Florence, Italy. https://doi.org/https://doi.org/10.1145/3095713.3095730

Ostwald, Michael J. 2011. The Mathematics of Spatial Configuration: Revisiting, Revising and Critiquing Justified Plan Graph Theory. *Nexus Network Journal* 13(2):445-470.

Pearl, Judea. 2018. Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint.*arXiv:1801.04016.

Rodrigues, Eugénio, David Sousa-Rodrigues, Mafalda Teixeira de Sampayo, Adélio Rodrigues Gaspar, Álvaro Gomes, and Carlos Henggeler Antunes. 2017. Clustering of architectural floor plans: A comparison of shape representations. *Automation in Construction* 80:48-65.

Sokal, Robert R and Charles D Michener. 1958. A Statistical Method for Evaluating Relationships. *University of Kansas Science Bulletin* 38:1409-1448.

Steadman, J Philip. 1983. *Architectural morphology: an introduction to the geometry of building plans*. London: Pion.

Turner, Alasdair, Maria Doxa, David O'Sullivan, and Alan Penn. 2001. From isovists to visibility graphs: A methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design* 28(1):103–121. https://doi.org/10.1068/b2684

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods 17*(3):261–272.

Yoshimura, Yuji, Bill Cai, Zhoutong Wang, and Carlo Ratti. 2019. Deep Learning Architect: Classification for Architectural Design Through the Eye of Artificial Intelligence. In *Computational Urban Planning and Management for Smart Cities*, ed. by Stan Geertman et al. 249–265. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-19424-6_14.2019

Yousif, Shermeen and Wei Yan. 2019. Shape clustering using k-medoids in architectural form finding. In *Computer-Aided Architectural Design. "Hello, Culture", CAAD Futures 2019*.ed. Ji-Hyun Lee. 459–473. Singapore: Springer Nature. https://doi.org/10.1007/978-981-13-8410-3_32

Ran Xiao is a PhD candidate at the Department of Architecture, University of Cambridge and a registered architect in the UK. He is interested in the means by which design knowledge is acquired through practical experience. This inspired his research topic, which explores the application of machine learning to acquire spatial knowledge from existing designs to support future design decisions. His research is informed by the legacy of his predecessors at Cambridge, who studied architectural morphology from the 1960s. He is currently collaborating with architectural practices in the UK on his research project, with a special interest in housing. By synthesising a morphological understanding of architecture with machine learning techniques, his research complements model-free deep-learning approaches, which are becoming increasingly popular in architectural research.