378

Ideal apartment layouts for mass production based on expected resident satisfaction

RÉKA MÁNDOKI & JOHN ORR,

UNIVERSITY OF CAMBRIDGE

ABSTRACT

As an answer to the challenges the construction industry is facing, off-site manufacturing is receiving increased attention from both the professional and the scientific community. The residential building sector may experience significant changes, as easy-to-predict functional requirements and structural loads make these buildings particularly suitable for mass production. Assuming that residential buildings are indeed going to enter mass production soon, objective requirements should be defined to ensure that the produced buildings will be socially acceptable and resilient. This paper presents a novel methodology to determine apartment layouts suitable for mass production by ensuring that the created homes would be as close as possible to what people find ideal. We argue that apartment layouts can be grouped based on their similarity by using graph editing distance (GED) and applying density-based clustering. Then, we show that from a sample of attractive apartment layouts collected from 71 Hungarian adults, those layouts that minimise the sum of GEDs within their group are the best candidates for mass production. Our results uncovered room arrangements that best fit the preferences of the sample, and more importantly, a new, data-driven design and review strategy was developed. As a demonstration tool, a sample script was created in the Revit-Dynamo environment that provides instant feedback on how close a layout is to what is considered ideal. The proposed methodology can aid both professionals and authorities in evaluating design solutions from the aspect of expected resident satisfaction.

KEYWORDS

mass production, ideal apartments, density-based clustering, justified plan graphs, resident satisfaction

1 INTRODUCTION

The right to adequate housing is a basic human right (UN General Assembly, 1984). However, today, housing crises are emerging in many European cities. In a market where demand exceeds supply, customer satisfaction that normally controls quality cannot work properly. New apartments sell easily irrespective of how good they are, leading to a natural decrease in quality. This affects most the poorest who are not able to choose freely from the market due to their financial constraints. Students, young families, and working-class people, essential to the cities' functioning, are increasingly being pushed out of the centres into suburban, often suboptimal real estate (Farha, 2017).

On the other hand, even though the problem has been known since the 1970s (Habraken, 1972), the way dwellings are created has become increasingly impersonal over time. Nowadays, most real estate developments are mass-housing projects without a direct connection between the designer and the customer that would help create better spaces. Unfortunately, decision-makers often focus on providing the quantity instead of the quality. It is a real shame, as low-quality residential projects are destroying the environment with buildings that experience rapid obsolescence and determine the look of neighbourhoods for decades.

As a general response to current problems, including housing shortages, a shortage of skilled labour, and environmental concerns, the construction industry is turning to the use of off-site manufacturing techniques that can accelerate construction time (McKinsey Global Institute, 2017). However, prefabricated and mass-produced buildings can only be part of the solution if they are socially acceptable (Mándoki and Orr, 2021). Therefore, in response to a common criticism regarding former prefab homes, the residential sector needs to provide people with comfortable and adaptable living spaces.

It is hard to objectively measure in the design phase how good an apartment is from the perspective of future resident satisfaction and determine minimum criteria beyond general health and safety requirements. Post occupancy quality assessments could help, but these are not carried out routinely (Hay et al., 2017), and it is especially rare to run these assessments over a longer period, which would provide a more accurate picture of how building users feel.

In contrast, to improve user experience in the age of information and big data, our computers, mobile phones, cars, and sometimes even vacuum cleaners collect data about what we prefer and how we use them. Our decisions are shaped by recommendation algorithms that offer options tailored to our taste in many fields of life. The automotive and computer industries also showed that by mass-manufacturing or mass-customising products, the optimisation potential of standardisation could be harnessed, leading to better and cheaper products. The construction industry is clearly in a lag with these contemporary trends.

This paper investigates how we could determine which apartment layouts were the best to be mass-produced to ensure that the resulting buildings were in line with the real needs of people. For this aim, the small graph matching method proposed by Dalton and Kirsan (2008) was developed further and tested on a sample of over 70 apartment layouts considered optimal by Hungarian citizens.

2 GRAPHS IN ARCHITECTURE

Graphs are mathematical representations of systems and consist of nodes and edges. Edges connect two nodes and can be directed or non-directed, depending on the system they represent. The mathematics describing the behaviour of graphs was born in the 18th century, when Euler, a swiss mathematician, proved about islands and bridges in Königsberg that independently of the starting point, it was impossible to reach all islands and return by going through all bridges only once (Shields, 2012). Since then, graphs have proven to represent very well various systems, including molecules in chemistry, family trees, etc. In architecture, graphs can represent layout arrangements, floor plans, processes like design and construction steps and even chains of requirements.

Graphs in the architectural literature were used first by Alexander (1964) for describing the hierarchy of design processes. Alexander et al. (1977) systematically described what is required for a good city by defining the essence of different design problems, called patterns, in a structure of an encyclopaedia using graphs to represent the system of patterns. Mándoki and Orr (2019) built on this method to evaluate the planned city of Cambourne, UK, and gain practical knowledge on what should be improved on our current practice to achieve real resilience in future neighbourhoods built of mass-produced buildings.

Steadman (1970) used graphs to present an analogy between wall length in rectangular floor plans and electrical circuits to generate minimum-standard house plans automatically. March and Steadman (1971) represented floor plans with graphs in two ways. In the plan graph, walls were represented as edges and wall joints as nodes, and in the connection graph, rooms were represented as nodes and the walls between two rooms as edges. The connection graph and the plan graph were shown to be duals. Such graphs, also called reciprocal diagrams, can also be found in graphostatics, frequently used for shape optimisation tasks (Rippmann et al., 2012).

Hillier and Hanson (1984) created the method to represent the syntactic structure of dwellings as justified plan graphs (JPG) and defined a set of metrics to analyse them. In their system, rooms were represented by nodes and doors as edges. Like many of their published case studies present, by using JPGs, it is possible to calculate how integrated specific rooms are and track cultural differences between groups of people (Hillier et al., 1987, Hillier 1996, Hanson 1999). However, no attempt was made to define the similarity of graphs mathematically or attach further information to nodes or edges.

By assuming that the generality of a plan was increasing with permeability, Herthogs et al. (2019) gave an assessment method, called SAGA, building on JPGs, that used weighted graphs to quantify a building's capacity to support changes. In comparison to earlier works, e.g., Hillier and Hanson (1984), they also attempted to add further information to the graph's edges: a metric describing the permeability of the walls.

Dalton and Kirsan (2008) used graph theory to define similarities between justified plan graphs and track cultural differences between homes. The graphs' similarity was defined as the required transformation steps to make two graphs isomorphic, called graph edit distance (GED). The method presented good results on a small set of Cyprian Turkish and Greek houses.

3 DATASETS AND METHODS

This paper aims to determine such room arrangements that would be worth mass manufacturing, building on the methodology described by Dalton and Kirsan (2008). It was hypothesised that if enough floor plans were collected of flats loved by their owners and it was possible to cluster them based on similarity, those flats could be found, which were the most similar on average to the others in their cluster. These flats would be the genotype signatures (the golden means) of their cluster. It was also hypothesised that these genotype signature flats would be the best candidates for mass manufacturing, as they are the fairest discretisation of the solution space of reasonably well-arranged flats. Based on these, the two hypotheses are:

- H1. Graph Edit Distance can serve as an input for clustering flats based on their similarity
- H2. Genotype signature flats from the created preference clusters are better candidates for mass manufacturing than others.

The validation of the hypotheses required three different steps: a reasonable sample that can be considered as good flats, the identification of the genotype signature flats, and the verification that the clustering makes sense and that the chosen flats are indeed better than other flats.

3.1 Data collection

Collecting unbiased data on ideal housing is a challenging task. It would be misleading to deduce the quality of an apartment from sales data alone since the available housing supply shapes the demand. In addition, multiple other factors affect sales data (e.g., the location of the project), which should not be ignored.

Two possible ways for collecting ideal homes were considered. Personalised flats were collected from Hungarian new developments, where people could rearrange their flats. However, when looking at such flats, it became evident that the final layouts were highly determined by the original design, e.g., some flats were designed so that it was almost impossible to change their layout reasonably. Therefore, the second option, developing our own dataset, was chosen.

An online survey was developed for this aim, where participants were asked to assemble their ideal apartment from freely variable room elements. This provided an opportunity to abstract away from reality and to think about homes in a maximally bias-free way.

As we aimed at testing a proposed method for finding ideal home layouts, the particular features of the sample were almost irrelevant for this paper. We decided to use a convenience sample of Hungarian adults for practical reasons. The survey was advertised on various online platforms, including the research project's webpage, mailing lists, and social media sites. As an incentive, the option to participate in a prize draw of a £50 Amazon voucher was offered to the respondents. All questions were optional, and responses were fully anonymised before the analysis. The survey was subject to an institutional ethical review and received approval.

In the survey, participants could choose from a wide range of room descriptions containing a size estimate, a comfort level, and some features. For example, amongst the entrance halls, participants could choose from the following 3 options, summarized in Table 1:

Table 1: Graph Node equivalents of room types

Room name	**Description**	**Size estimate**
L Entrance	Spacious entrance hall with space to unload everything comfortably and with a place to sit down to pull off your shoes. Comfortable for multiple people to use it simultaneously.	8 m^2 / 85 ft^2
M Entrance	A medium-sized entrance hall where you can unload everything, but it is a bit uncomfortable for more than two people to use it simultaneously.	6 m^2 / 65 ft^2
S Entrance	Minimum-sized entrance hall with a large built-in wardrobe.	4 m^2 / 45 ft^2

While constructing their ideal homes from room units, respondents were asked to keep the size of their current homes to avoid unrealistic solutions. Those responses were filtered out where the total area of the ideal home significantly exceeded the actual home of the respondents' based on the criteria summarised in Section 3.2.

After finishing the survey, participants were asked to present how they would arrange the chosen rooms. For this reason, each participant received an e-mail with a personalised task within 24 hours after completing the survey. This e-mail contained their unique set of rooms in an editable PowerPoint and PDF format.

The participants' task was to relocate the room bubbles and connect them with lines. A line between two bubbles meant that one would put a door between two rooms. Overlapping bubbles meant that the respondent wanted to put the functions in the same space. Touching bubbles meant separate spaces connected without doors. By connecting the room bubbles to a bubble named Exterior, respondents were also asked to indicate whether they would want windows in that room or not. The abstraction steps from the spatial arrangements to graphs are visualised in Figure 1 with three examples.

3.2 Data Filtering

Those ideal homes that were larger with 20 m^2 or more or exceeded 150% of the respondents' actual homes were excluded from further analysis. This meant that in the case of a 30 m^2 studio, respondents shouldn't have designed a 45 m^2 one, and in the case of a 60 m^2 apartment (mostly suitable for two people only), the new one shouldn't have exceeded 80 m^2 (suitable size for families with multiple children. Those answers were also excluded where the area of the ideal home was over 130 m^2 because flats with such a large size are not part of the typical apartment mix of contemporary condominiums and their mass-production is not realistic either.

3.3 Recoding

The collected drawings were simplified and coded as graphs. In this step, all main bedrooms, studies, additional bedrooms were treated simply as rooms, and laundries, wardrobes, pantries, etc., as storages. In those cases where respondents indicated they would merge two spaces, the larger space (in most cases the Living Room) was put in the graph. A separate node was defined as the starting point,' O', symbolising the exterior (e.g., the foyer in front of the flat).

Table 2: Graph Node equivalents of room types

Room type	Graph Node
Exterior	O
Entrance	E
Foyer	E
Bedroom	R
Living room	L
Kitchen	K
Dining room	R

Room type	Graph Node
Pantry	S
Storage	S
Wardrobe	S
Laundry	S
Bathroom	B
WC	W
Study / Hobby room	R

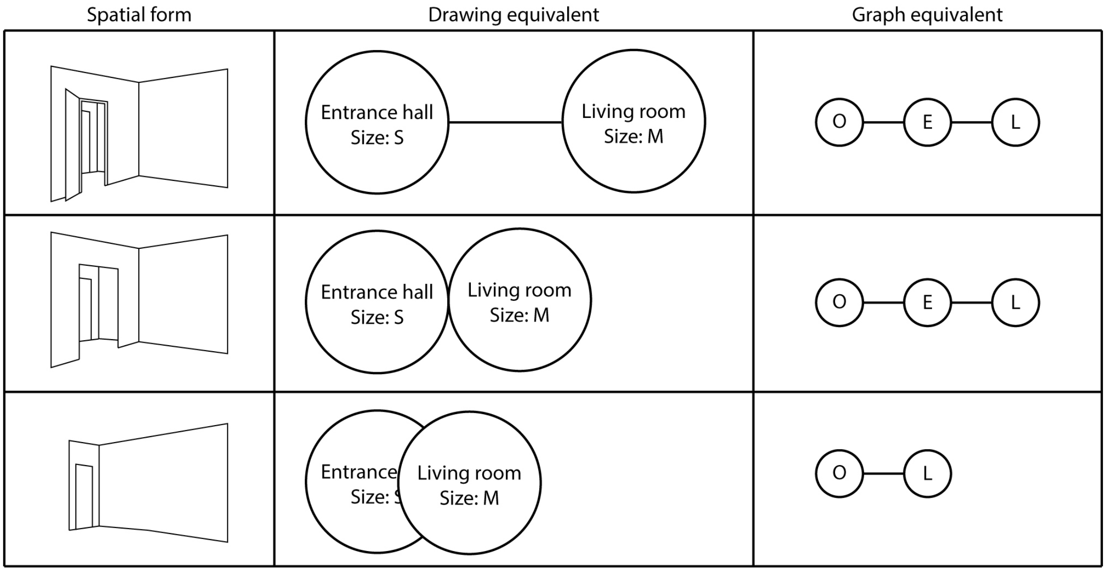

Figure 1: Abstraction from spatial forms to graphs

3.4 Clustering of room arrangements within predefined size groups

Graphs were grouped based on their size, using the number of main rooms, coded as R or L in the previous phase (e.g., bedrooms, living rooms, studies, etc.). This step saved a huge amount of computation time as in the next step we compared only those flats that could be considered real alternatives from the customer's perspective.

After creating these size groups, the next step was to make their spatial arrangements comparable by an objective similarity measure. This metric was the graph edit distance, a known metric to measure similarity in chemistry between molecules. For a similar architectural use case, Dalton and Kirsan (2008) used a different definition for GED than what is commonly used in other fields (e.g., in chemistry or machine learning). Their method permitted only such transformation steps that kept the graph interpretable to real-world configurations in each step, whereas traditionally, GED calculation methods only care about the final arrangement. During the trial runs, we found that even with using the traditional GED calculation method, the clustering represented reasonably well the professional intuition. Therefore, we simplified the process and used the traditional way. Still, it is a possible future research direction to implement an algorithm that can work with the special conditions Dalton and Kirsan (2008) defined.

A python package, Networkx 2.5, was used for the GED calculations. Because these calculations are computationally heavy, computing GED is NP-hard, the maximum GED to search for was set to 15, and the maximum time allowed for the search between two graphs was set to 10 minutes to make the calculations quicker. As plan graphs are relatively small compared to those frequently found in chemistry (only 9-10 nodes on average), the problem remained solvable, and the set time frame and max distance were enough to find the GED in all cases. However, with graphs over 15 nodes, the calculation would have become too slow, and approximation methods should have been considered.

After having an objective similarity metric, it was possible to cluster the graphs. This happened with a method called Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN Clustering, python package version 0.8.18) (Campello, Moulavi and Sander, 2013). This method used unsupervised machine learning to cluster data based on density, and it was flexible enough to use a GED matrix as an input (Berba, 2020).

Applying unsupervised machine learning techniques can be very tempting and useful to reveal patterns invisible otherwise. Still, its downside is that the parameter choices of the algorithm affect the outcome, in our case, the clustering, not in a fully predictable way.

The clustering parameters were set to allow single clusters. The *min samples* parameter was set to 1 to have a less conservative clustering, but the *minimum cluster size* was 4 to ensure more than three graphs would be put in each cluster. The *metric* was set to be precomputed to make it possible to use the graph edit distances, and the c*luster selection epsilon* parameter was set to 0. The clustering algorithm was run on each size group separately.

After clustering, those graphs were shortlisted that minimised the sum of GED within their groups. According to our hypothesis, the resulting graphs were the genotype signatures of their clusters and the best candidates for mass manufacturing.

4 RESULTS

During the data collection period, 73 Ideal Home graphs were collected. However, two graphs had to be excluded as the respondents clearly didn't understand the task, and their submitted solution didn't represent an interpretable layout. So, the final dataset contained 71 graphs. Surprisingly, all layouts were unique, and nobody could make their ideal home under 40 m^2. This result may indicate that we should not mass manufacture thousands of micro homes. The size distribution of the received layouts is summarised in Table 3.

Table 3: Size distribution of the received Ideal Homes

Flat size	Sample size
Apartments with 1 room	3
Apartments with 2 rooms	14
Apartments with 3 rooms	24
Apartments with 4 rooms	25
Apartments with 5 rooms	5

4.1 Genotype Signature Graphs

The samples of one-room and five-room apartments were considered too small for further analysis and excluded from the clustering. The algorithm was run on the remaining three subsets and resulted in one cluster in each category. The genotype signature graphs of the clusters are presented in the first column of Figure 2.

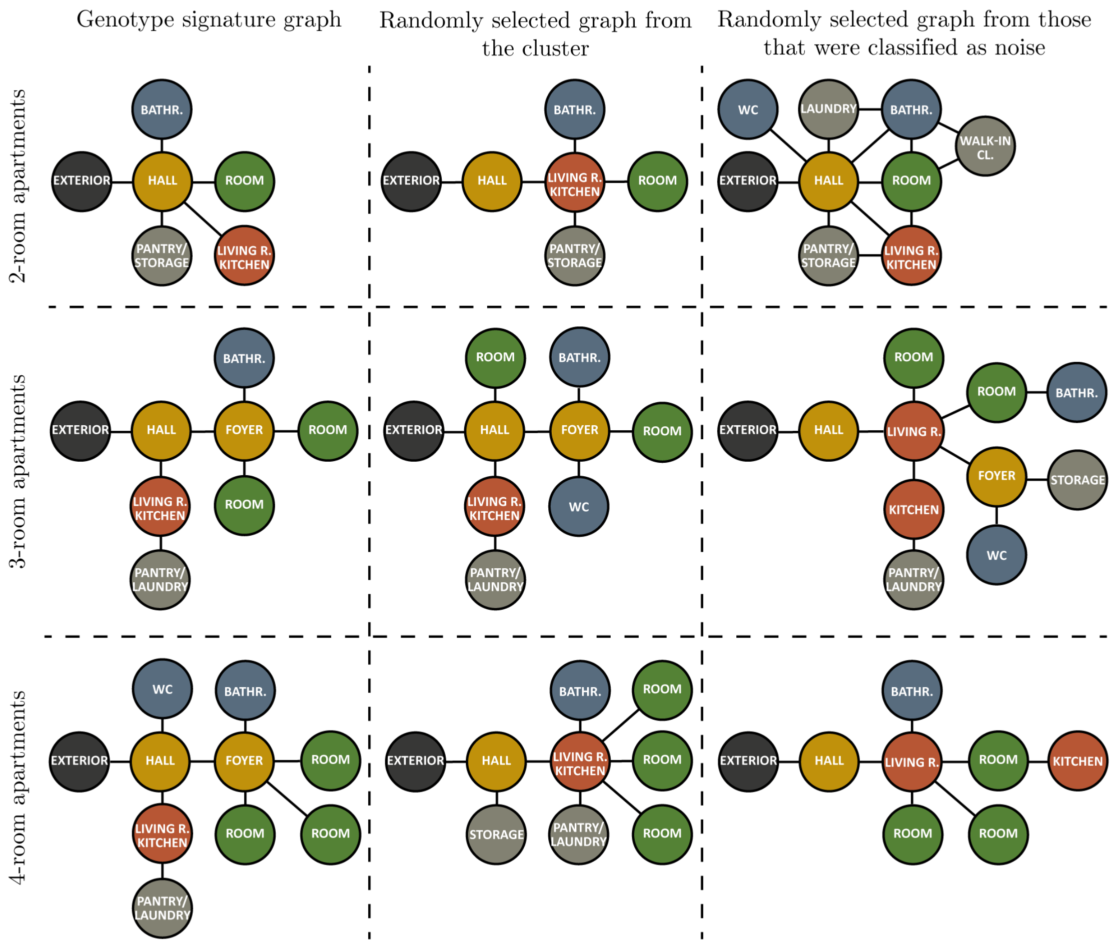

Figure 2: Genotype signature graphs and additional graphs selected for the evaluation process.

4.2 Ideal room sizes

As an interesting side result, it was also possible to use the collected data to create an Ideal Room Sizes database that can show how frequently each room option was chosen within each apartment size category. From this table, we can see how many people prefer a specific room type and the less popular room options also became clearer.

If we were up to mass manufacturing residential buildings and creating flat types, it would be advantageous to design them to meet the majority's needs. This Ideal Room Sizes database can give us information on the safe room options to incorporate and the ones to avoid.

5 EVALUATION OF THE METHOD

We conducted a second survey to examine how the genotype signature method performs. In this survey, the opinion of our former participants was asked about 3x3 graphs and apartment layouts, and we were interested in whether the genotype signature layouts do outperform the others.

5.1 Dataset and Method

We created a small test set for each of the three size groups where a genotype signature graph was found. These sets, shown in Figure 2, contained the genotype signature graph, a randomly selected graph from the cluster and another graph classified as noise by the algorithm. This way, we ended up with 3 graphs in each size group, amongst which we expected the genotype signature graph to perform the best and the graph that was considered to be noise to perform the worst.

Additionally, apartment layouts were developed based on each graph to test the performance in a less abstract form too. However, one graph can correspond to many different apartments, so the developed 3D models reflect not only on the underlying plan graph, but the architect's talent and boundary conditions, like the position and number of windows and the proportions of the floorplan. In addition, when people evaluate buildings, many different aspects are considered, including the shape, size, proportions, orientations, and furniture. This makes the comparison of the apartments much trickier than the comparison of the graphs. However, at the end of the day, the apartments are the ones people make decisions about.

Figure 3: Developed apartments in the 2-room category. From left to right: The genotype signature, the one from the cluster, and one from the noise

While developing the layouts, we tried to control as many variables as possible. Thus, the implementation of the apartments followed these guidelines:

- Each apartment had a rectangular shape, where the entrance door was located on the longer side.
- Within each size category, apartments had the same adjacent walls and size: 60 m^2 for 2-room apartments, 75 m^2 for 3-room apartments and approximately 90 m^2 for 4-room apartments.
- Each apartment had the same vertical shafts and windows in each size category.
- The furniture consisted of mainly the same elements. For instance, in all cases, the bathroom consisted of a shower, although a bathtub could have been better in some cases; and in the dining area, a separate dining table was preferred over the less common but sometimes really useful kitchen islands. All living rooms consisted of sofas for 4 people, a dining area for 6 and a home office area.

The final dataset contained 3 graphs and 3 apartment plans for each category of the 2, 3 and 4-room apartments. In the survey, first, respondents had to evaluate on a 7-point Likert scale the graphs in each size category separately, and when they finished, they could move on to the apartment layouts to do the same task.

5.2 Results of the evaluation

The data collection period lasted for two weeks between the 24th of November 2021 and the 8th of December 2021. As our original dataset was a convenience sample, we asked the same people who provided us with graphs earlier to verify the results. We collected 53 answers during the data collection period, with a remarkably high, 75% response rate.

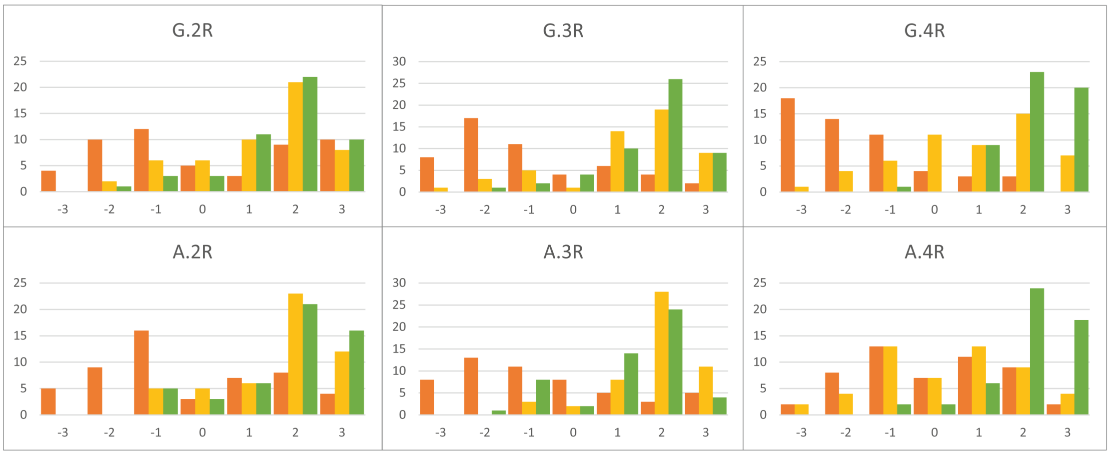

Figure 4: Preference distributions, based on Table 4. First row: Graphs in the 2, 3 and 4-room apartment size category, second row: Apartments in the 2, 3, and 4-room apartments size category. Genotype signatures are shown in green, cluster members in yellow, and noises in orange across all 6 cases.

Table 4: Responses received for the survey.

			Scores							N	Average	Rank
			-3	-2	-1	0	1	2	3			
Graphs	2-room	Genotype signature	0	1	3	3	11	22	10	50	1.60	1
		Cluster member	0	2	6	6	10	21	8	53	1.25	2
		Noise	4	10	12	5	3	9	10	53	0.13	3
	3-room	Genotype signature	0	1	2	4	10	26	9	52	1.63	1
		Cluster member	1	3	5	1	14	19	9	52	1.25	2
		Noise	8	17	11	4	6	4	2	52	-0.94	3
	4-room	Genotype signature	0	0	1	0	9	23	20	53	2.15	1
		Cluster member	1	4	6	11	9	15	7	53	0.81	2
		Noise	8	17	11	4	6	4	2	53	-1.58	3
Apartments	2-room	Genotype signature	0	0	5	3	6	21	16	51	1.78	1
		Cluster member	0	0	3	2	8	28	11	52	1.63	2
		Noise	5	9	16	3	7	8	4	51	-0.27	3
	3-room	Genotype signature	0	1	8	2	14	24	4	52	1.21	2
		Cluster member	0	0	3	2	8	28	11	53	1.81	1
		Noise	8	13	11	8	5	3	5	53	-0.66	3
	4-room	Genotype signature	0	0	2	2	6	24	18	52	2.04	1
		Cluster member	2	4	13	7	13	9	4	52	0.31	2
		Noise	2	8	13	7	11	9	2	52	0.00	3

The results of the evaluation of the method show that the respondents' preferences are more or less similar to the expected results. The average ratings projected well the expected preference based on the analysis.

As Figure 4 and Table 4 present, the genotype signature graphs (green) in all three cases outperformed the other graphs. As for the apartments, the genotype signature graph is only once was outperformed, in the case of the 3-room apartments, by the other graph in its cluster (yellow). In all 6 cases, the graph/apartment considered as noise (orange) had the most significant rejection and performed the worst.

6 APPLICATION

A script has been created to implement all new knowledge into a piece of software that could be turned into an easy-to-use tool helping designers and developers evaluate design solutions and identify potential issues with layouts. This script can compare home layouts to the genotype signature layouts and evaluate room sizes based on the Ideal Room Sizes database.

To make this script easily accessible for people who have only a basic programming background (e.g., experience with Grasshopper), the script was developed in Dynamo, a visual programming platform automatically installed with Autodesk Revit. This is a platform that many architects are familiar with, and it can take advantage of the data handling and visualisation a properly created BIM model can offer.

6.1 The programming environment

The first implementation was done on Autodesk Revit 2022 and Dynamo version 2.10.1. The code contained self-written Python codes and nodes from publicly available programming packages (Revit, Clockwork 2.3.0).

The room size evaluation part of the script worked without any more complicated setup, simply by running the script under Dynamo on any computer where Revit and Clockwork nodes had already been installed. However, the graph analysis required some additional setup. A new Python environment had to be created on the computer with all required packages installed in a specific folder. This enabled Dynamo to use Python packages, like Networkx, for the graph analysis. Nodes using these packages had to be run with CPython3 with the sys.path location appended to obtain the created folder instead of the previously common IronPython2 engine.

6.2 Pseudocode

When running the script, first, it obtained all rooms created in the model. It was important to ensure that the model contained only rooms that were placed indeed. These rooms were sorted then based on a parameter that described which room belonged to which flat. All balconies were removed from the list of rooms because they were not practically replaceable with any other rooms. At the end of the first phase, we had all real rooms grouped by flats.

At the second stage, the algorithm determined which rooms were neighbouring. These rooms were either separated by doors or with room separation lines. While the first option was easy to handle, the second was more complicated. Therefore, the following guidelines were created to deal with room separation lines: 1) one line shall separate two rooms only, but separations can be made of multiple lines, 2) room separation lines are considered similar to doors only if their total length is under 2 m. If a room separation line bordered more than 2 rooms, it should be divided into multiple lines during the modelling process. When their total length exceeded 2 m, the rooms would be considered separate functional units in the same room, so the algorithm merged the rooms, keeping the largest one’s name.

Neighbouring rooms were grouped then by the ID of the flat they belonged to. Rooms in front of the flats were also included in each group and considered as a starting point ‘O’ for the graphs. At the end of the second stage, the algorithm listed all neighbouring rooms as pairs for each flat.

The next step was to define both the nodes and the edges of each plan graph. This step required a strict naming convention that ensured that rooms with the same function got the same name. This convention was also required to link the Ideal Room Sizes database and the Revit model. Each flat was coded as graphs by the end of this phase.

The flats were grouped then based on the number of the main rooms, and in the last step, they were compared to the genotype signature graph of their size category. This step determined the Graph Edit Distance between layouts in the model and predefined genotype signatures.

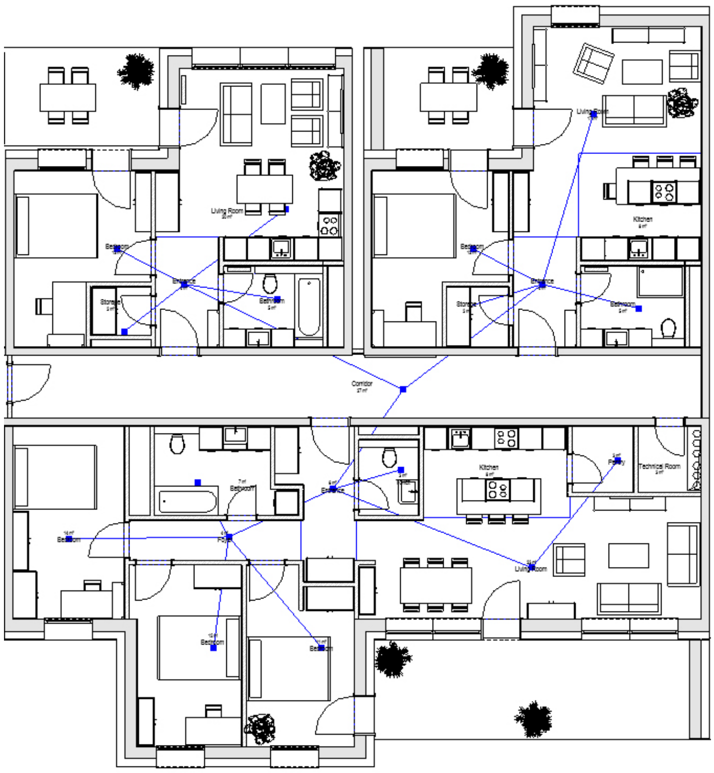

Figure 6: Graphs drawn over genotype signature apartment layouts by the Evaluator script in a Revit test file.

Figure 7: Genotype signature apartment layouts identified by the Evaluator script in a Revit test file, and rooms coloured based on the Ideal Room Sizes database.

The sizes of the rooms were compared to the Ideal Room Sizes database. When a room has been merged with other rooms, its size was calculated as the sum of the area of all merged ones.

At the end of the evaluation, rooms were put in 3 different lists. Those whose size was within limits got an olive-green colouring, those that exceeded the requirements were coloured in green, and those with smaller areas were coloured in red. Rooms that had been merged automatically got the largest included room's colour. The colourings were visualised within Dynamo, but it would have also been possible to create a view in Revit and overwrite the colours of the areas accordingly. The room lists could also be exported to excel.

7 DISCUSSION

Over 70 Hungarians were asked to create their ideal homes from predefined room options in an online experiment. The proposed data collection method turned out to be a great tool to examine the common characteristics of the homes that people find neat. By using gamification, people could get less biased by the realistic options they have currently.

Based on the resulting database, it was possible to understand the preferences' distribution, make data-based requirements for room sizes, and identify potentially risky layouts, where proportions deviated significantly from the majority's preferences.

The collected flats were clustered based on the similarity of their functional graphs. It was assumed that those layouts that were the most similar on average to the other layouts in their cluster were the ones that would maximise residential satisfaction if mass produced. These layouts were determined with graph analytical tools and developed further to real apartment layouts to verify the proposed methodology.

The evaluation generally verified the results. In 5 cases, among the 6 examined, the genotype signature graphs performed the best. The reasons for the only one mismatch include that, first, the randomly selected graph was very close to the genotype signature. Second, an even better layout may exist than the examined ones, which combines the most attractive features of both, but as it was not part of the dataset, the algorithm could not find it. For example, a graph like this could be one with the structure of the current genotype signature graph with an additional separate toilet. In all 6 cases, those layouts performed the worst that were classified as noise by the algorithm.

Therefore, our two hypotheses turned out to be true. It was possible to do the clustering, and we provided some evidence that those graphs that were the genotype signature flats outperformed the others. Furthermore, a script was produced to evaluate any new apartment designs created in Revit based on the developed guidelines.

This script compared the relevant genotype signature graph to each modelled apartment and coloured their rooms based on the Ideal Room Sizes database. However, the algorithm compared the apartments only to pre-given genotype signature graphs to make calculations quick, so it was not examined whether a modelled apartment should be considered as noise. Nevertheless, the GED still can be informative in cases where a strict convention exists to follow a specific design.

8 CONCLUSIONS

This paper showed that it is possible to use small graph matching on a larger scale than Dalton and Kirsan (2008) did by applying a density-based clustering algorithm and simplifying the GED calculation process.

By providing an initial database of unbiased preference data on layouts and room sizes for Hungarian adults, this project helps create homes that fit the demands of modern life, enhance their residents' well-being, and help them fully recharge. With the proposed methodology, further data could also be collected on other nations or specific social groups as well. The outlined design methodology can help us be more aware of what a well-designed apartment is like. The created genotype signature apartments could create an exemplar dwelling family that showcases good design. In addition, with the Evaluator script, it is possible to evaluate any new apartment designs created in Revit based on the developed guidelines and get instant feedback on how much a design fits people's preferences. Therefore, the tools created by this project can help designers and decision-makers evaluate design solutions from the perspective of expected resident satisfaction and thereby avoid creating apartments that do not meet the needs of the majority or are predictably substandard.

Future work can include an improved script for less traditional GED calculations and considerations about attaching further information to the graphs to include features like size, the proportion of rooms, or the openness of a connection.

REFERENCES

Alexander, C. (1964) *Notes on the Synthesis of Form*. Cambridge: Harvard University Press.

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fikidahl-King, I. and Angel, S. (1977) *A pattern language*. Oxford: Oxford University Press.

Bai, Y., Ding, H., Bian, S., Chen, T., Sun, Y., & Wang, W. (2019). 'SimgNN: A neural network approach to fast graph similarity computation', in *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, pp. 384–392.

Berba, P. (2020) *Understanding HDBSCAN and Density-Based Clustering*. Retrieved from: https://towardsdatascience.com

Campello, R., Moulavi, D. and Sander, J. (2013) 'Density-based clustering based on hierarchical density estimates', in *Pacific-asia conference on knowledge discovery and data mining*, pp. 160–172. Heidelberg: Springer Berlin. Retrieved from: https://hdbscan.readthedocs.io/en/latest/basic_hdbscan.html

Dalton, R. C. and Kirsan, C. (2008) 'Small-graph matching and building genotypes', *Environment and Planning B: Planning and Design*, 35 (5), pp. 810–830.

Farha, L. (2017) *Statement by the Special Rapporteur on the right to adequate housing, Leilani Farha, during the Interactive Dialogue at the Human Rights Council*. Retrieved from: https://www.ohchr.org/en

Habraken, N. J. (1972) *Supports: An alternative to mass housing*. London: The Architectural Press.

Hanson, J. (1999) *Decoding Homes and Houses*. Cambridge: Cambridge University Press.

Hay, R., Samuel, F., Watson, K. J. and Bradbury S. (2017) 'Post-occupancy evaluation in architecture: experiences and perspectives from UK practice', *Building Research & Information*, DOI: 10.1080/09613218.2017.1314692

Herthogs, P., Debacker, W., Tunçer, B., De Weerdt, Y. and De Temmerman, N. (2019) 'Quantifying the Generality and Adaptability of Building Layouts Using Weighted Graphs: The SAGA Method', *Buildings*, 9 (4), 92. Retrieved from: https://www.mdpi.com/2075-5309/9/4/92

Hillier, B. (1996) *Space is the machine*. Cambridge: Press Syndicate of the University of Cambridge.

Hillier, B. and Hanson, J. (1984) *The Social Logic of Space*. Cambridge: Cambridge University Press.

Hillier, B., Hanson, J. and Graham, H. (1987) 'Ideas are in things: an application of the space syntax method to discovering house genotypes', *Environment and Planning B: Planning and Design*, 14 (4), pp. 363–385.

Mándoki, R. and Orr, J. (2019) *The Sustainability of Standardisation* (Master of Research Thesis). Cambridge: University of Cambridge

Mándoki, R. and Orr, J. (2021) 'The social acceptance of mass produced residential buildings among Hungarian young adults', *Architectural Engineering and Design Management*, DOI: 10.1080/17452007.2021.1985425

March, L. and Steadman, P. (1971) *The Geometry of Environment: An Introduction to Spatial Organization in Design*. London: RIBA Publications.

McKinsey Global Institute (2017) *Reinventing Construction: A route to higher productivity*. Retrieved from: https://www.mckinsey.com

Raymond, J. W., Gardiner, E. J. and Willett, P. (2002) 'RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs', *The Computer Journal*, 45 (6), pp. 631–644.

Rippmann, M., Lachauer, L. and Block, P. (2012) 'Interactive Vault Design', *International Journal of Space Structures*, 27 (4), pp. 219–230.

Shields, R. (2012) 'Cultural Topology: The Seven Bridges of Königsburg, 1736', *Theory, Culture & Society*, 29 (4-5), pp. 43-57.

Steadman, P. (1970) *The Automatic Generation of Minimum-standard House Plans*. Cambridge: University of Cambridge Department of Architecture.

UN General Assembly (1984) *Universal Declaration of Human Rights, 217 A (III)*. Retrieved from: https://www.refworld.org/docid/3ae6b3712c.html