THE MANY FACES OF SIMILARITY

A Visual Analytics Approach for Design Space Simplification

AHMED M. ABUZURAIQ[1] and HALIL ERHAN[2]
[1,2]*Simon Fraser University*
[1,2]*{aabuzura|herhan}@sfu.ca*

Abstract. Generative design methods may involve a complex design space with an overwhelming number of alternatives with their form and design performance data. Existing research addresses this complexity by introducing various techniques for simplification through clustering and dimensionality reduction. In this study, we further analyze the relevant literature on design space simplification and exploration to identify their potentials and gaps. We find that the potentials include: alleviating the choice overload problem, opening up new venues for interrelating design forms and data, creating visual overviews of the design space and introducing ways of creating form-driven queries. Building on that, we present the first prototype of a design analytics dashboard that combines coordinated and interactive visualizations of design forms and performance data along with the result of simplifying the design space through hierarchical clustering.

Keywords. Visual Analytics; Design Exploration; Dimensionality Reduction; Clustering; Similarity-based Exploration.

1. Introduction

In the domains of design involving physical artifacts (e.g. architectural, mechanical and industrial design), the geometric forms of design alternatives and their performance are interrelated and equally important. Each presents information relevant to different design concerns. Generative design methods, including parametric design modeling, enable the creation of a large number of design alternatives. These methods can be augmented through tools that can estimate the performance of each alternative as they are created. This design process, guided with performance metrics, can be referred to as performance-based design (Shi, 2010). The combination gives rise to a design process where considerations that would traditionally take place at late design phases can now become part of the early formative phases. This warrants attention to research for systematically reviewing and building tools for interacting with and exploring design spaces considering design performance data beyond what design catalog systems currently offer (Brown and Mueller, 2017b).

An important artifact of this process that presents both opportunities and challenges for design is the large amount and different types of design data

RE: Anthropocene, Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2020, Paper 443 (Preprint).

involving both geometry and design performance. The field of Visual Analytics combines and leverages both interactive visualizations and data analysis for the sense-making of complex and abundant data (Thomas and Cook, 2006). We suggest that we are faced with a similar challenge in design decision-making when exploring design alternatives, as addressed by the visual analytics literature. Unlike data analysis, the main task in design is form-finding given design constraints and goals as part of a design brief. Focusing primarily on data may cause a lapse in this main task, while without data, design can be full of oversights.

For better supporting this process, in this research, we propose the use of design alternatives' form and performance similarity as a way for simplifying the design space and further investigating their affordances in an interactive data visualization setting. Although previous research (Beham et al., 2014; Erhan et al., 2015; Chen et al., 2015) were based on similar arguments, the conceptual justification that derived their solutions can be further expanded and applied to better demonstrate a seamless task integration between design exploration and data analysis, which we call 'design analytics'.

In this work, we build on the argument for design space simplification (Erhan et al., 2015; Brown and Mueller, 2017b) through the techniques of clustering and dimensionality reduction which expose and exploit the (dis)similarity between design alternatives. We further extend it by integrating the results of this simplification in a larger, interactive design analysis dashboard. We also find that this similarity-based simplification of the design space has many faces, in the sense that it can be achieved via varied means, represented in different ways and most importantly, has many potential applications for design space exploration. These applications include alleviating the choice overload problem, opening up new venues for interrelating design forms and data, creating visual overviews of the design space and introducing ways of querying alternatives through form-driven criteria. We create conceptual distinctions and remark about the use of design simplification by reflecting on the examples in the literature. Finally, we take this reflection one more step and present the first iteration in this line of research of a design analytics dashboard that combines the visual representations of design forms and performance data with similarity-based representations.

2. Information Visualizations: Key terms and Concepts

Visual Analytics combines interactive visualizations and data analysis. We address the use of both in design exploration. We will next introduce a few concepts from the information visualization literature next then proceed to data analysis in the form of data clustering and reduction in the next section.

Key terms: A view is a visual representation of some data (e.g. a scatterplot is a view representing two quantitative variables). Brushing is the act of highlighting elements on a view via user interaction. Multiple Coordinated Views (MCV) are views whose visual representations are changed in response to interactions (e.g. brushing, filtering) on the others. Multidimensional datasets are characterized by having multiple variables/metrics per record. For example, each design alternative is a single record and can be evaluated through multiple performance metrics.

Overview then Detail-on-Demand: To help us understand how data visualization tools can support the task of design exploration, we can be guided by the information-seeking mantra of "Overview first, zoom and filter, then details-on-demand" by Shneiderman (1996). In this work, we focus on the visual representations (also referred to as views) that provide an overview of information. We also consider the interactions between these views and on them. An overview view is a starting point for the exploration upon which the designer can focus their attention and attend to the details.

3. Design Space Simplification

Clustering and Dimensionality Reduction techniques group alternatives in a way that enables a hierarchy of choice and puts order into the design space.

3.1. CLUSTERING

Clustering is the process of grouping items based on their similarity. The similarity between any two alternatives can be defined in terms of a subset (or all) of their associated data such as input parameters, performance metrics or their geometric features. The subsets chosen will highlight certain aspects of the two alternatives and possibly suppress others. For example, we may compute the similarity between two tower buildings based on the difference in their heights. While this will enable us to cluster together buildings of similar heights, buildings within the same cluster can vary with respect to other aspects such as their surface area, or energy usage. Hierarchical clustering techniques produce a tree where the leaves are individual alternatives. Similar alternatives belong to clusters which, in turn, belong to a smaller number of clusters and so on until the root of the tree is reached.

In recent works, we find a number of examples using hierarchical clustering methods that are performed for visualization purposes. In Erhan et al. (2015), design alternatives are compared based on the distances between their input parameters after which they are hierarchically clustered and visualized as a dendrogram tree. The system Cupid (Beham et al., 2014) is aimed at exploring the generative space of a geometry generator by categorizing the different shapes it can produce and understanding the relationship between the input parameters and these categories. Initially, the distance between alternatives is computed by directly comparing their geometry (average of minimum distances between their corresponding mesh vertices). Next, hierarchical clustering is performed using the calculated distances and clusters are visualized on both a hierarchical radial tree and a composite parallel coordinates plot. The two views are coordinated together through brushing. Chen et al. (2015) first cluster alternatives based on the similarity between their performance metrics and then, within each cluster, they are clustered again based on chosen quantitative architectural features. Furthermore, the work by Brown and Mueller (2017a) uses clustering techniques for identifying families of similarly performing design alternatives and illustrates their use in design space exploration.

3.2. DIMENSIONALITY REDUCTION

Dimensionality Reduction (DR) is a set of techniques that assists in analyzing multidimensional data by minimizing the variables under consideration or mapping the data into a new, smaller space. This is done for the purposes of simplifying or denoising the data while retaining any intrinsic patterns found in the multidimensional data. For example, in the context of parametric design models, a large number of parameters usually control the generated geometric variations leaving designers with the challenge of discerning the range of geometries that a parametric model can express. Harding (2016) proposed applying a DR technique to map design alternatives, through their input parameters, into a two-dimensional grid. As a result, the geometric variations can be more readily pronounced as we move along any of the two, newly synthesized dimensions. The technique used by Harding is called Self-Organizing Maps (SOM), first introduced by Kohonen (1982), and it aims at creating a spatial map whereas inputs that are similar in their original multidimensional space are close to each other on the resulting map.

The spatial map resulting from SOM have been used in different ways such as: Creating a visual structure that can be colored or extruded to represent performance data (Fuchkina et al., 2018; van Kastel, 2018), or can be augmented with visuals (e.g. thumbnails or 3D representations) to give an overview of the geometric variations in the design space (Harding, 2016; Fuchkina et al., 2018; van Kastel, 2018; Pan et al., 2019). Furthermore, when the number of design alternatives is more than the cells in the map, SOM can be used for clustering (Erhan et al., 2015).

4. The Many Faces of Similarity

The simplification of the design space through clustering or dimensionality reduction has a number of applications for design space exploration as can be construed by reflecting on the examples we surveyed. To present these applications in context, we first discuss the relation between design forms and their associated performance data. Followed by their implications on design data visualization. We also discuss salient issues in design exploration tools that could then be tackled through simplification.

4.1. FORM AND DATA VISUALIZATIONS OF DESIGN ALTERNATIVES

In design, the geometric forms and their performance data are interrelated and equally important. Each presents information on different aspects of the design alternatives. Therefore, we expect that the interactive visualization for design analysis should enable the exploration of form and performance data both independently from each other and also in unity. Given the mutual importance of design forms and their associated data, we find it useful to break down the components of Shneiderman (1996) taxonomy to focus on either design forms or data, e.g. an overview task is then performed on either a data or form overview.

A common visualization technique of multidimensional data is the parallel coordinates plot (PCP). The PCP can provide an overview of the data patterns and correlation between alternatives. The existing design exploration interfaces make use of PCPs extensively for providing an overview over data (Beham et al., 2014; Abi Akle et al., 2017; Mohiuddin et al., 2018; Tomasetti, 2019). On the other hand, a grid of thumbnails in DreamLens (Matejka et al., 2018) or a radial tree visually representing clusters of similar forms in Cupid (Beham et al., 2014) provides an overview of the forms in the design space.

These visualizations can also be coordinated. For example, views of design forms in DreamLens and Cupid (a thumbnails grid and a radial tree respectively) are linked to the data views (a scatterplot and a PCP) via brushing. These are examples of Coordinated Multiple Views. The work by Javed and Elmqvist (2012) views (Javed and Elmqvist, 2012) identify a list of techniques for compositing these views. The aforementioned examples fall under Juxtaposed views. We also see other types of views' compositions (Nested) in the Design Space Explorer (DSE) (Fuchkina et al., 2018) and the thesis by van Kastel (2018).

In the Design Space Explorer (DSE) (Fuchkina et al., 2018), the design form views are nested in the data view. DSE creates a 2D hexagonal grid through SOM on which thumbnail images of forms are overlaid and the cells in the grid are colored based on the performance of the alternatives in each of these cells. In his thesis, van Kastel (2018) employs a similar technique of creating a combined overview but situates the visualization in a 3D digital environment where it encodes the design data geographically, e.g. by using soil layers, terrain variations, and water levels. By introducing this distinction between data and form views, whether they are juxtaposed or nested, we can clearly understand the kind of questions we can ask through interactions that link them. The left part of Figure 1 illustrates linking independent form and data views but can also be useful when thinking about combined views such as the ones in DSE or van Kastel's thesis.

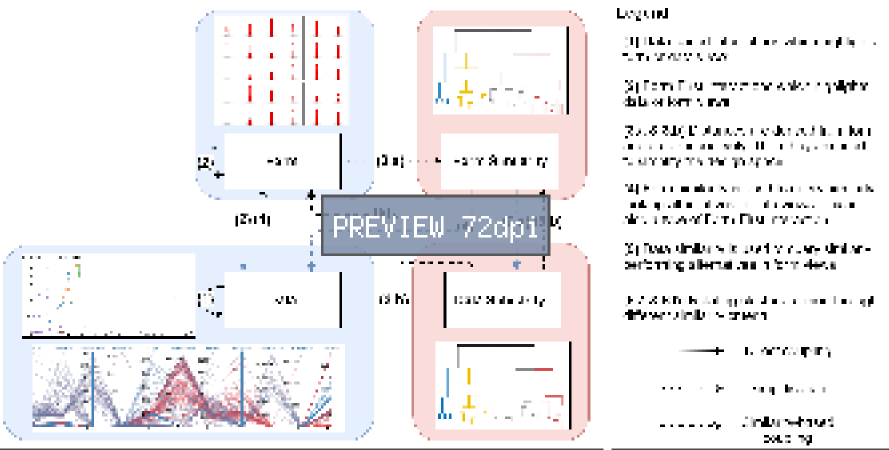

Figure 1. A conceptual model of the coupling of design data and forms in a visual analytics context. The Form and Data Similarity is derived by computing the distances between design alternatives. Examples of views for each are shown in the same colored block.

4.2. CALL FOR FORM-FIRST INTERACTIONS

We believe that the form overview visualizations in the systems discussed above can be further expanded to enable interactive form-driven queries that originate on these visualizations. This is analogous to the queries that are formulated on data overviews e.g. like brushing on parallel coordinates or scatterplot matrices. Examples of form-driven queries include: when a thumbnail in DreamLens is selected, then the point in the scatterplot associated with it is highlighted. Additionally, the chisel tool in DreamLens (Matejka et al., 2018) filters design alternatives by directly manipulating their superimposed forms. We also see an example in the radial tree in Cupid (Beham et al., 2014). Since the tree is created by clustering alternatives based on their geometric similarity, then brushing parts of that tree is essentially a form-driven query. We refer to these types of interactions, whether used for filtering or highlighting, as Form-First interactions. We expect that research into these interactions can open up new possibilities for supporting the explicit encoding of form-based design criteria.

4.3. SIMILARITY-BASED COUPLING OF FORM AND DATA

Although both DreamLens (Matejka et al., 2018) and Cupid (Beham et al., 2014) aim at visually representing design spaces, they show differences in how they enable interactions on them. The prior system implements a type of Form-First interactions on the raw 3D or 2D representations of the design forms through its 'chisel' tool. Cupid, on the other hand, first simplifies the design space through clustering and as a result, supports interactions on subsets of the design space. Both systems consider the visual similarity often found between generated design alternatives as an important factor in the design of their respective interactions.

We can refer to linking design forms and data in their raw format as direct coupling to distinguish it from the similarity-based coupling. Similarity-based coupling starts by computing the similarity between design alternatives. This computed similarity can be used implicitly or explicitly. An example of the implicit use can be found in DSE (Fuchkina et al., 2018), where it is possible to select an alternative on the SOM representation and highlight the thumbnails of alternatives that are visually similar to it. Explicit use of the similarity data starts by using it to simplify the design space through clustering or dimensionality reduction then visually represents their results as 2D maps, trees, or treemaps to name some. Interactions on the results of this simplification that then updates other data views can be referred to as an: explicit similarity-based coupling (ESC). An example of an ESC is the interaction on the radial tree in Cupid (Beham et al., 2014) which in turns highlights other data views in the system. Brushing on SOM representations (e.g. as in the DSE) that updates other views can also be considered an ESC. This is because of the similarity-preserving property of SOM, in that alternatives that are similar to each other are also closeby on the map.

These explicit representations can be used to provide an overview of the design space and to reduce the cognitive overload that can result from a large number of alternatives. Interacting with these representations enable us, depending on the clustering criterion, to either query the performance of alternatives with similar forms or the appearance of similarly performing alternatives. An example of the first can be seen in (Beham et al., 2014), in the use of radar charts for visualizing derived properties of similar geometries (i.e. they belong to the same cluster). We can see examples of the latter in the works by Chen et al. (2015), as well as Brown and Mueller (2017a) where clustering was shown to allow querying the different geometric variations that exhibit similar performances. In Figure 1, we illustrate both Similarity-based coupling and Form-First interactions.

4.4. A CASE FOR FORM-BASED SIMILARITY

The similarity between design alternatives can be computed in different ways depending on the aspects under comparison. If the similarity measured between alternatives correlate with their geometric or visual similarity, then we can refer to that as form-based similarity. Interactions that use form-based similarity, whether implicit or explicit, can be treated as Form-first interactions (which we argued for in Section 4.2). This applies to interactions with the results of both clustering and dimensionality reduction techniques that use form-based similarity data.

Form-based similarity can be computed by comparing input parameters if similar parameters also result in similar geometries. The applicability of this assumption is highly limited for design spaces with high complexity (Nagy et al., 2017) where the relationship between the input parameters and outputs is not predictable. Furthermore, in a design process that involves multiple generative design models, whether as iterations on a single model or resulting from a collaborative setting, we would like to be able to jointly explore them. Relying on the shared input parameters between them might not be possible or useful. Instead, we argue for comparing the geometric forms directly. This can be accomplished through shape similarity approaches which represent a geometric form (2D or 3D) numerically so that regular distance functions like the Euclidean distance can be used (Bustos et al., 2005). Alternatively, we can directly compare geometries by procedurally calculating the distances between their corresponding mesh vertices as is done in Cupid (Beham et al., 2014). In Figure 1, arrows (4) and (5) illustrate similarity-based coupling. Arrow (4) is also a type of Form-First interaction since it is initiated from a representation derived from form-related similarity.

5. First Design Iteration

Building on the arguments we outlined earlier, we describe our first design analytics dashboard prototype and the dataset we used to evaluate it.

5.1. DATASET FROM A CASE STUDY

During a SmartGeometry (2018) workshop, a group of designers was asked to develop proposals for a mixed-use high-rise tower design. The tower is to be located in a downtown context and the necessary geographical and climate data was provided. Participants were provided with a common set of performance calculation modules that produced thirteen different metrics covering aspects such as floor areas per function, solar energy gain, and energy usage among others. Six designers submitted design alternatives generated by the parametric design models they developed reaching a total of 250 alternatives. Each of the alternatives has a geometric form, performance metrics, and input values. For more details, the interested reader may refer to the case study by Erhan et al. (2020).

5.2. DESIGN DECISIONS

We developed a visualization dashboard to analyze the data mentioned above. We improved this dashboard to experiment with similarity-based exploration (Figure 2), which integrates Form and Data views along with Form and Data Similarity views. Here, we will motivate the choices made in terms of the views and the interactions between them with respect to the framework in Figure 1.

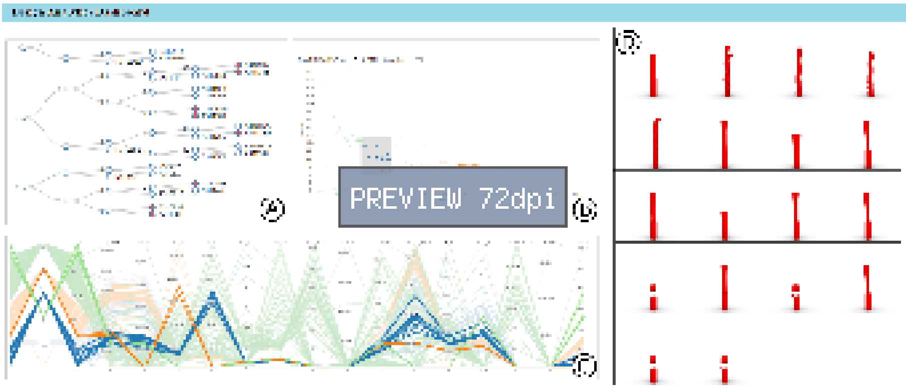

Figure 2. A: Dendrogram tree (Form or Data Similarity view), B: Scatterplot (Data view), C: Parallel Coordinates Plot (Data view), D: thumbnails grid (Form view).

5.2.1. Visualization Views

In this iteration, we have decided to provide both data and form views which are a parallel coordinates plot and a scatterplot as data views, and a grid of thumbnails as a form view. Dendrogram trees that represent hierarchies are chosen to present clusters in the design space with respect to either design forms or performance data. Because of the multidimensional nature of our performance data, we decided to use the parallel coordinates plot (PCP) as suggested by (Abi Akle et al., 2017). The PCP provides an overview visualization of performance data, enables filtering based on value ranges and can expose correlation patterns. In addition to the PCP, we have chosen to add a scatterplot chart to enable finding outliers and detecting a correlation between performance metrics. Since the design alternatives in our dataset come from multiple design models, we color the lines in the PCP and the scatterplot depending on the model they were generated from to facilitate inter-model comparison. We may also limit ourselves to a single model which then allows us to relate input parameters to performance metrics.

We have argued for design space simplification through dimensionality reduction and clustering. In this iteration, we decided to start our exploration with hierarchical clustering. We chose to compute the distances between design alternatives in two ways: The first takes the euclidean distance between the performance metrics of each of the alternatives, and the second compares their geometry directly using Hausdorff distance (i.e. form-based similarity as per Section 4.4). Upon computing these distances we used them to create a hierarchical clustering of the alternatives. The resulting hierarchy tree can be visualized in a number of ways, including treemaps, dendrogram trees or radial trees, of which we choose an implementation of dendrogram trees that supports zooming and tree manipulations. The leaves in this tree are the design alternatives whereas the intermediate nodes are subclusters. Finally, not all aspects of a design can be quantified. It is then important for designers to have visual access to the design forms so that they could identify these qualitative features and compare the designs forms with each other. To support that, a grid of 2D/3D thumbnails is filled with the currently selected alternatives in the scatterplot or the dendrogram tree.

5.2.2. Interactions

The main design goal we have is to facilitate communication between the four views as they represent different aspects of the design data. In most of the cases, we have favored highlighting the other views when alternatives in one view are brushed. This is to maintain the context of these alternatives with respect to the unselected. A design decision was made to start the thumbnails grid with no form views and fill up the grid upon selection instead of filtering views down to the size of the selection. The latter approach was taken in DreamLens (Matejka et al., 2018). This enables an interaction whereas a rectangular brush can be moved on the scatterplot view (as in Figure 2, view B) gradually while updating the grid of thumbnails with the currently brushed alternatives only. This gradual movement demonstrates how geometric forms are changing in relation to changes in performance values. Finally, the dashboard currently supports form-first interactions in the form of highlighting alternatives in all data views when their thumbnails are hovered over.

6. Future Work and Conclusion

We had demonstrated a series of techniques and concepts relevant to exploring design spaces of computationally generated alternatives. These are derived by reflecting on and extending the literature of design space simplification and the patterns of their use. Finally, we presented the first iteration of a design analytics dashboard that instantiates some of the ideas discussed. In the future, we intend to continue iterating on and evaluating this dashboard and explore the practical and design-specific concerns surrounding similarity-based exploration.

References

"SmartGeometry" : 2018. Available from <www.smartgeometry.org/inside-the-black-box>.

Abi Akle, A., Minel, S. and Yannou, B.: 2017, Information visualization for selection in Design by Shopping, *Research in Engineering Design*, **28**(1), 99-117.

Beham, M., Herzner, W., Gröller, M.E. and Kehrer, J.: 2014, Cupid: Cluster-Based Exploration of Geometry Generators with Parallel Coordinates and Radial Trees, *IEEE Transactions on*

Visualization and Computer Graphics, **20**(12), 1693-1702.
Brown, N. and Mueller, C.: 2017a, Automated performance-based design space simplification for parametric structural design, *In Proceedings of IASS Annual Symposia*.
Brown, N. and Mueller, C.: 2017b, Designing with data: Moving beyond the design space catalog, *Proceedings Catalog of the 37th Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2017*, 154-163.
Bustos, B., Keim, D.A., Saupe, D., Schreck, T. and Vranić, D.V.: 2005, Feature-based similarity search in 3D object databases, *ACM Computing Surveys (CSUR)*, **37**(4), 345-387.
Chen, K.W., Janssen, P. and Schlueter, A.: 2015, Analysing Populations of Design Variants Using Clustering and Archetypal Analysis, *Real time-extending the reach of computation: eCAADe 2015; proceedings of the 33. International Conference on Education and Research in Computer Aided Architectural Design in Europe*, 251-260.
Erhan, H., Abuzuraiq, A., Zarei, M., Alsalman, O., Woodbury, R. and Dill, J.: 2020, What do design data say about your model? a case study on reliability and validity, *Proceedings of the 25th Int. Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA*.
Erhan, H., Wang, I.Y. and Shireen, N.: 2015, Harnessing Design Space: A Similarity-Based Exploration Method for Generative Design, *International Journal of Architectural Computing*, **13**(2), 217-236.
Fuchkina, E., Schneider, S., Bertel, S. and Osintseva, I.: 2018, Design Space Exploration Framework-A modular approach to flexibly explore large sets of design variants of parametric models within a single environment, *Proceedings of the 36th eCAADe Conference*, 367-376.
Harding, J. 2016, Dimensionality reduction for parametric design exploration, *in* F.G. Sigrid Adriaenssens (ed.), *Advances in Architectural Geometry 2016*, vdf Hochschulverlag AG.
Javed, W. and Elmqvist, N.: 2012, Exploring the design space of composite visualization, *2012 IEEE Pacific Visualization Symposium*, 1-8.
van Kastel, J.: 2018, *Visual Analytics for Generative Design Exploration: An interactive 3D data environment for a computational design system facilitating the performance-driven design process of a nearly Zero-Energy sports hall*, Master's Thesis, TUDelft.
Kohonen, T.: 1982, Self-organized formation of topologically correct feature maps, *Biological cybernetics*, **43**(1), 59-69.
Matejka, J., Glueck, M., Bradner, E., Hashemi, A., Grossman, T. and Fitzmaurice, G.: 2018, Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 369:1-369:12.
Mohiuddin, A., Ashtari, N. and Woodbury, R. 2018, Directly Interactive Design Gallery Systems: Interaction Terms and Concepts, *in* J.H. Lee (ed.), *Computational Studies on Cultural Variation and Heredity*, Springer, 15-32.
Nagy, D., Villaggi, L., Zhao, D. and Benjamin, D.: 2017, Beyond Heuristics: A Novel Design Space Model for Generative Space Planning in Architecture, *Proceedings of the 37th Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2017*.
Pan, W., Sun, Y., Turrin, M., Louter, C. and Sariyildiz, I.: 2019, Design exploration of architectural geometries and structural performance for sports arenas based on SOM-clustering and structural performance simulation, *Proceedings of the Fourth International Conference on Structures and Architecture (ICSA 2019)*, 342.
Shi, X.: 2010, Performance-based and performance-driven architectural design and optimization, *Frontiers of Architecture and Civil Engineering in China*, **4**(4), 512-518.
Shneiderman, B.: 1996, The eyes have it: A task by data type taxonomy for information visualizations, *Proceedings 1996 IEEE symposium on visual languages*, 336-343.
Thomas, J. and Cook, K.: 2006, A Visual Analytics Agenda, *IEEE Cmpt. Graph. Appl.*, **26**, 3.
Thornton Tomasetti, C.O.R.E. s.t.u.d.i.o.: 2019, "Design Explorer" . Available from <https://github.com/tt-acm/DesignExplorer>.