

**CARNEGIE MELLON UNIVERSITY**

**School of Architecture**  
College of Fine Arts

**Thesis**

Submitted in Partial Fulfillment of the requirements for the degree of

**Master of Science in Computational Design**

TITLE:

**A Topology-Based Architectural Search Engine**

AUTHOR:

**Yuqian Li**

**ACCEPTED BY ADVISORY COMMITTEE:**

Daniel Cardoso Llach

\_\_\_\_\_  
Professor NAME

\_\_\_\_\_  
Principal Advisor

\_\_\_\_\_  
DATE

# A Topology-Based Architectural Search Engine

by

Yuqian Li

Submitted to the School of Architecture  
on May 18, 2018, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Computational Design

## Abstract

Architectural design involves the study and manipulation of complex spatial hierarchies and relations. Current search tools such as Google Image and websites like Pinterest are based on text-input from users, or use algorithms to determine visual similarity, and thus their results are not useful to explore spatial relations. There is thus a gap between the logic of raster or tag based search engines and the demands of architectural users. My thesis helps narrow down this gap by introducing a novel architectural search engine based on spatial relations. The system offers users a visual tool to depict the spatial structure of an architectural plan, which it uses to generate a graph of the building. The system then uses this information to quantify the degree of spatial similarity of the building in comparison with examples in a database, using a belief propagation algorithm. The thesis discusses the potential of this system to elicit new ways of searching architectural information and to scale into an open platform where users add new diagrams of existing and new designs.

Thesis Supervisor: Daniel Cardoso Llach  
Title: Assistant Professor



# Acknowledgments

A lot of thanks to my advisor, Professor Daniel Cardoso Llach for always understanding what I was trying to express and guided me when I was getting into troubles, and for opening a new door for me to learn new knowledge, to understand architecture in computational way, to form a new vision of architecture design process, also to see new possibility of my career further.

Thank you to all the friends that helped me during this thesis process, I feel like I grow up fast from someone who is so frustrated about what to do to now, someone who is more comfortable to consider a problem in computational way. Thanks Quan Sun, to help me collect data, and guide me when I was totally lost in the complicated mathematic papers. Thanks all my CS friends, you guys are awesome! I always bother every one of you to discuss the possibility algorithms that can be used in my projects.

Thanks to Code Lab guys, your great work also inspired me to be more intelligent and diligent to pursue my research.

Many thanks to my friends Kang Gao, for always supporting me when I feel exhausted.



# Contents

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	Thesis Outline . . . . .	9
1.2	Research Questions and Goals . . . . .	11
<b>2</b>	<b>Background</b>	<b>13</b>
2.1	Computer Representations of Architectural Problems . . . . .	14
2.2	Graph Theory in Architecture . . . . .	15
2.3	Planar Graphs and Relations . . . . .	16
2.4	Graphic Representation of Architecture Layouts . . . . .	17
<b>3</b>	<b>Architecture Data and Analysis Methods</b>	<b>19</b>
3.1	Analysis Methods . . . . .	19
3.2	Graph Model . . . . .	20
3.3	Belief Propagation . . . . .	21
3.3.1	PC Dataset . . . . .	23
3.3.2	My Customized Dataset . . . . .	24
3.3.3	Belief Propagation for Graph Similarity . . . . .	25
3.4	Workflow . . . . .	27
3.5	Compare Methods . . . . .	29
<b>4</b>	<b>Test and Evaluation</b>	<b>35</b>
4.1	Accuracy of the Algorithm . . . . .	35
4.2	Self Similar . . . . .	35

<b>5</b>	<b>Discussion</b>	<b>41</b>
<b>6</b>	<b>Limitation</b>	<b>51</b>
6.1	Limitation of Current Algorithm . . . . .	51
6.2	Limitation of Graph Representations . . . . .	52
<b>7</b>	<b>Conclusion</b>	<b>53</b>
7.1	Research Results and Contribution . . . . .	53
7.2	Future Possibility . . . . .	54
<b>A</b>	<b>Matrix of all the analysis results</b>	<b>59</b>

# List of Figures

2-1	Christopher philosophy of design process with recursion tree. . . . .	14
2-2	Intrinsic structures in FORTRAN [1] (left) Part of the LISP representation [1](right). . . . .	15
2-3	The Konigsberg bridge problems with Euler's graphic representation[3].	16
2-4	The floor plan of the house and its corresponding diagram[4]. . . . .	17
2-5	One adjacency graph and its possible plans. . . . .	18
3-1	For two architecture models, Room A share the same adjacent list but have different hierarchy. . . . .	21
3-2	Messages send to root . . . . .	22
3-3	PhoneCall graphic model illustration. . . . .	24
3-4	Room adjacency graph diagram. . . . .	25
3-5	The matrix for room adjacency graph. . . . .	26
3-6	Each different shade of green represents a different level of accessibility.	28
3-7	Matrix that convert from graph. . . . .	29
3-8	GUI interface. . . . .	30
3-9	Adding nodes and edges on the canvas. . . . .	31
3-10	Searching results. . . . .	32
3-11	Compare methods matrix. . . . .	33
4-1	Insert 2 new nodes and remove 2 original ones. . . . .	37
4-2	Arch1 and Arch4 with a similarity of 0.8 . . . . .	38
4-3	Arch1 and Arch18 with a similarity of 0.5 . . . . .	38
4-4	Arch1 and Arch19 with a similarity of 0.2 . . . . .	39

5-1	Their work is similar to bubble charts. . . . .	42
5-2	Compare result of Rolex Learning Center . . . . .	44
5-3	Compare result of Grace Farms Center . . . . .	45
5-4	Compare result of Glass Pavilion . . . . .	46
5-5	Compare result of 21st Century Museum of Contemporary Art . . . .	47
5-6	Compare result of Moriyama House . . . . .	48
5-7	Compare result of Towada Art Center . . . . .	49
5-8	Mariyama House . . . . .	50
5-9	Arch1 . . . . .	50
6-1	Architecture with different dimensions but share topological similarity to some extent . . . . .	52
A-1	. . . . .	60
A-2	. . . . .	61
A-3	. . . . .	62
A-4	. . . . .	63
A-5	. . . . .	64
A-6	. . . . .	65
A-7	. . . . .	66
A-8	. . . . .	67
A-9	. . . . .	68
A-10	. . . . .	69
A-11	. . . . .	70
A-12	. . . . .	71
A-13	. . . . .	72

# Chapter 1

## Introduction

### 1.1 Thesis Outline

#### Chapter 1: Introduction

This chapter briefly talks about problems in the traditional process of architectural initial design in the industry, and two possible reasons of why architectural design is hard to understand. On the one hand, architects tend to use abstract words to describe architecture. On the other, current mainstream search engines are not suitable to search architecture. Taking those two aspects together, one underlying research question is "Can we use spatial adjacent graph to analyze architecture topological similarity?"

#### Chapter 2: Background

Background introduces the classic theory of Alexander Christopher about diagrammatically parse architecture, which serves as an underlying theory of the following related work. A literature review to discuss the development of applying computational thinking and graph theory in architectural design field. In this chapter, the related literature discussion is divided into 4 sections, which were given to show the history and the transition from thinking architecture in a computational way to applying theory of graph to represent and solve architectural puzzles.

#### Chapter 3: Architecture Data and Analysis Methods

This chapter explores methods that are used in this project. Four following hypothe-

ses are raised up based on the tasks. Additionally, this chapter gives a detailed description about how the system works by describing the workflow including way to set up experiments, collect data, conduct tests and so on. It also mentions ways to set up model and a few pseudo codes to help understand. The last part displays the system interface and a workflow with instructions.

#### Chapter 4: Test and Evaluation

These two sections introduce two tests this thesis conducted to evaluate the accuracy of the system. By tracing the graphic representation of an architecture project, the test aims to see if the retrieve result will contain this project. In the second test, the existing graph diagram is modified by deleting two nodes, and then insert two random new nodes to test if the target project can still be searched out. It turns out that the system works for both cases, but the similarity results vary in an unfixed range. It is due to the complexity of the input graph, which means if the graph is complicated enough, it is easier to find out itself and get a higher similarity value, whereas if the input graph is common and general, other projects may become more similar than itself.

#### Chapter 5: Discussion

Discussion chapter illustrates the results of the last three hypotheses that are raised in the first chapter. It denotes that most of the hypotheses are correct to a great extent. Additionally, beyond the hypotheses, these tests also tried to find interesting design patterns, and compare different design methods over numbers of projects.

#### Chapter 6: Limitations

This chapter summarizes the current limitations of graph-based representations and the limitations of vector-based algorithms to compare similarity. The first hypothesis that brought up in Chapter one is also discussed here.

#### Chapter 7: Conclusion and Future Work

This chapter summarizes the thesis contributions as well as several next steps. To build up a more comprehensive graph model, this chapter gives an envision of future possibility and the reason why it is important.

## 1.2 Research Questions and Goals

In a traditional architectural design process, architects brainstorm, collect similar cases and draw inspiration from them. When architects collecting cases, spatial relations turn to be one of main concerns. Spatial layout is one of the most distinctive features that distinguish architecture as an art work from architecture as a science

<sup>1</sup> This is largely due to the various spatial relations that differentiate architecture projects topologically rather than morphologically may tend to imply more design intentions. For example, SANAA's work tend to be morphologically dissimilar while consistent in topology. Therefore, it would be hard to tell the design concepts of SANAA's work if people merely look into forms, since most of them are known as extremely free of shapes. Interestingly, architects also seem to have their own languages to express their feelings or opinions about architecture, which make architecture and architecture design process even harder to understand. The following quote from Takizawa captured the characteristic of one of the SANAA's architecture is like this:

*"A Sanaa building is complete only if inhabited. When we visited Kanazawa three months before it opened, it was so empty, I didn't feel any emotion," he says. "It was so cold. We just felt the material and the function. After the opening, it was completely changed, when I can concentrate on the pieces of art."* <sup>2</sup>.

It turns out that architecture work is hard to be understood by people who don't have architecture related backgrounds. As a result, the current architecture examples are always grouped or categorized by varied shapes instead of topological relationships. Correspondingly, since shapes tend to be too attractive at the first sight, many architecture design firms began to emphasize the importance of shapes and forms, which not only potentially leave troubles to subsequent spatial arrangements but also blur people's understanding of architectural design.

On the other hand, nowadays, most of the mainstream search engines such as Pinterest and Google, are based on texts as well as images. However, architectural

---

<sup>1</sup>Bill, H. and Adrian, L. The architecture of architecture. Models and Systems in Architecture and Building. The Construction Press Ltd, 1975.

<sup>2</sup>Lubow, A. (2005, OCT 9).Disappearing Act.  
Retrieved from <https://www.nytimes.com/2005/10/09/style/tmagazine/disappearing-act.html>

design work is not suitable to be searched by these, since it also contains spatial and visual sequences which are hard to describe in words and pixels. In this sense, there is a massive gap between the logic of existing tools, either raster-based search engine or the tag-based picture gallery and users to understand architectural information.

Much of the often-cited literature on architectural topology focuses on the theoretical level. I believe the primary benefit in this case is to help readers to better understand architecture from a heuristic perspective. However, since the use of graph theory is well established in solving various problems in all kinds of fields, my thesis looks into the graph representation of architectural topological issues by building a software with java which serves as a searching tool that allow users to input graphic representation of architecture cases. This workflow tends to create a new approach toward architecture searching process which not only devote to matching accurate retrievals, but also to generating a new perspective on architecture analysis.

Taking these two aspects together, an underlying question in this research is, "Can we use spatial adjacent graph to analyze architecture topological similarity?" This system tends to find a different perspective to analyze and understand architecture. as well as deeply explore the possibility of using graph-theoretic representation of architectural floor plans arrangement as an effective approach to search architecture.

# Chapter 2

## Background

A similar idea of understanding architecture by topological relations can be traced back to 1971, Christopher Alexander. In his PhD thesis *Notes on the Synthesis of Form*, he pointed out that design is a process from context to form. By this he means, ideally, if we can give a precise description to context, then we could also gain a best boundary corresponding to the form. This process provides an ideal design process. Contrasted with traditional design, Alexander parse design process into rules and procedures. Figure 2-1 shows Alexander's philosophy of resolving design problems. An entire village can be divided into A, B, C, D, 4 parts, and for each section, requirements can be further specified. This idea is actually similar to the idea of recursion in mathematics and computer science, where a condition being defined by its own definition.<sup>1</sup>Hence, design can be understood and mastered by machine, in other words, computer is able to analyze and generalize design work.

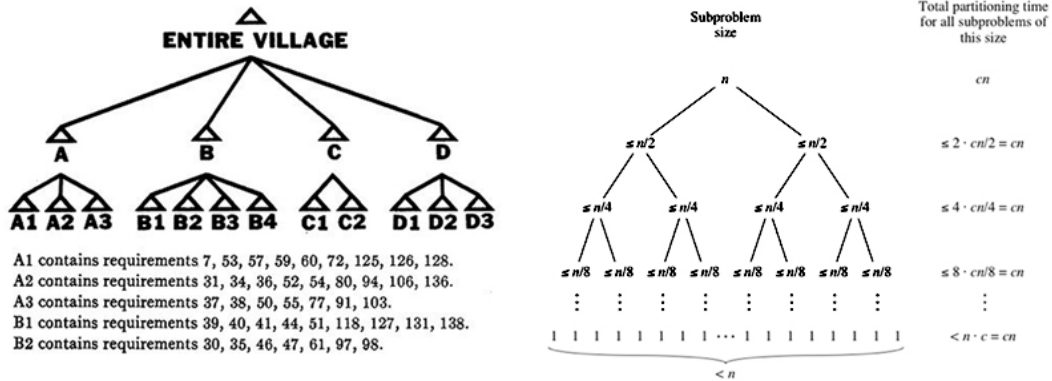
This became an underlying theory that influenced many subsequent researchers to explore ways of creating mathematical models of designed objects and their contexts.<sup>2</sup>

---

<sup>1</sup>Recursion. (27 March, 2018) In Wikipedia, the free encyclopedia. Retrieved April 26, 2018, from <https://en.wikipedia.org/wiki/Recursion>

<sup>2</sup>Steadman, P. Models in our heads, models in the material world, and models in the world of objective knowledge. *Models and Systems in Architecture and Building*. The Construction Press Ltd, 1975

Figure 2-1: Christopher philosophy of design process with recursion tree.



## 2.1 Computer Representations of Architectural Problems

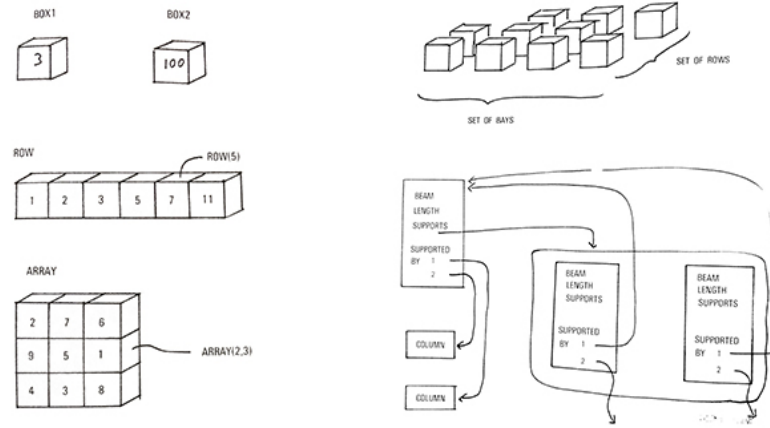
A classical work from Janet Tomlinson introduced a corresponding relationship between data structures and computer representation of architectural problems. Here architectural problems are mainly referring to the comprehensive representations of architecture as an object and the relations between them. Architecture can be classified by different layout patterns and architectural types, identically, programming languages also differentiate in terms of underneath data structures. On this subject, he proposed that the language where architectural information is encoded takes an important role, since if there is any kind of data that cannot be expressed then, that information cannot be encoded in the model. He also enumerated a series of examples to present different architecture cases and their most suitable programming languages. In case of Fortran, the underlying data structure is a collection of boxes, which can also be put into array or nested list. Another example, LISP, one of the earliest object-oriented programming language<sup>3</sup>, were proposed to be applied in the

<sup>3</sup>LISP. (27 March, 2018) In Wikipedia, the free encyclopedia. Retrieved April 26, 2018, from <https://en.wikipedia.org/wiki/Lisp>

layout of a town floor plan, as figure 2-2 displays.

Janet Tomlinson also mentioned that if the programming language itself can also serve as a tool for thinking, it would be more meaningful in solving architectural problems. It turns out that most of the problems she mentioned can nowadays be solved by modern OOP programming languages, but as an earlier approach, it is still interesting to see how previous scholars tried to break down architecture problems.

Figure 2-2: Intrinsic structures in FORTRAN [1] (left) Part of the LISP representation [1](right).



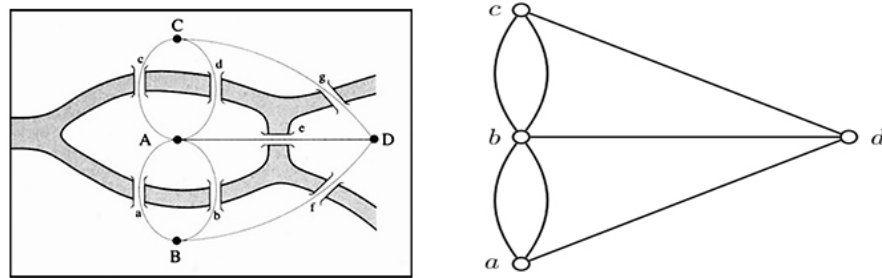
## 2.2 Graph Theory in Architecture

The first application of graph theory was the historically notable mathematic problem - Seven Bridges of Konigsberg. The reason why graph is trustworthy here lies in its ability to present the intrinsic structure of a set of relationships.<sup>4</sup> The problem was set to cross the seven bridges in the city without stepping on any of the bridges twice. Euler, used graph diagram to abstract the bridge and lands into graph model and then prove that there is no solution to this problem. In the Konigsberg problem, in order to get back to the starting point, we must go through each city the same

<sup>4</sup>March, L., Steadman, P. The geometry of environment: An Introduction to Spatial Organization in Design, pages 242-243. The MIT Press, 1971.

number of times. Hence, to come and leave each vertex without crossing the edges twice, each vertex needs to be connected by even number of edges, which is not the case of Königsberg.

Figure 2-3: The Königsberg bridge problems with Euler's graphic representation[3].



## 2.3 Planar Graphs and Relations

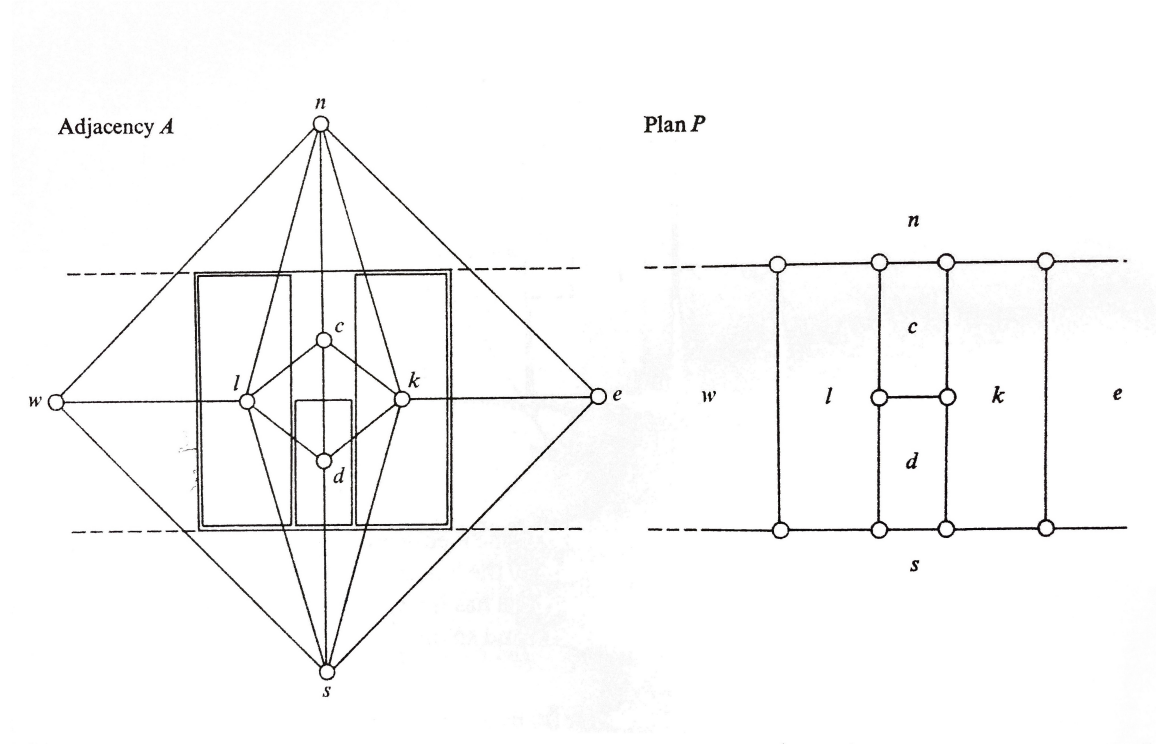
Under the inspiration of Königsberg bridge problems, a more architecture related problems that solved by graph theory is *"in the design of the floor plan of a tiny terrace house."* *"Assume that for the first floor of this house, there is a kitchen  $k$ , a dining room  $d$  and a living room  $l$  as well as some circulation space  $c$  which may refer to a hall or an aisle."*<sup>5</sup> As a common design procedure, they specified requirements and constraints to the floor plan, which in this case, are all in respect of adjacency. For instance, the living room is to be next to the dining room, and the dining room is adjacent to the kitchen<sup>6</sup> In his context, "adjacency" means that two rooms must

<sup>5</sup>March, L., Steadman, P. The geometry of environment: An Introduction to Spatial Organization in Design, pages 242-243. The MIT Press, 1971.

<sup>6</sup>Steadman, P. Models in our heads, models in the material world, and models in the world of objective knowledge. Models and Systems in Architecture and Building. The Construction Press Ltd, 1975

border on, or there are some partitions that shared in common. Then they are able to generate a graph that fulfill all the requirement.

Figure 2-4: The floor plan of the house and its corresponding diagram[4].



In addition, this paper also devoted much space to discussing how to easily amend and display architectural graph models in a single graph, which is able to solve by using force-directed graph algorithm now. The JAVA library that I used to visualize graph diagram also applied this algorithm, which will be mentioned in the implementation chapter.

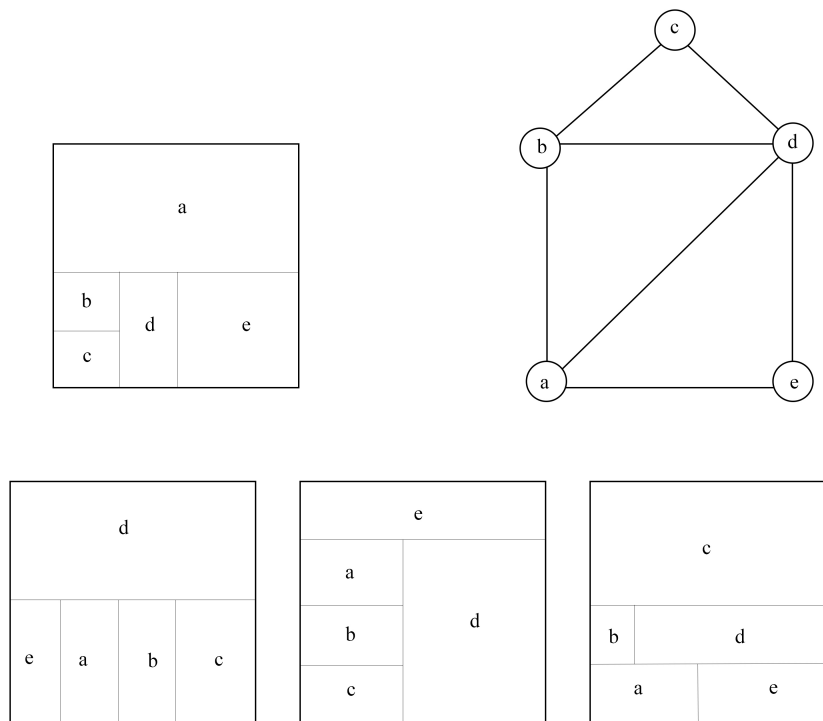
## 2.4 Graphic Representation of Architecture Layouts

In 1976, Philip Steadman in the paper Graph-theoretic representation of architectural arrangement<sup>7</sup> emphasized the feasibility of using graphs to represent the relation of

<sup>7</sup>Steadman, P. Graph-theoretic representation of architectural arrangement In L. March (Ed.). The Architecture of Form: Cambridge Urban and Architectural Studies. Cambridge: Cambridge

rooms. He elaborately introduced various approaches to represent floor plans with different attributes, such as ways to represent varied plans of the same adjacent graph and plans with different geometry, as figure 2-5 shows. Most importantly, he addressed two different design methods in producing architectural plans by graphic representations, which implies a completely different philosophy about what kind of position should computer occupy in architectural research and practice <sup>8</sup>.

Figure 2-5: One adjacency graph and its possible plans.




---

University Press, 1976.

<sup>8</sup>Steadman, P. Graph-theoretic representation of architectural arrangement In L. March (Ed.). The Architecture of Form: Cambridge Urban and Architectural Studies. Cambridge: Cambridge University Press, 1976.

# Chapter 3

## Architecture Data and Analysis Methods

### 3.1 Analysis Methods

The research approach is to design a software that allow users to draw graphic diagrams on the interface by clicking a series of room type buttons, for example, the bedroom, the living room, and then connect with edges to the rooms they want to be adjacent. The program will then compare similarity with all the topological graphs in my database and return the graph representation and the first-floor plans of the most similar 6 projects. In order to evaluate the system, below are a few hypotheses that addressed in this thesis:

H1: For architecture cases that emphasizes the importance of ambiguous space or free-flowing Space, such as Farnsworth and Germany Pavilion, they are also different from the most common housing topologically.

H2: Projects from the same architect/ architecture company are always under the influence of same kinds of topological relations.

H3: There is a specific kind of topology relations that occurs for most of the time as well as another kind of topological relations that never occurs.

H4: For the work from some of the famous architects, their projects show a certain kind of similarity in a certain range.

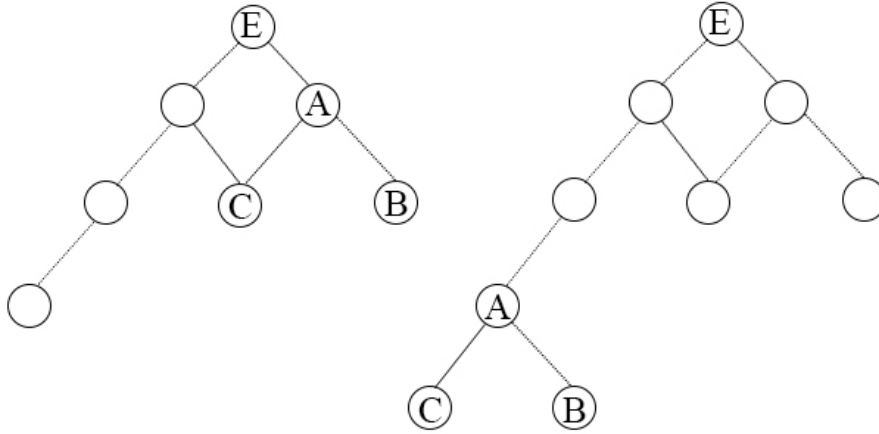
These four hypotheses are also interesting topics that deserves for research in the study of architecture topology.

## 3.2 Graph Model

For a classic graph searching problem, graph models can be further divided into undirected graph, bidirectional graph, acyclic graph and cycling graph. Through observations, it is found out that most of the cases in the dataset can be abstracted by polytree, a directed graph without any undirected cycles. In real world, architecture topological relations are supposed to be bidirectional, since Room A connects to Room B is identical to Room B connects to Room A. However, the problem of mapping spatial arrangement into this graph model lies in the loss of spatial sequences. For example, for a given spatial arrangement, Entrance  $\rightarrow$  Aisle  $\rightarrow$  Bedroom and Courtyard, as figure 3-1 shows, if we use an undirected graph to store the information, then we will lose the information to compare the different hierarchy of aisle(A) between these two graphs, since we cannot calculate how many nodes are below the current room.

Therefore, directed graph is applied to present the spatial sequence, I believe that the design logic of an architecture is related to the circulation from entrance to the others, since there is no architecture that starts from bedrooms or offices. Besides, since there could be multiple routines toward a certain destination, which involve the classical shortest path problems. To avoid running into unsolvable questions, the sequence of my directed graph is based on the circulation of my traverse an architecture project. This method can embody the differences between a public kitchen and a closed kitchen at the end of the aisle, which takes an important role in architecture spatial analysis.

Figure 3-1: For two architecture models, Room A share the same adjacent list but have different hierarchy.



### 3.3 Belief Propagation

In order to compare topological similarity across different examples of architecture, the main concern in this thesis is to compare room arrangements and find the most similar ones. Previous research analyzed the connection of rooms with graph theory<sup>1</sup>, which abstract and represent architecture by the adjacent relationships among different rooms. However, this method didn't involve the comparison of architecture or graphs. In addition, there is no effort has been expended on quantitative analysis with sufficient samples. I applied belief propagation algorithm to graph theory together with my customized architecture dataset as an application to graph theory as well as a compare method toward topological similarity.

The logic of graph similarity is that "a node in one graph is similar to a node in another graph if their neighborhoods are similar"<sup>2</sup> In other words, one room in the first architecture is similar to another room in the second architecture if their

<sup>1</sup>Bill, H. and Adrian, L.The architecture of architecture.Models and Systems in Architecture and Building.The Construction Press Ltd, 1975.

<sup>2</sup>Lubow, A.(2005, OCT 9).Disappearing Act.  
Retrieved from <https://www.nytimes.com/2005/10/09/style/tmagazine/disappearing-act.html>

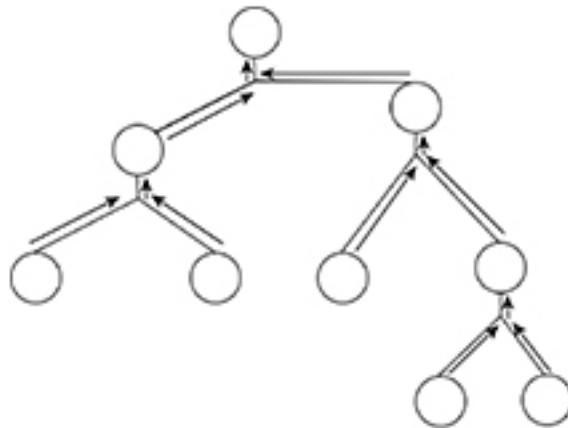
neighboring rooms are similar.

One of the key attributes of graph is the "score-passing" between connected nodes. To pass the scores between two nodes (rooms), I will briefly introduce an algorithm called Belief Propagation.

Belief propagation algorithm (BP) is a message passing algorithm for performing inference on graphical models <sup>3</sup>. It has been applied to various datasets to analyze systems that contain thousands of variables. Basically, there are three steps to implement:

1. Pick any node as the root of the graph.
2. Send messages from leaves to the root.
3. Send messages from root to the leaves.

Figure 3-2: Messages send to root



J.Pearl<sup>4</sup> proposed that, in order to calculate the conditional marginal probability<sup>5</sup>,

---

<sup>3</sup>Steadman, P. Models in our heads, models in the material world, and models in the world of objective knowledge. Models and Systems in Architecture and Building. The Construction Press Ltd, 1975

<sup>4</sup>Steadman, P. Graph-theoretic representation of architectural arrangement In L. March (Ed.). The Architecture of Form: Cambridge Urban and Architectural Studies. Cambridge: Cambridge University Press, 1976.

<sup>5</sup>Tomlinson, J. Computer representations of architectural problems. Models and Systems in Architecture and Building. The Construction Press Ltd, 1975.

this algorithm is able to build an inference network which has a huge number of variables to impact what happens, but each variable is only related to a few others. This network is close to architecture hierarchy, where architecture is affected by all rooms but each of the individual room is only related to a few of their adjacent rooms.

Based on this network, BP transferred the sum up of the entire graph into information that passing locally(nodes). In other words, each node in the network can assess its own probability by exchanging information with its adjacent nodes.

The original intention of this approach is to decrease time complexity for complicated and gigantic datasets and recently have been increasingly used to solve the problems of machine learning in big datasets. However, BP network could also be used to simulate a real-world situation, including architecture. Danai and his colleagues <sup>6</sup>applied this algorithm to PhoneCall and social network graph datasets. In both situations, BP has been proven successful and effective. I will briefly introduce PhoneCall example in the following section.

### 3.3.1 PC Dataset

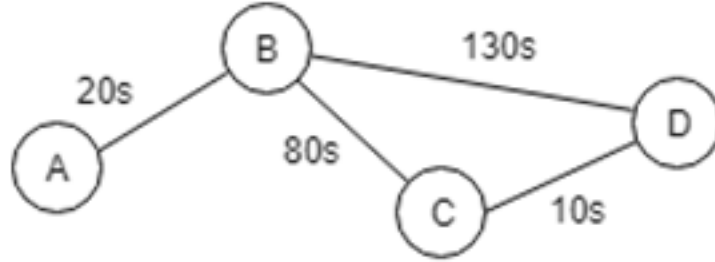
In this graphic model, the nodes represent people. If two people spoke to each other, then they will connect to each other with an edge, meanwhile, the total call duration would be the weight on that two edges. In total, the dataset consists of more than 30 thousand people in one single city using one cellphone operator. Each graph is the total of calls that took place within 24 hours. The goal is to compare if any two days of a week are similar in terms of a phone call made.

As I mentioned previously, the way to measure similarity is to evaluate if two nodes and their neighbors are similar. In this sense node B for example, is neighbored by A, C, D. After comparison, B will receive a number represents for how similar it is to another B nodes in the second graph. The average number of all nodes would be the Graph similarity. Graph similarity determines the extent of how similar two graphs are, which is a number between 0 and 1.

---

<sup>6</sup>Lubow, A.(2005, OCT 9).Disappearing Act.  
Retrieved from <https://www.nytimes.com/2005/10/09/style/tmagazine/disappearing-act.html>

Figure 3-3: PhoneCall graphic model illustration.

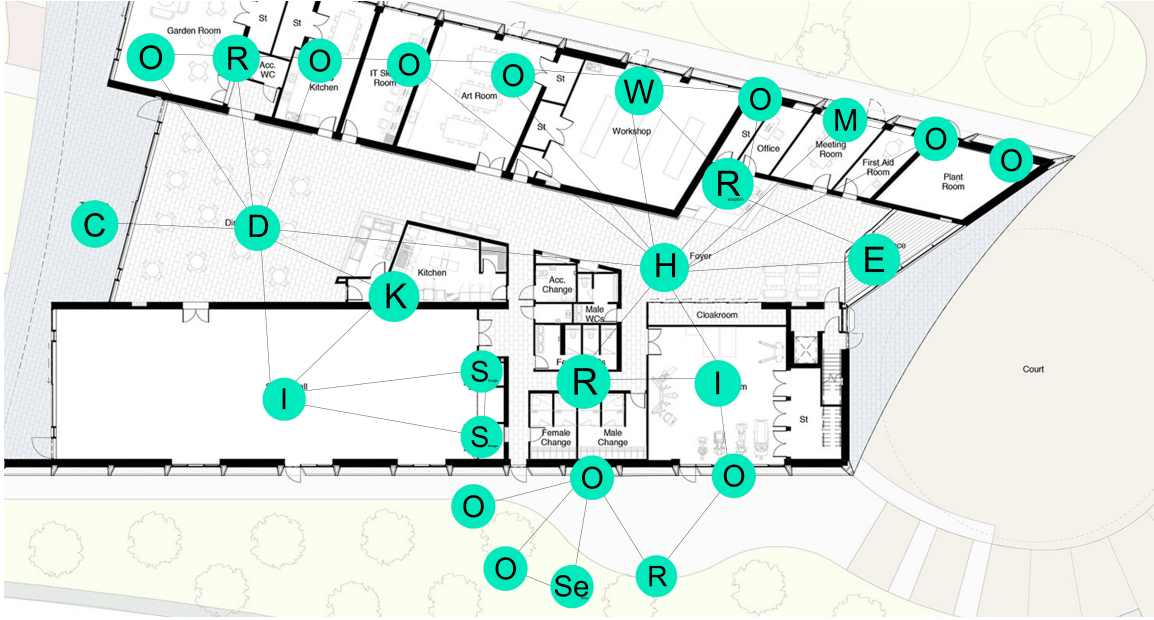


### 3.3.2 My Customized Dataset

In my case, I collect 100 architecture samples from ArchDaily, goood and other architectural websites. The criteria for choosing samples is 50 percentage being resident buildings, 30 percentage being public buildings such as art center, museum and restaurant, and another 20 percentage being high-rise buildings, and then convert them into an abstract graphic representation of spatial relationships. (As figure 3-4 shows). Each node represents a room type occurred in the building. Considering the complexity of room types, I group room types into a series of standard room types, which are "entrance, living room, bedroom, dining room, kitchen, study room, restroom, courtyard, reception, service room, storage, interactive, exhibition, hall, gallery, meeting room, workshop, office and studio." If two nodes are connected, they will be connected by an edge. This graph can be further converted into a 2-D matrix and be encoded into program.

To some extent, architecture is similar to PhoneCall networks and social networks. Since BP algorithm has been proven effective to quantify, compare PhoneCall networks and social networks, it worths attempting to apply BP algorithm to compare and measure architectural topological similarity.

Figure 3-4: Room adjacency graph diagram.



### 3.3.3 Belief Propagation for Graph Similarity

Graph similarity has been applied in various fields with a number of algorithms being raised up. In the original algorithm, each node conveys messages in form of matrix, prior belief is the sum production of adjacent nodes. Danai and his colleagues<sup>7</sup> applied BP based algorithm that is proposed by Yedidia in<sup>8</sup> to calculate graph similarity in various dataset. Based on their work, the setting for convey and update final belief in this thesis is as follows:

```

for i = 1, i < n, i++ do
#initialize node's i prior belief to p
prior belief = probability of occurrence
#get the bi1 and bi2 vectors of final beliefs
final belief = similarity measure bi1, bi2
similarity score i prior belief * final belief

```

<sup>7</sup>Lubow, A.(2005, OCT 9).Disappearing Act.  
Retrieved from <https://www.nytimes.com/2005/10/09/style/tmagazine/disappearing-act.html>

<sup>8</sup>Danai K., Ankur P., Aaditya R., Jing X. Algorithms for Graph Similarity and Subgraph Matching, 2011.

Figure 3-5: The matrix for room adjacency graph.

	A	B	C	D	E	F	G
1	entrance	public	entrance	2			
2	reception	public	reception	3	4	5	11
3	café	public	dining	10			
4	hall	public	hall	7			
5	exhibition space	public	exhibition	8			
6	play area	public	interactive				
7	restroom	private	restroom				
8	artists lodges	private	gallery	9			
9	service	private	service room				
10	open air café	public	dining	12			
11	exhibition hall	public	exhibition	12			
12	office	public	office				

end for

similarity of graphs = avg similarity score

The first step is to find two identical rooms in the graphs. For example, bedroom in architecture A versus bedroom in architecture B. The next step is calculating the prior belief of the two rooms. Prior belief in belief propagation is defined as the input information from the father node. It is the reflection of the room hierarchy in the architecture. For example, if a terrace is located next to the entrance and connects to most of other rooms in the architecture, it gets a high score of prior belief. If a terrace is only connected to a bedroom and serves as a balcony, it gets a lower score of prior belief. Most of the time, if a room is close to the major entrance, it would have a higher score of prior belief. If two rooms have similar prior belief, they would be considered comparable and receive a higher score of similarity. Prior belief score would be a number between 0 to 1.

With the prior belief, the second step is to get final beliefs for both nodes. In this thesis, final belief is calculated by the similarity score of its adjacency list. If two rooms connect to similar room list, they would be considered as similar. Thus, they would get a higher score for final belief. The final belief score would also be a number between 0 to 1. There are several approaches to calculate the score, section 5 will discuss more.

Then the most important step is to measure the similarity with the prior belief and the final belief. This thesis explores various ways of evaluating the similarity of the vectors, the ultimate goal is to get a number between 0 and 1, where 0 means completely dissimilar, meanwhile 1 means identical<sup>9</sup>. Since both prior belief and final belief are positive correlated to final similarity, the final similarity score would be calculated as the product of prior belief and final belief. The similarity of two architectural projects would be calculated as the mean value of the similarity score for each individual room.

## 3.4 Workflow

### Setup

Each project in the customized dataset has floor plans and sections. The way this thesis converted plans into graphic diagram follows Steadman's<sup>10</sup> setting for a small plan with 5 or 10 rooms only and a larger plan that contains small distinct sections, as figure 3-6 shown. For a classic directed graph traversal problem, circulation takes an important role, since for a single node, taking different paths will affect the adjacent matrix. In architecture design, the concept of circulation is similar, which refers to the way people move through space.<sup>11</sup> This idea can reference to the book *Architecture: Form, Space, and Order*, where Ching defined circulation as the movement through spaces<sup>12</sup>, which is not necessary to take everyone's flow into account. Instead, we often approximate the main routes of the majority of users. Hence, the logic of tracing in this project is based on my circulation routes to a building. Entrance is treated as the root of graph, as figure 11 shows, each different shade of green represents a different level of accessibility. The darkest green stands for root, while the lighter one

---

<sup>9</sup>Lubow, A. (2005, OCT 9). Disappearing Act.

Retrieved from <https://www.nytimes.com/2005/10/09/style/tmagazine/disappearing-act.html>

<sup>10</sup>Steadman, P. Graph-theoretic representation of architectural arrangement In L. March (Ed.). *The Architecture of Form: Cambridge Urban and Architectural Studies*. Cambridge: Cambridge University Press, 1976.

<sup>11</sup>In Portico.space. Retrieved April 29, 2018, from <http://portico.space/journal//architectural-concepts-circulation>

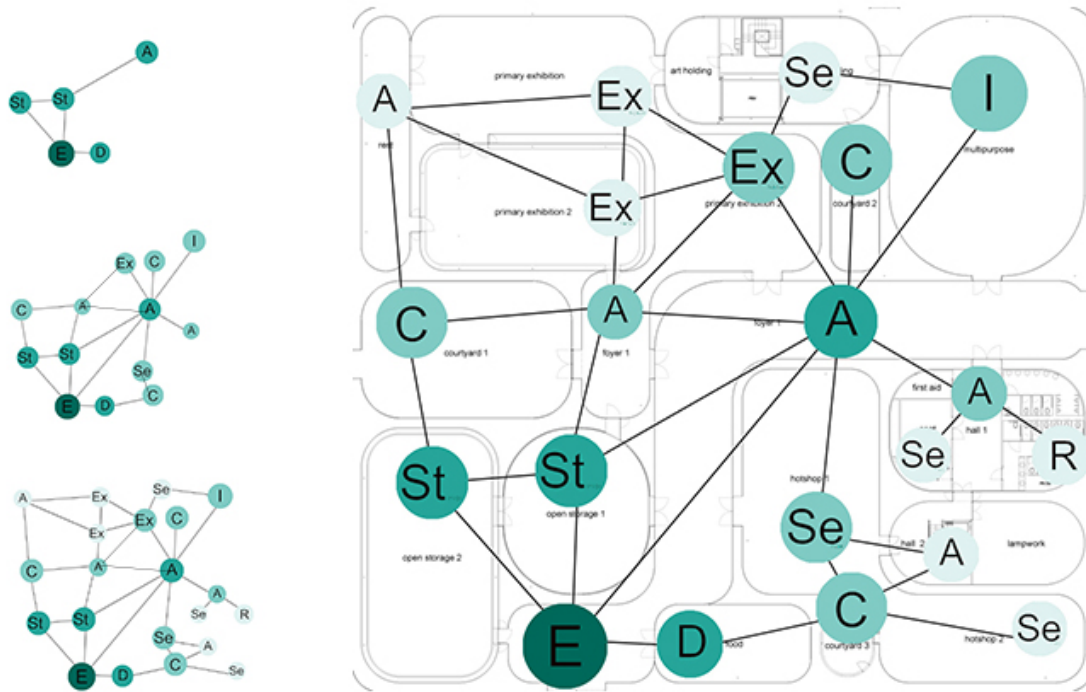
<sup>12</sup>Ching, F. *Architecture: Form, Space, and Order*. October 2014

represents rooms that connected to entrance and so on.

In case of multiple floor plans, circulation spaces are ignored. The implication behind the adjacency analysis is that two rooms that are next to each other, does not exactly allow for direct access. Besides, there is no comparative weight or value that is attached to these connections, which means each room is equivalent important to the architecture.

### Convert to matrix

Figure 3-6: Each different shade of green represents a different level of accessibility.



The next step is to convert graphic diagram into textual representation. The matrix is following the hierarchy of the graph. In figure 3-7, column A represents the original room name that describe on the floor plan. Column B implies the property of this room. To better organize and classify, column C is the standard name for column A. All the columns below, are index number of the current room's adjacent rooms. If two rooms adjacent with each other, only the front room will get the index, for example Room 1 and Room2, room 1's adjacent room index is 2, whereas Room 2 is none. In other word, indexes are in an ascending order with no overlap.

## Tracing the interface

Figure 3-7: Matrix that convert from graph.

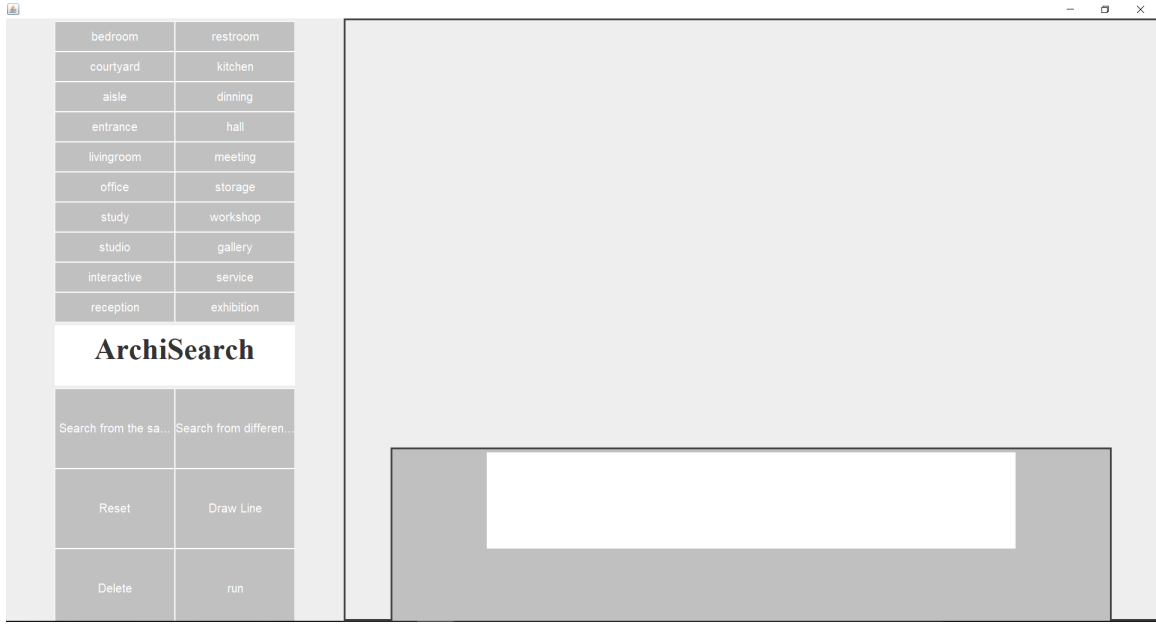
	A	B	C	D	E	F	G	H	I	J	K	L
1	Entrance	public	entrance	2								
2	Foyer	public	hall	3	4	5	6	7	9	10	20	22
3	Dining Area	public	dinning	12	15	16	17	21	29			
4	Reception	public	reception	8								
5	Plant Room	Private	service room	6								
6	First Aid Room	Private	office	7								
7	Meeting Room	Private	meeting	8								
8	Office	Private	office	9								
9	Workshop	public	workshop	10								
10	Artroom	public	office	11								
11	IT Skills Room	public	office	12								
12	Living Skills Kitchen	public	service room	13	15							
13	StorageL1	Private	storage	14	15	16						
14	StorageR1	Private	storage									
15	Restroom1	private	restroom	16								
16	Garden Room	public	service room									
17	Sports Hall	public	interactive	18	19	21						
18	StorageT	Private	storage	20								
19	StorageD	Private	storage	20								
20	Restroom2	Private	restroom	22								
21	Kitchen	Private	kitchen									
22	Gymnasium	public	interactive	23								
23	Staffroom	Private	office	24	25							
24	Restroom	Private	office	25								
25	Staff office	Private	office	26	27	28						
26	Managers office	Private	office	27	28							
27	Server room	Private	office	28								
28	Services well	Private	service room									
29	terrace	public	courtyard									

As Figure 3-8, 3-9 illustrates, users can click and draw nodes and edges on the right part of the canvas. Then the system will take 0.1s to calculate and retrieve the searching results back to them. The highest six similar projects will be displayed in a descending order together with their graph representations. The texts describe the range of the similarity value.

## 3.5 Compare Methods

As the last line of the pseudo code illustrates, in order to finally implement belief propagation, a suitable algorithm for the dataset is essential. The current chosen compare method is Cosine Similarity measure, which has been frequently used in text mining and search engines. However, the weakness of cosine similarity in this case is that, since not every architecture project contains the same number of rooms, the vectors may very different in size. Therefore, measuring the angle between two

Figure 3-8: GUI interface.



vectors is not informative about the distance in the  $n$ -dimensional space.<sup>13</sup> Figure 3-11 shows a part of related compare method that this thesis attempted to use as well as their configurations and limitations.

<sup>13</sup>Danai K., Ankur P., Aaditya R., Jing X. Algorithms for Graph Similarity and Subgraph Matching, 2011.

Figure 3-9: Adding nodes and edges on the canvas.

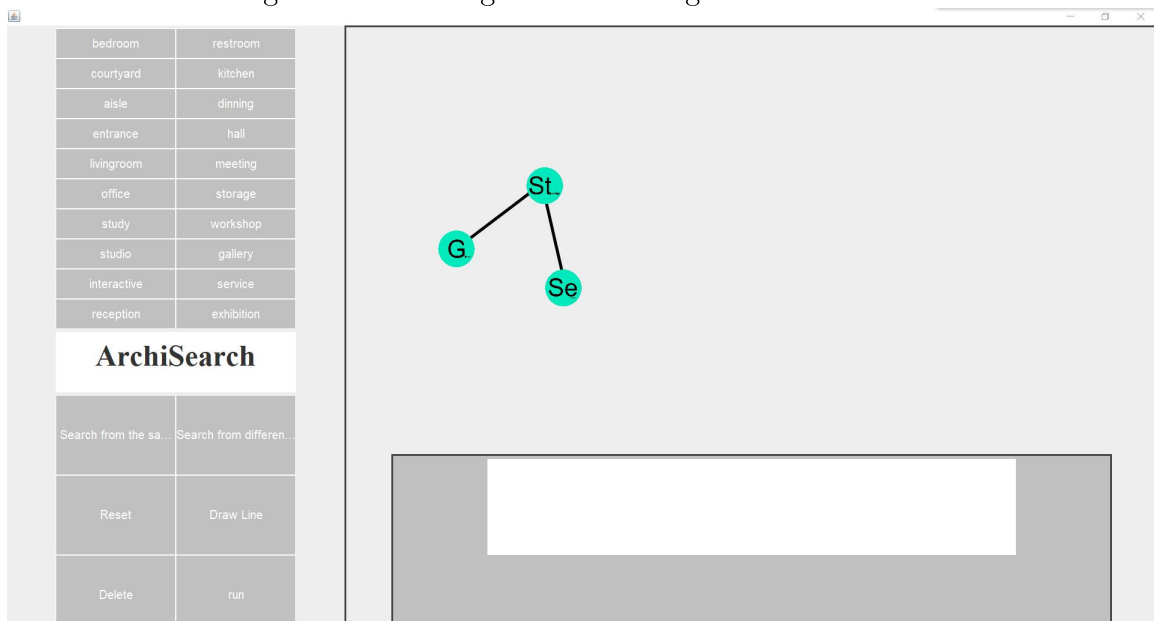


Figure 3-10: Searching results.

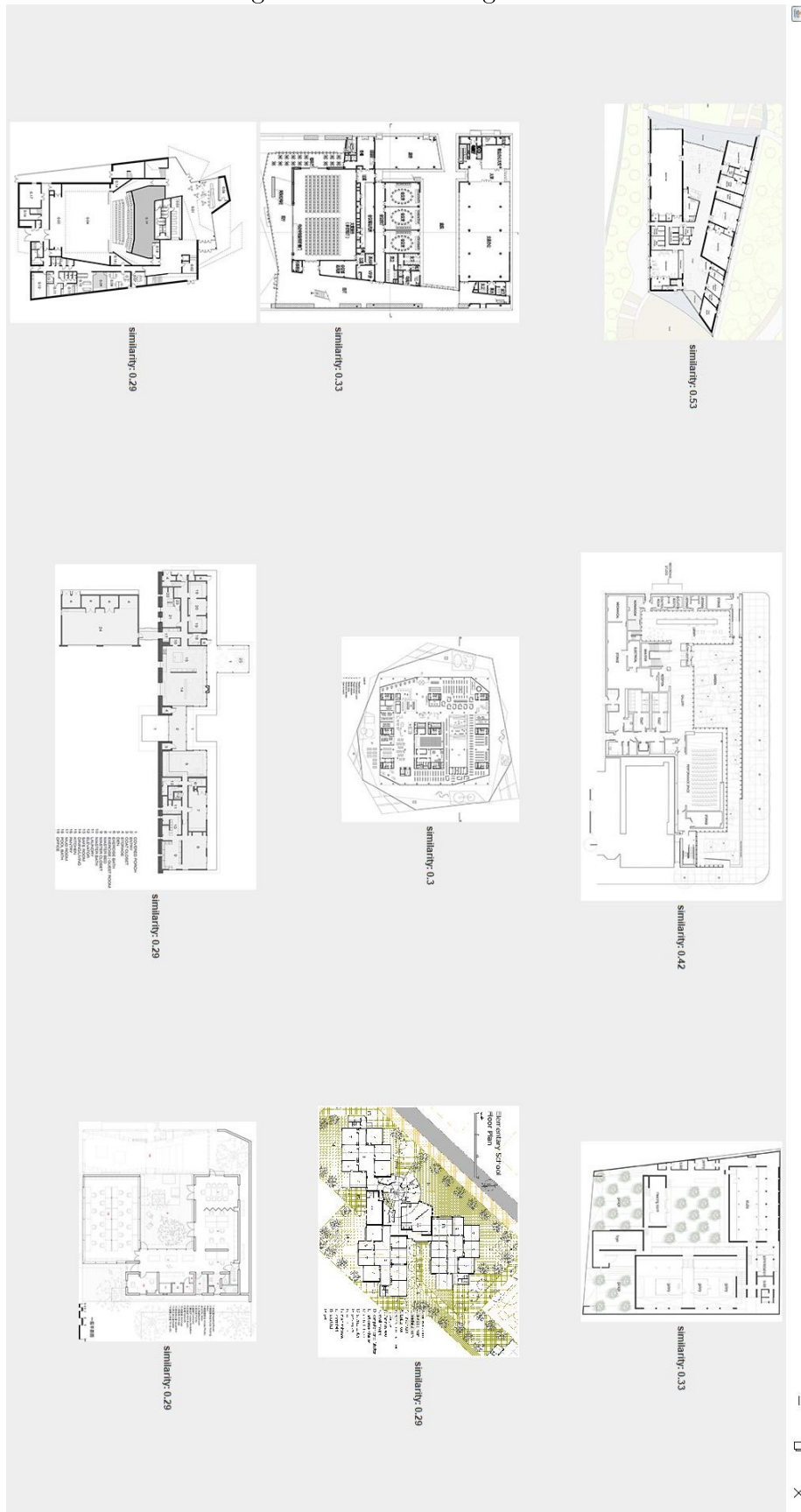


Figure 3-11: Compare methods matrix.

Algorithm name	Applied fields	Expression	Properties	Limitation
Dot Product	Text mining	$a \cdot b = \ a\  \cdot \ b\  \cos \theta$	Find numbers of the same words in two files.	not normalized, similarity value largely depends on the length of the vector
Cosine Similarity	Text mining	$a \cdot b = a \cdot b / \ a\  \cdot \ b\  \cos \theta$	Normalized dot product	If two vectors are very different in length, the result does not reflect the real similarity
Intersection	Text mining		Normalized vectors into $[0,1]$ , which allows horizontal comparison	If two vectors are very different in length, the result does not reflect the real similarity
Jaccard/Tanimoto	Sparse datasets	$J(A, B) =  A \cap B  /  A \cup B $	Binary vectors	
Dice/Sorensen/Czekanowski	Sparse datasets			No related information has been found



# Chapter 4

## Test and Evaluation

### 4.1 Accuracy of the Algorithm

To test the accuracy of the dataset and algorithms, the workflow that this project conducts is to trace an existing graphic diagram as figure 4-1 shows and modify it by removing and inserting some nodes. The first row is to take the original graph representation as a control group to prove that the system is able to detect it as the most similar case. The graph of the second row removes two nodes from the first one, as a result, the system can still detect it as the most similar one. For the last one, reception and restroom are changed to service room, as a result, the system labels the target architecture as the 5th similar one.

### 4.2 Self Similar

The form in appendix A shows part of the comparison results by taking 30 graphic diagrams from the dataset as input. The number is the overall similarity of two architectural projects, which is between 0 and 1. Basically, the higher the number is, the more similar two buildings are. For instance, compare to the same architecture project, figure 4-2 shows a project that has a similarity of 0.8. Compare the two graphs, it can be detected that both contain subcenter nodes, where a cluster of associated nodes are surrounding. This topological relation can be abstracted as

"clustering or compartmentalization"<sup>1</sup>, which is one of the basic topological organizations that can reference to many architecture projects.

Based on observation, if two have a similarity of 0.8, they are strongly similar in both the number of same adjacent rooms and room type, 0.5 means somewhat similar, which may indicate they are similar in either room type or a number of room layout sequences. 0.3 partly similar, 0 is dissimilar. It also denotes that most of the projects(24/30) are proved to be self similar, where 80% of them are the most similar project of them themselves, 10% of them are retrieved as one of the rest of the most similar 9 projects and 10% are not searched out.

---

<sup>1</sup>Architectural topology to explore the intrinsic of contemporary spaces. (2012, Aug 04). In Baidu WenKu.  
Retrieved April 26, 2018, from <https://wenku.baidu.com/view/c748df7e31b765ce05081465.html>

Figure 4-1: Insert 2 new nodes and remove 2 original ones.

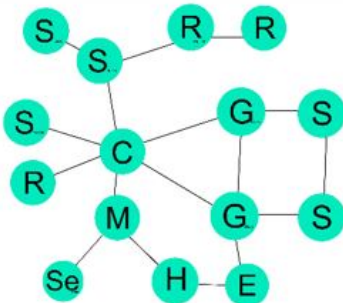

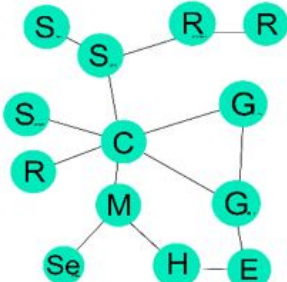

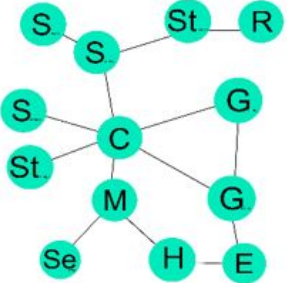

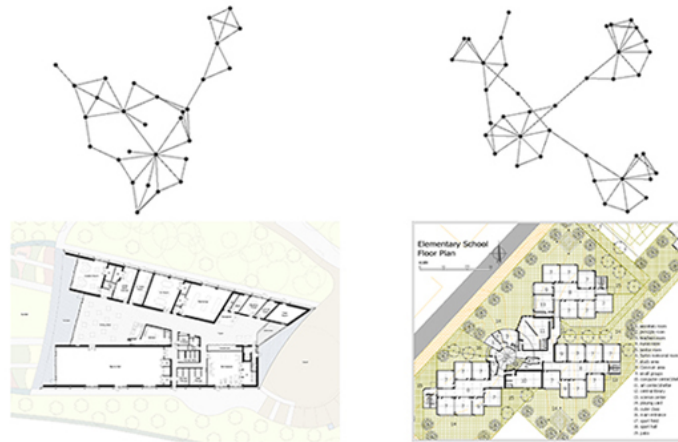
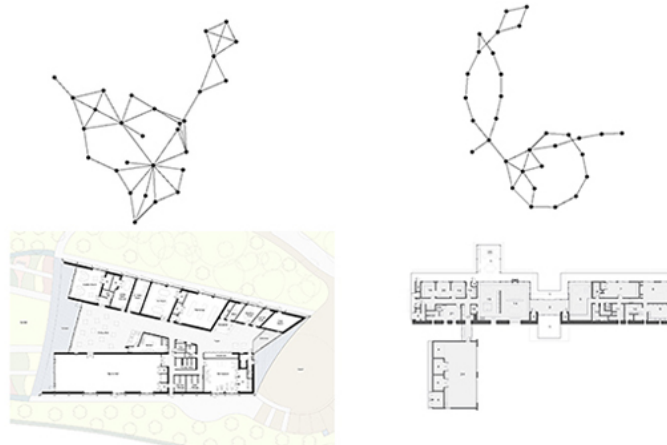
Graphic Diagram	Most Similar Project	Rank	Similarity
		No.1	1.00
		No. 1	0.89
		No. 3	0.36

Figure 4-2: Arch1 and Arch4 with a similarity of 0.8



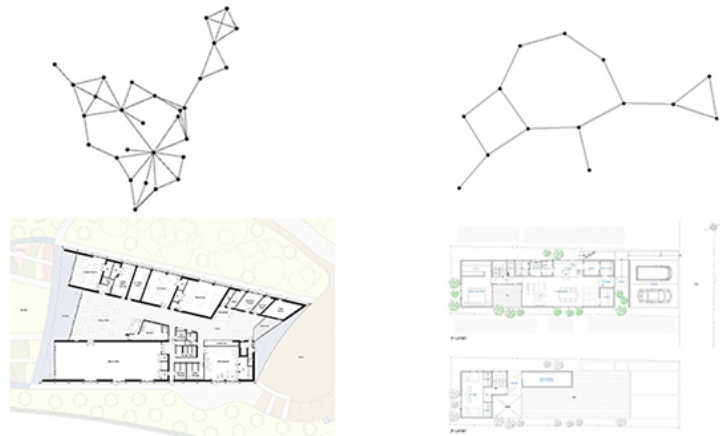
This figure shows a similarity value of 0.8. Compare the two graphs, it can be detected that both contain subcenter nodes, where a cluster of associated nodes are surrounding. This topological relation can be abstracted as "clustering or compartmentalization".

Figure 4-3: Arch1 and Arch18 with a similarity of 0.5



This figure shows a similarity value of 0.5. The two graphs share the same trends, where most of the nodes are only adjacent to two nodes. However, the second graph contains less sub-center nodes than the previous one.

Figure 4-4: Arch1 and Arch19 with a similarity of 0.2



This figure shows a similarity value of 0.2. The two graphs don't show distinct visual similarity in between, but their floor plans share a large part of room types in common.



# Chapter 5

## Discussion

(H2): Projects from the same architect/ architecture company are always under the influence of same kinds of topological relations.

### **Study 1: Topology issue in architecture and SANAA's work**

When referring to "diagrammatic", it is hard not to mention SANAA's work - as I mentioned a few times in this thesis, SANAA devotes to pursuing a kind of clear and rational spatial organization. Their architecture work is a kind of diagrammatic architecture, as figure 5-1 shows, which directly convert functionality and accessibility to spatial layout without any determinate shapes besides the permutation of rectangles and circles. Hence, when architecture topology is too obscure to be understood and referred from, their design was considered as a handbook for topology in architecture. In their work, topological issues can be divided into seven main groups: "clustering or compartmentalization", "concentration or dispersal", "compactness or breakup", "aperture or closure", "inside and outdoor", "restrict and connect", "continuous and break".<sup>1</sup> However, it is regretful that architects and architecture students love and imitate SANAA only because they are concise, austere which look different in shapes and plan layouts. In this context, a tool that is able to analyze architecture from a new perspective becomes meaningful.

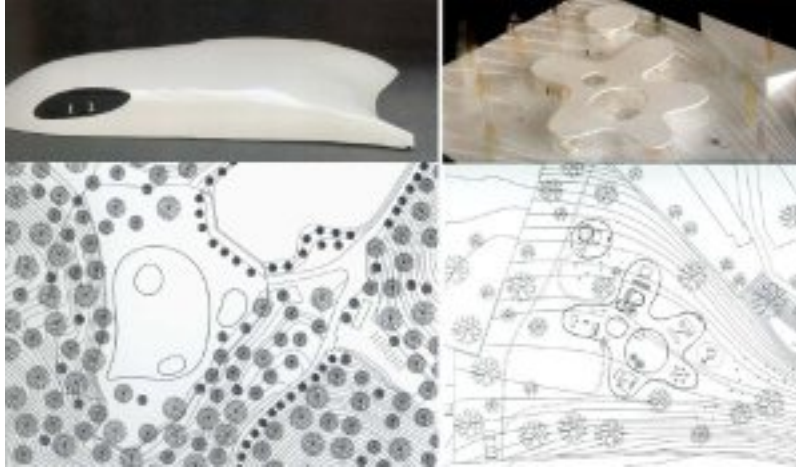
To prove the second hypothesis, this study uses 10 projects designed by SANAA as

---

<sup>1</sup> Architectural topology to explore the intrinsic of contemporary spaces. (2012, Aug 04). In Baidu WenKu.

Retrieved April 26, 2018, from <https://wenku.baidu.com/view/c748df7e31b765ce05081465.html>

Figure 5-1: Their work is similar to bubble charts.



the graphic inputs, since their work are famous for taking topology into consideration at the first stage. Figure 5-2 displays the entire matrix of the search results. It turns out that the system is able to match SANAA's work with other work from SANAA as well as present a relatively high percentage of similarity degree. For example, as the first row in the matrix, SANAA's work *Rolex Learning Center* has a very high degree of similarity with *Grace Farms*, even by looking at the graphic representations only, it is easy to detect the consistency of spatial layouts in between.

Interestingly, as the matrix shows, some work that appear in the retrieval results is not designed by SANAA, however, their graphic representations share distinct resemblance with SANAA's work. Therefore, it is arguably that if we understand a kind of topology pattern beforehand, this system can further become a promising tool to analyze architecture topologically.

Nevertheless, the matrix also presents that, *Mariyama House* (Figure 5-3), which is claimed to be similar to *Towada Art Center* topologically by the architects themselves, are not able to be searched out by the system. This is due to the different architecture types in between. In other words, if two projects contain the similar variety of room type, if they are similar topologically, it is relatively easier to be found out. On the contrary, if two projects are similar in design methods like this case, but totally different in terms of architecture type, the system doesn't have the capability to recognize it. For a further study, with the increasement of dataset, it would be

worthwhile to classify data by typology.

(H3): There is a specific kind of architecture that occurs for most of the time, whereas another kind of architecture never occurs.

Appendix A shows the frequency of each architecture's occurrence besides being similar to itself. With the 30 cases analyzed in this thesis, arch1 appears the most frequently as a result with a similarity index between 1.0 to 0.8. This implies that arch1 share common features with other arch samples more than any other architecture samples do. Probably this is due to the diversity of room types of arch 1. In this project, there are courtyard, hall, dining room as well as interactive space. Each of them is often designed as a principal space. Many other cases followed the same design pattern, thus arch1 became a general similar case for them.

For similarity of other ranges, there is no distinct dominant topology standing out. Additionally, some architecture samples, specifically Arch 25, 31, 36, 43, 46, 76, never occur as similar samples with the comparison among the 100 architecture sample data.

(H4) Similarity range among different architects/architecture companies

One approach to further comprehensively analyze architecture would be collecting specific architectural data from a certain kinds of architects/architecture firms to compare the similarity range among their projects. For example, we can collect 50 Le Corbusier's projects chronologically then encode them into the system to see if there is any unexpected pattern or similar range corresponding to the created time.

Figure 5-2: Compare result of Rolex Learning Center

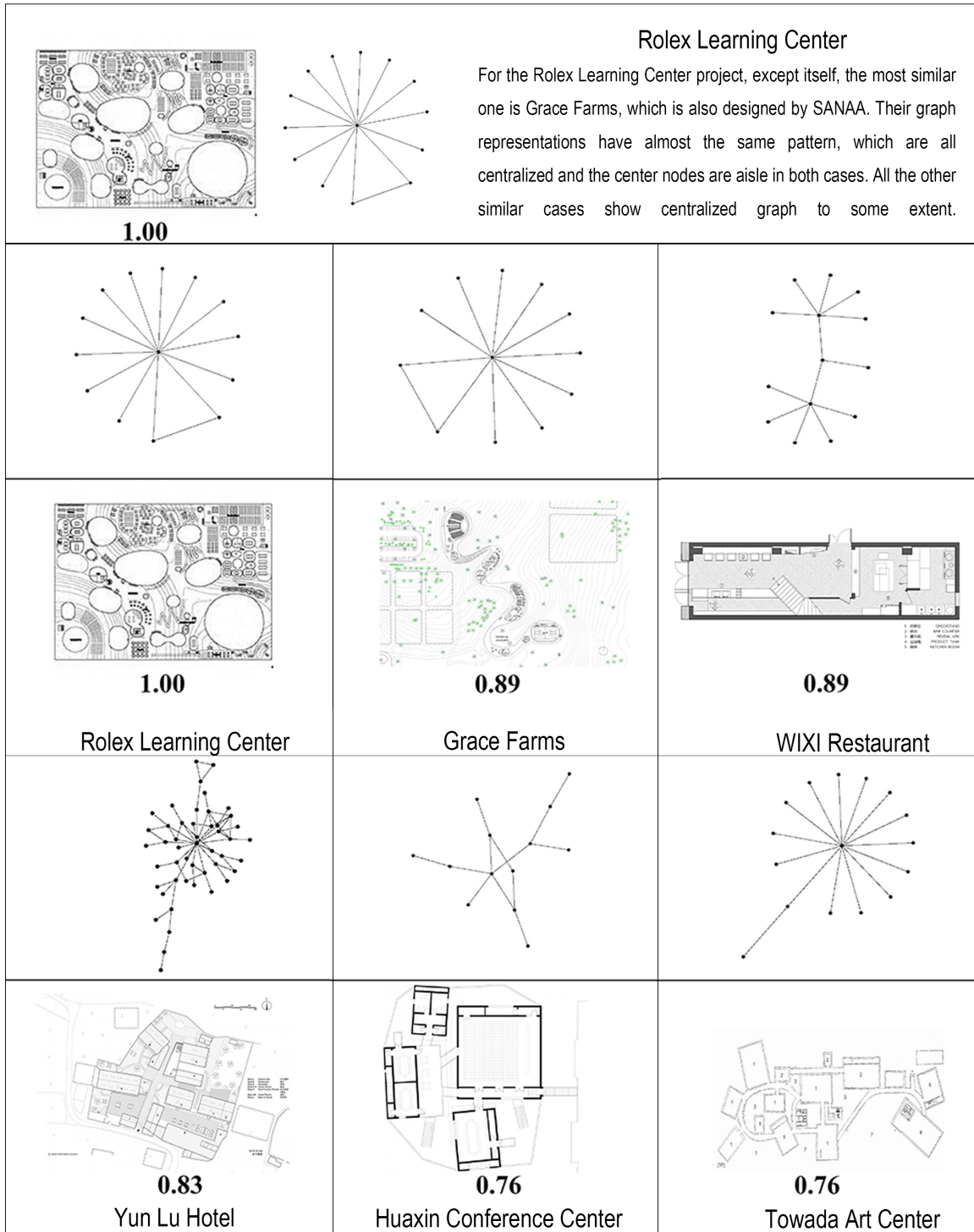


Figure 5-3: Compare result of Grace Farms Center


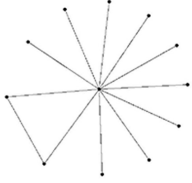
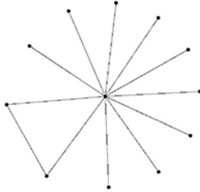
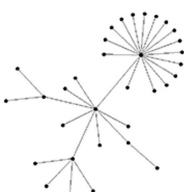
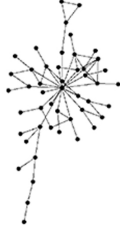



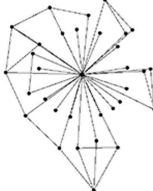
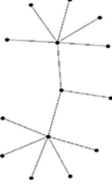




<div>   <div> <p><b>Grace Farms</b></p> <p>Grace Farms is a public art center, which contains lots of interactive spaces, thus, it is most similar to La Source, a media library and center for eco-citizenship in Le Bouscat. In this matrix, 21st century contemporary museum is designed by SANAA, their graph models share more similarity than their shapes and floor plans.</p> </div> </div>		
		
 <p><b>1.00</b></p> <p>Grace Farms</p>	 <p><b>0.7</b></p> <p>La Source</p>	 <p><b>0.64</b></p> <p>Yun Lu Hotel</p>
		
 <p><b>0.58</b></p> <p>21st Century Museum of Contemporary Art</p>	 <p><b>0.55</b></p> <p>WIXI Restaurant</p>	 <p><b>0.53</b></p> <p>Activation and Transformation Huxi Pottery Kiln</p>

Figure 5-4: Compare result of Glass Pavilion

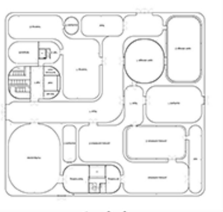


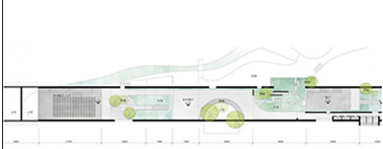


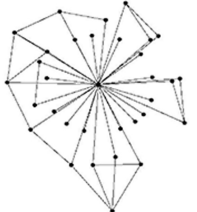
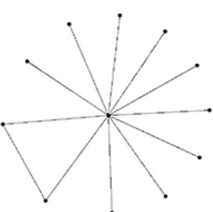



<p style="text-align: right;"><b>Glass Pavilion</b></p> <p>One thing 21st century contemporary museum and Grace Frames share in common is that the main hall that occurs in both cases connects with other main space. Glass Pavilion itself is an art center contains an extensive glass art collection, temporary exhibition galleries, which has the same functionality of the most similar project, Sanbaopeng art museum. Their overall circulations are linear, with courtyards and exhibition interleaved.</p>		
 <p><b>1.00</b></p>		
 <p><b>1.00</b> Glass Pavilion</p>	 <p><b>0.61</b> Sanbaopeng Art Museum</p>	 <p><b>0.52</b> Activation and Transformation Huxi Pottery Kiln</p>
		
 <p><b>0.52</b> La Source</p>	 <p><b>0.47</b> 21st Century Museum of Contemporary Art</p>	 <p><b>0.47</b> Grace Farms</p>

Figure 5-5: Compare result of 21st Century Museum of Contemporary Art

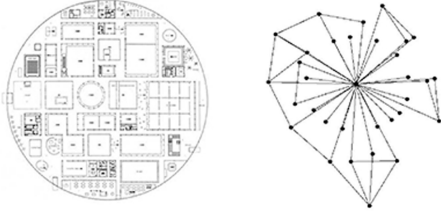
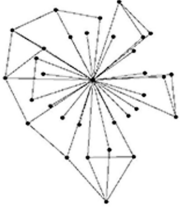
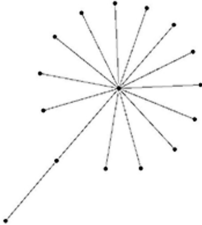
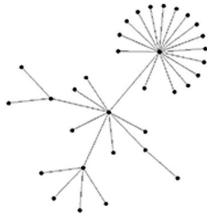





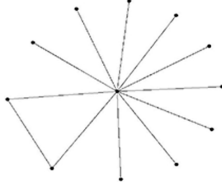



<div>  <p><b>1.00</b></p> </div> <div> <p><b>21st Century Museum of Contemporary Art</b></p> <p>Across this matrix, all the retrievals are art museums, which share the same architecture type with 21st Century Museum. Besides, they all get a relatively higher similarity values, since 21st Museum has a very clear graphic pattern: 1. mainly exhibition-oriented room type 2. all rooms connect to the main center node. Thus, museums that share the same room variety and spatial layout are easy to be detected.</p> </div>		
		
 <p><b>1.00</b> 21st Century Museum of Contemporary Art</p>	 <p><b>0.89</b> Towada Art Center</p>	 <p><b>0.80</b> La Source</p>
		
 <p><b>0.74</b> Schmidt Hammer Lassen</p>	 <p><b>0.62</b> Sanbaopeng Art Museum</p>	 <p><b>0.60</b> Grace Farms</p>

Figure 5-6: Compare result of Moriyama House


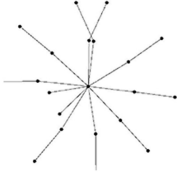


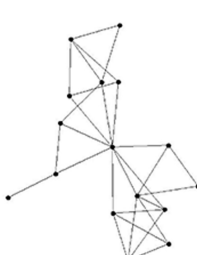



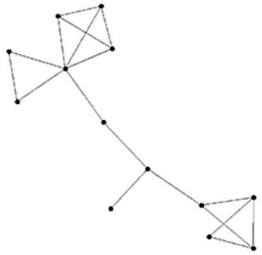


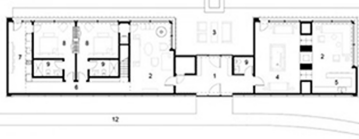
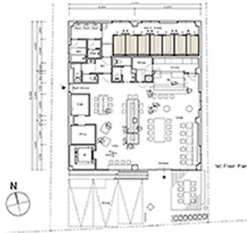

<div>   </div> <div> <h3>Moriyama House</h3> <p>Moriyama House which is claimed to be similar to Towada Art Center topologically by the architects themselves, are not able to be searched out by the system. This is due to the different architecture types in between. Since Moriyama House is a collection housing project which is more similar to other resident projects, this is also the reason Funabashi Apartment Building- another housing project from SANAA, is searched out.</p> </div>		
		
<div>  <p><b>0.9</b> Anonymous Architecture</p> </div>	<div>  <p><b>0.9</b> House on Cedar Hill</p> </div>	<div>  <p><b>0.87</b> Shade Court House</p> </div>
		
<div>  <p><b>0.8</b> Woodshed</p> </div>	<div>  <p><b>0.8</b> Unplan Kagurazaka</p> </div>	<div>  <p><b>0.8</b> Funabashi Apartment Building</p> </div>

Figure 5-7: Compare result of Towada Art Center


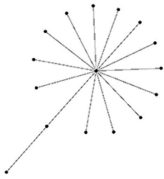
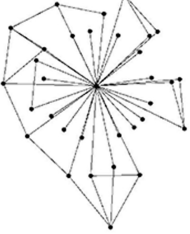
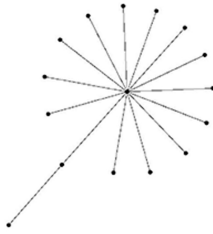
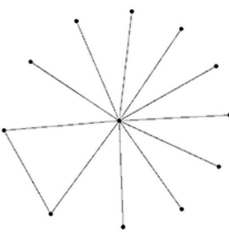
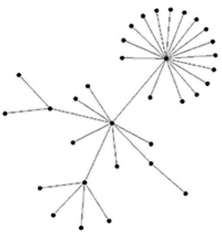
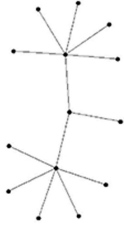

<div>   </div> <div> <p><b>Towada Art Center</b></p> <p>The top three most similar projects of Towada Art Center are from SANAA as well. Interestingly, the last three projects designed by different architecture companies can be abstracted to the same kind of graphic representation - a series of main nodes connects to each other with a cluster of associate nodes in the surrounding. The only difference is the number of nodes that are around those main nodes.</p> </div>		
		
<p><b>1.00</b></p> <p>21st Century Museum of Contemporary Art</p>	<p><b>0.94</b></p> <p>Towada Art Center</p>	<p><b>0.93</b></p> <p>Grace Farms</p>
		
<p><b>0.76</b></p> <p>La Source</p>	<p><b>0.6</b></p> <p>Activation and Transformation Huxi Pottery Klin</p>	<p><b>0.56</b></p> <p>Schmidt Hammer Lassen</p>

Figure 5-8: Mariyama House

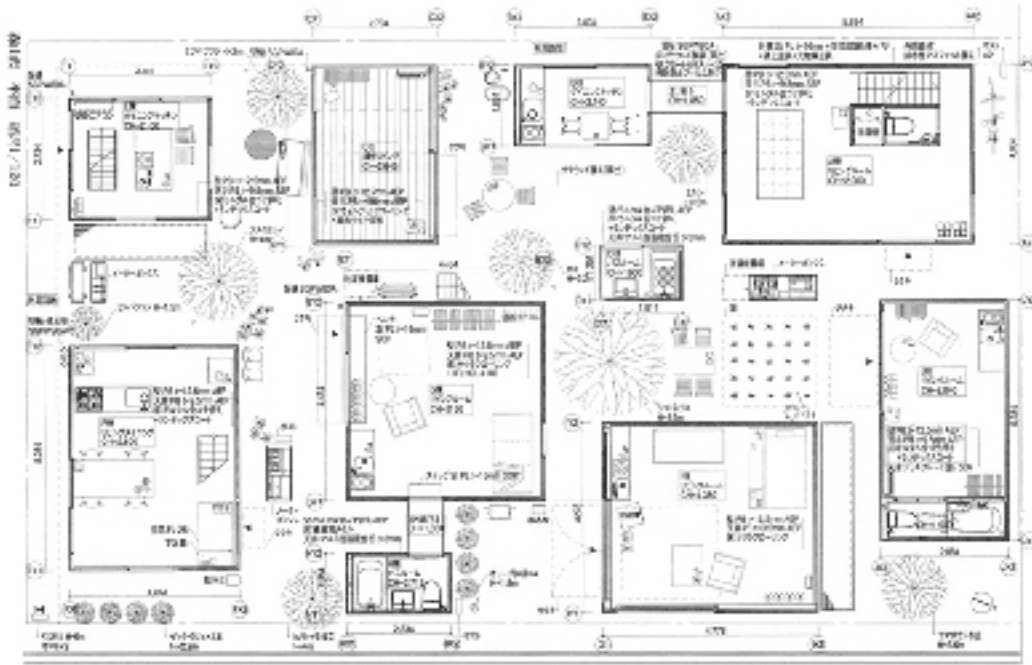
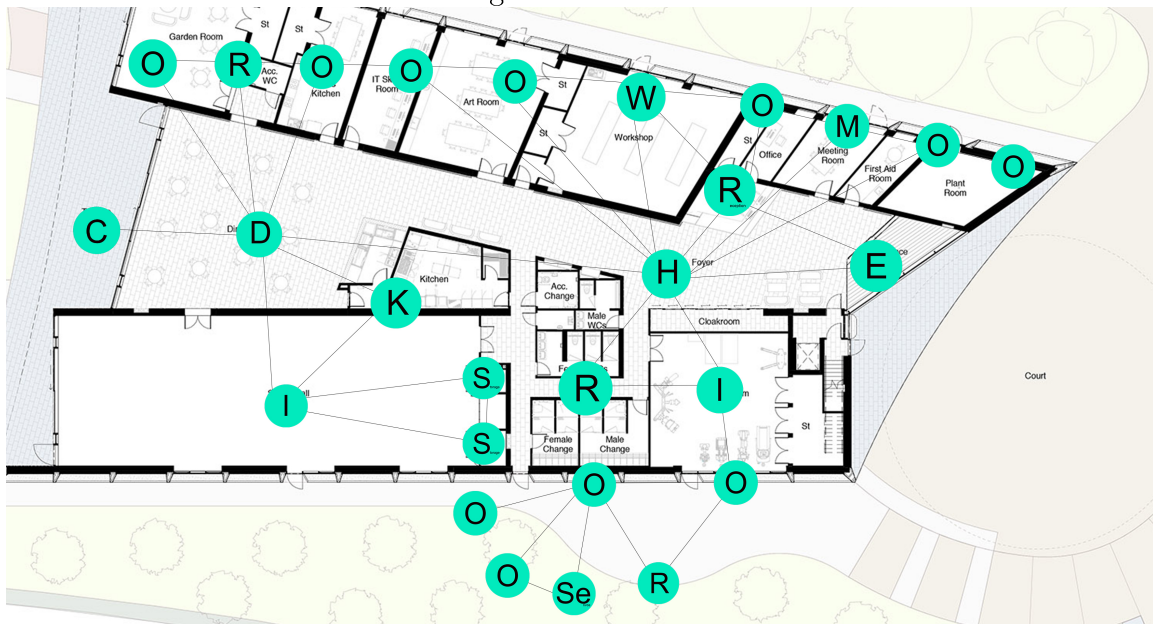


Figure 5-9: Arch1



# Chapter 6

## Limitation

### 6.1 Limitation of Current Algorithm

The form in Chapter 3 shows all the algorithms that were attempted to apply in this project, however, all of them are based on vectors, which inevitable involve the mapping from architecture to vectors. A uniform idea in this thesis is that, since architecture was mapped into vector, while the length of vector represents the room number in one single architecture, the only different is whether we normalize it or not (meaning two comparing projects will get a normalized value based on the total room numbers they have). However, in case of wide difference between two projects, for example Architecture A contains 5 rooms only whereas B has 100, there would be large errors here. This is because in text mining, if the size of two documents are very different, then they are very likely to be dissimilar. In contrast, architecture topological similarity is not sufficient to be measured by dimensions, even a small house with only 5 rooms could share some kinds of similarity with a complex office building. In this regard, it is still worthwhile to keep exploring other algorithms to conduct more experiments.



# Chapter 7

## Conclusion

To review the entire process, this thesis first presents the reason why architecture is not suitable for searching by current searching tools. Then, it focuses on workflow and methods for compare algorithms. Three tests are conducted to prove the accuracy of this method. Finally, this thesis discusses the possible applications in architecture analysis, as well as the considerations and limitations.

### 7.1 Research Results and Contribution

This thesis intends to provide a new attempt to architecture searching, by doing this, the focus is on proposing a feasible workflow that has been applied to a number of cases and proposing a new perspective to analyze architecture topologically. Challenges that are iterated a lot come from the right graph to choose for storing the adjacent data and the algorithms to compare similarity. As Chapter 3 mentioned, different approaches largely depend on how to define similarity. In this regard, there is no precise graph model that is suitable for all situations. This thesis provides and evaluates a solution which is based on the author's understanding (as an architect) of architecture topological similarity and the theory of graph. For this reason, it may differ from people from other areas. As what I find out in the communications with peers from other departments, it is interesting to see how people with various backgrounds tend to build different architecture graph models based on their definitions.

In fact, I would propose that architectural similarity itself may vary from person to person. This becomes the reason why the system supports upload files option for users to search architecture cases based on their customized dataset.

Perhaps the most distinct contribution of this thesis is its implication of one valuable architectural application based on theory of graph. Scholars have been researched theory of graph in architecture problems solving for decades, but rarely of them incorporated with computational knowledge or illustrated with considerable examples. Although the search engine is limited in the current stage, tests and proposals that discussed in this thesis shows promising future for the further studies.

Another important outcome is the reconsideration of architecture analysis. Although topological problems have been brought to the table in architecture design pedagogy, typology and morphology analysis is still dominant in current architecture schools. This thesis illustrates a way to quantify architecture topological similarity by a pre- given spatial sequence. It is arguably that more cases will come out based on other focus of architecture spatial components.

## 7.2 Future Possibility

To generalize and summarize the adjacent attributes among various spatial sequences, it is necessary to collect ten or more times the size of data. Due to the sparse data in my current dataset, it is hard to give a value to the edge which can be used to represent the cost (or the weight) of this kind of adjacency. For a further step, assume that we collect all kinds of adjacent relations in all variety of the plans to a matrix. Then once we encode one project into graph, we can refer to that matrix to learn about the probability of this adjacency. If 90% restrooms are adjacent to living room, it means the cost of restroom to living room is relatively lower, which reflects to a low value for the edge. In the opposite, if two rooms only occur 10% times to be connected, it reflects a higher weight. This proposal offsets the weakness of the current vector-based algorithm as well that might provide more vision to the study of graph-based topological similarity.

Through the tests, I believe that the issue that this thesis is starting to explore may become more positive as we are taking a new look at analyzing architecture and understanding architecture design concern. Additionally, the combination of this idea with computer vision knowledge to get graphic diagram could be an approach to relief people from the labor work of labelling data manually. A variety of applications of machine learning and other techniques are also promising to address the issue of obtaining more accurate results.



# Bibliography

- [1] Bill,H.and Adrian,L. *The architecture of architecture. Models and Systems in Architecture and Building*. The Construction Press Ltd, 1975.
- [2] Lubow, A. *Disappearing Act*.  
Retrieved from <https://www.nytimes.com/2005/10/09/style/tmagazine/disappearing-act.html>. 2005, OCT 9.
- [3] Steadman,P. Models in our heads, models in the material world, and models in the world of objective knowledge. *Models in our heads, models in the material world, and models in the world of objective knowledge. Models and Systems in Architecture and Building*. The Construction Press Ltd, 1975.
- [4] March,L.and Steadman,P. *The geometry of environment: An Introduction to Spatial Organization in Design*. pages 242-243. The MIT Press, 1971.
- [5] Steadman,P. Graph-theoretic representation of architectural arrangement In L. March (Ed). *The Architecture of Form: Cambridge Urban and Architectural Studies*. Cambridge: Cambridge University Press, 1976.
- [6] Tomlinson,J. *Computer representations of architectural problems. Models and Systems in Architecture and Building*. The Construction Press Ltd, 1975.
- [7] Danai K.and Ankur P.and Aaditya R.and Jing,X. *Algorithms for Graph Similarity and Subgraph Matching*. 2011.
- [8] Belief algorithm. (2018, Feb 16). In Wikipedia, the free encyclopedia. *Retrieved February 20, 2018, from https://en.wikipedia.org/wiki/Belief\_propagation*.
- [9] Manuel J. Fonseca, Alfredo F. and Joaquim A. Jorge *Sketch-Based Retrieval of Vector Drawings INESC-ID / TU Lisbon, R. Alves Redol. 9, 1000-029 Lisboa, Portugal*
- [10] Pearl J. *Comment: graphical models, causality and intervention..* Stat Sci 1993;8;266-269
- [11] Marginal distribution.(2018,Feb 21)In Wikipedia, the free encyclopedia. *Retrieved February 19, 2018, from https://en.wikipedia.org/wiki/Marginal\_distribution*.

- [12] Architectural topology to explore the intrinsic of contemporary spaces.(2012,Aug 04).In Baidu WenKu. *Retrieved April 26, 2018, from <https://wenku.baidu.com/view/c748df7e31b765ce05081465.html>.*
- [13] Recursion. (27 March,2018)In Wikipedia,the free encyclopedia. *Retrieved April 26, 2018, from <https://en.wikipedia.org/wiki/Recursion>.*
- [14] Ching,F. *Architecture: Form, Space, and Order*. October 2014.
- [15] In Portico.space. *Retrieved April 29, 2018, from <http://portico.space/journal//architectural-concepts-circulation>.*
- [16] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. *Understanding belief propagation and its generalizations, pages 239–269*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.

# Appendix A

## Matrix of all the analysis results

Taking 13 cases from the database and using their graphic representations as the program inputs, this appendix contains samples of data matrix displays that were developed during the similarity analysis phase of this project.

Figure A-1:





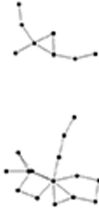







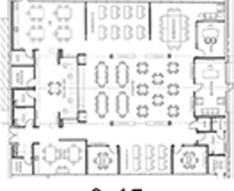

		<p><b>The Hawkhead Centre</b></p> <p>The Hawkhead Centre is an activity center. All the six similar projects are the same architecture type. The most similar project (House on Cedar Hill) shares the same trends with The Hawkhead Centre, where most of the nodes are only adjacent to two nodes. The only different is that the graph of House on Cedar Hill project contains less sub-center nodes than the target one.</p>
		
 <p><b>1.00</b> The Hawkhead Centre</p>	 <p><b>0.57</b> House on Cedar Hill</p>	 <p><b>0.49</b> Dominique Coulon</p>
		
 <p><b>0.48</b> Poets Society Building</p>	 <p><b>0.45</b> Reggio Emilia School</p>	 <p><b>0.44</b> Unplan Kagurazaka</p>

Figure A-2:

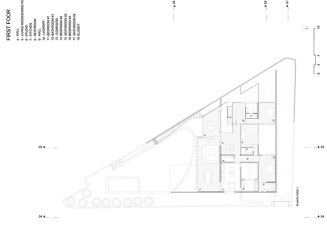













		<p><b>ED&amp;JO House</b></p> <p>ED&amp;JO House is a typical resident building, the spatial layout resembles The Hawkhead Centre. In this case, it's similar projects are also like The Hawkhead Centre's. Across those projects, the most similar ones are also resident buildings that have the same kind of scale.</p>
		
 <p><b>1.00</b> ED&amp;JO House</p>	 <p><b>0.34</b> The Hawkhead Centre</p>	 <p><b>0.32</b> House on Cedar Hill</p>
		
 <p><b>0.25</b> M House</p>	 <p><b>0.24</b> Unplan Kagurazaka</p>	 <p><b>0.23</b> Dominique Coulon</p>

Figure A-3:

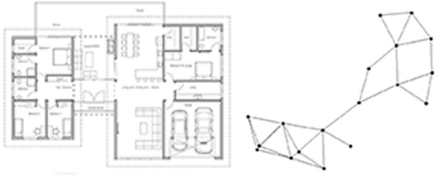












<div>  <p><b>Modern House</b></p> <p>This project is an anonymous common resident building floor plan that I collect from the Internet. Like many common houses, it is regular even mediocre spatially. However, one thing that is interesting to me is that, it matches to those anonymous resident buildings that I collect in the same way. It is hard to match these kinds of topology into any of the named ones, but it is obviously that they don't share topological similarity with architecture from the famous companies.</p> </div>		
		
 <p><b>1.00</b></p> <p>Modern House</p>	 <p><b>0.31</b></p> <p>ED&amp;JO House</p>	 <p><b>0.31</b></p> <p>Celebration Homes</p>
		
 <p><b>0.30</b></p> <p>Shade Court House</p>	 <p><b>0.30</b></p> <p>Afsharian's House</p>	 <p><b>0.28</b></p> <p>House on Cedar Hill</p>

Figure A-4:

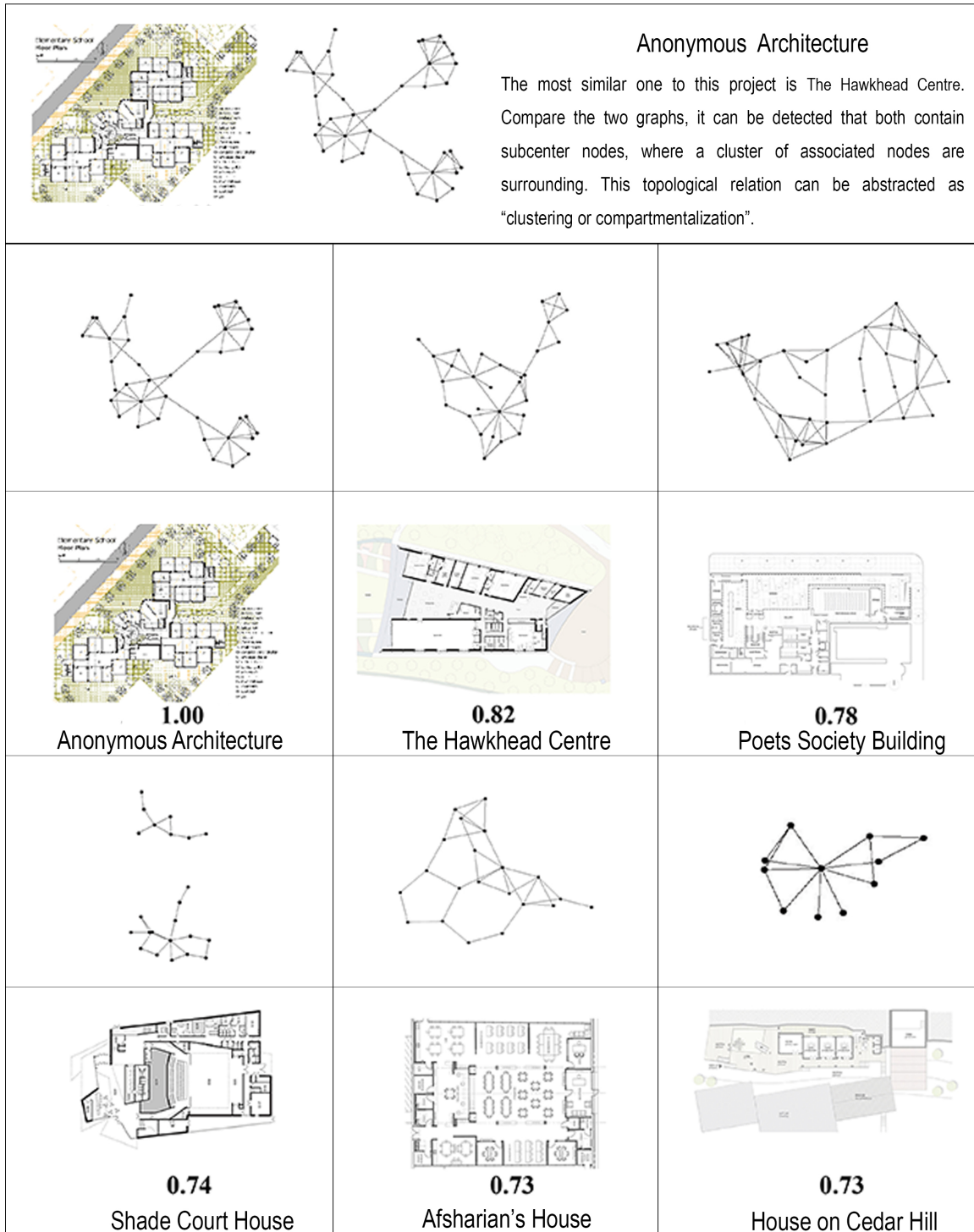


Figure A-5:









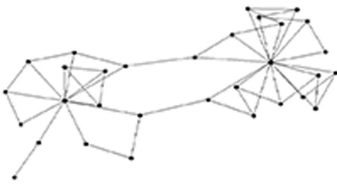




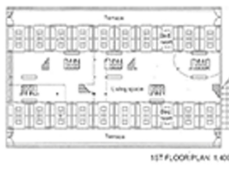
<div>   <div> <h3>Celebration Homes</h3> <p>Celebration Homes is another anonymous common resident building floor plan that I collect from the Internet. As the Modern House project I mention previously, this project shares similarity with it and other common resident buildings</p> </div> </div>		
		
 <p><b>1.00</b> Celebration Homes</p>	 <p><b>0.37</b> ED&amp;JO House</p>	 <p><b>0.31</b> Modern House</p>
		
 <p><b>0.31</b> Youth Hotel</p>	 <p><b>0.31</b> Afsharian's House</p>	 <p><b>0.25</b> Saishunkan Seiyaku Women's Dormitory</p>

Figure A-6:

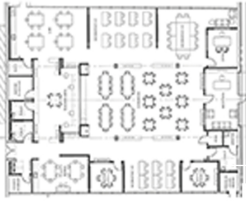





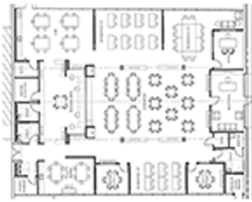







<div>   <div> <h3>Reggio Emilia School</h3> <p>Reggio Emilia School is a middle school that contains many public classroom spaces. In this context, it is similar to other projects that also conclude many interactive hotels, such as hostel, collection housing and activity center. One distinct feature is that across all projects, they all have a centralized, big-scaled main hall in the middle.</p> </div> </div>		
		
 <p><b>0.73</b></p> <p>The Hawkhead Centre</p>	 <p><b>0.58</b></p> <p>Reggio Emilia School</p>	 <p><b>0.50</b></p> <p>House on Cedar Hill</p>
		
 <p><b>0.43</b></p> <p>Youth Hotel</p>	 <p><b>0.42</b></p> <p>Anonymous Architecture</p>	 <p><b>0.42</b></p> <p>Poets Society Building</p>

Figure A-7:

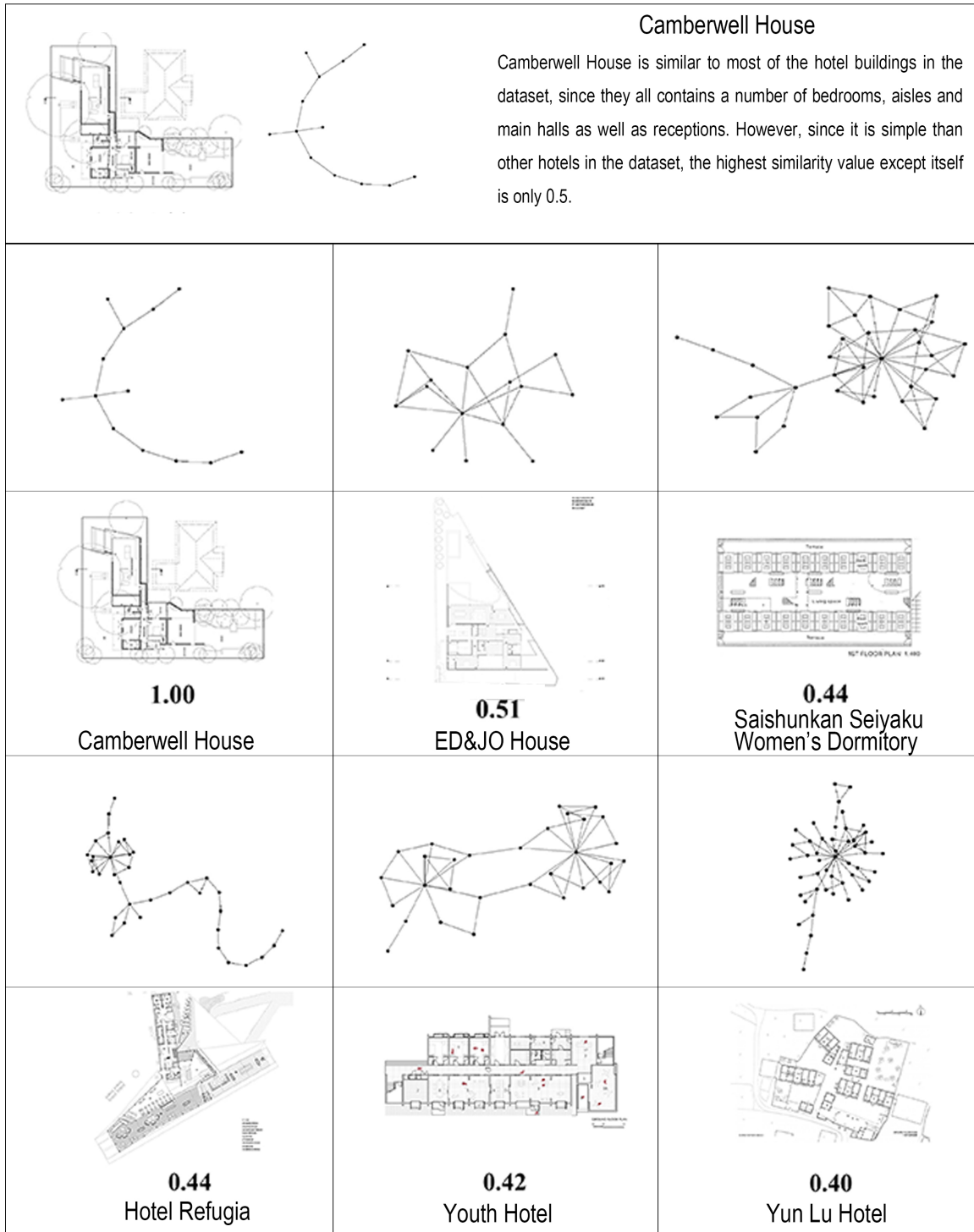


Figure A-8:

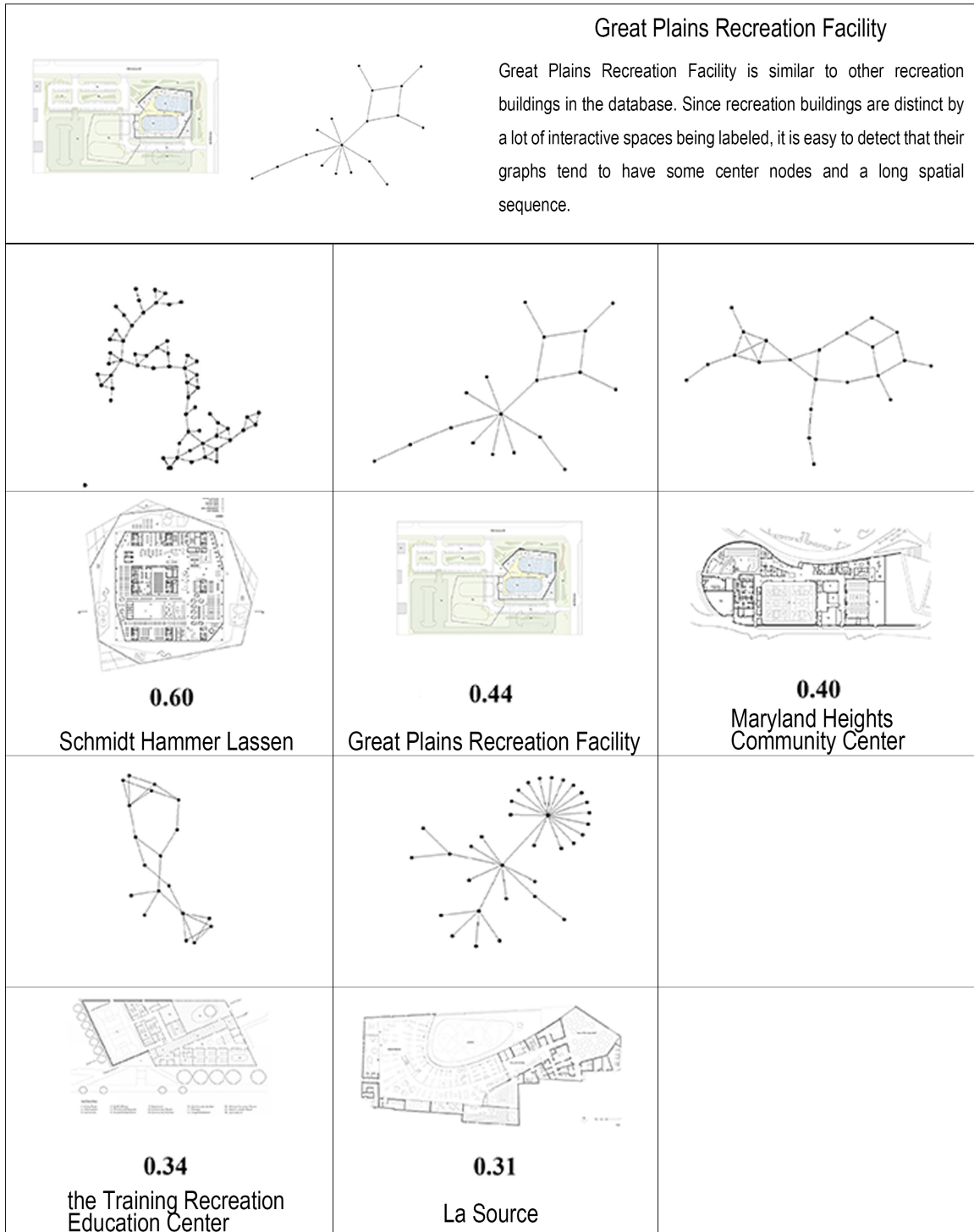


Figure A-9:

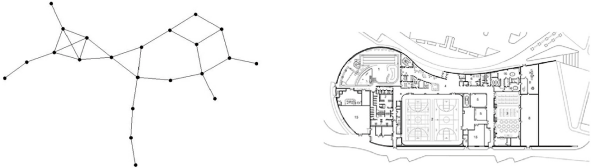


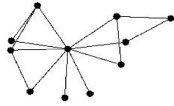





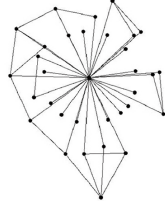
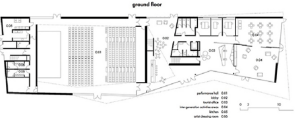


		<p><b>Maryland Heights Community</b></p> <p>This project is a complex community building that contains a large number of rooms. Across the searching results, it can be detected that most of them are also complex buildings and three of them share the same architecture type.</p>
		
 <p><b>1.00</b></p> <p>Maryland Heights Community</p>	 <p><b>0.62</b></p> <p>Schmidt Hammer Lassen</p>	 <p><b>0.5</b></p> <p>Activation and Transformation of Huxi Pottery Kiln</p>
		
 <p><b>0.48</b></p> <p>Art and Conference Centre Inter Generation Centre and Office</p>	 <p><b>0.4</b></p> <p>TREC the Training Recreation Education Center</p>	 <p><b>0.38</b></p> <p>Saishunkan Seiyaku Women's Dormitory</p>

Figure A-10:



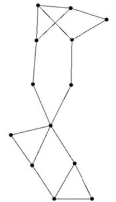
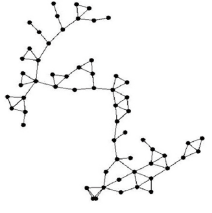
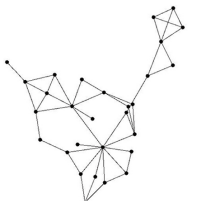

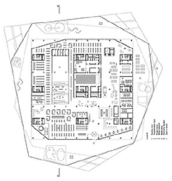

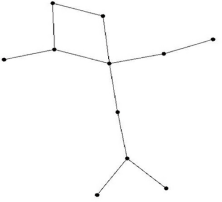
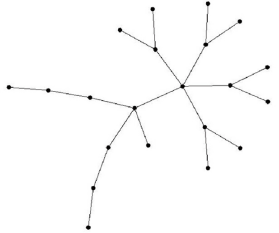
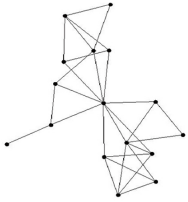

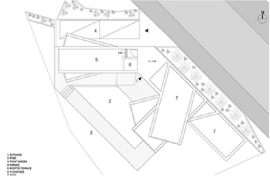

		<p><b>Bamboo Forest on the Roof</b></p> <p>Bamboo Forest on the Roof has the spatial feature of traditional Chinese architecture. Every room is surrounding by a courtyard that belongs to itself. The last five similar projects also use “courtyard” as their design elements. Especially the Façade reading room project, is located in the same city with it.</p>
		
 <p><b>1.00</b> Bamboo Forest on the Roof</p>	 <p><b>0.33</b> Schmidt Hammer Lassen</p>	 <p><b>0.29</b> The Hawkhead Centre, Paisley</p>
		
 <p><b>0.25</b> Façade Reading Room</p>	 <p><b>0.22</b> Afsharian's House</p>	 <p><b>0.19</b> Saishunkan Seiyaku Women's Dormitory</p>

Figure A-11:


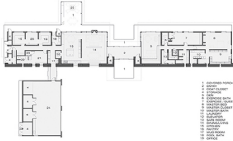

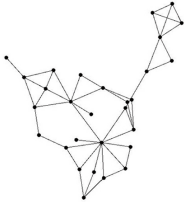
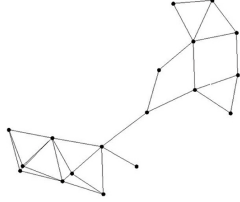
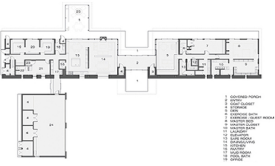



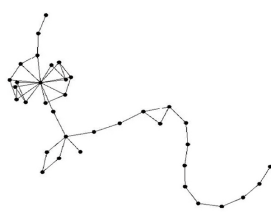
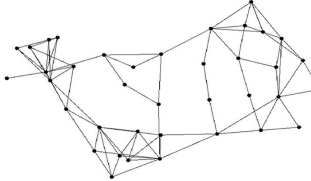
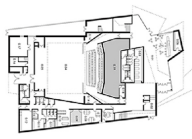

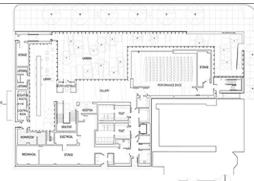
		<p><b>House on Cedar Hill</b></p> <p>House on Cedar Hill is a kind of resort hotel that contains a variety of functions. Thus, it shares similarity with other projects that have more or less the same kind of functionality.</p>
		
 <p><b>1.00</b></p> <p>House on Cedar Hill</p>	 <p><b>0.36</b></p> <p>The Hawkhead Centre</p>	 <p><b>0.28</b></p> <p>Modern House</p>
		
 <p><b>0.28</b></p> <p>Théodore Gouvy Theatre</p>	 <p><b>0.27</b></p> <p>Afsharian's House</p>	 <p><b>0.25</b></p> <p>Poetry Foundation</p>

Figure A-12:




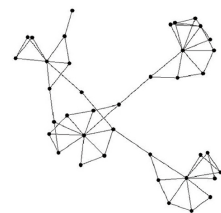
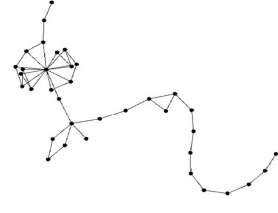
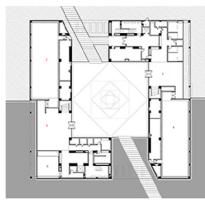
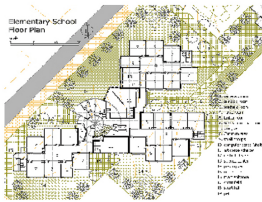

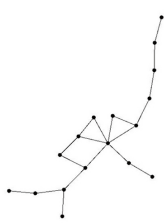
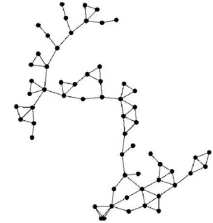

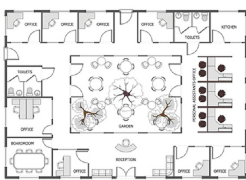


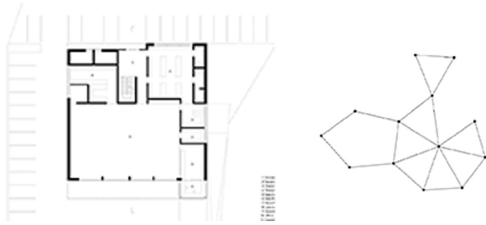

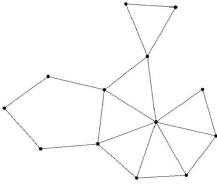
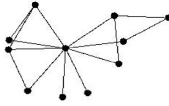



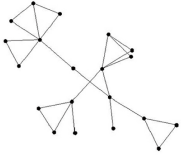
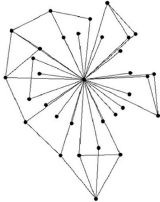

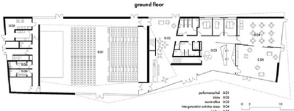


<div>   <div> <h3>Fan Zeng Art Gallery</h3> <p>Fan Zeng Art Gallery is a traditional large-scale art gallery in China. Since there is not enough cases of architecture under the same design pattern, the matrix doesn't denote a distinct similarity across all the projects.</p> </div> </div>		
		
<div>  <p><b>1.00</b> Fan Zeng Art Gallery</p> </div>	<div>  <p><b>0.28</b> Gekko / Moke Architecten</p> </div>	<div>  <p><b>0.28</b> Hotel Refugia</p> </div>
		
<div>  <p><b>0.27</b> Tales Pediatric Dental Office</p> </div>	<div>  <p><b>0.27</b> Afsharian's House</p> </div>	<div>  <p><b>0.26</b> The Hawkhead Centre</p> </div>

Figure A-13:

<div>  <div> <h3>Anlong County Tourist Center</h3> <p>As a tourist center, this project has concise circulation and room types, which is similar to some of the small-scaled building such as the pottery kiln and activity center.</p> </div> </div>		
		
 <p><b>0.73</b></p> <p>The Hawkhead Centre</p>	 <p><b>0.60</b></p> <p>Anlong County Tourist</p>	 <p><b>0.60</b></p> <p>Activation and Transformation of Huxi Pottery Kiln</p>
		
 <p><b>0.60</b></p> <p>Art and Conference Centre Inter Generation Centre and Office</p>	 <p><b>0.60</b></p> <p>21st century museum of contemporary art</p>	 <p><b>0.57</b></p> <p>Schmidt Hammer Lassen</p>