Modeling Computational Design

An Interactive Visual Analysis of the Field based on Topic Modeling and Bibliometrics

Zhuoni Yang

Committee Members:

Prof. Daniel Cardoso Llach, Carnegie Mellon University

Prof. John Stamper, Carnegie Mellon University

Prof. Adam Perer, Carnegie Mellon University

Abstract

Design is everywhere in people's life, and as the recent development of computation technologies, *Computational Design*, a relatively new research field that integrating concepts from computer science to design process, has been becoming increasingly popular. Because the field is highly interdisciplinary, quickly updating and relatively new, there is a lack of an effective and systematic window to observe it. To find principles and insights about topics and keywords, geographic distribution, the collaboration of researchers and historical evolution in computational design, my thesis utilizes interactive visualization techniques to analyze the field based on topic modeling and bibliometrics.

Keywords

Computational Design, Topic Modeling, Bibliometrics, Data Visualization, Data Analysis

Acknowledgment

During my two years at Carnegie Mellon University, I have received invaluable bits of help from many people. They not only give me beneficial suggestions and feedback to my research, but also encouragement and support to my life.

First and foremost, my sincere thanks flow to my advisor, Daniel Cardoso Llach, for his helpful guidance, valuable suggestions, and constant encouragement. His impressive insights, profound theoretical knowledge, and enlightening opinions taught me so much, which even inspire me the primary direction for my thesis. He always gives me quick and accurate feedback for my research progress and draft, with a strong planned milestone for my thesis.

I also want to express my heartfelt thanks to the other committee members who have helped and taught me in this university. I am particularly grateful to Professor John Stamper and Professor Adam Perer. They provided much guidance and broadened the technical depth for my thesis.

Then, I owe many thanks to all the CodeLab members and friends in CMU, who gave me wonderful feedbacks and inspired discussions in the two years. Best wishes to you guys and hope we all will have the best future.

Last but not least, I am deeply indebted to my parents, and my boyfriend, Di Wang, for their loving considerations and great confidence to me. Their support and comfort gave me much strength.

The thesis could not be finished without their patient helps. Thank you all for reading this article.

Table of Contents

1 Introduction

> *Design is a roadmap or a strategic approach for someone to achieve a unique expectation. It defines the specifications, plans, parameters, costs, activities, processes and how and what to do within legal, political, social, environmental, safety and economic constraints in achieving that objective.*
>
> Y. DON KUMARAGAMAGE
>
> *Design Manual.* Vol 1 (2011)

Design is pervasive in people's everyday life. Skyscraper construction involves complicated architectural design. Popular mobile apps require delicate user experience design. Chairs, pens, mugs, as well as almost all the daily necessities, are highly related to industrial design. Design projects could be very large and complex, especially for the Built environment, and the design processes could cost a lot of human and time resources.

The recent rapid evolvement of Computer Science poses both opportunities and challenges to improve the process of design. On the one hand, computers become a wonderful substitution of pencils and papers as people nowadays commonly draw on computers, as well as provide powerful visualization tools, such as *Computer-Aided-Design* (CAD) software, including AutoCAD and Rhino. On the other hand, computers can not only be used as virtualization tools but also carry substantial computational power, which is particularly good for *automated* processing and reasoning over *tons of* data. For example, one idea of *algorithmic* design is to develop an algorithm with certain design principles formalized in it and let the computer to enumerate and evaluate different concrete design representations, following those principles. Here comes the field of *Computational*

Design, which is aimed to apply newly developed computation concepts, ideas and technologies to inspire design ideas, help design process and resolve design problems in a systematic and effective way.

The purpose of my thesis is to provide an overview of knowledge in Computational Design using related academic publications. Section 1.1 sketches the history of Computational Design and an introduction to this field. In Section 1.2, I explain why a bibliometric approach with visual interaction is desirable to help obtain an unbiased and quantitative review of the field, as well as describe the goal and methodology of my thesis. Finally, I give a survey of related work in Section 1.3 on literature reviews of Computational Design, bibliometric methods for other fields, as well as interactive visualization technologies.

The first challenge is to obtain reliable bibliographic data. I found *CuminCAD*, an online library that is one of the most reliable, accurate and accessible online data sources for publication records in Computational Design. Using a web crawler, I scratched the data from CuminCAD and found some of the data are quite messy. I clean the data to fix glitches, such as inconsistent author names, misspelling information, etc. and verify the data with official websites to generate a complete, correct and coherent database.

The second challenge is to conduct meaningful analyses on the bibliographic data. In my thesis, I perform multidimensional interactive visual analyses. To analyze topic evolution, I apply the Latent Dirichlet allocation algorithm to abstracts and keywords of the publications in different timeframes. To study the geographic distribution of publications, I collect institution information from the authors' email addresses and then map institutions to countries to obtain geographic data. To investigate collaboration among individual researchers, institutions, and regions, I implement a visualization process based on Simple JS to generate sensible networks.

My work could help new learners in Computational Design know about the history of the area and understand the research topics, and help researchers keep eyes on specific core problems. It could also provide a guide for researchers to explore trending topics and to find their interested collaborators.

1.1 Computational Design as a Research Field

Computational Design arose as an *interdisciplinary* research field around the 1960s. One of the pioneer research groups was the Key Centre of the Design Computing and Cognition at the University of Sydney in Australia. As pointed out by (Bayazit 2004), researchers in Computer Science started to develop systematic design methods during the 1970s, such as programming to evaluate design decisions. A lot of researches were conducted in *architecture*: According to (Bayazit 2004) and (Rocha 2004), Thomas A. Markus and Thomas Maver had started working on building performance at Strathclyde University, Peter Cowan initiated the building research center at the University of Sydney, as well as Herbert Simon and Alan Newell studied the intersection of design science and CAD at Carnegie Mellon University. Along this line, researchers contribute to a new research discipline named C*omputer-Aided Architectural Design* (CAAD).

In the 1980s, international organizations were formed to support academic conferences on CAAD research (Celani and Veloso 2015). Interestingly, these conferences were created in a way that a continent has its own regional organization. The Association for CAD in Architecture (ACADIA) was created in 1981 in North America. Education and research in CAAD in Europe (eCAADe) appeared in 1983. Also founded in Europe, CAAD Futures was created in the Netherlands.[1] Similar organizations were created in other continents a bit later. The Association for CAAD

[1] CAAD Futures tends to be organized *internationally* since it has been also held in North/South America and Asia.

Research in Asia (CAADRIA) was initiated in 1996, the Sociedad Iberoamericana de Gráfica Digital (SIGraDi) in 1997, and the Arab Society for CAAD (ASCAAD) in 2001. These six sister organizations form a major community of CAAD research (Celani and Veloso 2015).

As a result of the broad list of conferences, there have been plenty of research topics in the CAAD community, and some of them are even not restricted to architecture. For example, CAD (Chilton, Wester, and Yu 1993) and automated fabrication (Adel et al. 2018) are two classic research domains in architectural studies. In a wider sense, the principle of Computational Design is generalized to collaborative design (Achten and Beetz 2009), interactive design (Santos and Derani 2003), web design (Anders 1996), user experience design (Chen Rui and Aurel 2011), etc. These research topics lead to more generic design paradigms, including generative design (Abdelmohsen 2013) and design powered by artificial intelligence (Achten and Jessurun 2003). Eventually, researchers reach a very abstract end about computational thinking (Menges and Ahlquist 2011).

In my thesis, I aim to model Computational Design with a concentration on CAAD. My approach seems to be highly biased at first glance, but it does turn out to work properly. First, the research on CAAD takes a large portion of Computational Design, because it has a relatively long history (dated back to the 1980s), and many topics originated from architectural research (e.g., CAD). Second, a lot of research groups for Computational Design are affiliated with departments of architecture, such as the Design and Computation group at Massachusetts Institute of Technology's School of Architecture and the Design Computing Program at Georgia Institute of Technology's College of Architecture. Third, the international conferences on CAAD are generally not restricted on architecture, as ACADIA declares on its website,[2]

[2] See http://acadia.org/ for more details.

"ACADIA is an international network of digital design researchers and professionals (...) encouraging innovation in design creativity, sustainability, and education."

1.2 Interactive Visual Modeling

Computational Design is a relatively new, highly interdisciplinary, and rapidly evolving research field, thus it is quite difficult for researchers, especially the junior ones, to obtain insights about this field. When I entered the Master's program, I tried to search "Computational Design" on Google but obtained more than 3 million results and I could not find one that presents a clear, systematic, and up-to-date image of this field. Then I decided to review and model Computational Design as a research field by myself. Roughly speaking, in my thesis, I hope to learn the concepts, history, and trends in Computational Design by studying the following three research questions:

- How do the research topics evolve over time?
- How do the academic publications distribute geographically?
- How do the researchers and institutions collaborate?

Meanwhile, I want to study these questions in a systematic manner. I started with reading papers from the literature but soon realized that the quality of such manual reviews highly depends on the number of articles to be reviewed. To achieve my goal, I adopt a *bibliometric* approach, i.e., a *statistical* analysis of academic publications in Computational Design. Bibliometrics are commonly used to quantitatively analyze academic literature (De Bellis 2009; Price and De Solla Price 1976). Compared to manual reviews, this approach is more like a *distant reading* because it takes all existing academic publications into consideration, and is also more *systematic* because it requires developing a logically consistent statistical algorithm to collect, organize, and analyze the bibliographic data. In addition, bibliometrics present *quantitative* findings, e.g., in terms of the number of citations, co-authorships, etc. Recently, bibliometric analyses have been applied to modeling the landscape of Human-Computer Interaction based on publications from CHI conferences (Y. Liu et al. 2014). More related work is discussed in Section 1.3. It should be noted that the bibliometric approach should be accompanied by a manual review, in order to interpret and understand the findings from bibliographic data.

I base my work on a public online bibliographic database named *CumInCAD* (Martens and Turk 2003; Cerovsek and Martens 2016), which provides a cumulative index of publications related to CAAD. This index is *comprehensive* in the sense that it collects metadata of a large portion of papers published on CAAD-related conferences and journals, as well as a lot of technical reports and Ph.D. theses (Celani and Veloso 2015). CumInCAD is also reliable and up-to-date since the six sister organizations are still actively making large investments in it.

The first step of modeling Computational Design is to collect bibliographic data from CumInCAD. Because CumInCAD is an online database, I implement a web crawler to scrape data from its website.[3] Although this process seems straightforward, a significant challenge arises when I realize that there are multiple errors in the scraped data that hinder meaningful bibliometric analyses. For example, the data format could be inconsistent and some part of the data could be incomplete. I illustrate such data errors in detail and describe my data pipeline, especially the data clean process, in Section 2.

In order to study the three research questions proposed above, I conduct multidimensional interactive visual analyses based on the bibliographic data. Although the three questions look independent, *multidimensional* analyses can uncover more meaningful insights, e.g., the evolution of research topics in each

[3] CumInCAD is currently maintained at http://papers.cumincad.org/.

institution over time indicates the pioneers of different research directions. From the perspective of knowledge representation, *visualization* is desirable because it can provide informative and intuitive conceptualizations of the field. Moreover, *interaction* makes visual analyses even better because it can produce findings in an on-demand manner, e.g., a user can select a researcher to extract their collaboration networks. Concretely, I employ the *Latent Dirichlet allocation* (LDA) algorithm for topic modeling and perform several bibliometric analyses based on published time, authors, institutions, belonging topics, collaboration relations of the academic publications. I present the detail of these analyses and my findings related to the three research questions in Section 3, Section 4, and Section 5, respectively.

1.3 Related Work

Literature Reviews of Computational Design From the perspective of education, books about Computational Design such as (Menges and Ahlquist 2011; Agkathidis 2012; Leach and Yuan 2018) present professional, informative, and detailed manual reviews of the field. While the editors of these books summarize the field from their own research experience, my thesis is aimed to provide an overview via bibliometrics. (Celani and Veloso 2015) presents a brief history of CAAD-related conferences, which helps me figure out the scope of academic publications needed to establish a meaningful model of Computational Design. (Koutamanis 2005) also discusses the history of CAAD research using information about academic conferences. These overviews follow a manual approach and focus on the historical development of the field. (Martens and Turk 2003; Cerovsek and Martens 2016) found the CumInCAD database that provides the bibliographic data used in my thesis, but they just perform some straightforward quantitative analyses such as keyword frequency statistics.

Bibliometrics for Research Areas One of the most related work is (Turk, Cerovsek, and Martens 2001), which employs machine learning algorithms such as clustering to extract topics of CAAD from CumInCAD. Their work does not make use of state-of-art topic extraction algorithms such as LDA and indeed leaves a statistical analysis for future work. Another related work focuses on the evolution of topics in CAAD (Turk, Cerovsek, and Martens 2001). As for other research areas, (Y. Liu et al. 2014) utilizes multiple bibliometric analyses, including a co-word analysis, to reason about the intellectual landscape of the CHI conferences on the field of Human-Computer Interaction. (Suominen and Toivanen 2016) provides a Finnish scientific map based on unsupervised learning classifications and discuss the advantages and disadvantages of this approach over human reasoning. The automatic classification scheme can generate a classification model from the available text corpora, thereby is able to identify a trusted new knowledge body. They can also perform multi-functional large-scale data analysis and support a range of big data possibilities. Interestingly, some sub-disciplines of Computational Design have been reviewed via bibliometrics, such as *Building Information Modeling* (BIM) (Li et al. 2017) and *Green Building* (Li et al. 2017; Zhao et al. 2019). My work shares the same goal as these reviews but is aimed at a broader scope.

Interactive Visualization Technologies (Sievert and Shirley 2014) proposes a web-based interactive visualization technique for LDA analyses, which not only illustrates the scales of the extracted topics but also indicates the relevance among topics. A team from Indiana University researches on major information visualization about authors, papers and topics in the ACM Library, aiming to identify major research topics, co-authorships, and trends (Ke, Borner, and Viswanath 2004). They visualize co-authorships as an undirected graph where nodes represent authors where the size of a node corresponds to the number of publications of the associated author, and edges represent co-author relations where the thickness of an edge indicates the frequency of the associated collaboration. A lot of bibliometrics-based visualization tools have also been developed, such as CiteSpace (Cobo et al. 2011) and Sci2 (Lewis and Alpi 2017). All of these tools support co-citation analysis and keyword co-occurrence analysis (Z. Liu et al. 2015). In my work, I implement data processing and analysis scripts in Python, HTML and JavaScript and then make use of multiple visualization tools, such as Tableau, Photoshop, Gephi, and sigma.js, in order to obtain a beautiful and informative presentation of the findings.

2 Data Pipeline: Collecting Bibliographic Data

In this section, I describe the data pipeline to collect bibliographic data in Computational Design from CumInCAD. I follow a standard approach that integrates five key facets of using data (illustrated in figure 1):

1. **Question**: Put forward the right and answerable question.
2. **Collect data**: Data wrangling and sampling to get a suitable data set.
3. **Clean data**: Cleaning data and find errors.
4. **Explore data**: Generating hypotheses and building intuition.
5. **Use data**: Summarizing results through visualization, stories, etc.

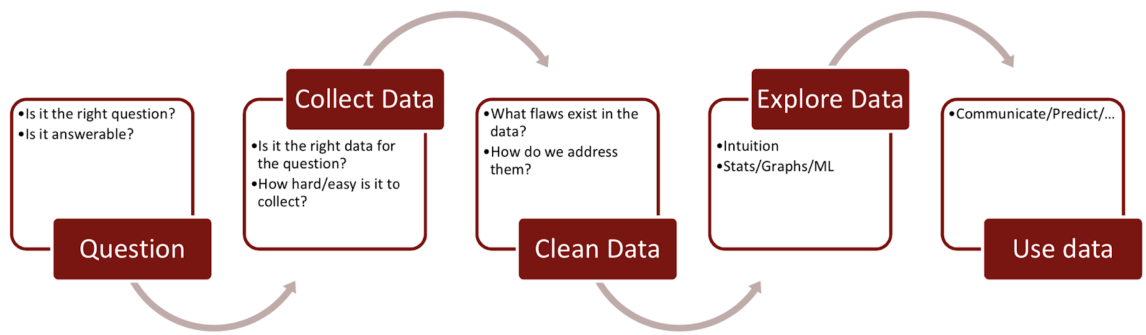


Figure 1. A standard data pipeline (from Stamper and Perer 2018)

First, I give an overview of the CumInCAD database in Section 2.1, which serves as the data source of my work. Then I describe the implementation of a web crawling script to scrape data from CumInCAD in Section 2.2. In Section 2.3, I elaborate the most challenging part of the data pipeline: Data cleaning. Finally, after resolving errors in the data, I conduct some quality checks on the obtained data in Section 2.4.

2.1 Data Sources

At the moment I accessed and obtained data from CumInCAD, there were 13,990 entries in the index, each of which stands for a CAAD-related publication. The database contains comprehensive metadata of academic conferences including the six major sister organizations: ACADIA, eCAADe, CAAD Futures, CAADRIA, SIGraDi, and ASCAAD (Martens and Turk 2003; Cerovsek and Martens 2016; Celani and Veloso 2015). For each academic publication, the database provides multiple useful fields, such as authors, year, title, source, summary, keywords and email addresses.

Table 1 shows the number of entries at CumInCAD at the moment I accessed the online database. Other conferences including Added Value of Computer Aided Architectural Design (AVOCAAD), Design and Decision Support System in Architecture and Urban Planning (DDSS), etc. All others include journal papers, theses, technical reports, and books. Papers from the six sister organizations take up about 85.25% of conference papers and about 68.68% of all the academic publications. The facts justify my claim in Section 1.1, which is also pointed out in (Celani and Veloso 2015), that the six sister organizations form a major community of CAAD research.

Table 1. Entries found at CumInCAD

Source	**#Publications**	**Source**	**#Publications**
ACADIA	1,475	SIGraDi	2,161
eCAADe	2,980	ASCAAD	302
CAAD Futures	828	Other conferences	1,662
CAADRIA	1,862	All others	2,720

2.2 Data Crawling

Data crawling is usually a complicated and subtle process because the websites might have unstructured HTML source code and inconsistent data format. Fortunately, the data in CumInCAD are well maintained such that I am able to write an automated script to scrape the data and store them locally.

I implement the data crawling script in Python with the Selenium tool[4], which provides web browser automation. The online CumInCAD database provides a search interface, so my script starts with searching publications that contain an empty string as a substring (i.e., these publications are indeed all the available publications). The CumInCAD website then returns the search results as a group of pages, each with which contains at most 20 entries. Then my script iterates through these pages and scrapes the entries on these pages. By manually investigating the HTML source code, I identify the data patterns and then use XPath to extract the desired information, such as the title, summary, authors of the publications. Finally, my script collects all the scraped data and stores them in an SQLite database.

My script took about one day to crawl all the 13,990 entries from CumInCAD. It is not very efficient but works properly for my purpose. One of the performance bottlenecks is that the Selenium tool invokes a real web browser, such as Firefox, to fetch data, which might lead to considerable overheads. A possible optimization is to substitute the Selenium package with some lightweight tools like Scrapy.[5]

[4] See https://www.seleniumhq.org/.
[5] See https://scrapy.org/.

2.3 Data Cleaning

Although CumInCAD provides a comprehensive and up-to-date database of bibliographic data in Computational Design, some of the data are unfortunately kind of messy and inconsistent, especially the author information. For example, figure 2 arranges names in the format “First Last” and separates them using a comma, while figure 3 uses semicolons as separators and a mixture of “First Last” and “Last, First” formats for names. Even worse, the same author might have different identifiers in the data, e.g., in figure 4, two papers from the same author record different author information, though they look quite identical.

authors A. Fatah gen. Schieck, A. Penn, V. Kostakos, E. O'Neill, T. Kindberg, D. Stanton Fraser, and T. Jones

Figure 2. Author information of the paper with id “DDSS2006-HB-467”

authors Abad, Gabriel; Adriane Borde; Mónica Fuentes; Virginia Agrielav; Adriana Granero and Jacqueline Fernández

Figure 3. Author information of the paper with id “sigradi2006_c129b”

authors Abadi Abbo, Isaac

authors Abadí Abbo, Isaac

Figure 4. Author information of papers with id “bacd” and “a9ca”. Note the difference between the English “i” and the Spanish “í”

I tried to clean the data to make author information consistent but failed. First, it is so time-consuming for me to manually check all the 13,990 publications. Second, I wrote more than ten small scripts, some of which are sketched in table 2, to identify and fix certain patterns of data inconsistency, which seemed to work for a lot of data, however, still left tons of dirty data with unrecognized patterns. Third, I experimented some human parsers, such as the Nameparser package[6] in Python,

[6] See https://nameparser.readthedocs.io/.

which are aimed to check whether two English names in different formats are the same, but these parsers cannot process non-English names properly.

Table 2. Some small scripts to parse names from the string of authors

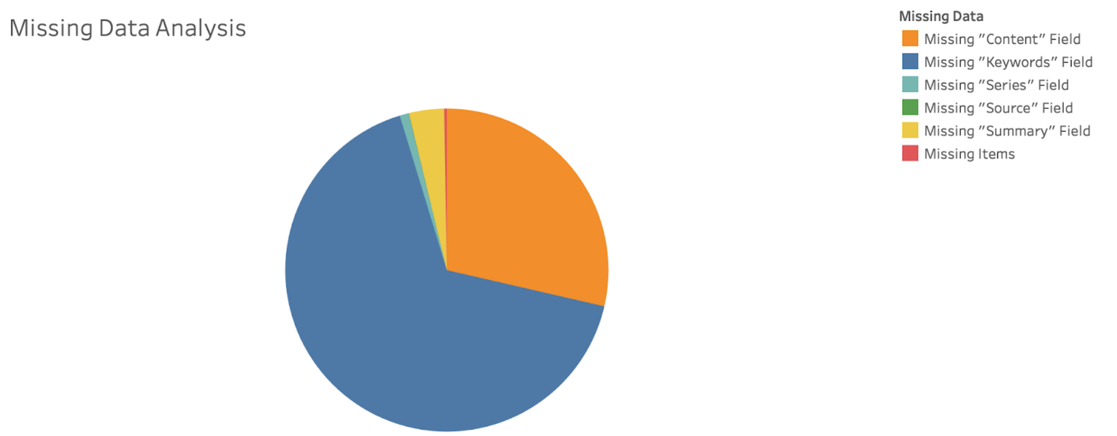
Script #1	If the string contains "_and_" or "_And_", then there is a name after such separators.
Script #2	If the string contains both commas and semicolons, then semicolons must be used as the separator.
Script #3	If the string does not contain commas, then the names are in the "First Last" format.
Script #4	If the string does not contain semicolons but commas, and the number of segments separated by commas is even, then each name should be in the "Last, First" format.
Script #5	Abbreviations should only occur in first/middle names. Make use of such information to tag different segments in the string.

Nevertheless, I found another online CumInCAD database,[7] which is, in fact, an out-of-date index of CAAD-related publications. A big advantage of this database against the latest CumInCAD is that it provides a reasonable list of author names and links each author to all their publications. Therefore, I decided to use both bibliographic indices, but for different purposes. On the one hand, I conduct topic modeling and geographic analyses using the latest database because these analyses do not require accurate information on author names. On the other hand, I analyze collaboration relations using the other slightly out-of-date database, because it provides more accurate information for the authors. To make sure that the analysis on the out-of-date database is meaningful, I perform several checks to justify the database's quality with the latest CumInCAD as an oracle. Details of the checks are shown in Section 2.4.

[7] The database is located at https://cumincad.architexturez.net/.

2.4 Data Quality

Currently, the CumInCAD database still lacks some information of the academic publications. In other words, the data have some missing values. I evaluate the completeness of the data by counting missing items and as shown in figure 5, only about 2.44% of the publications don't have a summary. As shown later in Section 3.1, the information about content, keywords, series, and sources is not needed in topic modeling. Therefore, I claim that the data is reasonably complete.



Missing "Content" Field	Missing "Keywords" Field	Missing "Series" Field	Missing "Source" Field	Missing "Summary" Field	Missing Items
2,785	6,494	88	3	342	26

Figure 5. Missing data analysis

Besides completeness, it is also very important to validate the correctness of the data. I follow a common principle in statistics—*random sampling*—to check the correctness of the data. In short, I randomly picked 100 researchers and 100 publications and checked whether they are correct with respect to online records about them. I repeated the process for 5 times and found no errors. Therefore, I get highly confident in the correctness of this bibliographic index.

Finally, as mentioned in Section 2.3, I also use an out-of-date version of CumInCAD in some analyses, so I need to justify the quality of this database. In the scraped data, there are 11,815 publications and 10,105 researchers, which means that this index is about 84.45% complete, compared to the latest CumInCAD. Though the percentage does not look very good at first glance, the index lacks a lot of latest publications and I instead compute the completeness of the publications from time frames provided by this index (presented in table 3). As a result, the index is about 99.22% complete for the six major sister organizations, compared to the latest CumInCAD. Therefore, I am confident that the out-of-date database is also reasonably complete, at least at the time it stopped maintenance.

Table 3. Time frames of six sister organizations in the out-of-date CumInCAD

Conference	ACADIA	eCAADe	CAAD Futures	CAADRIA	SIGraDi	ASCAAD
Year before	2012	2014	2011	2014	2013	2012

3 Topic Analysis

The first research question is to discover *what people in Computational Design have been working on*. As proposed in (Price and De Solla Price 1976), the literature of a research discipline can be categorized as *persistent* and *transient* elements. The persistent part, often known as *knowledge domains*, characterizes structural information of unstructured contents of the research. The transient part, often known as *knowledge evolutions*, describes the dynamic development of the research discipline. In this section, I follow the principle above and conduct a topic analysis in the field of Computational Design. In Section 3.1, I apply state-of-art topic extraction algorithms to model knowledge domains for Computational Design. Then in Section 3.2, based on the topic model, I analyze the knowledge evolution pattern in Computational Design.

3.1 Topic Modeling

My goal is to extract a topic model from the bibliographic data in CumInCAD. The major idea is that the database contains titles and summaries of almost all the publications, and my hypothesis is that it should be possible and reasonable to cluster publications by solely looking at the texts of titles and summaries. Before performing the analysis, I review the manual topic extraction in (Martens and Turk 2003) and summarize them as four sub-disciplines: computation technologies, design processes, applications, and society-related topics. Table 4 presents a priori categorization of topics into these four sub-disciplines.

Table 4. A priori knowledge domains of Computational Design

Computation Technology	**Design Process**	**Application**	**Society**

• Animation • Artificial Intelligence • Constraint-Based Design • Database Systems • Design Methodology • Generative Design • Human-Computer Interaction • Image Processing • Shape Grammars	• 2D Representation • 3D Modeling • Case-Based Reasoning • Collaborative Design • Communication • Design Process • Interactive Design • Performance Simulation • Visualization • Web Design	• 3D City Modeling • Computer Integrated Construction • Digital Media • Knowledge Modeling • Object-Oriented Modeling • Virtual Design Studio • Virtual Reality	• Digital Design Education • Environmental Simulation • Learning Environment • Virtual Environments

The task of topic clustering from texts is a classic task in Machine Learning. One of the state-of-art commonly used techniques is the *Latent Dirichlet Allocation* (LDA) algorithm (Blei, Ng, and Jordan 2003). Given a collection of documents as input, the idea of LDA is to view each document as a mixture of possible topics. LDA assumes that each document is treated as a bag of words, and there are a small number of topics as well as these topics only use a small set of words frequently. Formally, LDA is a generative probabilistic model for text corpora (Hoffman et al. 2013). Figure 6 illustrates the three-level mathematical structure of LDA, where D stands for a collection of documents, each document $d \in D$ contains N_d words, and K is the collection of topics. Each node in the graphical representation is associated with a random variable, while *observed* ones are shaded and *latent* variables are unshaded. The vector α is a model parameter corresponding to a prior distribution of topics, and θ_d, sampled from a Dirichlet distribution parameterized by α, represents topic proportions for the document d. By the definition of Dirichlet distributions, θ_d is also vector-valued and has the same length as α. Intuitively, θ_d gives a prior distribution on the topic of the document d. Next, for each word n in the document d, the random variable $z_{d,n}$, drawn from a multinomial distribution parameterized by θ_d, represents a topic assignment for the word. In order to generate a concrete word, i.e., the only observed data $w_{d,n}$, a distribution over the words for a specific topic is needed. Therefore, for each topic $k \in K$, a distribution β_k on words is sampled from a Dirichlet distribution parameterized by a prior η. Then the observed word $w_{d,n}$ is drawn from the distribution $\beta_{z_{d,n}}$. Now it becomes clear that the boxes in figure 6 represent repeated sampling processes. The goal of LDA is then to infer the latent structures based on the observed text corpora.

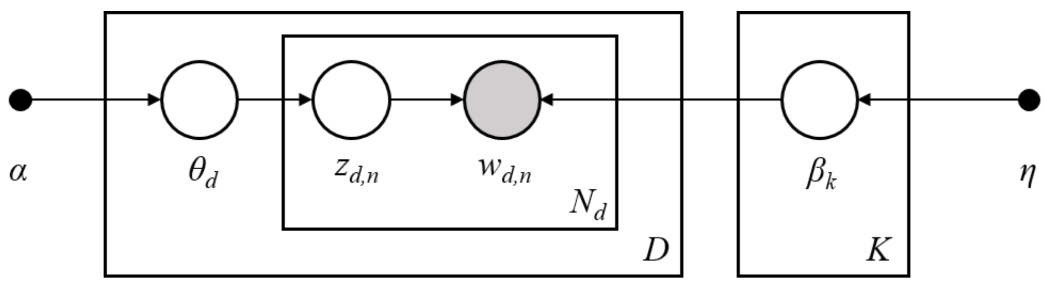


Figure 6. The graphical representation of LDA (from Hoffman et al. 2013)

To apply the LDA algorithm to extract topics from bibliographic data of Computational Design, I set up a document for each publication by concatenating its title and summary. To make the analysis more reasonable, I further remove stop words and frequently used words in English, get rid of all the punctuations, retain only the nouns and adjectives, as well as perform *lemmatization* that converts words to its root.

I implement the topic modeling algorithm on CumInCAD using Python with the scikit-learn package[8] for machine learning and also the spaCy package[9] for natural language processing. After several iterations of parameter searching (for the learning rate, mini-batch size, number of topics, etc.), I ended up in categorizing the publications into 21 topics. Figure 7 visualizes the clustering results using the pyLDAvis package[10] in Python, which implements the technique from (Sievert and Shirley 2014), where circles represent research topics, the size of a circle reflects the number of publications under the associated topic, and the distance of two circles denotes the relevance of two corresponding topics. The inter-topic distances is computed by the Jensen-Shannon divergence, which is based on Kullback-Leibler divergence from information theory. It should be noted that the LDA algorithm only clusters documents by their topics, but does *not* tell you what the topics actually are. The situation reflects my claim in Section 1 that manual analysis is still needed to interpret results from bibliometrics. After looking into the publications in the clusters, I assigned a dominant topic to each cluster. As you can see in figure 7, some top topics are CAAD Education, Generative Design, and Knowledge Representation.

[8] See https://scikit-learn.org/.
[9] See https://spacy.io/.
[10] See https://pyldavis.readthedocs.io/.

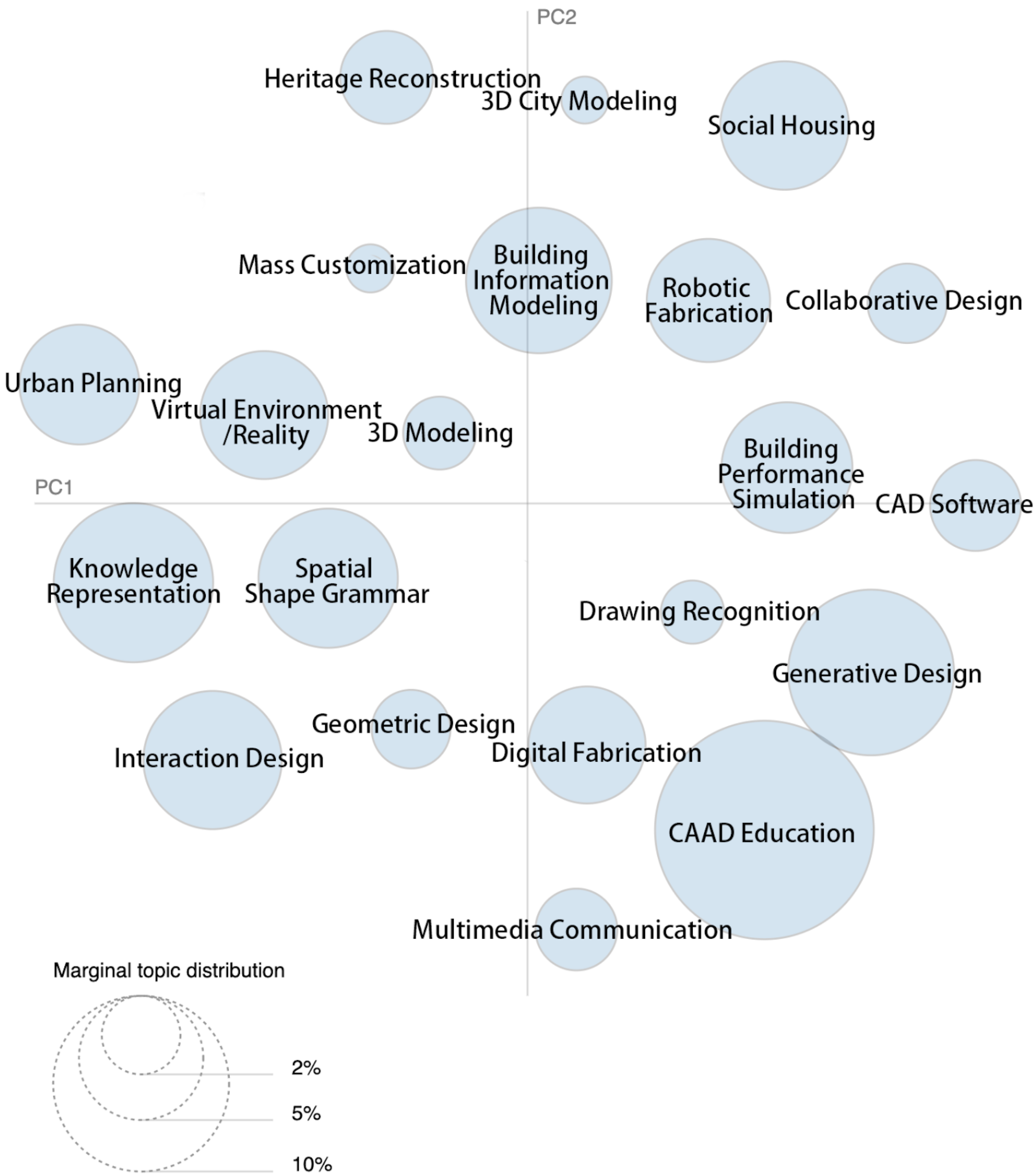


Figure 7. Results of the LDA analysis on all CAAD-related publications

Recall that before performing topic extraction I came up with a priori categorization of topics in Computational Design as shown in table 4. Now I hope to check to what extent my hypothesis that the LDA algorithm can reasonably identify topics from tons of papers is correct. Table 5 presents the knowledge domains coming from the

LDA analysis in the same format as table 4. It turns out that the bibliometric method works pretty well as 12 out of the recognized 21 topics coincide with manual reviews in (Martens and Turk 2003). Building Information Modeling (BIM), etc. are also quite meaningful. Intuitively, some are new or trending topics in recent years. For example, BIM started to receive much attention around 2002 (Li et al. 2017), and User Interface Design becomes popular with the rapid development of smartphones since 2007. Also, there are some other interesting trends such as that more society-related topics arise, e.g., heritage reconstruction and social housing. A more detailed analysis of the evolution of these topics is presented in Section 3.2.

Table 5. Knowledge domains of Computational Design by bibliometrics

Computation Technology	**Design Process**	**Application**	**Society**
• CAD Software • **Drawing Recognition** • **Generative Design** • Geometric Design • **Special Shape Grammars**	• **3D Modeling** • Building Performance Simulation • **Collaborative Design** • User Interface Design • **Multimedia Communication**	• **3D City Modeling** • Building Information Management • **Digital/Robotic Fabrication** • GIS and Urban Planning • **Knowledge Representation** • **Virtual Environment/Reality**	• **CAAD Education** • Heritage Reconstruction • Mass Customization • Social Housing

There are some limitations to the LDA analysis and the visualization process. Although figure 7 plots the circles with respect to the relevance among the associated topics, they do not seem to follow the paradigm with four sub-disciplines as I proposed earlier. One possible reason is that the LDA analysis is based on word frequencies but does not reason about the *semantics* of the words rigorously. Nevertheless, delicate account for word semantics is also a challenging research problem in the field of natural language processing.

3.2 Topic Evolution over Time

Besides topic modeling, the other important aspect of a research literature review is to study topic evolution patterns in the research field. The evolution patterns can help uncover trends of research and provide directions for future studies. As for Computational Design, my hypothesis is that one trend should align with the evolution of computational technologies, such as cloud computing and artificial intelligence. For example, Generative Design should be becoming popular because the increasing computational power should make it possible for algorithms to explore a larger search space for designs. Another trend could be about society related applications like urban planning, green building, etc. Despite the widely used citation burst detection (Kleinberg 2002), I adopt a fully visual approach to observe topic evolution over time. I use Tableau Public[11] to plot trend charts on the number of publications under each research topic on the time axis. Tableau Public can not only provide informative visualization but also support user interaction such that the user can click on the broken lines in the charts to view detailed information, such as the belonging topic and the accumulated number of publications.

[11] See https://public.tableau.com/.

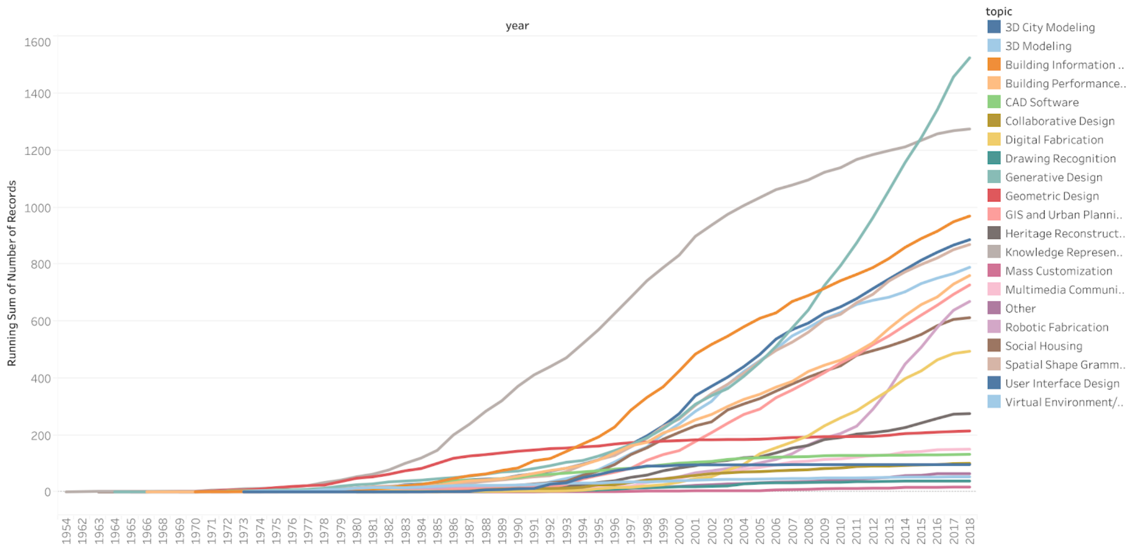


Figure 8. Topic evolution over time by counting the number of publications

Figure 8 presents the visualization of the overall topic evolution in Computational Design.[12] The broken lines illustrate the accumulated amounts of publications in those topic categories on a time axis. The chart uses different colors for different topics and picks colors carefully to produce a beautiful presentation. Moreover, the visualization is also *interactive*. One can click on either a curve in the chart or a topic in the legend to highlight the trend for a topic. For example, I select the topics with increasingly fast development and demonstrate them in figure 9. Intuitively, the curved lines of these topics represent *convex* functions. There are four such topics: Generative Design, Building Performance Simulation, GIS and Urban Planning, as well as Robotic Fabrication. Generative Design is the topic with the most significant trend. I think the reason for this fact is two-fold. First, the development of computational power makes it possible to develop generative algorithms to explore and synthesize more design prototypes from the search space in a fixed amount of time. Second, the recent evolution in artificial intelligence,

[12] CAAD Education is excluded because the number of publications under this topic is substantially larger than the number of publication under any other topic. Removing CAAD Education from the chart makes the trends more clear.

especially those on deep learning, e.g., (Silver et al. 2016, 2017), computer algorithms can explore the search space for the design more strategically and more efficiently. A similar trend is observed for the topic of Robotic Fabrication because the field of Robotics also benefits from the advances in artificial intelligence. Another topic that becomes increasingly popular is GIS and Urban Planning, including researches on *smart cities*. Overall, I summarize the trends in the research of Computational Design as the following three bullets:

- **Algorithmic Design**: Take advantage of recent achievements of artificial intelligence, especially deep learning.
- **Computer-Aided Construction**: Build robots for automated/semi-automated fabrication.
- **Applications with Social Impacts**: Lift the scope of design to larger-scale objects, such as city construction.

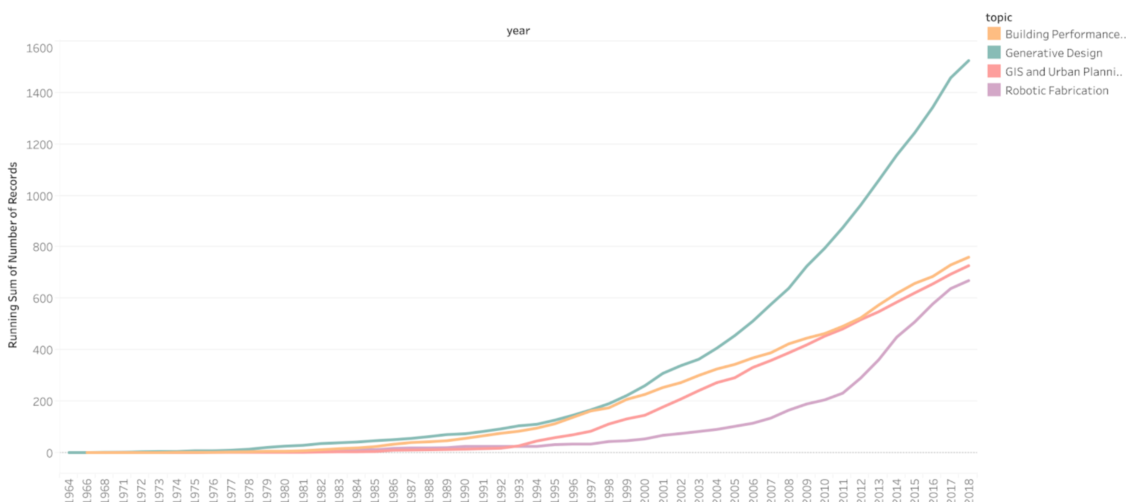


Figure 9. Topic evolution over time for trending research problems

Besides the trending topics, I am also interested in the topics that expose a slower or a constant development speed in recent years. I identify the topics with those features and plot them in figure 10 and figure 11, respectively. Intuitively, a topic with a milder trend could be seen as well studied research areas and the research ideas of the topic are getting saturated. On the other hand, a topic with a constant development speed should still be interesting for researchers to work on. From the visualization, I conclude that CAD Software, Geometric Design, Knowledge Representation, and Virtual Environment are less trending than other topics (and indeed these topics are classic ones in Computational Design), and Building Information Management (BIM), Digital Fabrication, Heritage Reconstruction, Social Housing, Spatial Shape Grammars, and User Interface Design are still good directions to conduct research in the future.

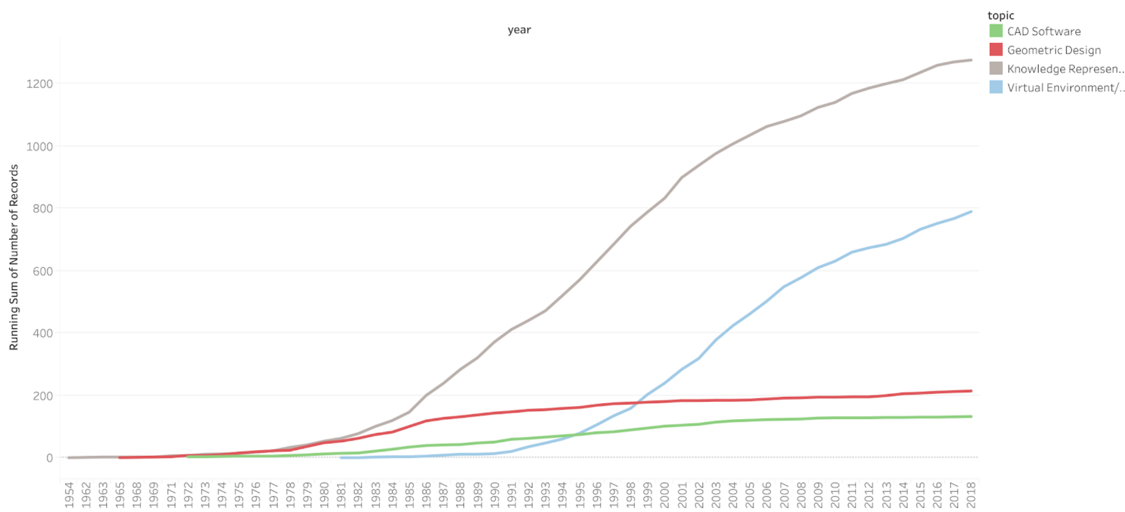


Figure 10. Evolution of topics with a slower development speed recently

In Section 4.3, I again apply the trend visualization technique to analyze a research institution's interest evolution over time. This kind of interactive visualization proves to be powerful and informative when multiple dimensions (e.g., time, location, topic) are taken into consideration since it can structure data to provide insights in a more intuitive way and allow users to zoom in specific parts of the data to gain more information.

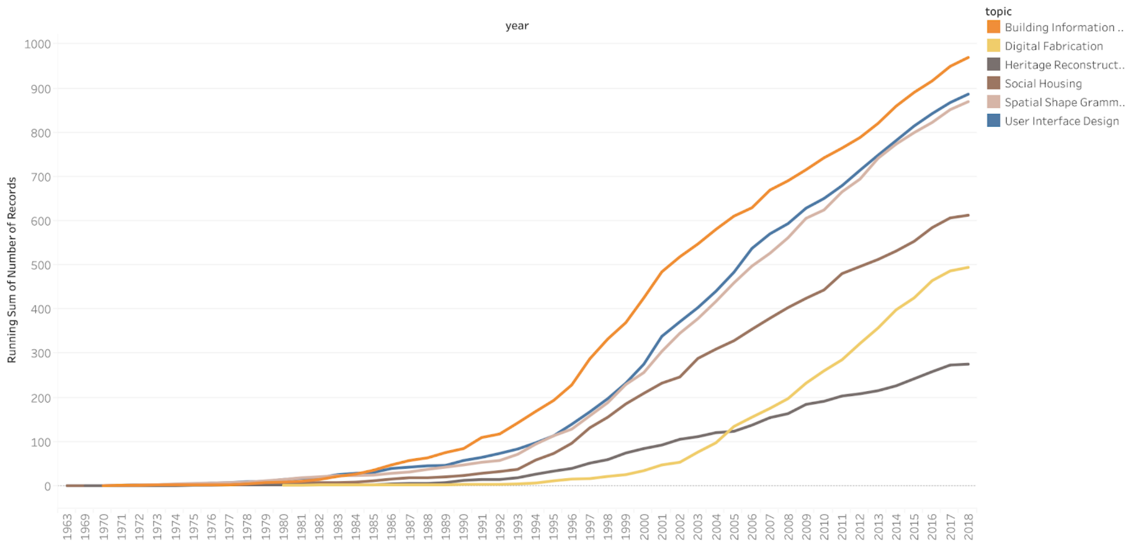


Figure 11. Evolution of topics with a constant development speed recently

4 Geographic Distribution: Country/Region and Institutions

The second research question is to study *how research publications in Computational Design distribute around the world*. The reason why this question is interesting is two-fold. First, it can help reveal research institutions with great contributions, as well as identify pioneers among them. Second, combining with topic modeling described in Section 3, it can uncover favored research topics of different institutions. These kinds of findings can help junior researchers seek collaboration opportunities at proper institutions with matched research interests. In Section 4.1, I elaborate on the process to obtain geographic data associated with academic publications. With the data, I conduct geographic distribution analyses on both a country/region level and an institutional level in Section 4.2. As a meaningful and informative multidimensional extension of the geographic analysis, I utilize topic evolution modeling to discover research profiles of some institutions in Section 4.3.

4.1 Obtaining Geographic Data

Although the CumInCAD database provides a lot of information about researchers, they do not contain the information about institutions that the researchers belong to. Moreover, researchers might change their institution from time to time, therefore, it is not always reliable to scrape their current information from their websites. Fortunately, I manage to sidestep the problem because the last CumInCAD database includes email addresses of the authors. Thus, I am able to determine each researcher's belonging institution by looking up their email addresses in an online database of university domains and names.[13] Though I succeeded in implementing a Python script that uses the database to resolve most of the institutions in the United States, there are still lots, but a manageable number of unresolved email addresses from other countries. I then resolved all these email addresses manually by Google Search.

As a result, I identify 648 academic institutions from 64 countries across the world, and 7,324 out of the 13,990 publications recorded in CumInCAD are affiliated with at least one author from academic institutions.

4.2 Geometric Distribution of Academic Publications

To start with, I am interested in *which countries/regions contribute most to the research of Computational Design*. My hypothesis, based on the history of the development of Computational Design sketched in Section 1.1, is that Australia and the United States should be two candidates because they are home of several institutions that pioneered the field of Computational Design. To model the contribution of a country/region or an institution, for each publication, I accumulate the percentage of the number of authors from the country/region or the institution. In other words, for a publication, if k out of n authors come from the same affiliation, the affiliation obtains k/n points in its score.

[13] The database is accessible at https://github.com/Hipo/university-domains-list.

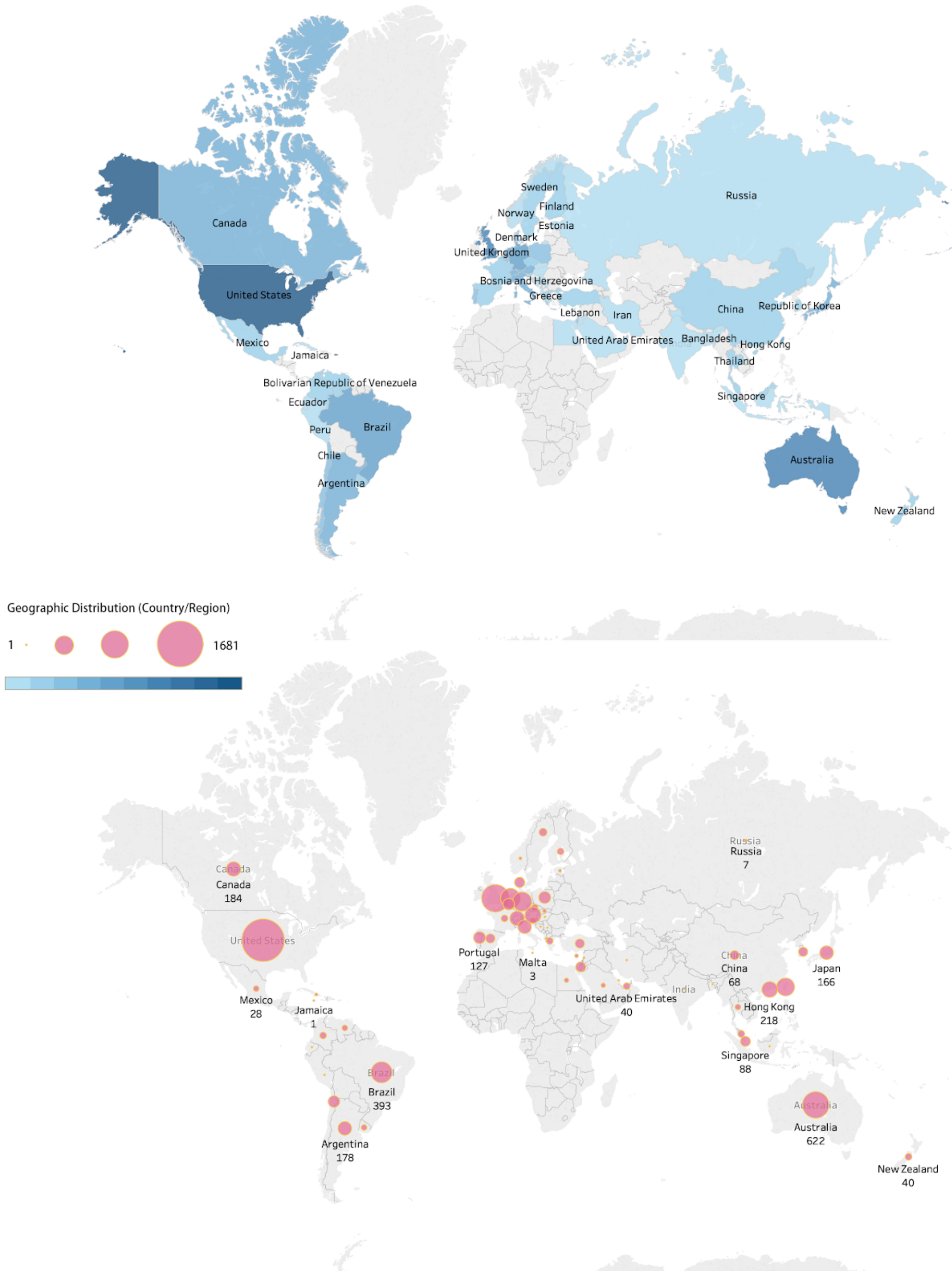


Figure 12. A heat map of publications from research institutions in countries/regions

Figure 12 illustrates a heat map on the globe, generated in Tableau Public, to demonstrate the contribution of countries/regions. I use two kinds of visualization in the heat map. Red circles whose radius is determined by the calculated scores are placed on the map, i.e., a bigger circle indicates a larger number of publications. Also, the countries/regions are colored by a stepped blue palette. The color depth is determined by the square root of scores. With the two metrics visualized on the same figure, it becomes easy to identify countries/regions that have a great contribution. Indeed, the United States and Australia are two countries with significant publication scores. This finding aligns with the history of Computational Design that the University of Sydney in Australia and Carnegie Mellon University, Massachusetts Institute of Technology in the United States are some of the first institutions to combine computation technologies in the design process.

However, because there are a lot of countries/regions in Europe, figure 12 does not give a clear presentation about the contributions of these countries/regions. Therefore, I adopt another visualization technique to generate a treemap based on the publication scores. Figure 13 demonstrates a treemap of publications from research institutions in countries/regions. Accompanied by figure 12, I find out that the United Kingdom, Netherlands, and Germany are countries/regions with top contributions in Computational Design in Europe. Another fact is that Brazil stands out in South America, as well as Taiwan has a significant contribution in Asia.

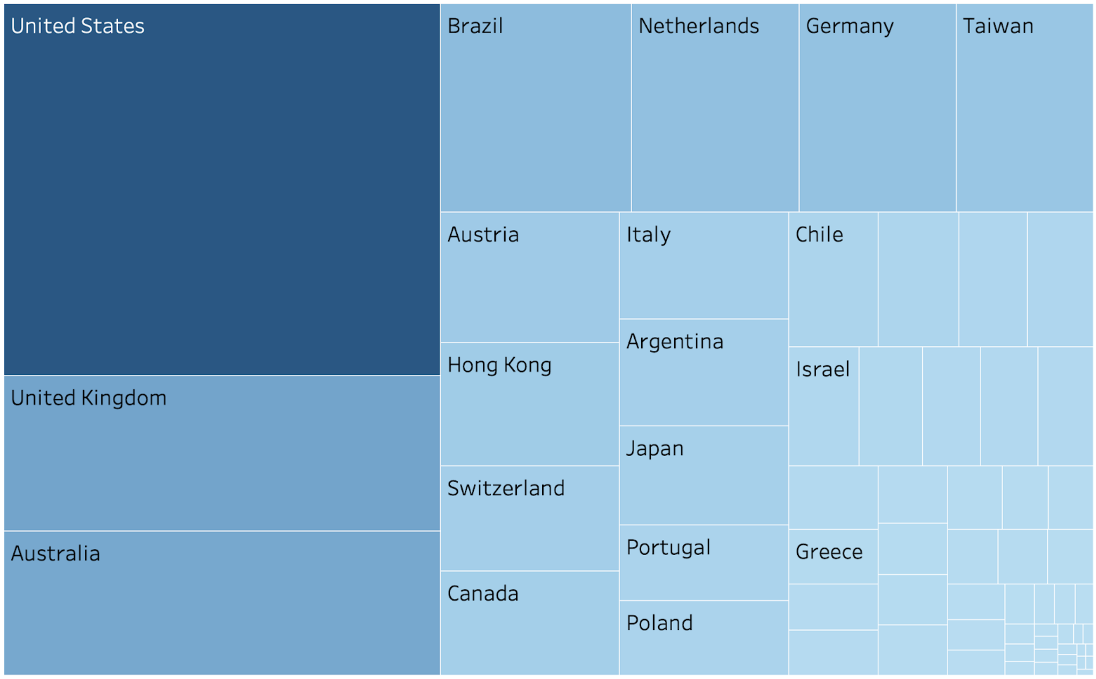


Figure 13. A treemap of publications from research institutions in countries/regions

Country/Region level statistics seems too abstract to derive interesting findings, so I zoom in the map and investigate geographic distribution of publications in Computational Design at an institutional level. Using the same metrics introduced above, I generate a treemap for research institutions directly, as shown in figure 14. The University of Sydney is significant exactly as I suppose. It is also noted that the University of Strathclyde in the United Kingdom, the Delft University of Technology in the Netherlands, and ETH Zurich in Switzerland are some of the productive research institutions in Europe, and besides Massachusetts Institute of Technology and Carnegie Mellon University, Georgia Institute of Technology is another active university that research on Computational Design in the United States. Indeed, I find Georgia Institute of Technology has a Design Computing program in their School of Architecture.

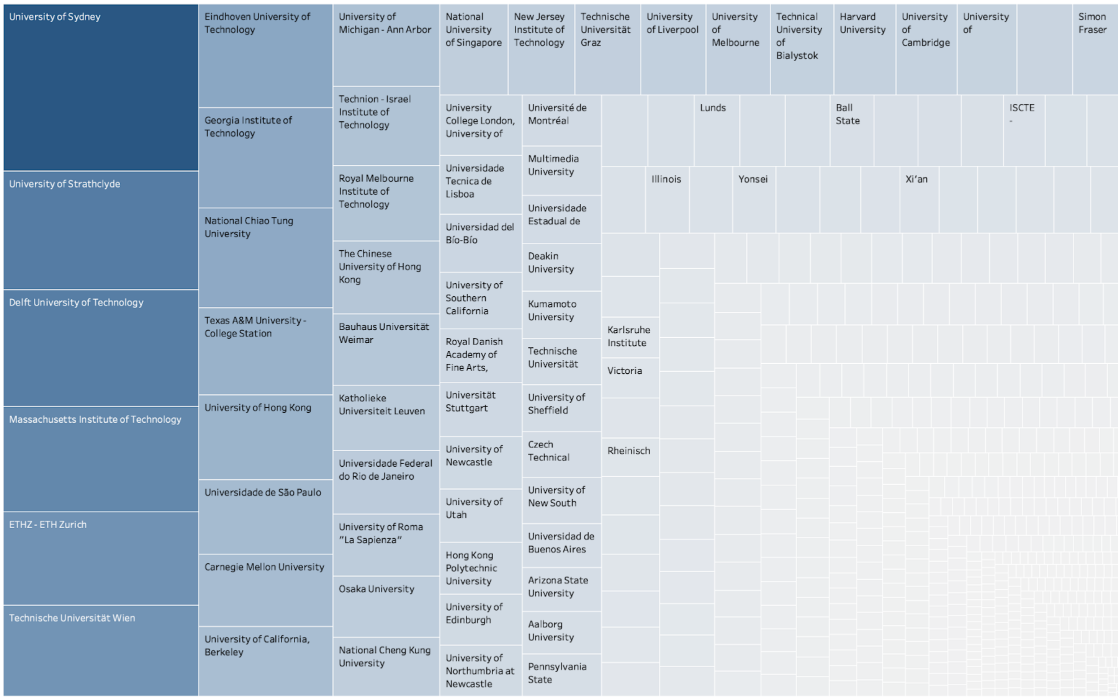


Figure 14. A treemap of publications from research institutions across the world

4.3 Adding a Time Dimension

At this point, I am not quite satisfied with these results. Although the findings point out outstanding research institutions around the world, I find it more informative to discover the research topics in the field of Computational Design in which the institutions are interested and even identify the research trend in the institutions. Thus I conduct an analysis by combining techniques from Section 3 on topic modeling. The multidimensional integration is quite straightforward, as I just need to filter out the publications by the affiliations of their authors and visualize the trend in the same way as used in Section 3.2. In this section, I demonstrate and discuss findings for the University of Sydney, the University of Strathclyde, Massachusetts Institute of Technology, and Carnegie Mellon University.

Figure 15 shows the research trend in Computational Design at the University of Sydney. As a pioneer of the field of CAAD, classic research topics such as Knowledge Representation and Spatial Shape Grammars have received much attention. Some samples of papers on knowledge-based design are (J. S. Gero and Rosenman 1990; J. S. Gero 1990), and some on shape grammars are (J. Gero and Jun 1995; J. S. Gero and Cha 1997). Professor John S. Gero, affiliated with many of these publications, was indeed formerly Professor of Design Science and Co-Director of the Key Centre of Design Computing and Cognition at the University of Sydney, which was one of the earliest organizations to research on Computational Design.

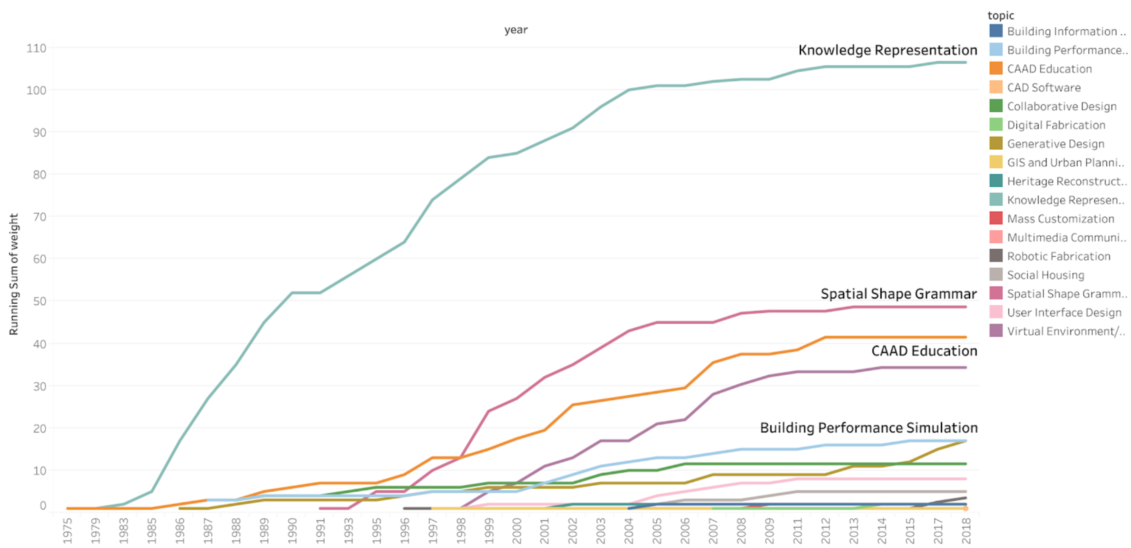


Figure 15. Topic evolution for research at the University of Sydney

Figure 16 shows the research trend in Computational Design at the University of Strathclyde. CAAD Education takes up a large portion of the papers, and many of them are published on eCAADe. Professor Alan Bridges had been working in that direction for years (Bridges 1994, 1986, 1997, 2006).

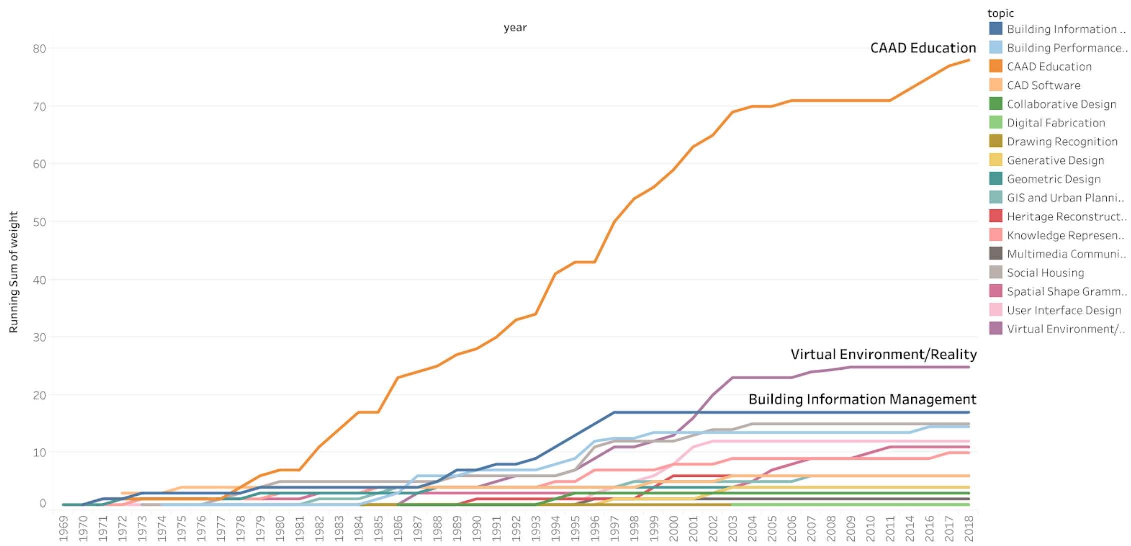


Figure 16. Topic evolution for research at the University of Strathclyde

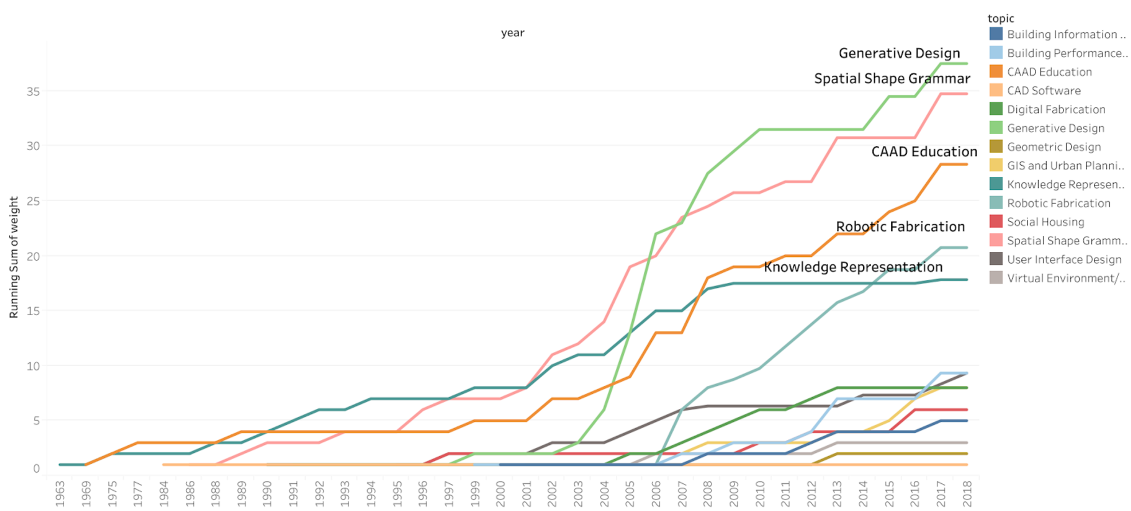


Figure 17. Topic evolution for research at Massachusetts Institute of Technology

Figure 17 shows the research trend in Computational Design at the Massachusetts Institute of Technology. Classic research topics including Shape Grammars and Knowledge Representation are significant and start to receive attention in the early years. Two special trending research topics, compared to other universities, are Generative Design and Robotic Fabrication. As one of the top research institutions in Computer Science, architecture researchers at Massachusetts Institute of Technology seem to benefit from state-of-art research about algorithms and robotics. Some papers on the algorithmic direction, including constraint-based and parameterized design, are (Kilian 2006b, [a] 2006; Loukissas and Sass 2004). As for the intersection of robotics and architecture, some examples are (Ariza and Gazit 2015; Staback et al. 2017).

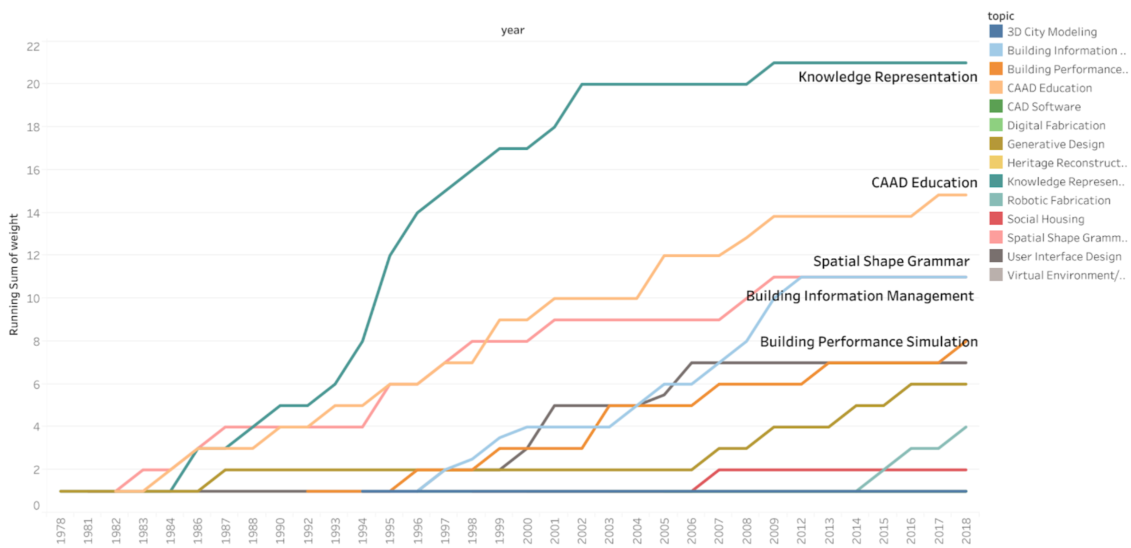


Figure 18. Topic evolution for research at Carnegie Mellon University

Figure 18 shows the research trend in Computational Design at Carnegie Mellon University. Similar to former universities, classic research topics such as Knowledge Representation and Shape Grammars have been interesting for the researchers there. Also, benefiting from close cooperation with the Department of Civil Engineering, Building Information Management and Building Performance Simulation receive researchers' attention as well. For example, (Kui and Krishnamurti 2007) builds a system to predict interior structures of a building and

(Park et al. 2013) performs energy performance modeling to reason about buildings' efficiency and sustainability.

5 Collaboration Analysis: Individuals, Institutions, and Countries/Regions

The third research question is to discover *how researchers in Computational Design collaborate among individuals, institutions, and countries/regions.* Such analyses can uncover collaboration patterns in the community and help find out "core" researchers. As mentioned in Section 2.3, I use an out-of-date version of CumInCAD to perform collaboration analysis because the database contains a consistent and correct list of author names, which is crucial for the analysis.

5.1 Individual-Level Collaboration

There are 10,105 authors and 23,870 authorships (author to paper) in the database. To visualize the huge network is one of the biggest challenges in my thesis. I utilized Gephi[14] (for graph construction) and sigma.js[15] (for graph visualization) in my implementation of the visualization process. In the network (shown in figure 19), one node is for one author, and the edge between two nodes represents that they had collaboration in the past. The closer two nodes are, the more times of collaboration they have. The color and size of a node represent his/her co-author times. The larger the node and the darker the color describe the more collaboration of the author. The intuitive visualization derives a clean presentation of the collaboration network such that it is easy to identify researchers with a large number of collaborations. The user is able to interact with the visualization by clicking on the graph to select a node and then highlighting the sub-part of the graph "centered" at

[14] See https://gephi.org/.
[15] See http://sigmajs.org/.

the selected node. For example, figure 19 is generated by clicking on Mark Burry and David Kurmann, respectively.

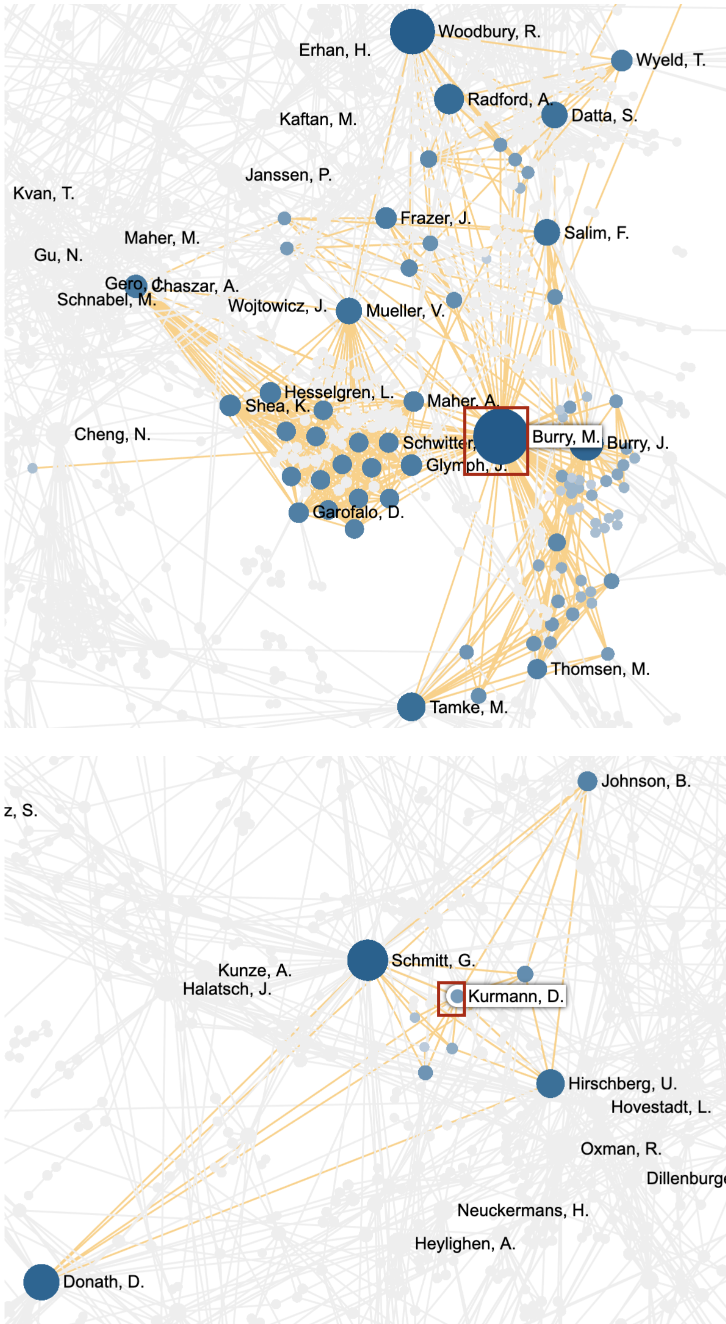


Figure 19. Co-author network graphs, where the left is obtained by selecting an author with plenty of co-authorships, and the right by picking an author with fewer co-authorships

Besides discovering the collaboration pattern for a specific researcher, the visualization also exposes some phenomenons in the community. For example, figure 20 demonstrates two highly dense sub-graph in the collaboration network. For the researchers in the sub-graphs, they usually collaborate with each other. Nevertheless, most of the co-author network is not that dense so it indicates that researchers in the community have been actively collaborating to advance the field of Computational Design.

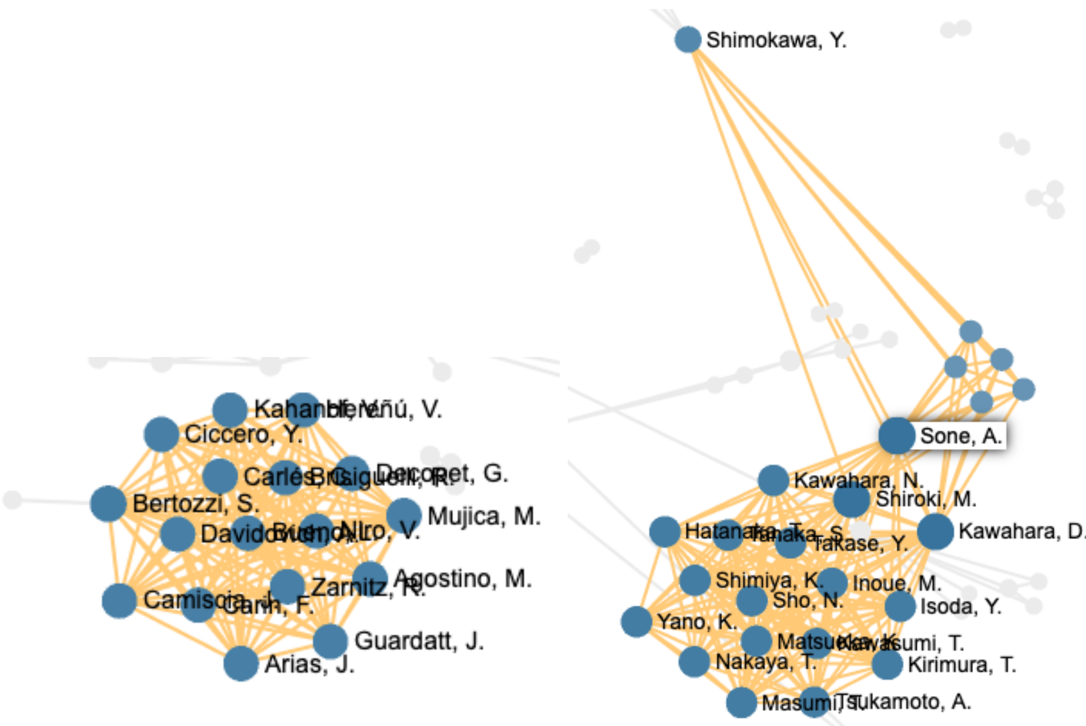


Figure 20. Two dense sub-graphs in the individual collaboration network

5.2 Institution-Level Collaboration

Using similar techniques to Section 5.1, I construct an institutional collaboration network. Instead of using the sigma.js tool for visualization, I exported figure 21 directly in Gephi. As before, the color and size of a node represent the collaboration frequency of the associated institution and the thickness of an edge denotes the collaboration times of two connected institutions. However, to achieve a better presentation of the network, I had to manually adjust some of the node positions to reduce overlaps among node labels as well as to make it more clear to observe the connections.

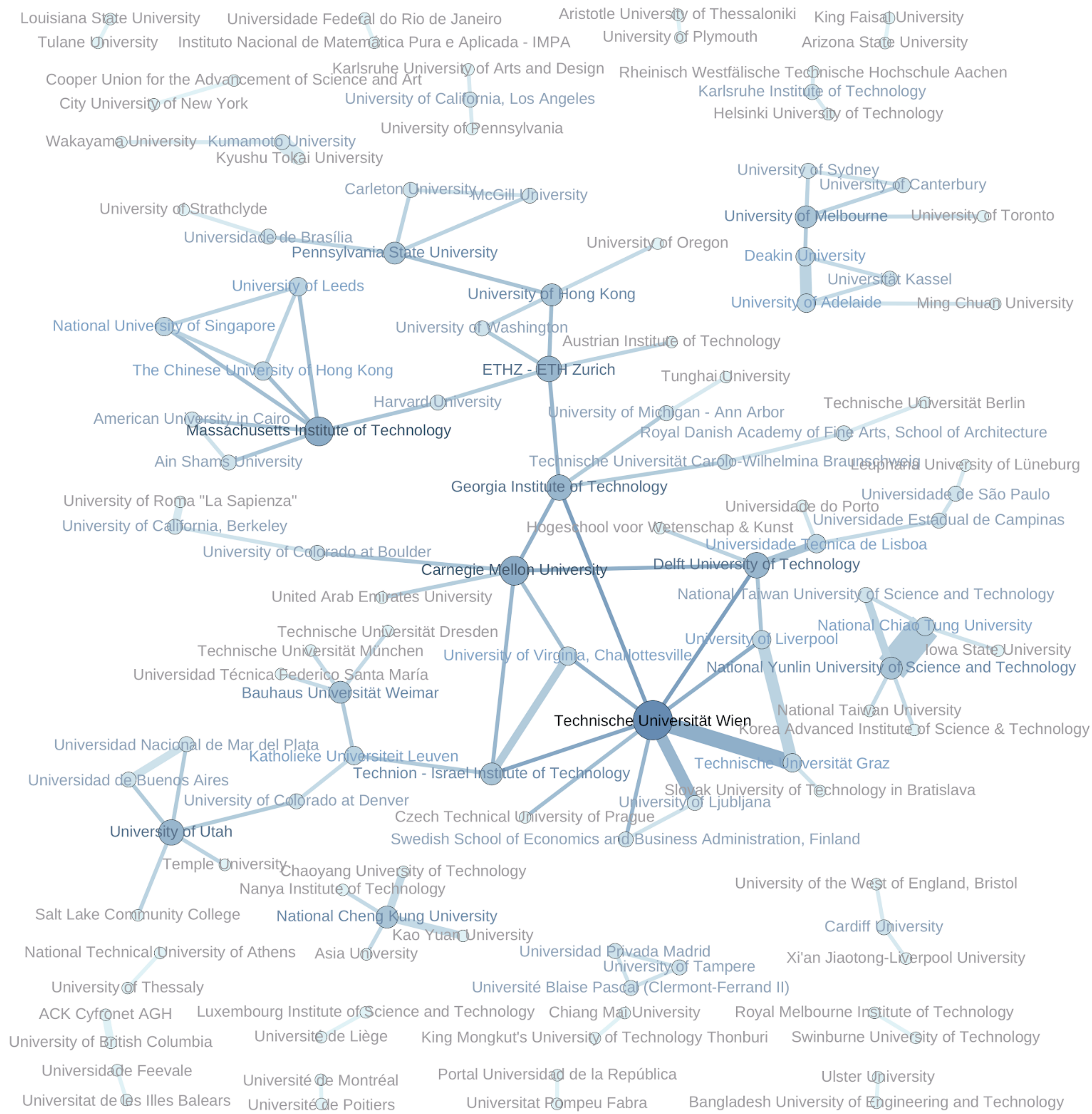


Figure 21. A collaboration network across research institutions

It should be noted that institutions with a larger number of publications do *not* necessarily collaborate more with other institutions. Compared with the findings in Section 4.2 and Section 4.3, the University of Sydney and the University of Strathclyde have lots of academic papers but are *not* significant in figure 21. On the other hand, Massachusetts Institute of Technology and Carnegie Mellon University are among the research institutions that are both productive and collaborative. In addition, Georgia Institute of Technology, ETH Zurich, and Technische Universität Wien are also very collaborative in the community.

5.3 Country/Region-Level Collaboration

By the same means as Section 5.2, I construct a country/region level collaboration network, which is demonstrated in figure 22. The number of nodes at this level is much less than that at the institutional level, and as a consequence, the default visualization provided by Gephi is intuitive and informative enough.

The United States is the largest node in the graph. Clearly, this fact could be caused by that the United States has several pioneer institutions in the field of Computational Design, as well as has established a highly inclusive and collaborative academic community. After I zoom in the graph a little bit and highlight the part of the graph centered at the United States, as shown in figure 23, I find out a very thick academic collaboration edge between the United States and Switzerland. This fact aligns pretty well with the institutional collaboration network, as many universities in the United States have been working with ETH Zurich, one of the top research institutions all over the world that is located at Switzerland.

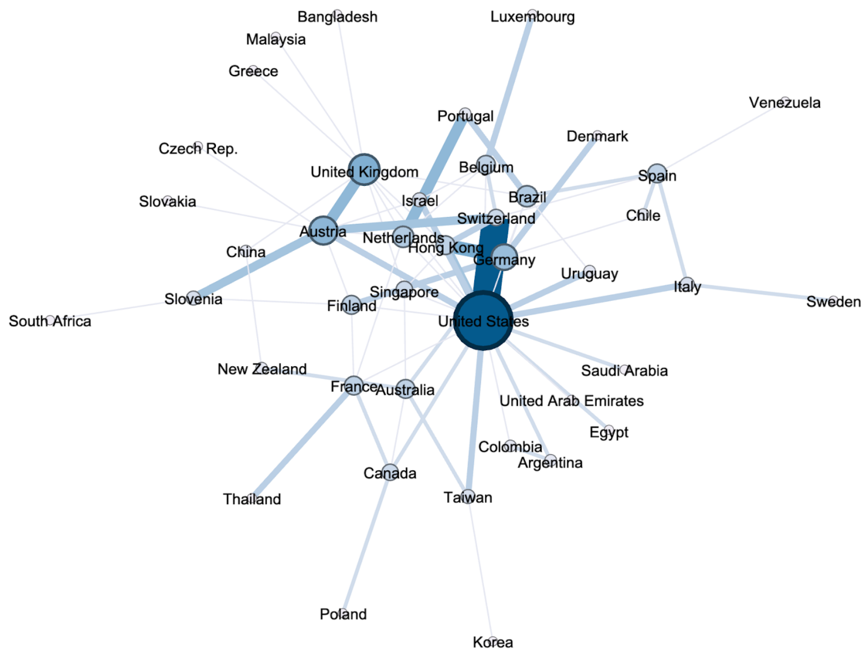


Figure 22. A collaboration network across countries/regions

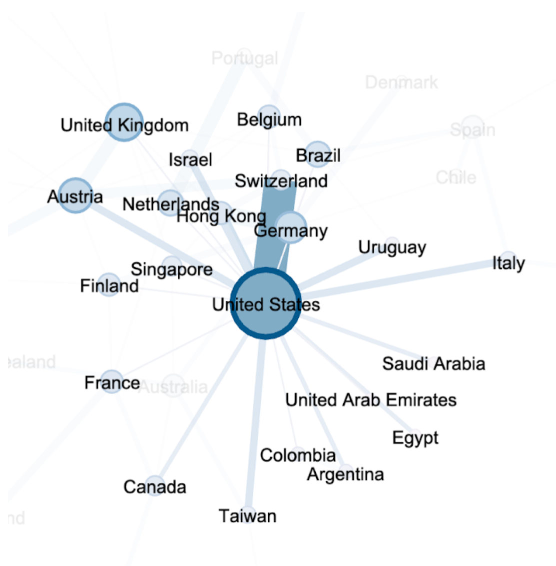


Figure 23. The collaboration sub-network centered at the United States

6 Conclusion

Design is pervasive and fascinating. The recent achievements in Computer Science, especially those on artificial intelligence, are inspiring and exciting. Computational Design, as an intersection and integration of design science and computation technologies, is a rapidly developing interdisciplinary research field, which combines art and mathematics.

I can still remember the moment I searched "Computational Design" in Google Scholar when I just started to study this field, I found out there are over 3,800,000 results and I could not decide how to start my exploration in this area. After spending almost two years in reading papers in this field and getting involved in multiple CAAD-related projects, I eventually have a picture of Computational Design in my mind, but I hope to construct a systematic and detailed review of Computational Design as a research field, with the goal to provide informative, intuitive, interactive visualization of the field to assist novices and junior researchers to obtain a high-level overview of the area, look for trending research directions, and also discover possible collaborations around the world.

Realizing that manual review of the whole field is biased and highly depends on the reviewer's own experience, I adopt a bibliometric approach to model Computational Design from all the academic publications in this field. In my thesis, I choose CumInCAD as the data source and perform data scrawling, data cleaning, and data quality analysis to ensure the bibliographic data are meaningful and reliable. To construct a multidimensional model of Computational Design, I propose and study three questions:

- How do the research topics evolve over time?

- How do the academic publications distribute geographically?
- How do the researchers and institutions collaborate?

To study the research topics in Computational Design, I apply state-of-art topic extraction algorithms in machine learning on the text corpora gathered from the titles and summaries of all the academic publications. To study the evolvement of the interests of researchers, I use a fully visual interaction technique to generate broken lines for trends in the number of publications under different topics.

After obtaining institutional information of the researchers through the email addresses recorded in the academic publications, I am able to study the geographic distribution of these publications on both a country/region level and an institutional level. With the help of topic modeling, I conducted the research topic analysis and trend analysis for each institution to uncover research profiles for the institutions.

Finally, collaboration is crucial for a healthy and inclusive research community. I use the data on co-authorships from the academic publications in Computational Design to construct collaboration networks on three levels: among individuals, research institutions, and countries/regions. The analysis results, as an informative visualization, indicate patterns for researchers' taste of collaboration. Also, from the results, I find out that the community of Computational Design is quite collaborative.

Although topic modeling and bibliometrics are unbiased and accurate in general, my work has several limitations. First of all, My work relies on CumInCad, which results in the main hard bias of the conclusions. Second, the topic extraction algorithm is based on syntactic information of the text and does not account for the semantics of the text. Third, the collaboration analysis does not use the latest CumInCAD database, so the findings might not be consistent with the current Computational

Design community. Last but not least, I do not perform some other widely used bibliometric-based analyses, such as the keyword co-occurrence analysis and the burst detection analysis, which could provide more information about the research field. In the future, I hope to resolve these limitations and build a user interface for all the analyses in my thesis to share my findings and methodology with other researchers in Computational Design.

References

Abdelmohsen, Sherif M. 2013. "Reconfiguring Architectural Space Using Generative Design and Digital Fabrication: A Project Based Course." In *Proceedings of the 17th Conference of the Iberoamerican Society of Digital Graphics*.

Achten, Henri, and Jakob Beetz. 2009. "What Happened to Collaborative Design?" In *27th eCAADe Conference Proceedings*.

Achten, Henri, and Joran Jessurun. 2003. "Learning From Mah Jong - Towards a Multi-Agent System That Can Recognize Graphic Units." In *Proceedings of the 10th International Conference on Computer Aided Architectural Design Futures*.

Adel, Arash, Andreas Thoma, Matthias Helmreich, Fabio Gramazio, and Matthias Kohler. 2018. "Design of Robotically Fabricated Timber Frame Structures." In *Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture* (ACADIA).

Agkathidis, Asterios. 2012. *Computational Architecture: Digital Designing Tools and Manufacturing Techniques*. Bis Pub.

Anders, Peter. 1996. "Envisioning Cyberspace: The Design of On-Line Communities." In *ACADIA Conference Proceedings*.

Ariza, Inés, and Merav Gazit. 2015. "On-Site Robotic Assembly of Double-Curved Self-Supporting Structures." *Anais Do XIX Congresso Da Sociedade Ibero-Americana de Gráfica Digital* 2015. https://doi.org/10.5151/despro-sigradi2015-110316.

Bayazit, Nigan. 2004. "Investigating Design: A Review of Forty Years of Design Research." *Design Issues*. https://doi.org/10.1162/074793604772933739.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research: JMLR* 3: 993–1022.

Bridges, Alan. 1986. "Alternative Approaches Towards the Teaching of Computer Aided Architectural Design." In *4th eCAADe Conference Proceedings*.

———. 1994. "Architectural Computing Education." In *Proceedings of the 12th European Conference on Education in Computer Aided Architectural Design*.

———. 1997. "Building Systems Integration and the Implications for CAD Education." In *15th eCAADe Conference Proceedings*.

———. 2006. "A Critical Review of Problem Based Learning in Architectural Education." In *24th eCAADe Conference Proceedings*.

Celani, Gabriela, and Pedro Luís Alves Veloso. 2015. "CAAD Conferences: A Brief History." In *16th International Conference CAAD Futures*.

Cerovsek, Tom, and Bob Martens. 2016. "CumInCAD 2.0: A Redesigned Scalable Cloud Deployment - Towards Higher Impact with Openness and Novel Features." In *Proceedings of the 34th eCAADe Conference*.

Chen Rui, Irene, and Schnabel Marc Aurel. 2011. "Multi-Touch - the Future of Design Interaction." In *Proceedings of the 14th International Conference on Computer Aided Architectural Design Futures*.

Chilton, J. C., T. Wester, and J. Yu. 1993. "Exploring Structural Morphology Using CAD." In *eCAADe Conference Proceedings*.

Cobo, M. J., A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera. 2011. "Science Mapping Software Tools: Review, Analysis, and Cooperative Study among Tools."

Journal of the American Society for Information Science and Technology. https://doi.org/10.1002/asi.21525.
De Bellis, Nicola. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Scarecrow Press.
Gero, John, and Han J. Jun. 1995. “Getting Computers to Read the Architectural Semantics of Drawings.” In *ACADIA Conference Proceedings*.
Gero, John S. 1990. “Design Prototypes: A Knowledge Representation Schema for Design.” *AI Magazine* 11 (4).
Gero, John S., and Myung Yeol Cha. 1997. “Computable Representations of Patterns in Architectural Shapes.” In *Proceedings of the Second Conference on Computer Aided Architectural Design Research in Asia*.
Gero, John S., and Michael A. Rosenman. 1990. “A Conceptual Framework for Knowledge-Based Design Research at Sydney University’s Design Computing Unit.” *Artificial Intelligence in Engineering*. https://doi.org/10.1016/0954-1810(90)90003-m.
Hoffman, Matthew D., David M. Blei, Chong Wang, and John Paisley. 2013. “Stochastic Variational Inference.” *Journal of Machine Learning Research: JMLR* 14: 1303–47.
Ke, Weimao, K. Borner, and L. Viswanath. 2004. “Major Information Visualization Authors, Papers and Topics in the ACM Library.” *IEEE Symposium on Information Visualization*. https://doi.org/10.1109/infvis.2004.45.
Kilian, Axel. 2006a. “Design Exploration through Bidirectional Modeling of Constraints.” Massachusetts Institute of Technology.
———. 2006b. “Design Innovation through Constraint Modeling.” *International Journal of Architectural Computing*. https://doi.org/10.1260/147807706777008993.
Kleinberg, Jon. 2002. “Bursty and Hierarchical Structure in Streams.” *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’02*. https://doi.org/10.1145/775047.775061.
Koutamanis, Alexander. 2005. “A Biased History of CAAD.” In *23nd eCAADe Conference Proceedings*.
Kui, Yue, and Ramesh Krishnamurti. 2007. “Feature-Based Prediction of the Interior Layout of Building.” In *Proceedings of the 12th International Conference on Computer Aided Architectural Design Research in Asia*.
Leach, Neil, and Philip F. Yuan. 2018. *Computational Design*.
Lewis, Danica M., and Kristine M. Alpi. 2017. “Bibliometric Network Analysis and Visualization for Serials Librarians: An Introduction to Sci2.” *Serials Review*. https://doi.org/10.1080/00987913.2017.1368057.
Liu, Yong, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. “CHI 1994-2013: Mapping Two Decades of Intellectual Progress through Co-Word Analysis.” In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI ’14*, 3553–62. New York, New York, USA: ACM Press.
Liu, Zhigao, Yimei Yin, Weidong Liu, and Michael Dunford. 2015. “Visualizing the Intellectual Structure and Evolution of Innovation Systems Research: A Bibliometric Analysis.” *Scientometrics*. https://doi.org/10.1007/s11192-014-1517-y.
Li, Xiao, Peng Wu, Geoffrey Qiping Shen, Xiangyu Wang, and Yue Teng. 2017. “Mapping the Knowledge Domains of Building Information Modeling (BIM): A

Bibliometric Approach." *Automation in Construction*. https://doi.org/10.1016/j.autcon.2017.09.011.

Loukissas, Yanni, and Lawrence Sass. 2004. "Rulebuilding (3D Printing: Operators, Constraints, Scripts)." In *Proceedings of the 23rd Annual Conference of the Association for Computer Aided Design in Architecture*.

Martens, Bob, and Ziga Turk. 2003. "Cumulative Index of CAAD: Current Status and Future Directions." *International Journal of Architectural Computing*. https://doi.org/10.1260/147807703771799193.

Menges, Achim, and Sean Ahlquist. 2011. *Computational Design Thinking: Computation Design Thinking*. John Wiley & Sons.

Park, Jihyun, Azizan Aziz, Kevin Li, and Carl Covington. 2013. "Energy Performance Modeling of an Office Building and Its Evaluation – Post-Occupancy Evaluation and Energy Efficiency of the Building." In *Proceedings of the 18th International Conference on Computer-Aided Architectural Design Research in Asia*.

Price, Derek De Solla, and Derek De Solla Price. 1976. "A General Theory of Bibliometric and Other Cumulative Advantage Processes." *Journal of the American Society for Information Science*. https://doi.org/10.1002/asi.4630270505.

Rocha, A. J. M. 2004. "Architecture Theory, 1960-1980: Emergence of a Computational Perspective." PhD Thesis, MIT.

Santos, Eduardo T., and Luis A. Derani. 2003. "An Immersive Virtual Reality System for Interior and Lighting Design." In *Proceedings of the 8th International Conference on Computer Aided Architectural Design Research in Asia*.

Sievert, Carson, and Kenneth Shirley. 2014. "LDAvis: A Method for Visualizing and Interpreting Topics." *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. https://doi.org/10.3115/v1/w14-3110.

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, et al. 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search." *Nature* 529 (7587): 484–89.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. 2017. "Mastering the Game of Go without Human Knowledge." *Nature* 550 (7676): 354–59.

Staback, Danniely, Mỹdung Nguyễn, James Addison, Zachary Angles, Zain Karsan, and Skylar Tibbits. 2017. "Aerial Pop-Up Structures." Massachusetts Institute of Technology.

Stamper, John, and Adam Perer. 2018. "Lecture 2 Asking Questions."

Suominen, Arho, and Hannes Toivanen. 2016. "Map of Science with Topic Modeling: Comparison of Unsupervised Learning and Human-Assigned Subject Classification." *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23596.

Turk, Ziga, Tomo Cerovsek, and Bob Martens. 2001. "The Topics of CAAD." *Computer Aided Architectural Design Futures* 2001. https://doi.org/10.1007/978-94-010-0868-6_41.

Zhao, Xianbo, Jian Zuo, Guangdong Wu, and Can Huang. 2019. "A Bibliometric Review of Green Building Research 2000–2016." *Architectural Science Review*. https://doi.org/10.1080/00038628.2018.1485548.