The Study of Chaos Theory and Information Theory in Enhancing Data Standard towards Smart Infrastructure

Thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy
By
Yi AN

September 2023

School of Engineering

Dedicated to my beloving parents Prof. Zhimei AN

and Ms. Yuqiu WU.

Acknowledgement

My greatest debt is to my Supervisor, Prof. Haijiang Li, not only for his encouragement and guidance over the last 4 years including his outstanding scientific insight and scholarship has made our research fruitful, especially considering I had the history of persistence on research and stubborn enough that if anything went south I would not even blink my eyes before drop the PhD to pursue another one, but also his vision that always been there supporting me to give me hope of keeping myself together for the research whenever I feel down, his trust that allows me to dive into a field that no other fellow researcher would have even heard about it, and allow me to keep the work on my pace even I had nothing serious to demonstrate due to huge mass of mathematics to study before I started chaos-BIM research. I have learned so much from him that I firmly believe that my whole life from now on would benefit extensively from things he taught me. I would like to share a little story between us. In my first 2 years of PhD, I had been working on the first Chaos-BIM paper for a long time before submission to a prestigious journal, even the submission is way more complicated than other journals. The decisions from editor and reviewers are quite controversial and aggressively insulting. I was quite upset and negative so that I was shockingly being the first time of not doing any research for the week. Prof. Haijiang Li, instead, regardless of also shocked by insulting and reckless decision from edtior/reviewers, he started to write an explanation letter to further clarify the standouts of paper and attempted to cheer me up. There were 2 lines I remember the most, he said,

you always want to be an excellent researcher in the future, but do you know what the excellence is? Excellence is not defined by the good work they have done, rather excellence is defined by how they bounce back from their bottom. Ever since that, I never felt giving up. One would never expect a better supervisor than him.

I am very grateful to Dr. Lorenzo Morini for his enlightening question about the paradigm of chaos I am supposed to discuss, and his valuable comments on my research progress. Ultimate tribute to late Prof. Bolin Hao, who is the very first one to open the world of chaos and nonlinearity for me. His knowledgebase, solid mathematics/physics, and excellent understanding of research/science truly motivates me to better myself and follow my superhero’s footstep.

I am greatly owned to my parents. Without their continuing encouragement and sacrifices, I would have no chance to start academic research. Their role modeling, encouragement, support both financially and emotionally, are always there walk me through transition from one country to another country, from one university to another, from one field to another, through the whole period of Covid. My family is very similar to those middle-class Chinese households, parents kept trying their best to support kids especially when it comes to education. The education does not limit to academic aspects such as to best primary school in the region, best middle school and high school in the city at all costs. They also paid my football training, basketball training and such, in hope of me keeping fit meanwhile understanding the importance of endurance, persistence and teamworking. Most importantly, they taught me how to behave, to be polite, to be kind to others, to survive and etc. Although, I was never the obedient son. So many conflicts, arguments among us, mostly I don’t agree what they asked me to do. But deeply, in my heart, I am very thankful for my parents showing me that much of patience, and I am extremely shamed of myself that I always failed to control my temper. In the last 10 years since I left home, went abroad. I have encountered plenty people from different places, with various cultural context, work on different sectors. Still, I would claim, my mother, Miss Yuqiu WU, is the purest soul I have ever seen, the most kind-hearted person I have ever heard of, the warmest and greatest female I have ever met. All these, by showing me how to be a better person, taught me and guided me to be a better man. To all these, I cannot express my appreciation more. Without you guys, I would not even achieve 1% of what I am today. My special thanks must go to my fiancé, Yuchen Wang, not only for her graciously understanding and support, but also for her inspiration so that direct me to this field. As a way more outstandingly intelligent being, her kindness, patience and acceptance does create and lead my way of being a better man. I have undying gratitude towards her for that.

As much learnt as I have, these 4 years is tough, especially with experience of Covid situation. Besides aforementioned, so many friends and colleagues are so kind, they inspired me, fought with me, stood by me, warmed me in many ways. These people should be credited, Prof. Guoqian Ren, Dr. Xiaoyu Liu, Dr. Zhjiang Chen, Dr. Xiaofeng Zhu, Dr. Jiaxin Ling, Dr. Xuhui Lin, Dr. Weicheng

Peng, Tianyun Jiang, Future JD. Zhihui Shen, Zhaokui Huang, Yuhan Zhang, Nan Xiang, Yuchen Lin, Yalin Tan, Kaiyan Zhou, Yingxue Chen, Youwei Xu, Zeyu Liu, Siyi Chen, Guohong Luo, Justin, Bahni, Suraby, Kevin, Kate, Paul, Shawn, Liming Xuan.

Y.AN

Home in London,

31st Sep 2023

Abstract

This dissertation explores the complex dynamics underlying Building Information Modeling (BIM) data standard development, with the aim of enhancing efficiency and effectiveness in the Architecture, Engineering, Construction and Operations (AECO) industry. The research integrates chaos theory and information theory to elucidate hidden patterns and principles governing BIM standards.

This thesis establishes the methodological framework, combining theoretical research and design science approaches. Information theory provides a quantitative lens to analyze BIM information flows, while chaos theory recognizes inherent complexity and unpredictability. Philosophically, the research embraces interdisciplinarity and pragmatism. Through sandpile simulations, the dynamics of BIM standard development are modeled computationally. Innovative mapping techniques connect simulation patterns to actual BIM standards topologies, represented as tree structures. Analyses reveal "similarity cross-scalability," indicative of chaos and self-organized criticality. This suggests BIM standards evolve akin to a chaotic system, with sensitivity to initial conditions. Mathematical techniques rigorously prove chaotic properties in BIM standard development. Time series data from simulations enable phase space reconstruction. Determining optimal time delay and dimensionality allows creating an accurate phase space capturing system dynamic. Calculation of a positive Lyapunov exponent provides definitive evidence of chaos.

New methodologies emerge from the chaos-driven perspective. Information theory and sandpile principles generate novel Model View Definitions (MVDs) for tunnel linings, embracing dynamism while reducing ambiguity. Comparative analysis shows improved consistency over conventional standards. System attractors within reconstructed phase space form the basis for a chaos-informed performance indicator for BIM models, using distance to attractors as a stability metric.

In summary, this pioneering research makes significant contributions:

It proves, mathematically and empirically, the presence of chaos in BIM standard development related to information flows. Chaos theory and information theory are shown to offer valuable perspectives for enhancing BIM standards. Innovative techniques are proposed for generating adaptable, robust MVDs and evaluating BIM model stability. Philosophy of interdisciplinarity and pragmatism is embraced to integrate diverse concepts. Computational modeling and mapping reveal new insights into complex BIM standard dynamics.

The implications are profound. Identifying chaos enables harnessing advanced techniques from disparate disciplines to optimize BIM processes. The proposed methodologies demonstrate enhanced efficiency, consistency, and performance. This research lays the foundations to utilize chaos theory for next-generation innovations in the AECO industry. The transformative potential is to fundamentally evolve BIM standards to be highly adaptive and responsive to the industry’s dynamic needs.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AEC(O)= Architectural Engineering and Construction (Operation)

BIM=Building Information Modeling

IFC=Industry Foundation Class

IoT=Internet of Things

MVD=Model View Definition

NLP=Natural Language Processing

IFD= International Framework for Dictionaries

LoD=Level of Details

IDM= Information Delivery Manual

AI=Artificial Intelligence

SOC= Self-organized Criticality

IAI= International Alliance for Interoperability

XML= eXtensible Markup Language

CIS/2 =CIMsteel Integration Standards, Second Release

GBXML = Green Building XML

CAD = Computer-Aided Design

Chapter 1—Introduction

1.1 Problem Statement

The AEC industry is upgrading its traditional methods from operation model to business model and as such, with the development of digitalization. However, moving forward with this digitalization within the industry requires more than just new techniques and technologies to help them collaborate more effectively and enhance the decision-making process. also It also requires a way of looking into these new techniques and technologies from a more top-down hierarchical thinking with more fundamental theories.

Building Information Modeling (BIM) has been proposed, utilized and considered as a core and ground concept as well as technology to provide better digitalization for AECO project (Amor, 2008; Patacas, Dawood, & Kassem, 2020). BIM is characterized as a comprehensive, intelligent, and parametric digital depiction of a facility, enriched with data. This representation allows users to extract and analyze tailored views and data, facilitating informed decision-making and enhancing the facility delivery process (An, 2017). However, a significant challenge arises in the form of interoperability, defined as the seamless data exchange capability across diverse disciplines and stakeholders. This challenge stems from the use of varied tools in BIM (Eastman, Jeong, Sacks, & Kaner, 2010). As a result, there is a pressing need to convert the BIM model into a format that is compatible with other design and analytical tools.

In response to these, Industry Foundation Class (IFC) development along with Information Delivery Manual (IDM), International Framework for Dictionaries (IFD), Model View Definition (MVD), Level of Detail (LoD), Linked data emerged in the last decade Andre Borrmann, Koch, Beetz, 2018).

Such development and associated information delivery process plays a significant part in enhancing collaboration through information requirements, exchange requirements, etc. However, the data exchange process faces several concerns that restrict its prosperity, mainly related to the understanding of information modeling, which requires an informed understanding of information exchange from the perspective of information theory and complex nature. In a typical project, multiple stakeholders produce various BIM models. Each of these models captures a distinct segment of the whole building and is termed a domain-specific partial model (An, Lin, Li & Wang, 2023b). Although these sub-models serve different purposes, they possess overlapping features that are not limited to a particular domain. This leads to lacking robustness in terms of standards development (not limited sets of solution for potential applications) meanwhile creating information redundancy. On the other hand, although several attempts have been made to develop data exchange standards for BIM models, various scales of projects that contains different amount of information would be expressed with same data exchange standard, which apparently result in failure in guaranteeing the efficiency and performance of associated BIM models. It is due to limited understanding of complex nature of system about data exchange standard against information mass of the project.

Despite the effort from extensive scholars to develop a multi-objective knowledge base within the BIM context, most research did not provide comprehensive and fundamental thinking that can comprehend data/information in AEC industry in a relatively more generic and critical way. Hence, an attempt is required to lever information theory and comprehend complex nature of data exchange standard within BIM context on the purpose of optimizing existing data exchange standard development.

1.2 Motivation and Background

In recent years, building information modelling (BIM) has become very popular, providing means widely used in the architecture, engineering, and construction industry to describe conventional information about building/infrastructure system (C. Eastman, 2011). To achieve the goal of BIM, a vendor-neutral, open, and standardized data exchange format was needed. The international organization buildingSMART has dedicated many years to the development of IFC as an open, vendor-neutral data exchange format. This is a complex data model with which it is possible to represent both the geometry and semantic structure of a building model using an object-oriented approach. The building is broken down into its building components on the one hand and its spaces on the other, both of which are described in detail along with the interrelationships between them. Thanks to its comprehensive data structure, it can be used for almost any data exchange scenario in the life cycle of a building. For example, the design and construction of shield tunnel linings in the field of civil engineering require careful consideration of various factors such as structural integrity, material properties, and construction feasibility. To facilitate the design process, Model View Definitions (MVDs) play a crucial role in representing and managing the relevant information associated with shield tunnel linings (Koch, Vonthron, & König, 2017). MVDs serve as a means to define subsets of a building model that fulfill specific purposes and support interoperability among stakeholders. They enable the exchange and coordination of information between different software applications and facilitate efficient collaboration.

As the boosting development of BIM standard, data exchange standard assessment has been a tenacious concern. The starting point for this is to develop a better and more comprehensive way to assess the quality of an IFC file. The assessment criteria derive from mainly two perspectives, one is being applicable and reliable as a data exchange such as syntactically checking, and the other one is being applicable for fulfilling end user's requirements, such as testing, satisfying programmatic requirements, and building code-oriented requirements. The concerns on BIM standard assessment are mainly applicability oriented. The consideration of lacking efficiency about BIM standard has risen doubts among standard development institutions, software vendors and end users. This necessitates an imperative matter regarding defining an assessment criterion for various developed BIM standard in terms of efficiency. However, defining an assessment criterion for BIM standard is challenging as it could not take out the BIM model-based performance evaluation, ergo, the validating data quality would be involved which contains multiple characteristics of data. Such validations as well as evaluation of BIM standard spotlighted the data source, domains including systems, geometry, personnel, process etc. which regard BIM model as a media rather than separating BIM standard as independent variable against the BIM model performance.

It, apparently, can be seen that development data exchange standard play a significant role in the development process of the BIM ecosystem. However, there are two major defects that has been overlooked. Consequently, the BIM ecosystem has been facing extensive challenges. The very first defect that is overlooked is existing research in the field of IFC/BIM was primarily focused on improving data interoperability and use case scenario adoption without extensively incorporating information theory-based approaches. However, the limitation such as significantly diverse versions for applications and ambiguity generated by information requirements in IDM is risen those would impede BIM development. The main reason behind this would be lack of a rigorous theoretical framework to understand the fundamental limits of IFC and to guide the design of efficient BIM systems. Referring to the nature of information theory that would quantify information and provide a measure of its fundamental properties, it would be helpful to establish a theoretical foundation for the design and analysis of BIM/IFC development systems. However, there is a research gap in understanding the potential benefits and contributions of information theory in enhancing data interoperability within the IFC framework. Information theory provides a robust foundation for quantifying and analyzing information flow, data compression, and optimization, which can potentially offer valuable insights and solutions for enhancing data interoperability in IFC. By acknowledging the existing research that does not heavily involve information theory, there is an opportunity to explore the incorporation of information theory principles and techniques. The second overlooked defect is the missing complex nature exploration respect to data exchange standard development against the changing information mass of projects. Without explicit understanding of complex nature of data exchange standard development, the capability of quantifying the development process is limited. Ergo, it hinders the capability of generalization. Consequently, in this research, the information theory and the complex nature exploration in BIM standard development are investigated by reviewing the existing concepts, technologies, tools to see how these could benefit BIM community and AEC industry in terms of MVD generation and BIM modeling testing.

In IFC data model, the wealth of information that can be captured in attributes, properties and at a geometric level often exceeds the intended use at a particular stage in the life cycle of a building project, which make it difficult to capture and retrieve information in an appropriate form for other scenarios. To address the problem, the Information Delivery Manual (IDM)/Model View Definition (MVD) framework is developed to agree on uniform and standardized means to further specify the contents expected from a building model instance. These specifications regulate which information is delivered by whom, when, and to which recipient (buildingSMART, 2012, 2013). This helps reduce room for interpretation and makes it easier to implement specific use cases and application areas. The framework distinguishes content-related requirements captured in IDM and technical implementations and mappings of these requirements in the form of MVD. The technical implementation of these agreed requirements in the form of partial IFC Models is based on the Model View Definition standard. An MVD is a technical means of checking the validity of instance models for a particular exchange scenario. Specifications in a Model View range from the definition of required Property Sets to restrictions on allowable forms of geometry representations. According to ISO 29481, IDM predefined uniform structure and method for presenting process models enables users to develop, agree on and accurately document their BIM processes. The corresponding technical counterpart to the individual IDM specifications are MVD that define the specific sub-element of the overall IFC data model that can support the specific exchange requirements of the IDMs (Lee, Park, & Ham, 2013). The relations among IFC, MVD, IDM and International Framework for Dictionaries (IFD) is given in Figure 1 (Borrmann, Koch,Liebich, Muhic, 2018)._To this step, it can be concluded that the IFC is used to bridge the connections between stakeholders and project phases in a fragmented project environment typical of the construction industry through IFD and MVDs. A main reason is that there are vast objects (building types, infrastructure types) and subjects (facility management, structural analysis/design, energy performance) brought by various parties.

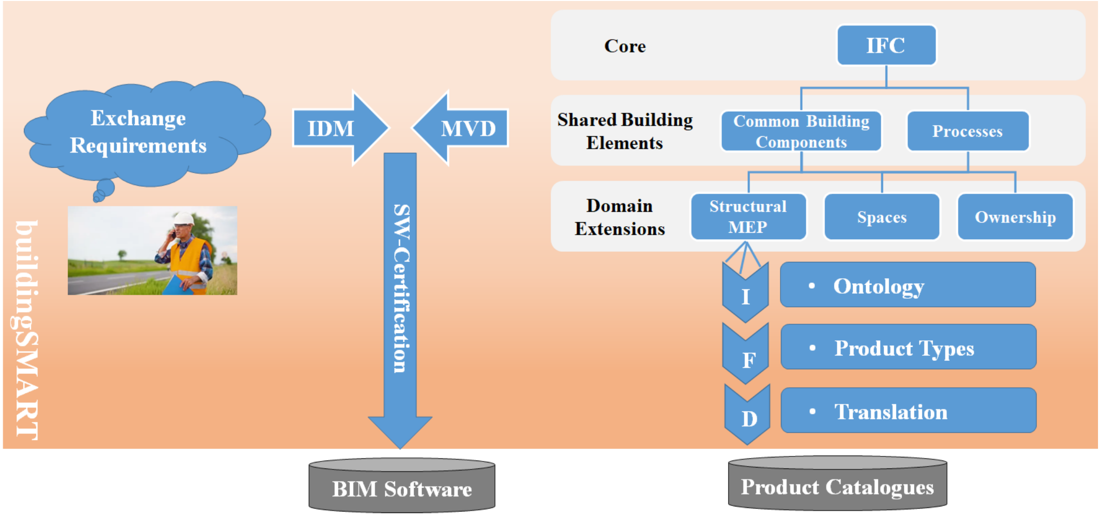


Figure 1 The relationship between IDM, MVD, IFC and IFD

While it comes to the more complicated projects like underground tunnel facilities, the complexity could ascend more. To better collaborate the BIM with underground tunnel facilities, fellow researchers make tremendous effort on it. Unlike other buildings and infrastructures, the design process, construction, and operations of mechanized tunnels requires more comprehensive, detailed information models that collaborate with varied and different information technology tools in an efficient way (Borrmann, Flurl, Jubierre, Mundani & Rank, 2014). To address this, as an alternative and complementary approach to specifying design and planning requirements using IDM/MVD, a multi-level information representation of the built environment was proposed, and a conceived collaboration platform was developed to support early-stage design/planning and a conceived collaboration platform was developed to support early-stage design/planning (Borrmann, Flurl, Jubierre, Mundani & Rank, 2014). Later, Abualdenien and Borrmann completed it further by presenting a multi-LoD meta-model to explicitly describe an LoD’s requirements, incorporating the potential fuzziness of both, geometric and semantic information of individual elements. The explicitly defined fuzziness can be taken into account when applying simulations or analyses for assessing the performance of different building design variants (Abualdenien & Borrmann, 2019).

Put in a nutshell, it is promising to further develop a complex thinking-based BIM standard assessment method that is quantifiable and outperforms existing BIM standard test methods in terms of efficiency, generalization, and robustness.

1.3 Research Hypotheses

In light of the research problems and motivation identified above, this research aims to develop a further understanding from information theory and complex nature that could refine the current understanding of data exchange standard to eliminate inefficiencies, inconsistencies in data exchange standard development stage. The overarching aim and hypothesis adopted in the research are as follows: The aim of this research is to enhance the data exchange standards in Building Information Modeling (BIM) by integrating advanced theories and technologies from information theory and other domains. The hypothesis of this research is that applying principles of information theory and incorporating successful techniques from other technical domains can significantly reduce inefficiencies and inconsistencies in BIM data exchange standards, thereby improving the overall quality and utility of BIM models.

1.4 Research Question

Based on the research aim in this thesis, the research hypothesis is broken down into the following research questions:

Q1: What are the concepts, frameworks, and process existing within the data exchange standard development to improve BIM ecosystem development? And how are those concepts and such backing BIM development while considering their scope and limitations? These are discussed in Chapter 2

Q2: What are general information theory applications and why information theory would have potential to be employed on data exchange standard development in BIM? What is the possible complex nature of the data exchange standard development in BIM? These are discussed in Chapter 2

Q3: what is required to employ information theory and identify the complex nature to support data exchange standard? These are discussed in Chapter 3

Q4: why are these required to employ information theory and identify the complex nature to support data exchange standard? These are discussed in Chapter 4

Q5: What is the possible complex nature through exploration? These are discussed in Chapter 5

Q6: How to covert the possibility of discovered complex nature to certainty of discovered complex nature? These are discussed in Chapter 6

Q7: Can the findings and refined understanding with respect to information theory and complex nature benefit the data exchange standard development? These are discussed in Chapter 7

1.5 Thesis Overview

This thesis is divided into eight chapters, each pursuing answers to the main research questions.

Following the introduction in Chapter 1, Chapter 2 aimed to answer the first and second research question by introducing a thorough literature review that is relevant to the research topic. Section 2.2 elaborated on the data exchange standards development in BIM and the benefits as well as limitations of it at the moment. In section 2.3, information theory was investigated in terms of its achievements in other domains' application and the potential of employing it in BIM. Section 2.4 discuss the chaos as a typical nonlinear complex nature and its application in other domains so that could reveal the possibility of being the complex nature of data exchange standard in BIM. However, due to the large-scale of BIM, information theory and chaos theory topics, and considering the focus of this thesis, this literature is by no means an exhaustive review. On the flip side, it indicates the extensive developments taking place in these areas and their potentials to advance further. At the end of this chapter, the main findings of the review are given in the similar form of research gaps.

Chapter 3 presented the overarching methodology through which this research was carries out to clarify the principles and methods utilized in this research. The main methodology adopted in this research is the integration of Theoretical Research and Design Science Research (DSR). Theoretical Research is used here to bridge gaps between difference disciplines (ICT and BIM) by providing overarching theories (information theory and chaos theory) and concepts. And DSR is usually utilized in categories of artefacts referring to engineering and computer science disciplines to solve a generic challenge experienced in practice. Together, Theoretical Research and Design Science Research could both provide theoretical foundations and empirical studies for reshaping the data exchange standard development in BIM. The chapter answered the Question 3.

Chapter 4 provided theoretical and philosophical underpinnings that guide the research methodology and approach. It provides a rationale for the chosen methods, situating them within broader philosophical traditions and paradigms. This chapter started with exploration of what constitutes knowledge in the realm of Building Information Modeling and data standards, then introduced the idea that BIM standards and data structures can be viewed as complex systems. Also, reviewed how information theory provides a bridge between quantitative data (e.g., measurements from simulations) and qualitative understanding (e.g., the meaning and implications of BIM standards). Afterwards, discussed philosophical and theoretical implications of embracing chaos theory in the study of BIM and data standards. The findings of the chapter answered the Question 4.

Chapter 5 set the cornerstone of bridging techniques and theories from other discipline with tangible BIM applications. Therefore, the findings of this chapter are used to discover the possible complex nature of data exchange standard development in BIM ecosystem, which is meant to answer Question 5. The technical framework includes the information quantification to complete the initial setup for linking Sandpile Simulations with BIM Model (Section 5.2); sandpile simulations to assist studying MVD which acts the proxy of data exchange standards' abstract form (Section 5.3); proposing new mapping system to convert proxy to applicable MVD (Section 5.4); Section 5.5 provided simulation results and indicated the possible complex nature would be chaos paradigm.

Chapter 6 shows Theoretical Research to prove the chaos is not merely possible but a firmed fact in a rigorously and crucially mathematical way, which answer the Question 6. The proof was conducted with known critical process, starting from data preparation and sampling (Section 6.1), then reconstructed phase space (Section 6.2), determined the Lyapunov Exponent to prove the syste chaotic. It was the definitive proof, the seal of authenticity, confirming the system's chaotic nature. In essence, this exploration was not just about proving a point but about embarking on a scientific odyssey.

Chapter 7 addressed the testing and validation of the previous findings in response to the last research question. After Chapter 5 and 6, we developed information theory and sandpile simulation-based framework to generate MVD for reducing vagueness and inconsistency (Section 7.2). Sandpile simulation and information theory, the new approach of generating MVD could be able to evolve to keep pace with the dynamic nature of the AECO industry. Conventional MVD generation approaches often struggle to accommodate the nuances of complex projects, leading to inefficiencies, errors, and lost opportunities for optimization. Our methodology transcends the limitations of traditional methods, embracing the complex nature of construction processes to create MVDs that are agile, robust, and adaptable. Moreover, with the findings of Chapter 6, a new BIM model performance evaluation method was developed. The attractor within the context of chaos theory formulates the ground for proposed key indicator that enables the evaluation method quantifiable and outperform existing BIM standard test methods in terms of efficiency, generalization, and robustness (Section 7.3). Both developments were accompanied with case studies.

Chapter 8 wrapped up the discussions and analyses from earlier chapters, highlighting the primary outcomes in relation to the research hypothesis. This is followed by a discussion on the limitations of the research and potential directions for future studies. The chapter concludes by summarizing the significant contributions of the research.

Chapter 2 -Literature Review

2.1 Introduction

This chapter presents a literature review in 3 sections. Section 2.2 elaborates on the data standard development for building information modeling. Also, along with the data standard development, evolution on testing methods of BIM data standard is investigated. Through comprehensive review of data standard development for building information modeling, the advancements and limitations of existing development are discussed. Then, in Section 2.3, information theory and information theory application on other information related domains is conducted to provide a comprehensive grasp of rationale about employing information theory in information related subject to optimize the existing problems. Based on these, the research gaps would be uncovered in the end of this section. Finally, Section 2.4 discusses chaos theory and its application on other domains. However, given the vast scope of data standard topics and the specific focus of this research, the literature presented does not claim to be a comprehensive review. Nonetheless, it underscores the numerous advancements and their inherent constraints in this domain. The chapter concludes by highlighting the key insights from the review, which align closely with the identified research gaps.

2.2 Data Standard Development for Building Information Modeling

2.2.1 Standards Development

The evolution of BIM data exchange standards commenced in the 1990s, concomitant with the burgeoning application of computer technology within the architectural industry, thereby crystallizing the necessity for standardized descriptions and exchanges of building information models. In 1997, the International Alliance for Interoperability (IAI) initiated a project (Eastman *et al.*, 2011) with the objective of establishing a unified data exchange standard for the architectural industry, aimed at resolving prevalent incompatibility issues among various CAD software at the time.

In 1999, IAI released version 1.0 of the Industry Foundation Classes (IFC) (Costa and Madrazo, 2015), representing the inaugural open data model standard oriented towards the architectural industry and signalling the inception of the BIM data exchange standard era. For instance, Froese (2003) utilized IFC 1.0 for model exchange in steel structural engineering. Transitioning into the 21st century, with the rapid development of BIM technology and its applications, the limitations of IFC 1.0 became increasingly evident, eliciting additional demands for data exchange standards. In 2005, IAI was rebranded as buildingSMART, introducing the IFC 2.x versions (Eastman *et al.*, 2011), a series that significantly expanded the definitions and relational descriptions of architectural elements, thereby becoming the mainstream standard for BIM data exchange. The IFC enables BIM models from disparate software platforms to be exchanged and interoperated, significantly enhancing the efficiency of collaborative design. For example, Coates et al. (2010) utilized IFC to facilitate collaboration and evaluation in architectural design schemes. Also, the USA General Services Administration (GSA) has, since 2007, employed IFC 2x3 to realize collaborative design in over 250 government projects (US General Services Administration, 2007). In the construction domain, the IFC has also been implemented; for example, Tan, Hammad and Fazio (2010) explored the application of IFC in monitoring concrete protective layers, whereby the IFC standard allows for the smooth integration of monitoring data into BIM models, thereby providing support for quality management.

Thus, serving as a universal BIM data model standard, the IFC employs an object-oriented approach to describe architectural elements and systems, possessing a rich model architecture to encompass the principal information in the architectural domain (Sun *et al.*, 2015). The IFC is not only capable of representing models from various architectural disciplines but also articulates the topological relationships between architectural elements, supporting analyses of space and architectural performance (Costa and Madrazo, 2015). Consequently, the advantage of the IFC standard lies in its capacity as an open data model to support the integration and interactive operation of multi-domain and multi-system information.

Concurrently, with the widespread application of BIM (Building Information Modeling) across various specialized fields, data exchange demands in professional domains, such as steel structure and architectural energy analysis, have emerged, ading to the advent of standards like CIS/2 and GBXML in the 2000s. Regarding CIS/2, specifically formulated for the steel structure domain, employs an efficient parametric approach to meticulously model the geometry and manufacturing of steel components and their connections (Lee, Eastman and Solihin, 2018). As a BIM data exchange standard oriented towards the steel structure domain, CIS/2 has currently been applied to multiple stages within this field to realize the digitization of information flow. In the manufacturing stage, Shan et al. (2012) constructed a CIS/2 parametric steel component factory modelling system, capable of automatically outputting precise manufacturing information. Simultaneously, CIS/2 has also been applied to bridge component design details with overall modelling, as Eastman et al. (2005) proposed its interoperability with IFC through an intermediary conversion server. It also demonstrates commendable compatibility with mainstream steel structure software. For instance, Lipman (2009) researched and implemented a bidirectional conversion between CIS/2 and Tekla software, supporting a collaborative workflow from steel structure design to manufacturing. In general, the advantage of the CIS/2 standard lies in its ability to precisely articulate the manufacturing and construction information of steel structures, directly outputting detailed data for digital manufacturing and modelling (Kamat and Lipman, 2007).

On the other hand, as a BIM data exchange standard oriented towards architectural energy and environmental analysis, GBXML finds its applications extensively spanning architectural design optimization, compliance checks with standards, and building system performance simulation, among others. For instance, Elnabawi (2020) employed GBXML to achieve modeling conversion from Revit to DesignBuilder software, conducting an analysis of energy-saving potential in school renovations. Kim et al. (2015) utilized GBXML to output building thermal load information, conducting compliance checks with the ASHRAE standard. Furthermore, Xu et al. (2019) designed a framework to map the architectural information in GBXML format to the thermal properties needed in EnergyPlus. The mapping covers building envelopes, schedules, internal loads etc. GBXML has been widely applied in green building rating and certification, such as LEED and BREEAM, for conducting architectural energy-saving simulations (Jalaei and Jrade, 2014). In addition, the GBXML standard boasts its merits through adopting a lightweight XML format to describe the geometry, components, materials, and performance data of buildings (Dong, Lam and Huang, 2007), information that can be directly imported and utilized for architectural energy and environmental performance analysis, with the file being compact and easily shared (Jeong *et al.*, 2014). GBXML simplifies the conversion process from BIM to architectural energy modeling. Predominantly applied in the field of architectural energy conservation, the GBXML standard, as utilized by Garwood et al (2018), facilitates information conversion from BIM to architectural energy simulation software, conducting analyses of architectural energy conservation. GBXML standardizes the description of the geometry and performance information of buildings, automating the modelling and simulation processes.

In summary, the GBXML standard simplifies the conversion from BIM to architectural energy models, rendering the architectural environmental analysis workflow more efficient and intelligent. Its application scope encompasses the digital processes related to energy conservation and emission reduction throughout the entire lifecycle of a building, from early design optimization to operational assessment.

With the advent of the 2010s and the emergence of complex 4D and 5D BIM applications, new demands were placed on exchange formats. In this context, in 2010, buildingSMART introduced the concept of Model View Definition (MVD), aimed at extracting information required for specific domains from the comprehensive IFC model, thereby expanding the application scope of the IFC standard (BuildingSMART, 2016). MVD not only provides the capability to extract customized subset data from the complete IFC model but also offers the customization of IFC views for various industry application processes and domains (Pinheiro *et al.*, 2018). This implies that users can define views according to business requirements, acquire specific modelling information, and further broaden the application range of the IFC standard. The applications of MVD are diverse and multifaceted. For instance, Lai, Zhou and Deng (2019) defined an MVD view, mapping IFC architectural information with CIS/2 structural steel information, realizing the integrated application of BIM and

CIS/2, thereby enhancing the information interoperability of structural steel design. In another instance, Lee, Eastman and Solihin (2021) defined an MVD from IFC to Revit, achieving the conversion of the architectural structural model between the two. Wang et al. (2016) defined an IFC-MVD view to extract relevant spatial and geometric information for four-dimensional collision detection, based on the business requirements of MEP system collision prevention. Moreover, MVD has also played a role in data security control. As an extension standard of IFC, MVD has demonstrated its value and potential in various domains. In the realm of architectural energy, Pinheiro et al. (2018) utilized MVD to extract the subset of physical building information required for the IFC to BEM conversion. Abualdenien, Pfuhl and Braun (2019) extracted data views of the human flow simulation sub-model from hospital BIM. These instances showcase the potential of MVD in supporting novel applications, signalling the further expansion of the BIM application domain.

In 2013, the BCF standard emerged to support complex BIM software collaborative workflows. As a BIM information delivery standard, COBie has been extensively utilized to convey project completion information to owners and facility managers (Alnaggar and Pitt, 2018), fulfilling the requirements for project information delivery by organizing key project information in the form of Excel spreadsheets (Gao and Pishdad-Bozorgi, 2019). This structured data can be directly imported into facility management systems, enhancing the quality of information delivery. COBie focuses on the delivery at the project completion stage, effectively meeting the owners' needs for facility management data. For instance, Anderson et al (2012) implemented the COBie standard to realize a delivery platform for school facility management information. Additionally, COBie has also expanded its application to infrastructure engineering, as Hüthwohl et al. (2018) managed bridge and tunnel asset information based on COBie. In general, COBie standardizes the delivery of information needed for facility management in an easily accessible and communicative manner (Hosseini *et al.*, 2018). With owners paying attention to operational data, COBie plays an irreplaceable role throughout the entire building lifecycle. It effectively organizes the key information delivered to the owner after project completion, satisfying the owner's data needs for utilizing BIM in facility management.

In 2015, buildingSMART unveiled the Linked Data version of IFC, heralding a new paradigm for BIM data exchange based on Semantic Web technologies (BuildingSMART, 2017). As an emerging standard for BIM data exchange, Linked Data is propelling the connection of Building Information Modelling with data from various other fields, constructing digital collaboration. Pauwels (2014) developed a design decision support system by linking BIM models with architectural component catalogues, recommending suitable prefabricated components. Utilizing Semantic Web technologies, the Linked Data standard is conducive to the integration of BIM applications with other information systems, establishing more digitalized and intelligent means of managing the building lifecycle. For instance, Wang, Pan and Luo (2019) combined IFC models with geospatial information, supporting comprehensive analysis of buildings and urban environments. Also, Yao et al. (2018) achieved collaborative simulation and evaluation of buildings and their surrounding environment by linking IFC structural information and CityGML urban models. With the advancement of information technology, various non-BIM data have been deeply integrated with BIM, enriching building information applications. For instance, integrateing material databases to achieve automatic updates of material data in BIM models (Fenz *et al.*, 2021; Kebede *et al.*, 2022); Terkaj, Schneider and Pauwels (2017) evolved BIM models into digital twins of building facilities by linking perceptual device data, realizing status monitoring; Davila Delgado et al. (2020) connected BIM with monitoring input, supporting closed-loop control of intelligent building systems. Concurrently, Linked Data has also expanded the application scope of BIM into more fields, such as the integration of BIM and maintenance information realized by Marmo et al. (2020), and the connection between BIM and earthquake analysis conducted by Shi et al. (2023). It is evident that the Linked Data standard is driving the innovative application of integrating BIM with more heterogeneous data sources. Overall, the Linked Data standard facilitates the interlinking and interoperability of BIM data with external data sources, employing Semantic Web technologies to organize and associate distributed architectural data in a linked manner (Pauwels, Zhang and Lee, 2017). Its prowess is notably demonstrated in enhancing interoperability between heterogeneous data sources (Costa and Madrazo, 2015), enabling the association of dispersed BIM and non-BIM data to formulate a knowledge graph, and supporting the integrated management and intelligent analysis of building information throughout the entire lifecycle. This provides pivotal support for innovative BIM applications.

Thus far, throughout data standard development in BIM, it could find that there have been several successful attempts to (a) demonstrate the need of a comprehensive, systematic, robust information model, (b) automate the link from information model to all relevant information that could meet the requirements of the complicated projects planning, design, construction and maintenance, and (c) develop a comprehensive applicable information modelling framework within an open IFC environment including Proxies, Property Sets and Model View Definitions. Regardless of these successful research ideas and outcomes, it would be easily to review the way of developing the existing data exchange standards for fitting specific use-case scenarios is diverse. Within existing buildingSMART approved framework, there are common approaches to choose, such as, LoDs, MVDs and linked data. And even with the same approach, the proposed contents could be completely different from one to another, such as, with the same IDM, the Li's MVD for tunnelling BIM model is far different from Koch's. This, a standard framework which is supposed to specify the limited set of solutions to actual or potential matching problems and intended/expected to be used repeatedly or continuously by a substantial number of the parties for whom they are meant creates vast different solutions for the same case, could be definitely considered as a standard that has certain potential to grow (Vries, 2005). Without a good a-priori agreement among implementing manufacturers, the standard would not see a more drawn-out process on the market (Laakso, 2012). As aforementioned, in the case of complicated projects, the project environment gets even more fragmented. Hence, lack of principles to regulate the MVD/linked data/LoD could lead to even more significantly diverse version for applications. All these would hinder the development of data standard, thus impede prosperity of BIM, which result in obstruction of Internet of Things (IoT) and smart city advancing.

Though, highly fragmented project environment, being the nature of complexity, is the driven force to the problem. The direct reason to that is, main principle to follow is the information requirements from IDM while developing MVD/LoD for a specific case scenario and such principle is vague in a way if it involves the complex and super fragmented project. Hence, despite the information requirements remain same, different MVDs have been developed, not even mention the lack of comprehensive and unified information requirements for various complicated projects so far. These emerge the needs to find a proper way to derive consistent, comprehensive, and unified data standard development for complicated projects that can reduce ambiguity generated by information requirements in IDM.

2.2.2 Evolution of BIM Data Standard Testing Methodologies

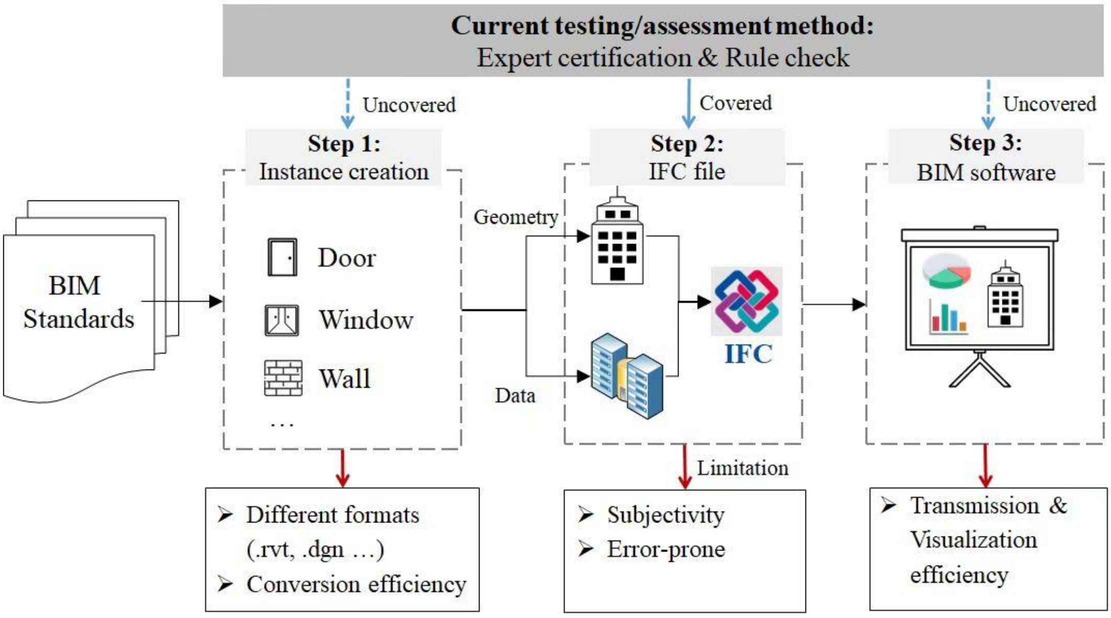


Figure 2 Simplified diagram of the BIM standard based typical BIM model generation and application.

The early concerns about the BIM standard testing focus on the interoperability of Industry Foundation Class (IFC), the certification approaches and novel ideas are mirrored from the healthcare and manufacturing domains for the AEC industry, in the paper, Amor suggested to establish the third party which is independent of the IAI to accomplish testing and certification, further he promoted the development an open tool for file checking (Amor, 2008). The similar urge was pointed out by Kiviniemi in 2008 (Kiviniemi, 2008). To address the interoperability issue more precisely, Solihin et al examined the standard user workflow and refined the issues as certification—closed loop process, criteria for assessment—not fully well-defined, all errors—different levels of impact on downstream, geometry tests—rely on manual evaluation(Solihin et al., 2015). With the evolvement of certification, the complete Model View

Definition is documented as a contribution to define test cases with precise exchange requirements and instructions for software developers to optimize their application, the exported files are checked based on 3 categories of tests: IFC schema syntax and where rules, rules from implementer's agreement, and limited numbers of simple semantic checks based on the MVD (Chipman, 2013). Similarly, Digital Alchemy developed the process to conduct GSA Concept CD BIM 2010 which involves spatial program validation, energy performance analysis, etc. (Sanguinetti et al., 2009). Another validation promoted by the US Army Corps Engineers in 2013 passed down a same believe that using MVD to assist facility management(William East et al., 2013). Likewise, British scholars managed to integrate existing technologies and use open standards as the basis of common data environment to validate the delivery of information models, information models against the requirements and the applicability of information in facility management(Patacas et al., 2020). To make it more specific, Liu et al trialed the theory by integrating BIM and fabrication workflow to interoperate the computerized design and prefabrication automation of steel reinforcement(Y. Liu et al., 2021). While considering the MVD, IFC shcema and user case scenarios, Yi brought up ease of implementation, ease of use, expressivity, and performance as main metrics to comparatively assess BIM model(Y. An, 2017).

Followed the idea of validating the IFC model, fellow scholars started to put more focus on well-formedness/syntactic correctness of data in an IFC model. They select a few data quality dimensions those fit IFC quality criteria to adopt, these dimensions mainly concern intrinsic (complete, meaningful, correct) and representational (conformance, ambiguous). And based on these dimensions, the rules for measuring the quality of IFC data was proposed(Solihin et al., 2015; Strong et al., 1997; Wand & Wang, 1996). Thus far, there have been several successful attempts on evolution process of BIM standard testing methodologies, including a) certification reduced plenty of issues and notably improve the quality of the exchanged models; b) manual effort is vastly reduced with the automated rule checking and the correctness of model data is highly assured; c) applicability to the actual project is enhanced.

To understand the gap of BIM standard testing technologies in the use case, it is important that we look at the typical user workflow. Figure 2 shows simplified diagram of the BIM standard based typical BIM model generation and application in use case scenario. It can be seen from An et al, there are 3 major stages from BIM standard development stage to recipient utilizing BIM model stage with various BIM software for their needs. The first major issue about current testing/assessment technologies is that failure of covering the multiple stages. In other words, it targets at IFC rather than other involved BIM standards (.rvt, .dgn) from previous stages, which result in the negligence of mistakes generated in the instance creation stage or the conversion process between relatively less developed BIM standard to exchange IFC file. The limitation would be amplified while the projects get complicated, requires more exchange cases with large data from various sources. Thus, the current testing/assessment methods are not robust enough on that term. Secondly, while use existing testing methodologies to assess BIM standard, the data interoperability check is the main subject. Though, there are a few research discussing other aspects such as expressivity, ease of implementation. The measurement is not fully well-defined and quantifiable. The lack of quantifiability hinders the capability of generalization. Moreover, current validation in the use case scenario relies on the applicability, and the validating cases are primarily with limited data. The first level of comfort for the end users is reducing issues undoubtedly. Yet, great applicability cannot equal to the great comfort as performance of BIM model should have considered the transmission and visualization efficiency while large data involved. The large data cases are often seen in practice. In stage 3, working with huge BIM model on various BIM software struggles in terms of visualization/transmission and storage/maintenance, and the current testing/assessment methods have not dabbled those two aspects. Ergo, it is lack of comprehension in certain way.

Despite the primary contributions previous BIM testing related research made, more inspirational advancement they brought up is to attempt referring other domains those face same issues of the same magnitude at AECO. And these attempts led the further development of BIM standard greatly. In 2008, Amor surveyed Healthcare and Manufacturing domains and proposed a) to establish a body independent of the IAI to manage conformance testing and certification against the IFC standard, b) to encourage companies within the ISO-STEP community to repurpose the tools developed there for model checking and comparison within A/E/C-FM, c) Investigate the possibility of a data interoperability lab for software vendors in the major software development regions(Amor, 2008). Later, Beetz et al discussed the applicability of ontologies to the problems in the building information modelling, in which, the ontologies, underlying principles are obtained from communication domain(Beetz et al., 2009). With such, great deal of research to enrich the BIM standard are conducted by the fellow BIMers(Y. An, 2017; H. Liu et al., 2016; Y. Liu et al., 2021; Ma & Liu, 2018; Ren et al., 2021). In 2015, Solihin et al introduced intrinsic data quality focused IFC testing criteria from Information System domain, they shrink the data quality dimensions from 4 categories to 2, i.e., the intrinsic quality of the IFC model and the domain context of representational quality(Solihin et al., 2015; Wand & Wang, 1996).

From these successfully attempts regarding implanting novel ideas and technologies from other domains to BIM, it can be clearly seen, the industries those provide lessons for BIM possess similar topological features to the BIM, i.e., ontologies are applied semantic enabler of communication between both users and applications in fragmented, heterogeneous multinational business, Building Information Modelling deals with the same fragmented, heterogeneous data; data quality is an important topic in the information system domain, BIM model cannot focus less on data quality. And when it comes to the image processing domain, topological similarity of the problem between the image processing and BIM modelling is, considering BIM standard would affect associate BIM model performance(visualization), how to compress the intricate detail while consume less computer memory to improve visualization and transmission efficiency. To address that matter, Barnsley and Sloan proposed fractal image compression to significantly improve efficiency televised image and computer image transmission(Barnsley & Sloan, 1988). The idea behind fractal image compression is utilizing the property of self-similarities within images those consist of small parts similar to itself or to some big part in them(Barnsley, 1993). The method is improved by Jacquin with recurrent iterated function systems, with which, the fractal image compression was empowered with more flexibility(Jacquin, 1992). And the partition of Jacquin was improved by Fisher(Fisher, 2012). Followed the development of fractal image compression, Sheridan proposed a hexagonal structure called the Spiral Architecture(Sheridan, 1996). Bouboulis et al introduced an image compression scheme using fractal interpolation surfaces which are attractors of some recurrent iterated function systems(Drakopoulos et al., 2006). Hence, it is reasonable to learn from image processing domain regarding the utilization of fractal to improve efficiency information transmission and visualization for BIM researchers developing BIM standard testing.

2.3 Information Theory and Its Engineering Application in Complex Information/Dataset Management Domain

In the ever-evolving landscape of the Architecture, Engineering, and Construction (AEC) industry, the quest for efficient and effective communication technologies remains paramount. Building Information Modeling (BIM), a pivotal communication technology, has emerged as a cornerstone in this endeavor, acting as a bridge that translates real-world projects into machine-readable data. However, as with any transformative technology, BIM is not without its challenges. The development of data standards, crucial for the seamless integration and interoperability of BIM across various platforms and stakeholders, has been a topic of intense research and debate. This chapter delves into the intricate dynamics of BIM data standard development, exploring its complexities through the lens of information theory and the concept of entropy.

Information theory, with its roots in quantifying the essence of information and uncertainty, offers a unique perspective on the challenges faced in BIM standard development. The concept of entropy, in particular, serves as a powerful tool to measure and manage the vast amounts of information inherent in BIM projects. However, the application of information theory to BIM is not straightforward. The dynamic nature of the AEC industry, coupled with the multifaceted challenges of BIM, calls for a nuanced and interdisciplinary approach. Information theory, a domain pioneered by Claude Shannon in the mid-20th century, has revolutionized our understanding of communication, data transmission, and the very essence of information. Central to this theory is the concept of 'entropy,' a measure that captures the uncertainty or unpredictability associated with random variables. In the context of information theory, entropy quantifies the amount of uncertainty or surprise associated with outcomes (Shannon, 1948). The problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the message has a meaning that is a key to or is correlated to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not only the one which will be chosen since this is unknown at the time of design. If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley, the most natural choice is the logarithmic function (Shannon, 1948). In the nature of communication, building information modelling can be regarded as one of the main information and communication technologies in the AEC industry. The building information model is media that carries the information from the real-world project to the machine-readable data, hence, the information model like the tunnel lining data model, fits the description of a communication fundamental. Ergo, it cannot be more natural to apply information-theory based tools to measure the information in the tunnel lining model. The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly "bits", a word suggested by John Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states 2^N and $\log_2 2^N = N$ (Shannon, 1948).

With such definition of information entropy, the information theory is further enriched in many aspects. The first of many would be quantifying uncertainty. Entropy serves as a metric to quantify the uncertainty inherent in a set of outcomes. A system with higher entropy is more unpredictable, and thus, on average, more bits are required to represent its outcomes(Bensky, 2019). The second one is optimal coding. Shannon's source coding theorem posits that the entropy of a source is the lower bound on the average length of the shortest possible representation of the source's outcomes. This has profound implications for data compression, suggesting that entropy determines the limit to how much a set of data can be compressed without loss. The third significance would be channel capacity and Transmission: In the realm of communication, entropy plays a pivotal role in determining the channel capacity – the maximum rate at which information can be transmitted over a communication channel without error (An, Lin, et al., 2023b; Shannon, 1948). The difference between the entropy of the input and the conditional entropy of the input given the output represents the mutual information, a measure of the amount of information shared between the input and output. Another significance would be bridging thermodynamics and Information. The concept of entropy in information theory shares parallels with the entropy in thermodynamics, a measure of disorder or randomness. This has led to interdisciplinary explorations, particularly in the domains of statistical mechanics and quantum computing (Shannon, 1948).

On the other side, in many other fields, existing literatures have clearly point out the use of information theory thinking proficiently improves information expression and management (Barbosa, Maciel, & de Azevedo Marques, 2010).

Since the basis of mange information efficiently would be quantification of information. And as described by Shannon's information theory, the quantification of information is a basic and powerful tool that can be applied to various fields, such as communication, statistics, and generic computer science (Uda, 2020; G. Wang, 2011; Yusof & Man, 2016), and its impact has been crucial to the success of the Voyager missions to deep space, the invention of the compact disc, the feasibility of mobile phones and the development of the Internet. (Ben-Naim, 2023; Haken, 2006; Jean-Yves Chouinard, Paul Fortier, & Gulliver, 1996; Pierce, 1980; Yeung, 2008)Leite et al employ information theory as a basis of quantifiers to increase capability of detecting components' faults while analyzing noise in wind turbines operations (de Novaes Pires Leite et al., 2021).

Blokh and Stambler bring up the measures of entropy and mutual information from information theory for the study of aging and aging-related diseases. They prove the information theory could provide insight into the nature of aging as a problem of deregulation (Blokh & Stambler, 2017). Angulo et al incorporation of the concepts developed in the Information Theory (entropy, complexity, etc.) with the aim of quantifying the variation of the uncertainty associated with a stochastic physical system resident in a spatiotemporal region (Angulo, 2018). Zhang et al proposed an error bit rate analysis of digital image watermarking based on information theory. Their analysis indicate it will keep a lower lever if its payload capacity is less than the channel capacity (Zhang, Zhang, Cao, Li, & Li, 2013). Preston builds an framework based on information theory to motivate child welfare case managers (Preston, 2013). Raloczy applies informational theory to the analysis of the grinding process under action of transverse rotating magnetic field. His model could produce adequate prediction of the particle size distribution (Rakoczy, 2010). Moniz et al use information theory techniques on time series of abundances to determine the topology of a food web. It shows the efficacy of their methods with decreasing time series size (Moniz, Cooch, Ellner, Nichols, & Nichols, 2007).

Ludwig et al introduce a new information-theoretic methodology for choosing variables and their time lags in a prediction setting in the case of predicting oil flow, particularly when neural networks are used in non-linear modeling (Ludwig, Nunes, Araújo, Schnitman, & Lepikson, 2009). Burnham and Anderson used Shannon's information theory/ the quantification of information to conduct statistical inference more accurately (Anderson, 2002), Delgado-Bonal and Martic-Torres applied it on the human vision research (Delgado-Bonal & Martín-Torres, 2016), Passalis and Tefas optimized information retrieval with information theory (Passalis & Tefas, 2016), and Gerhing et al. utilized it on quantum computing very recently (Gehring et al., 2021). The key to these fundamental applications is the measure of information—entropy, which was brought up by Claude Shannon in "A Mathematical Theory of Communication", he defined the quantity of information produced by a source using a formula similar to the equation that defines thermodynamic entropy in physics, thereafter, set the revolutionary cornerstone of digitizing of information. Besides that, Shannon offered a way to evaluate the ability to send information through a communication channel—the bandwidth of the channel (Shannon, 1948).

Thus far, it can be obviously seen that using information theory in research would have many significant advantages. First one would be the quantitative analysis, Information theory provides a quantitative framework for analyzing data, which is particularly useful for researchers who want to quantify the amount of information contained in a dataset or communication system (Pierce, 1980). Second known leverage of information theory that could be added to research is Interdisciplinary approach, given Information theory is a highly interdisciplinary field, encompassing mathematics, statistics, computer science, engineering, and other areas. This means that researchers from a wide range of backgrounds can use information theory to address diverse research questions (Ben-Naim, 2023). Also, information theory could be universally applicable. Information theory has universal applicability to any system or process that involves the transmission or processing of information. This means that it can be applied to a wide range of research questions and problems (Pierce, 1980). Further, information theory provides efficient methods and algorithms for analyzing large amounts of data. This can save time and resources for researchers who need to analyze large datasets (Jean-Yves Chouinard et al., 1996). By the same token, Information theory provides insights into the complexity of systems, including communication systems, biological systems, and social systems. This can help researchers to better understand how these systems function and how they can be improved (Haken, 2006; Yeung, 2008).

2.4 Chaos Theory and Its Application in Real World

Chaos theory, often referred to as the study of unpredictable and complex systems, has its roots in mathematics and physics (Lorenz, 1963). At its core, chaos theory seeks to understand the behavior of deterministic nonlinear systems that appear random or chaotic. Despite the inherent unpredictability of these systems, chaos theory posits that there is an underlying order and deterministic laws that can be discovered (Spencer, 2021).

One of the foundational concepts in chaos theory is the butterfly effect, which suggests that a small change in one state of a deterministic nonlinear system can result in vast differences in a later state (Lorenz, 1963). This concept underscores the sensitivity of chaotic systems to initial conditions, making long-term prediction nearly impossible. However, it also highlights the interconnectedness of elements within a system, suggesting that seemingly insignificant factors can have profound impacts on outcomes. Another essential concept in chaos theory is the attractor. Attractors are sets toward which a system tends to evolve over time, regardless of the starting conditions (Ruelle & Dewitt-Morette, 1990). They represent the long-term behavior of a system. In the context of chaos, strange attractors are often observed, which are characterized by their fractal structures. Fractals, in turn, are complex structures built from simple repetitions, and they exhibit self-similarity across different scales (Mandelbrot, 1983). This means that a small part of a fractal can resemble the whole, a property that has been observed in various natural phenomena. Following the idea, Mezei and Sarosi ocuses on the out-of-time ordered four-point function as a probe of chaos and operator growth in many-body quantum systems. The authors discuss the growth of chaos effects and their organization along rays, characterized by a velocity-dependent Lyapunov exponent (Mezei & Sárosi, 2020).

The concept of renormalization, borrowed from the realm of statistical mechanics and quantum field theory, has also found its place in the study of chaotic systems (Wilson, 1983). Renormalization provides a technique to study the scale invariance and self-similarity of systems, allowing researchers to understand the behavior of systems at different scales. Self-organizing criticality (SOC) is another principle closely related to chaos theory. It describes how complex systems can evolve to a critical state, where a minor event can lead to a significant system-wide change or cascade (Bak, Tang, & Wiesenfeld, 1987b). This concept has been observed in various natural systems, from avalanches to forest fires, and underscores the inherent unpredictability yet deterministic nature of chaotic systems.

The real-world applications of chaos theory are vast and varied. In the realm of business strategy, chaos theory has been applied to understand the recurrent patterns and exceptions in models, emphasizing the need for adaptability in strategic planning (Levy, 2007). The application of chaos theory in supply chain management has also been explored, suggesting that the principles of chaos can enhance supply chain solutions by adapting to inconsistent and non-static conditions (Stapleton, Hanna, & Ross, 2006). In meteorology, Lorenz's work on the butterfly effect has led to a better understanding of weather systems and the inherent challenges in long-term weather prediction (Lorenz, 1963). In biology, chaos theory has been used to study the population dynamics of species, revealing that seemingly random fluctuations in population sizes can be attributed to deterministic yet chaotic factors (May, 1976). The field of psychology has not remained untouched by chaos theory. The fluidic nature of the mind and its complex behaviors have been analyzed through the lens of chaos, offering a fresh perspective on psychoanalytic phenomena (Mg, 1994).

In the realm of engineering, chaos theory has been applied to optimize various processes. For instance, the principles of SOC and attractors have been used to enhance network designs, ensuring robustness and adaptability in the face of unpredictable demands (Y.-Y. Liu, Slotine, & Barabási, 2011). Similarly, the concepts of fractals and renormalization have been employed in materials science to develop materials with unique properties that are scale-invariant (Meakin, 2011). Furthermore, the paradigm of complexity introduced by chaos theory has been applied in various real-life situations, emphasizing that uncertainty is a norm in our daily lives (Cambel, 1992). Advances in chaos theory have also led to the development of intelligent control techniques, showcasing its multidisciplinary applications (Azar & Vaidyanathan, 2016). Hellen and Thomas’s research explores the production of chaotic behavior from difference equations with unstable fixed points. The authors present predictions and experimental results for the transient responses of a first difference-based feedback control method applied to a chaotic finite difference 1-dimensional map (Hellen & Thomas, 2008). Choueiri et al emphasizes the theoretical foundations of chaos predominantly laid out for finite-dimensional dynamical systems. The authors demonstrate the hydrodynamic pilot-wave systems as a bridge between low- and high-dimensional chaotic phenomena (Choueiri et al., 2022).

In economics, chaos theory has provided insights into the unpredictable nature of financial markets. Researchers have identified chaotic behaviors in stock prices, suggesting that while these systems are inherently unpredictable, they are not entirely random and can be understood using the principles of chaos theory (Peters, 1996). Poon et al studies the challenge of distinguishing chaotic from stochastic fluctuations in short experimental recordings, a fundamental issue in nonlinear dynamics and statistical physics. The authors emphasize the ambiguity arising from real-world data corrupted by measurement noise or perturbed by deterministic or stochastic inputs (Poon, Li, & Wu, 2010). Similarly, Vlad et al highlights the significance of nonlinearities in mathematical models and the surprising symmetry found within chaos. The authors present three models with chaotic features, including a nonlinear feedback profit model, a simulation model for exchange rates, and an application of chaos theory in capital markets(Vlad, Dumitru, & Pascu, 2007).

In conclusion, chaos theory, with its foundational concepts like the butterfly effect, attractors, fractals, renormalization, and self-organizing criticality, offers a lens through which the complexity and unpredictability of various systems can be understood. While the applications of chaos theory are vast, spanning from meteorology to economics, its principles provide a framework that can potentially be applied to optimize processes in various fields, setting the stage for future research endeavors. And Building Information Modeling (BIM) has revolutionized the construction and architectural industries by providing a comprehensive digital representation of the physical and functional characteristics of a facility. The development of BIM data standards is crucial for ensuring interoperability, consistency, and accuracy in the representation and exchange of information. However, the process of developing these standards is inherently complex, given the multitude of factors, stakeholders, and dynamic variables involved.

Drawing parallels from the literature review, it becomes evident that systems with intricate interdependencies, such as BIM data standard development, can benefit from the principles of chaos theory. the principles of chaos theory, as elucidated in the literature review, offer a promising framework for navigating the complexities of BIM data standard development. By recognizing the inherent unpredictability and intricacies of the process, and by harnessing the insights provided by chaos theory, it becomes possible to develop BIM data standards that are robust, adaptive, and future-proof. These outline the nature of uncertainty in BIM standard, wherein research would be built on endogenous uncertainty. It refers to uncertainty that originates from within the system or model. It is the inherent variability or unpredictability in a system's internal components or dynamics. This is different from exogenous uncertainty, which is related to external factors that influence the system or model. These are uncertainties that arise from outside the system, such as environmental changes, economic shifts, or policy changes, which are not controlled by the system itself. In a chaotic system, uncertainty is a fundamental characteristic, not just due to lack of knowledge, but because of the system's sensitive dependence on initial conditions. Even if variable information mass of project appears to be independent and data structural form of data exchange standard dependent, in a chaotic system, small variations in information mass can lead to disproportionate and unpredictable changes in data structure of exchange data standard. This sensitivity within BIM model means that long-term prediction of the system's state becomes practically impossible. Understanding this form of uncertainty requires acknowledging that the system's current state can only offer limited information about its future states, regardless of how precisely these variables are measured or modeled. As the construction and architectural industries continue to evolve, the marriage of chaos theory and BIM data standard development holds the potential to drive innovations and advancements that are both transformative and sustainable.

2.5 Conclusion and Remarks

The literature review has provided a comprehensive exploration into the intricate dynamics of Building Information Modeling (BIM) within the Architecture, Engineering, and Construction (AEC) industry. By delving into the evolution of data standards and the profound implications of information and chaos theories, a holistic understanding of the challenges and opportunities that lie ahead for BIM and the broader AEC landscape is provided.

Evolution of BIM and Data Standards

The historical trajectory of BIM data exchange standards, beginning in the 1990s, underscores the industry's commitment to achieving seamless interoperability. The inception of the Industry Foundation Classes (IFC) and its subsequent iterations have been instrumental in ensuring that BIM models maintain consistency across diverse software platforms. The adoption of such standards, especially by significant entities like the USA General Services Administration, is a testament to their importance and relevance. However, as BIM continues to evolve, the development of these standards cannot remain static. The diversity of projects and the increasing complexity of architectural endeavors necessitate a continuous refinement of these standards.

Implications of Information Theory

Claude Shannon's information theory, with entropy at its core, offers invaluable insights into the challenges of BIM standard development. The quantification of information and the understanding of uncertainty are pivotal in navigating the complexities of data exchange in BIM. The AEC industry's dynamic nature demands a nuanced, data-driven approach, and information theory provides the tools for such an endeavor. The universality of this theory, as evidenced by its wide-ranging applications, suggests its potential in refining and enhancing

BIM processes. Starting from an information science perspective, not quantifying uncertainty in information modeling and management can significantly impair efficiency as the volume of data increases within an existing information model. Without proper measures to account for and mitigate uncertainty, the influx of new information can lead to inaccuracies, misinterpretations, and decision-making delays. This inefficiency arises because the system becomes overloaded with data that cannot be effectively processed, analyzed, or utilized, hindering the overall performance and reliability of the information model. Consequently, addressing uncertainty is pivotal in ensuring that information systems remain robust, agile, and capable of supporting complex decision-making processes. On the other hand, incorporating the diversity of stakeholders and their varying data sources significantly increases the complexity of data structures and volumes within existing data/information models. The myriad of perspectives and requirements from different stakeholders introduces a range of data types and formats, making the management and integration of this information a challenging task. By quantifying uncertainty, it becomes feasible to organize and prioritize the input from this diversity more effectively. This methodological approach helps in streamlining data processing, enhancing the clarity of information flow, and ensuring that the information model remains coherent and aligned with project goals, despite the inherent complexity and volume of data involved.

Chaos Theory and the AEC Landscape

Chaos theory, emphasizing deterministic yet unpredictable systems, provides a fresh lens to view the challenges inherent to the AEC industry. The mature techniques to deal with complex, which highlights the profound impacts of dynamical systems, is especially variety in variables and involvements of stakeholders/restrictions in construction projects where initial decisions can have cascading effects. This theory serves as a reminder of the intricate interplay of variables in BIM and the need for meticulous planning and adaptability.

Looking Ahead

The confluence of data standards, information theory, and chaos theory in the realm of BIM suggests a future AEC industry that is interconnected, adaptive, and profoundly data-centric. As we move forward, several considerations emerge:

Collaborative Standardization: The evolution of the AEC industry demands ongoing collaboration among stakeholders to refine and develop comprehensive data standards.

Interdisciplinary Integration: The application of diverse theories to BIM highlights the potential of interdisciplinary research. There's immense value in fostering collaborations across fields to drive innovation in BIM processes.

Capacity Building: As BIM becomes increasingly complex, there's a pressing need to equip professionals with the skills to navigate its intricacies. This calls for targeted educational initiatives and training programs.

In summation, the literature underscores a transformative phase for the AEC industry. By leveraging the insights from data standards, information theory, and chaos theory, there's an opportunity to redefine processes, foster collaboration, and achieve unparalleled efficiency in construction projects. The path forward, while laden with challenges, holds the promise of innovation and growth for all industry stakeholders.

Chapter 3 – Methodology

3.1 Introduction

In the dynamic field of BIM data standard development, the quest for structured order amid complexity is paramount. Standardized data structures are fundamental to the Architecture, Engineering, Construction, and Operations (AECO) industry's transformation facilitated by Building Information Modeling (BIM). However, beneath the apparent order lies a complex and fluid system influenced by intricate underlying dynamics. This chapter lays the methodological foundation for an exploration into the dynamics inherent in BIM data standard development, seeking to unveil hidden patterns and relationships that shape the industry's landscape. The objective is twofold: firstly, to detect and understand the underlying principles that govern this intricate system and secondly, to provide rigorous mathematical proof of their existence.

Our journey begins with an inquiry into sandpile simulations, Chaos Theory, Information Theory, and Methodological Pragmatism, which underpin this research. These theoretical frameworks guide our subsequent phases, framing our investigation within a broader context. The following phase examines the underlying dynamics within data standard developments for BIM, endeavoring to connect theoretical principles with practical standards. Through computational simulations and innovative mapping techniques, we connect abstract concepts with tangible industry standards, revealing previously obscured patterns and structures. The third phase focuses on the rigorous mathematical verification of these patterns. Using applied mathematical techniques such as time series data generation, time delay determination, dimension analysis, and Lyapunov exponent calculation, we seek to provide undeniable proof of the underlying principles. This verification forms a robust foundation for the subsequent application of these principles to BIM data standards. Finally, the last phase extends the research to practical applications within the field of advanced engineering. Through MVD generation and BIM data standard evaluation, we demonstrate how these newly identified principles can optimize BIM processes, enhance system performance, and drive innovation in the AECO industry.

This chapter is an exploration into the dynamic nature of BIM data standard development, aimed at revealing the intricacies and uncovering the principles that underpin it. The results promise to reshape our understanding of BIM, opening doors to new opportunities and efficiencies within the AECO sector. As we delve deeper into the components of this methodological framework, it is essential to maintain our focus on the ultimate goal: to harness these principles for the benefit of the AECO industry. In the forthcoming sections, we will dissect each phase of the framework, revealing insights, addressing challenges, and considering implications. Through this methodological inquiry, we aim to contribute valuable knowledge that will advance the BIM industry into a more efficient, effective, and adaptable future.

3.2 Research Design and Framework

Considering the research process involves many domains' theories (ID3, information theory, chaos theory, sandpile simulation) and tools other than the ones from AECMO, like all methodological framework, the fundamentals to them are the assumptions those set the basic propositional logic for the whole system. It is necessary to set reasonable research assumption as basis, build dynamical system to simulate the relation between information mass and information exchange standards' topological structure, develop new mapping system and notion system for conversion from non-mathematical language in practical term to mathematical language as preparation of mathematical modelling. Then, with the new mapping system and notion system, the proof would be provided to indicate dynamical system chaotic. To ensure the research process is exploratorily/explanatorily conducted, this section specifically articulates major research steps/hypotheses in a clear and direct way. Hence, the assumptions are made at the beginning of methodology, and their True value of proposition are justified through firstly verifying them with real world settings and logical thinking which are provided right after the assumption making.

3.2.1 Research Assumption

To ensure the research process is exploratorily/explanatorily conducted, this paper only consider the standard topological structures. Six general research assumptions are imposed as follow:

i. The different types of information contained in the tunnel lining are presumed independent and identically distributed random variables, this assumption is used to eliminate the misunderstanding when calculating the information mass with Shannon information entropy.

ii. Topology of IFC schema, in this paper, refers to the tree structure which is consist of nodes and branches. Theoretically, the inheritance hierarchy of a BIM standard is tree structure while the semantic aspects are excluded from the specific cases. Considering the application of BIM standards in practice, this tree structure is a weighted directed graph, that can be referred as tree hierarchical combination according to semantic meaning and the ontology, and each node corresponds to different engineering entities. The weight is the information mass that the node accommodates, and the direction represents its level in the tree. Based on that, the topological structure/topology of data exchange standard is predefined as a tree structure that can be used to create links (branches-Ifcrelations) among objects/subjects (nodes-Ifcentity).

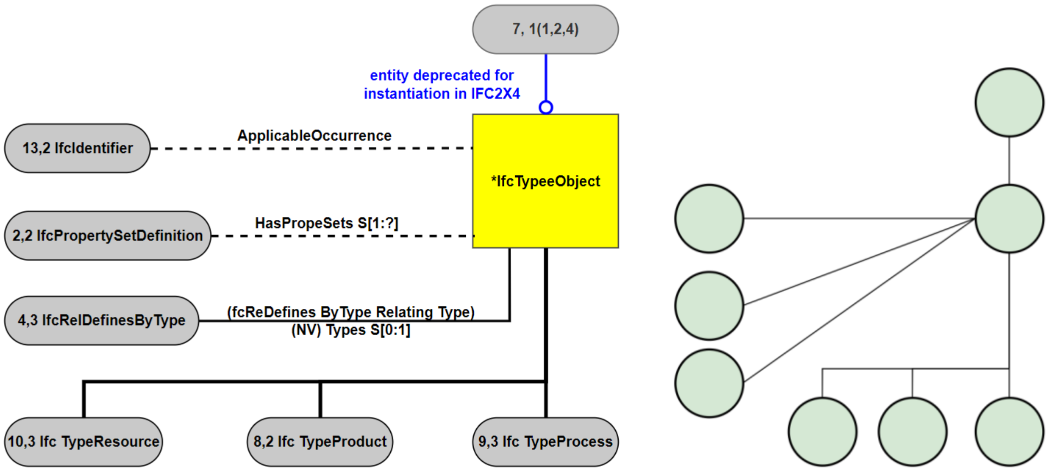
iii. Considering the assumption of Iterative Dichotomiser 3 (ID3) algorithmic thinking, entropy can determine if the node within the schema (consider the schema as a structure) needs to be further extended. While this paper relies only on the value of entropy, ID3 can split nodes when nodes have reached a maximum information gain as nodes assumed to have become too “informative” require more information to clarify themselves. And information entropy would act as an information topological invariant of a project, which could be seen as a metaphorical URI of a project, different information requirements of a project would produce different information mass (information entropy) hence rising different needs for organizing information. Accordingly, the driven force for further development on topological tree structure is the information entropy value of the entity.

iv. With above reasonable assumption that the topology of BIM standard (tree structure) is essentially a weighted directed graph, the layout of sandpile lattice is topologically same as tree structure. The research conducted by An et al (2023) exploited the topological resemblance to propose a mapping method to map the layout of sandpile lattice to BIM standard. On the basis of that, the BIM standard development could be studied further with sandpile simulation. In such system, the information mass could be regarded as grains of sand.

v. Time series is an ordered sequence of values of a variable at equally spaced time intervals. To address the issue this discussed, this dissertation would put attention on discrete time maps. This is really no restriction as in some sense all analysis of physical systems takes place in discrete time: we never sample anything continuously. And while conduct sandpile simulation, the layout of sandpile lattice would varies with the drop of sands. Hence, each gap between grains of sand drop would be taken as the sampling interval. In that sense, the time intervals are equally replaced by the grain of sand dropping speed.

vi. Greater projects contain more information and higher standard projects require more detailed information, as the source of information, the BIM model file becomes larger, hence the need for organizing information is greater. The research is built on the basis of assuming that efficiency of different information standard varies with information mass. An example that would be that neither big information mass project fitting in simple information standard nor little information fitting in a complex information standard account for efficient. Hinge on that assumption, presume the dynamical system would be the evolution of node split entropy which is used to evaluate the complexity of tree structure over information mass (grains of sand drop). Noted that the state variable is node split entropy of tree structure, and it will evolve over information mass increasing.

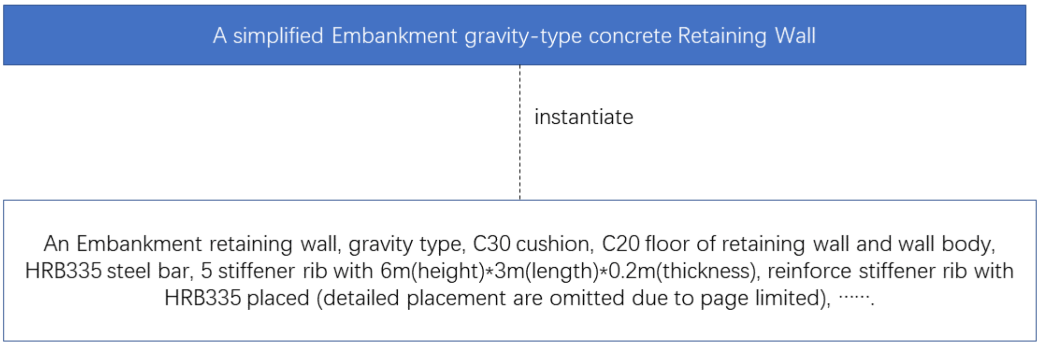
In a real-world BIM setting, the assumption (i) is default as true to comply Ockham’s Razor. Assumption (ii) is consistent with analogy of BIM standard development initiative. Especially, in the case of IFC, metaphorize IFC entity as node and relationship as branches, and it aligns with each other as shown in Figure 3. The tree structure is a directed graph, the IFC inheritance schema is a directed graph too. The similar analogy is used in Gan’s research to extract BIM-based graph data model and in Borrmann et al ‵s book-Building Information Modelling (Borrmann, Koch,Beetz, 2018; Gan, 2022).



(a) (b)

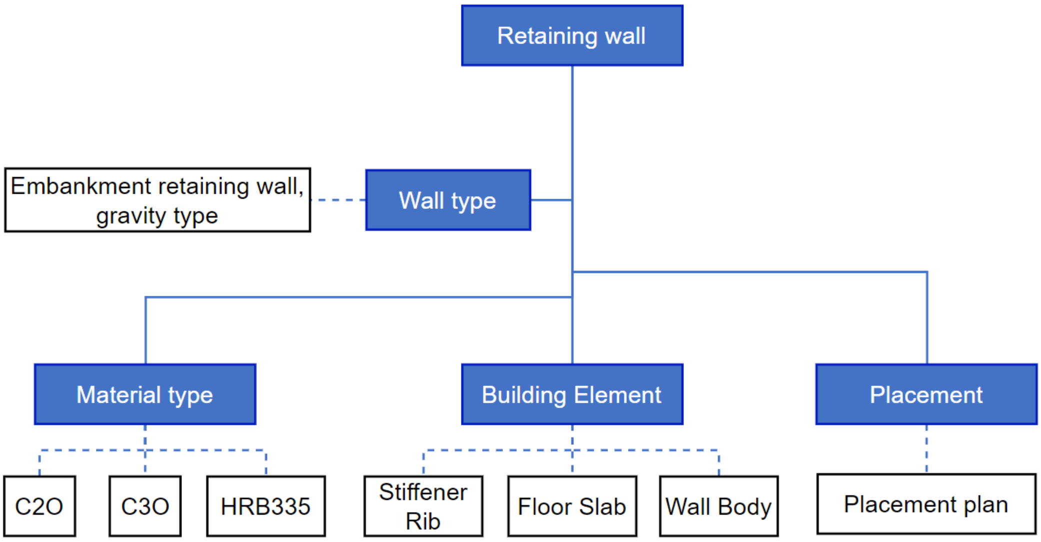
Figure 3 A IFC Kernel schema b associate topology of IFCKernel in tree structure

Assumption (iii) is very near to reality, take the Koch's IFC model for Shield Tunnel and Li's IFC model for shield Tunnel (Koch, Vonthron, & König, 2017; Lin, 2016). Both models are developed to manage lifecycle information of shield tunnel projects. And both of them are verified through the practical project, Koch's worked in metro project Wehrhahn-linie in Dusseldorf, Germany and Li's worked in Shanghai Metro Line 12. The information mass for the whole metro line is considerably bigger than a metro station. Both models specifically developed the tunnel lining information model. Yet, their need for organizing information is different, consequently, their product model varies. This assumption is, in essence, an abstract form in theoretical level of BIM-based thinking in practical level. While information of a building element entity is much, reorganizing and restructuring the way of expressing information is considered as better option in terms of efficiency in management. For instance, a retaining wall design, without assumption (iv) it would be organized as Figure 4a and with assumption (iii) would be organized as Figure 4b. It can obviously be seen that even with information omitted, the way organizing the information as Figure 4b is much better than the one shows in Figure 4a. Hence, as stated in assumption, the bigger information mass it is, more splitting it will be. To be more concise, the metaphorized example given in Figure 4 stresses the need of organizing while dealing with massive information mass. Noticeably, the sole purpose of assumption (iii) is to make this point clear, rather than discussing specific stage/phase of project. No matter it is design phase, construction phase, or information modeling phase, it always requires more organizability to deal with bigger information mass. This can be, again, verified by Koch's Shield Tunnel IFC model and Lin's Tunnel IFC model, which both target the lifecycle of actual projects (Koch et al, 2017; Lin 2016). Since the information mass of their projects is different, the IFC they developed for instantiating their models is different, which revealed different levels of organizability of their data structure. Therefore, the assumption (iii) is proposed reasonably.



An Embankment retaining wall, gravity type, C30 cushion, C20 floor of retaining wall and wall body, HRB335 steel bar, 5 stiffener rib with 6m(height)*3m(length)*0.2m(thickness), reinforce stiffener rib with HRB335 placed (detailed placement are omitted due to page limited), ······.

(a)



(b) some information is omitted but will not affect the demonstration

Figure 4 a retaining wall design (a). organizing information without assumption iv, (b). organizing information with assumption (iii).

Assumptions (iv) and (v) will be validated as reasonable via proposed mapping system in Section 5.3.2. Assumption (vi) is obviously intuitive.

3.2.2 Methodological Framework

Figure 5 shows the methodological framework of the whole research. The overarching narrative of the dissertation unfurls with a clear and compelling mission—to elucidate the presence of chaos within the dynamical system of BIM data standard development, particularly in the context of information mass. This mission is not merely an academic exercise; it carries profound implications for the BIM industry at large. By showcasing the chaotic nature of this system, the research aspires to open the doors to a wealth of chaos-based techniques, algorithms, and theories that have already reshaped industries and sectors far beyond the realm of construction.

Information theory initiated configuration

To embark on this intellectual journey, a structured approach is imperative. The methodological framework is systematically divided into four key components, each serving as a foundational building block. These components collectively lead to a comprehensive comprehension of chaos within the realm of BIM data standard development. The initial component, denoted as 'Information Theory Initiated Configuration,' encompasses the preparatory and configurational phase, laying the groundwork for subsequent empirical investigations. i. In the context of dissertation aimed at unraveling the complexities of chaos within Building Information Modeling (BIM) data standard development, the initial phase of our methodological journey serves as a fundamental cornerstone. This stage, referred to as 'Initiating Information Theory-Based Configuration', establishes the essential foundation for a comprehensive exploration that bridges the gap between theoretical abstraction and practical inquiry. The methodological journey commences with a detailed examination of Model View Definitions (MVDs), a foundational element within the realm of BIM data standard development. MVDs, often likened to the DNA of the BIM ecosystem, encapsulate the core principles of data standards. This initial phase aims to unravel the intricacies, reveal the underlying logic, and identify information requirements and end user needs within their structural frameworks.

Subsequently, the research shifts its focus to identifying information requirements, a critical aspect of governing BIM data standard development. These requirements, derived from industry regulations and specifications, serve as the foundational bedrock upon which entities and relations among them will be explored. This sub-task unveils the interplay between information theory and regulatory frameworks, shedding light on their mutual influence.

As the exploration deepens, regulations in the form of precast specifications, PPVC handbooks, and codes of practice emerge as pivotal elements within the crucible of BIM data standard development. And these regulations are designed and agreed upon different stakeholders' need across different stages of life cycle. Therefore, it can be seen that employing these regulations and practice of code would not overlook stakeholders' needs. These regulations act as external forces, shaping and molding the evolving chaos within the system. Subsequent exploration involves a thorough examination of these regulations, recognizing their integral role in influencing the fabric of information mass. The research seamlessly transitions into the analytical phase, where a nuanced understanding of information requirements and regulations forms the basis for further investigation. This phase encompasses a comprehensive exploration of identified information requirements, dissecting them to uncover nuances, complexities, and interdependencies lurking beneath the surface. As the analysis of information requirements unfolds, chaos takes on a quantifiable form known as "information mass." The subsequent sub-task centers on quantifying contained/required information, measuring its influence, and providing empirical evidence of its existence. Here, the profound insights of information theory find tangible expression, bridging the gap between theory and empirical evidence.

The culmination of this inaugural phase involves the initiation of sandpile simulations, an innovative approach led to identification of chaos theory. These simulations are configured to replicate the dynamics of chaos within BIM data standard development, marking a significant milestone. Meticulous calibration of these simulations ensures their fidelity in reproducing chaotic dynamics. This orchestration of sub-tasks fosters a harmonious interplay, where theory and practice, abstraction and empiricism converge to establish the foundational framework of the dissertation. This framework holds the promise of unveiling the secrets of chaos within the intricate domain of BIM data standard development. As we progress further in the methodological journey, this initial phase serves as a springboard, propelling us into the heart of empirical exploration. It is the stage where chaos theory and information theory intersect, setting the stage for a transformative exploration that has the potential to reshape our understanding of BIM and its associated standards.

Discovering chaos in data standard development for BIM

The second segment of our process encompasses a critical phase in our research, dedicated to investigating chaos within BIM data standard development. This phase builds upon the foundational work established in the preceding section, advancing our exploration of the complex domain of BIM standards. At the onset of this phase, we are presented with the results of our sandpile simulations, precisely labeled as "Simulation Results." These simulations transcend numerical exercises; they provide insights into dynamic interactions and emergent behaviors governed by chaos theory within the context of BIM data standards. Here, theoretical concepts intersect with empirical observations as we encounter tangible manifestations of chaos. Our research progresses methodically with the development of the "Proposed Mapping System." This integral component acts as the linchpin bridging the abstract world of sandpile simulations to the concrete domain of BIM data standards. The objective is to establish a seamless connection, crafting a mapping system that proficiently translates the intricate patterns unveiled in the simulation results, embedded within the sandpile lattice, into the organized structure of BIM standards, often likened to a tree structure.

The significance of this undertaking cannot be overstated. The mapping system serves as the metaphorical Rosetta Stone of our research—a tool designed to decode the language of chaos and translate it into the vocabulary of BIM standards. This translation process is pivotal as it signifies the initial point of convergence between the chaos elucidated in our simulations and the established standards governing the AECO industry. Following the establishment of the mapping system, we embark on a rigorous analytical exploration into the patterns, trends, and anomalies disclosed by our simulations. Our objectives extend beyond surface-level observations of chaos; instead, we delve profoundly into its dynamics. Within this analytical framework, chaos assumes a more structured identity, revealing underlying principles and behaviors.

Our analysis extends further into the core of chaos theory, specifically, "similarity cross scalability." This phenomenon represents a pivotal tipping point in complex systems, where stability suddenly transforms into rapid, cascading changes. This concept, intricately linked to chaos theory, becomes the focal point of our inquiry. Identifying "similarity cross scalability" within our simulation results represents a pivotal milestone in our research journey. The essence of this phase is the "Discovery of Chaos." This marks the focal point of our research—a moment of profound significance. Identifying "similarity cross scalability" within our simulations signifies the recognition of chaos itself within the BIM data standard development domain. It is akin to discovering the proverbial needle of chaos within the vast haystack of data and standards. This discovery represents more than an intellectual triumph; it signifies a transformational moment wherein chaos transitions from an abstract concept to a tangible, identifiable force within the BIM ecosystem. With chaos identified, we can confidently assert its presence within the realm of BIM data standards. The once-theoretical chaos now stands unveiled and documented within the systems underpinning the AECO industry.

In summary, the second segment of our process embodies an exploration—an expedition commencing with tangible results from sandpile simulations, progressing with the creation of a mapping system connecting chaos to BIM standards, conducting in-depth analyses of simulation results to uncover "similarity cross scalability," and culminating in the identification of chaos within the domain of BIM data standard development. As we advance in our exploration of chaos, this phase sets the stage for further investigations, analyses, and applications. The chaos we have uncovered holds the potential to reshape our comprehension of BIM standards, unlocking innovative approaches, problem-solving techniques, and optimizations previously concealed within the depths of chaos theory. It is a transformative journey, with upcoming phases building upon the foundations established here, unraveling more of the secrets concealed by chaos in BIM data standard development.

Proving Chaotic Properties in Data Standard Development for BIM

The third phase of our methodological framework represents a significant stage in our research, where we aim to empirically confirm the presence of chaos within the domain of Building Information Modeling (BIM) data standard development. This phase builds upon the foundations laid in the previous steps, advancing our exploration of BIM standards. This phase begins with the generation of time series data. This task leverages the dataset obtained from the earlier sandpile simulations. These simulations have provided us with essential time series data illustrating the dynamic evolution of BIM data standard development, effectively capturing the underlying chaos. To execute this task, we process the data obtained from the simulations, ensuring its suitability for time series analysis. We then generate time series data representing the temporal progression of key variables within BIM data standard development. Time series data generation is vital as it allows us to examine how system parameters change over time, providing a window into the dynamic behavior of BIM standards. This temporal perspective is essential for uncovering chaotic dynamics.

With the time series data in hand, our focus shifts to determining the optimal time delay. This step involves a detailed analysis of the data to identify the time delay parameter that best characterizes the system's dynamics. The time delay parameter plays a crucial role in constructing the reconstructed phase space, a fundamental concept in chaos theory. Selecting the right time delay ensures that the reconstructed phase space accurately represents the underlying chaos. To determine the time delay, we employ advanced mathematical techniques, such as mutual information functions or autocorrelation analysis. These methods help us identify the time delay that best captures temporal relationships within the data. Having established the time delay parameter, we proceed to determine the dimension of the reconstructed phase space. This dimensionality is a fundamental aspect of chaos theory and affects the accuracy of our empirical verification. The dimension of the reconstructed phase space indicates the number of variables needed to adequately represent the system's dynamics. It guides the construction of an accurate mathematical model of chaotic behavior within BIM data standard development. To determine dimensionality, we apply techniques like the false nearest neighbors algorithm or correlation dimension analysis. These methods uncover the underlying structure of the system by identifying the minimum number of variables required to capture its complexity.

With the time delay and dimension established, we create the reconstructed phase space. This phase space serves as a mathematical representation of chaotic behavior within BIM data standard development, transforming the abstract concept of chaos into a tangible framework. The reconstructed phase space is a multi-dimensional space where each dimension corresponds to a variable characterizing the system's behavior. It allows us to gain insights into the complex dynamics of the system. To construct the phase space, we use embedding techniques that transform time series data into a suitable format for analysis. These techniques, involving sliding windows and lagged coordinates, map the system's trajectory within the phase space. The reconstructed phase space is central to our empirical verification, providing a platform for applying advanced mathematical tools to validate the presence of chaos. The empirical verification terminates at calculating the Lyapunov exponent. This exponent is a fundamental indicator of chaos, providing quantifiable insight into the system's sensitivity to initial conditions. To calculate the Lyapunov exponent, we employ specialized algorithms and numerical techniques. This calculation involves evaluating the rate at which initially close trajectories within the reconstructed phase space diverge over time. A positive Lyapunov exponent indicates the presence of chaos, as it signifies exponential divergence—a hallmark of chaotic systems. The Lyapunov exponent is the conclusive evidence of chaotic properties within the BIM data standard development system. Its calculation empirically validates chaos, bridging the gap between theory and observation.

In summary, this phase represents an empirical journey, encompassing time series data generation, determination of critical parameters, and empirical validation of chaos within BIM data standard development. Empirical verification is essential for integrating chaos-based techniques and methodologies into BIM data standard development, going beyond theoretical concepts to practical applications. It marks the transformation of chaos from an abstract idea to a tangible, identifiable force within the realm of BIM standards. These empirical insights will serve as the foundation for further investigations and practical implementations, enhancing our understanding of BIM standards and offering innovative solutions within the Architecture, Engineering, Construction, and Operations (AECO) industry. As we progress, the knowledge gained from this phase will guide our research toward practical applications, where chaos-based techniques can reshape the landscape of BIM data standard development.

Verification and case studies with advanced engineering applications

In final phase of our methodological framework, we delve into practical applications and validation. Here, we apply the theoretical underpinnings and discoveries from earlier phases of our research to real-world engineering scenarios. The primary aim is to substantiate the relevance and applicability of chaos theory in the context of Building Information Modeling (BIM) data standard development. Within this part, two interrelated applications take the forefront. Central to our first application is the transformation of data, an essential step in making chaos theory applicable to BIM data standards. This transformation involves converting complex tree-structured graph data into Model View Definitions (MVDs), specifically within the context of tunnel lining representation. Through this conversion, we bridge the gap between abstract chaos theory and the practical world of BIM, showcasing how chaos theory can address real-world challenges. In the process of data transformation, we encounter the intricate interplay between the chaos-driven representation and the established BIM standards. It's here that we demonstrate the adaptability of chaos theory, as we translate complex data structures into MVDs that embrace the dynamism inherent in chaotic systems.

Following data transformation, we proceed to generate novel MVDs informed by chaos-driven insights and principles developed in earlier research phases. These new MVDs represent departures from conventional standards, embracing the dynamism inherent in chaotic systems. This application not only demonstrates innovation in BIM data standards but also highlights chaos theory's adaptability in practical engineering. As we generate these novel MVDs, we observe how chaos theory infuses new perspectives into the representation of tunnel lining data. The resulting MVDs, shaped by the principles of chaos, become a testament to the transformative potential of chaos theory in BIM data standard development. This application concludes with a critical comparative analysis. Here, newly generated chaos-informed MVDs undergo rigorous comparison with existing MVDs proposed by other scholars and industry experts. This comparison serves a dual purpose. Firstly, it evaluates the effectiveness and relevance of chaos-informed MVDs against established standards. Secondly, it provides empirical evidence of chaos theory's value in BIM data standard development. By showcasing improvements or efficiencies in chaos-driven MVDs, we further advocate for the integration of chaos theory into the BIM landscape.

The second application delves into BIM data standard evaluation. We employ the reconstructed phase space from prior research to localize and describe system attractors. These attractors serve as reference points that reveal the system's dynamics and stability, shedding light on its behaviors and tendencies. Within this process, we encounter the tangible manifestations of chaos theory within BIM data standard development. The localization and description of system attractors unveil patterns and trends that were previously hidden. This application reinforces chaos theory as a powerful tool for understanding and optimizing BIM standards. Building on our understanding of attractors, we define a novel chaos-informed indicator. This indicator quantifies the system's divergence from the attractor, using the Euclidean distance as a metric. The underlying principle is that systems further from their attractors demand more "energy" to maintain stability and functionality. This indicator offers a tangible measure of a BIM model's performance and stability within a chaos-driven framework.

As we define this indicator, we bridge the theoretical underpinnings of chaos theory with practical engineering assessments. The indicator provides a bridge that spans the gap between abstract theory and real-world BIM model evaluations. The final phase of this application centers around a practical case study. Here, a real-world BIM model undergoes evaluation using the chaos-informed indicator. By measuring the Euclidean distance to the attractor, we assess the model's performance and stability. This verification process is essential in demonstrating the real-world applicability and reliability of our chaos-driven indicator. In this case study, chaos theory's impact becomes palpable. It transforms how we evaluate BIM model performance, emphasizing the importance of stability and proximity to attractors. This practical validation reinforces chaos theory's relevance in BIM data standard evaluation.

In this phase, we transition from theoretical exploration to practical validation. Through these two applications, we bridge the gap between chaos theory and BIM data standards, demonstrating how chaos can be harnessed for innovation and efficiency. Moreover, these applications set the stage for the integration of chaos theory into the BIM ecosystem, offering pathways to improved standards and performance evaluation. As we progress in this final part of our methodological framework, we are poised to demonstrate that chaos theory is not just an abstract concept but a practical tool that can revolutionize Building Information Modeling. The applications presented here represent the culmination of our research, where theory meets practical application, and innovation transforms the landscape of BIM data standard development. This research, taken as a whole, serves as proof of chaos theory's potential to enhance and revolutionize BIM data standards and their applications. By weaving chaos theory into BIM's fabric, we anticipate far-reaching implications for the architecture, engineering, construction, and operations (AECO) industry. It is a transformative journey where theoretical exploration, computational simulations, data analysis, and practical applications merge to usher in a new era of BIM innovation and excellence. In summary, the last part of our methodological framework, rooted in advanced engineering applications, brings our research to life. By introducing chaos theory to BIM data standards, we uncover opportunities for innovation, efficiency, and enhanced performance. Through data transformation, new MVD generation, and comparative analysis, we demonstrate chaos theory's practical relevance. Our exploration of system attractors and the creation of a chaos-informed indicator offer fresh perspectives on BIM model evaluation. The case studies within this part lay the foundations for a future where chaos theory is integral to BIM standard development and evaluation.

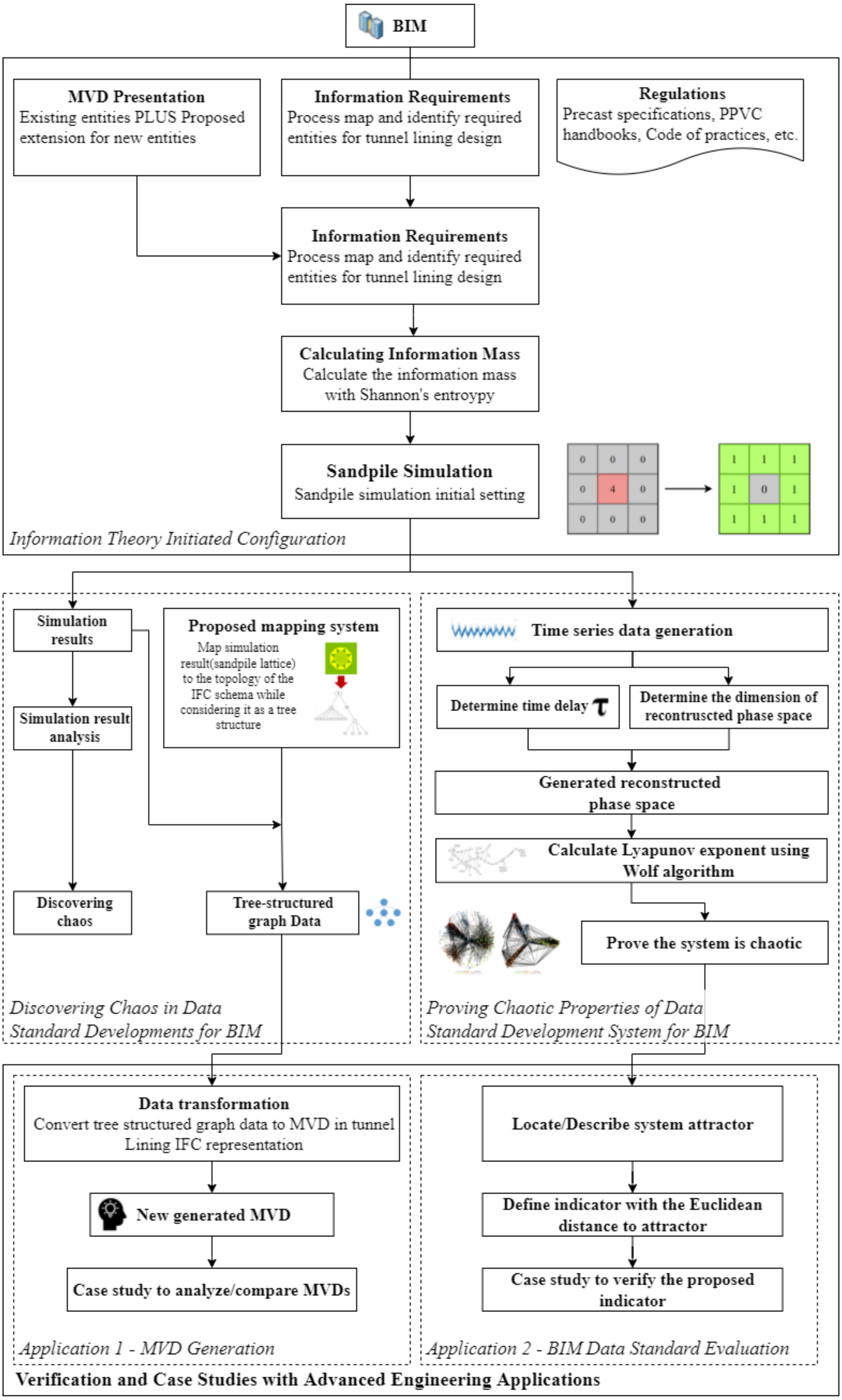


Figure 5 Overarching Methodological Framework of Research

3.3 Conclusion and Remarks

In this chapter, we have navigated through the methodological framework that underpins our research endeavor to uncover chaos within the domain of BIM data standard development. Our framework serves as a blueprint for our research journey, guiding us through the intricate landscape of data standards in the Architecture, Engineering, Construction, and Operations (AECO) industry. As we conclude this chapter, several key points and overarching themes emerge.

Integration of Chaos Theory and Information Theory: Central to our approach is the integration of chaos theory and information theory. Chaos theory provides the lens through which we view the complex dynamics of BIM data standards, recognizing the presence of chaos as a driving force. Information theory bridges the gap between quantitative measurements and qualitative insights, offering a holistic framework for comprehending the intricacies of information mass within BIM standards. The synergy between these theories forms the philosophical foundation of our research.

Foundational Components: The chapter has unveiled the four foundational components of our research framework. These components prepare the groundwork for subsequent explorations, from the initial preparation and configuration to the chaos discovery process and rigorous mathematical proof. They guide us through the integration of chaos theory into BIM standards and extend our research into practical applications.

Transformative Potential: This methodological framework not only lays the foundation for our research but also holds transformative potential for the AECO industry. By identifying chaos within data standards, we open doors to innovative approaches, problem-solving techniques, and optimizations. Our findings have the capacity to reshape the way BIM standards are understood and applied, introducing a new era of efficiency and effectiveness.

Future Directions: As we move forward, this framework will continue to guide our research. It offers a structured path to uncovering chaos within BIM data standards and leveraging this knowledge for practical applications. The subsequent chapters will delve deeper into each component, providing detailed insights, analyses, and case studies to substantiate our claims.

In conclusion, this chapter has established the methodological underpinning of our research, illustrating the intricacies and interconnectedness of the components within the framework. It sets the stage for a rigorous, comprehensive, and transformative exploration of chaos within the domain of BIM data standard development. As we proceed, the knowledge gained through this framework will serve as a beacon, illuminating the way forward in our pursuit of understanding and harnessing chaos in the AECO industry.

Chapter 4 – Theoretical Essence of Research Methodology

4.1 Introduction

The purpose of this chapter is to articulate the objectives and emphasize the critical role of philosophical and theoretical underpinnings in guiding the broader research endeavor. Specifically, this chapter serves as a nexus for the philosophical and methodological frameworks that inform our investigation into Building Information Modeling (BIM) and data standards. It aims to clarify the complex relationships among chaos theory, information theory, and methodological pragmatism, each of which uniquely contributes to both the intellectual and practical aspects of the research.

The primary objectives of this chapter are manifold:

1. To elucidate the epistemological landscape of BIM and data standards, thereby facilitating a nuanced comprehension of these intricate domains.
2. To scrutinize the philosophical traditions of rationalism and empiricism, assessing their respective merits and limitations within the context of BIM research.
3. To advocate for the incorporation of chaos theory and information theory as transformative paradigms, offering novel insights into the complexities inherent in BIM and data standards.
4. To outline the pragmatic methodological approach employed in this research, highlighting its results-oriented focus and its congruence with the study's philosophical foundations.

The importance of philosophical and theoretical discussion in research is paramount. It serves as the foundational structure upon which the entire scholarly endeavor is built, offering a coherent and robust framework for inquiry (Cuthbertson, Robb, & Blair, 2020; Rovelli, 2018; Serpico, Lynch, & Porter, 2023). Far from being a mere assemblage of isolated facts or observations, such discussion and justification ensures that the research constitutes a cohesive, academically rigorous body of work. It establishes criteria for method selection, data interpretation, and the validation of research findings. Furthermore, it imbues the research with a level of depth and rigor that elevates it from mere investigation to scholarly discourse (Booth, Colomb, Williams, Bizup, & FitzGerald, 2016; Okasha, 2002). In the specific context of BIM and data standards, the role of philosophical and theoretical discussion gains heightened relevance. Given the complex, dynamic, and often volatile nature of these fields, a well-defined philosophical stance is indispensable for providing the necessary grounding and direction. This ensures that the research remains methodologically robust and intellectually rigorous.

4.2 The Nature of Knowledge in BIM and Data Standards

4.2.1 Discussion on the epistemological stance of the research

Constructivist Approach in the Context of Chaos Theory and Information Theory in BIM Standards

Constructivism posits that knowledge is not an objective entity but is constructed through human experiences and interpretations (Rorty, 1991; Savery & Duffy, 1995). This perspective becomes especially pertinent when studying complex systems like BIM standards, where multiple variables and stakeholders interact in intricate ways. These formulate the foundational beliefs of research. In the realm of Building Information Modeling (BIM), the standards are not static or universally agreed-upon templates. They evolve based on technological advancements, industry needs, and stakeholder feedback. Through a constructivist lens, BIM standards can be seen as dynamic constructs that are continuously shaped and reshaped by the collective experiences and inputs of the AECO (Architecture, Engineering, Construction, and owner operated) community. In such sense, BIM standards are considered as dynamic constructs. Accordingly, chaos theory, with its emphasis on non-linearity, unpredictability, and emergent properties, aligns well with the constructivist belief in a continuously evolving knowledge landscape. Just as small changes in initial conditions can lead to vastly different outcomes in chaotic systems (the butterfly effect), slight shifts in stakeholder perspectives or technological innovations can significantly alter the trajectory of BIM standard development. While it comes to Information entropy, metaphorized as grains of sand in the research, represents the inherent uncertainty and complexity in information systems. In a constructivist context, this entropy can be seen as the myriad interpretations, experiences, and inputs that stakeholders bring to the table. Each "grain" or unit of information entropy is a piece of constructed knowledge, and its interaction with other units leads to the emergent properties observed in sandpile simulations. Additionally, understanding BIM standards through a constructivist lens means recognizing the fluidity and subjectivity inherent in these standards. It suggests that there's no "one-size-fits-all" BIM standard but rather a continuously evolving framework that reflects the constructed knowledge of its community. This perspective underscores the importance of stakeholder engagement, iterative feedback loops, and adaptability in the research and development of BIM standards (An, Chen, Li, & Wang, 2023; An, Lin, et al., 2023b).

Empirical Realism: A Tangible Path to Understanding Chaos and Information in BIM Standards

Empirical realism, in nature, asserts that knowledge is rooted in observable phenomena and can be acquired through direct observation and experience. This philosophical stance contends that reality exists independently of human perception and can be studied through systematic empirical methods (Allzén, 2021; Lennox & Jurdi-Hage, 2017). In the context of this research, empirical realism becomes a guiding principle that ensures investigation into the chaotic properties of BIM standards is firmly grounded in tangible, measurable evidence. Meanwhile, it stands out due to such roots of empirical realism. This empowered the importing of sandpile simulation into study becoming more intrinsic regardless that no other fellow researchers ever studied on this but could accept this. The research's use of sandpile simulations exemplifies empirical realism in action. By designing and conducting these simulations, author is engaging in a process of active observation and experimentation. Each simulation run is an instance of direct engagement with the virtual representation of chaotic behavior within BIM standards. The simulations provide a controlled environment where author can manipulate variables, observe outcomes, and gather data that reflects real-world scenarios (An, Chen, et al., 2023; An, Lin, et al., 2023b). Also, in the realm of empirical realism, data is the currency of knowledge. The data generated from sandpile simulations serves as empirical evidence of the chaotic properties author is investigating. The patterns, trends, and emergent behaviors observed in the simulation outcomes become concrete indicators of how chaos manifests within the intricate landscape of BIM standards. This data-driven approach enhances the credibility and rigor of the research. Moreover, empirical realism ensures that the understanding author developed regarding chaotic properties in BIM standards is firmly grounded in concrete experiences and observations. The simulations provide a tangible window into the behavior of these standards under varying conditions. This understanding goes beyond theoretical speculation, offering a practical grasp of how chaos and information theory interplay within the context of smart infrastructure (An, Chen, et al., 2023; An, Lin, et al., 2023b).

The empirical realism perspective underscores the practical application of research findings (Allzén, 2021). The simulations offer insights that can have practical implications for optimizing data standards in smart infrastructure. By relying on empirical evidence, this research aims to produce knowledge that is not only theoretically sound but also relevant and actionable in real-world scenarios. Through these, adhering to empirical realism safeguards the epistemological integrity of this research. It ensures that conclusions are not based on conjecture or assumptions, but on observable phenomena. By embracing empirical realism, author demonstrate a commitment to a rigorous and evidence-based exploration of chaotic properties in BIM standards.

Interdisciplinary Integration: Navigating Complexity through Diverse Epistemological Tools

Interdisciplinary integration refers to the deliberate weaving of insights, methodologies, and perspectives from different disciplines to enhance understanding of complex phenomena (Wang, Mao, Lu, Cao, & Li, 2021; Zhang, 2023). It signifies a departure from a single-discipline approach and embraces the belief that multifaceted challenges like optimizing data standards in smart infrastructure require a diverse toolkit of epistemological lenses. By integrating concepts from chaos theory and information theory (Chaos and Information Theory as Epistemological Tools), this research manifests a profound epistemological openness. Chaos theory, with its focus on non-linearity, sensitivity to initial conditions, and emergence, provides a lens through which to explore the unpredictable behavior within BIM standards. Information theory, on the other hand, serves as a framework to quantify and manage the complexity inherent in data structures (An, Chen, et al., 2023; An, Lin, et al., 2023b). The integration of these concepts acknowledges the multidimensional complexity inherent in optimizing data standards for smart infrastructure. Traditional disciplinary boundaries often fall short when tackling such multifaceted challenges. By drawing from both chaos theory and information theory, this research navigates the intricate terrain where technological innovation, industry dynamics, and human interactions converge. The integration of chaos theory and information theory is not just juxtaposition; it's synergy. Chaos theory underscores the potential for unforeseen disruptions and nonlinear interactions within BIM standards. Information theory complements this by providing a framework to quantify uncertainty and complexity. Together, they offer a holistic perspective on how data standards evolve and adapt within dynamic environments.

Besides, Interdisciplinary integration embodies a respect for diverse ways of knowing and problem-solving. It recognizes that complex phenomena cannot be fully understood through a singular lens (Wang et al., 2021; Zhang, 2023). This stance values the contributions of different disciplines, fostering a more comprehensive and nuanced understanding of the intricate interplay between chaos, information, and BIM standards. The willingness to integrate concepts from diverse disciplines reflects an epistemological humility—an acknowledgment that no single perspective holds all the answers. This openness encourages researchers to adapt and refine their approaches based on the insights gained from various fields, creating a more robust foundation for understanding and addressing complex challenges. Interdisciplinary integration enriches this research by expanding the scope of knowledge horizons. It encourages us to draw connections between seemingly disparate ideas, revealing new insights that might remain hidden within disciplinary silos. The result is a more holistic, multidimensional understanding of chaos, information, and BIM standards.

Pragmatism: A Practical Path to Understanding and Application

Pragmatism is rooted in the belief that the value of knowledge is determined by its practical utility. It emphasizes the importance of focusing on methods and approaches that yield tangible results and can be applied to real-world challenges (Katsenelinboigen, 1980; Maiväli, 2015). In the context of this research, pragmatism serves as a guiding principle that places practical applicability at the forefront of exploration. This research's pragmatic approach is evident in its methodological choices, particularly the use of sandpile simulations. Pragmatism encourages researchers to choose methods based on their effectiveness in addressing research questions, rather than rigid adherence to a specific philosophical tradition. By opting for simulations, author prioritize a method that can provide actionable insights into the chaotic properties of BIM standards, this can be seen as methodological flexibility and efficacy (An, Lin, et al., 2023b). Another way of reviewing the pragmatism within the scope of this research, a. pragmatism seeks to strike a balance between theoretical exploration and practical application. While theoretical frameworks are important, pragmatism insists that their value is derived from their ability to generate insights that can inform decisions and actions (Katsenelinboigen, 1980; Maiväli, 2015). By focusing on the practical implications of chaos theory and information theory within BIM standards, the research aligns with the pragmatic philosophy; b. pragmatism aligns with the notion that knowledge holds value when it can be applied to real-world scenarios. This research's pragmatic orientation suggests a commitment to producing knowledge that has meaningful implications for the field of smart infrastructure. It recognizes that the insights gained from chaos theory and information theory can only truly fulfill their potential when they inform practical strategies for data standard optimization (An, Chen, et al., 2023; An, Lin, et al., 2023b); c. the pragmatic stance highlights author’s interest in generating solutions that address real challenges. This research aims to translate theoretical insights into actionable steps that enhance the quality and efficiency of data standards in smart infrastructure. This orientation resonates with pragmatism's core belief that knowledge's worth lies in its ability to bring about positive change; d. pragmatism encourages an open-mindedness towards adopting approaches that work best for the research's goals. It allows author to leverage insights from various philosophical traditions and methodologies, as long as they contribute to the practical advancement of understanding and optimizing BIM standards.

Last but not the least, Pragmatism aligns with research's epistemological blend of constructivism, empirical realism, and interdisciplinary integration. It reinforces the idea that knowledge is most valuable when it can be integrated across different perspectives and disciplines to create solutions that are both theoretically sound and practically applicable.

4.2.2 Knowledge in BIM and Data Standards: A Constructivist Perspective

Conceptualizing BIM Standards as Constructed Knowledge

Within the intricate realm of Building Information Modeling (BIM) and data standards lies a profound paradigm shift in how we perceive and interact with knowledge. Gone are the days when knowledge was a static entity; instead, it emerges as a dynamic construct, perpetually evolving through the intertwining threads of human interpretation, industry exigencies, and technological progress. An incisive exploration is described here, peering beyond the façade of BIM standards to unveil their true essence as products of collective cognition, interaction, and evolution.

The Canvas of Constructed Knowledge

At the heart of our constructivist perspective lies the resounding affirmation that knowledge is not a monolithic edifice, but a malleable, ever-changing tapestry woven through dynamic engagements. In the context of BIM standards, this perspective liberates them from the confines of being mere technical blueprints. Instead, they take on the hues of a living artwork, a canvas upon which architects, engineers, planners, and stakeholders inscribe their insights, experiences, and aspirations. This perspective illuminates the iterative nature of BIM standard development. It acknowledges that these standards are not rigid absolutes, but rather responsive entities that adapt to the shifting tides of industry demands. By embracing a constructivist lens, we recognize that BIM standards mirror the dynamic evolution of the Architecture, Engineering, Construction, and owner operated (AECO) community itself, perpetually shaped and reshaped by the currents of collective wisdom.

Stakeholders: Architects of Knowledge

In this constructivist journey, stakeholders emerge as active architects of knowledge, rather than passive recipients of standardized doctrines. Architects, engineers, contractors—all become co-authors of the evolving narrative that is BIM standards. This collaborative process finds a vivid manifestation in the challenges depicted in the first paper—namely, the intricate implementation of the IFC data model and the pivotal role of Model View Definitions (MVDs). The constructivist lens magnifies the collaborative nature of knowledge construction. Stakeholders convene, engage in discourse, and coalesce their perspectives into the Model View Definitions—a testament to the communal endeavor of shaping BIM standards. This discourse underscores the co-creation of knowledge, as each stakeholder weaves their threads of expertise into the fabric of BIM standards.

Navigating the Seas of Collaborative Wisdom

This constructivist exploration navigates the seas of collaborative wisdom that infuse BIM standards. It envisions knowledge not as an isolated pursuit, but a collaborative symphony, where diverse voices harmonize to craft standards that resonate with the complexity and dynamism of the industry. Stakeholders, in this view, become the artisans of knowledge, collectively sculpting a multidimensional narrative that encapsulates the evolving needs and ambitions of the AECO landscape.

As mentioned ahead, this constructivist perspective serves as our compass. It guides us to interpret BIM standards not as static artifacts, but as living testimonies to collective endeavor. The collaborative dance between stakeholders and the fluid evolution of these standards illuminates the intricate and ever-evolving tapestry of knowledge that underlies the realm of BIM and data standards.

4.3 Rationalism and Empiricism in BIM Research

In this segment, we delve into a profound juncture of philosophical inquiry—where rationalism and empiricism, two cardinal philosophical traditions, converge to shape the methodological and theoretical landscape of our research. These traditions, far from being abstract dispositions, are epistemological cornerstones that orchestrate the trajectory of our quest to decipher the intricate dance of chaos theory and information theory within the realm of optimizing data standards for smart infrastructure.

Rationalism: Illuminating the Realm of Abstract Constructs

Rationalism, revered for its celebration of reason and intellectual abstraction, emerges as a guiding compass (Angel, 1980). At its zenith is the conceptual framework of information entropy—a rationalist edifice that serves as a lodestar to navigate the labyrinthine corridors of BIM standards' complexity. This theoretical scaffold is not just an intellectual ornament; it is a cognitive framework that articulates the conceptual architecture of chaos and information theory within our context.

A Critical Lens on Rationalism's Scope:

Nonetheless, even as rationalism casts its illuminating light, we must adopt a critical gaze. The rationalist lens, while potent in facilitating abstract comprehension, demands scrutiny in its ability to encapsulate the dynamic and chaotic landscape of smart infrastructure. Its propensity to elevate reason may inadvertently veil us from the nuanced cacophony of chaos embedded in the multifaceted data standards.

Empiricism: Embodied Insight through Sensory Engagement

In tandem, empiricism unfurls its empirical tapestry—an epistemological stance rooted in sensory engagement and direct experience (Angel, 1980; Hage, 2007). This tangible manifestation finds embodiment in sandpile simulations—a calculated endeavor to distill chaos theory's abstract conjectures into observable phenomena. Empiricism endeavors to extract empirical insights from the alchemical crucible of simulations, grounding the abstract in the palpable.

A Critical Lens on Empiricism's Scope

However, as empiricism takes center stage, it is essential to exercise a critical perspective. While simulations offer empirical insights, their controlled environment grapples with the complexities inherent in the uncontrolled reality of the architecture, engineering, construction, and owner operated (AECO) domains. Empiricism's stronghold in observation encounters the challenge of navigating the intricate interplay of human choices, environmental variables, and industry dynamics.

Harmonizing Rationalism and Empiricism: The Nexus of Intellectual Inquiry

The crux of our research lies in the nuanced harmony achieved by fusing rationalism and empiricism—a philosophical and methodological synthesis that bridges the chasm between theoretical constructs and empirical observations. Rationalism lays the conceptual foundation, delineating the theoretical contours of chaos and information theory. Empiricism, in turn, bestows the empirical texture, enriching theoretical constructs with tangible data points.

This bridge between traditions is meticulously erected through a methodological concerto, choreographing a symphony where rationalism sets the theoretical stage, and empiricism provides the empirical resonance. It is an intricate dialectic that transcends the confines of philosophical categorization, yielding an understanding that transcends the abstract-real binary.

Practical Implications and Epistemological Enrichment

The matrimony of rationalism and empiricism extends beyond the realm of philosophical discourse, echoing in the practical corridors of AECO domains. The theoretical frameworks, when interwoven with empirical insights, become navigational instruments for the intricate waters of smart infrastructure. The research's value resounds not only in its theoretical elegance but also in its practical contributions, resonating in the nexus of chaos and information theory's intersection.

4.4 Complex Systems and Constructivist Epistemology

In this pivotal segment, our scholarly voyage embarks upon a profound exploration, traversing the intersecting realms of complex systems theory and constructivist epistemology. As we navigate this intellectual terrain, a resounding revelation awaits—a realization that transcends the conventional view of BIM standards and data structures as discrete entities. Instead, we unveil them as integral constituents of intricate, dynamic, and adaptive complex systems. These systems, resplendent with interwoven intricacies, transcend the static confines of their individual components, evolving through interconnectivity, emergence, and the enigmatic orchestration of non-linear interactions.

Delving Deeper into Complex Systems Theory

The profound tapestry of complex systems theory, a cornerstone of modern scientific discourse, presents itself as the bedrock upon which our exploration rests. An intricate web of concepts and paradigms, this theory proffers a lens through which we unravel the rich fabric of interconnected components that collectively embody BIM standards. With this vantage point, we transcend the limitations of reductionism, embracing the essence of the whole system that emerges from the intricate interplay of its parts.

Constructivist Epistemology: A Philosophical Portal

Intimately intertwined with this exposition is the mantle of constructivist epistemology—a philosophical architecture resonant with the very dynamics of complex systems (Knorr-Cetina, 1981). Constructivism, echoing through the corridors of epistemological discourse, asserts that knowledge is not a static commodity but a product of dynamic interactions within complex systems

(Ladyman, 2007). In the realm of BIM standards, this postulation asserts that knowledge is not a predetermined endowment; rather, it emerges through the ceaseless dance of iterative engagements, collaborative endeavors, and contextual negotiations among diverse stakeholders(An, Lin, et al., 2023b).

Critical Inquiry into Constructivist Epistemology

The constructivist tenet assumes a weighty significance in our pursuit. The metaphorical sands of sandpile simulations and the nuanced cadence of information entropy resonate profoundly within the constructivist paradigm. Information entropy morphs from a mathematical abstraction into a conceptual apparatus—allowing us to apprehend the intricate symphony of interactions that constitute BIM standards. This transformation from abstraction to applicability serves as a testament to the constructivist perspective's prowess in capturing the essence of complex systems' intricate dynamics.

Harmonizing Complexity and Constructivism: A Theoretical Confluence

The harmonious union of complex systems theory and constructivist epistemology bestows upon us a nuanced prism through which to decipher the tapestry of chaos and information theory within data standards. Complex systems theory grants us the audacity to peer into the labyrinthine corridors of chaos, illuminating the interwoven threads that intricately bind BIM standards. Constructivist epistemology, in turn, furnishes the cognitive scaffolding required to decipher the choreography of these threads—transforming chaos into orchestrated interactions that culminate in meaningful knowledge construction.

Scholarly Implications and Intellectual Significance

The amalgamation of complex systems theory and constructivist epistemology does not confine itself to intellectual musings—it reverberates across the academic sphere with profound implications. It elevates our comprehension of BIM standards from mere technical blueprints to vibrant, living entities ensconced within the dynamic embrace of complex systems. The constructivist epistemological lens imbues our empirical insights from sandpile simulations with philosophical depth, intertwining theory and practice in a symphony of scholarly discourse.

4.5 Chaos Theory: A Paradigm Shift in Understanding

Unveiling Chaos Theory's Transformative Impact

In this section, we embark on an intellectual odyssey that takes us beyond the realm of linear and deterministic models of understanding. At the forefront of this journey lies chaos theory—an avant-garde paradigm that redefines our perception of complexity, uncertainty, and emergence within the domain of BIM and data standards(An, Chen, et al., 2023; An, Lin, et al., 2023b). As we traverse this uncharted terrain, we confront the seismic shift that chaos theory introduces, unshackling us from traditional confines and inviting us to embrace a new lens through which to decipher the intricacies of chaotic systems.

Chaos Theory: Departure from Linear Determinism

Central to our discourse is the profound departure from the mechanistic dogmas of linear determinism. Chaos theory, with its roots tracing back to the pioneering work of Edward Lorenz, transmutes our understanding of dynamic systems. Rather than adhering to linear predictability and deterministic outcomes, chaos theory celebrates the unpredictable dance of non-linear interactions (Lv et al, 2005; Sprott, 2003; Williams, 1997). It introduces us to a world where minuscule perturbations can unleash a cascade of profound consequences, rendering predictability futile in the face of chaotic dynamics (Sprott, 2003).

Critical Scrutiny of Linear Determinism

Linear determinism, while once an intellectual stronghold, now stands under the scrutiny of chaos theory's incisive critique. The deterministic paradigm's attempt to reduce complexity to predictable cause-and-effect chains is shaken by the revelation that seemingly insignificant initial conditions can reverberate into wildly divergent outcomes (Schuldberg & Guisinger, 2020). The deterministic facade crumbles as chaos theory unveils the inherent limits of our ability to predict and control complex systems with unwavering precision (S. E. Vlad, 2023).

Philosophical Implications: Chaos as a Catalyst

The embrace of chaos theory in the study of BIM and data standards engenders profound philosophical implications. The juxtaposition of chaos and order becomes a philosophical conundrum—a contemplation of how intricate patterns emerge from what appears to be randomness. The philosophical ripples extend beyond the scientific realm, echoing in the corridors of epistemology, ontology, and even our perception of reality itself.

Navigating Chaos's Philosophical Landscape

The very act of integrating chaos theory within our discourse demands philosophical introspection. Chaos theory challenges our conventional understanding of knowledge, questioning the boundaries of determinism and indeterminism. It beckons us to embrace uncertainty as an inherent facet of our quest for comprehension (Schuldberg & Guisinger, 2020; Vlad, 2023). The philosophical voyage thus encompasses not only the unraveling of chaotic phenomena but the contemplation of knowledge's very essence within the turbulence of chaos.

Chaos Theory's Transcendent Impact

In essence, chaos theory's emergence as a paradigm shift signifies more than a theoretical revolution—it is a clarion call to reevaluate our methods, assumptions, and aspirations. It prompts us to traverse the intellectual horizon with renewed vigor, prepared to navigate the uncharted waters of complexity (Schuldberg & Guisinger, 2020; Vlad, 2023). Chaos theory begets not just an evolution of understanding, but a metamorphosis in our philosophical stance—a realization that our journey within the realm of BIM and data standards is a perpetual dance of chaos and order, uncertainty and insight. Chaos theory, with its emphasis on non-linearity and unpredictability, offers a fresh perspective on BIM and data standards. By moving away from linear and deterministic models, and by embracing the complexities and intricacies of chaos, researchers can gain a deeper and more comprehensive understanding of BIM systems. This paradigm shift not only has practical implications for the development of better BIM systems and standards but also challenges us philosophically, prompting us to reconsider our epistemological assumptions and our understanding of reality in the realm of BIM research.

4.6 Information Theory: Bridging Quantitative and Qualitative Realms

In this subsection, we discuss the intricate terrain of information theory—a domain that holds the key to bridging the chasm between the quantitative and qualitative realms within the context of BIM standards. This exploratory quest delves not merely into the mathematical intricacies, but also the profound philosophical underpinnings that underlie the very fabric of information theory. We unravel how this theory becomes a metaphysical bridge—a conduit through which quantitative measurements find resonance with qualitative meaning, enabling a holistic comprehension of the multifaceted landscape of BIM standards (An, Lin, et al., 2023b). Information theory, pioneered by Claude Shannon in the mid-20th century, has traditionally been associated with the quantification of information, particularly in the context of communication systems(Shannon, 1948). However, its philosophical underpinnings extend far beyond mere quantification, delving into the very nature of knowledge, understanding, and meaning. At its core, information theory grapples with the dichotomy between the tangible and intangible, the measurable and the immeasurable, offering a framework that transcends these boundaries. Central to our discourse is a meticulous examination of the philosophical bedrock upon which information theory stands. A product of Claude Shannon's pioneering insights, information theory transcends its mathematical origins to assume a profound philosophical dimension. It propounds that information is not just data points or measurements, but a vessel of meaning—a bearer of significance that transcends numerical values (Shannon, 1948).

In the context of Building Information Modeling (BIM) and data standards, quantitative data often emerges from simulations, measurements, and standardized metrics. This data, with its precise values and structured format, provides a tangible foundation upon which decisions can be based. It offers a semblance of objectivity, allowing for systematic analysis, comparison, and validation. Yet, while its precision is invaluable, quantitative data alone can be devoid of context, meaning, and nuance. Contrasting the precision of quantitative data is the depth and richness of qualitative understanding (An, Chen, et al., 2023; An, Lin, et al., 2023b). In the realm of BIM standards, qualitative understanding delves into the implications, interpretations, and meanings associated with data. It concerns itself with the "why" and "how" rather than just the "what." Qualitative insights provide context, illuminating the nuances and intricacies that quantitative data might overlook. It captures the experiential, interpretative, and subjective dimensions of knowledge.

Information theory emerges as a bridge between these two realms. It recognizes that information is not merely about data points but also about the context and meaning associated with those data points. The bridge crafted by information theory is not just a pragmatic tool—it is a theoretical construct that offers insights into the very essence of knowledge itself. As quantitative data streams from simulations, information theory imbues these streams with semantic resonance. It unravels the latent patterns, unveiling the hidden stories and insights that these data streams encapsulate, transforming them from numerical abstractions into conduits of understanding. In doing so, it offers a holistic framework that integrates the quantitative and qualitative, grounding abstract concepts in measurable metrics while preserving the depth and nuance of qualitative insights, offering a framework to transmute numerical data into a coherent narrative—a narrative woven from the threads of meaning, significance, and implication.

The philosophical significance of this bridge extends beyond the pragmatic utility. It delves into the epistemological essence of knowledge acquisition—how information theory serves as a mediating force, intertwining empirical data with conceptual constructs. This interaction between quantitative and qualitative domains mirrors the intricate choreography of chaos within BIM standards, revealing the profound interplay between seemingly disparate elements. In essence, information theory becomes a mediator between the dimensions of quantity and quality, transforming raw numerical data into knowledge rich with meaning and significance. This philosophical and methodological harmony signifies a transcendence—a movement beyond mere measurement to the realm of comprehension and insight. For BIM and data standards, the integration of information theory signifies a paradigm shift. It suggests that standards should not merely be about prescribing precise measurements and metrics but also about capturing the broader context, implications, and meanings associated with those metrics. By embracing information theory, BIM researchers and practitioners can develop standards that are both rigorous in their precision and rich in their interpretative depth.

In a nutshell, Information theory, with its philosophical roots and practical applications, offers a transformative framework for BIM and data standards. By bridging the quantitative and qualitative realms, it challenges us to rethink our approach to information, knowledge, and understanding. In doing so, it paves the way for more holistic, comprehensive, and meaningful standards that capture the full spectrum of knowledge, from the tangible to the intangible, from the measurable to the meaningful.

4.7 Methodological Pragmatism

Methodological pragmatism, at its core, emphasizes the practical application of knowledge. It prioritizes the utility and efficacy of methods over strict allegiance to any singular philosophical or theoretical tradition (Charmaz, 2001; Shannon-Baker, 2023). In the context of this research, pragmatism manifests in the selection of methods like sandpile simulations, chosen not for their adherence to a particular philosophical stance, but for their relevance and effectiveness in addressing the research questions at hand.

Traditional research methodologies often advocate for a strict adherence to established paradigms, potentially at the expense of the research's applicability or relevance (Biswas, Rhodes, McCrea, & Zino, 2023; Charmaz, 2001; Smith, 2020). However, the dynamic and multifaceted nature of BIM and data standards necessitates a more flexible approach. By adopting sandpile simulations and other innovative methods, this research sidesteps the constraints of traditional paradigms, focusing instead on the tools most apt for the task. This pragmatic approach acknowledges that the most effective methodological tools may span across various philosophical traditions. Also, such pragmatic approach envisions methods as tools in a dynamic toolkit, to be wielded not merely for the sake of tradition, but in pursuit of actionable insights and meaningful outcomes. The embrace of methodological pragmatism presents a trove of benefits. It shatters the confinement of adhering strictly to one philosophical tradition, allowing the research to tap into diverse methodologies that resonate with the complexity of the research questions (Biswas et al., 2023; Charmaz, 2001; Shannon-Baker, 2023). This eclectic blend bridges the chasm between theoretical frameworks and real-world applications, grounding abstractions in tangible methodologies that yield empirical data. To be more specific, pragmatism would bring in adaptability in terms of offering the flexibility to adapt to the evolving demands of the research, ensuring that the methodology remains relevant and effective throughout the research process; pragmatism would provide holistic understanding by drawing from a diverse array of methods, it facilitates a more comprehensive understanding of the research topic, capturing its nuances from multiple angles; pragmatism allocate practical relevance to the research so that ensures that the research remains grounded in practicality, producing findings that are not only theoretically sound but also directly applicable in real-world contexts.

In adopting a pragmatic approach, the research acknowledges both its strengths and its potential pitfalls. It becomes imperative to strike a balance, ensuring that the flexibility of pragmatism does not compromise the research's depth or coherence. By judiciously selecting methods based on their relevance and efficacy, and by critically reflecting on their integration, the research ensures that methodological pragmatism serves as an asset rather than a liability. The research's adoption of methodological pragmatism harmonizes with the philosophical depth that underscores the entire dissertation. Instead of dwelling solely within the realms of constructivism, empirical realism, or other individual traditions, methodological pragmatism emerges as a chalice that captures the essence of multiple epistemologies. The synthesis amalgamates the practicality of chosen methods with the intellectual rigor of philosophical contemplation. In the grand tapestry of research, methodological pragmatism stitches together disparate threads into a cohesive whole. It forges a dynamic balance between philosophical traditions and practical applications, fostering a symbiotic relationship that fortifies the research endeavor. This pragmatic ethos resonates with the very essence of the research—chaos theory, information theory, and their interplay within BIM standards—a dynamic dance that transcends rigidity and embraces the fluidity of complexity.

In conclusion, methodological pragmatism stands as a beacon that guides the research towards its ultimate destination. It invites us to navigate the ever-evolving waters of inquiry with a compass that values both theoretical foundations and pragmatic applicability. By embracing this approach, the research manifests as a holistic voyage—a scholarly expedition that transcends the boundaries of philosophical singularities and embarks on a pragmatic odyssey toward comprehensive understanding. While it challenges traditional notions of theoretical adherence, it also opens up new avenues for exploration, innovation, and understanding. In the realm of BIM and data standards, where the landscape is constantly evolving, such a pragmatic approach proves not only beneficial but essential.

4.8 Conclusion

As we reach the terminus of this intellectually invigorating chapter, it becomes imperative to pause and reflect on the intricate philosophical and methodological tapestry that has been meticulously woven. This chapter has served as an intellectual crucible, fusing the transformative paradigms of chaos theory, information theory, and methodological pragmatism into a cohesive narrative that underpins the research on Building Information Modeling (BIM) and data standards.

Summarizing the Chapter's Key Points

Chaos Theory: Our scholarly expedition has delved into the labyrinthine contours of chaos theory, challenging the traditional bastions of linear determinism. We have unearthed its profound implications for understanding the complex dynamics inherent in BIM and data standards. The embrace of chaos theory has not only ushered in a new dawn of understanding but has also harmonized with the constructivist epistemology discussed in section 4.2, both emphasizing the emergent, non-linear, and unpredictable nature of complex systems.

Information Theory: Our journey further led us to the philosophical underpinnings of information theory, which serves as a bridge between the quantitative and qualitative realms. This theoretical scaffold, as elaborated in section 4.6, extends beyond mere quantification, offering a holistic framework that melds empirical data with the profound nuances of qualitative insights. It resonates with the empirical realism and rationalism discussed in section 4.3, providing a balanced approach to understanding the complexities of BIM and data standards.

Methodological Pragmatism: As we navigated the pragmatic waters of research methodology, we embraced an efficacy-driven approach, as detailed in section 4.7. This pragmatic stance allowed us to transcend the limitations of singular philosophical traditions, aligning perfectly with the interdisciplinary integration emphasized in the third uploaded file. It has fortified our research endeavor, providing a robust methodological framework that is both practically applicable and intellectually rigorous.

The Imperative of Philosophical Justification

The philosophical justifications illuminated in this chapter serve as both the anchor and the compass for our research. They ground the inquiry within established paradigms, providing a robust foundation upon which the research is built. Simultaneously, these justifications guide the research through the labyrinthine complexities of BIM and data standards, ensuring alignment with objectives while navigating the inherent unpredictability and non-linearity. The philosophical underpinnings thus serve a dual role: they not only provide a coherent and relevant framework but also infuse the research with depth and resonance.

Looking Forward

As we venture into the uncharted territories that lie ahead, the insights gleaned from this chapter will serve as guiding stars. The interplay between chaos theory, information theory, and methodological pragmatism has furnished the research with a rich palette of possibilities. It sets the stage for a rigorous, enlightening, and holistic exploration of BIM and data standards—a realm where philosophy and practice converge in an elegant dance of complexity and understanding.

In summation, this chapter emerges as the philosophical and methodological cornerstone upon which the research can ascend. Through the interplay of paradigms and the meticulous justification of philosophical choices, the chapter sets the stage for a scholarly expedition into the domain of BIM and data standards.

Chapter 5 – Discovering Chaos in Data Standard Development for Building Information Modeling

5.1 Introduction

This chapter introduces the sandpile simulation as a representative model of complex systems, drawing parallels with the challenges and dynamics of BIM standard development. The sandpile model, characterized by its Self-Organized Criticality (SOC), provides insights into the inherent chaos and unpredictability of BIM data standard development. By mapping the results of the sandpile simulation to the Model View Definition (MVD) within BIM, this chapter proposes a novel approach to understanding and managing the complexities of BIM standard development.

As we navigate through this chapter, we will embark on a journey that bridges the gap between abstract theoretical constructs and their tangible applications in BIM. Through a rigorous exploration of information theory, entropy, and the sandpile simulation, this chapter aims to shed light on the challenges of BIM data standard development and propose innovative solutions grounded in scientific research.

5.2 Information Entropy: The initial setup for linking Sandpile Simulations with BIM Model

Owing to above significance of information entropy, considering the findings of

Chapter 2 with respect to the limitation of existing data standard development for building information modeling—mainly lack of quantifiability and consisteny to either regulate MVD development or assess/test developed BIM standard, information entropy would be utilized to lever the feature of quantification and consistency. Take MVD development as an example, the major limitation of MVD development in tunneling project is lack of principles to regulate the MVD/linked data/LoD could lead to even more significantly diverse version for applications. All these would hinder the development of IFC, thus impede prosperity of BIM, which result in obstruction of Internet of Things (IoT) and smart city advancing. In other words, the ambiguity and inconsistency exist in the process of MVD development, which leads to the current situation as pointed out—same project would have different MVDs to comply. This, evidently, can hardly perform the function of standards which limited set of solutions to actual or potential matching problems and intended/expected to be used repeatedly or continuously by a substantial number of the parties for whom they are meant creates vast different solutions for the same case, could be definitely considered as a standard that has certain potential to grow. The main reason leads to such cause could be categorized into 3 aspects, a. current MVD development relies on end use case scenarios, which is mainly qualitative actions, while it is not quantifiable, it is merely possible to unify the information requirements hence subjectivity rise; b. the nature of tunneling project is fragmented in terms of information management, hence the existing MVD development mechanism is limited while dealing with large complex datasets; c. though the building information modelling is based on interdisciplinary approach involving computer science, AEC and etc, the focus is still too AEC which has not encompassed other domains in a efficient way. To further think about these and review on information theory application, regardless of that from the perspective of empirical realism, discussing information modeling without involving information theory is seemingly absurd, the significant advantages of using information in research are providing quantitative framework for analyzing data, particularly quantifying the amount of information contained in a dataset, which is the first reason causing limitation of existing MVD development—lack of quantifiable approach. Also, the MVD for tunneling project is undeniably about managing large datasets, during review on information theory and its application, information theory proves its excellency on providing efficient methods and algorithms for dealing large amounts of data. Moreover, information theory enables complex system gaining insights into the complexity of systems which could help researchers to better understand how MVD development in tunneling project function and how they can be improved. On top of that, information theory has been proven its universal applicability to any system or process that involves the transmission or processing of information which is exactly BIM/IFC/MVD about—information sharing/processing. Last but not the least, information theory is known for encompassing mathematics, statistics, computer science, engineering, and other areas so that researchers from a wide range of backgrounds can use information theory to address diverse research questions. All these explicitly elaborate the promising potential of involving information theory in MVD development in the case of complicated engineering project.

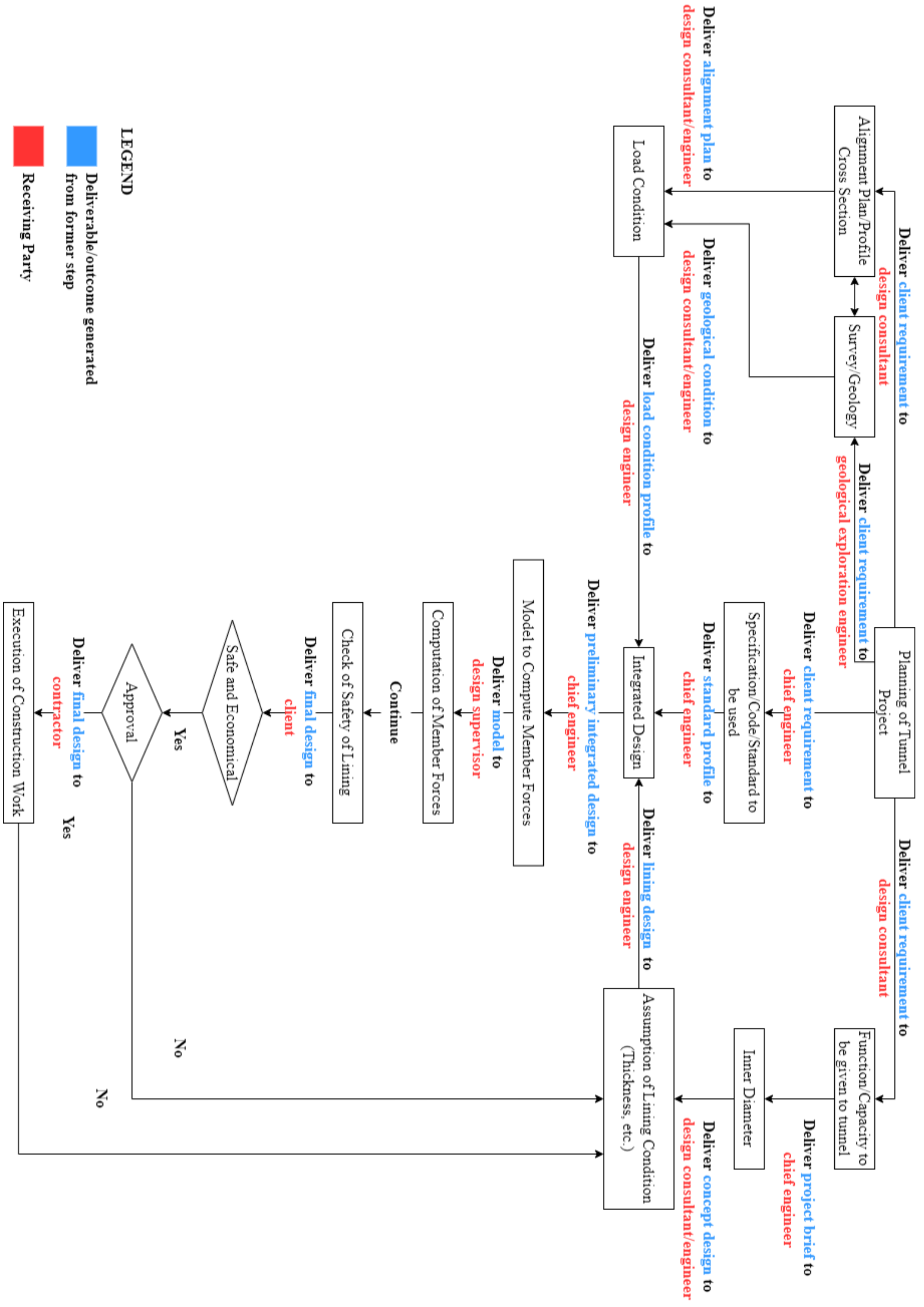
Yet, existing research in the field of IFC/BIM was primarily focused on improving data interoperability and use case scenario adoption without extensively incorporating information theory-based approaches. However, the limitation such as significantly diverse versions for applications and ambiguity generated by information requirements in IDM would be brought up, which would impede BIM development. The main reason behind this would be lack of a rigorous theoretical framework to understand the fundamental limits of IFC and to guide the design of efficient BIM systems. Referring to the nature of information entropy that would quantify information and provide a measure of its fundamental properties, it would be helpful to establish a theoretical foundation for the design and analysis of BM/IFC development systems. However, there is a research gap in understanding the potential benefits and contributions of information theory in enhancing data interoperability within the IFC framework. Information entropy provides a robust foundation for quantifying and analyzing information flow, data compression, and optimization, which can potentially offer valuable insights and solutions for enhancing data interoperability in IFC. By acknowledging the existing research that does not heavily involve information theory, there is an opportunity to explore the incorporation of information theory principles and techniques. This research gap calls for an investigation into how information theory and entropy can be effectively integrated into the IFC framework to improve data interoperability, transmission efficiency, semantic integration, and performance optimization. Through this research, the potential contributions of information theory/entropy in understanding of the role of information theory in enhancing data interoperability within the IFC framework, leading to the development of more robust, efficient, and standardized approaches for data exchange and integration in the architecture, engineering, and construction (AEC) industry.

The most obvious way of levering information entropy in the research is to quantify the information mass of built project/required to be digitalized object. Since the basic assumption of research is that performance of information standard should varies with the information mass as indicated in Section 3.2. The quantification process makes information entropy acting as a topological invariant of a specific IFC model as long as the information requirements are figured, this allows the reduction of ambiguity. Also, according to Section 3.2, the sandpile simulation would be involved in the research as a major tool, with proposed mapping system and justified metaphorized construct, the grain of sand is information mass of project. Thus, the information entropy with the unit "bit" is used to represent the grain of sand while conducting sandpile simulation. One bit of information is one grain of sand. It is extremely vital as an initial setup of simulation. Those sand waiting to be dropped in the simulation is the information mass contained in the project. In formation theory's concept of entropy provides insights into the complexity and uncertainty of data. MVD development can benefit from understanding the entropy of the information being modeled. It helps in identifying critical information, optimizing data flow, and prioritizing resources for data processing and storage (this has been proven in the follow up paper, An et al, 2023). In present work, a measuring scheme based on a "bit" system is proposed to measure the quantity of information for tunnel lining. The bits of which is substituted with the value of possible information entries of designed and need-to-be constructed linings. The information mass for the tunnel ling components is based on the information requirements. In other words, according to proposed information requirements in MVDs, what information is required to give for every entity of shield tunnel lining needs to be figured out. This would naturally involve information requirement analysis which is commonly regulated by certain association/council/administration. In order to obtain the information mass, one needs to determine the code of practice/guidelines aforehand. Take MVD in tunneling as an example, the tunnel lining information analysis of a simplified tunnel lining element is supposed to be given based on information requirement analysis referring "Guidelines for the Design of Shield Tunnel Lining" in Figure 6 (Working Group No.2, 2000) brought up by International Tunneling Association, who specifically indicate the steps as 1. Adherence to specification, code or standard, 2. Decision on inner dimension of tunnel, 3. Determination of load condition, 4. Determination of lining conditions, 5. Computation of member forces, 6. Safety check, 7. Review, 8. Approval of the design. View the information requirements from all its aspects including specifications/codes of practice/standards, ITA tunnel ling design flow chart, buildingSMART IFC-Tunnel Project, research of fellow scholars who focus on the shield tunnel lining design as well as the information modelling for it (Hao, 2016; Koch et al., 2017; Rives, 2020; Working Group No.2, 2000), the information requirement for shield tunnel lining segment is brought up followingly. There are mainly 5 categories of information, information of segments and their rebar construction, information of segment bolt, information of second lining, information of void element, information of backfill grouting. The detailed information contains in these categories is summarized in Table 1 The basic design information is regulated by specific national code of practice.

Table 1 Information requirements of shield tunnel lining

Segment	types of segments geometry information of segment ring types of assembling position of assembling sequential positioning block Information segment Block
Rebar construction	steel bar construction material grade mechanical parameters of material
Segment Bolt	bolt joint types shape of bolt material grade dimensions
Second lining	material type thickness
Backfill grouting	elastic modulus of grouting poisson ratio of grouting layer density of grouting layer depth of grouting layer grouting position
Void element	void position void shape/type void dimension

From this, it is evident that information theory provides a valuable framework for optimizing the development of model view definitions within building information modeling, which aligns with the positivist perspective By leveraging information entropy, researchers can enhance the representation and exchange of information, leading to improved objectivity and reliability of research findings. Especially, the concept of information entropy can be applied to quantify the level of uncertainty and redundancy in data models, enabling researchers to streamline the representation by eliminating redundant information and emphasizing critical data elements. This reduction in redundancy enhances the clarity and efficiency of information exchange among stakeholders. Additionally, the measurement techniques derived from information theory, such as information gain and mutual information, can facilitate the identification and prioritization of essential information components within the model view definition. By objectively measuring the relevance and influence of information elements, researchers can establish a more robust foundation for hypothesis testing and causal relationships analysis. This integration of information theory into model view definition development, grounded in the positivist perspective, not only strengthens the validity and generalizability of research findings but also enhances the overall effectiveness of building information modeling processes.

Planning of Tunnel Project
Deliver client requirement to design consultant
Alignment Plan/Profile Cross Section
Survey/Geology
Deliver client requirement to geological exploration engineer
Deliver client requirement to chief engineer
Specification/Code/Standard to be used
Deliver client requirement to design consultant
Function/Capacity to be given to tunnel
Deliver project brief to chief engineer
Inner Diameter
Deliver concept design to design consultant/engineer
Assumption of Lining Condition (Thickness, etc.)
Deliver alignment plan to design consultant/engineer
Deliver geological condition to design consultant/engineer
Load Condition
Deliver load condition profile to design engineer
Deliver standard profile to chief engineer
Deliver lining design to design engineer
Integrated Design
Deliver preliminary integrated design to chief engineer
Model to Compute Member Forces
Deliver model to design supervisor
Computation of Member Forces
Continue
Check of Safety of Lining
Deliver final design to client
Safe and Economical
Yes
No
Approval
Yes
No
Deliver final design to contractor
Execution of Construction Work
LEGEND
🟦 Deliverable/outcome generated from former step
🟥 Receiving Party

Following such reasoning in employing ID3 and sandpile simulation to develop MVD, the study trying to establish between the ID3 algorithm and the sandpile simulation, as well as the analogy between information entropy and information mass in the context of ID3 and MVD, respectively. In the ID3 algorithm, information entropy is used to measure the impurity or disorder of a dataset. The algorithm aims to split the dataset based on attributes that minimize the entropy and maximize the information gain at each node of the decision tree. This splitting process helps to organize and represent the information contained in the dataset more effectively. Similarly, in the context of MVD, an entity may contain a significant amount of information, which can be seen as an information mass. When an entity becomes too large or complex, it may be necessary to split it into smaller sub-entities to describe and manage the information more efficiently. This process of splitting an entity into sub-entities is analogous to the sandpile simulation, where the pile of sand (representing the entity) is divided into smaller piles (representing sub-entities) to prevent excessive information accumulation. By drawing this analogy (Figure 3 and 错误!未找到引用源。), the study is highlighting the common underlying principle of managing and organizing information by dividing it into smaller, more manageable units. Both the ID3 algorithm and MVD development process involve the concept of splitting or partitioning to effectively handle information entropy or mass.

The attempt to draw an analogy between grain of sand and information entropy is same as sandpile simulation and MVD development, which aims to illustrate the process of managing and organizing information within a complex entity. In MVD development, practitioners encounter entities that contain a significant amount of information. This information can become overwhelming and difficult to manage efficiently. By comparing it to sandpile simulation, the study highlights the concept of information mass. Just as a pile of sand can become unstable and collapse if it accumulates too much mass, an entity within MVD can become unwieldy and challenging to work with when it contains excessive information. To address this, the analogy suggests that we can split the entity into smaller sub-entities. These sub-entities represent more manageable units that allow for a more effective representation and organization of the information contained within the entity. This process of splitting an entity parallels the idea of dividing a sandpile into smaller piles to prevent information overload or collapse. By applying this analogy, the study aims to emphasize the importance of breaking down complex entities in MVD development, allowing for better information management and facilitating the design and implementation of the data model and associated views.

5.3 Sandpile Simulations to BIM data standard

5.3.1 Applied Sandpile Simulation

This section presents the applied sandpile simulation that could help to develop the building information modeling data standard with MVD as a proxy through tree-structured graph data model. The sandpile simulation is used because the toppling process in this simulation is analogous to the node splitting process in tree generation, based on assumption iv. When a vertex contains the sand grains over the capacity, it topples by passing the sand in the vertex to the neighboring vertices. It is exactly how a tree generation on the basis of assumption iv. The entity (node) contains too much information, it splits to multiple nodes in next level. Through Figure 7, the resemblance between these 2 subjects can be easily demonstrated. Through the demonstration, the similarities between the toppling process of sandpile and node splitting process of tree (MVD schematic structure) is perceived so that analogical inference could be made to draw argument from analogy. In other words, the perceived similarities are used as a basis to infer some further similarities those have yet to be observed. These formulate the basis of rationale justification of levering sandpile simulation to development of MVD for tunnel lining BIM model.

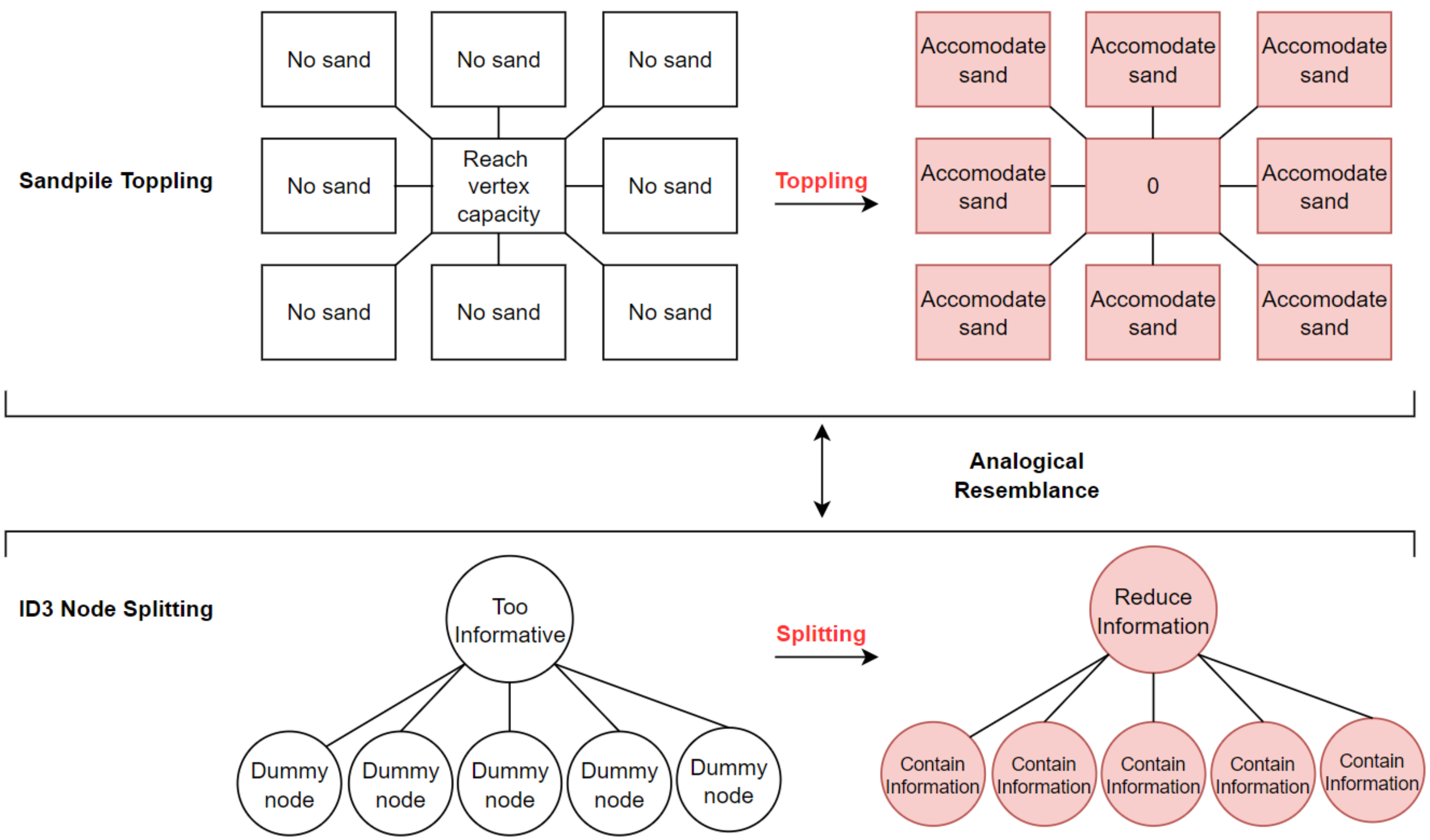


Figure 7 Analogy demonstration of Node Splitting and Sandpile Toppling

The sandpile model proposed by Bak, Tang and Wiesenfeld in 1987, revealed that the frequent occurrence of fractal structures is the generic spatial characteristic of a dynamical systems with many spatial degrees of freedom evolve naturally (Bak et al., 1987b; Cannon, 1984; Feder, 2013). They stressed the importance of this discovery with the claim "it develops complexity out of simplicity in contrast with the attempt to reduce complexity to simplicity". Mostly, a sandpile is used as a paradigm of an extended many-body system. Also, the stabilization of chip configurations obeys a form of least action principle: each vertex topples no more than necessary in the course of the stabilization. This can be formalized as follows. Call a sequence of topples legal if it only topples unstable vertices and stabilizing, if it results in a stable configuration. The standard way of stabilizing the sandpile is to find a maximal legal sequence, i.e., by toppling so long as it is possible. Such a sequence is obviously stabilizing, and the Abelian property of the sandpile is that all such sequences are equivalent up to permutation of the toppling order; that is, for any vertex v, the number of times v topples is the same in all legal stabilizing sequences. According to the least action principle, a minimal stabilizing sequence is also equivalent up to permutation of the toppling order to a legal (and still stabilizing) sequence. In particular, the configuration resulting from a minimal stabilizing sequence is the same as results from a maximal legal sequence. This property of sandpile model does not only reduce complexity to simplicity, but also reduce the ambiguity through the term "minimal—least action". It is a vivid analogy of an assumed case, which is, taking a square table and a large bucket of sand, then start sprinkling grains of sand on the table, one grain at a time, the start point is random and repeat the act till all motion terminated. It can be easily seen that there is no particular pattern. After a while, small local avalanches are triggered in order to decrease the local slopes whenever they become too steep, and eventually we end up with only one big sandpile. At some point (the transient time) this pile ceases to grow. The (global) average slope has reaches a steady state corresponding to the angle of repose which the sandpile cannot exceed no matter how much sand is added. The pile reaches a statistically stationary state and additional grains of sand will ultimately fall off the pile. The avalanches induce the transport of sand which is necessary to relax the sandpile. To examine the phenomenon, a cellular automation is introduced (Bak et al., 1987b).

To better understand the dynamical system of sandpile simulation so that the MVD development methodological framework could be more rigorous, the mathematical language is given. In that language, sandpile model is a cellular automation defined on a rectangular domain of the standard square lattice. Each vertex (i, j) of the domain carries a nonnegative number C_{ij} of particles ("grains of sand"), with c referring to the configuration of the sandpile. Starting from some initial configuration, particles are slowly dropped onto vertices chosen at random. If during this process the number of particles of any vertex exceeds three, this vertex becomes unstable and "topples", decreasing the number of its particles by four and increasing the number of particles of each of its direct neighbor by one. Thus, the toppling of vertices in the interior of the domain conserves the total number of particles in the sandpile, whereas the toppling of vertices at the sides and the corner of the domain decreases the total number by one and two. The redistribution of particles due to the toppling of a vertex can render other vertices unstable (Lang & Shkolnikov, 2019), resulting in subsequent toppling in a process referred to as an "avalanche." Due to the loss of particles at the boundaries of the domain, this process eventually terminates (Creutz, 1991), and the "relaxed" sandpile reaches a stable configuration which is independent of the order of toppling. The basic sandpile model can also be expressed as a Cyclic group of 4, considering its modulus 4. Hence, the order of the sandpile group $|G| = 4,$ and the proper expression of the group would be sandpile groups $G = \mathbb{Z}/4\mathbb{Z}, \mathrm{G} \subset H, H = \mathbb{R} / 4\mathbb{Z}$. To comprehensively elaborate the typology of the group of a vertex, the process of generating group G out of group Q ($Q = \mathbb{Z}$) is given along with the Cayley graph. First, considering the group $Q = \mathbb{Z}$, defining an arrow represent the generator <1>, its Cayley graph would be shown as Figure 8.

Figure 8 Infinite cyclic group $\mathbb{Z}$

With the Isomorphism theorems(Noether, 1927), $\exists G, H$, define $\varphi: G \to H$ as a homomorphism. Then: $\ker(\varphi) \lhd G$ (the kernel of the mapping φ) is a normal subgroup of; $\mathrm{Im}(\varphi) \cong \frac{G}{\ker(\varphi)}$, and if φ is surjective, $H \cong \frac{G}{\ker(\varphi)}$.

Based on this, it can easily conclude a group H', which is a group contains 4 and it is a normal subgroup, the quotient group Z/H' is corresponding to a mapping which maps 4 to 0. To avoid eliminating too many elements, the smallest normal subgroup that contains 4 would be used, in this case it would be <4>, in which 4 is the generator. The elements in <4> would be multipliers of 4, hence, it can be expressed as <4>={…, -12, -8, -4, 0, 4, 8, 12, …}. . Since ℤ is an Abelian group, its subgroups are all normal, <4> is a normal subgroup. Other than that, the cosets are needed Also, Z is defined as an additive group. Ergo, $k+< 4 >$ is used to express the co-sets. Figure 9 is given to visualize the <4> and co-sets as a Cayley graph. As expected, the 1+<4> contains all the numbers those have modulo 1 after divide 4, and 2+<4> represents all the numbers have modulo 2 after divide 4, and so forth.

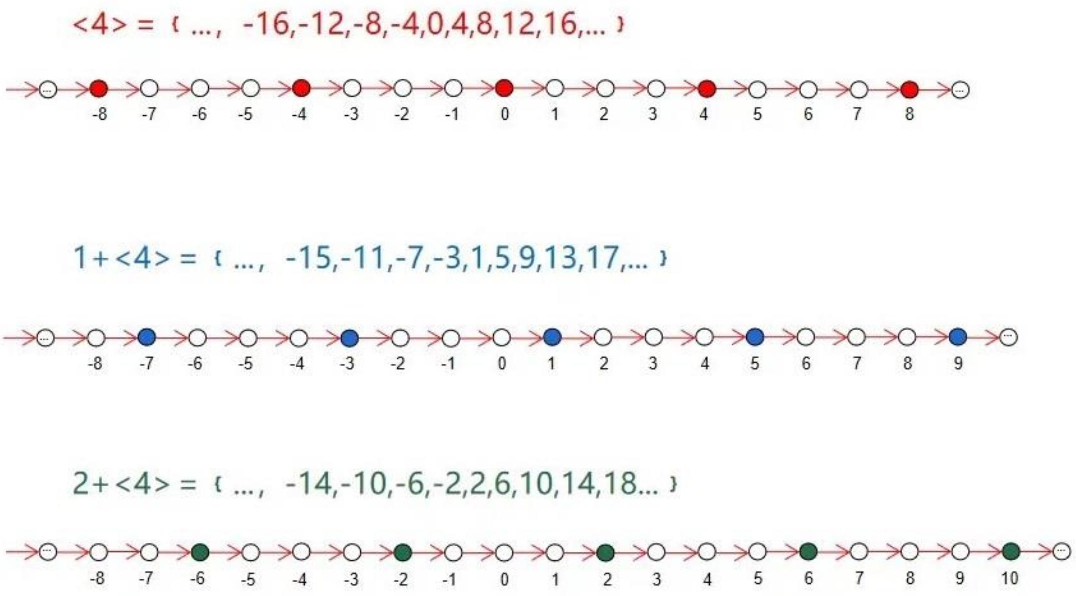


Figure 9 $k+< 4 >$ cosets，$k = 0,1,2$

All these co-sets are named congruence modulo 4. If 2 elements are in the same congruence class, they would be in an equivalence class. While they in an equivalence class, they are denoted as $a \equiv 4b$, for instance, $1 \equiv$45, $2 \equiv$ 46, $3 \equiv$ 47. Using these cosets to reorganize group ℤ, the structure would be an infinite spiral as shown in Figure 10. The gathering mode of cosets in the layout reveals that $\frac{\mathbb{Z}}{\langle 4\rangle} \cong C_4$. And this gathering mode is the topology of the used and extended sandpile groups on a domain which consists of a single vertex: space of nonnegative configurations, and the corresponding sandpile groups $G = \mathbb{Z} / 4\mathbb{Z}$. To understand Figure 10 better, a few extra notations are given as below, homeomorphism mapping $\phi: \mathbb{Z} \to c_4$ and isomorphism mapping $i: \frac{\mathbb{Z}}{\langle 4\rangle} \to c_4$; quotient mapping: $q: \mathbb{Z} \to \frac{\mathbb{Z}}{\langle 4\rangle}$.

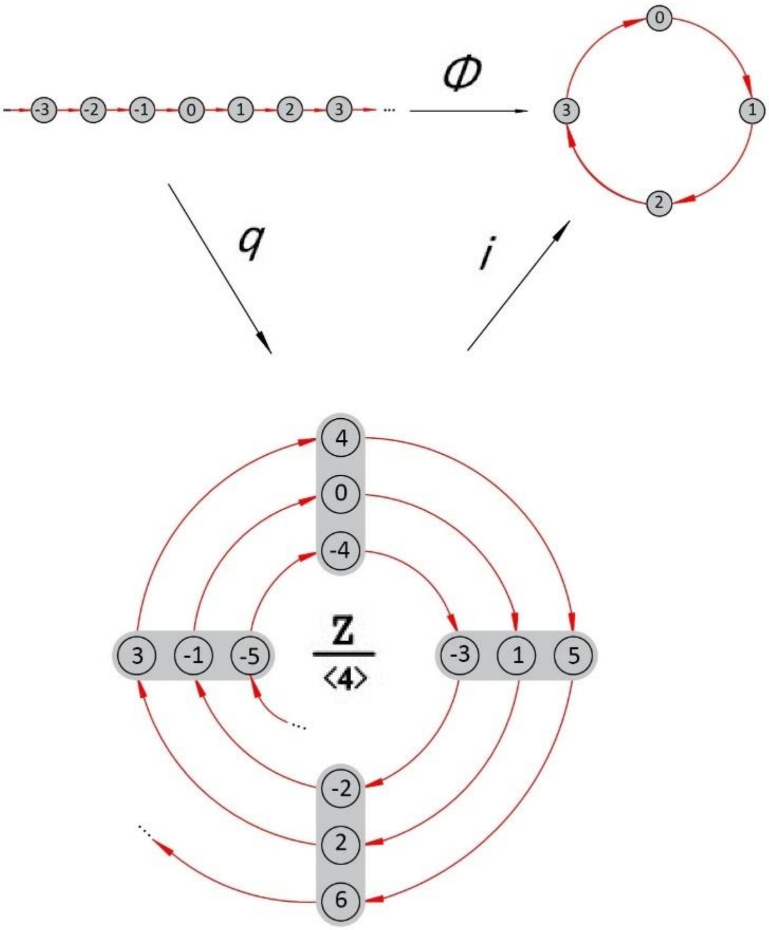
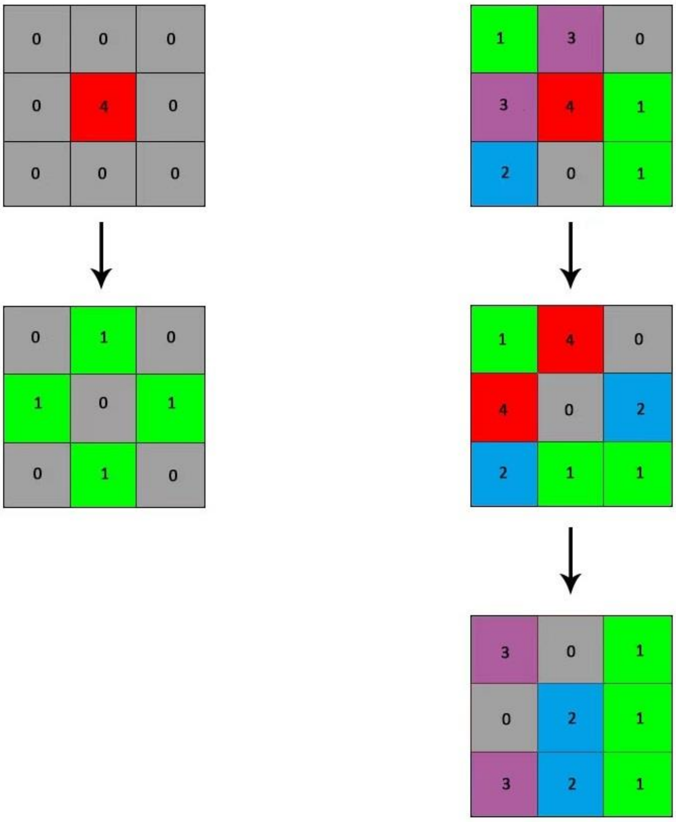


Figure 10 Integration of quotient mapping and renamed isomorphism equals to mapping ϕ. For any integer k, $\phi(k)$ is remainder of k modulo 4

The sandpile groups are a multiplicative group of integers modulo 4. To visualize the toppling process of sandpile model, Figure 11 below shows a process of a vertex toppling and a series of toppling involve 3×3 vertices.



0	0	0
0	4	0
0	0	0

0	1	0
1	0	1
0	1	0

1	3	0
3	4	1
2	0	1

1	4	0
4	0	2
2	1	1

3	0	1
0	2	1
3	2	1

Figure 11 Toppling of vertices-1. A vertex of the sandpile (3×3) carries 4 or more particles, it becomes unstable and topples, decreasing the number of its particles by 4 and increasing the number of particles carried by each of its neighbors by one. The toppling of one vertex can render other, previously stables unstable, resulting in an avalanche of subsequent toppling. Red is used to mark vertex with 4 grains, and grey is for 0 grain, green for 1 grain, blue for 2 grains and purple for 3 grains.

In this research, the sandpile simulation is used to simulate the topology development of MVD with the information quantity increasing. However, to make the simulation close to the practice, the initial configurations, avalanche threshold and the falling directions would be redesigned as following. To better fit the information standard development case, the settings of sandpile model is given as: Set a 2-dimensional lattice that contains $L^2 = L \times L$ square vertices.

Each vertex in the lattice gets a unique coordinate (i, j). For instance, in the case of $L = 100$ lattice, (i, j) ranges from 1 to 100. Then define $E(\mathrm{i}, j)$ as the number of sands in a vertex, and define e as the sands adding to the appointed vertex (i, j) at the time t, then there will be:

$$E(\mathrm{i}, j) = E(\mathrm{i}, j) + e \qquad \text{equation (1)}$$

Each vertex has a capacity of sand contained, denoted as E_{max}. At a given time point τ, the sand amount of vertex (i, j) $E(\mathrm{i}, j) \geq E_{\mathrm{max}}$, then sandpile model would activate the event with the following rules: $E(\mathrm{i}, j)$ would redistribute sands to its 8 neighbors, meanwhile $E(\mathrm{i}, j)$ is reset as 0. Its dynamical equations are listed as follow and the toppling demonstration of vertices is given in Figure 12,

$$E(\mathrm{i}, j) = 0 \qquad \text{equation (2)}$$
$$E(\mathrm{i}+1, j) = E(\mathrm{i}+1, j) + E(\mathrm{i}, j)/8 \qquad \text{equation (3)}$$
$$E(\mathrm{i}-1, j) = E(\mathrm{i}-1, j) + E(\mathrm{i}, j)/8 \qquad \text{equation (4)}$$
$$E(\mathrm{i}, j+1) = E(\mathrm{i}, j+1) + E(\mathrm{i}, j)/8 \qquad \text{equation (5)}$$
$$E(\mathrm{i}, j-1) = E(\mathrm{i}, j-1) + E(\mathrm{i}, j)/8 \qquad \text{equation (6)}$$
$$E(\mathrm{i}-1, j+1) = E(\mathrm{i}-1, j+1) + E(\mathrm{i}, j)/8 \qquad \text{equation (7)}$$
$$E(\mathrm{i}+1, j+1) = E(\mathrm{i}+1, j+1) + E(\mathrm{i}, j)/8 \qquad \text{equation (8)}$$
$$E(\mathrm{i}-1, j-1) = E(\mathrm{i}-1, j-1) + E(\mathrm{i}, j)/8 \qquad \text{equation (9)}$$
$$E(\mathrm{i}+1, j-1) = E(\mathrm{i}+1, j-1) + E(\mathrm{i}, j)/8 \qquad \text{equation (10)}$$

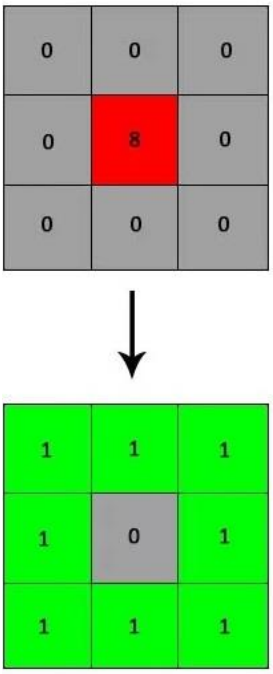


Figure 12 Toppling of vertices-2. A vertex of the sandpile (3×3) carries 8 or more particles, it becomes unstable and topples, decreasing the number of its particles by 8 and increasing the number of particles carried by each of its neighbors by one. The toppling of one vertex can render other, previously stables unstable, resulting in an avalanche of subsequent toppling. Red is used to mark vertex with 8 grains, and, green for 1 grain.

The redistributed sand would be added to a neighboring vertex, while the sand of the neighboring vertex is far less than the capacity of the toppling stop. While sand from the neighboring vertex is close to capacity, with the added sand, the toppling would be activated again, and the sand in would be redistributed to its neighbors. This process would last till sand in all the vertices is at less than capacity. Along with this process, the avalanche would be triggered to express the sands' redistribution. There are two types of avalanches, inner avalanche represents the avalanche occurring within the lattice, and outer avalanche happens while the sand is redistributed to the outer space of lattice. Hence, it is easy to learn that the order of the sandpile group $|G| = 8$; the proper expression of the group would be sandpile groups $G = \mathbb{Z} / 8\mathbb{Z} \subset H = \mathbb{R} / 8\mathbb{Z}$.

Applying Isomorphism theorems, assume group H' which is a group that contains 8 and is a normal subgroup, the quotient group $\frac{\mathbb{Z}}{H'}$ corresponds to a mapping which maps 8 to 0. To avoid eliminating too many elements, the smallest normal subgroup that contains 4 would be used, in this case it would be <8>, in which 8 is the generator. The elements in <8> are multipliers of 8, hence, it can be expressed as <8>={…, -16, -8, 0, 8, 16, …}. Since $\mathbb{Z}$ is an Abelian group, its subgroups are all normal, making <4> a normal subgroup. Other than that, the cosets are needed. Also, $\mathbb{Z}$ is defined as an additive group. Ergo, $k+< 8 >$ is used to express the co-sets. The figure s is given to visualize <4> and the co-sets. As expected, 1+<4> contains all the numbers that have modulo 1 after dividing 8, and 2+<8> represents all the numbers that have modulo 2 after dividing 8, and so forth. The gathering mode of cosets in the layout reveals that $\frac{\mathbb{Z}}{\langle 8 \rangle} \cong C_8$. And this gathering mode is the topology of the used and extended sandpile groups on a domain consisting of a single vertex: space of non-negative configurations, and the corresponding sandpile groups $G = \mathbb{Z} / 8\mathbb{Z}$. To understand Figure 13 better, a few extra notations are given as below, homeomorphism mapping $\phi\colon \mathbb{Z} \to c_8$; isomorphism mapping $i\colon \frac{\mathbb{Z}}{\langle 8 \rangle} \to c_8$; quotient mapping $q\colon \mathbb{Z} \to \frac{\mathbb{Z}}{\langle 8 \rangle}$. The topology of this sandpile model is given in Figure 12. The sand drops to the lattices continuously till a critical point where one more grain of sand would trigger a certain scale of avalanche including catastrophic avalanche that involves the entire lattice. For the most part, a large-scale avalanche would not be triggered by dropping a single grain of sand. Catastrophic avalanches happen seldomly. To measure the scalability of avalanche, the number of affected vertices in an iteration need to be counted. The relation between avalanche scalability and occurrence shows the statistical feature of power law distribution. Denoting S as avalanche scalability and $D(S)$ as the according occurrence, then the relation between scalability and occurrence would be expressed as:

$$D(S) \propto S^{a} \qquad \text{equation (11)}$$

In the log-log plot, the fitting function of power law distribution is linear, as:

$$\ln\big(D(S)\big) = a\ln(S) + b \qquad \text{equation (12)}$$

In which, a is power law index, and b is the intercept. This rule applies to the avalanche frequency and avalanche lifetime as well, denoted as:

$$D(L) \propto L^{c} \qquad \text{equation (13)}$$

In which, L is the avalanche scalability and D(L) is the according occurrence. In the log-log plot, the fitting function of power law distribution is linear, as:

$$\ln\big(D(L)\big) = c\ln(L) + d \qquad \text{equation (14)}$$

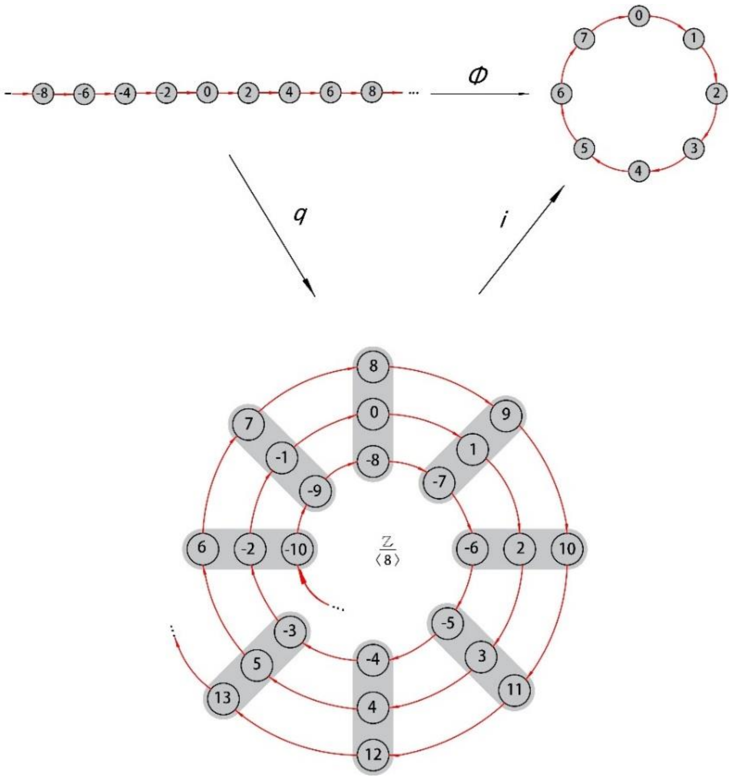


Figure 13 Integration of quotient mapping and renamed isomorphism equals to mapping ϕ. For any integer k, $\phi(k)$ is remainder of k modulo 8.

5.3.2 Mapping Sandpile Simulation to Topology of BIM Data Standard

In the research assumption of Section 3.2.1, it attempted levering sandpile simulation to BIM data standard (MVD development as proxy). Back then, it is given through drawing argument from analogy. The interlink between the sandpile simulation and development of MVD is detected. Yet, the interlink is not implicit. To lay out the interlink more precisely, the mapping that could map layout (phase) of sandpile simulation lattice to topology is defined. And to better understand the mapping, a simplified theoretical graphical demonstration is provided.

Firstly, abstracting and symbolizing the representation system of the sandpile model and the typology of BIM standard is necessary. For the sandpile models, every drop of grain of sand generates a new layout of the lattice, and it is regarded as one of the simulation results. Hence, each iteration (dropping grain of sand) would have a result. This layout of lattice can be defined as a phase of a system. The whole phase (layout of the lattice) can be composed by each component which in this case would be phase of each vertex. Each vertex has 3 independent features, location of it in the whole lattice, whether there are any grains of sand in it, and if there are, the grains of sand it accommodates. For a rigorous definition of location, a polar coordinate is imported to the system rather than keeping (i,j) Cartesian coordinate system. The origin point is the center vertex, and the radial coordinate is denoted by r , the angular coordinate is denoted by θ. Due to the generalized feature of the sandpile system, the coordinates of system is discontinuous, thus, r can only be a natural number $\{r|r \in \mathbb{N}\}$, and θ can only range from set $\left\{\theta_n \middle| \theta_{n-1} + \frac{360^0}{n}\right\}$, where, n stands for the number of toppling directions. The vertex not accommodating sand is regarded as a "False" value, denoted as "0", the vertex accommodating sand is considered a "True" value denoted as "1"; the corresponding component is denoted as m, $\{m|m = 0 \; or \; m = 1\}$. As for the grains of sand it accommodates, the component is denoted by g, if m=0, then $g = 0$; else, $g = n$, where n is the number of grains, ranging from 1 to vertex accommodation capacity that is predefined by the modeler. The phase of a vertex p can be denoted by $p(\theta, r, m, g)$. Figure 14 is given to visualize the expression of a phase for a random vertex in the lattice.

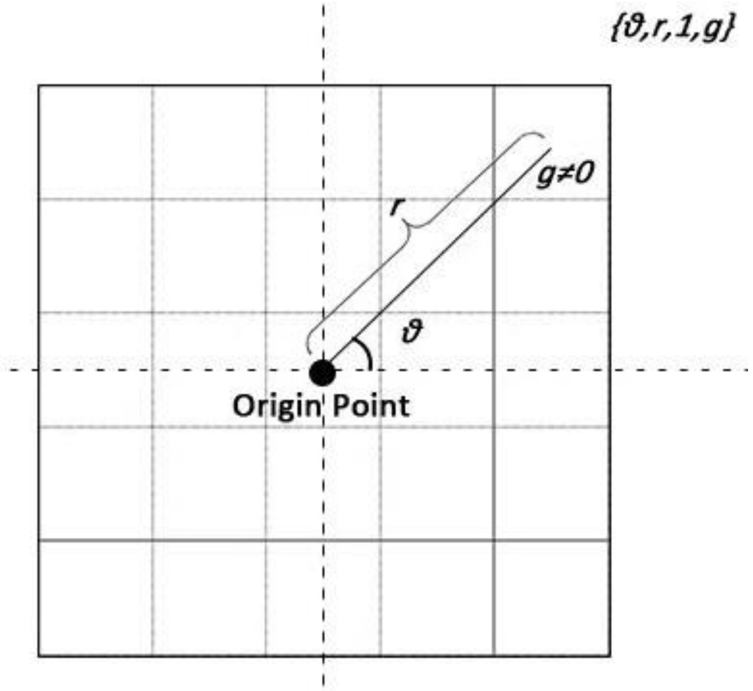
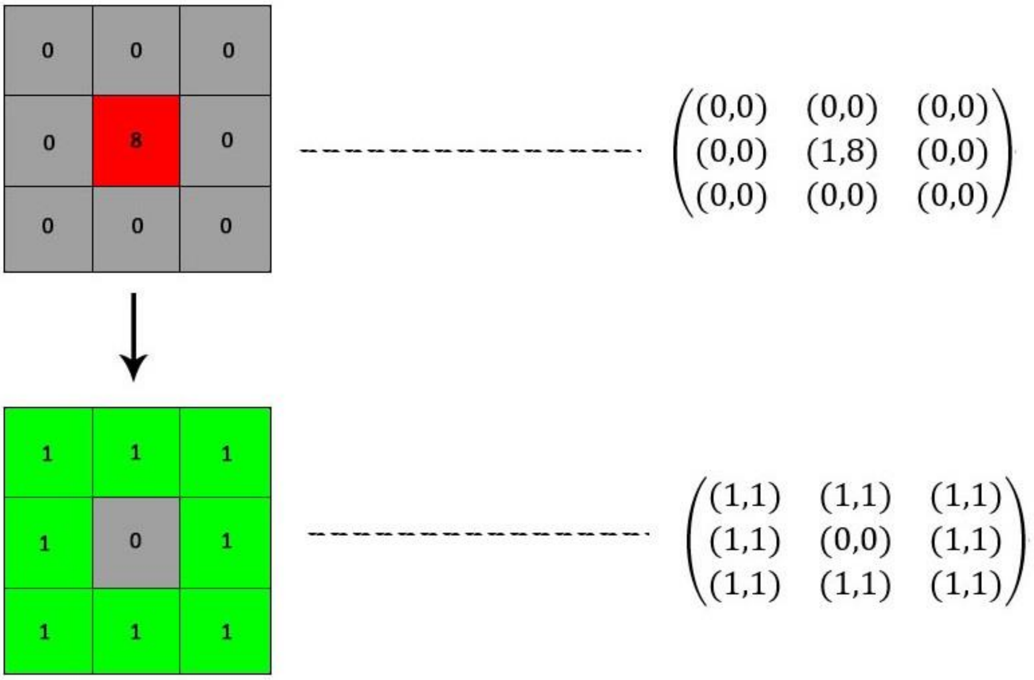


Figure 14 Expression of phase for a random vertex in lattice

With the defined notations, the resulting lattice layout of the sandpile simulation can be converted to a matrix with a 2-dimensional tensor. The column is θ , and the row is r, each element in the matrix is an according vertex and its phase is represented with a 2-dimensional tensor (g, m). An example of a 3×3 vertices as a result of a sandpile simulation following rules of applied sandpile model defined in 5.3.1 equation(2)-equation(10) is given in Figure 15.



$$\begin{pmatrix} (0,0) & (0,0) & (0,0) \\ (0,0) & (1,8) & (0,0) \\ (0,0) & (0,0) & (0,0) \end{pmatrix}$$

$$\begin{pmatrix} (1,1) & (1,1) & (1,1) \\ (1,1) & (0,0) & (1,1) \\ (1,1) & (1,1) & (1,1) \end{pmatrix}$$

Figure 15 Symbolizing sandpile simulation result to 2-dimensional tensor.

Theoretically, the inheritance hierarchy of a BIM standard is tree structure while the semantic aspects are excluded from the specific cases. Considering the application of BIM standards in practice, this tree structure is a weighted directed graph, that can be referred as tree hierarchical combination according to semantic meaning and the ontology, and each node corresponds to different engineering entities. The weight is the information mass that the node accommodates, and the direction represents its level in the tree. To design a topology of MVD, a few details ought to be clarified. The structure of the tree refers to the number of nodes, their associated levels, and the overall number of levels in the tree. Once these elements are predetermined, the weight (information mass) of each node must also be specified. The number of nodes in level h is denoted by n_h, and the weight (information mass) of node is denoted by w. After symbolizing the representation systems of sandpile simulation and topology of BIM standards, the mapping between these two is regulated as:

i. the origin point of sandpile lattice is the root node of tree.

ii. level parameter h in tree corresponds to r of $p(\theta, r, m, g)$.

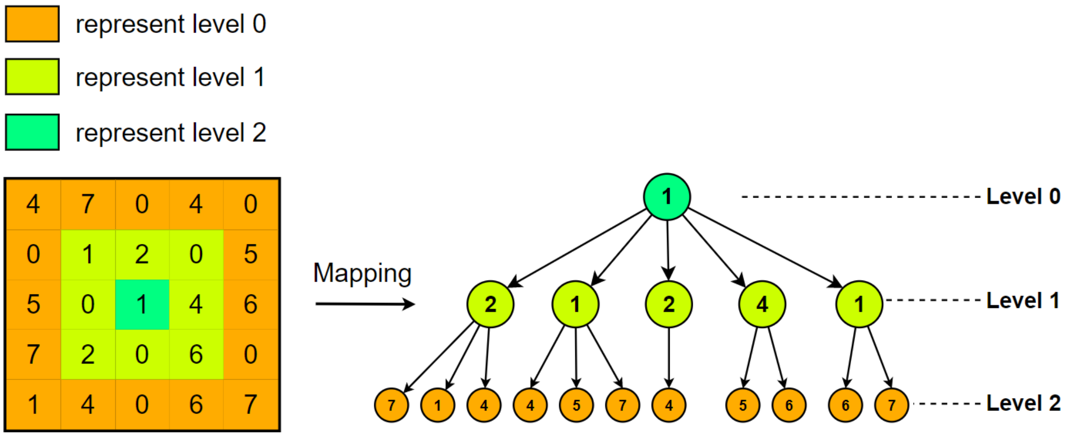
iii. information mass (weight) w associates with g of $p(\theta, r, m, g)$.

iv. n_h equals to the summation of m while r (h) remains the same.

v. The links among different nodes are created based on the semantic aspects of inclusion relation as well as assumption iv which implies the branches prioritizes linking higher entropy value possessed nodes.

Through proposed mapping system, the tree-structured graph data model comprehensively describes the essential characteristics, topological relations of entities in tunnel linings, which can be leveraged to generate and optimize the possible MVDs. With the simulation results (layout of sandpile lattice) and proposed mapping, the basic data to support MVD development could be obtained since the layout of sandpile lattice contains the key information of the tree-structure graph which is a topology of MVD. And the proposed mapping offers a way to extract the information from the simulation results. On the basis of these, number of nodes which represents the entities needed for the MVD, number of levels that tree contains which represents for the height of tree-structured graph, number of nodes in every level which represents for the complexity of associated entity, weight of node which represents for the information quantity contained in the entities could be all obtained from the simulation results. Moreover, there are some other advanced data could be obtained from that, such as node fission entropy, complexity coefficient, proportion of fission nodes. Following such ideology, a new definition of node fission entropy is given in the Section 3.5 and employed in following research.

The root node of the tree is regarded as level 0. A sandpile lattice is given in Figure 16, the initial layout of sandpile lattice is generated with randomizer, and the toppling rule is aligned with the sandpile group $G = \mathbb{Z} / 8\mathbb{Z} \subset H = \mathbb{R} / 8\mathbb{Z}$ and the gathering mode of cosets in the layout reveals that $\frac{\mathbb{Z}}{\langle 8 \rangle} \cong C_8$. The demonstration about toppling process of a vertex is given in Figure 15. The topology of associated MVD is given as Figure 16. It can be seen, there are 3 levels of tree, level 0, level 1 and level 2. Therefore, the height of tree would be 3. And the level 0 only contains root node, it is also the root node of tree. In level 1, there are 8 nodes including 3 nodes with 0 weight. Hence, it could only be considered as 5 nodes in the level 1 as shown in Figure 16. Similarly, it can be easily to conclude that there would be 11 nodes in level 2 (5 nodes assigned weight 0 among 16 nodes). At all three levels, there are 17 nodes, including the root node, as shown in Figure 16. And the weight assigned to nodes is the information quantity contains in the node, which is the entity in MVD. Accordingly, the information quantity associated with nodes is indicated. All these basic data related tree-structured graph is easily extract from simulation results. And it provides the fundamental features of MVD. Noticeably, in this way, the universality is revealed. As long as the information quantity of a project is predetermined, initial layout and the toppling rule remain consistent, the MVD could be universally consistent cross the users. This forms the theoretical basis of unifying MVD development then making a standard "standard".



4	7	0	4	0
0	1	2	0	5
5	0	1	4	6
7	2	0	6	0
1	4	0	6	7

Figure 16 Mapping of a layout of sandpile simulation result to tree-structured graph

Beside the function of linking to BIM data standard development, the sandpile simulation has other vital implication in the dissertation. Over the past decades, the study of complex systems has evolved from a niche area to the epicenter of interdisciplinary research. These systems, be they biological, social, or physical, exhibit a rich, multi-scale behavior that often transcends the purview of traditional linear theories. To capture and characterize the behavior of these systems, a plethora of theories and models have been proposed. Among them, Self-Organized Criticality (SOC) has emerged as a pivotal concept in this domain. The central tenet of SOC is that certain systems, devoid of overt external driving or tuning, can spontaneously reach a dynamic equilibrium. This state is neither absolutely stable nor entirely chaotic but resides at a "critical" juncture between the two. A salient feature of this critical state is that events within the system, regardless of their magnitude, adhere to a specific statistical distribution, typically a power-law distribution. This suggests that large-scale events, such as massive avalanches or market crashes, are not outliers but intrinsic components of the system.

Power-law distributions have been observed across a myriad of natural and societal phenomena, from the magnitude distribution of earthquakes to the population distribution of cities, and even the distribution of links on the internet. This ubiquity hints at a universal mechanism propelling these disparate systems towards this critical state. A key postulate of the SOC theory is that the internal interactions and dynamics of a system are sufficient to naturally usher it into this critical state, obviating the need for external modulation or driving. To elucidate and describe SOC more comprehensively, various models have been advanced. Among them, the sandpile model proposed by Bak, Tang, and Wiesenfeld stands as a quintessential and widely recognized example. At first glance, the model might appear deceptively simplistic, but beneath its elementary facade lies a profound insight into the inherent dynamics of many natural systems. The essence of the model revolves around the idea of incremental accumulation and systemic response. Sand grains are progressively added to a pile. As one might intuitively expect, there exists a threshold — be it in terms of a specific angle of repose or a particular number of grains — beyond which the pile cannot maintain its stability. Once this threshold is breached, a grain or a cluster of grains begins to topple, initiating a cascade of subsequent toppling. This cascading effect, reminiscent of a chain reaction, is termed an "avalanche" within the context of the model.

What makes this model particularly intriguing is its ability to reach a critical state without any external intervention. Instead of being driven to this state by external factors, the system, through its intrinsic dynamics, naturally evolves to a point where it teeters on the brink of stability. At this juncture, any minor perturbation, such as the addition of a single grain, can potentially trigger avalanches of varying magnitudes. The distribution of these avalanche sizes, interestingly, follows a power-law, a hallmark of SOC. Furthermore, the sandpile model underscores an important facet of complex systems: the interplay between local events and global responses. A local disturbance, such as the toppling of a single grain, can propagate through the system, leading to a large-scale rearrangement. This local-to-global dynamic is not only central to the sandpile model but is also observed in various other systems, from neural networks to financial markets.

Deepak Dhar's contributions to the field of Self-Organized Criticality (SOC) have been instrumental in elucidating the intricate mathematical intricacies inherent in the sandpile model. While the model, as proposed by Bak, Tang, and Wiesenfeld, provided a conceptual framework for understanding the dynamics of complex systems, it was Dhar's rigorous mathematical analysis that truly illuminated the underlying principles governing these dynamics. Dhar's exploration into the model's mathematical structure unveiled profound connections with other areas of physics and mathematics. For instance, his work drew parallels between the sandpile model and certain lattice models in statistical mechanics, highlighting the universality of the principles underlying SOC. Additionally, he identified deep-rooted links between the model and combinatorial mathematics, particularly in the context of configurations and their corresponding state spaces. One of the most striking revelations from Dhar's research was the identification of the model's conformal invariance at its critical state. This property, often associated with critical phenomena in two-dimensional systems, provided a robust mathematical toolset for analyzing the model's behavior near its critical point (Dhar, 2006).

In essence, the Bak-Tang-Wiesenfeld sandpile model serves as a potent instrument for comprehending the phenomenon of Self-Organized Criticality in complex systems. It underscores the notion that profound complexity and universality can be inherent even in systems that outwardly appear simplistic. he impetus behind the integration of this model into the present investigation stems from its unmatched prowess in demystifying the multifarious facets of SOC. This phenomenon, over the years, has been at the forefront of academic discourse, primarily due to its profound implications in deciphering the dynamics of intricate systems. Recent scholarly endeavors, especially those spearheaded by Selvam (Selvam, 2001), have cast a spotlight on the symbiotic relationship between SOC and the realm of chaotic behaviors. Delving into these works reveals a compelling narrative: systems that are emblematic of SOC, despite exuding an aura of orderliness and predictability, possess an underbelly teeming with chaotic tendencies. This juxtaposition presents a fascinating duality (Selvam, 2001). For instance, when we consider system of information mass against BIM data standard development, its predominant characterization by SOC belies an underlying bedrock of chaos, manifesting in its behavior and responses (Selvam, 2004). Such a duality is not merely an academic curiosity but holds profound ramifications for our understanding of complex systems. It suggests that beneath the veneer of stability and order, there lies a dynamic interplay of forces, oscillating between the realms of predictability and unpredictability. This intricate dance between SOC and chaos forms the crux of BIM data standard development system behavior, making it a subject of immense research interest.

Given this backdrop, the sandpile model becomes an invaluable asset. Its mathematical and conceptual framework offers a lens through which the chaotic substratum of System A can be meticulously examined. By harnessing the power of this model, this dissertation embarks on a journey to unearth the chaotic elements that underpin information mass-BIM data standard development, all the while emphasizing the confluence of SOC and chaos as pivotal forces that mold its behavior.

5.4 Data Transformation for MVD Transformation –Elaboration on Sandpile Analogy

This subsection introduces the data transformation that could convert graph data to MVD in tunnel lining IFC representation. Figure 17 demonstrates the process of data transformation. Since the transformation process targets at transforming information from BIM to the tree-structured graph data model. The whole operation is based on graph data model. It is essential to derive the data model through sandpile simulation. Thus, the whole process starts with initial settings of sandpile simulations and such those are all introduced in section 5.3. Following this, the operation to covert the required information in BIM to the graph representation is conducted. To accomplish that, the information entropy is calculated as the indicator for information quantity, and the semantic information is extracted. Subsequently, topological relationships could be given with the information requirement analysis. Provided the information quantity, semantic information and topological relationships, the new entities those required to be proposed are constructed into graph data model in compliance with the information quantity. In this step, it is mainly matching the entities properties with nodes' attributes in the tree-structured graph. For example, the number of proposed entities is supposed to align with the number of nodes in the graph. And the way to position entity in nodes layout would consider the consistency of entity's information mass and weight of node. The branches among the nodes are regulated by the topological relationships as well as semantics of associated elements.

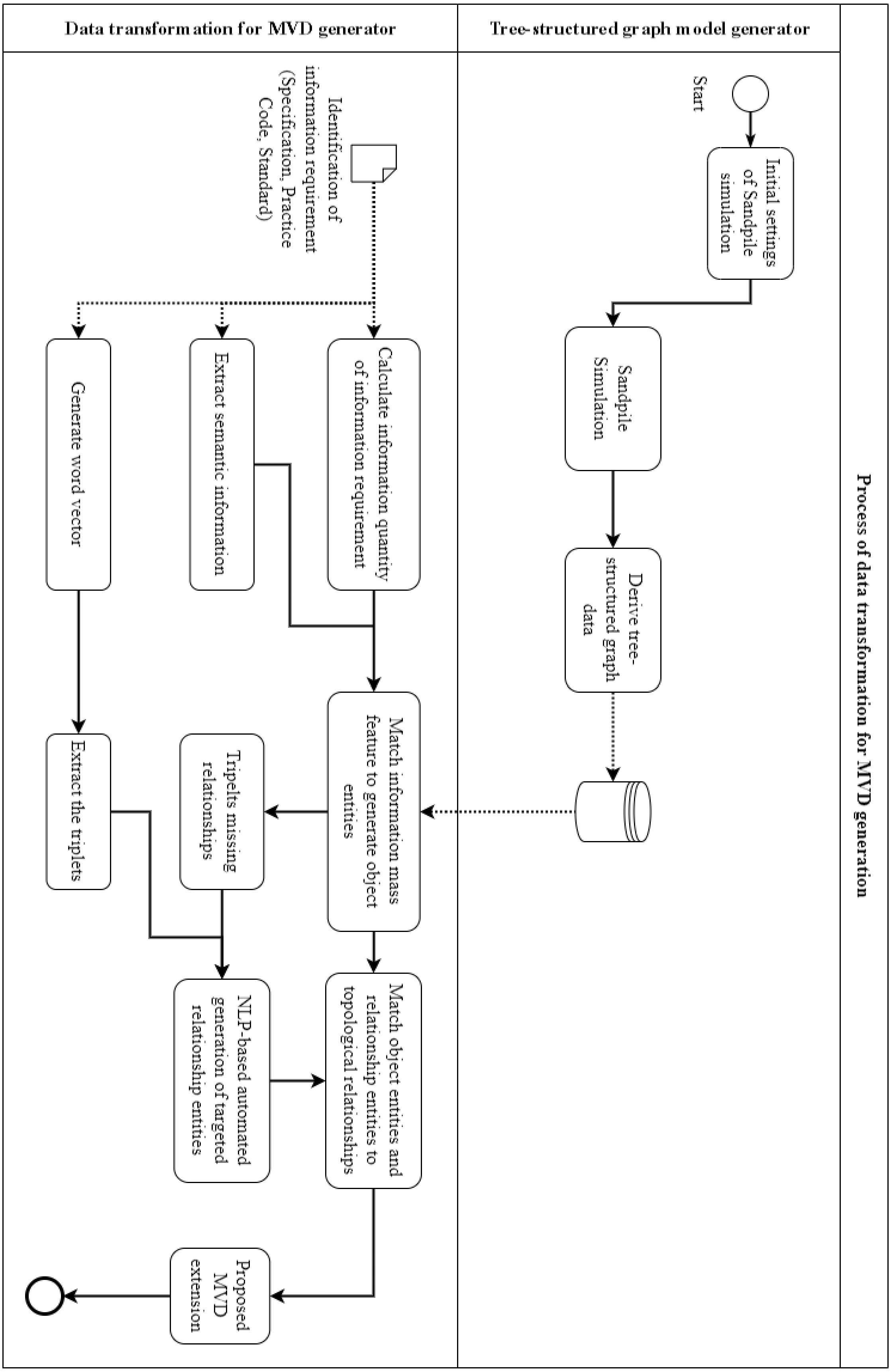


Figure 17 Process of data transformation for MVD transformation

Then the Natural Language Processing (NLP) based technique could be used to create the relationships between entities generated by sandpile simulations. The NLP based system/framework generating ontologies in BIM has been widely discussed in the high level of maturity last few years (Chen & Luo, 2019; Li et al, & Liu, 2022; H. Wang, Meng, & Zhu, 2022; Wu et al, 2019; Zhou et al, 2022). Naturally, such technique is used in the study to complete the MVD by creating suitable links between generated entities. The proposed integrated entity and relationship completion model is given in Figure 18. Sandpile simulation would generate the informed information mass for potential entities. Followingly, preprocess required specific code of practice and specifications to tokenize and segment text so that the entity pairs would be generated, such as $< h_{ei}, ?, t_{ej} >$. h_e captures the entity's behavior as the head of a relation and t_e captures the entity's behavior as the tail of a relation. These two compose a pair which is a partial triplet that missing a relation, the missing relation in the pair is represented by "?". Meanwhile, word embedding/vector technique is applied to chosen specifications/code of practice to preprocessed text for the purpose of identifying as well as classifying entities as "named entity recognition. With outputs from named entity recognition and dependency parsing, the NLP pipeline identifies entity pairs related to MVD concepts within the same sentence or a specific distance from each other. Based on that, the NLP pipeline extracts the relevant triplets based on the MVD concepts and the code of practice/specification. For example, a triplet could be formed, so does the specific relationship between them such as decomposes, assigns, etc. The triplet is represented with $< h_{ei}, v_r, t_{ej} >$, in which h_{ei} captures head entity, t_{ej} captures tail entity and v_r captures relations. All generated triplets would be regarded as template to compare with the entity pairs created from earlier steps (sandpile simulation-based generation) through the NLP pipeline. The NLP pipeline extracts the relationships between the matched entity pairs based on the generated triplets. This involves analyzing the context, linguistic patterns, and dependencies to identify the specific relationship. For instance, the relationship between "SegmentTunnelElement" and "Segment Lining" could be "decomposes". Accordingly, the sandpile simulation generated entity pair would be captured by:

$< SegmentTunnelElement, ?, SegmentLining >$,

and associated triplet in triplet generator would be represented as:

$< SegmentTunnelElement, decomposes, SegmentLining >$.

The extracted relationship in the case would be "decomposes" through the NLP pipeline. By means of these, the proposed entities are selected and positioned associating with the graph model to provide a comprehensive expression of proposed MVD extension.

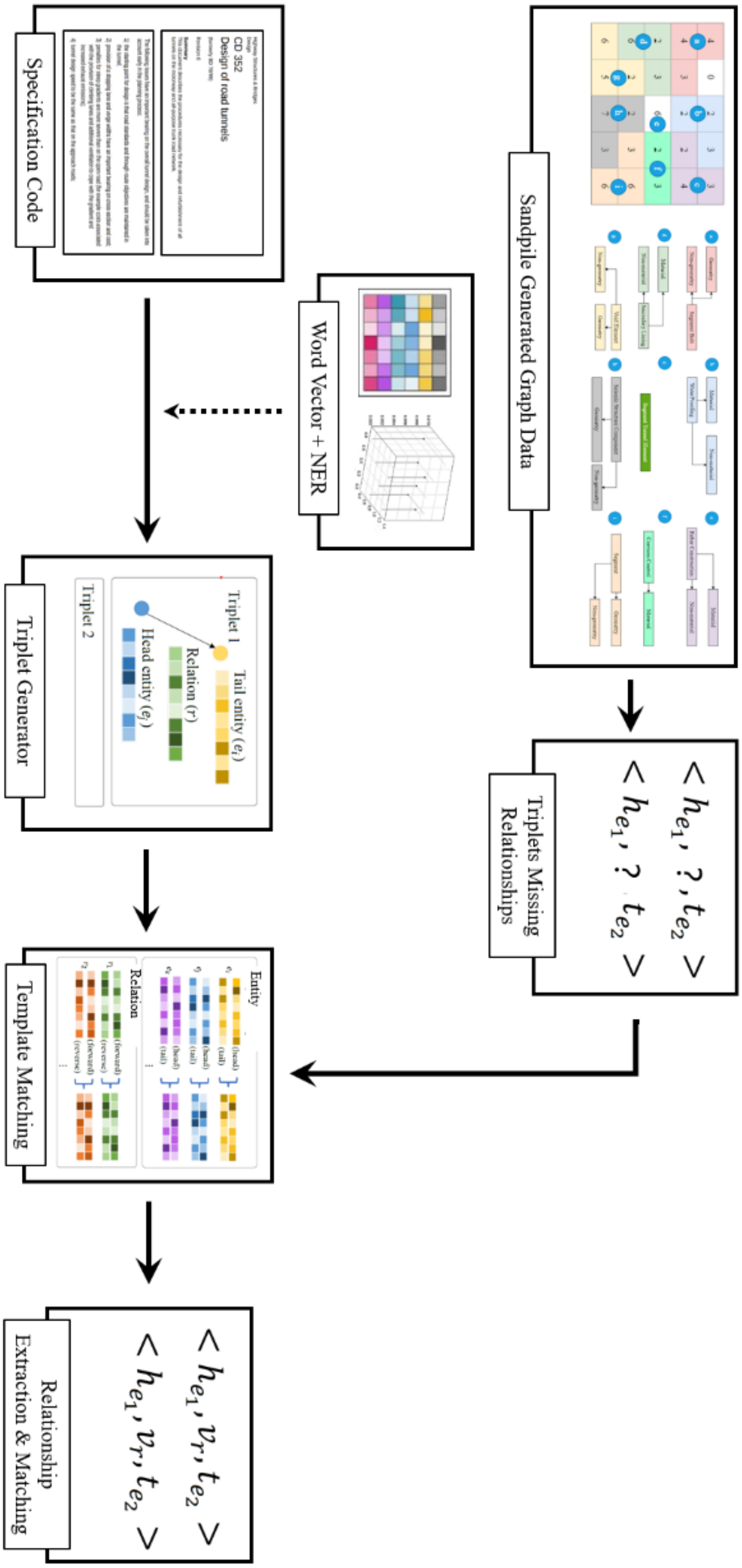


Figure 18 The architecture of proposed integrated entity and relationship completion model

Consider Xiaojun's MVD, illustrated in Figure 32, as a simplified example. Assume a project entails constructing a Q km shield tunnel, interpreted using tree-structured graph data as outlined in the proposed MVD.

- Shield tunnel element is a root node.
- 2nd level of tree has 7 nodes which means we would need 7 entities as extended sub-entities to better describe shield tunnel element for a project with certain level complexity. And in the project, shield tunnel connected aisle element has A_1 bits of information, shield tunnel void element has A_2 bits, shield tunnel segment bolt has A_3 bits, shield tunnel segment has A_4 bits, shield tunnel secondary lining has A_5 bits, shield backfill grouting has A_6 bits, tunnel element proxy has A_7 bits. Noticeably, the A_1 to A_7 is determined by the Qkm shield tunnel project information requirement analysis, calculated with the formula provided in section 5.2.
- 3rd level of tree has 10 nodes which means we would need 10 entities in this level to assist describe upper sub entities. And the information mass is B_1 to B_{10} accordingly. B_1 to B_{10} is also predetermined by the Qkm shield tunnel project information requirement analysis, calculated with the formula provided in section 5.2.
- Matching the topological relationships among entities and the pairing information mass to the weight on the nodes.

Given above interpretations, the tree-structured graph that aligns with the

Xiaojun's MVD is given in Figure 19.

There would be another several things to notice while conducting the data transformation:

a. The example given here is a simplified example, it tries to sketch the graph data model that the Xiaojun's MVD transform from, the original process should have been reversed (get the graph data model first and then match). The process is reversable.
b. Under the framework of IFC, the IfcProxy is defined to serve as a placeholder for representation objects that do not correspond to any of the semantic types so that they can still be defined in the IFC model. This offers flexibility to the data transformation, if there are further entities that may exceed the lattice capacity, it could be defined using IfcProxy so that could help to further develop MVD. Such as, in a 2-dimensional lattice, the second layer of the lattice only contains 8 vertices, which means that the second level of tree could only accommodate 8 nodes. In that case, the MVD developers would have to consider reorganizing the structure based on semantics that related to practical project information requirements so that could make the more detailed breakdown entities to the 3rd level on which could accommodate 16 nodes meanwhile consolidating the entities on the 2nd level. Else, reconsolidating, conforming entities then extending some of them into IfcProxy would match the capacity of the lattice.
c. The calculated information mass would possibly not match weights exactly. In that case, the weight of node in the graph data model would be recalculated as the proportion which stands for the relative value rather

than the absolute value of information mass. Following this, the flexibility for matching is increasing while the universality remains. Another alternative for such matter would be similar to the item b, split the entity again to make it fits the weight of positioned nodes. For example, the information mass of entity 1 has A_1 bits of information based on the project information requirements analysis, and the node 1 from the graph data model has the weight G_1. While $A_1 = G_1$, entity 1 match the node 1. While $A_1 > G_1$, then keep the information mass G_1, and then put the extracted information in another entity and then match it with another node in graph data model. Interestingly, the grain of sand is defined as the information mass of the project. Hence, the closure property fits the scenario, which means, there supposed to be consistent between the sum of nodes' weight and sum of information mass. After reorganizing the information mass distribution, the new entity would be defined based on the semantics. In the graph, the nodes are linked with their nearest.

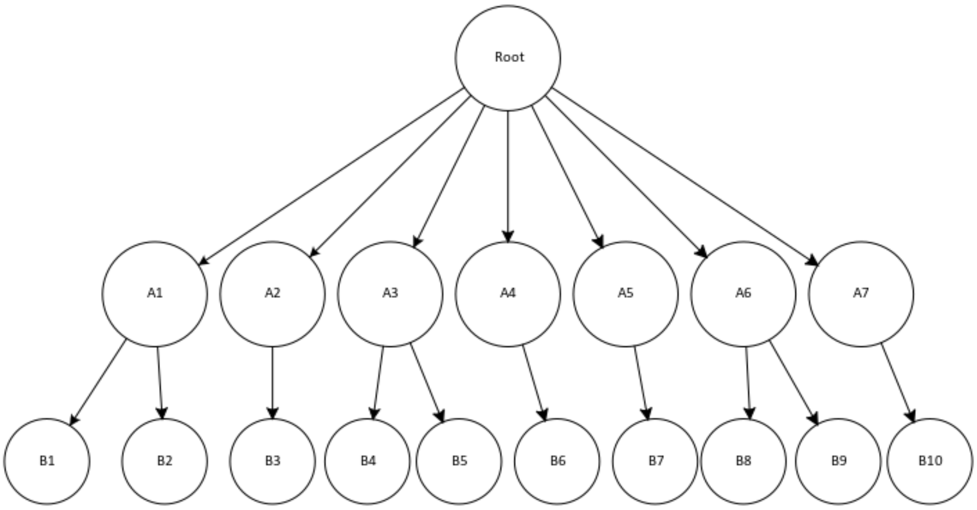


Figure 19 Tree structured graph that aligns with Xiaojun's MVD

5.5 Simulation settings and Results

Initial Setting

IFC, as a well-developed and mainly considered as a building information modelling standard, is designed to further extend through 3 mechanisms which are IfcProxy, IfcPropertySet and IFC entity (An, 2017; Wang & Li, 2014). With such mechanisms, considering the feature of two dimensional sandpile models, the numerical sandpile model interpretation of MVD to building Information model standard development is given as follows:

1) The vertices composing the two-dimensional lattice represent the entities. The initial configuration of all the vertices is set to 0, which stands forneeds to sort information by giving the data schema, and while the information is little, it requires instances directly to explain information rather than with metadata.

2) The central vertex is the dropping point; this point stands for the IFC Root. The grains of sand dropping into the vertices are the analogical figure of information mass contained in the building information model, and the grains of sand in the vertex express the information contained in entity.

3) The avalanche triggered by a grain of sand are abstractly associate with restructuring the standard, such as extending IfcPropertySet to enrich the entity and recategorizing the information stored in according entity. Larger the scale of avalanche triggered; higher proportion of standard would require to update. Or, longer the avalanche last, more urgent needs for updating the standard would show.

4) Keep dropping sands continuously after certain interval, the average of ratio

for sands and vertices would tend to remain the same level. In such cases, it would be defined as critical state (also called self-organized criticality), its corresponding amount of information is an exact value. And this information mass could trigger the new further development of the standard. Otherwise, the expressivity of information would be compromised.

5) According to self-organized criticality and the modelling process, one can reckon that the different standards or different iteration of standards are the results of the optimal expression of the BIM model. While information mass exceeds the according standard (common data environment) capacity, the performance of the model will be negatively affected.

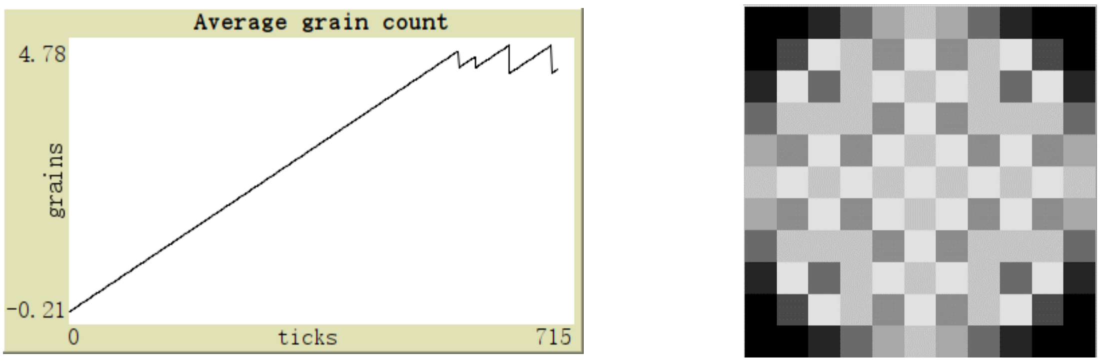
Based on the self-organized criticality and the sandpile model, a few regulations are given while the avalanche involves a quarter of the whole lattice, for instance the lattice is $L^2 = 51 \times 51$, 650 vertices start toppling, the accumulation of information noise decreases and the effectivity of information increases. While the toppling vertices is over one third, for instance the lattice is $L^2 = 51 \times 51$, with 867 vertices toppling, the limits of standard for expressing the information mass prevails. Two considerations shape these regulations, a. grains of sand dropping are limited, as aforementioned, certain amounts of information mass can ensure the data standard development could be efficiency improvement, also the information mass cannot be unlimited since the scale of the project is limited in practice; b. if a single grain of sand triggers one third of vertices toppling, the scalability is great enough to be catastrophic. To summarize, along with the general research hypotheses, from the perspective of the sandpile model and the BIM model standard development, the way of interpreting the BIM model standard development is rational. The sandpile model application abstracts the relationship of the BIM model information mass and standard development, attempting to explore the endogenous mechanisms of standard extension from a different perspective.

The proposed mechanism is demonstrated through an illustrative example that focuses on shield tunnel lining. The sandpile system, as a dissipative system implicitly exchanges between mass and energy (toppling rules), which can be analogically regarded as a simplified system showing the process of restructuring (recategorizing entities) and extending MVD. To make the demonstration more comprehensive, ensure a rigorous statistical sense of simulation and realistically reflect the trend of MVD development processes, "bit" is used to represent the information mass, and the two-dimensional lattice are set as $L^2 = 11 \times 11$, with the initial starting layout as no vertex contains sand. Followingly, the ticks (sands dropping) are total grains of sands. Since the case is a simplified version based on a single layer shield tunnel as underpass , the whole line is 0.4 kilometers. All the tunnels are underground, considering the width of a ring is roughly 1.2 meters, the first underpass would have 334 rings and each ring contains 16 bits of information. The total information mass is approximately 5344 bits. While dropping grains of sand, one at a time, sand grains can fall off the edge of the table, helping ensure the avalanche eventually ends. Considerably, there are limited vertices, 11×11. Even all the vertices reach their full capacity, it would be only 800 grains. The greater number of grains of sand falling off the edge of the table would lack value to the study. Thusly, the grains need to be limited. Constructing isomorphism can be of great assistance. With isomorphism, the grains of sand can easily be converted to 334 (5344/16) grains of sand. Technically, the sandpile model is implemented in the Netlogo software developed by Uri Wilensky (1999). It is a free and open-source software platform with a simplified and flexible programming language (Castilla-Rho, Mariethoz, Rojas, Andersen, & Kelly, 2015). Hence, it is chosen to be used for the current study. And the sandpile model developed by Weintrop et al (2011) is imported to Netlogo for further developments to suit the current study. And the information requirements analysis complies with "China National Design Code of Shield Tunnel Engineering".

Given the initial configurations of the sandpile simulation described earlier, each simulation involves generating a tree-structured graph data model. Additionally, several observations warrant further attention, as will be reported here (with code provided in Appendix A). These include the average height of vertices, calculated using the average grain count to indicate the critical state; plots of avalanche sizes and lifetimes on a log-log scale; occurrences of catastrophic avalanches where one-third of the vertices topple; and the first extreme catastrophic avalanche, occurring after the system enters the critical state, where over 90% of the vertices topple.

Results and Observations

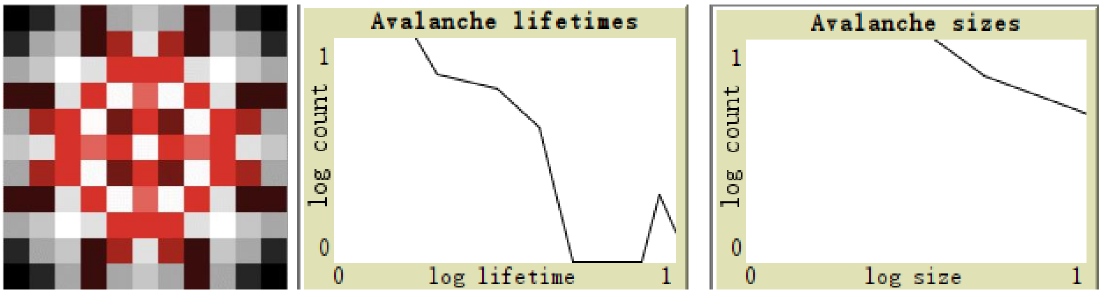
There are some other observations those would also be useful for the BIM standard development are reported during the sandpile simulation. These observations include average height of vertices using average grain count which indicates the critical state, avalanche sizes plots using log-log scale, avalanche lifetimes plots using log-log scale, the catastrophic avalanche (1/3 vertices toppling), the first extreme catastrophic avalanche after entering critical state (over 90% vertices toppling). For the settings provided in section 5.3.1, on the lattice of $L^2 = 11 \times 11$: the critical state reaches while the grain of sand is 550, the average height stopped increasing and started to show the fluctuance in a level and the distribution before the critical state is given as shown in Figure 20. The height of vertex is demonstrated with grey scale, the toppling height 0 is represent with color black ■, and the toppling height 7 is white □, and the rest is in between with equally distributed with different grey scale. The average height fluctuates around the 4.39 and range from 4.9 to 3.9. before the critical state, the average grain count shows the linear increasing, the slope is 123.2, and the intercept is - 0.21.



(a) (b)

Figure 20 Simulation result of 550 grains of sand, 20-a shows the average grain count with sand dropping, and 20-b shows the layout of sandpile at critical state

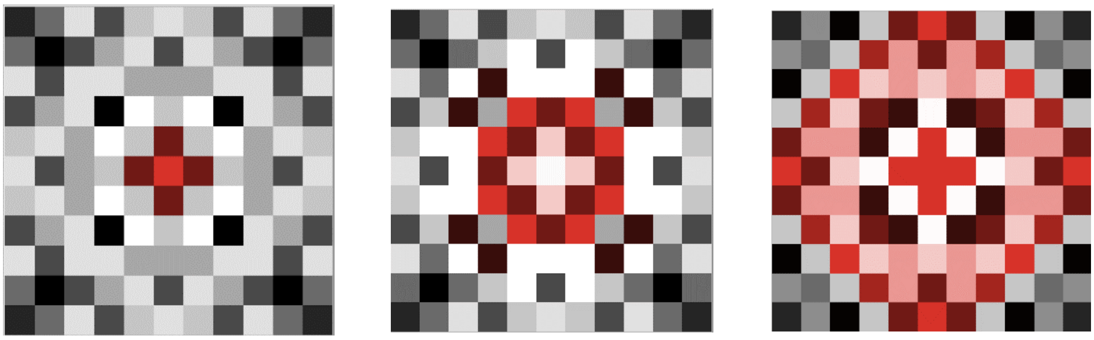
After entering the critical state, the first catastrophic avalanche happens at 575 grains of sand. The toppling vertices are given as Figure 21a, the color red covers the toppling vertices. And the avalanche lifetimes (the ratio between the frequency logarithm and lifetime logarithm) for the time being shows the tendency of sharp decrease, from 0.294 drops to 0.127 drops (Figure 21b). Accordingly, avalanche sizes (the ratio between the frequency logarithm and avalanche size logarithm) descend to 0.659 (Figure 21c). This catastrophic avalanche lasts longer than before, and during the toppling process, the vertices involved more than it appears.



(a) (b) (c)

Figure 21 Simulation result of first catastrophic avalanche. a - shows toppling scale and pattern; b shows the log-log plot of avalanche lifetimes and c- provides the log-log plot of avalanche

The first extreme catastrophic avalanche after entering critical state occurs at 623 grains of sand. The toppling involves over 90% of vertices for 11 iterations as shown in Figure 22, hence the size of avalanche is large which results in an avalanche size close to 0.



(a) (b) (c)

Figure 22 The layout of first extreme catastrophic avalanche toppling process, a shows the first iteration of toppling and the b shows the 6th iteration of toppling, c shows the 11th iteration of toppling

Other related coefficients of the sandpile simulation regarding both simulations are given in Table 2. The symbols a, b are the coefficients in equation (12), and c, d are the coefficients in equation (14).

Table 2 Other related coefficients of sandpile simulation

	Average height (5344)	(a,b)	(c,d)
$L^2 = 11 \times 11$	4.31	(-0.73, 1.35)	(-2.09, 0.47)

Other configurations are given to examine the isomorphism since the grains of sands is simplified to 334 rather than original 5334 grains. Hence, the extra control group simulations which ease the limitation on the grains of sand. With the scale of lattice enlarged, under the 1,000,000 grains dropping, the average height for $L^2 = 11 \times 11$ lattice is about the 4.55, and the average height for $L^2 = 21 \times 21$ lattice is approximately 4.59. The difference is minor. Besides, the log-log plot of each is given in Figure 23, it can be seen the log-log plot of both avalanche lifetime and avalanche size for different lattices are similar.

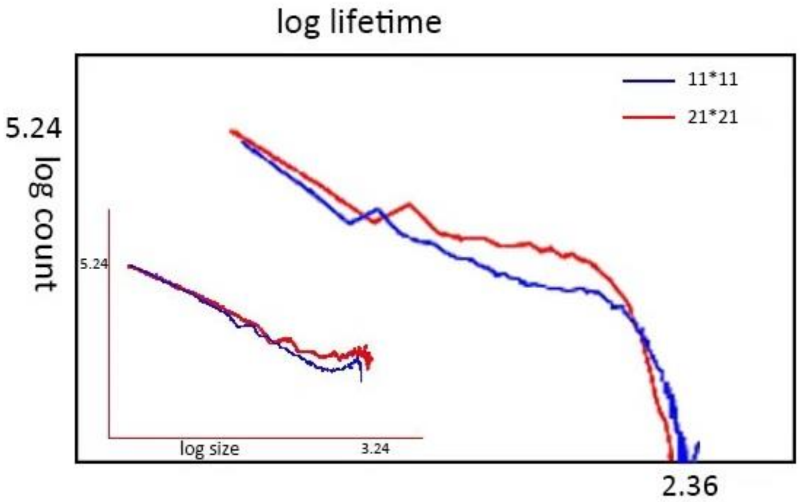


Figure 23 Comparison of log-log plot (lifetime and size) between 2 simulations

As can be seen in Figure 23, the curve of log-log plot of all avalanches shows the existence of the power law distribution. The existence of power law distribution is the fingerprint of self-organized criticality according to SOC theory (Bak, Tang, & Wiesenfeld, 1987a). The cut-off of power law curve is correlated to the scalability. Apart from minor differences on the cut-off, the log-log plot of 2 different sandpile simulations reveal great degree of fitting, so does the average grains in vertex. This indicates that there are some universalities across the scalability. It generally prevails chaos. Notwithstanding, the fingerprint of chaos—sensitivity on initial conditions and recurrence are not outward. Hence, it is considered as a stage of weak chaos which is a stage before the pure chaos (Zhao & Chen, 2002). Markedly, once the scale is set as $L^2 = 21 \times 21$ with 18,280 grains of sand and $L^2 = 101 \times 101$ with 1,000,000 grains of sand, the complete uncertainty predominates the system. As a consequence, the factual could be observed (shown in Figure 24). Thus, whether weak chaos or chaos, the proposed research brings a great step forward by detecting the existence of chaos in the BIM model development.

Chaos as a paradigm, manages to apply its related fractals to explore many other disciplines. This paper succeeds in being the first within the scholars to bridge chaos with BIM by presenting the existence of chaos in the BIM standard development. Besides, the SOC shown in Figure 23 and Table 2 shows that the information mass every entity (vertex) in MVD is determined by the general information mass, which result in changes on complexity of MVD (lattice). But initial settings of lattice and vertex capacity could cause the average size and count of avalanche, which result in average information mass accounted in each entity (vertex). Hence, it indicates that the part is determined by the whole and the whole is influenced by part. The latent order is constructed by both part and body. In other words, the form of MVD and information mass of project emerges synergetic. This, again, verifies the assumption i that expects different information mass requiring different MVD to accommodate. Moreover, another intriguing inspiration would be the information adaption. Noted, from perspective of information theory, the purpose of generating a piece of information is to expect the recipient's response, or process the information, or store the information for future use. This is called practical information; it means something to recipient. The information mass contained in the project is Shannon information before the MVD is developed to accommodate them. When the amount of Shannon information exceeds the limit, the average information accommodate in entity would decrease sharply with increasing information. Such property of information adaption could explain assumption i--why the bigger and higher standard project (Shannon information) requires greater developed MVD to adapt (practical information), which is to ensure the recipient could utilize the information with more efficiency and less ambiguity.

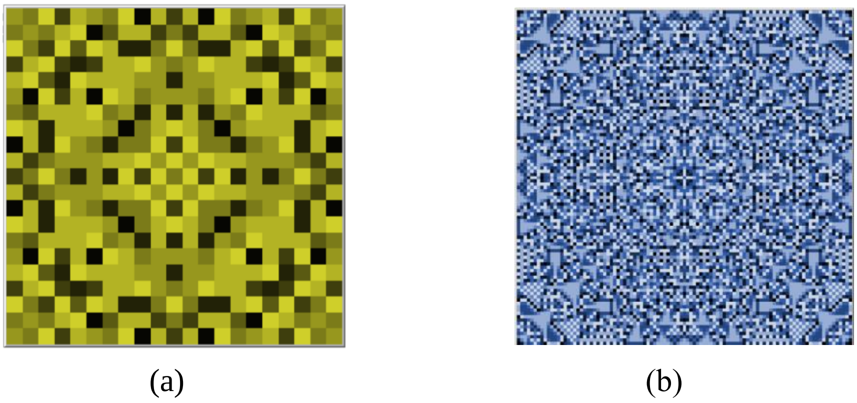

(a) (b)

Figure 24 a). Layout of $L^2 = 21 \times 21$ with 18280 grains and (b). Layout of $L^2 = 101 \times 101$ with 1,000,000 grains

5.6 Conclusion and Remarks

The exploration of information theory, with a particular emphasis on entropy, in the context of Building Information Modeling (BIM) has been a profound journey into the intricacies of data standard development. This chapter has sought to bridge the gap between abstract theoretical constructs and their tangible applications in the realm of architecture, engineering, and construction (AEC). As we reflect upon the insights garnered, several critical observations and implications emerge.

One of the most salient findings of this chapter is the discovery of chaos in BIM data standard development. This chaos, characterized by the Self-Organized Criticality (SOC) inherent in complex systems, offers a unique lens through which we can understand the challenges and opportunities in BIM. The sandpile simulation, a representative model of SOC, serves as a poignant metaphor for the dynamic and often unpredictable nature of BIM standard development. Just as a sandpile reaches a critical state of balance, teetering on the edge of stability, the BIM framework too grapples with the delicate equilibrium of managing vast amounts of information without succumbing to chaos.

The universality of the Model View Definition (MVD), grounded in information theory principles, offers a beacon of hope for the future of standardized approaches in the AEC industry. The chapter's innovative proposal of a mapping system to convert the sandpile simulation results to MVD is a testament to the interdisciplinary nature of this research. By integrating certain information requirement analyses into this mapping, we can pave the way for a more objective, reliable, and efficient BIM process.

Furthermore, the parallels drawn between the ID3 algorithm, the sandpile simulation, and the MVD development process emphasize the holistic approach required in BIM research. The potential of information theory, especially entropy, in enhancing data interoperability within the IFC framework cannot be understated. By quantifying uncertainty and redundancy, we can provide a more comprehensive understanding of BIM which focus on the dynamics of BIM standard evolvement and development from perspective of information theory, highlighting what and how information system/expression cope with BIM standard while instantiating developed standard in practice. Considering the objectivity, reproducibility and generalization are undoubtful general benefits brought in with quantification regardless of subjects, quantifying uncertainty and redundancy is definitely beneficial, as aforementioned in Section 2.2, it hinders the development of standard once lacking generalization. And solely user-centered development of standards tends to create confusion and inconsistency, hence rising redundancy. With quantification, it could be resolved.

However, as with all research endeavors, this exploration is not without its limitations. While the sandpile model and the proposed mapping system offer valuable insights, they represent just one facet of the multifaceted challenges of BIM. The AEC industry is a complex ecosystem, with myriad stakeholders, objectives, and constraints. Future research must endeavor to integrate multiple models, theories, and perspectives.

The implications of the findings in this chapter are vast. The discovery of chaos in BIM data standard development challenges the traditional linear approaches that have dominated the field. It calls for a paradigm shift, urging researchers and practitioners alike to embrace the complexities and uncertainties inherent in the AEC industry. The proposed mapping system, grounded in the principles of information theory, offers a structured and systematic approach to navigate these complexities. By leveraging the power of entropy and the insights from the sandpile simulation, we can usher in a new era of BIM research, one that is more aligned with the realities of the AEC industry.

In conclusion, this chapter has illuminated the potential of information theory in revolutionizing BIM standard development, especially when viewed through the lens of SOC. The discovery of chaos in BIM data standard development and the proposed mapping system are pivotal contributions to the field.

Chapter 6 – Proving Chaotic Properties of Data Standard Development System for Building Information Modeling

6.1 Introduction

In the evolving landscape of architectural and construction research, Building Information Modeling (BIM) has emerged as a transformative paradigm, reshaping the way we conceptualize, design, and implement built environments. At the heart of BIM lies the intricate web of data standards, which, while foundational, presents a complex and dynamic system. This chapter delves into the heart of this system, exploring its chaotic properties and shedding light on its inherent intricacies. The journey of scientific exploration is often marked by moments of revelation, where established paradigms are challenged, and new horizons are unveiled. In the realm of BIM research, understanding the underlying dynamics of the data standard development system is one such frontier. While the system's complexities are evident, the nature of its behavior – whether deterministic, random, or chaotic – remains an open question. Last chapter discovered the existence of chaos in the BIM data standard development, but it is not the definite proof of system being chaotic. This chapter seeks to address this gap, embarking on a methodical journey to unravel the system's chaotic properties.

The motivation for this exploration stems from the recognition that chaos, while often perceived as disorder, is a manifestation of underlying patterns and deterministic laws. If the BIM data standard development system exhibits chaotic behavior, it implies a delicate interplay of order and unpredictability. Understanding this interplay is crucial, not just from a theoretical standpoint but also for its practical implications. It can guide the development of robust data standards, inform strategies for BIM implementation, and shape the future trajectory of architectural and construction research. Central to our exploration is the concept of the Lyapunov exponent, a metric that serves as a litmus test for chaos. A positive value for this exponent signifies the system's inherent sensitivity and propensity to exhibit chaotic behavior. But arriving at this metric is not straightforward. It requires a meticulous approach, from data preparation to phase space reconstruction, each step building upon the previous, culminating in a comprehensive understanding of the system's dynamics.

This chapter is structured methodically, mirroring the systematic approach adopted in our exploration. We begin with data preparation, leveraging sandpile simulations to capture the essence of the increasing information mass. The time series derived from this data serves as our guiding narrative, chronicling the system's behavior over successive intervals. The reconstruction of the phase space follows, providing a multi-dimensional view of the system's dynamics and paving the way for the identification of strange attractors. The culmination of this process is the calculation of the largest Lyapunov exponent, which, as we shall see, provides definitive proof of the system's chaotic nature. Essentially, this chapter is more than just an exploration; it is a testament to the power of rigorous scientific inquiry. It showcases the depth and breadth of research that goes into understanding complex systems, highlighting the challenges, revelations, and moments of epiphany that define the journey. As we navigate the intricate landscape of the BIM data standard development system, we invite readers to join us on this odyssey, to witness the dance of chaos and order, and to appreciate the beauty that lies in complexity.

6.2 Data Preparation and Sampling from Sandpile Simulations

Every groundbreaking scientific endeavor is built upon the foundation of meticulously prepared data. It is akin to the blueprint for an architect or the notes for a musician. In our exploration into the chaotic properties of the data standard development system for Building Information Modeling (BIM), this foundational step took on paramount importance.

The sandpile simulations, while metaphorical in nature, were more than just a creative representation. They were a bridge, connecting abstract concepts to tangible observations. By visualizing the increasing information mass as grains of sand, we were able to grasp the dynamic interplay of variables within the system. Each grain, each shift, mirrored the complexities and nuances of the BIM data standard development process. However, beyond this visual metaphor, lay a more intricate and structured framework: the time series. This was not just a collection of data points but a narrative, a story that unfolded over successive intervals. Each data point in the time series was like a snapshot, capturing a moment in the ever-evolving dance of numbers and patterns. Together, they wove a tapestry that chronicled the standard development process in all its complexity.

This chronological narrative, with its ebbs and flows, peaks and troughs, provided invaluable insights. It shed light on patterns that might have otherwise gone unnoticed, highlighting trends, anomalies, and critical junctures. In a realm where data can often seem overwhelming, akin to a vast, uncharted ocean, the time series served as our compass. It ensured that our exploration was not aimless but was guided by precision, purpose, and clarity. Furthermore, the meticulous nature of our data preparation underscored a broader principle: the importance of grounding scientific inquiry in robust and reliable data. It reinforced the idea that true understanding stems not just from observation but from deep, structured analysis. In the grand tapestry of our exploration, data preparation was not just a starting point but the very thread that wove the narrative together.

Data preparation and sampling starts with constructing dynamic system that could stimulate the standard development over information mass changes. Referring to Section 5.2, the proposed method to simulate the standard development process over information mass changes with sandpile simulation. The proposed method regards the sand as information, information mass increasing is metaphorized as dropping grains. And with proposed denotation system, the information standard topology is metaphorized as sandpile lattice layout. In the study, this analogy is justified through information entropy definition specified by Shannon and the tree graph's nature—directed weight graph prescribed by Reinhard. Under these conditions, the dynamics of sandpile simulation is vested into the information mass –information standard development. On basis of this, the time series data could be generated with following regulations:

i. From the point of view of extract quantitative information from observations, the characteristic features would pose an interesting feature to the observer. First of all, it is typical to observe only one dynamical variable which govern the behavior of system. The characteristic feature just outlined before is required to be capable of describing standard development. And previous research by An developed a mapping system to map the standard topology to tree structure, which enables the factors for tree structure from graph theory – node fission entropy to present the significance of nodes (demonstration given in Figure 25), which could govern the behavior of tree structure development to be used as the dynamical variable. Another interesting factor about the tree structure is the node amount, which sum up all the nodes in different levels of tree structure. It is also an indicator for complexity of tree structure. Hence, the study regulates the developed standard topology –tree structure 's node fission entropy integrates node amount by assigning different weights to nodes in different level considering the alignment with Level of Development concept (nodes in different level indicate different level of development) as target observation dynamical variable. The fission entropy is presented by weights that determined by in LoD conceptual context and information mass (grains of sand) the entity (node/vertex) would contain. In other words, nodes in different level would present entities that contained different level of detailed information mass aligned with associated LoDs in BIM instantiation, it is denoted as D.

ii. To address the problem, we will have to focus our attention on discrete time maps first. This is really no restrictions as in some sense all analysis of physical takes place in discrete time: we never sample anything continuously. If we sample a scalar signal $s(t)$ at time intervals τ_s starting at some time t_0, then our data is actually of the form $s(n) = s(t_0 + n\tau_s)$, and the evolution we observe takes us from $s(k)$ to $s(k+1)$. So, the observations take $s(t_0 + k\tau_s) \rightarrow$ s $(t_0 + (k+1)\tau_s)$, $s(k) \rightarrow s(k+1)$. And owing to these, the sand dropping operation can be regarded as time series. The would be dropped in a constant interval, which in the case, τ_s at some amount of dropped sand t_0, here denoted as d_0. The scalar signal associated with sand dropping would be development of standard, denoted as $D(d_0)$. Accordingly, the data is actually of the form $D(n) = s(d_0 + n\tau_s)$, and the evolution we observe takes us from $D(k)$ to $D(k+1)$. So, the observations take $D(d_0 + k\tau_s) \rightarrow$ s $(d_0 + (k+1)\tau_s)$, $s(k) \rightarrow s(k+1)$.

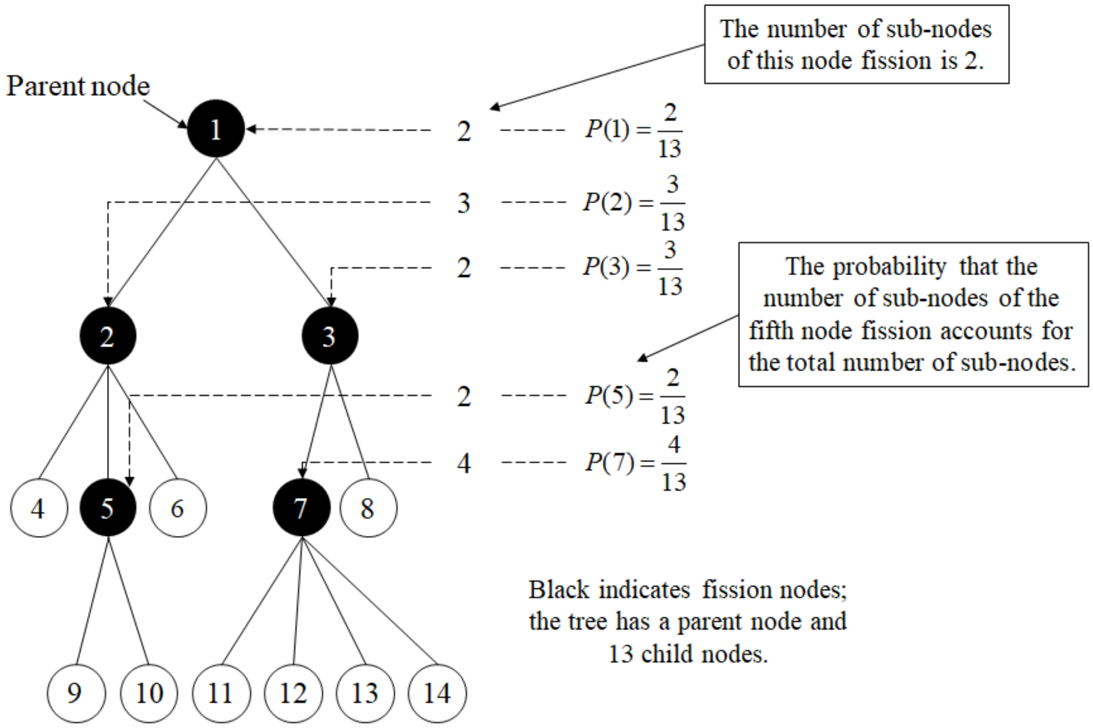


Figure 25 Node fission entropy in a tree graph

Simulation Initial settings and data sampling

The data sampling starts with the sandpile simulation, the simulation results would be the layout of lattices. And then with the mapping system in Section 5.3.2, it could be converted into a tree structure. Afterwards, the new indicator that considers both node amount and node fission entropy conceptual thinking is defined as an observed scalar signal. With sand dropping (more information added to the information model), different layout of lattice would be generated. Accordingly, different tree structures which would derive different scalar signals for observation while the instantiating defined indicator to the system. These collected scalar signals forms the data samples. The first setting for the sampling would be sandpile simulation initial settings. The initial settings of sandpile simulation in this study is the succession of Chapter 5. All the vertices start with 0 grain of sand. The dropping starts and stays on the center vertex.

The capacity for vertex to start toppling is 8 grains. And the toppling rule is to give a grain to each of its neighbor for vertex that reaches capacity. After the dropping is completed, convert the generated layout of lattice to tree structure. And the tree structure could be translated into MVDs, thanks to previous scholars. Considering a vital concept LoDs, the upper level of node can be seen as an entity that contains more information and lower level of node can be interpreted as entities containing less information. Due to that, the study would accordingly assign weight to different level of nodes to represent the informative level these nodes can be. The higher-level node locates, more weight would be assigned to. The settings are to assign the 1 to the lowest level of node, and then linearly increasing by 1 with level goes up. The indicator that would be defined as observed scalar signal is the weighted summation of all nodes, the sand grains/information mass accommodated by each node/entity/vertex multiply by their assigned weights, and then accumulated all the nodes. Sampling process is documenting every observed scalar signal (weighted summation) with sand dropping from start to end.

To better understand the method here and above, the proposed mechanism is demonstrated through an illustrative example that focuses on shield tunnel lining. The sandpile system, as a dissipative system implicitly exchanges between mass and energy (toppling rules), which can be analogically regarded as a simplified system showing the process of restructuring (recategorizing entities) and extending MVD which is a regarded as proxy of BIM data standard in the study. To make the demonstration more comprehensive, ensure a rigorous statistical sense of simulation and realistically reflect the trend of MVD development processes, “bit” is used to represent the information mass, and the two-dimensional lattice are set as $L^2 = 21 \times 21$, with the initial starting layout as no vertex contains sand. Followingly, the ticks (sands dropping) are total grains of sands. The case is a simplified version based on a single layer shield tunnel as underpass, the whole line is 0.4 kilometers. All the tunnels are underground, considering the width of a ring is roughly 1.2 meters, the first underpass would have 334 rings and each ring contains 16 bits of information. The total information mass is approximately 5344 bits. While dropping grains of sand, one at a time, sand grains can fall off the edge of the table, helping ensure the avalanche eventually ends. Considerably, there are limited vertices, 21×21. Even all the vertices reach their full capacity, it would be only 3200 grains. The greater number of grains of sand falling off the edge of the table would lack value to the study. Thusly, the grains need to be limited. Constructing isomorphism can be of great assistance. With isomorphism, the grains of sand can easily be converted to 334 (5344/16) grains of sand (An et al, 2023a). Since the vertices is 21×21, according to assumption before, the furthest layer (lowest level of tree) away from center (highest level) would be assigned with weight 1, and the center layer (highest level) would be assigned with weight 11. Sampling is documenting every weighted summation with sand dropping from 1 to 334. Then the sample data would be given as follow (code given in appendix B):

```
 [(1, 11), (2, 22), (3, 33), (4, 44), (5, 55), … , (328, 2688),
(329, 2699), (330, 2710), (331, 2721), (332, 2732), (333, 2743),
(334, 2754)]
```

With the regulation ii, the sand dropping is “time passing”. The sampled data could be reorganized directly to a time series data as follow:

```
[11, 22,33,44,55, … , 2688, 2699, 2710, 2721, 2732, 2743, 2754]
```

Hamiltonian Mechanics

For better understanding the way of expressing dynamic system via phase space, Hamiltonian Mechanics is introduced into this dissertation due to its excellence on time evolution of dynamical system during analysis. Hamiltonian mechanics is a theoretical framework in classical mechanics, formulated by Sir William Rowan Hamilton in the 19th century. It offers a reformulation of classical mechanics and provides a powerful and elegant way of describing the evolution of physical systems over time. Unlike Newtonian mechanics, which focuses on forces and acceleration, or Lagrangian mechanics, which is centered around the principle of least action and deals with energy, Hamiltonian mechanics uses the concept of the Hamiltonian to describe systems. Hamiltonian (H): The Hamiltonian of a system is a function that represents the total energy of that system, expressed as the sum of its kinetic (T) and potential (V) energies. The Hamiltonian is usually denoted by (H), and it's a function of the system's generalized coordinates (q) and their conjugate momenta (p), i.e., $H = H(q, p, t)$. Hamiltonian mechanics uses generalized coordinates (q_i) and their conjugate momenta (p_i) to describe the state of a system. These variables can be more abstract than simple position and momentum, allowing for a more generalized and flexible description of systems. In the research, the evolution of system state is denoted with such way, which is expressed the sum of its kinetic (T) and potential (V) energies. The abstract and generalized potential (V) energies is the node fission entropy given in earlier this section. And the kinetic (T) is the information mass changing 1 bit/nat/lin each time (similar to sand drop one grain each time).

6.3 Reconstructing Phase Space

The challenge of understanding complex systems often lies in their intangible nature. Much like trying to decipher the intricate flight pattern of a bird solely from the fleeting shadows it casts, our endeavor to grasp the dynamics of the data standard development system for Building Information Modeling (BIM) presented similar challenges. The system's behavior, with its myriad interactions and variables, was like a dance hidden in the shadows, elusive and intricate. Reconstructing the phase space was our beacon in this quest for understanding. This wasn't just a mathematical exercise but a transformative process. By mapping the system's state variables into a multi-dimensional space, we were essentially creating a stage where the hidden dance of the system could be visualized and studied. Each dimension in this reconstructed space represented a facet of the system, and together, they provided a holistic view of its dynamics.

The emergence of the strange attractor within this space was nothing short of revelatory. These attractors, with their intricate and often mesmerizing patterns, served as focal points, drawing our attention to the inherent rhythms and patterns of the system. They were like the footprints of the system's dance, capturing its essence in a tangible form. The strange attractor's patterns, looping and spiraling, were a testament to the system's inherent complexities, revealing the delicate balance between order and chaos. Moreover, this reconstruction process underscored the importance of having the right tools and perspectives in scientific exploration. Without the phase space, the system's dynamics would have remained a mystery, hidden in the shadows. But with it, we were granted a front-row seat to one of nature's most intricate choreographies. It was a reminder that in the realm of scientific inquiry, the right perspective, combined with the right tools, can illuminate even the most elusive of phenomena. For all intents and purposes, reconstructing the phase space was not just a step in our exploration but a paradigm shift. It transformed our understanding, turning the invisible visible and the intangible tangible. It was our window into the heart of the system, revealing its rhythms, patterns, and inherent beauty.

The essence to proving an observed system chaos is to find its positive largest Lyapunov Exponent. Prior to determining largest Lyapunov Exponent, the reconstruction of phase space is a requisite. The purpose of reconstructing phase space is to rebuild strange attractor in high dimensional phase space. The strange attractor, as one of basic features that chaotic system reveals, scratch the regularity of chaotic system. In other words, the chaotic system would finally fall into a certain orbit. And such orbit is the strange attractor. The evolution of every component within system is determined by other components those would interact with. Hence, the information of these related components would be hidden in the evolution process of the component. With such, it enables extracting and rebuilding original patterns of system from a time series data of the component. The pattern is an orbit in high dimensional space by its nature. Put differently, over time, an orbit generated by a chaotic system would evolve itself to a motion with pattern, and then form a regular trajectory. While transforming this trajectory to relate time series through stretching and folding, the trajectory appears feature of chaotic and complicated. The data of time series is correlated due to the interaction among factors which drive the system chaotic. Packard at al suggested the idea of reconstructing a phase space through delayed measurements (Packard, Crutchfield, Farmer, & Shaw, 1980). Embedding theorems proposed by Takens ensure that locating a suitable embedding dimension m enables strange attractor unfolded in such embedding dimensional subspace as long as $m \geq 2u+1$ in which u represents dimension of strange attractor in original phase space (Takens, 1981).

$$Definition\ 1\ (N,p),(N_1,p_1)\ are\ two\ metric\ space, if\ \exists\ \varphi\ N \to N_1 s.t.(1)\ \varphi\ surjective;\ (2)\ p(x,y) = p_1\big(\varphi(x),\varphi(y)\big)(\forall\ x,y \in N), then\ (N,p),(N_1,p_1)\ is\ isometry.$$

$$Definition\ 2\ If\ (N_1,p_1)\ and\ the\ subspace\ (N_0,p_2) of\ another\ metric\ space\ (N_2,p_2)\ is\ isometry,\ then,(N_1,p_1)\ could\ be\ embedded\ into\ (N_2,p_2)\ .$$

A time series data with length N, $\{s(t_0+k\tau_s): k=0,1,\dots,N-1\}$, reconstructed phase space is $\mathbb{R}^m$, then the reconstructed trajectory would be

$$X_i = \big(x(t_i), x(t_i+\tau_s), \dots, x(t_i+(k-1)g\tau_s)\big),\quad i = 1,\tau_s,3,\dots,M, \qquad \text{equation (15)}$$

$$x(t_r) = s(t_0+(r-1)\Delta t), r = 1,2,\dots,N, \qquad \text{equation (16)}$$

X_i represents the ith point among M points on reconstructed trajectory in reconstructed phase space $\mathbb{R}^m$. And $M = N-(m-1)g$, m is embedded dimension, $\tau = g\tau_s$ is time delay, $\tau_w = (m-1)\,\tau$ is time window, and τ_s is sampling interval. Obviously, to reconstruct phase space, 2 parameters (τ, m) need to be determined at first.

Choosing time delay

Since the embedding theorem is silent on the choice of time delay to use in constructing m-dimensional data vectors. Some methods are developed to choose optimal time delay, including average mutual information method, C-C method, autocorrelation method, small-window solution method, fill factor method, average displacement method, and such. This study would rely on "eliminated partial complex autoregressive function". The rationale of employing eliminated partial complex autoregressive function to determine time delay is that a) its mathematical expression is concise; b) it is easy to compute as well as moderate complexity; c) it does not attach to the length of series (sample size) (C. Wang & Yi, 2007). The eliminated partial complex autoregressive function is given as:

$$R_{XX} = \frac{1}{N}\sum_{i=1}^{N-\tau}(X_i - \bar{X})(X_{i+\tau} - \bar{X}) \qquad \text{equation (17)}$$

N represents for the number of samples, X_i stands for the i^{th} sample, $\bar{X}$ is the mean value of series. R_{XX} would decrease with τ increasing, and τ would be chosen time delay while R_{XX} decreases to $initial\ value \times (1 - e^{-1})$. Based on the sample provided in chapter 5.4, the chose time delay would be 61 (code given in appendix C).

Determining the dimension of reconstructed phase space

G-P algorithm proposed by Grassberger and Procaccia is widely used to determine embedding dimension for reconstruction of phase space(Grassberger & Procaccia, 1983). The G-P algorithm is conducted through:

(1) Given time series $x_1, x_2, \dots, x_{n-1}, x_n, \dots$, firstly start with a small integer m_0. Then a m_0 dimensional space is constructed, the first vector in the space would be $X_1 = (x_1, x_2, \dots, x_{m_0-1}, x_{m_0})$.

(2) Take any point X_i =$\{X_{i1}, X_{i2}, \dots, X_{im_0}\}$ from m_0 dimensional space as a reference point. Calculate the distance between reference point and the rest $N-1$ points. Take reference point as center of circle and take r as the radius of circle. Then count the points fall into the circle, form the cumulative distribution function $C_m(r)$. Calculate correlation function:

$$C_m(r) = \lim_{N\to\infty} \frac{1}{N\sum_{i\neq j} H(r-|X_i-X_j|)} \qquad \text{equation (18)}$$

$H(*)$ is Heaviside step function, $|X_i - X_j|$ is the distance between phase point X_i and X_j. Then the cumulative distribution function $C_m(r)$ could express the probability of distance between 2 points within attractor smaller than r.

(3) With proper range of r , the dimension of attractor d is log-linear related to $C_m(r)$, that is $d(m) = lnC_m(r)/lnr$. Hence, $d(m_0)$ regards to m_0 could be solved through fitting.

(4) Increasing $m_1 > m_0$, repeat step (2) and (3) until $d(m)$ no longer increase significantly as m increases and remain constant in a range of considerable error. The $d(m)$ is the correlation dimension of attractor. Noticeably, if $d(m)$

keeps increasing as m increasing, then the system would be a stochastic time series.

With the given time series data in section 6.1, the embedding dimension would be 3 (code given in appendix D).

To conclude, the chosen time delay and embedding dimension for reconstructing phase space would be 61 and 3 respectively. The reconstructed phase space would be a dimensional space, and with the chose time delay, the phase point/ vector of would be expressed $X_i = (x_i,\ x_{i+61}, x_{i+122})$.

6.4 Determining the Lyapunov Exponent

In scientific research, certain milestones distinctly shape the trajectory of inquiry. One pivotal point in our investigation into the chaotic properties of the data standard development system for Building Information Modeling (BIM) was the calculation of the largest Lyapunov exponent. The Lyapunov exponent, a term perhaps arcane to those unfamiliar with chaos theory, stands as a pivotal metric. It quantifies the rate at which system trajectories diverge or converge, effectively serving as an indicator of system predictability. A positive value for this exponent signifies that trajectories, even those extremely close initially, will diverge over time, underscoring the system's inherent sensitivity. The process of calculating this exponent transcended mere mathematical formalism; it represented a pursuit of clarity. We aimed to identify a robust metric encapsulating the system's core behavior. The emergence of a positive value was revelatory, corroborating the system's chaotic tendencies, as suggested by other aspects of our research.

This discovery can be analogized to a detective pinpointing the conclusive evidence in a convoluted case or an archaeologist retrieving the final shard of an ancient relic. Amidst the expansive terrain of data and patterns, the Lyapunov exponent emerged as a guiding beacon, elucidating our comprehension and affirming our hypotheses. Beyond mere affirmation, the significance of this metric extends to its predictive power. It sheds light on the system's prospective behavior, offering insights into its inherent dynamics, stability, and potential paths. This exponent serves not just as a metric but as a navigational tool, steering future BIM research and innovations. In summary, the Lyapunov exponent is more than a mere numerical value; it epitomizes the essence of methodical scientific investigation. It stands as irrefutable evidence of the chaotic nature inherent in the BIM data standard development system.

In other words, in the realm of dynamical systems, the Lyapunov exponent stands as a quintessential metric, offering profound insights into the inherent nature of a system's trajectories. This exponent, denoted as a real number, quantifies the average rate at which trajectories, originating from proximate initial conditions in the phase space, either converge or diverge over the entire attractor. The attractor, in this context, refers to a point or an ensemble of points in the phase space that delineates the potential steady states of the dynamical system under consideration. The significance of the largest Lyapunov exponent, in particular, cannot be overstated. When this exponent manifests a positive value, it bears testament to a pivotal characteristic of the system: two trajectories, though initiated from infinitesimally close points, will inevitably diverge as time progresses. This divergence is not a mere random drift but is emblematic of the system's inherent chaotic nature. Such behavior underscores the celebrated 'butterfly effect,' a phenomenon where minuscule perturbations in the initial conditions can lead to vastly different outcomes, rendering long-term predictions virtually impossible (Sprott, 2003; Williams, 1997).

The presence of a positive largest Lyapunov exponent is not just an academic curiosity; it serves as a rigorous litmus test for chaos in the system. This metric's importance becomes even more pronounced when we transition into the realm of complex systems, such as the sandpile model, which exemplifies the principles of self-organized criticality (SOC). The sandpile model, a paradigm of emergent complexity, showcases how local interactions can culminate in global patterns, often teetering on the brink of instability. The interplay between the Lyapunov exponent and the principle of SOC provides a rich tapestry of understanding, bridging the gap between deterministic chaos and emergent phenomena in complex systems (Abarbanel, 1996; Chen, 2005; Sprott, 2003; Williams, 1997). Within the vast domain of dynamical systems, the Lyapunov exponent emerges as a pivotal metric, shedding light on the intricate behaviors of system trajectories. Determining this exponent, though it may appear direct at first glance, is layered with complexities that require rigorous scrutiny. Historically, numerous techniques have been introduced to ascertain this value, each boasting its unique strengths and challenges. Notably, the Jacobian and Wolf algorithms have risen to prominence, capturing the interest and rigorous examination of scholars in the field.

The Jacobian algorithm, rooted in the principles of differential geometry, leverages the Jacobian matrix to compute the rate of divergence or convergence of trajectories in the phase space. Its robustness in the face of noisy time series data makes it an invaluable tool for real-world applications where noise is often an inescapable reality. However, its prowess is not without limitations, especially when dealing with pristine, noise-free data. Enter the Wolf algorithm. This methodology, pioneered by Wolf et al. (1985), offers a refined approach, especially tailored for time series data devoid of noise. At its core, the Wolf algorithm meticulously measures the divergence between neighboring points in the reconstructed phase space. The procedure commences with the selection of two proximate points within this space. As time progresses, these points evolve, tracing their respective trajectories. After a predetermined time interval, these initial points metamorphose into two new positions, and the divergence between them is meticulously quantified. For the purposes of this study, our data is derived from simulations, a realm where noise is conspicuously absent. Given this pristine nature of our dataset, the Wolf algorithm emerges as the methodology of choice. Its ability to accurately compute the Lyapunov exponent in noise-free environments makes it an indispensable tool for our research.

In the application of the Wolf algorithm within reconstructed phase space, divergence between neighboring points is measured to assess dynamic stability. Initially, two neighboring points are selected. Over a known time interval, these points evolve into new positions. The divergence is quantified by comparing the distances between the original and the evolved points, labeled as d_1 and d_2, respectively. The ratio of d_1 to d_2 illustrates the trajectory's divergence. If d_1 significantly exceeds d_2, indicating substantial divergence, a new nearby point is chosen to minimize the angular difference between the trajectory of the original and the new points. This process is iteratively repeated, updating the distances and their ratio to reflect changing dynamics. Figure 26 visually depicts this methodology. The logarithmic average of these ratios ultimately determines the largest Lyapunov Exponent, L_1, a measure of the system's sensitivity to initial conditions.

$$L_1 = \left(\frac{1}{t}\right) \sum_{i=1}^{m} \log_2 \left[\frac{L^{'}\,(t_{i+1})}{L(t_i)}\right] \qquad \text{equation (19)}$$

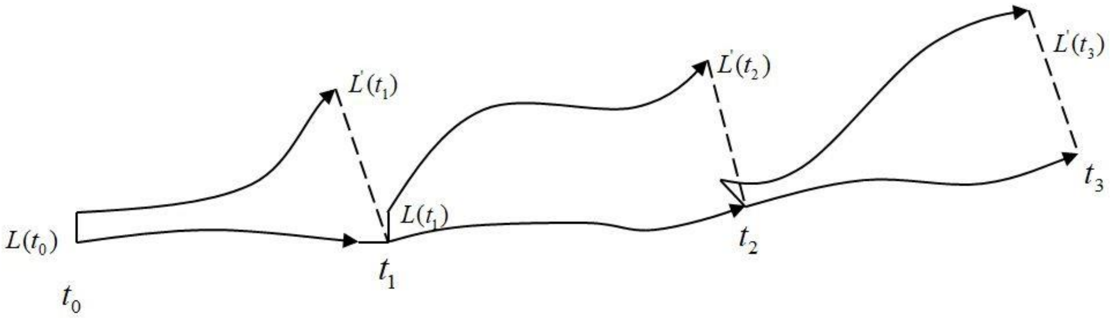


Figure 26 Wolf algorithm visualized demonstration for calculating largest Lyapunov Exponent

The computation process is scripted in python (Appendix E). With Wolf algorithm, under the condition embedding dimension $m = 3$, time delay $\tau = 61$, the largest Lyapunov exponent $L_1 = 2.039$. The scenario is with sample size

500 grains of sand as aforementioned. Apparently, the largest Lyapunov exponent is positive and not close to zero, which is the sign that dynamical system reveals strong chaos. Hence, it is a rigorous mathematical proof of chaos existence in the dynamical system information standard development against information mass.

6.5 Conclusion and Remarks

The meticulous exploration into the chaotic properties of the data standard development system for Building Information Modeling (BIM) stands as a beacon of rigorous scientific investigation. This exploration wasn't just a cursory glance into the realm of BIM but a deep dive, seeking to unravel the intricate patterns and behaviors that define its very essence. At the heart of this exploration was the recognition that understanding such a complex system required a structured and methodical approach. The sandpile simulations, for instance, served as a foundational tool, allowing us to metaphorically represent the increasing information mass. This wasn't just a creative representation but a calculated endeavor to capture the dynamic interplay of variables within the system. By viewing the information mass changes through the lens of these simulations, we were able to glean insights into the system's inherent tendencies and behaviors.

Furthermore, the reconstruction of the phase space was not just a step but a pivotal juncture in our investigation. This reconstruction provided a multi-dimensional view of the system's dynamics, offering a visual representation of the strange attractors that underpin its chaotic nature. These attractors, with their intricate patterns, served as a roadmap, guiding our understanding of the system's long-term behavior. The choices made during this exploration, such as the selection of time delay and the reconstructed dimension, were not arbitrary. They were the result of careful deliberation, grounded in solid theoretical principles. Each choice was akin to adjusting the focus on a microscope, ensuring that we captured the finest details of the system's behavior.

The culmination of this methodical approach was the calculation of the largest Lyapunov exponent. This metric, while a single value, encapsulated the essence of our findings. It was the definitive proof, the seal of authenticity, confirming the system's chaotic nature. In essence, this exploration was not just about proving a point but about embarking on a scientific odyssey. Each step, each decision, was a testament to the power of systematic inquiry, showcasing the depth and breadth of scientific rigor that went into understanding the chaotic properties of the BIM data standard development system.

Data Preparation and Its Significance

The foundation of our exploration was laid with the careful preparation of data. Using sandpile simulations, we metaphorically represented the increasing information mass, providing a tangible framework to understand the abstract dynamics at play. This data, while serving as a representation, was pivotal in capturing the essence of the standard development process and its inherent complexities.

Reconstructing Phase Space: A Crucial Step

The reconstruction of the phase space was a significant milestone in our journey. This step, which aimed to rebuild the strange attractor in a high-dimensional phase space, was instrumental in visualizing the system's dynamics. By understanding the system's behavior in this reconstructed space, we gained insights into its inherent tendencies and potential trajectories.

Time Delay and Dimension Selection

The choice of time delay and the reconstructed dimension were not arbitrary but were grounded in methodical considerations. These choices, which influenced the accuracy and reliability of our subsequent analyses, were made after careful deliberation. The time delay, in particular, played a crucial role in ensuring that the reconstructed phase space accurately captured the system's dynamics.

Lyapunov Exponent: The Definitive Proof

The culmination of our methodical approach was the calculation of the largest Lyapunov exponent. This metric, a definitive indicator of a system's chaotic nature, provided the rigorous proof we sought. A positive value, especially one as significant as we observed, was a clear testament to the system's sensitivity and its propensity to exhibit chaotic behavior.

Reflecting on the Journey

In retrospect, this chapter has been a methodical and systematic journey towards proving the chaotic nature of the BIM data standard development system. Each step, from data preparation to the calculation of the Lyapunov exponent, was crucial in building our case. The meticulous nature of our approach ensured that our conclusions were grounded in solid scientific principles, offering a robust and comprehensive proof of the system's chaotic nature.

As we conclude this chapter, it is essential to appreciate the broader implications of our findings. Proving the system's chaotic nature not only challenges traditional paradigms but also paves the way for future research and innovations in the realm of BIM. The journey, while challenging, has been enlightening, and the insights gained will undoubtedly shape the future trajectory of BIM standardization.

Chapter 7 – Advanced Engineering Applications for Chaos in Building Information Modeling

7.1 Introduction

Scientific research often concludes with the practical application of its findings. In this chapter, we bridge the gap between theory and application, as we harness the insights of chaos theory to advance Building Information Modeling (BIM) in the Architecture, Engineering, Construction, and Operations (AECO) sector. Here, we delve into advanced engineering applications that leverage the latent dynamics within BIM data standard development, with the overarching goal of reshaping how we conceive, design, and construct our built environment. Our foray into this chapter beckons us to revisit our primary research objective — the quest to unearth and exploit the concealed dynamics within BIM data standards to enrich the AECO industry. We have traversed the intricacies of chaos theory, decrypted the chaotic essence of BIM data, and deepened our comprehension of how chaos can be a catalyst for innovation. Now, it is time to witness the transformation of theory into pragmatic solutions that can redefine the efficiency, quality, and sustainability of AECO projects.

Model View Definitions (MVD) serve as the foundational elements of BIM data standards. Their accurate formulation is critical for successful project outcomes. By applying insights from information theory and sandpile simulation, we propose a method for MVD generation that surpasses traditional techniques. This methodology not only incorporates principles of quantifiable but also provides a framework for creating MVDs that are adaptable, robust, and representative of the intricate dynamics of AECO projects. Chaos theory offers an alternative lens for BIM Data Standard Evaluation, presenting novel methods to assess data standard quality. We will elucidate how chaos theory translates into practical metrics and indicators, providing an enhanced capacity to evaluate the efficacy and dependability of BIM data standards. Our assessment goes beyond surface-level analysis, delving into the essence of BIM data and offering valuable insights for stakeholders. This is achieved via better understanding system attractors that is essential to grasp the complex behavior within BIM. Chaos theory provides methodologies to detect and characterize these attractors, revealing their impact on project dynamics. This knowledge enables the recognition of nuanced changes within BIM data standards, enhancing our ability to fine-tune project processes. The climax of our research is the introduction of indicators based on the Euclidean distance to system attractors. Representing the practical aspects of chaos theory, these indicators provide an exact method for assessing BIM model performance. By quantifying the energy needed to sustain a system, its efficiency and resilience can be determined. The utility of these indicators is validated through real-world case studies, providing stakeholders and scholars with a robust method to improve BIM model outcomes as well as a perspective to look into data standard development not only in term of Building Information Modeling but also other kinds of data standard in other domains/sectors/industries.

Subsequent sub-chapters will intertwine these advanced engineering applications, showcasing their interdependent relationship and potential to bring transformative changes to the AECO sector. Chaos theory evolves from being a mere academic interest to a driving force that informs our understanding and fosters innovation. This research bridges the gap between chaos theory and BIM, heralding a new era in AECO practices. This chapter sets the context for the ensuing discussions by highlighting the interplay between theoretical insights and practical applications. Emphasis is placed on the transformative potential of chaos theory within the AECO realm. With this foundation established, the forthcoming sections will present a detailed analysis of the proposed methodologies and their implications.

7.2 Application 1--Model View Definitions (MVD) Generation

In the ever-evolving landscape of Architecture, Engineering, Construction, and Operations (AECO) industries, Building Information Modeling (BIM) has emerged as a transformative force. Central to the efficacy of BIM lies the concept of Model View Definitions (MVDs). This, often likened to the core part of the BIM ecosystem, are instrumental in shaping the digital representation of projects, governing information exchange, and ensuring data standards compliance. As we venture into the uncovered findings into BIM practices, we would have to delve into the core of MVDs, exploring their generation through the lens of information theory and sandpile simulation.

To appreciate the significance of MVDs, one must first understand their pivotal role in the AECO industry. MVDs are akin to blueprints that prescribe how information about a building project should be structured and exchanged. They provide the framework for data interoperability among different software applications, facilitating seamless collaboration among various stakeholders, from architects and engineers to contractors and facility managers. However, traditional approaches to MVD generation often fall short in capturing the complexities inherent to modern construction projects and retaining consistency of MVDs' development. The dynamism of AECO endeavors necessitates MVDs that are not only accurate and comprehensive but also adaptive to evolving project requirements, which direct the subjectivity while it is not guided under definite and implicit development systems. This is where our exploration takes a transformative turn, embracing the known strength of information theory and sandpile simulation to revolutionize MVD generation.

To eliminate the inconsistency and vagueness from MVD generation, we would have to develop a more unified and topological development approach that would incorporate different end user's needs with a rational and agreeable tool to shift the traditional paradigm of developing MVD. In which case, it is sandpile simulation collaborating with quantifiable information mass using information theory. With Sandpile simulation and information theory, the new approach of generating MVD could be able to evolve to keep pace with the dynamic nature of the AECO industry. Conventional MVD generation approaches often struggle to accommodate the nuances of complex projects, leading to inefficiencies, errors, and lost opportunities for optimization. Our methodology transcends the limitations of traditional methods, embracing the complex nature of construction processes to create MVDs that are agile, robust, and adaptable.

Case study

Scenario: 100 grains, drop from central and no vertex has sand on it, The $L^2 = 5 \times 5$ lattice.

The final layout of lattice is shown in Figure 19 (Section 5.4). It could be seen that there are three layers accommodate sands including root node, 8 vertices in second layer, 16 vertices in third layer. The sands accommodate in the vertex stands for the information quantity. The layout of lattice after sandpile simulation is given in Figure 27. With the mapping system proposed in Section 5.3.2, the associated tree-structure graph is given step by step. Then following the process proposed in Figure 17, based on information requirements given in Table 1 and the regulations in China National Design code of Shield Tunnel Engineering, the information mass is calculated so that the entity pair could be generated. The illustrative process takes the root node and nodes in first layer to clearly demonstrate transformation/generation process. As Figure 27 shows, the center vertex of layout represents the root node, and eight vertices surround the center vertex are the nodes of first layer under the root node. So as speak, the root node contains information mass measured with six, and the eight nodes contains, 6,1,6,1,6,1,6,1 respectively. Employing information mass to capture generated entity pair as section 3.3 regulated, these entity pairs would be: <6, ?,1> , <6,?, 6>, <6,?,1> , <6,?, 6>, <6,?,1> , <6,?, 6>, <6,?,1> , <6,?, 6>. These information entropy-based entity pairs would be translated and enriched to applicable entity through matching with tunnel terminology in specification/code of practice.

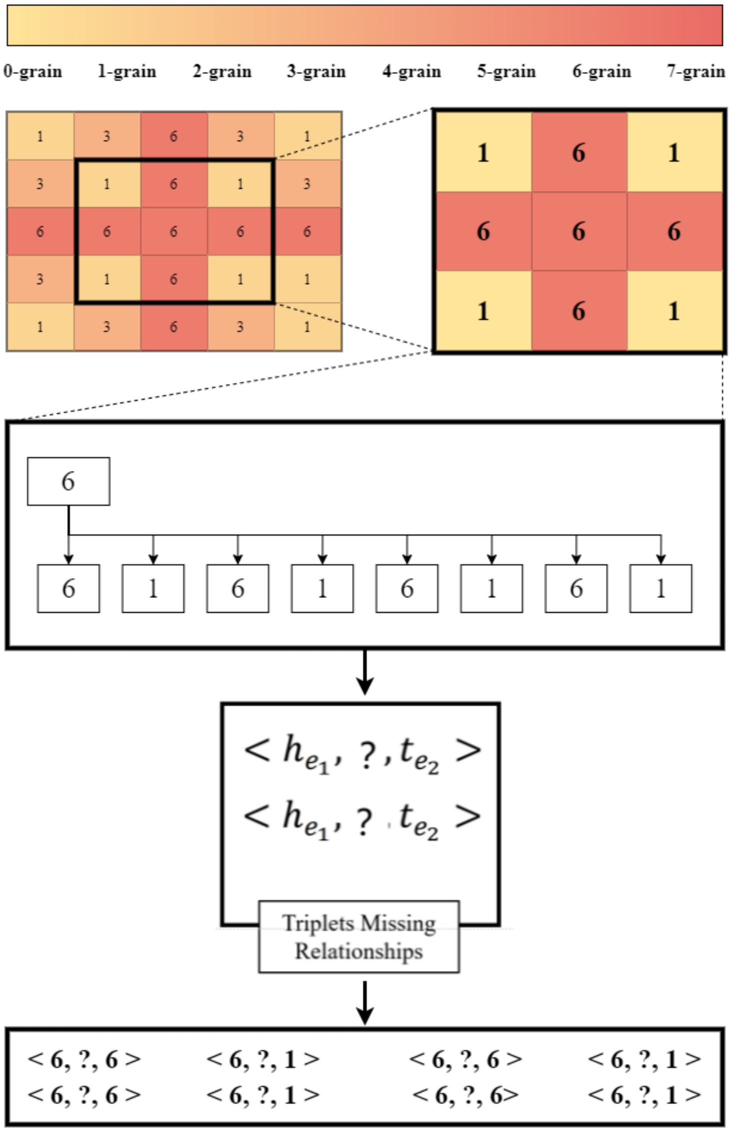


Figure 27 The layout of lattice after Sandpile Simulation and transformation to tree structure taking root node and first layer as a demonstrative example.

To develop proper entities for tunnel MVD, information entropy-based triplets need to be enriched with tunnel terminology semantics from specification/code of practice via information requirements and information mass calculation. According to calculation process described in Section 5.2, the information requirements given in Table 1, and the regulations in China National Design code of Shield Tunnel Engineering, the information mass is calculated as following.

The whole calculation starts with the nodes on first level. The root node also acts as the absolute head entity of the whole MVD (top of tree structure). As a tunnel, according to code of practice GB/T51438-2021, besides the shield segment tunnel implied here, there are many other alternatives, from the perspective of potential project type, underwater tunnel, submerged floating tunnel, road tunnel, pressure tunnel, pipeline tunnel, and such in total twelve alternatives. Hence, according to information entropy formula, the total information mass would be $\log_2 2^N = N = \log_2 12 = 3.58\ bit$; from perspective of construction method, there are 5 alternatives including TBM, Shield tunnel, Immersed tunnel, Cut and Cover tunnel, Rock Blasting tunnel, the total information mass would be $\log_2 2^N = N = \log_2 5 = 2.32\ bit$. In general, the information mass entity SegmentTunnelElement contained is $3.58 + 2.32 \approx 6bits$. As for the tail entity, For the segment, the whole tunnel needs to decide whether it is a single layered segment or double layered segment. Hence, then it is the 1 bit $=\log_2 2$. Hence, the information mass entity TunnelSegmentEnvelope contained 2 bits. In such way, the first information entropy-based triplet <6,?,1> would be enriched as <SegmentTunnelElement, ?, TunnelSegmentEnvelope >. Similarly, the head entity for the triplet <6,?,6> is till also SegmentTunnelElement, the tail entity is supposed to be a entity that contains six bits of information given the circumstances those specified in code of practice, considering the dimensions of void, and types of void, the total information mass for the entity void would be 6 bits. Ergo, this triplet could be translated to <SegmentTunnelElement,?,Void >. Can reduce from this, the translation results for eight generated information entropy-based entity pairs would be given in Table 3.

Table 3 Translating information entropy-based entity pair to actual entity pair according to specifications

Generated entity pairs	Information mass contained in head entity	Reasonable Entity contains associated information mass in Specification	Information mass contained in tail entity	Reasonable Entity contains associated information mass in Specification	Translated triplet (missing relationship)
<6,?,1>	6	`SegmentTunnelElement`	1	`TunnelSegmentEnvelope`	`<SegmentTunnelElement, ?, TunnelSegmentEnvelope >`
<6,?,6>	6	`SegmentTunnelElement`	6	`Void`	`<SegmentTunnelElement, ?, Void >`
<6,?,1>	6	`SegmentTunnelElement`	1	CorrosionControl	`<SegmentTunnelElement, ?, CorrosionControl >`
<6,?,6>	6	`SegmentTunnelElement`	6	Segment	`<SegmentTunnelElement, ?, Segment >`
<6,?,1>	6	`SegmentTunnelElement`	1	WaterProofingType	`<SegmentTunnelElement, ?, WaterProofingType >`
<6,?,6>	6	`SegmentTunnelElement`	6	RebarConstruction	`<SegmentTunnelElement, ?, RebarConstruction >`
<6,?,1>	6	`SegmentTunnelElement`	1	SeismicStructureComponent	`<SegmentTunnelElement, ?, SeismicStructureComponent >`
<6,?,6>	6	`SegmentTunnelElement`	6	SegmentBolt	`<SegmentTunnelElement, ?, SegmentBolt >`

In accordance with Section 5.4, an NLP-based technique has been utilized to extract triplets from the same specifications or codes of practice as those used to construct entity pairs. For example, within the China National Design Code of Shield Tunnel Engineering, the original description states that "The segment tunnel element can be constructed with a single-layer or double-layer design." To align with the naming conventions for IFC relationships, which are typically denoted as "RelatingEntity_RelatedEntity," a new relationship entity for the case involving "Segment Tunnel Element" and "Tunnel Segment Envelope" has been defined as:

"SegmentTunnelElement_ConstructsWith_TunnelSegmentEnvelope."

This method of deriving new relationship entities follows the procedure outlined in the NLP-based technique and is depicted in Figure 28.

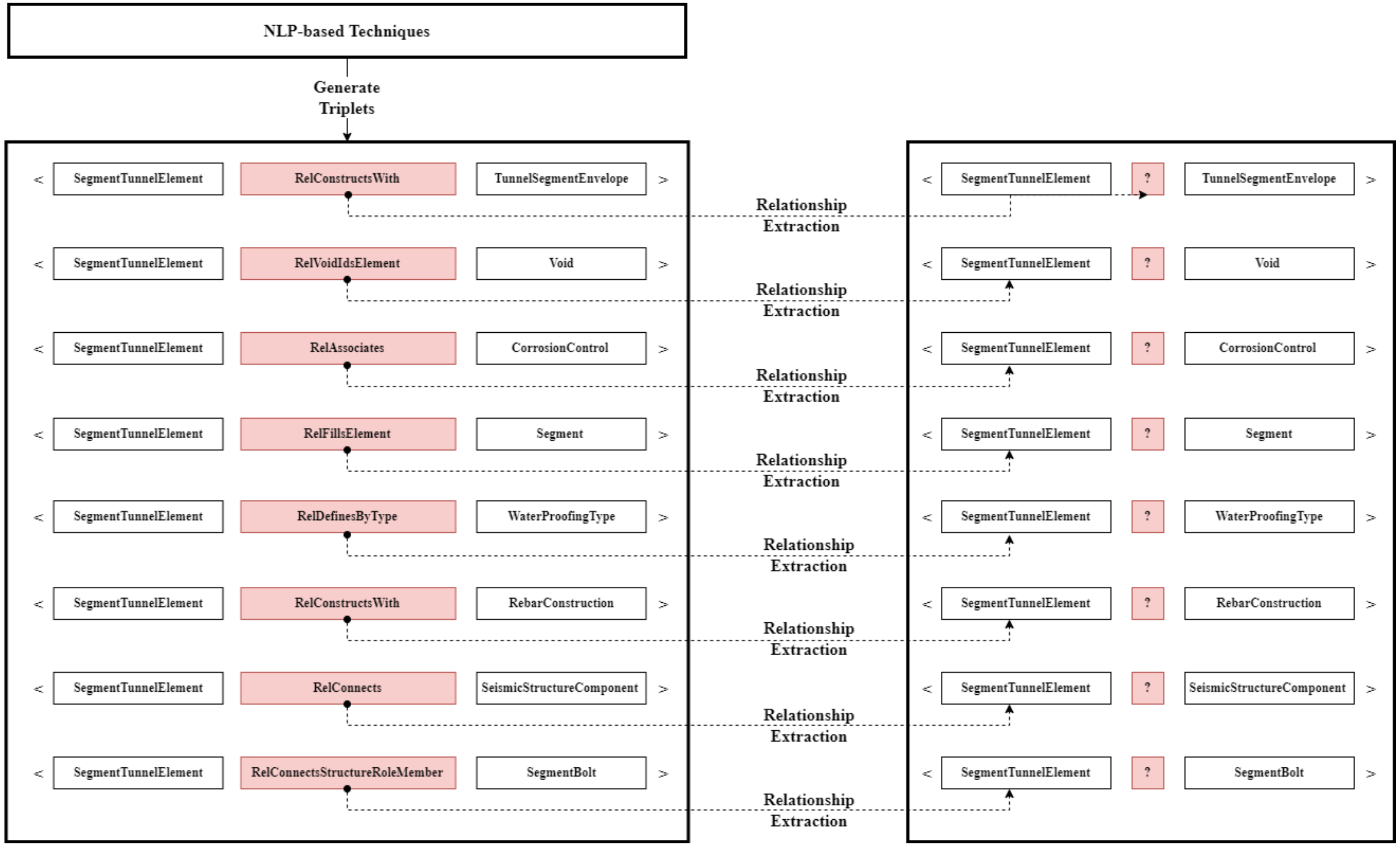


Figure 28 Completing triplet by extracting relationship entity via NLP.

The remaining calculations of information mass for the entities on the second level can be completed in the same manner. To date, we have concluded the existing calculations and converted these nodes from Figure 19 to MVD (given in Figure 29).

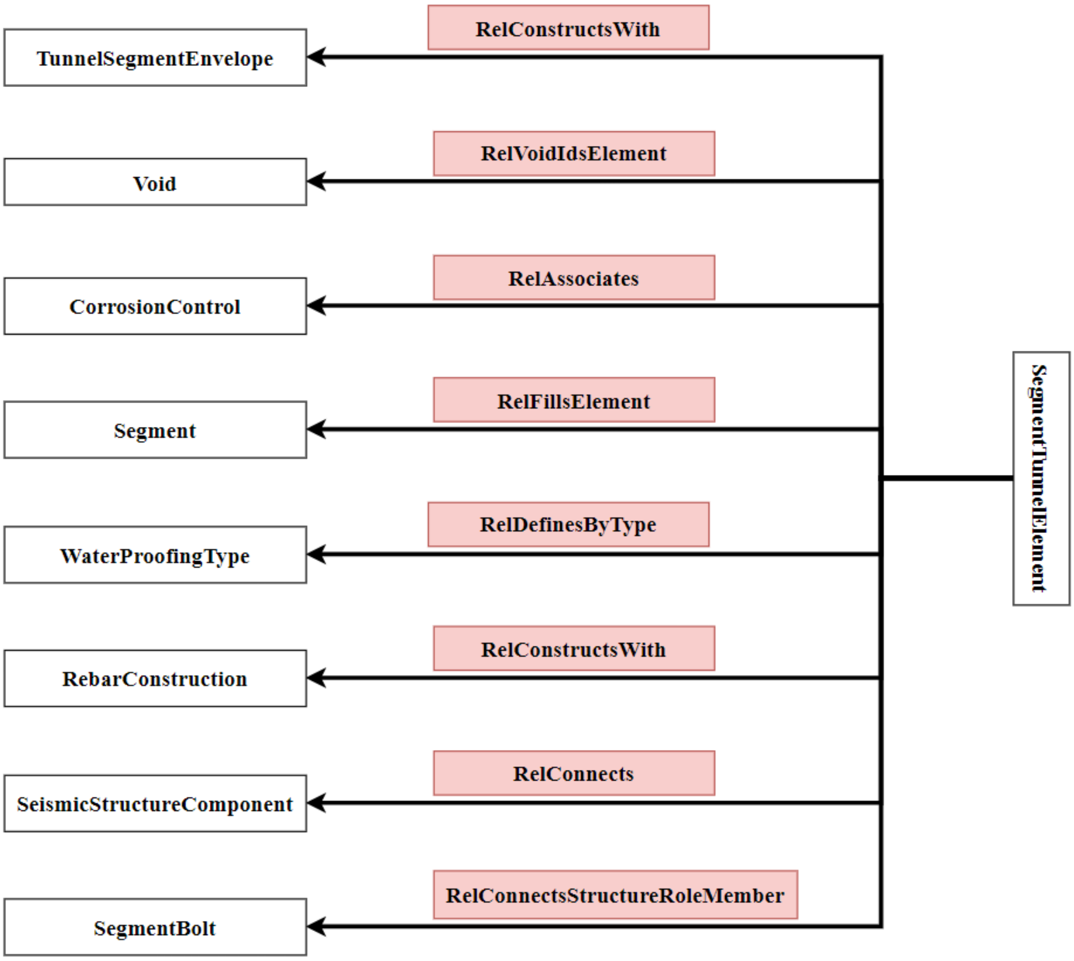


Figure 29 Converted MVD for the scenario.

Figure 29 shows the new MVD for the scenario, there are 8 entities under the segment tunnel element, secondary lining, segment, rebar construction, segment bolt, seismic structural component, void element, corrosion control, water proofing. Very interestingly, the topological structure of proposed MVD for the scenario is similar to MVD structure for shield tunnel proposed by Koch (the right half started with “Element”). Besides the topological structure, there are some entities are the same. The main reason causing the differences are the semantics, after all, Koch’s research is based on the different codes to design the shield tunnel, the semantics are naturally different. Unlike Li’s MVD, the semantics of entities are mostly the same since Li's group obviously followed the same code while designing the shield tunnel. The major difference to Li’s MVD is that the topological structure of the MVD is completely different, the main reason causing this would be the scale of the use case scenario, Li’s scenario is a large-scale project, and the scenario to make illustration here is a relatively smaller one which contains much less information quantity than Li’s. The findings about resemblances between Koch’s and proposed MVD structure as well as between Li’s and proposed MVD structure implies that the MVD development in BIM can be properly integrated with the tree-structured graph model and sandpile simulation for generating a general MVD for making it uniformly standard. In other words, in a way, it provided the evidence that the proposed methods for MVD development is producing a general solution, and the Koch’s and Li’s MVD are special solution under certain circumstances. With the sandpile simulation, information entropy, code for design, and the proposed mapping system, it establishes ways of creating and analyzing MVD, providing the decision-support basis for MVD development.

It can be seen that the methodology is firmly rooted in information theory and sandpile simulation. Our research harnesses the power of information theory, a technique that capture the information mass so that we would be able to analyze the intricate interactions within construction projects. The generated data forms the bedrock upon which our innovative approach to MVD generation is built. The findings from our investigation provide substantial advancements in our understanding, representing a significant breakthrough in the field. By harnessing the principles of chaos theory, we generate MVDs that reflect the true dynamism of AECO projects. These MVDs are not static blueprints; they are living documents that adapt to project requirements in real-time.

One of the key innovations stemming from our approach is the adaptive nature of MVDs. Traditional MVDs often fail to account for the quantifiable and unifiable nature of standard to comply project. In contrast, our MVDs are robust and essential in terms of information management of the project which align with digitalization purpose of industry. This adaptability reduces errors, enhances collaboration, and streamlines project workflows. The practical implications of our MVD generation approach are profound. It aligns BIM data standards with the realities of modern construction projects, fostering greater efficiency, flexibility, and resilience. Stakeholders across the AECO spectrum stand to benefit from MVDs that truly reflect the complexities of their projects. Architects and engineers gain access to MVDs that adapt seamlessly to evolving design requirements. Contractors benefit from MVDs that anticipate and mitigate construction challenges. Facility managers receive MVDs that enhance the management and maintenance of built assets. Ultimately, our approach empowers the entire AECO industry to operate with greater precision and effectiveness by embracing information theory and sandpile simulation.

As we embark on this exploration of advanced engineering applications in complexity for BIM, we invite you to delve deeper into the transformative world of MVD generation. Our journey will take us through the intricate landscape of BIM data standards, revealing how chaos theory can reshape the very foundations of the AECO industry. Together, we will witness the evolution of MVDs from static blueprints to dynamic tools that drive innovation and excellence in practical projects.

7.3 Application 2—Chaos-based BIM Data Standard Evaluation

The contemporary Architecture, Engineering, Construction, and Operations (AECO) industry operates within an ever-evolving digital landscape. In this dynamic arena, the efficacy of Building Information Modeling (BIM) data standards holds immense significance. These standards, like the silent conductors of a digital orchestra, harmonize diverse stakeholders, technologies, and processes, thereby orchestrating the symphony of modern construction projects. However, as projects grow in complexity and scale, the ability to evaluate and adapt these standards becomes increasingly pivotal.

The early concerns about the BIM standard testing focus on the interoperability of Industry Foundation Class (IFC), the certification approaches and novel ideas are mirrored from the healthcare and manufacturing domains for the AEC industry, in the paper, Amor (2008) suggested to establish the third party which is independent of the IAI to accomplish testing and certification, further he promoted the development an open tool for file checking. With the evolvement of certification, the complete Model View Definition is documented as a contribution to define test cases with precise exchange requirements and instructions for software developers to optimize their application, the exported files are checked based on 3 categories of tests: IFC schema syntax and where rules, rules from implementer's agreement, and limited numbers of simple semantic checks based on the MVD (Chipman, 2013).

Followed the idea of validating the IFC model, fellow scholars started to put more focus on well-formedness/syntactic correctness of data in an IFC model. They select a few data quality dimensions those fit IFC quality criteria to adopt, these dimensions mainly concern intrinsic (complete, meaningful, correct) and representational (conformance, ambiguous). And based on these dimensions, the rules for measuring the quality of IFC data was proposed (Solihin et al., 2015; Strong et al., 1997; Wand & Wang, 1996). Thus far, there have been several successful attempts on evolution process of BIM standard testing methodologies, including a) certification reduced plenty of issues and notably improve the quality of the exchanged models; b) manual effort is vastly reduced with the automated rule checking and the correctness of model data is highly assured; c) applicability to the actual project is enhanced. Though, there are a few research discussing other aspects such as expressivity, ease of implementation. The measurement is not fully well-defined and quantifiable. The lack of quantifiability hinders the capability of generalization. Moreover, current validation in the use case scenario relies on the applicability, and the validating cases are primarily with limited data. The first level of comfort for the end users is reducing issues undoubtedly. Yet, great applicability cannot equal to the great comfort as performance of BIM model should have considered the transmission and visualization efficiency while large data involved. The large data cases are often seen in practice. In stage 3（Figure 2）, working with huge BIM model on various BIM software struggles in terms of visualization/transmission and storage/maintenance, and the current testing/assessment methods have not dabbled those two aspects. Ergo, it is lack of comprehension in certain way.

BIM data standards serve as the digital blueprint of the AECO industry, shaping the way information flows, collaborations occur, and projects evolve. The significance of evaluating these standards cannot be overstated. In an era characterized by rapid technological advancements, the role of BIM data standards transcends conventional documentation; they represent the critical infrastructure of the digital construction landscape. They ensure that a design in an architect's studio seamlessly transforms into a tangible structure, guided by engineers, constructed by contractors, and maintained by facility managers. The complexity of modern construction projects requires BIM data standards to adapt continually. They must cater to an ever-expanding universe of variables: from architectural intricacies to structural engineering nuances and from supply chain management challenges to facility maintenance demands. Ensuring that these standards remain agile and responsive to the evolving needs of the industry is not a mere academic pursuit; it's a fundamental requirement to maintain the efficiency, precision, and sustainability of construction projects.

Traditionally, BIM data standard evaluation has been a formidable task. Conventional methodologies, often rooted in static compliance checks and manual assessments, struggle to keep pace with the dynamic, chaotic nature of modern construction projects. The sheer intricacies of today's structures, coupled with the manifold challenges in project execution, necessitate a paradigm shift in how we perceive, assess, and refine these standards. It's here that chaos theory emerges as a beacon of hope.

Put in a nutshell, it is promising to further develop a fractal thinking based BIM standard assessment method that is quantifiable and outperforms existing BIM standard test methods in terms of efficiency, generalization, and robustness. Since the fractal thinking based BIM standard assessment method put attention on the topology of the information standard rather than solely IFC’s data structure, hence it could be applied to other BIM standards those involved in stage 1 (Figure 2). Thereby, it improves the robustness. And with the development of fractal theory in the last 60 years, there are many mature theories and tools to study it, such as attractors, fractal dimension and etc. Thereupon, the proposed fractal thinking based BIM standard assessment could be more quantifiable and better defined. In addition, the fractal thinking was set from a more systematic perspective. This chapter embarks on a captivating exploration of how chaos theory, an interdisciplinary scientific principle, can revolutionize the assessment and refinement of these vital BIM data standards.

Key Performance Indicator Definition

As it has been proven in section 6, the dynamical system (information standard development over information mass) contains chaos. Hence, the properties of the chaos in such dissipative system would be embedded in the system. In chaotic system, there would be a strange attractor that would attract the phase point (state) moving towards attractor. Since the system is a dissipative, the attraction would require the external source providing energy exchange. On basis of this property, this section would propose an indicator that could assess the overall performance of information standard development in different information mass scenario. The principle of assessment would be that information standard development is supposed to accommodate the given information mass in the best way. Here, the "best" refers to the attractor of the system. Further the given information standard is away from supposed attractor, more energy would consume for the attraction movement. Ergo, the distance between the phase point derived from given information standard and supposed attractor generated by system given information mass would be defined as an indicator. Value of Indicator would equal to Euclidean distance, greater distance (indicator value) indicates the overall performance of information standard is poorer in terms of accommodating given information mass.

Consider a set, Ψ, the elements are the phase points of system described in section 6.2, one dimensional component of time series generated by each step of sandpile simulation, at time t, the component would be $s(t)$, $s(t) \in \Psi$. And a subset $A \subseteq \Psi$. Let G be a collection of subsets of Ψ. Provided Following conditions hold:

i. $\emptyset \in G$,

ii. If $A \in G$ then $A^C \in G$,

iii. If $A_1, A_2, \dots \in G$ then $\bigcup_{i=1}^{\infty} A_i \in G$

Then G is a σ – algebra defined on Ψ. Then the pair consisting of a set and a σ – algebra defined on that set, (Ψ, G), is referred to as a measurable space. Then we define measure on (Ψ, G).

Given a measurable space (Ψ, G), a measure on (Ψ, G) is a function, $anm : G \to \mathbb{R}^+$, such that,

i. $anm(A) \geq 0$ for all $A \in G$,

ii. $anm(\emptyset) = 0$,

iii. If $A_1, A_2, \dots \in G$ are disjoint then $anm(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} anm(A_i)$.

Noticeably, the subset A here is defined as

$$A = \{A_i | the\ phase\ point\ representation\ given\ information\ standard\ under\ embedded\ dimension\ m\ and\ time\ delay\ \tau\ predetermined\ by\ provided\ information\ mass\ \} = \{A_i |\ A_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})\} \qquad \text{equation (20)}$$

And the defined indicator function anm to assess information standard development against given information mass for a set A is where

$$anm\left(A_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})\right) = \begin{cases} d = \sqrt{(x_{a1} - x_i)^2 + (x_{a2} - x_{i+\tau})^2 + \cdots + (x_{am} - x_{i+(m-1)\tau})^2} & for\ d > r_a \\ d = 0 & for\ d \leq r_a \end{cases}$$

equation (21)

Where $(x_{a1}, x_{a2}, \dots x_{am})$ is the center of attractor, and r_a is radius of attractor. Since the dynamical system has been proven chaotic, there would be an attractor of the system that could be located.

Case Study

With defined indicator, the performance of information model associate with instantiated information standard could be assessed. The process diagram of evaluation is shown in Figure 30. The process is initiated by stakeholders, they would provide project information and code of practice. With the code of practice and information entropy from Shannon's information theory, following the process of An's proposal (An, Lin, et al., 2023a), information mass could be calculated. This could be used to conduct sandpile simulation to generate the time series sample as Section 6.2 demonstrated. Afterwards, reconstruct phase space of generated data samples. Locate the attractor in reconstructed phase space. This is the basis of the assessment which acts as the reference origin. Then the existing information standards (developed and promoted by various parties including software vendors, standard development institutions and organizations) those would be used for instantiation of information modelling will be converted into tree structured graph with Yi's mapping system (Yi et al, 2022). Followingly, the tree structured graph would be converted to a sample of time series data with approach proposed in Section 6.2 and then a phase point in reconstructed phase space with approach proposed in Section 6.3. On basis of these, use the measure defined previously to obtain the indicator of the standard assessment result. Theoretically, larger the indicator shows, poorer performance the information model would have in a sense.

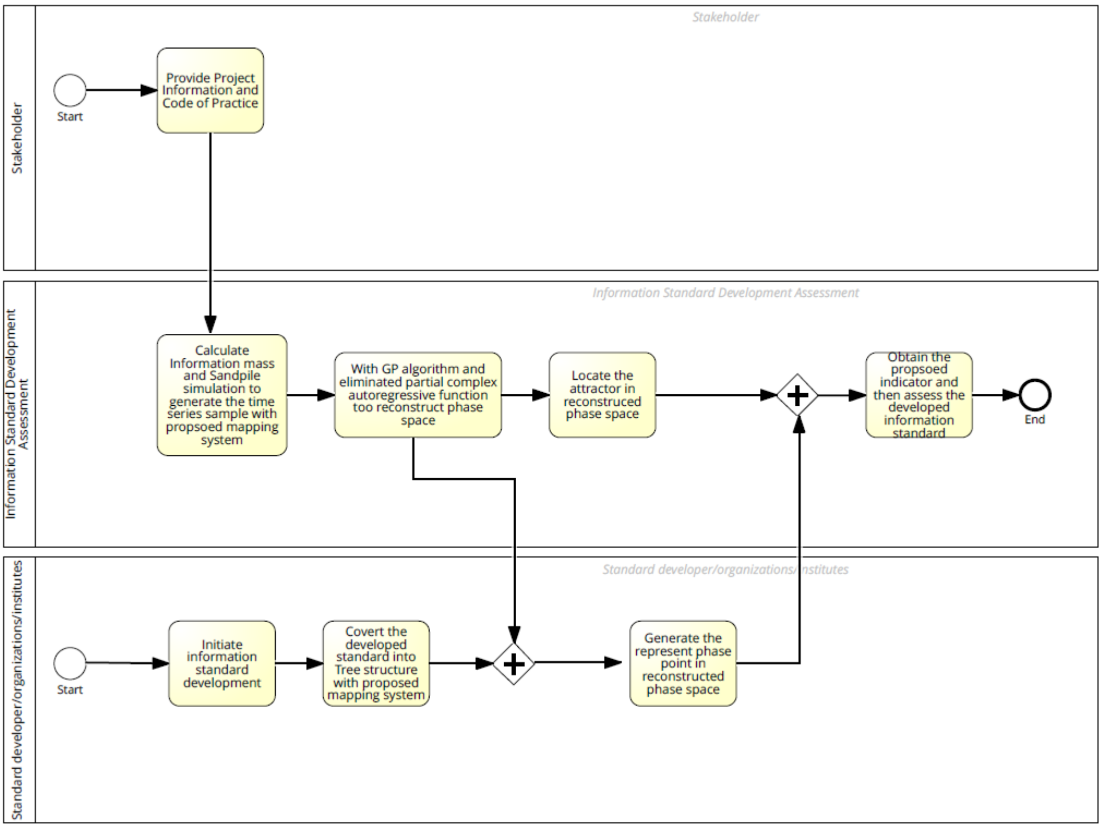


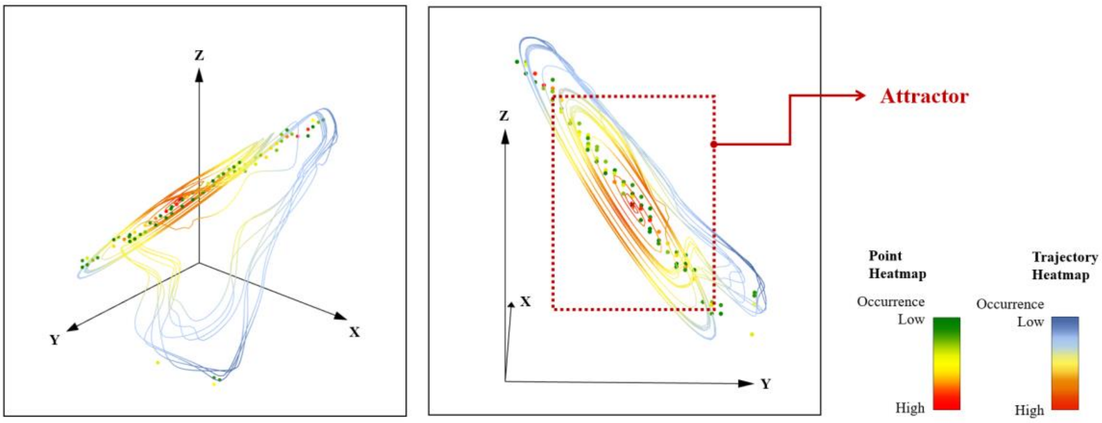
Figure 30 Process diagram of evaluation process

To verify the defined measure, a case study would be conducted here to check if measure indicated results would align with actual information model performance. The case is used as the same as previous series of research (An, Lin, et al., 2023a). Initial Settings is given as following:

a tunnel lining element is given based on information requirement analysis referring “Guidelines for the Design of Shield Tunnel Lining” in Section 5.2 (Working Group No.2, 2000). The basic design information is regulated by “China National Design code of Shield Tunnel Engineering-- GB/T51438-2021” (Development, 2021)and listed as follows: 1. single layer shield segment (code) ; 2. material is reinforced concrete; 3. assemble method is bolted; 4. 8 segments consist a ring; 5. segment form type is flat, need 8 botts; 6. Ring’s width 900mm; 7. capped joint form is radial wedge; 8. assemble type is continuous joint A; 9. segment joint: circumstantial; 10. ring rotation to attach last ring: 20 degree (range 10-40, considering integer, it would be 31 alternatives); 11. stiffener: material consistent with segment material; 12. cushion pad: exist, material: rubber; 13 Grouting/lifting hole: inner diameter 50mm; 14. Stagger joint; 15. bolts : material, dimension, construction pretention force; 16: segment dimension; 17. water proof gasket (doubled, one in inner side, another in outer side), material, dimension. According to the design description of an assumed simplified tunnel lining, the number of information entries are calculated. Noticeably, item 1 contains 2 information entries, single layer shield segment and its code as URI; item 2 contains 3 information entries, segment material type, concrete grade, steel grade; since in item 4, 8 segments consist a ring, there are A_8^8 alternatives for the layout, hence there are A_8^8 information entries; item 5 contains 8 information entries; item 10 contains 31 information entries; item 12 contains 2 information entries; item 15 contains 3 information entries; item 17 contains 4 information entries. All other items contain only 1 information entry each. Therefore, the total information entries contained in the assumed tunnel lining is 40372, making the information quantity $\log_2 40372 = 15.30$, which rounds up to 16 bits. To make the demonstration more comprehensive, ensure a rigorous statistical sense of simulation and realistically reflect the trend of MVD development processes, “bit” is used to represent the information mass, and the two-dimensional lattice are set as $L^2 = 11 \times 11$, with the initial starting layout as no vertex contains sand. Followingly, the ticks (sands dropping) are total grains of sands. Since the case is a simplified version based on a single layer shield tunnel as underpass, the whole line is 0.4 kilometers. All the tunnels are underground, considering the width of a ring is roughly 1.2 meters, the first underpass would have 334 rings and each ring contains 16 bits of information. The total information mass is approximately 5344 bits. While dropping grains of sand, one at a time, sand grains can fall off the edge of the table, helping ensure the avalanche eventually ends. Considerably, there are limited vertices, 11×11. Even all the vertices reach their full capacity, it would be only 800 grains. The greater number of grains of sand falling off the edge of the table would lack value to the study. Thusly, the grains need to be limited. Constructing isomorphism can be of great assistance. With isomorphism, the grains of sand can easily be converted to 334 (5344/16) grains of sand.

Then the 2 verified tunnel lining standards are Koch's and Li's. These 2 tunnel lining standards are coded as XML schema then instance it as XML file with project information given above. The redundancy (size) of instantiated file could be considered as a vital indicator of XML schema, in other words, smaller file size is, more efficiency XML schema would be in the case of given project.

Reconstruction of phase space for given project: as the general information mass is 334. Hence, go through the section 4, it could be known that the reconstructed phase space is a space with time delay τ =42, embedded dimension $m = 3$. Since the standard of both Koch's and Li's are static, the phase point of their standard under such time delay and embedded dimension would be $(D_{Koch}, D_{Koch}, D_{Koch})$ and (D_{Li}, D_{Li}, D_{Li}). To create an attractor landscape map from time series data, the Attractor Landscapes method is used. First, divide the phase space into an $n \times m$ grid $(30 \times 30\ grid\ has\ worked\ well)$. Next, compute how many data points are present in each grid box. Finally, expand the two-dimensional phase space into three dimensions, with the third dimension being a count of how frequently data falls into a particular grid space (Michaels, 2013). And for the given information mass and reconstructed phase space, the attractor landscape is located through python, given the attractor visualized as follow in Figure 31. Figure 31a shows the trajectory heatmap and phase point heatmap in reconstructed space (code given in Appendix F). The occurrence level is ranked from low with cool color to high with warm color. And attractor, by its nature, is supposed to be a state which dynamic system tends to evolve. Hence, the high occurrence points and traits would profile the contour of attractor. The clearer view of attractor is given in Figure 31b.



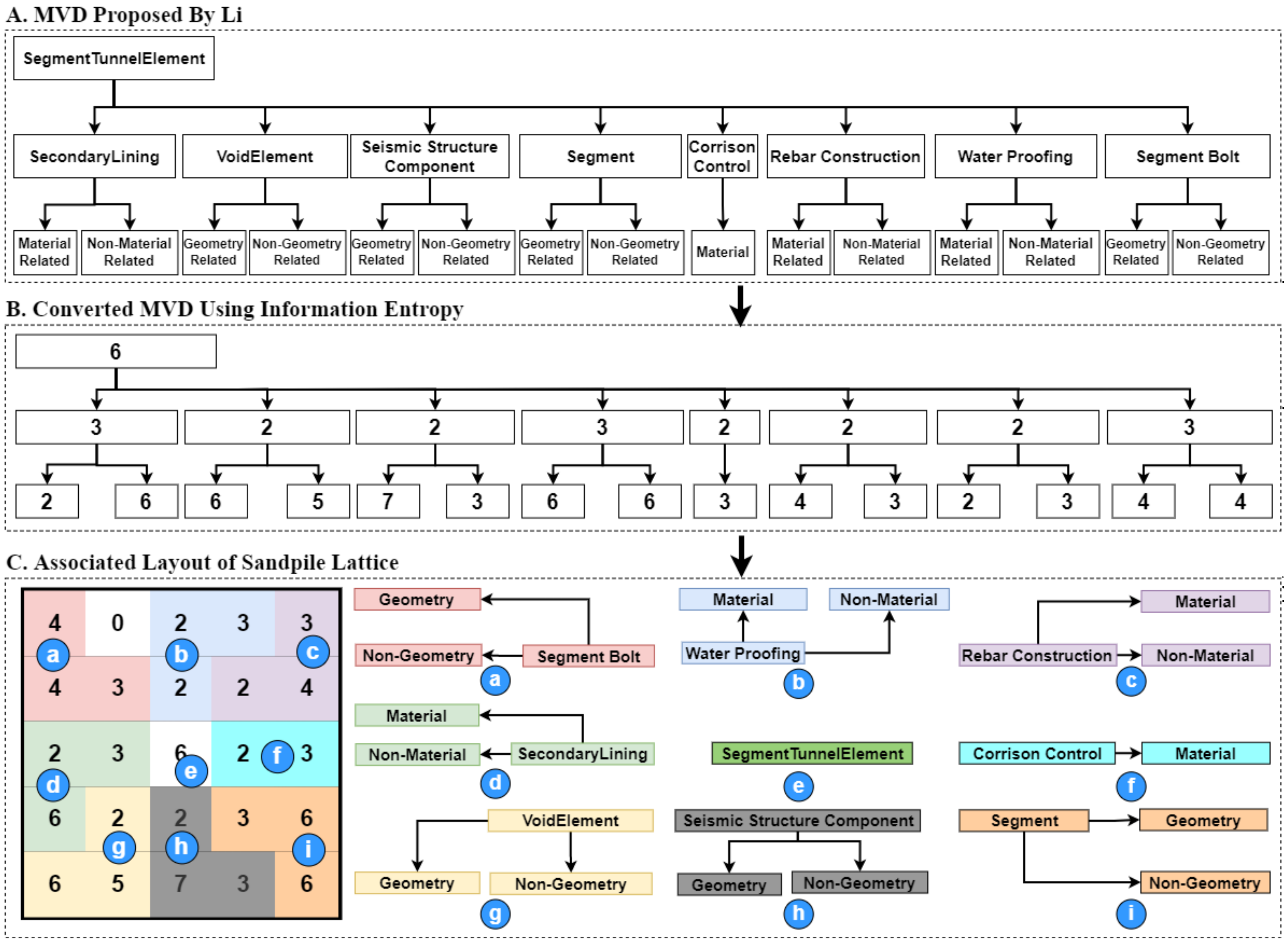
a

b

Figure 31 a. Phase point heat map and Trajectory heatmap in reconstructed phase space; b. clear attractor

Converting the developed MVD into tree structure starts with calculating the information mass that associated entities containing. The calculation is based on information requirements along with given code of practice. This study is in accordance with "Guidelines for the Design of Shield Tunnel Lining" (Working Group No.2, 2000), and the basic design information is regulated by "China National Design code of Shield Tunnel Engineering-- GB/T51438-2021" (Development, 2021). For example, the first entity in Li's MVD is SegmentTunnelElement. According to code of practice GB/T51438-2021, besides the shield segment tunnel implied here, there are many other alternatives, from the perspective of potential project type, underwater tunnel, submerged floating tunnel, road tunnel, pressure tunnel, pipeline tunnel, and such in total 12 alternatives. Hence, according to information entropy formula, the total information mass would be $\log_2 2^N = N = \log_2 12 = 3.58\,bit$; from perspective of construction method, there are 5 alternatives including TBM, Shield tunnel, Immersed tunnel, Cut and Cover tunnel, Rock Blasting tunnel, the total information mass would be $\log_2 2^N = N = \log_2 5 = 2.32\,bit$. In general, the information mass entity SegmentTunnelElement contained is $3.58 + 2.32 \approx 6bits$. Accordingly, while converting MVD proposed by Li (Figure 32a) to tree structure (Figure 32b), the first node is represented with a node numbered 6, which stands for such entity would contain 6 bits of information. Following such method, the SecondaryLining entity contains 3 bits of information, VoidElement contains 2 bits of information, etc. The detailed conversion mapping can be seen in Figure 32a and Figure 32b. With mapping system developed by Yi (An, Lin, et al., 2023a), the tree structure would be converted to the toppling lattice with sands in vertex, given in Figure 32c.

Similarly, Koch's MVD is firstly converted to a tree structured graph using information entropy, then transformed to sandpile lattice. The detailed diagram is given in Figure 33.


A. MVD Proposed By Li

B. Converted MVD Using Information Entropy

C. Associated Layout of Sandpile Lattice

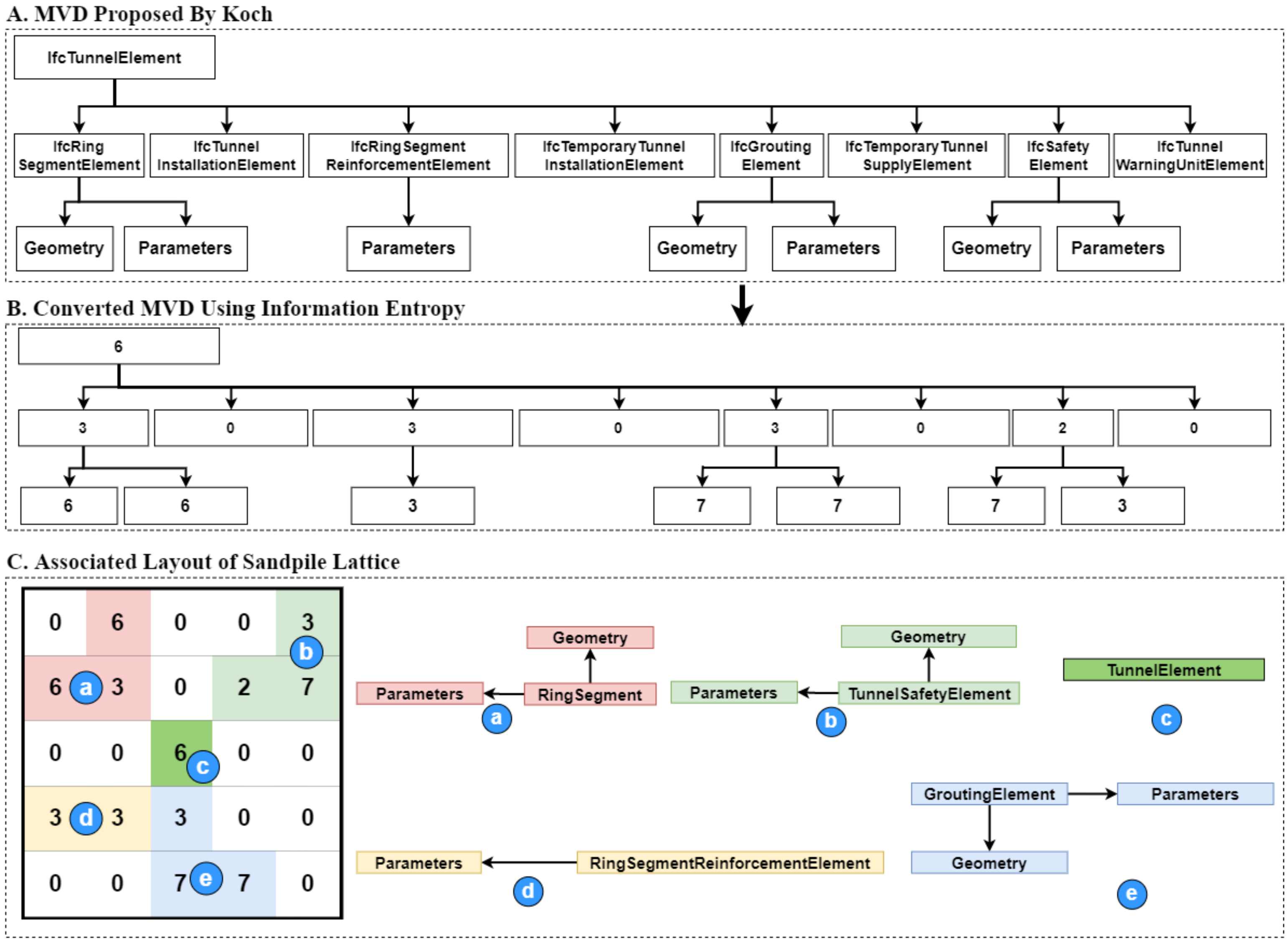

A. MVD Proposed By Koch
IfcTunnelElement
IfcRing SegmentElement
IfcTunnel InstallationElement
IfcRingSegment ReinforcementElement
IfcTemporaryTunnel InstallationElement
IfcGrouting Element
IfcTemporaryTunnel SupplyElement
IfcSafety Element
IfcTunnel WarningUnitElement
Geometry
Parameters
Parameters
Geometry
Parameters
Geometry
Parameters
B. Converted MVD Using Information Entropy
6
3
0
3
0
3
0
2
0
6
6
3
7
7
7
3
C. Associated Layout of Sandpile Lattice
0 6 0 0 3
6 3 0 2 7
0 0 6 0 0
3 3 3 0 0
0 0 7 7 0
a
b
c
d
e
Geometry
Parameters
RingSegment
Geometry
Parameters
TunnelSafetyElement
TunnelElement
GroutingElement
Parameters
Geometry
Parameters
RingSegmentReinforcementElement

Figure 33 Covert developed MVD by Koch to tree structure and generated associated phase point. a. MVD proposed by Li; b. converted MVD schema using information entropy; c. associated layout of Sandpile Lattice

The weights' set of nodes remain consistent with attractor generation process. In the study, the weights of node set are $\{w_1, w_2, w_3\} = \{11,10,9\}$. With node fission entropy defined in the study (Section 6.2), the partial MVD of Li's is selected to demonstrate calculation process given in Figure 34. Then the developed MVD could transformed into a time series data. Hence, go through the chapter 6, it could be known that the reconstructed phase space is a space with time delay τ =42, embedded dimension $m = 3$. Since the standard of both Koch's and Li's are static, the phase point of their standard under such time delay and embedded dimension would be $(D_{Koch}, D_{Koch}, D_{Koch})$ and (D_{Li}, D_{Li}, D_{Li}). In the case, the calculation shows $D_{Li} = 841$, hence associated phase point is $(841, 841,841)$, and $D_{Koch} = 497$, associated phase point is $(497, 497,497)$ (code given in Appendix G）.

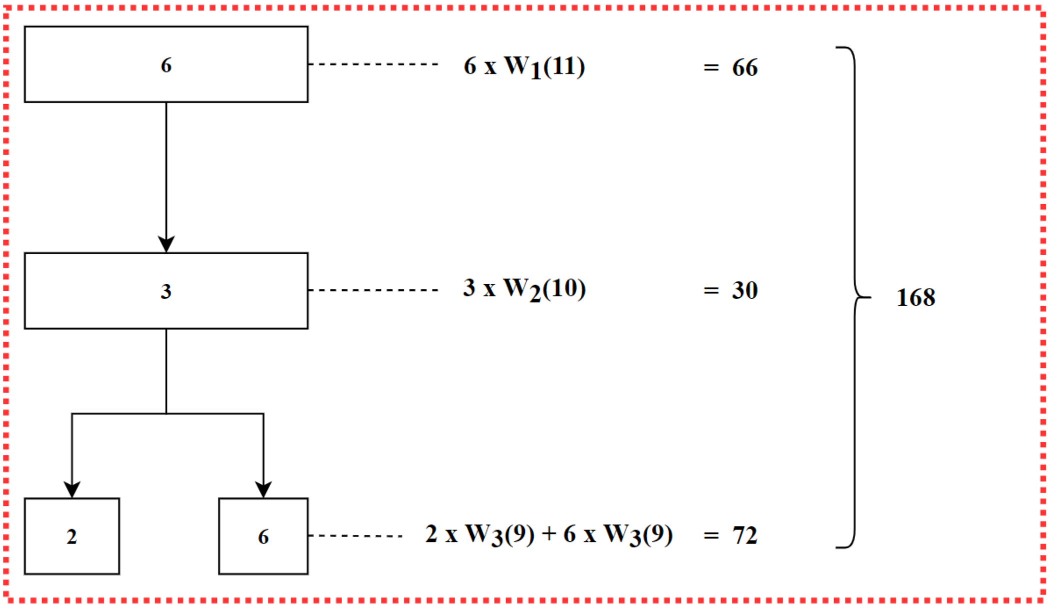


Figure 34 Demonstrate calculation process of proposed node fission entropy in partial Li's MVD

Plotting a phase point in a 3-dimensional phase space involves representing the point using same three coordinates that correspond to the x, y, and z-axes as the one that locating attractor used. The point (497, 497, 497) can be plotted by measuring the distance of the point from each of the axes along the corresponding coordinate axis. A Cartesian coordinate system was used for locating attractor, which is a three-dimensional grid with three perpendicular axes intersecting at the origin. For better visualization, the y-axis runs horizontally, the x-axis runs vertically, and the z-axis runs perpendicular to the x and y axes. To plot the point (497, 497, 497) in this coordinate system, we first locate the x-axis and measure 497 units from the origin in the positive x direction. Next, we locate the y-axis and measure 497 units from the origin in the positive y direction. Finally, we locate the z-axis and measure 497 units from the origin in the positive z direction. The point is then located at the intersection of the three axes where the three measurements meet. The resulting plot will show the point (497, 497, 497) as a single dot in phase space (Figure 35), with its position relative to the origin and the orientation of the axes providing information about its location in three-dimensional phase space. Analogously, the phase point reflecting Li's MVD $(841, 841, 841)$, the resulting plot will show the point as a single dot in space (Figure 35). With $anm\left(A_i = \left(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\right)\right)$ measure function—defined indicator, the domain set in this case is $\{A_i | \, A_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})\}$={(841,841,841), (497,497,497)}, then referring $anm\left(A_i = \left(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\right)\right) = d = \sqrt{(x_{a1} - x_i)^2 + (x_{a2} - x_{i+\tau})^2 + \cdots + (x_{am} - x_{i+(m-1)\tau})^2}$ □ d $=0$

$for\ d \leq r_a$ $for\ d > r_a$ equation (21), $anm(A_i = (841{,}841{,}841)) = 1221$, $anm(A_i = (497{,}497{,}497)) = 1798$. Since, the proposed measure, in nature, is a type of Euclidean distance but in reconstructed phase space. It could be visualized as Figure 35. Besides, theoretically, larger the indicator shows, poorer performance the information model would have in a sense. Consequently, in theory, Koch's MVD is supposed to show poorer performance (1798) than Li's MVD in the given case.

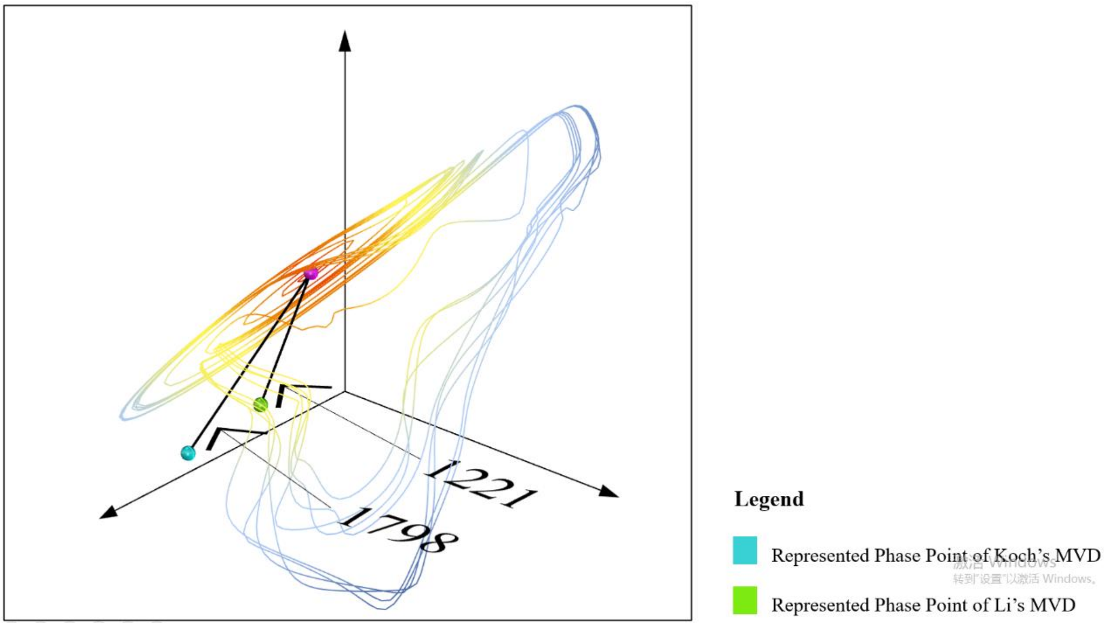


Figure 35 Visualized attractor and representation of proposed measure for phase point of Koch's MVD and Li' MVD in phase space.

The efficiency of defining XML element with properties or sub-elements depends on the use case and the size of the data. There are 4 major criteria for efficiency evaluation of XML element defining, file size, parsing time, memory usage and maintenance (code provided in Appendix H) (Fan, Ma, Wang, & Liu, 2018; Lawrence, 2004).

File size: File size of XML document is used to compare the efficiency. The smaller the file size, the faster it can be transmitted and processed. The smaller the file size, the faster it can be transmitted and processed. This could be directly observed via instantiated file.

Parsing time: the time it takes to parse an XML document can also be a factor in determining its efficiency. XML parsing tools is used to measure the time it takes to parse each document and compare the results. The study conducts XML file parsing with ElementTree module in Python 3.9.

Memory usage: When processing an XML document, it is vital to consider the memory usage. The document with the lower memory usage is generally considered more efficient. This is documented with memory profiler in Python 3.9.

Maintenance: the ease of maintaining the XML document structure is also a factor to consider. A simpler structure is generally easier to maintain, making it more efficient in the long run. And one of main test for maintenance is XML Load Testing. The essence of load test in the case is to create the same amount of instances/sub elements in both Li's and Koch's instantiated MVD using XML. Then compare the running time of operations. It is programmed with Python 3.9. And all efficiency comparison results are given in Table 4.

Table 4 Actual efficiency comparison

	Parsing Time	Memory Usage	Load Testing Running Time	Load Testing Memory Usage	File Size before Load Testing	File Size after Load Testing
Instantiated Li's MVD using XML	0.35106 s	82.9 MB	1.17636 s	103.5 MB	9551KB	11974KB

Instantiated Koch’s MVD using XML	0.38953 s	84.2 MB	1.18066 s	102.9 MB	10761KB	13538KB

It can be clearly seen from the table, when it comes to practice, in general, it is clear that Instantiated Li’s MVD using XML outperforms Instantiated Koch’s MVD using XML in every aspect. First, in Parsing Time, Instantiated Li’s MVD using XML proves to be excellent as its processing speed is faster than Instantiated Koch’s MVD using XML, which means that it can reduce data processing time and improve data processing efficiency. Second, in Memory Usage, Instantiated Li’s MVD using XML uses less memory than Instantiated Koch’s MVD using XML, which can improve system performance and stability. In Load Testing Running Time, Instantiated Li’s MVD using XML performs better than Instantiated Koch’s MVD using XML, indicating that it can respond to loads faster and improve performance. Although Instantiated Li’s MVD using XML has higher Load Testing Memory Usage than Instantiated Koch’s MVD using XML, the difference is not significant, and Instantiated Li’s MVD using XML still shows better performance. In File Size before Load Testing, Instantiated Li’s MVD using XML's size is much smaller than Instantiated Koch’s MVD using XML, indicating faster data processing and shorter file transfer time. Lastly, although Instantiated Li’s MVD using XML is slightly smaller than Instantiated Koch’s MVD using XML in File Size after Load Testing, it still performs well. In conclusion, Instantiated Li’s MVD using XML is a superior choice, which performs better in terms of speed, performance, stability, and resource utilization. It can significantly improve both data processing efficiency and the overall system's running speed. These is quite consistent with the theoretical evaluation using proposed indicator.

7.4 Conclusion and Remarks

This chapter has been a profound journey into the nexus of chaos theory, information theory, and Building Information Modeling (BIM). This chapter has not only illuminated the transformative potential of chaos theory in advancing BIM but has also showcased its practical applications in the AECO sector. The chapter began with the audacious proposition that chaos theory, often relegated to abstract mathematical discussions, can be pragmatically applied to drive advancements in BIM. This was exemplified in the exploration of Model View Definitions (MVD) generation using sandpile simulation and information theory. Traditional methods, while foundational, sometimes miss the intricate nuances of modern construction projects. The proposed methodology, deeply rooted in chaos and information theory, offers a dynamic, adaptive, and responsive approach, ensuring BIM data standards are in sync with the evolving demands of the AECO sector. The adaptability of the proposed MVD generation method stands out as a significant advancement. In the ever-shifting landscape of the AECO industry, having a methodology that can evolve in tandem with project requirements is invaluable. This adaptability not only minimizes errors but also promotes enhanced collaboration and optimizes workflows.

The chapter's case study, involving a 5x5 lattice and sandpile simulation, serves as a tangible manifestation of how theoretical constructs can be translated into actionable insights. The transformation of information into a tree structure, leveraging information theory, exemplifies the potential of this interdisciplinary approach. A pivotal advancement presented in this chapter is the concept of defining an attractor to evaluate BIM data standard performance. Attractors, in the realm of chaos theory, represent stable states that a system tends to evolve towards. By identifying and characterizing these attractors for BIM data standards, we can gain profound insights into the system's behavior, stability, and performance. The application of attractors in evaluating BIM data standard performance is groundbreaking. It offers a novel lens through which we can assess the efficacy, resilience, and adaptability of BIM data standards. By understanding where the system's attractor lies, we can predict its future behavior, identify potential pitfalls, and make informed decisions to steer the system towards desired outcomes. Furthermore, the concept of attractors brings a level of quantifiability to the evaluation process. Instead of relying solely on qualitative assessments, attractors provide a metric that can be measured, analyzed, and compared, enhancing the rigor and objectivity of the evaluation process.

The methodologies and insights presented in this chapter, while revolutionary, are just the beginning. As the AECO industry continues its march towards smart infrastructure, the role of optimized data standards will become paramount. The insights from this chapter provide a robust foundation but also underscore the need for continuous research, refinement, and adaptation. Chaos theory, at its core, is about understanding the unpredictable and seemingly random behaviors of complex systems. In the context of BIM, this unpredictability is not just a theoretical construct but a daily reality. Projects often face unforeseen challenges, from design alterations to construction delays, all of which can introduce chaos into the system. By applying chaos theory, we can better anticipate these challenges, understand their underlying causes, and develop strategies to mitigate their impact. The concept of using attractors to evaluate BIM data standard performance has far-reaching implications. Attractors, as stable states that systems gravitate towards, provide a benchmark against which we can measure system performance. In the context of BIM, this means understanding the ideal states of data standard development and implementation and assessing how closely current practices align with these ideals. Furthermore, the attractor-based approach introduces a level of predictability into the evaluation process. By understanding where a system's attractor lies, we can anticipate potential challenges and develop strategies to address them proactively. This proactive approach can lead to more efficient project workflows, reduced errors, and enhanced collaboration among stakeholders.

Moreover, as the industry grapples with the challenges of integrating AI, IoT, and other emerging technologies, the principles of chaos and information theory will become even more crucial. The methodologies proposed in this chapter, from MVD generation to attractor-based evaluation, will need to be expanded, refined, and integrated with these new technologies. as the AECO sector continues to evolve, new challenges will emerge. The integration of emerging technologies, such as AI and IoT, will introduce new layers of complexity into BIM data standards. The methodologies proposed in this chapter will need to be refined and expanded to address these challenges. However, these challenges also present opportunities. The principles of chaos and information theory provide a robust foundation upon which we can build the future of BIM. By continuing to research, innovate, and collaborate, we can harness the full potential of these methodologies and drive transformative change in the AECO sector.

In wrapping up this chapter, it's evident that the confluence of chaos theory, information theory, and BIM holds immense promise for the AECO sector. The journey has been enlightening, but it's also clear that it's just the beginning. As researchers, practitioners, and industry stakeholders, the onus is on us to navigate the challenges ahead, ensuring that the transformative potential of this research is fully actualized in real-world applications.

Chapter 8 -Conclusion and Future Work

Reflecting on the observations and findings from previous sections, this chapter summarizes the research by revisiting the hypothesis and the pre-defined research questions. After that, the research limitations and further improvements that can be made are discussed. Finally, a summary of the research contributions of this thesis is presented.

8.1 Revisiting the hypothesis

In order to clarify the aim of this research, a proposed research hypothesis was assumed as follows:
To building a bridge based on refined understanding of data exchange standard between data standard development and information theory as well as its complex nature that mitigates different advanced theories, techniques, technologies those successfully employed in other domains to support BIM development.

The hypothesis was then decomposed into seven research questions, which are discussed below, based on findings from the previous chapters. Although it was initially envisaged that each chapter would concentrate on particular research questions, the findings from all chapters combined are utilized to remind more compressive answers.

Q1: What are the concepts, frameworks, and process existing within the data exchange standard development to improve BIM ecosystem development? And how are those concepts and such backing BIM development while considering their scope and limitations?

With the rise of computer technology in the architectural domain, recognizing the need for standardized descriptions and exchanges of building information models, the International Alliance for Interoperability (IAI) initiated a project in 1997. This led to the release of the Industry Foundation Classes (IFC) 1.0 in 1999, marking the dawn of the BIM data exchange standard era. FC has been a cornerstone in BIM development. The limitations of IFC 1.0 paved the way for the IFC 2.x versions introduced by buildingSMART. These versions expanded the definitions and relational descriptions of architectural elements, making it the mainstream standard for BIM data exchange. With the advent of complex 4D and 5D BIM applications, buildingSMART introduced the MVD concept in the 2010s. MVD allows users to extract specific modeling information, broadening the application range of the IFC standard. Later, buildingSMART unveiled the Linked Data version of IFC, connecting BIM with data from various fields and constructing digital collaboration (ection 2.2.1).

Along with these existing concepts and frameworks, there are also process carrying on supporting the BIM ecosystem development. Interoperability: The IFC standard has significantly enhanced collaborative design by enabling BIM models from different software platforms to be exchanged and interoperated. BIM Standard Testing: Early concerns revolved around the interoperability of

IFC. The certification approaches mirrored from other domains like healthcare and manufacturing. The focus was on establishing an independent third party for testing and certification. Data Quality Dimensions: Emphasis was placed on the well-formedness and syntactic correctness of data in an IFC model. The dimensions mainly concerned intrinsic and representational data quality (Section 2.2.2).

The BIM data exchange standards have undeniably propelled the BIM ecosystem forward. They have demonstrated the need for a comprehensive, systematic, and robust information model. However, the development of these standards for specific use-case scenarios is diverse. Even with the same approach, the proposed contents can differ significantly. This diversity, while showcasing the potential for growth, also highlights the challenges in achieving a universally accepted standard. Furthermore, while the current testing methodologies have made significant strides in ensuring data quality and interoperability, they are not robust enough for scenarios involving large data from various sources. The lack of quantifiability and understanding in complex nature of data exchange standard development in some testing methodologies hinders their capability for generalization (Section 2.2).

In conclusion, the BIM ecosystem has witnessed remarkable growth and evolution, thanks to the development of data exchange standards. These standards, with their concepts, frameworks, and processes, have provided a solid foundation for BIM development. However, as the architectural and construction landscape becomes more complex, there's a pressing need to address the limitations of these standards. By doing so, we can ensure that BIM continues to be a robust and efficient tool for the industry, paving the way for innovations like the Internet of Things (IoT) and smart city advancements (Section 2.2).

Q2: What are general information theory applications and why information theory would have potential to be employed on data exchange standard development in BIM? What is the possible complex nature of the data exchange standard development in BIM?

Information theory is generally applied to 1. Quantifying Uncertainty: Entropy serves as a metric to quantify the uncertainty inherent in a set of outcomes. A system with higher entropy is more unpredictable; 2. Optimal Coding: Shannon's source coding theorem suggests that entropy determines the limit to how much data can be compressed without loss; 3. Channel Capacity and Transmission: Entropy plays a pivotal role in determining the maximum rate at which information can be transmitted over a communication channel without error; 4. Interdisciplinary Explorations: The concept of entropy in information theory shares parallels with entropy in thermodynamics, leading to explorations in statistical mechanics and quantum computing; 5. Diverse Domains: Information theory has been applied in various fields, from aging-related diseases to predicting oil flow, showcasing its versatility (Section 2.3).

BIM serves as a bridge, translating real-world AEC projects into machine-readable data1. The dynamic nature of the AEC industry, coupled with the multifaceted challenges of BIM, calls for a nuanced approach. Information theory, with its roots in quantifying information and uncertainty, offers insights into: 1. Quantitative analysis–information theory provides a framework for analyzing data, useful for quantifying the amount of information in a dataset or communication system; 2. Interdisciplinary approach – given its interdisciplinary nature, information theory can address diverse challenges in BIM from various angles; 3. Universal applicability – information theory’s principles can be applied to any system involving information transmission or processing, making it apt for BIM’s diverse applications; 4. Efficient methods – information theory offers efficient algorithms for analyzing large datasets, crucial for BIM projects that handle vast amounts of data (Section 2.3).

The possible complex nature of the data exchange standard development in Building Information Modeling (BIM) can be understood by drawing parallels with the principles of chaos theory, just as chaos theory studies unpredictable and complex systems, the development of BIM data standards is inherently complex. This complexity arises from the need to provide a comprehensive digital representation of the physical and functional characteristics of a facility. Also, as chaotic systems have dynamic variables that influence outcomes, BIM data standard development is influenced by dynamic variables (information mass increment) like technological advancements, industry needs, and regulatory changes (Section 2.4).

Q3: what is required to employ information theory and identify the complex nature to support data exchange standard?

The first requirement supporting information theory and identify the complex nature of data exchange standard is to set the proper assumptions those would contribute to extract the topological form of data exchange standard. To better comprehend the dynamics between project information mass and such topology, the ID3 thinking and information entropy are introduced to the research (Section 3.2.1). With the help of ID3 and information entropy, the sandpile simulation is imported to simulate the plausible dynamics. However, this simulation is hard to interpret directly in terms of BIMer's universal terminologies. To tackle the problem, a proxy of data exchange standard (MVD) is chosen to represent the data exchange standard, and a mapping system is proposed to help converting the simulation results to the associated proxy (Section 3.2.2).

Q4: why these are required to employ information theory and identify the complex nature to support data exchange standard?

Information Theory as a Bridge: Information theory serves as a bridge between quantitative measurements and qualitative meaning, enabling a holistic comprehension of the multifaceted landscape of Building Information Modeling (BIM) standards. It suggests that information is not just data points or measurements but a vessel of meaning—a bearer of significance that transcends numerical values. This theory becomes a metaphysical bridge—a conduit through which quantitative measurements find resonance with qualitative meaning. In the context of BIM and data standards, quantitative data often emerges from simulations, measurements, and standardized metrics. However, while its precision is invaluable, quantitative data alone can be devoid of context, meaning, and nuance. Information theory recognizes that information is not merely about data points but also about the context and meaning associated with those data points. It suggests that standards should not merely be about prescribing precise measurements and metrics but also about capturing the broader context, implications, and meanings associated with those metrics (Section 4.6).

Chaos theory, with its emphasis on non-linearity, unpredictability, and emergent properties, provides a lens through which to explore the unpredictable behavior within BIM standards. The integration of chaos theory and information theory acknowledges the multidimensional complexity inherent in optimizing data standards for smart infrastructure. By drawing from both chaos theory and information theory, this research navigates the intricate terrain where technological innovation, industry dynamics, and human interactions converge. Chaos theory underscores the potential for unforeseen disruptions and nonlinear interactions within BIM standards. Interdisciplinary integration embodies a respect for diverse ways of knowing and problem-solving. It recognizes that complex phenomena cannot be fully understood through a singular lens. This stance values the contributions of different disciplines, fostering a more comprehensive and nuanced understanding of the intricate interplay between chaos, information, and BIM standards (Section 4.2).

In summary, employing information theory and identifying the complex nature of BIM standards can provide a more comprehensive understanding of data exchange standards. This approach not only captures the precise measurements but also the broader context, implications, and meanings, ensuring that the standards are both rigorous in their precision and rich in their interpretative depth.

Q5: What is the possible complex nature through exploration?
The chaos is the possible complex nature of BIM data exchange standard development. This is enlightened by the discovery of SOC and fractal in the development of MVD—proxy of data exchange standard which are deemed as the indication of chaos exists in the system. Thus, whether weak chaos or chaos, the proposed research brings a great step forward by detecting the existence of chaos in the BIM model development. Chaos as a paradigm, manages to apply its related fractals to explore many other disciplines. This succeeds in being the first within the scholars to bridge chaos with BIM by presenting the existence of chaos in the BIM standard development. The discovery of chaos in BIM data standard development challenges the traditional linear approaches that have dominated the field. It calls for a paradigm shift, urging researchers and practitioners alike to embrace the complexities and uncertainties inherent in the AEC industry (Section 5.5).

Q6: How to covert the possibility of discovered complex nature to certainty of discovered complex nature?

Start with data preparation, leveraging sandpile simulations to capture the essence of the increasing information mass. The time series derived from this data serves as guiding narrative, chronicling the system's behavior over successive intervals. The reconstruction of the phase space follows, providing a multi-dimensional view of the system's dynamics and paving the way for the identification of strange attractors. The culmination of this process is the calculation of the largest Lyapunov exponent using Wolf algorithms, which, as indicated, provides definitive proof of the system's chaotic nature (Section 6.4).

Q7: Can the findings and refined understanding with respect to information theory and complex nature benefit the data exchange standard development?

Chaos theory, traditionally confined to abstract mathematical discourse, presents a pragmatic avenue for advancements in Building Information Modeling (BIM). This is vividly illustrated in the innovative approach to Model View Definitions (MVD) generation through the lens of sandpile simulation and information theory. While conventional methodologies provide a foundational understanding, they occasionally overlook the intricate subtleties inherent in contemporary construction endeavors. The introduced methodology, anchored in the principles of chaos and information theory, presents a fluid, adaptive, and context-aware strategy, ensuring that BIM data standards align seamlessly with the dynamic needs of the AECO domain. The inherent flexibility of the suggested MVD generation approach emerges as a noteworthy innovation. Given the fluid nature of the AECO sector, a methodology that can organically adapt to project nuances is of paramount importance. Such flexibility not only reduces discrepancies but also fosters improved teamwork and streamlines processes. The empirical studies, utilizing a 5x5 grid and sandpile simulation, epitomize the practical application of theoretical frameworks. The conversion of data into a hierarchical tree structure, underpinned by information theory, underscores the vast potential of this cross-disciplinary methodology. A seminal contribution elucidated in chapter 7 is the introduction of an attractor as a metric for gauging BIM data standard efficacy. Within chaos theory, attractors symbolize the equilibrium states a system naturally gravitates towards. Recognizing and delineating these attractors for BIM data standards offers deep insights into system dynamics, robustness, and operational efficiency. The incorporation of attractors as a metric for BIM data standard evaluation is pioneering, providing a fresh perspective for assessing resilience, adaptability, and overall performance of BIM data standards. By pinpointing the system's attractor, we can forecast its trajectory, discern potential challenges, and make strategic decisions to guide the system towards optimal outcomes. Moreover, the introduction of attractors infuses a quantitative dimension to the evaluation, complementing qualitative analyses and enhancing the evaluation's precision and objectivity.

While the methodologies and perspectives articulated in this dissertation are groundbreaking, they signify just the initial steps. As the AECO sector progresses towards intelligent infrastructure, the imperative for refined data standards intensifies. The revelations from chapter 7 lay a solid groundwork but also highlight the imperative for ongoing inquiry, enhancement, and evolution. At its essence, chaos theory seeks to decipher the erratic and ostensibly arbitrary behaviors of intricate systems. Within the BIM landscape, such unpredictability is not merely theoretical but a tangible challenge. Projects frequently encounter unexpected hurdles, from design modifications to operational setbacks, each introducing elements of chaos. By integrating chaos theory, we can more adeptly foresee these obstacles, comprehend their root causes, and devise mitigation strategies. The proposition of utilizing attractors for BIM data standard assessment holds extensive ramifications. As equilibrium states, attractors serve as performance yardsticks, enabling us to gauge the alignment of current practices with idealized standards. Additionally, the attractor-centric approach infuses a degree of foresight into the assessment, allowing for the proactive identification and addressal of challenges, leading to streamlined project operations, diminished inaccuracies, and bolstered stakeholder collaboration.

8.2 Research Limitations and Future Works

The limitations and future work of this research are discussed below:

Amalgamating cutting-edge technologies with Building Information Modeling (BIM) into a unified framework has emerged as a pivotal area of scholarly inquiry. Harnessing the synergies of these technologies presents a promising avenue to construct a robust support infrastructure, enhancing both BIM capabilities and the decision-making mechanisms throughout the building lifecycle. However, the pattern of integrating advanced technologies is both conservative and radical. It is "radical" since the most integration is application oriented; the application is merely built on sound theoretical ground but empirical results. It is "conservative" as regardless of extensive attempts with integration of various technologies in BIM but rare attempts to dig deeper with respect to the reason behind these empirical results. This thesis makes attempts to transplant an entire existing knowledge system of chaos theory and information theory into BIM domain that has previously not engaged with or discussed this system requires a significant amount of work. Notwithstanding, tribute to Vapnik and Chervonenkis 's contribution regarding the investigation of uniform convergence in functional spaces in 1974 which enables the learning theory application on the basis of observed data by providing rigorous theoretical proof to that, this thesis provided rigorous and comprehensive proof of employing chaos theory related techniques, methods to BIM. It means that, in theory, all the chaos based/oriented tools, techniques, technologies remain potential to integrate with BIM. This thesis provides two applications, and it is not meant to be exclusive. It may take several years to achieve the whole transplanting. Numerous technologies, principles, and notions could augment BIM's capabilities, though they remain outside the scope of this research. Additionally, the evaluation in Chapter 2 relied on data sourced from WoS, focusing solely on English-language literature. Beyond the quantitative assessment, a qualitative exploration was undertaken. However, due to the sheer volume, it was infeasible to encompass all the gathered literature. A selection of the amassed papers was examined to provide a glimpse into prevailing research. Consequently, this investigation might not encapsulate the full breadth of BIM literature pertaining to these subjects.

The beginning of this thesis established “bit” as the unit for measuring information. However, with the simulation results, it can be seen the actual amount of information is insignificant. What really matters are ratios of information, as long as consistency is maintained of the information measurement, other units of information are applicable as well, such as “nat”, then a grain of sand represents a “nat” information. Or even simpler, information contained in a lining can be defined as “lin”, and “lin” can be associated with grains of sand. Hence, the unit system varies from specific projects. Another practical approach to calculating information mass is to use natural language processing algorithms that capture the occurrence of information entries as the material to recalculate the importance/weight/quantity of information. As one can see, the branching condition for topology development of BIM model is associated with the toppling rules. In assumption, different types of information contained in the tunnel lining do not have to be taken as identically distributed. Since it is natural that the usefulness of information varies with end-users and use case scenarios. This thesis designed assumption for the purpose of simplicity as the author intended to bring more methodological thinking rather than the simulation results. Also, as aforementioned, the branching triggering tendency is tightly linked to objective collapse. In this context, it is use case sensitive. All the assumptions made in the thesis enable the simplicity and self-consistency throughout the whole research, the viable application in practice could be further discussed. The more unified detailed information requirement analysis is supposed to advocated, which could serve the information mass calculation consistent across borders. Studies on different toppling rules need to be promoted so that valuable outputs in practice can be arranged in a more applicable and explanatory way. Meanwhile, the interpretation of toppling along with fluctuation -dissipation relation in the case of building information modelling development deserves more effort.

The current proposed application with information theory and chaos theory is the "proof of concept" stage and requires further improvement. The developed framework is semi-automated, and the knowledge base covers the most necessary concepts and aspects. There are supposed to be more cases involved. To some extent, without considering the thesis is theoretical study oriented, some would argue that the limited case studies would limit the generalizability. Future work will include fully automating the framework and involve more cases along with various models. Also, studies on integrating information theory with BIM should gain more attention. This paper shows the aptitude of employing information theory to improve the BIM model development in terms of topology, the comprehensive capability of information theory in BIM fields has foreseeable space to explore. The discovered topology of BIM data model can be practiced with associated engineering semantics, which responds to

Shannon's information theory as engineering semantics are endowed with Information Delivery Manual (IDM) and International Framework for Dictionary (IFD). Based on these this thesis would contribute to and haven't discussed in the thesis would be the automated generation of IFC. Since the typical steps of automated IFC generations would be data extraction, data mapping, conversion and transformation, validation and quality assurance, integration and collaboration. The proposed research ideas and framework would be utilized to unify data mapping process, conversion and transformation with less priori agreement considering the process would only require toppling rules, initial state of sandpile lattice and information requirements. On top of that, as the study provides a new dimension for validation and quality assurance, the associated step would be more objective and comprehensive. Therefore, research on automated IFC generation could be advanced in the future. Especially, given the current matching process of entropy entity pair and specification entity pair has not been further discussed the possibility of perfect matching, such automated generation could be limited at the moment. Further research on number theory, interrelationship through mathematical functions, combinatorial mathematics and unique pairings would require more effort, so that a verifiable mathematical link between the entity pairs would establish, indicating that their correspondence is both intentional and mathematically substantiated. The use of a deterministic mathematical function to define the relationship between the original and translated pairs could ensure that the transformation of each pair is underpinned by well-defined mathematical operations, thus providing a solid basis for verification. Finally, the application of combinatorial mathematics could highlight the deliberate complexity and uniqueness of the pairing process, reinforcing the assertion that the matching of the pairs is not coincidental but is the outcome of a meticulously defined, mathematically rigorous procedure.

The whole research method is built on deduction using symbolic logic after mathematical modelling—creating certain symbols to represent system. This would fetch another limitation; the mathematical modeling presumes the relations between information mass and information model standard development at the beginning which is empirical realism. From the perspective of empirical realism, discussing information modeling and not involving information theory is rather absurd regardless of certain amount of BIMers would not even agree on such perspective. This study offers a way out of this, by using information mass as a dimension to develop information standard. Yet, the viewpoint is undoubtful, the solution this study proposed is limited. Not only there might be other ways of relating information theory to information standard development those this study did not cover, but also there might be unspecified entities within existing specification in certain cases. Despite these are not wrong either if the research is justifying from the perspective of positivism as the validation results is consistent with proposed theoretical deduction. Furthermore, the research relies heavily on quantitative data, at the early-stage development of this field, it does not provide a comprehensive analysis of qualitative or subjective factors that may also impact BIM modeling performance. Besides, the whole thesis heavily relies on methods of mathematical physics, which is relatively complex and requires advanced mathematical knowledge, which may limit its accessibility to non-expert users. For the very same reason, the other BIM standard performance evaluation techniques are rarely developed on the basis of theoretical foundation in terms of mathematics/physics but more prone to application oriented. Hence, it makes the direct comparison between proposed metrics and existing techniques inaccessible. A more integrated comparison approach is supposed to develop for compatibilizing chaos and information theory with BIM.

8.3 Research Contribution

This research contains work related to several and practical developments in an effort to transplant the mature existing knowledge system of chaos theory and information theory into BIM data exchange standard development to eliminate inefficiencies and inconsistencies in the process. Taking into account the findings and development presented in this dissertation, the main contributions resulting from this thesis are listed below:

Utilizing the sandpile model to simulate the topology of MVD is self-consistent under the provided assumptions. And tangibly, it is not only the approach to interpret BIM standard development, but also a very intuitional method to develop a powerful tool for meta standard promotion. Though, vast research indicated that graph theory can be used to analyze BIM standards, many scholars attempted to exploit tools within graph theory to study BIM standards and made valuable contributions. From the results, the proposed research including mapping system and sandpile simulations yields more compatibility between graph theory and the BIM standard in favor of information theory. In other words, the proposed intuitive method for developing the topology of the BIM model can import powerful tools from both graph theory and information theory to BIM, which is a considerable contribution of this research. Unlike most other fellow scholars' work which mainly built on semantic of engineering entities and civil engineering's subjectivity-based BIM standard development framework, the proposed approach empowers the dynamics among information self-organizing topology of BIM model. The subjectivity has been eliminated which is considered significant addition of this thesis. A detailed article entitled "Sandpile-Simulation-based Graph Data Model for MVD generative Design of Shield Tunnel Lining using Information Entropy" was published in journal "*Advanced Engineering Informatic*"*.* And the NLP related research used in the thesis is mentioned in another article entitled "Determining Uncertainties in AI Applications in AEC Sector and their Corresponding Mitigation Strategies" was published in journal "*Automation in Construction*".

One of the main contributions is to transplant an entire existing knowledge system of chaos theory and information theory into BIM by rigorously proving the chaos does exist in BIM. Through such theoretical proof that discovering the complex nature of BIM data standard development, the mature techniques, tools, technologies of chaos theory those have already been developed and employed to other disciplines could be empowered with huge potentials in the BIM domain. It enlightens the fellow scholars to embrace a huge world of technique base, knowledge base in a more direct, comprehensive and rigorous way. The chaos-based BIM data standard evaluation proposed in this thesis is a typical example of that. It levers the concept of "attractor" from chaos to help appraise the data standard. Also, such way of evaluating data standard does not limit to BIM domain but all potential industry data standard. A detailed article entitle "Time Series Analysis of Chaotic Properties for Assessing BIM Modeling Performance" is submitted to journal "*Nonlinear Analysis, Theory, Methods and Applications*".

Another contribution of the research is involving sandpile simulation and proposed a mapping system to study the proxy of BIM data exchange standard (MVD). This is pivotal due to the significance of importing group theory and tensor into BIM study. In other words, Sandpile simulation and proposed mapping system being the lens to bring mathematical concepts as well as modeling into BIM study. As most scholars acknowledged, a dimension of mathematics in application-oriented study could not only optimize the ways of studying subjects but also provide more comprehensive ways of justifying subjects. It could easily to better generalize the findings, to allow more precise expression, to provide more structured and logical framework, to conduct quantitative analysis so as eliminating subjectivity as well as remain consistent, to further integrate with computers and advanced software. The contribution of studying MVD through group theory is currently in the preparation stage and will be submitted to the journal "*Nonlinear Analysis: Real World Applications*", in the paper, it discusses the identity element of group that could generate/interpret MVDs via sandpile simulation and how the identity element could affect the MVDs' development.

References

Abarbanel, H. D. I. (1996). *Analysis of Observed Chaotic Data*. New York: Springer-Verlag.

Abualdenien, J., & Borrmann, A. (2019). A meta-model approach for formal specification and consistent management of multi-LOD building models. *Advanced Engineering Informatics, 40*, 135. doi:10.1016/j.aei.2019.04.003

Allzén, S. (2021). Scientific realism and empirical confirmation: A puzzle. *Studies in History and Philosophy of Science Part A, 90*, 153-159. doi:https://doi.org/10.1016/j.shpsa.2021.10.008

Amor, R. (2008). A Better BIM: Ideas from other Industries. *Proceedings of CIB W78 conference on Improving the management of construction projects through IT adoption*, Santiago, Chile, 15-17 July, pp. 48-53.

An, Y. (2017). *Geological data extension for subway tunnel BIM models using a Linked Data Approach.* [Master dissertation, Eindhoven University of Technology]. Cardiff Repository

An, Y., Chen, N., Li, H., & Wang, Y. (2023). Time Series Analysis of Chaotic Properties for Assessing BIM Modeling Performance. *Advanced Engineering Informatics*, 102108. Retrieved from https://www.sciencedirect.com/science/article/pii/S1474034623002367

An, Y., Lin, X., Li, H., & Wang, Y. (2023b). Sandpile-simulation-based graph data model for MVD generative design of shield tunnel lining using information entropy. *Advanced Engineering Informatics, 57*, 102108. doi:https://doi.org/10.1016/j.aei.2023.102108

Andre Borrmann, J. B., Christian Koch, Thomas Liebich, Sergej Muhic. (2018). Industry Foundation Classes: A Standardized Data Model for the Vendor-Neutral Exchange of Digital Building Models. In M. K. Andre Borrmann, Christian Koch, Jakob Beetz (Ed.), *Building Information Modeling*. Switzerland: Springer International Publishing AG.

Andre Borrmann, M. K., Christian Koch, Jakob Beetz. (2018). *Building Information Modeling*: Springer.

Angel, R. B. (1980). 4 - Empiricism, Rationalism and Special Relativity. In R. B. Angel (Ed.), *Relativity* (pp. 103-122): Pergamon.

Azar, A. T., & Vaidyanathan, S. (2016). *Advances in Chaos Theory and Intelligent Control* (Vol. 337).

Bak, P., Tang, C., & Wiesenfeld, K. (1987a). Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: An explanation of 1/f noise. Phys. Rev. Lett. 59, 381-384. *Physical Review Letters, 59*, 381-384. doi:10.1103/PhysRevLett.59.381

Bak, P., Tang, C., & Wiesenfeld, K. (1987b). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters, 59*(4), 381-384. doi:10.1103/PhysRevLett.59.381

Barnsley, M. F. (1993). Chapter X - Recurrent Iterated Function Systems. In M. F. Barnsley (Ed.), *Fractals Everywhere (Second Edition)* (pp. 379-411): Academic Press.

Barnsley, M. F., & Sloan, A. D. (1988). A better way to compress images. *BYTE, 13*(1), 215–223.

Beetz, J., van Leeuwen, J., & Vries, d. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. *AI EDAM, 23*, 89-101. doi:10.1017/S0890060409000122

Bensky, A. (2019). Chapter 9 - Introduction to information theory and coding. In A. Bensky (Ed.), *Short-range Wireless Communication (Third Edition)* (pp. 211-236): Newnes.

Biswas, S., X. Rhodes, H., O. McCrea, A., & Zino, C. (2023). Chapter 55 - Large, pragmatic trials: designs, measures and classic examples. In A. E. M. Eltorai, J. A. Bakal, P. C. Newell, & A. J. Osband (Eds.), *Translational Surgery* (pp. 333-338): Academic Press.

Booth, W. C., Colomb, G. G., Williams, J. M., Bizup, J., & FitzGerald, W. T. (2016). *The Craft of Research* (Fourth Edition ed.): University of Chicago Press.

Borrmann, A., Flurl, M., Jubierre, J. R., Mundani, R.-P., & Rank, E. (2014). Synchronous collaborative tunnel design based on consistency-preserving multi-scale models. *Advanced Engineering Informatics, 28*(4), 499-517. doi:https://doi.org/10.1016/j.aei.2014.07.005

buildingSMART. (2012). An integrated process for delivering IFC based data exchange. Retrieved from http://iug.buildingsmart.org/idms/methods-and-guides/Integrated_IDM-MVD_ProcessFormats_14.pdf.

buildingSMART. (2013). Construction operations building information exchange, MVD definition for IFC4. Retrieved from http://docs.buildingsmartalliance.org/MVD_COBIE/

Cambel, A. B. (1992). *Applied Chaos Theory: A Paradigm for Complexity* (1st ed.): Academic Press.

Cannon, J. W. (1984). The Fractal Geometry of Nature. [The Fractal Geometry of Nature., Benoit B. Mandelbrot]. *The American Mathematical Monthly, 91*(9), 594-598. doi:10.2307/2323761

Castilla-Rho, J. C., Mariethoz, G., Rojas, R., Andersen, M. S., & Kelly, B. F. J. (2015). An agent-based platform for simulating complex human-aquifer interactions in managed groundwater systems. *Environmental Modelling and Software, 73*, 305-323. doi:10.1016/j.envsoft.2015.08.018

Charmaz, K. (2001). Grounded Theory: Methodology and Theory Construction. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 6396-6399). Oxford: Pergamon.

Chen, H., & Luo, X. (2019). An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing. *Advanced Engineering Informatics, 42*, 100959. doi:https://doi.org/10.1016/j.aei.2019.100959

Lv, J., Lu, J., Chen, S. (2002) *Chaotic Time Series Analysis and Application*. Wuhan, China: Wuhan University.

Choueiri, G., Suri, B., Merrin, J., Serbyn, M., Hof, B., & Budanur, N. B. (2022). Crises and chaotic scattering in hydrodynamic pilot-wave experiments. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 32*(9), 093138. doi:10.1063/5.0102904

Cuthbertson, L. M., Robb, Y. A., & Blair, S. (2020). Theory and application of research principles and philosophical underpinning for a study utilising interpretative phenomenological analysis. *Radiography, 26*(2), e94-e102. doi:https://doi.org/10.1016/j.radi.2019.11.092

China National Design code of Shield Tunnel Engineering, (2021).

Dhar, D. (2006). Theoretical studies of self-organized criticality. *Physica A: Statistical Mechanics and its Applications, 369*(1), 29-70. doi:https://doi.org/10.1016/j.physa.2006.04.004

Drakopoulos, V., Dalla, L., & Bouboulis, P. (2006). IMAGE COMPRESSION USING RECURRENT BIVARIATE FRACTAL INTERPOLATION SURFACES. I*nternational Journal of Bifurcation and Chaos* 16(07), 2063-2071. https://doi.org/10.1142/S0218127406015908

Eastman, C. M., Jeong, Y. S., Sacks, R., & Kaner, I. (2010). Exchange Model and Exchange Object Concepts for Implementation of National BIM Standards. *Journal of Computing in Civil Engineering, 24*(1), 25-34. doi:10.1061/(ASCE)0887-3801(2010)24:1(25)

Fan, H., Ma, Z., Wang, D., & Liu, J. (2018). Handling distributed XML queries over large XML data based on MapReduce framework. *Information Sciences, 453*, 1-20. doi:https://doi.org/10.1016/j.ins.2018.04.028

Feder, J. (2013). Fractals. Springer New York, https://doi.org/10.1007/978-1-4899-2124-6

Fisher, Y. (2012). *Fractal image compression: theory and application*: Springer Science & Business Media.

Gan, V. J. L. (2022). BIM-based graph data model for automatic generative design of modular buildings. *Automation in Construction, 134*, 104062. doi:https://doi.org/10.1016/j.autcon.2021.104062

Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena, 9*(1), 189-208. doi:https://doi.org/10.1016/0167-2789(83)90298-1

Hage, J. (2007). - The intersection of philosophy and theory construction: The problem of the origin of elements in a theory. In S. P. Turner & M. W. Risjord (Eds.), *Philosophy of Anthropology and Sociology* (pp. 121-156). Amsterdam: North-Holland.

Hellen, E., & Thomas, J. (2008). Prediction and measurement of transient responses of first difference based chaos control for 1-dimensional maps. arXiv:abs/0807.2637v3.

Jacquin, A. E. (1992). Image coding based on a fractal theory of iterated contractive image transformations. *IEEE Transactions on Image Processing, 1*(1), 18-30. doi:10.1109/83.128028

Katsenelinboigen, A. (1980). 1 - Ideology, Pragmatism, and the Development of Mathematical Economics in the USSR. In A. Katsenelinboigen (Ed.), *Soviet Economic Thought and Political Power in the USSR* (pp. 1-27): Pergamon.

Kiviniemi, A. (2008). IFC Certificationn process and data exchange problems. In A. Kiviniemi (Ed.), *eWork and eBusiness in Architecture, Engineering and Construction* (pp. 6): Taylor&Francis.

Knorr-Cetina, K. D. (1981). Chapter 1 - The Scientist as a Practical Reasoner: Introduction to a Constructivist and Contextual Theory of Knowledge. In K. D. Knorr-Cetina (Ed.), *The Manufacture of Knowledge* (pp. 1-32). Amsterdam: Pergamon.

Koch, C., Vonthron, A., & König, M. (2017). A tunnel information modelling framework to support management, simulations and visualisations in mechanised tunnelling projects. *Automation in Construction, 83*, 78-90. doi:https://doi.org/10.1016/j.autcon.2017.07.006

Laakso, M. K., AO. (2012). The IFC standard: A review of History, development, and standardization, Information Technology. *ITcon*. Retrieved from http://usir.salford.ac.uk/id/eprint/28373

Ladyman, J. (2007). - Ontological, Epistemological, and Methodological Positions. In T. A. F. Kuipers (Ed.), *General Philosophy of Science* (pp. 303-376). Amsterdam: North-Holland.

Lawrence, R. (2004). The space efficiency of XML. *Information and Software Technology, 46*(11), 753-759. doi:https://doi.org/10.1016/j.infsof.2004.02.003

Lee, G., Park, Y. H., & Ham, S. (2013). Extended Process to Product Modeling (xPPM) for integrated and seamless IDM and MVD development. *Advanced Engineering Informatics, 27*(4), 636-651. doi:https://doi.org/10.1016/j.aei.2013.08.004

Lennox, R., & Jurdi-Hage, R. (2017). Beyond the empirical and the discursive: The methodological implications of critical realism for street harassment research. *Women's Studies International Forum, 60*, 28-38. doi:https://doi.org/10.1016/j.wsif.2016.11.010

Levy, D. (2007). Chaos theory and strategy: Theory, application, and managerial implications. *Strategic Management Journal, 15*, 167-178. doi:10.1002/smj.4250151011

Li, X., Yang, D., Yuan, J., Donkers, A., & Liu, X. (2022). BIM-enabled semantic web for automated safety checks in subway construction. *Automation in Construction, 141*, 104454. doi:https://doi.org/10.1016/j.autcon.2022.104454

Lin, H. (2016). *IFC based IFC shield tunnel structure life-cycle information modelling research and application.* [Master dissertation, . Eindhoven University of Technology]. Cardiff Repository

Liu, H., Lu, M., & Al-Hussein, M. (2016). Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry. *Advanced Engineering Informatics, 30*(2), 190-207. doi:https://doi.org/10.1016/j.aei.2016.03.001

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature, 473*(7346), 167-173. doi:10.1038/nature10011

Liu, Y., Li, M., Wong, B. C. L., Chan, C. M., Cheng, J. C. P., & Gan, V. J. L. (2021). BIM-BVBS integration with openBIM standards for automatic prefabrication of steel reinforcement. *Automation in Construction, 125*, 103654. doi:https://doi.org/10.1016/j.autcon.2021.103654

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences, 20*(2), 130-141. doi:https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Ma, Z., & Liu, Z. (2018). Ontology- and freeware-based platform for rapid development of BIM applications with reasoning support. *Automation in Construction, 90*, 1-8. doi:https://doi.org/10.1016/j.autcon.2018.02.004

Maiväli, Ü. (2015). Chapter 2 - The Basis of Knowledge: Causality and Truth. In Ü. Maiväli (Ed.), *Interpreting Biomedical Science* (pp. 55-108). Boston: Academic Press.

Mandelbrot, B. B. (1983). *The fractal geometry of nature* (3 ed.). New York: W. H. Freeman and Comp.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature, 261*(5560), 459-467. doi:10.1038/261459a0

Meakin, P. (2011). Fractals, Scaling and Growth Far From Equilibrium. Cambridge University Press.

Mezei, M., & Sárosi, G. (2020). Chaos in the butterfly cone. *Journal of High Energy Physics, 2020*(1), 186. doi:10.1007/JHEP01(2020)186

Mg, M. (1994). [Chaos theory and psychoanalysis: the fluid nature of the soul]. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse, 40*, 384-403.

Michaels, J. (2013). *Identifying Attractors in Time Series Data*. Paper presented at the Innovation Lab on Dynamical Systems Theory, Peace and Conflict, New York.

Noether, E. (1927). Abstrakter Aufbau der Idealtheorie in algebraischen Zahl- und Funktionenkörpern. *Mathematische Annalen, 96*(1), 26-61. doi:10.1007/BF01209152

Okasha, S. (2002). *Philosophy of Science: A Very Short Introduction* (First Edition ed.): OUP Oxford.

Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a Time Series. *Physical Review Letters, 45*(9), 712-716. doi:10.1103/PhysRevLett.45.712

Patacas, J., Dawood, N., & Kassem, M. (2020). BIM for facilities management: A framework and a common data environment using open standards. *Automation in Construction, 120*, 103366. doi:https://doi.org/10.1016/j.autcon.2020.103366

Peters, E. E. (1996). *Chaos and order in the capital markets*: Wiley.

Poon, C.-S., Li, C., & Wu, G.-q. (2010). A unified theory of chaos linking nonlinear dynamics and statistical physics. *arXiv: Chaotic Dynamics*.

Ren, G., Li, H., Liu, S., Goonetillake, J., Khudhair, A., & Arthur, S. (2021). Aligning BIM and ontology for information retrieve and reasoning in value for money assessment. *Automation in Construction, 124*, 103565. doi:https://doi.org/10.1016/j.autcon.2021.103565

Rorty, R. (1991). *Objectivity, relativism, and truth* (Vol. 1): Cambridge University Press.

Rovelli, C. (2018). Physics Needs Philosophy. Philosophy Needs Physics. *Foundations of Physics, 48*(5), 481-491. doi:10.1007/s10701-018-0167-y

Ruelle, D., & Dewitt-Morette, C. (1990). Elements of Differentiable Dynamics and Bifurcation Theory. *Physics Today, 43*(2), 120-120. doi:10.1063/1.2810458

Sanguinetti, P., Eastman, C., & Augenbroe, G. (2009). Courthouse energy evaluation: BIM and simulation model interoperability in concept design. *IBPSA 2009 - International Building Performance Simulation Association 2009*.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational technology, 35*(5), 31-38.

Schuldberg, D., & Guisinger, S. (2020). Nonlinear Dynamical Systems, Chaos Theory, and Creativity☆. In S. Pritzker & M. Runco (Eds.), *Encyclopedia of Creativity (Third Edition)* (pp. 245-254). Oxford: Academic Press.

Selvam, A. M. (2001). Signatures of Quantum-like Chaos in Spacing Intervals of Non-trivial Riemann Zeta

Zeros and in Turbulent Fluid Flows. *arXiv: General Physics*.

Selvam, A. M. (2004). Fractal space-time fluctuations: A signature of quantumlike chaos in dynamical systems. *arXiv: General Physics*.

Serpico, D., Lynch, K. E., & Porter, T. M. (2023). New historical and philosophical perspectives on quantitative genetics. *Studies in History and Philosophy of Science, 97*, 29-33. doi:https://doi.org/10.1016/j.shpsa.2022.11.009

Shannon-Baker, P. (2023). Philosophical underpinnings of mixed methods research in education. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education (Fourth Edition)* (pp. 380-389). Oxford: Elsevier.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), 379-423. doi:https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sheridan, P. J. (1996). *Spiral architecture for machine vision*. https://arxiv.org/pdf/0807.2637

Smith, S. J. (2020). Pragmatism. In A. Kobayashi (Ed.), *International Encyclopedia of Human Geography (Second Edition)* (pp. 1-4). Oxford: Elsevier.

Solihin, W., Eastman, C., & Lee, Y.-C. (2015). Toward robust and quantifiable automated IFC quality validation. *Advanced Engineering Informatics, 29*(3), 739-756. doi:https://doi.org/10.1016/j.aei.2015.07.006

Spencer, H. (2021). CHAOS : Review/Essay of "CHAOS" 'Making a New Science' by James GLEICK (1987).

Sprott, J. C. (2003). *Chaos and Time-Series Analysis*. New York: Oxford University Press.

Stapleton, D., Hanna, J. B., & Ross, J. R. (2006). Enhancing supply chain solutions with the application of chaos theory. *Supply Chain Management: An International Journal, 11*(2), 108-114. doi:10.1108/13598540610652483

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Commun. ACM, 40*(5), 103–110. doi:10.1145/253769.253804

T.Chipman, L., T., T.Wiese,. (2013). *buildingSMART Model Support Group*. Retrieved from https://standards.buildingsmart.org/MVD/RELEASE/mvdXML/v1-1/mvdXML_V1-1-Final.pdf

Takens, F. (1981, 1981//). *Detecting strange attractors in turbulence.* Paper presented at the Dynamical Systems and Turbulence, Warwick 1980, Berlin, Heidelberg.

Vlad, S., Dumitru, O., & Pascu, P. (2007). Chaos Models in Economics. *Journal of Applied Computer Science & Mathematics,* Volume 1, Number 2, pp. 43-46(4).

Vlad, S. E. (2023). 16 - Chaos. In S. E. Vlad (Ed.), *Boolean Systems* (pp. 145-149): Academic Press.

Vries, d. (2005). IT Standards Typology. In J. K. (Ed.), *Advanced Topics in Information Technology Standards and Standardization Research* (pp. 11-36). Hershey, PA, USA: Idea Group Publishing.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Commun. ACM, 39*, 86-95.

Wang, C., & Yi, S. (2007). Empirical Evidence of chaotic time series analysis in Gross Domestic Production

Prediction *Journal of Tianjin University 9*(2), 137-139. doi:10.3969/j.issn.1008-4339.2007.02.009

Wang, H., Meng, X., & Zhu, X. (2022). Improving knowledge capture and retrieval in the BIM environment: Combining case-based reasoning and natural language processing. *Automation in Construction, 139*, 104317. doi:https://doi.org/10.1016/j.autcon.2022.104317

Wang, S., Mao, J., Lu, K., Cao, Y., & Li, G. (2021). Understanding interdisciplinary knowledge integration through citance analysis: A case study on eHealth. *Journal of Informetrics, 15*(4), 101214. doi:https://doi.org/10.1016/j.joi.2021.101214

Wang, Y. Z., J.; Li, J.;. (2014). IFC based Building Structure Construction Design Information Model Presentation. *Information Technology in Civil Engineering*, 30-35.

Weintrop, D. T., S.; Tinker, R.; Head, B. and Wilensky, U. (2011). NetLogo Sandpile model. Retrieved from http://ccl.northwestern.edu/netlogo/models/Sandpile

Wilensky, U. (1999). NetLogo. Retrieved from http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling

William East, E., Nisbet, N., & Liebich, T. (2013). Facility Management Handover Model View. *Journal of Computing in Civil Engineering, 27*(1), 61-67. doi:10.1061/(ASCE)CP.1943-5487.0000196

Williams, G. (1997). *Chaos Theory Tamed*: CRC Press.

Wilson, K. G. (1983). The renormalization group and critical phenomena. *Reviews of Modern Physics, 55*(3), 583-600. doi:10.1103/RevModPhys.55.583

International Tunnel Association. (2002). Guidelines for the Design of Shield Tunnel Lining. Retrieved from https://doi.org/10.1016/S0886-7798(00)00058-4

Wu, S., Shen, Q., Deng, Y., & Cheng, J. (2019). Natural-language-based intelligent retrieval engine for BIM object database. *Computers in Industry, 108*, 73-88. doi:https://doi.org/10.1016/j.compind.2019.02.016

Zhang, X. (2023). Knowledge integration in interdisciplinary research teams: Role of social networks. *Journal of Engineering and Technology Management, 67*, 101733. doi:https://doi.org/10.1016/j.jengtecman.2023.101733

Zhao, X., & Chen, T. (2002). Type of self-organized criticality model based on neural networks. *Physical Review E, 65*(2), 026114. doi:10.1103/PhysRevE.65.026114

Zhou, Y.-C., Zheng, Z., Lin, J.-R., & Lu, X.-Z. (2022). Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Computers in Industry, 142*, 103746. doi:https://doi.org/10.1016/j.compind.2022.103746

Appendix A Netlogo Simulation Code for Sandpile Simulation with Applied Model

```
globals [
  ;; By always keeping track of how much sand is on the table, we can compute
the
  ;; average number of grains per patch instantly, without having to count.
  total
  ;; We don't want the average monitor to updating wildly, so we only have
it
  ;; update every tick.
  total-on-tick
  ;; Keep track of avalanche sizes so we can histogram them
  sizes
  ;; Size of the most recent run
  last-size
  ;; Keep track of avalanche lifetimes so we can histogram them
  lifetimes
  ;; Lifetime of the most recent run
  last-lifetime
  ;; The patch the mouse hovers over while exploring
  selected-patch
  ;; These colors define how the patches look normally, after being fired,
and in
  ;; explore mode.
  default-color
  fired-color
  selected-color
]

patches-own [
  ;; how many grains of sand are on this patch
  n
  ;; A list of stored n so that we can easily pop back to a previous state.
See
```

```
  ;; the NETLOGO FEATURES section of the Info tab for a description of how
stacks
  ;; work
  n-stack
  ;; Determines what color to scale when coloring the patch.
  base-color
]

;; The input task says what each patch should do at setup time
;; to compute its initial value for n.  (See the Tasks section
;; of the Programming Guide for information on tasks.)
to setup [setup-task]
  clear-all

  set default-color yellow
  set fired-color red
  set selected-color blue

  set selected-patch nobody
  ask patches [
    set n runresult setup-task
    set n-stack []
    set base-color default-color
  ]
  let ignore stabilize false
  ask patches [ recolor ]
  set total sum [ n ] of patches
  ;; set this to the empty list so we can add items to it later
  set sizes []
  set lifetimes []
  reset-ticks
end

;; For example, "setup-uniform 2" gives every patch a task which reports 2.
to setup-uniform [initial]
  setup [ -> initial ]
```

```
end

;; Every patch uses a task which reports a random value.
to setup-random
  setup [ -> random 8 ]
end

;; patch procedure; the colors are like a stoplight
to recolor
  set pcolor scale-color base-color n 0 16
end

to go
  let drop drop-patch
  if drop != nobody [
    ask drop [
      update-n 1
      recolor
    ]
    let results stabilize animate-avalanches?
    let avalanche-patches first results
    let lifetime last results

    ;; compute the size of the avalanche and throw it on the end of the sizes
list
    if any? avalanche-patches [
      set sizes lput (count avalanche-patches) sizes
      set lifetimes lput lifetime lifetimes
    ]
    ;; Display the avalanche and guarantee that the border of the avalanche
is updated
    ask avalanche-patches [ recolor ask neighbors [ recolor ] ]
    display
    ;; Erase the avalanche
    ask avalanche-patches [ set base-color default-color recolor ]
    ;; Updates the average monitor
```

```
    set total-on-tick total
    tick
  ]
end

to explore
  ifelse mouse-inside? [
    let p patch mouse-xcor mouse-ycor
    set selected-patch p
    ask patches [ push-n ]
    ask selected-patch [ update-n 1 ]
    let results stabilize false
    ask patches [ pop-n ]
    ask patches [ set base-color default-color recolor ]
    let avalanche-patches first results
    ask avalanche-patches [ set base-color selected-color recolor ]
    display
  ] [
    if selected-patch != nobody [
      set selected-patch nobody
      ask patches [ set base-color default-color recolor ]
    ]
  ]
end

;; Stabilizes the sandpile. Reports which sites fired and how many iterations
it took to
;; stabilize.
to-report stabilize [animate?]
  let active-patches patches with [ n > 7 ]

  ;; The number iterations the avalanche has gone for. Use to calculate
lifetimes.
  let iters 0

  ;; we want to count how many patches became overloaded at some point
```

```
  ;; during the avalanche, and also flash those patches. so as we go, we'll
  ;; keep adding more patches to to this initially empty set.
  let avalanche-patches no-patches

  while [ any? active-patches ] [
    let overloaded-patches active-patches with [ n > 7 ]
    if any? overloaded-patches [
      set iters iters + 1
    ]
    ask overloaded-patches [
      set base-color fired-color
      ;; subtract 8 from this patch
      update-n -8
      if animate? [ recolor ]
      ;; edge patches have less than four neighbors, so some sand may fall
off the edge
      ask neighbors [
        update-n 1
        if animate? [ recolor ]
      ]
    ]
    if animate? [ display ]
    ;; add the current round of overloaded patches to our record of the
avalanche
    ;; the patch-set primitive combines agentsets, removing duplicates
    set avalanche-patches (patch-set avalanche-patches overloaded-patches)
    ;; find the set of patches which *might* be overloaded, so we will check
    ;; them the next time through the loop
    set active-patches patch-set [ neighbors ] of overloaded-patches
  ]
  report (list avalanche-patches iters)
end

;; patch procedure. input might be positive or negative, to add or subtract
sand
to update-n [ how-much ]
```

```
  set n n + how-much
  set total total + how-much
end

to-report drop-patch
  if drop-location = "center" [ report patch 0 0 ]
  if drop-location = "random" [ report one-of patches ]
  if drop-location = "mouse-click" and mouse-down? [
    every 0.3 [ report patch mouse-xcor mouse-ycor ]
  ]
  report nobody
end

;; Save the patches state
to push-n ;; patch procedure
  set n-stack fput n n-stack
end

;; restore the patches state
to pop-n ;; patch procedure
  ; need to go through update-n to keep total statistic correct
  update-n ((first n-stack) - n)
  set n-stack but-last n-stack
end
```

Appendix B Code for Data Preparation and Sampling from Sandpile Simulations

```
import numpy as np
import matplotlib.pyplot as plt

class Sandpile():
    def __init__(self, arr = None ,rows = 3, cols = 3, max_sand = 8):
        """
        arr - 2d array of values
        rows - height of sandpile
        cols - width of sandpile
        max_sand - max count of sandpile grains (must be div by 8)
        """
        self.max_grains = max_sand
        if arr == None:
            self.rows = rows
            self.cols = cols
            self.grid = np.zeros((cols,rows), int)
        else:
            self.rows = len(arr)
            self.cols = len(arr[0])
            self.grid = np.array(arr)

    def topple(self, elem_x, elem_y):
        p = self.grid[elem_x, elem_y]
        b = p // self.max_grains
        o = p % self.max_grains
        self.grid[elem_x, elem_y] = o

        # increase height of neighbor piles
        self.grid[elem_x-1, elem_y] += b
        self.grid[elem_x+1, elem_y] += b
        self.grid[elem_x, elem_y-1] += b
        self.grid[elem_x, elem_y+1] += b
```

```
        self.grid[elem_x-1, elem_y-1] += b
        self.grid[elem_x+1, elem_y+1] += b
        self.grid[elem_x+1, elem_y-1] += b
        self.grid[elem_x-1, elem_y+1] += b

        # border
        self.grid[0] = self.grid[-1] = 0
        self.grid[:, 0] = self.grid[:, -1] = 0



    def run(self):
        #from time import time

        #start_time = time()
        iterations = 0
        topple = self.topple
        where = np.where

        while np.max(self.grid) >= self.max_grains:
            elem_x, elem_y = where(self.grid >= self.max_grains)
            topple(elem_x, elem_y)
            iterations += 1

        #print("--- %d iterations %s seconds ---" % (iterations, time() -
start_time))

    def get_pile(self):
        return self.grid

    def set_sand(self, x, y, number):
        self.grid[x,y] = number

    def __add__(self, other):
        result = Sandpile(rows = self.rows, cols = self.cols)
        try:
            result.grid = self.grid + other.grid
```

```
            return result.run()
        except ValueError:
            print("ValueError: sandpile grid sizes must match")

    def show(self):
        """
        plot sandpile and/or save it in the file
        save - true = save picture, false = dont save picture
        filename - name of the file, where would be picture of sandpile
        """

        heatmap = plt.pcolor(self.grid)
        plt.axis('off')
        plt.imshow(self.grid)
        plt.colorbar(heatmap, ticks=range(self.max_grains))
        #if save:
            #plt.savefig(filename, bbox_inches='tight')
        plt.show()

    #def save(self, filename = "sandpile.png"):
        #from PIL import Image
        #colors = [(255, 255, 0), (0, 185, 63), (0, 104, 255), (122, 0, 229)]
        #img = Image.fromarray(color_grid(self.grid, colors), "RGB")
        #img.save(filename)

def color_grid(grid, colors):
    new_grid = np.zeros((len(grid), len(grid[0]), 3), dtype=np.uint8)
    for i in range(len(new_grid)):
        for j in range(len(new_grid[0])):
            new_grid[i, j] = colors[grid[i, j]]
      return new_grid
if __name__ == '__main__':
    weights = [11,10,9,8,7,6,5,4,3,2,1]
    result = []
    number = int(input("input the number of sands:"))+1
    pile = Sandpile(rows = 21, cols = 21)
```

```
    for i in range(number):
        sum_weights = 0
        if i == 0:
            continue
        pile.set_sand(10 , 10, i)
        pile.run()
        for y in range(21):
            for x in range(21):
                #coordinate (abs(x-10),abs(y-10))
                dis = max(abs(x-10),abs(y-10))
                sum_weights += pile.grid[y][x]*weights[dis]
        print("The number of sands is "+str(i)+", the sum of weights is "+str(sum_weights))
        result.append((i,sum_weights))
        if i%8==0 or i==number-1:
            pile.show()
        pile.grid = np.zeros((21,21), int)
    print(result)
```

Appendix C Code for Choosing Time Delay

```
def result_to_seq(result):
    # The result is a list of tuples, this function will return a sequence
of weights
    weights=[]
    for item in result:
        weights.append(item[1])

    return weights

def compute_R(seq, tau):
    # Compute the R
    mean = np.mean(seq)
    N = len(seq)
    total_sum = 0
    for i in range(len(seq)-tau):
        total_sum = total_sum + (seq[i]-mean)*(seq[i+tau]-mean)

    return total_sum/N

def compute_tau(result):
    seq = result_to_seq(result)
    tau = 0
    R_0 = compute_R(seq, tau)

    while True:
        tau = tau + 1
        R_next = compute_R(seq, tau)
        if R_0*(1-1/math.e) >= R_next:
            break

    return tau

def convert_seq_to_point(seq, dim, tau):
    weights = result_to_seq(seq)
```

```
    points = []
    for i in range(len(weights)):
        point = []
        for j in range(dim):
            index = i + j * tau
            if index < len(weights):
                point.append(weights[i+j*tau])

        if len(point) == dim:
            points.append(point)

  return np.array(points)
tau = compute_tau(result)
print(tau)
```

Appendix D Code for Determining the dimension of reconstructed phase space

```
# compute the dimension m
from scipy.spatial.distance import pdist

def convert_seq_to_point(seq, dim, tau):
    weights=[]
    for item in seq:
        weights.append(item[1])



    points = []
    for i in range(len(weights)):
        point = []
        for j in range(dim):
            index = i + j * tau
            if index < len(weights):
                point.append(weights[i+j*tau])

        if len(point) == dim:
            points.append(point)

    return np.array(points)



def Cmr(points, r):
    #compute the pair distance
    pair_distance = pdist(points, 'euclidean')

    N = len(points)

    diff = r - pair_distance
    Cmr = (1/(N**2))*np.sum(np.absolute(np.heaviside(diff, 0)-1))
```

```
    return Cmr



def D2mr(points, r):
    delta = 0.01

    bottom = math.log((r+delta)/r)
    top = Cmr(points, r+delta) - Cmr(points, r)

    slope = top/bottom

    return slope

def compute_dim(result, tau):
    r = float(input("input the radius r:"))
    m = 2
    points = convert_seq_to_point(result, m, tau)
    D2 = D2mr(points, r)
    print(D2)
    if math.isnan(D2) or D2:
        return m
    else:
        while True:
            m = m + 1
            points = convert_seq_to_point(result, m, tau)
            new_D2 = D2mr(points, r)
            print(new_D2)
            if math.isnan(new_D2):
                break
            elif (new_D2-D2)<0.01:
                break

        return m
compute_dim(result, tau)
```

Appendix E Code for Determining the Lyapunov Exponent

a. Original Code for Lyapunov Exponent Solver:

```
def convert_seq_to_point(seq):
    weights=[]
    for item in seq:
        weights.append(item[1])

    dim = int(input("input the dimension:"))
    time_delay = int(input("input the time_delay:"))

    points = []
    for i in range(len(weights)):
        point = []
        for j in range(dim):
            index = i + j * time_delay
            if index < len(weights):
                point.append(weights[i+j*time_delay])

        if len(point) == dim:
            points.append(point)

      return np.array(points)

def time_forward(seq):
    points = convert_seq_to_point(seq)
    pair_distance = squareform(pdist(points, 'euclidean'))
    threshold = float(input("input the threshold:"))
    bound = len(points)-1
    t = 0
    distance = pair_distance[t]
    min_dis = np.partition(distance, 1)[1]
    pair = np.where(distance == min_dis)[0][-1]
    ratio_list = np.empty(0)
    count = 0
```

```
    while True:
        t = t + 1
        pair = pair + 1
        count = count + 1
        if (t>bound) or (pair>bound):
            break
        min_for_dis = pair_distance[t][pair]
        ratio = min_for_dis/min_dis
        if ratio <= threshold:
            ratio_list = np.append(ratio_list, ratio)
            avg_log_ratio = np.sum(np.log(ratio_list))/count
            print('forward time: '+ str(count)+', the average of log ratio:
'+str(avg_log_ratio))
            min_dis = min_for_dis
        else:
            t = pair
            distance = pair_distance[t]
            min_dis = np.partition(distance, 1)[1]
            pair = np.where(distance == min_dis)[0][-1]
time_forward(result)
```

b. Test code for validating correctness of LE solver with another set of data of similar size/structure that known to be pure stochastic

```
def time_forward(seq):

    points = convert_seq_to_point(seq)

    pair_distance = squareform(pdist(points, 'euclidean'))
    threshold = float(input("input the threshold:"))
    bound = len(points)-1
    t = 0
    distance = pair_distance[t]
    min_dis = np.partition(distance, 1)[1]
```

```
    pair = np.where(distance == min_dis)[0][-1]
    ratio_list = np.empty(0)
    count = 0

    while True:
        t = t + 1
        pair = pair + 1
        count = count + 1
        if (t>bound) or (pair>bound):
            break
        min_for_dis = pair_distance[t][pair]
        ratio = min_for_dis/min_dis
        if ratio <= threshold:
            ratio_list = np.append(ratio_list, ratio)
            avg_log_ratio = np.sum(np.log(ratio_list))/count
            print('forward time: '+ str(count)+', the average of log ratio: 
'+str(avg_log_ratio))
            min_dis = min_for_dis
        else:
            t = pair
            distance = pair_distance[t]
            min_dis = np.partition(distance, 1)[1]
            pair = np.where(distance == min_dis)[0][-1]

periodic_values = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100]

# Assuming the original sequence has indices from 1 to n (or more)
n = 500  # Change this as per the actual number of tuples you need

# Generate seq2 with periodic second values
seq2 = [(i, periodic_values[(i - 1) % len(periodic_values)]) for i in range(1, 
n + 1)]

random_numbers = np.random.normal(loc=1000, scale=100, size=500)  # Mean=1000, 
SD=10
# Assuming the original sequence has indices from 1 to n (or more)
```

```
n = 500  # Change this as per the actual number of tuples you need
# Create seq3
seq3 = [(i + 1, num) for i, num in enumerate(random_numbers)]

# Print the first few elements to verify
print(seq3[:10])  # Display the first 10 tuples

result=seq3
time_forward(result)
```

```
forward time: 1361050, the average of log ratio: 0.011986359958694516
forward time: 1361052, the average of log ratio: 0.011986599410820477
forward time: 1361075, the average of log ratio: 0.011986209347962741
forward time: 1361087, the average of log ratio: 0.011986395001526892
forward time: 1361088, the average of log ratio: 0.011986366881666087
forward time: 1361089, the average of log ratio: 0.011986273519820883
forward time: 1361092, the average of log ratio: 0.01198618444397359
forward time: 1361108, the average of log ratio: 0.011986208040766716
forward time: 1361109, the average of log ratio: 0.011986018893474113
forward time: 1361117, the average of log ratio: 0.011986204849866807
forward time: 1361118, the average of log ratio: 0.011986299617412332
forward time: 1361122, the average of log ratio: 0.011986314078020048
forward time: 1361125, the average of log ratio: 0.01198625520936866
forward time: 1361142, the average of log ratio: 0.011986360015907006
forward time: 1361144, the average of log ratio: 0.011986599451848266
forward time: 1361167, the average of log ratio: 0.011986209415353818
forward time: 1361179, the average of log ratio: 0.011986395056369335
forward time: 1361180, the average of log ratio: 0.011986366938409068
forward time: 1361181, the average of log ratio: 0.011986273582873996
forward time: 1361184, the average of log ratio: 0.011986184513047041
forward time: 1361200, the average of log ratio: 0.011986208108244509
forward time: 1361201, the average of log ratio: 0.011986018973735826
forward time: 1361209, the average of log ratio: 0.011986204917559818
forward time: 1361210, the average of log ratio: 0.011986299678700246
forward time: 1361214, the average of log ratio: 0.011986314138330437
```

```
forward time: 1361217, the average of log ratio: 0.011986255273657646
forward time: 1361234, the average of log ratio: 0.011986360073111765
forward time: 1361236, the average of log ratio: 0.011986599492870507
forward time: 1361259, the average of log ratio: 0.011986209482735787
forward time: 1361271, the average of log ratio: 0.011986395111204366
forward time: 1361272, the average of log ratio: 0.011986366995144377
```

it turned out the LE always stay positive, which indicates divergence. In other words, the dynamical system is stochastic. Considering the seq3 is generated by randomizer (`random_numbers = np.random.normal(loc=1000, scale=100, size=500)`) . The algorithm managed to recognize the system stochastic. Hence, the code is correct.

Appendix F Code for Locating Attractor

```
def compute_lenint(num_intervals, points):
    max_cor = np.max(points)
    len_interval = math.ceil(max_cor/num_intervals)

    return len_interval

def cube_pos(num_intervals, points):
    len_interval = compute_lenint(num_intervals, points)
    cube_pos = {}

    for x in range(num_intervals):
        for y in range(num_intervals):
            for z in range(num_intervals):
                pos = str(x) + str(y) + str(z)
                cube_pos[pos]  =  (np.array([x*len_interval,  y*len_interval,
z*len_interval]),     np.array([(x+1)*len_interval,     (y+1)*len_interval,
(z+1)*len_interval]))
    return cube_pos

def cube_center(num_intervals, cube_pos):
    cube_center = {}

    for x in range(num_intervals):
        for y in range(num_intervals):
            for z in range(num_intervals):
                pos = str(x) + str(y) + str(z)
                cube_center[pos] = (cube_pos[pos][0] + cube_pos[pos][1])*0.5
    return cube_center

def cube_count(num_intervals, points):
    len_interval = compute_lenint(num_intervals, points)
    cube_count = {}
    for x in range(num_intervals):
        for y in range(num_intervals):
```

```
            for z in range(num_intervals):
                pos = str(x) + str(y) + str(z)
                cube_count[pos] = 0

    pos_list = np.floor(points/len_interval).astype(int)
    pos_list[pos_list == num_intervals] = num_intervals - 1
    for item in pos_list:
        pos = str(item[0]) + str(item[1]) + str(item[2])
        cube_count[pos] = cube_count[pos] + 1

    return cube_count

def point_location(num_intervals, point):
    len_interval = compute_lenint(num_intervals, points)
    pos_list_cor = np.floor(np.array(point)/len_interval).astype(int)
    pos_list_cor[pos_list_cor == num_intervals] = num_intervals - 1
    pos = str(pos_list_cor[0]) + str(pos_list_cor[1]) + str(pos_list_cor[2])

    return pos

def compute_center(cube_count, cube_pos, attractor):

    weight_sum = np.array([0,0,0])
    count_sum = 0
    for pos, count in cube_count.items():
        if count >= attractor:
            weight_sum              =              weight_sum              +
count*(cube_pos[pos][0]+cube_pos[pos][1])/2
            count_sum = count_sum + count
    center = weight_sum/count_sum

    return center

def compute_dis(point, num_intervals, cube_count, cube_pos, attractor):
    point_pos = point_location(num_intervals, point)
    if cube_count[point_pos] >= attractor :
```

```
        return 0
    else:
        center = compute_center(cube_count, cube_pos, attractor)
        dis = math.sqrt(np.sum((center - np.array(point))**2))
        return dis

num_intervals = int(input("input the number of intevals: "))

len_edge = compute_lenint(num_intervals, points)
cube_count = cube_count(num_intervals, points)
cube_pos = cube_pos(num_intervals, points)
cube_center = cube_center(num_intervals,cube_pos)

pos_list = []
for (pos, freq) in cube_count.items():
    if freq > 0:
        pos_list.append(pos)
output_dict = {}
for pos in pos_list:
    center_coor = tuple(cube_center[pos])
    freq = cube_count[pos]
    output_dict[pos]= (center_coor, freq)
output_dict
# the index mean the order: (center position, frequency)

#data preparation
cube_center_freq = []
for x in range(num_intervals):
        for y in range(num_intervals):
            for z in range(num_intervals):
                pos = str(x) + str(y) + str(z)
                if cube_count[pos] != 0:
                    center = cube_center[pos].tolist()
                    freq = cube_count[pos]
                    center.append(freq)
                    cube_center_freq.append(center)
```

```
import pandas as pd
df = pd.DataFrame(cube_center_freq, columns =['x', 'y', 'z', 'freq'], dtype
= float)
df_max_scaled = df.copy()

# apply normalization techniques on Column 1
column = 'freq'
df_max_scaled[column]                =                df_max_scaled[column]
/df_max_scaled[column].abs().max()
print(df)
def color_map_color(value, cmap_name='coolwarm', vmin=0, vmax=1):
    # norm = plt.Normalize(vmin, vmax)
    norm = matplotlib.colors.Normalize(vmin=vmin, vmax=vmax)
    cmap = cm.get_cmap(cmap_name)  # PiYG
    rgb = cmap(norm(abs(value)))[:3]  # will return rgba, we take only first
3 so we get rgb
    color = matplotlib.colors.rgb2hex(rgb)
      return color
import matplotlib
from matplotlib import cm
from matplotlib.colors import Normalize
def get_cube():
    phi = np.arange(1,10,2)*np.pi/4
    Phi, Theta = np.meshgrid(phi, phi)

    x = np.cos(Phi)*np.sin(Theta)
    y = np.sin(Phi)*np.sin(Theta)
    z = np.cos(Theta)/np.sqrt(2)
    return x,y,z



fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
L = len_edge
```

```
for i in df.index:
    x,y,z = get_cube()

    # Change the centroid of the cube from zero to values in data frame
    x = x*L + df.x[i]
    y = y*L + df.y[i]
    z = z*L + df.z[i]
    ax.plot_surface(x, y, z, color=color_map_color(df_max_scaled.freq[i]),
                      linewidth=0,     antialiased=False,     shade=True,
alpha=0.75)
    ax.set_zlabel("z")

plt.xlabel("x")
plt.ylabel("y")
plt.show()
```

Appendix G Code for Computing anm of D_{Koch} and D_{Li}

```
attractor = int(input("input the selected frequency: "))
dis = compute_dis([841,841,841], num_intervals, cube_count, cube_pos,
attractor)
print("the distance is", dis)

attractor = int(input("input the selected frequency: "))
dis = compute_dis([497,497,497], num_intervals, cube_count, cube_pos,
attractor)
print("the distance is", dis)
```

Appendix H Code for Load Testing of Li’s data standard and Koch’s data standard

```
import time
from xml.etree import ElementTree
from copy import deepcopy
from memory_profiler import profile

file1_path = "./koch xml.xml"



# @profile
def main():
    start = time.time()
    tree = ElementTree.parse(file1_path)
    # root = tree.getroot()
    ring_node = tree.find(".//ringsegmentelement")
    ring_node_father = tree.find("element")
    for i in range(335, 435):
        new_node = deepcopy(ring_node)
        new_node.set("attached_tunnelelement_id", str(i))
        ring_node_father.append(new_node)

    tree.write("./koch xml_3.xml")
    print("k Consumetime", time.time() - start)



main()

file1_path = "./Li segmenttunnelelemtn xml.xml"



# @profile
def main():
    start = time.time()
```

```
    tree = ElementTree.parse(file1_path)
    father = tree.getroot()
    node = tree.find(".//Segmentunnelelement")
    for i in range(335, 435):
        new_node = deepcopy(node)
        new_node.set("id", str(i))
        father.append(new_node)

    tree.write("./Li segmenttunnelelemtn xml_3.xml")
    print("L 消耗时间", time.time() - start)


main()

import sys
import time
from memory_profiler import profile
from xml.etree import ElementTree

file1_path = "./koch xml.xml"
file2_path = "./Li segmenttunnelelemtn xml.xml"


class TailRecurseException(BaseException):
    def __init__(self, args, kwargs):
        self.args = args
        self.kwargs = kwargs



def tail_call_optimized(g):
    """
    重复使用栈顶 2 个数据，以解决递归溢出的问题
    """
    def func(*args, **kwargs):
        f = sys._getframe()
        if f.f_back and f.f_back.f_back and f.f_back.f_back.f_code ==
```

```
f.f_code:
            raise TailRecurseException(args, kwargs)
        else:
            while 1:
                try:
                    return g(*args, **kwargs)
                except TailRecurseException as e:
                    args = e.args
                    kwargs = e.kwargs

    func.__doc__ = g.__doc__
    return func



# @profile
def loops_parse(tree):
    """
    :param node:
    :return:
    """
    for node in tree.findall(".//*"):
        s = node.tag
        # print(s)

    return



# @profile
def timer(file_path):
    """
    统计遍历所有子节点的时间
    :param file_path:
    :return:
    """
    start = time.time()
    tree = ElementTree.parse(file_path)
```

```
    # root_node = tree.getroot()
    loops_parse(tree)
    return time.time() - start


# @profile
def main():
    t = timer(file1_path)
    print(t)


main()
```