Advanced Engineering Informatics 57 (2023) 102108

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

# Sandpile-simulation-based graph data model for MVD generative design of shield tunnel lining using information entropy

Yi An [a], Xuhui Lin [b], Haijiang Li [a,*], Yitong Wang [c]

[a] Cardiff School of Engineering, Cardiff University, United Kingdom
[b] The Bartlett School of Sustainable Construction, University College London, United Kingdom
[c] China International Engineering Consulting Corporation, China

## ABSTRACT

BIM standard development is central to the performance and behavior of BIM model application across transmission, visualization, and information management perspectives. Tremendous effort has been made to ease the implementation of IFC data model in practice. Yet, the complexity of IFC data model hurdles the implementation of the import and export functionality by software vendors. To overcome this, buildingSMART introduced the concept of Model View Definitions to define which parts of an IFC data model need to be implemented for a specific data exchange scenario. With such, the certification of compatibility for software products with the IFC standard is formed. The Model View Definition is use case orientated to determine whether the specific information should be included in an IFC partial model. With the creation of ad-hoc, project-specific Exchange Requirements increasing, associated MVD development requires much more work to incorporate standard development. To resolve this issue, this paper attempts to exploit the potential of information entropy which has proven itself extremely crucial in many other industries in terms of information management, and then integrates it with sandpile simulation to propose a Top-down hierarchy to structure as well as interpret IFC partial model via Model View Definition. The proposed information entropy shifted MVD development approach would manage to unify the MVD development process that enables the reduction on confusion for various end users, specific organization, or project needs. Moreover, to better translate the BIM standard topology into sandpile simulations, a new notion system is proposed. Sandpile simulations are further implemented to prove their applicability, during the simulation, self-organized criticality is identified, and the existence of chaos is observed.

## 1 Introduction

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 [14]. 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 the Industry Foundation Classes (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. 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 [29]. 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. However, existing research in this domain often overlooks the potential application of information entropy, a concept derived from information theory, in optimizing the generative design of MVDs for shield tunnel linings. The inclusion of information entropy in the generative design process of MVDs for shield tunnel linings remains relatively unexplored. This research paper aims to explore and propose a novel approach that incorporates information entropy into the generative design of MVDs for shield tunnel linings. Information entropy provides a quantitative measure of the uncertainty and complexity within a system. By applying information entropy measures to the generative design process of MVDs, this research intends to identify and prioritize the most informative and critical subsets of the building model. This incorporation of information entropy can aid in the selection and configuration of entities for MVDs that optimize the representation of

---

* Corresponding author at: Cardiff School of Engineering, Cardiff University, Queen's Building, The Parade, Cardiff CF24 3AA, Wales, United Kingdom.
*E-mail address:* LiH@cardiff.ac.uk (H. Li).

https://doi.org/10.1016/j.aei.2023.102108
Received 10 November 2022; Received in revised form 14 June 2023; Accepted 14 July 2023

shield tunnel lining designs. By leveraging the principle of ID3 integrating information entropy and sandpile simulation, this study seeks to enhance the efficiency and effectiveness of the generative design process, resulting in optimized MVDs that better capture the essential design information for shield tunnel linings.

The key objectives of this research paper are as follows: a. develop an approach that integrates information entropy into the generative design process of MVDs for shield tunnel linings; b. apply information entropy measures to evaluate and select the most informative subsets of the building model for inclusion in the MVDs; c. Explore the practical implementation of the proposed approach in generating optimized MVDs for shield tunnel linings, considering various design considerations and constraints. Through this research, we aim to contribute to the advancement of MVD design methodologies specifically tailored for shield tunnel linings. By integrating information entropy into the generative design process of MVDs, this study seeks to enhance the design of MVD and reduce the ambiguity generated during MVD development. The outcomes of this research can inform engineers and scholars in making more informed decisions, leading to more efficient and sustainable BIM development for shield tunnel lining. This paper employs the information entropy to quantify the information mass required for tunneling IFC model then interpret the topology of IFC model to develop the associate MVD with sandpile simulation. Section 2 conducts a critical literature review regarding MVD development and information theory application in many other information related domains to reveal the rationale of employing information theory in MVD development so that research gap can be uncovered. Section 3 provides the research assumption, then justify the methods, concepts, models those used in the paper before introducing the whole research framework at the end of section; section 4 introduces the illustrative example about the simulation and proposes a mapping system to covert the sandpile simulation result to associate MVD development, followed by the simulation results and observations generated from a multi-agent programmable modeling environment – Netlogo; section 5 discuss the findings, contributions, and limitations of the research, later point out the future directions; at the end, the conclusions is given in section 6.

## 2. Literature review

The construction industry has witnessed a paradigm shift with the introduction of Building Information Modeling (BIM), revolutionizing the way projects are planned, designed, and executed. Within the BIM framework, the Industry Foundation Class (IFC) has emerged as a widely adopted data model for representing and exchanging information throughout the project lifecycle. However, as much convenient as the IFC can be, the process of establishing IFC is lengthy. While the current IFC4 can be considered largely mature and ready for use as a standard, it can only be used in practice once the different software vendors have implemented it as an import and export interface. The quality of the implementation of such interfaces is crucial for its take-up in the industry. In the past, errors in these import and export modules led to data errors or even data loss, impacting on the reputation and market acceptance of the IFC data format. One reason for the inadequate implementation of the import and export functionality by software vendors is also the complexity of the IFC data model. For example, it is possible to represent 3D geometry in an IFC model in several different ways. For software vendors, this means that they need to support all geometric representation methods to offer full IFC compatibility, which is an immense implementation task. To overcome this hurdle, buildingSMART introduced the concept of Model View Definitions (MVD) that define which parts of an IFC data model need to be implemented for a specific data exchange scenario [4]. This section aims to firstly critically analyze the existing literature, providing a holistic understanding of the significance of MVDs in facilitating efficient communication, collaboration, and decision-making in the construction industry. By examining the available literature, the research would identify drawbacks of MVD currently. Then the review for information theory application on other information related domains would be 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.

### 2.1. Review on MVD

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 [12–13]. 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. A 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 [32]. The relations among IFC, MVD, IDM and International Framework for Dictionaries (IFD) is given in Fig. 1 [5]. 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.

These are only the buildings/infrastructures BIM related above the ground. While it comes to the 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. 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 [11]. 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 [1].

Thus far, throughout IFC/IDM/MVD application in buildings/infrastructures to underground facilities, it could find that there have been several successful attempts to (a) demonstrate the need of a comprehensive, systematic, robust tunnel information model, (b) automate the

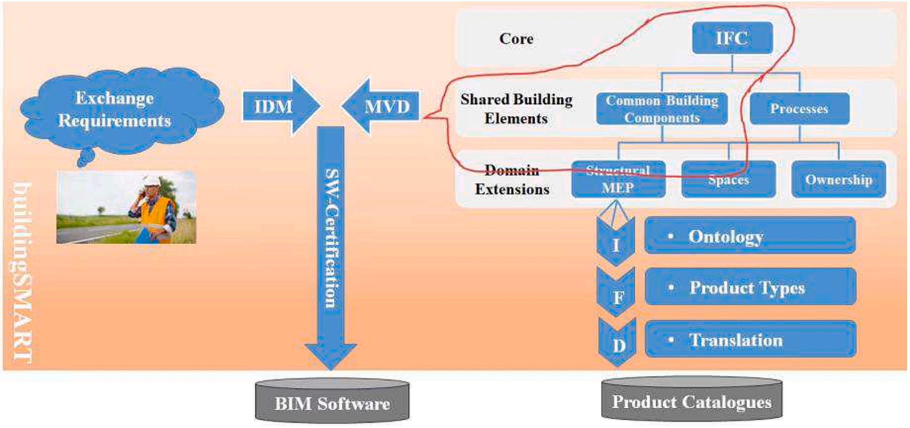


**Fig. 1.** The relationship between IDM, MVD, IFC and IFD.

link from information model to all relevant information that could meet the requirements of the mechanized tunnel projects planning, design, construction and maintenance, and (c) develop a comprehensive applicable tunnel 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 extending the existing IFC4 for fitting specific use-case scenarios is diverse. Within existing buildingSMART approved framework, there are 3 common approaches among Lods, MVDs and linked data to choose. 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 tunnel ling 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 [43]. Without a well a-priori agreement among implementing manufacturers, the standard would not see a more drawn-out process on the market [30]. As aforementioned, in the case of tunnel engineering, 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 IFC, 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 tunneling projects, 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 like tunneling. Hence, despite the information requirements remain same, different MVDs have been developed, not even mention the lack of comprehensive and unified information requirements for tunneling projects so far. These emerge the needs to find a proper way to derive consistent, comprehensive, and unified MVD for tunneling projects that can reduce ambiguity generated by information requirements in IDM.

### 2.2. Review on information theory and its application

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 [23]. 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 [42,44,52], 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. [38,9,26,28;38,51] 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 [19]. 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 [10]. 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 [6]. 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 [53]. Preston builds an framework based on information theory to motivate child welfare case managers [39]. 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 [40]. 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 [36]. 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 [34]. Burnham and Anderson used Shannon's information theory/ the quantification of information to conduct statistical inference more accurately [3], Delgado-Bonal and Martic-Torres applied it on the human vision research [20], Passalis and Tefas optimized information retrieval with information theory [37], and Gerhing et al. utilized it on quantum computing very recently [25]. 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 [41].

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 [38]. 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 [9]. 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 [38]. 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 [28]. 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 [26,51].

### 2.3. *Rationale of employing information theory in MVD and identifying research gap*

As aforementioned in section 2.1, 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 tunnleing 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 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 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 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. This research gap calls for an investigation into how information theory 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 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.

## 3. Methodologies

The study tries to resolve the problem with the help of ID3, sandpile simulation, and information theory, which 3 concepts might be commonly used for some related research separately but not together. Hence, it is important to build a comprehensive methodological framework so that each of these research methods would be rationale to collaborate with each other for solving the research problem. Like all methodological framework, the fundamentals to them are the assumptions those set the basic propositional logic for the whole system. 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 and secondly providing analogical inference in the subsections. Since the objective is developing proper MVDs for tunnel lining while reducing ambiguity, information entropy could be brought in thanks to assumption i and ii as its nature of quantifying information mass. 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. Another ambiguity source contributing to MVD development is the way to define entities. This is dealt by using ID3 thinking and sandpile simulation, the ID3 thinking firstly manages to convert the MVD to a tree structure (directed weighted tree diagram). And in graph theory, the directed weighted tree diagram could be described with more features, in other words, more detailed regulated. This could, in a way, reduce the ambiguity. And then the sandpile simulation is used to eliminate most ambiguity on the basis of comprehending the assumption i, that is indicating the MVD development system is a dynamic system who driven force to further MVD supposed to the information mass. And with ID3 thinking, considering the end product of MVD as a tree structure, with the analogical inference of sandpile simulation (provided in section 3.2.2, Fig. 8), the dynamical system of sandpile simulation would be used to represent the dynamical system of MVD development. Through this, ambiguity could be reduced due to the nature of sandpile simulation, since the simulated layout of lattice must follow the least action principle, which indicates that, given the information mass, the dynamical system can only possess one least action state which is also the stationary state. The eliminate other alternatives of MVD topological structure, thereby the ambiguity is eliminated. Owing to these, the methodological framework is bult. The whole section is about to explain why these would be involved and how MVD, ID3, information entropy, sandpile simulation are collaborated to target at the research problem.

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

i. 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.

ii. 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.

iii. Topology of IFC schema, in this paper, refers to the tree structure which is consist of nodes and branches. While regardless of semantics, the entities are the “nodes” of the tree, and the relations are the “branches” of the tree.

iv. 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. According to the sandpile simulation, the avalanche point is when the value of entropy reaches certain limits, the node splits. It is worth noticing that apart from the extensive information case, there is another scenario with extreme condition, while the information within BIM model is extreme limit, such as only 1 or 2 entries of information, there should be no need to extend the schema as the applicable model would only hold the instances rather than embedding a complicated schema. Although such cases are nearly impossible, the principles are still consistent with the assumption.

In a real-world BIM model setting, the assumption (i) is very near to reality, take the Koch’s IFC model for Shield Tunnel and Li’s IFC model for shield Tunnel [27,29]. 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. Assumption (ii) matches with practice, the information is considered as the independent and identically distributed while the Modeling Support Group of buildingSMART makes great deal of effort to collect feedback from different user groups (active users, software developers and researchers) for developing MVDs [30]. Though, the feedback covers all the aspects, but there is no weight distribution considered in their feedback, hence the assumption (ii) is the practice. The hypothesis (iii) The assumption 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 Fig. 2. 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 [4,24]. And its rationale again is mentioned above in assumption (iv).

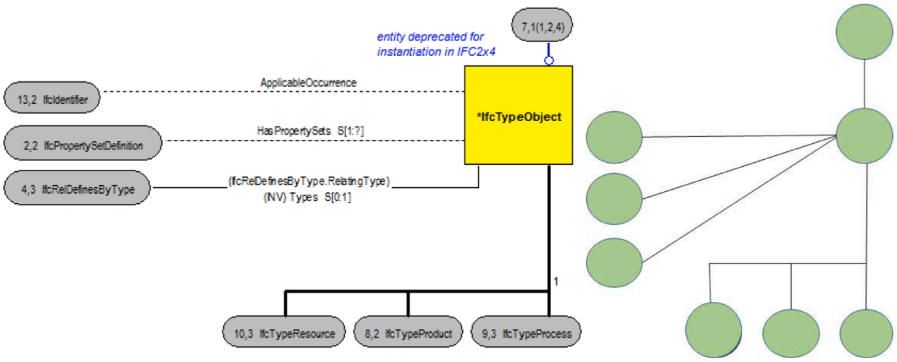
The assumption (iv) is, in essence, an abstract form in theoretical level of BIM-based thinking in practical level. While information of a building element entity is much, reorganizing and restructuring the way of expressing information is considered as better option in terms of efficiency in management. For instance, a retaining wall design, without assumption (iv) it would be organized as Fig. 3a and with assumption (iv) it would be organized as Fig. 3b. It can obviously be seen that even with information omitted, the way organizing the information as Fig. 2b is much better than the one shows in Fig. 2a. Hence, as stated in assumption, the bigger information mass it is, more splitting it will be. Therefore, the assumption (iv) is proposed reasonably.

Fig. 4 shows the methodology framework for developing the new MVD. The proposed framework consists of three phases: (1) Identification of information requirements in tunnel lining, (2) Tree-structured graph model, (3) Data transformation for MVD generation.

1. The required information for designing tunnel linings is identified with reference to communications with practioners as well as published building regulations (such as precast specifications, document-based design guidebooks, code of practices, etc.). The information requirements shall include spatial, topological, geometric, and semantic aspects, which are relevant to the design and construction of tunnel linings. IFC MVD is established to provide a structured overview of the required information in the BIM environment taking into consideration the existing MVD.
2. Following this, the information mass is calculated based on the IDM/MVD information requirements using Shannon’s information entropy. Then the entropy value would be imported to sandpile simulation helping determination of sand grains, such as 1 lining segment contains 100 bits, the initial setting for amount of sand grains needed to drop would be set as 100. Thereafter, working with assumption iv, complete the initial settings by defining toppling/node splitting rules. Directly after, the simulation is conducted. Simulation results consist of layout of lattice, log–log plot for avalanche scalability and occurrence. Successively, to interpret the simulation results (layout of sandpile lattice), a mapping system is proposed to map layout sandpile lattice to a tree structure which can stands for the topology of IFC schema. As a result, the tree-structured graph model is generated.
3. Lastly, the tree-structured graph model only contains the topological structure of the IFC schema rather than all required information to generate MVD. Hence, the data transformation using the information entropy obtained from the first phase and the IDM of tunnel lining is used to convert the tree-structured graph model into a new MVD.

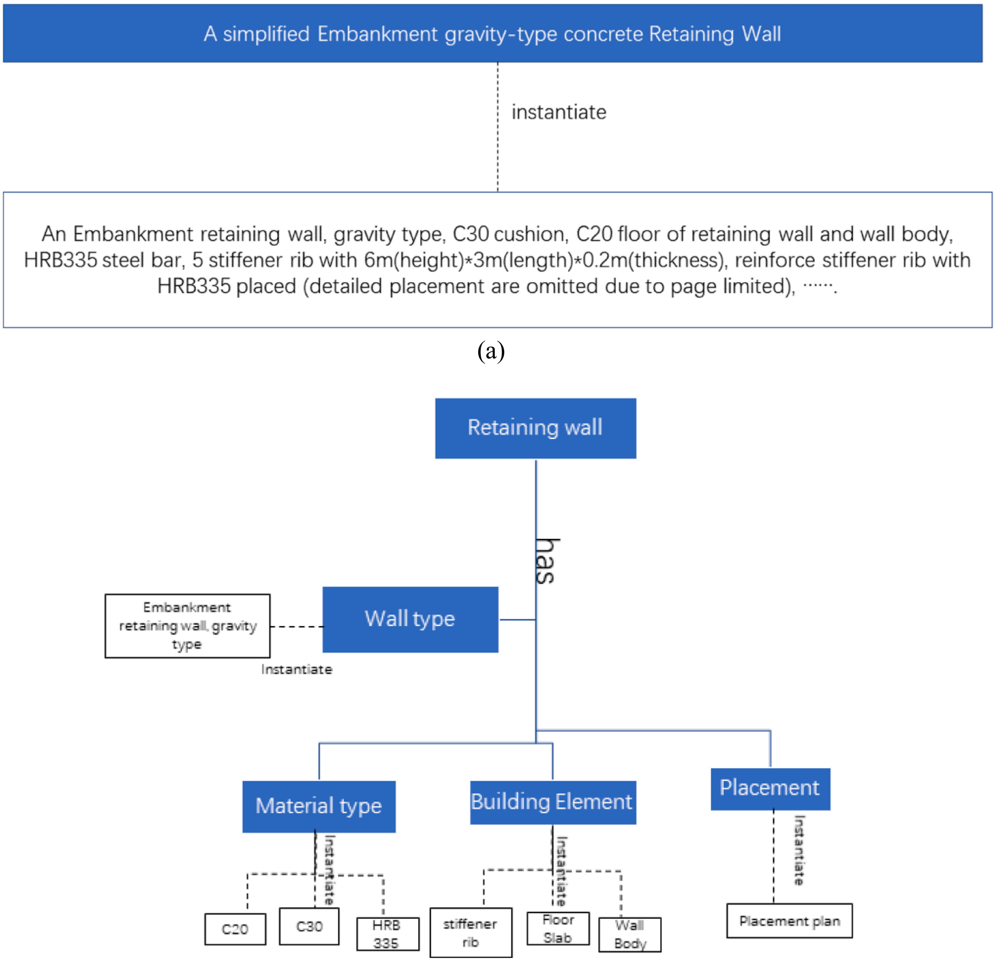
Philosophy of involving ID3, information theory and sandpile simulation in MVD development

Besides aforementioned in section 2.3, the philosophy of involving information theory, ID3 and sandpile simulation in MVD development is mainly on basis of positivism. The basic principle of positivism is that all factual knowledge is based on the “positive” information gained from observable experience, and that any ideas beyond this realm of demonstrable fact are metaphysical. With such way, it would explicitly elaborate the argument “discussing BIM/IFC/MVD without considering information theory tends to limit the development of itself” in a more philosophical way. The literature review in section 2.2 firstly introduced successful applications of information theory, where information theory has been effectively employed in various domains. These examples include data organization and retrieval databases, information exchange in communication networks, decision -making processes utilizing



(a) (b)

**Fig. 2.** a. A IFC Kernel schema b associate topology of IFCKernel in tree structure.



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

**Fig. 3.** A retaining wall design (a). organizing information without assumption iv, (b). organizing information with assumption iv.

information entropy and uncertainty analysis. With information theories, these applications from various domains levered the fundamental features of information theory (listed in section 2.2) and then enhanced the rigor and generalizability of their research findings which manages to align with positivist objective. By now, so as to justify philosophy of employing information theory in MVD development with positivism, one should retrospect the MVD with ontological form. As stated in section MVD, there are 4 ontological features of MVD as following:

1. Data Abstraction and Representation: MVD development involves defining the structure and representation of data within a model.
2. Data Compression and Efficiency: MVDs often deal with large volumes of data.
3. Data Integrity and Reliability: In MVD development, ensuring data integrity and reliability is crucial.
4. Communication and Information Flow: MVDs often involve multiple views and stakeholders.

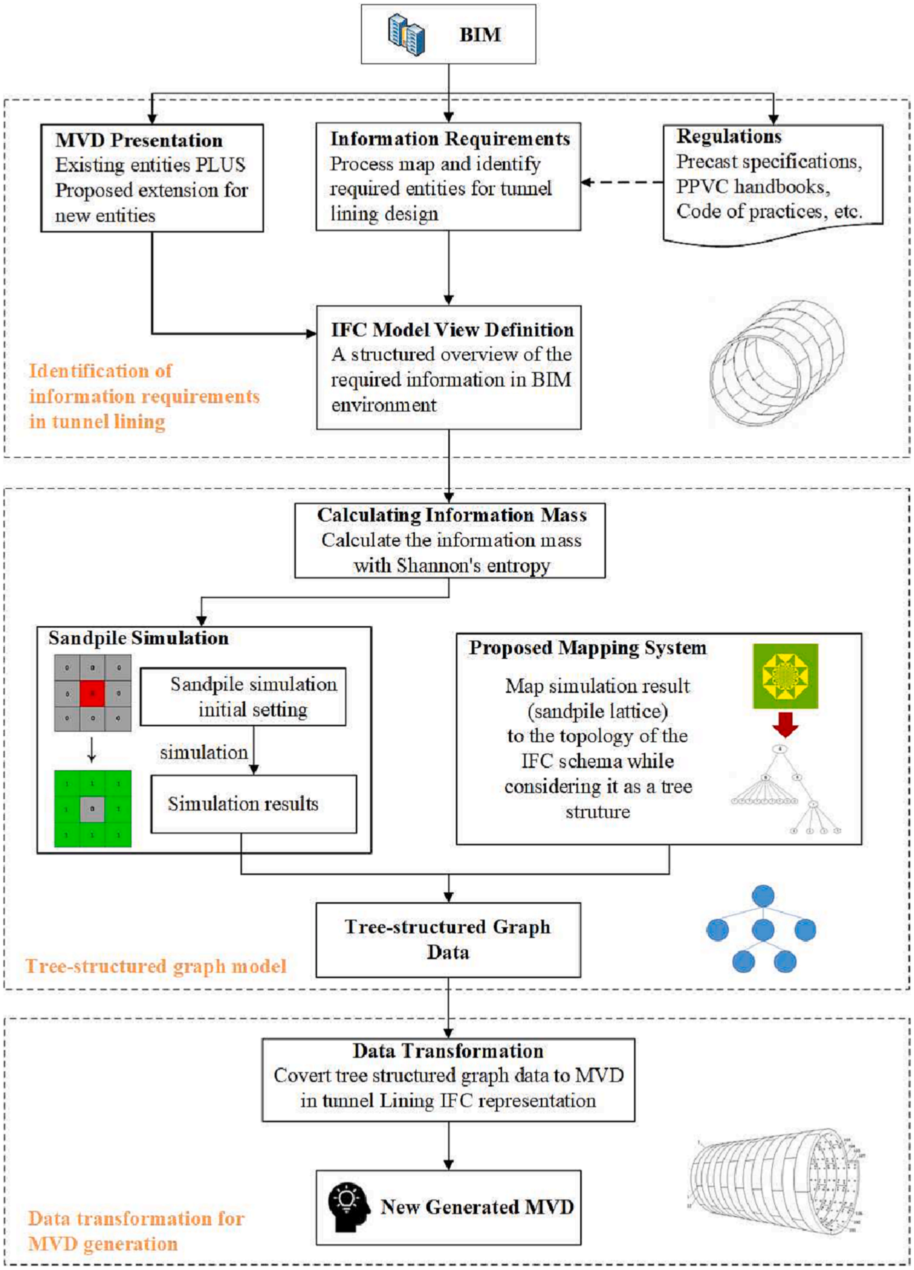


**Fig. 4.** Proposed methodology framework.

These ontological features of MVD show similarity to those information theory-oriented applications in various domain. To be more specific, to address feature 1–Data Abstraction and Representation, information theory helps in understanding the inherent complexity and entropy of the data. By quantifying the information content and identifying patterns and dependencies, information theory aids in creating effective data abstractions and representations that capture the essential characteristics of the system (this is one of research objective of this study); feature 2—Data Compression and Efficiency, information theory techniques such as data compression algorithms help reduce the storage and transmission requirements of the model. By leveraging entropy coding and other compression methods, MVD development can optimize the storage and transmission efficiency, leading to more scalable and performant systems; feature 3– Data Integrity and Reliability, information theory provides tools for error detection and correction, By applying error-correcting codes, checksums, and other techniques, MVDs can incorporate mechanisms for detecting and recovering from errors during data transmission or storage; feature 4– Communication and Information Flow, Information theory helps in analyzing and optimizing the information flow between different views and components. By quantifying the capacity and limitations of communication channels, MVD development can ensure efficient and reliable information exchange. In addition, information theory could help in quantifying and managing redundancy effectively. By strategically incorporating redundancy within the MVD, the system can handle data loss or component failures and ensure continuous operation. Further, Information 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, Yi et al, 2023). From these, it can be seen that Information theory offers a valuable framework for optimizing model view definition development within building information modeling, aligning with the positivist perspective. By leveraging information theory, researchers can enhance the representation and exchange of information, leading to improved objectivity and reliability of research findings. For example, 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.

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 (Figs. 3and 8), 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 sandpile simulation and MVD development 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.“.

In summary, the analogy with sandpile simulation is used to explain the need for splitting entities within MVD development to manage and organize information effectively. It highlights the importance of breaking down complex entities into smaller, more manageable units for better information handling and system design.

### *3.1. Identification of information requirements in tunnel lining*

#### *3.1.1. Required information for tunnel lining design*

Identification of information starts by reviewing the flow chart of shield tunnel lining design which configures the tasks. The flow chart (Fig. 5) is 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 [49].

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 [27,29,35,49], 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.

#### *3.1.2. IFC model view definition (MVD) for tunnel lining segment*

This subsection presents the new entities and attributes for MVD proposed by Prof. Christian Koch from Bauhaus-University Weimar and Prof. Xiaojun Li from Tongji University in accordance with the information requirements in section 3.1.1. Both of their works are selected for 3 reasons. Firstly, their works manage to develop the integrated comprehensive tunnel information modelling framework. Secondly, both of their works are used in real-world projects which authenticate their applicability and ensure a relatively objective and comprehensive information analysis. Thirdly, both elaborate the MVD on the same subject but raise different MVD structures, which highlights the differences of topology of their associate schema. The Koch's result is shown in Fig. 6 and Xiaojun's is shown in Fig. 7 as follow. To avoid displaying redundant information, the new addition and existing entities are differentiated by blue and yellow color respectively. It can be easily seen that MVD extension starts from the product extension. And Koch's model apparently contains more entities (60 entities-nodes in Koch's and 21 entities-nodes in Xiaojun's), both existing entities and extension entities are more than Xiaojun's model. As aforementioned, these schemas can be regarded as tree-structure graph (topology of schema) which is direct weight graph. If we consider the product node is the root node of tree structure, koch's MVD developed to 6th level which Xiaojun's only to 5th. For both proposed MVD, the designing aspects are expanded first by introducing new entities related to the definition of structural form for shield tunnel. As the shield tunnel is categorized as mechanized tunnel, it is very different of any other form of infrastructures. Consequently, the structural forms and the components it used are far different from the existing entities developed for others. Also, shield tunnel lining segments can be disaggregated into a finite number of components (such as segment block, joint bolt, void element and etc.). Noticeably, the most of types in proposed MVD are the new extended. A main reason of that existing types cannot be used as common types is the special construction of the mechanized tunnel like shield tunnel. Take segment block as an example, a ring segment consist of 8 segment blocks, and there would be 4 types of segment block for it to choose from, standard segment block, adjacent segment block, top sealing segment block, arch inverted segment block. Apparently, these types are the proprietary types for shield tunnel lining segments blocks. Similar to it, joint bolt types, tunnel warning unit types, tunnel installation types, tunnel temporary installation types, tunnel safety types. As for the material, some materials are common among the infrastructures, concrete, steel, cast iron. They just happen to be the main material for the ring segment, ring segment reinforcement. Ergo, the material for

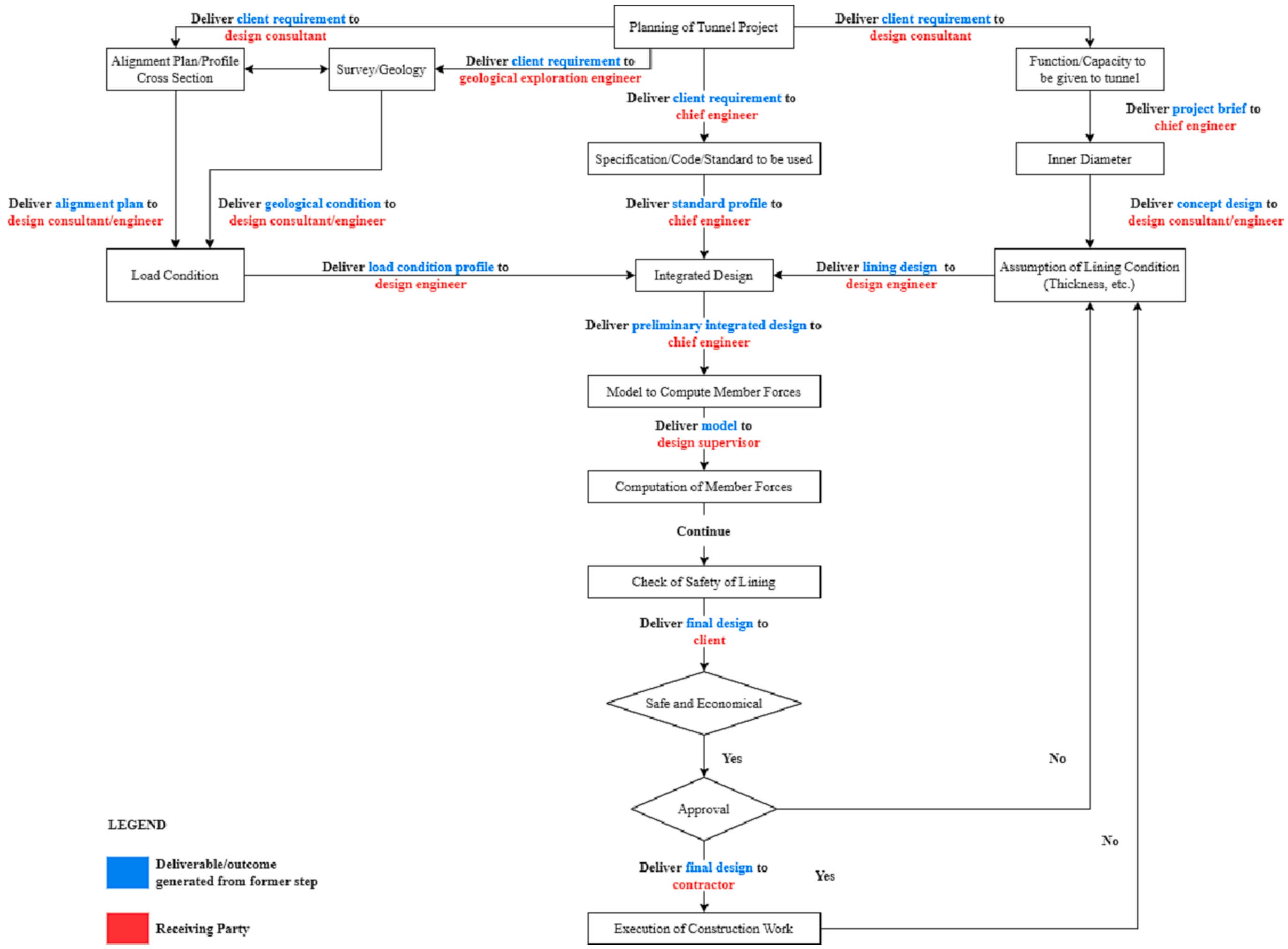


**Fig. 5.** Flow chart of shield tunnel lining design.

**Table 1**
Information requirements of shield tunnel lining.

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

these 2 entities is not extended with new. However, material for procedure like grouting is way more different from procedures in other infrastructure, as the material for shield tunnel grouting is normally a very sophisticated technique that has to associate with drilled soil, with drilled soil on site, it needs to go through a further fabrication process to make it proper material for grouting.

Noticeably, according to the assumption iii and iv, it can be seen firstly, the new proposed MVDs emerged the new tree structures with new proposed nodes those connected by new proposed branches. Regard the IfcShieldtunnel as product model, it apparently contains loads of information, to better express as well as manage the information, Koch supposed it would be better to describe it with 2 sub models—spatial element and element. Ergo, it can be seen in the Fig. 6, the root node (product) split to 2 nodes (spatial element and element). On account of that, the spatial element node and element node split followingly. Therefore, the topological structure of proposed MVD is directed, top-down hierarchically speaking, the direction is from high information mass breakdown to low information mass. And information mass is related to the information requirements of the entities. The level of information requirements acts as the weight distributed to the nodes. Again, it proves the MVD structure is actually a directed weighted structure—tree structure. Notwithstanding, two proposed MVDs shared some similarities on the topological relations, the detail layout is different as aforementioned earlier, it confirms the research gap

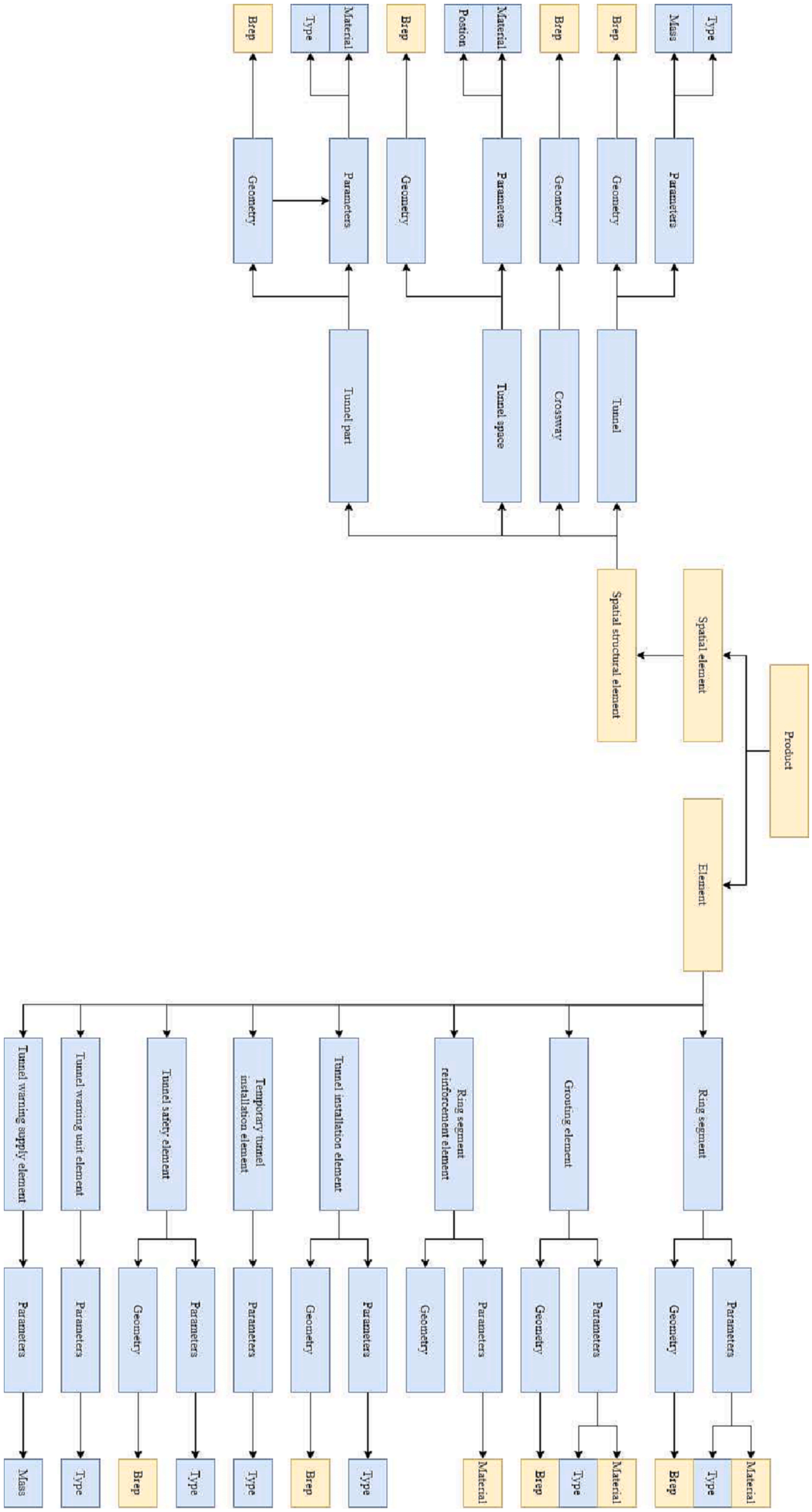


**Fig. 6.** Proposed MVD structure for Shield Tunnel by Koch.

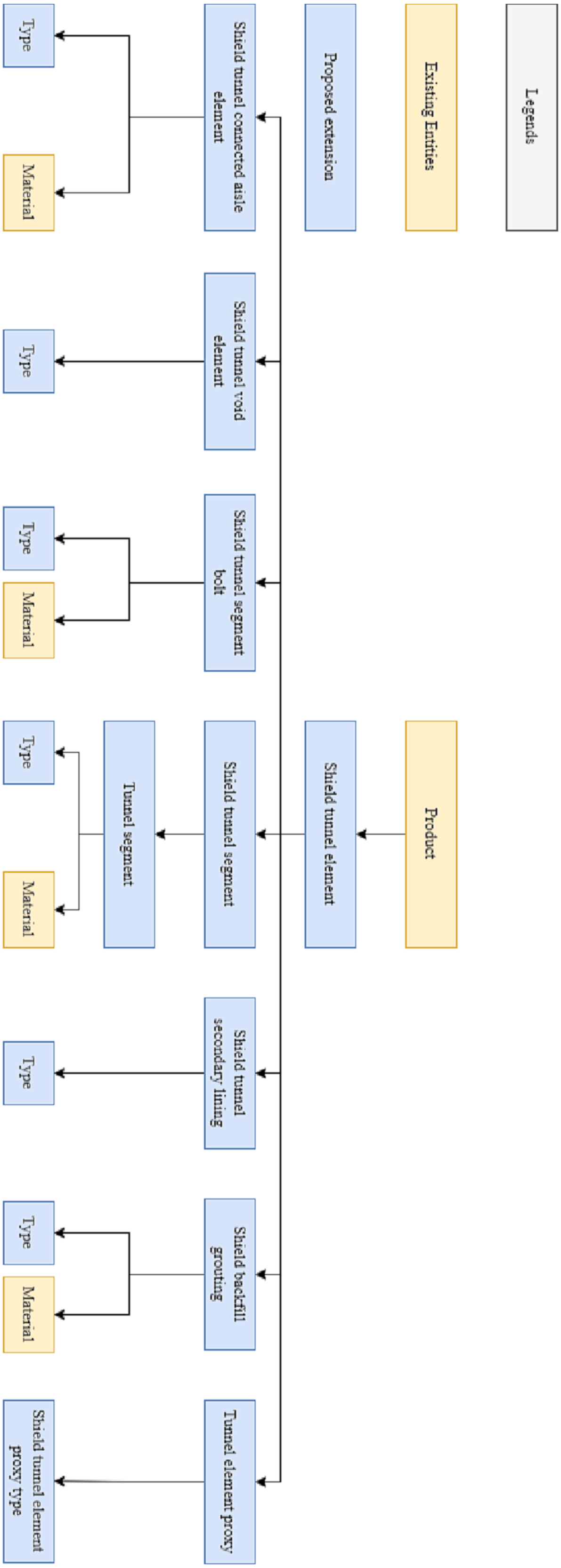


**Fig. 7.** Proposed MVD structure for Shield Tunnel by Xiaojun Li.

identified in the section 2, that is lacking principles to regulate the MVD/linked. To overcome this, the information entropy is used to measure the information requirements, and thereupon would be the principles to regulate the MVD development.

## 3.2. *Tree-structure graph data modelling for tunnel lining*

### 3.2.1. *Calculating information mass*

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 [41]. 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 $log_2 2^N = N$.

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 figure out. To crystalize the transformation process, a demonstration with simplified a tunnel lining element is given based on information requirement analysis referring “Guidelines for the Design of Shield Tunnel Lining” in 2.1.1 [49]. The basic design information is regulated by “China National Design code of Shield Tunnel Engineering– GB/T51438-2021” [21] 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 900 mm; 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 50 mm; 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 $A_8^8$

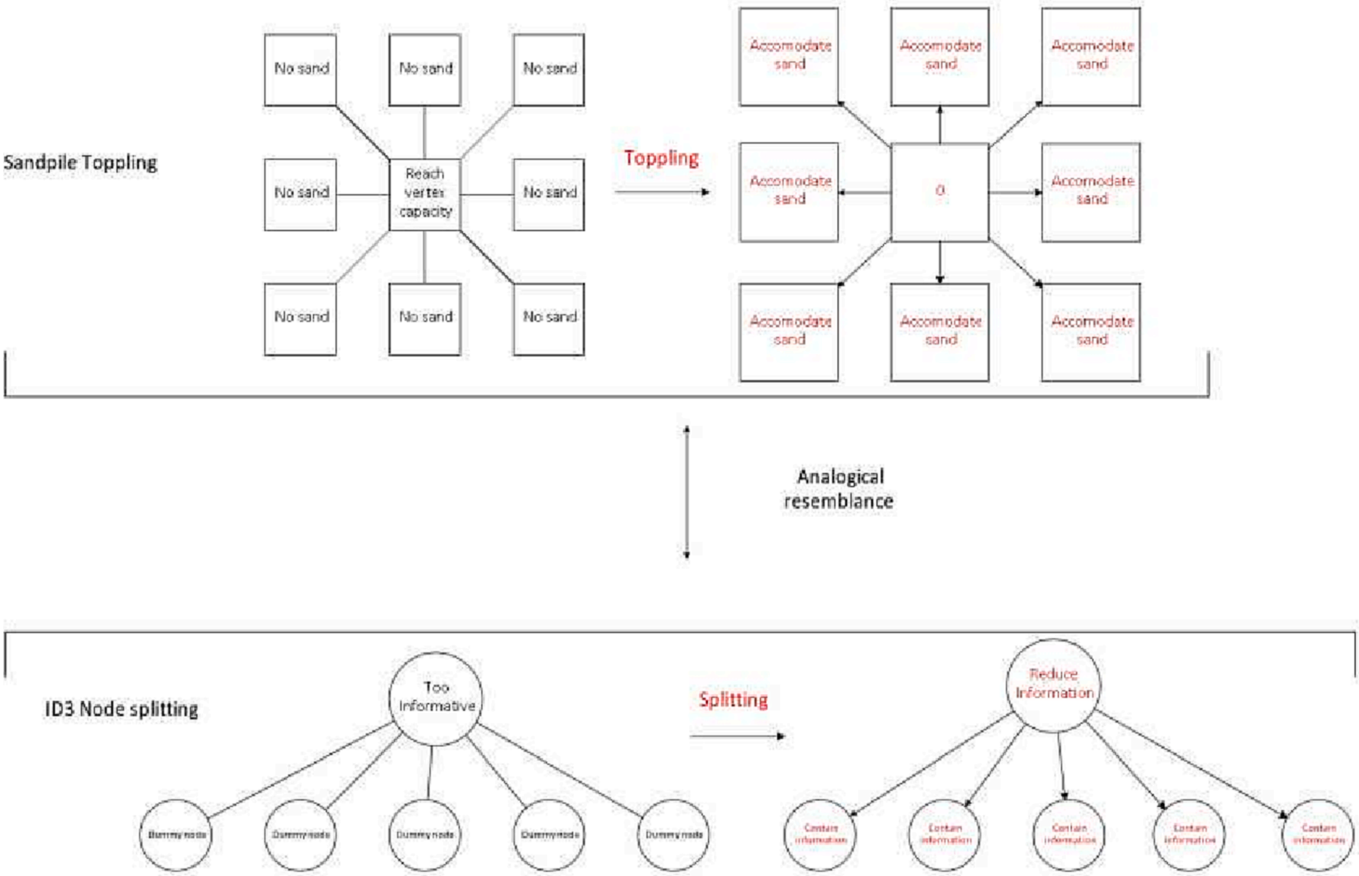


**Fig. 8.** Analogy demonstration of Node Splitting and Sandpile Toppling.

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. In consequence, a simplified tunneling lining ring segment contains 16 bits information mass. By this way, the information is quantifiably measured with Shannon information entropy.

### 3.2.2. *Sandpile simulation*

This section presents the applied sandpile simulation that could help to develop the MVD through tree-structured graph data model. The sandpile simulation is leveraged since the analogy of toppling process while conduct sandpile simulation is same as the node splitting process of tree generation on the basis of 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 many information, it splits to multiple nodes in next level. Through Fig. 8, 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.

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 [8,15,22]. They stressed the importance of this discovery with the claim “it develops complexity out 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 [8].

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 $(\mathrm{i},j)$ 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 [31], 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 [18], 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 $G = Z/4Z \subset H = R/4Z$. To comprehensively elaborate the typology of the group of a vertex, the process of generating group $G$ out of group Q ($Q = Z$) is given along with the Cayley graph. First, considering the group $Q = Z$, defining an arrow represent the generator $< 1>$, its Cayley graph would be shown as Fig. 9.

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

Based on this, it can easily conclude that assumes a group $H^{'}$, which is a group contains 4 and it is a normal subgroup, the quotient group $\frac{\mathbb{Z}}{H}$ 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 $< 4 >= \{\dots, -12, -8, -4, 0, 4, 8, 12, \dots\}$. Since $\mathbb{Z}$ is an Abelian group, its subgroups are all normal, $<4>$ is a normal subgroup. Other than that, the cosets are needed Also, $\mathbb{Z}$ is defined as an additive group. Ergo, $k + < 4 >$ is used to express the co-sets. Fig. 10 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.

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 $a \equiv 4\ b$, for instance, $1 \equiv 45$, $2 \equiv 46$, $3 \equiv 47$. Using these cosets to reorganize group $\mathbb{Z}$, the structure would be an infinite spiral as shown in Fig. 9. 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 $G = Z/4Z$. To understand Fig. 11 better, a few extra notations are given as below, homeomorphism mapping $\phi : Z \rightarrow c_4$ and isomorphism mapping $i : \frac{\mathbb{Z}}{\langle 4 \rangle} \rightarrow c_4$; quotient mapping: $q : Z \rightarrow \frac{Z}{\langle 4 \rangle}$.

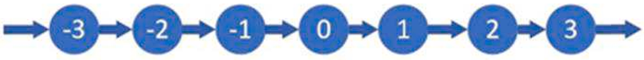


**Fig. 9.** Infinite cyclic group $\mathbb{Z}$.

The sandpile groups are a multiplicative group of integers modulo 4. To visualize the toppling process of sandpile model, Fig. 12 below shows a process of a vertex toppling and a series of toppling involve 3 × 3 vertices.

In this paper, 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 $(\mathrm{i},j)$. For instance, in the case of $L = 100$ lattice, $(\mathrm{i},j)$ 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 $(\mathrm{i},j)$ at the time $t$, then there will be:

$$E(\mathrm{i},j) = E(\mathrm{i},j) + e \tag{1}$$

Each vertex has a capacity of sand contained, denoted as $E_{\max}$. At a given time point $\tau$, the sand amount of vertex $(\mathrm{i},j)$ $E(\mathrm{i},j) \geq E_{\max}$, 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 Fig. 13.

$$E(\mathrm{i},j) = 0 \tag{2}$$

$$E(\mathrm{i}+1,j) = E(\mathrm{i}+1,j) + E(\mathrm{i},j)/8 \tag{3}$$

$$E(\mathrm{i}-1,j) = E(\mathrm{i}-1,j) + E(\mathrm{i},j)/8 \tag{4}$$

$$E(\mathrm{i},j+1) = E(\mathrm{i},j+1) + E(\mathrm{i},j)/8 \tag{5}$$

$$E(\mathrm{i},j-1) = E(\mathrm{i},j-1) + E(\mathrm{i},j)/8 \tag{6}$$

$$E(\mathrm{i}-1,j+1) = E(\mathrm{i}-1,j+1) + E(\mathrm{i},j)/8 \tag{7}$$

$$E(\mathrm{i}+1,j+1) = E(\mathrm{i}+1,j+1) + E(\mathrm{i},j)/8 \tag{8}$$

$$E(\mathrm{i}-1,j-1) = E(\mathrm{i}-1,j-1) + E(\mathrm{i},j)/8 \tag{9}$$

$$E(\mathrm{i}+1,j-1) = E(\mathrm{i}+1,j-1) + E(\mathrm{i},j)/8 \tag{10}$$

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 $< 8 >= \{\dots, -16, -8, 0, 8, 16, \dots\}$. 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

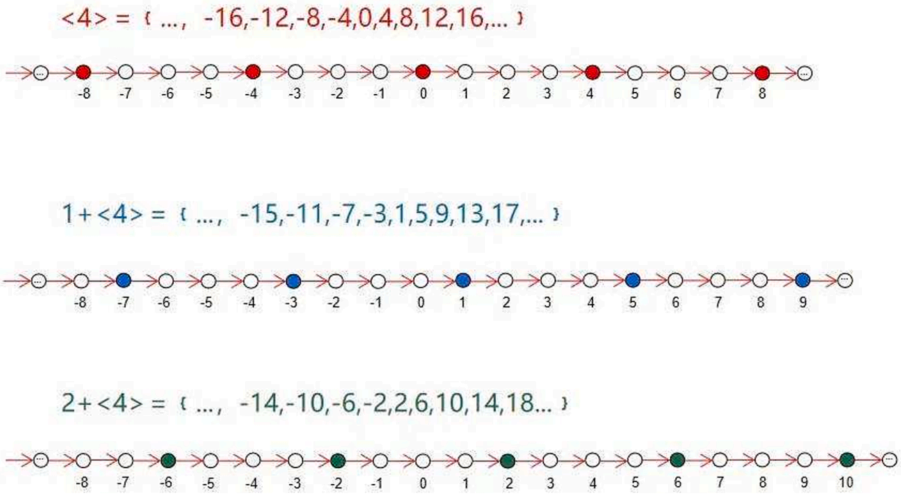


**Fig. 10.** $k+ < 4 >$ cosets, $k = 0, 1, 2$.

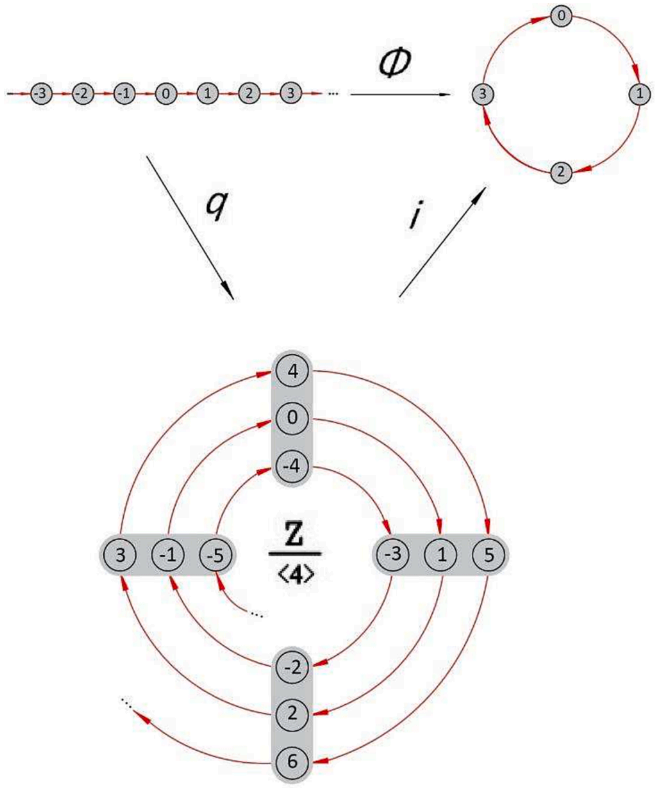


**Fig. 11.** Integration of quotient mapping and renamed isomorphism equals to mapping $\phi$. For any integer $k, \phi(k)$ is remainder of $k$ modulo 4.

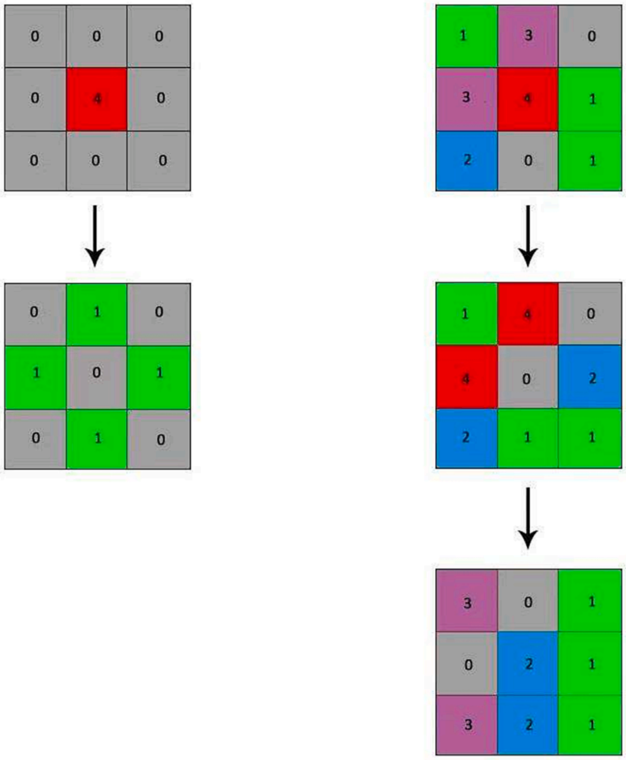


**Fig. 12.** Toppling of vertices-1. A vertex of the sandpile ($3 \times 3$) carries 4 or more particles, it becomes unstable and topples, decreasing the number of its particles by 4 and increasing the number of particles carried by each of its neighbors by one. The toppling of one vertex can render 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 Fig. 14 better, a few extra notations are given as below, homeomorphism mapping $\phi : \mathbb{Z} \to c_8$; isomorphism mapping $i : \frac{\mathbb{Z}}{\langle 8\rangle} \to c_8$; quotient mapping $q : \mathbb{Z} \to \frac{\mathbb{Z}}{\langle 8\rangle}$ . The topology of this sandpile model is given in Fig. 13. 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

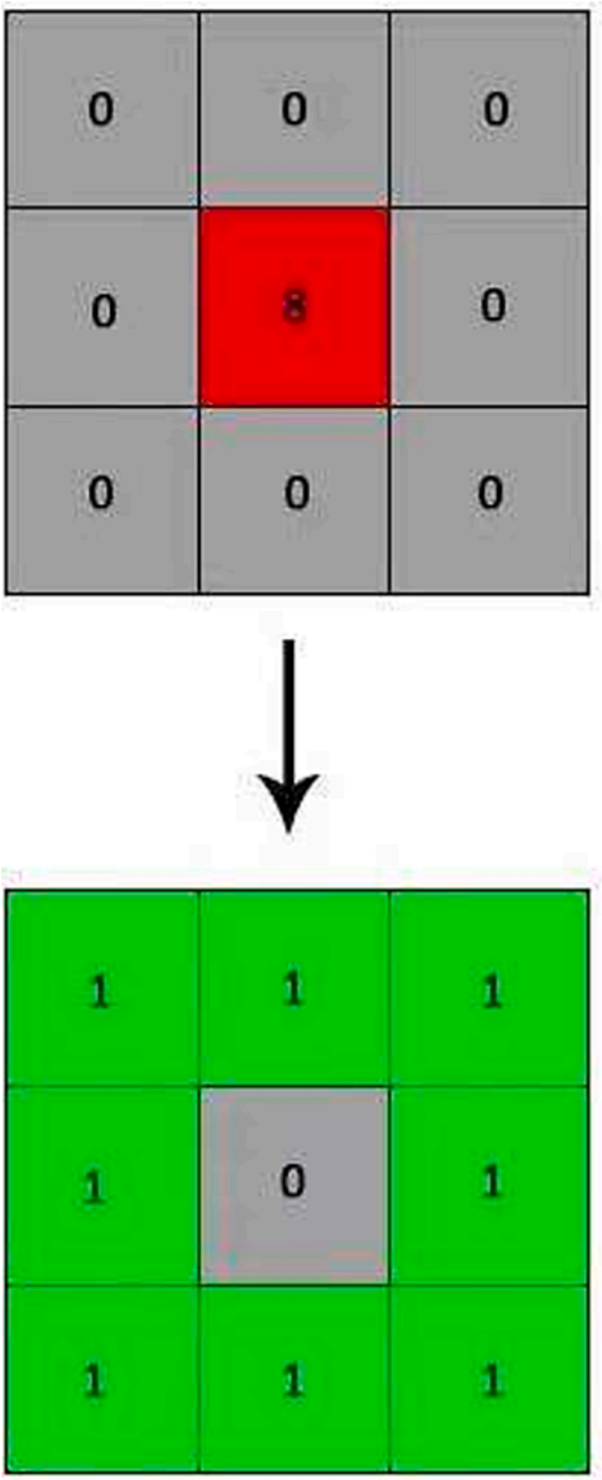


**Fig. 13.** Toppling of vertices-2. A vertex of the sandpile (3 × 3) carries 8 or more particles, it becomes unstable and topples, decreasing the number of its particles by 8 and increasing the number of particles carried by each of its neighbors by one. The toppling of one vertex can render other, previously stables unstable, resulting in an avalanche of subsequent toppling. Red is used to mark vertex with 8 grains, and, green for 1 grain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

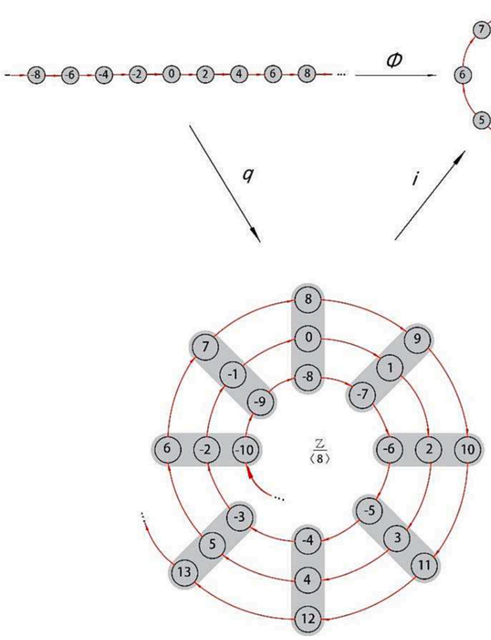


**Fig. 14.** Integration of quotient mapping and renamed isomorphism equals to mapping ϕ. For any integer k, ϕ(k) is remainder of k modulo 8.

occurrence, then the relation between scalability and occurrence would be expressed as:

$$D(S) \propto S^{a} \tag{11}$$

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

$$\ln(D(S)) = a\ln(S) + b \tag{12}$$

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:

$$D(L) \propto L^{c} \tag{13}$$

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:

$$\ln(D(L)) = c\ln(L) + d \tag{14}$$

### 3.2.3. *Mapping sandpile simulation to topology of MVD development*

In the first part of section 3.2.2, the rationale justification of levering sandpile simulation to MVD development 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, the angular coordinate is denoted by $\theta$. 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 $\theta$ can only range from set $\left\{\theta_n|\theta_{n-1} + \frac{360^0}{n}\right\}$, wherein, 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 $m$, $\{m|m = 0orm = 1\}$. 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 $p(\theta, r, m, g)$. Fig. 15 is given to visualize the expression of a phase for a random vertex in the 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, and the row is $r$ ow, 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 $3 \times 3$ vertices as a result of a sandpile simulation following rules of applied sandpile model defined in 3.1 equation(2)-equation(10) is given in Fig. 16.

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, by structure, it means the number of nodes in the tree, and their associated level, the number of levels in the tree, after all these predetermined, the weight (information mass) of each node, are all required to provide. 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:

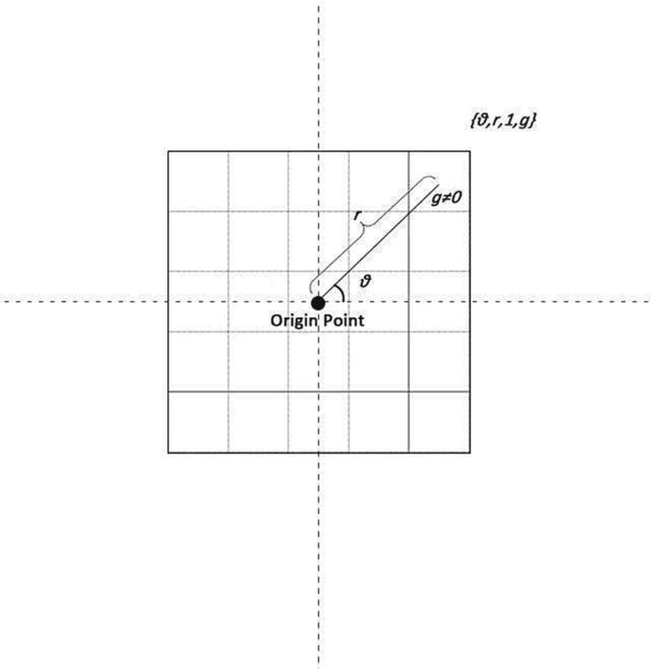


**Fig. 15.** Expression of phase for a random vertex in lattice.

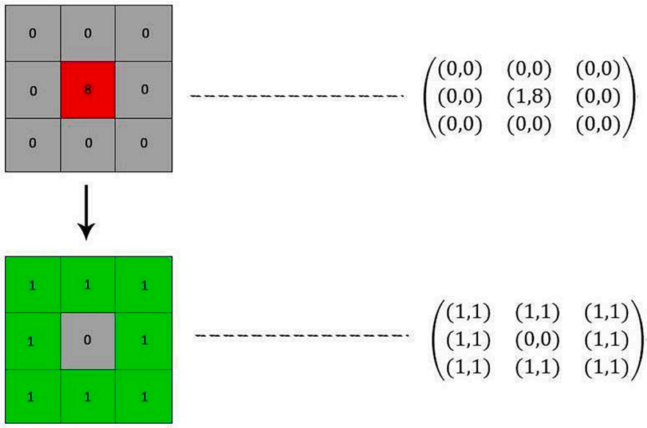


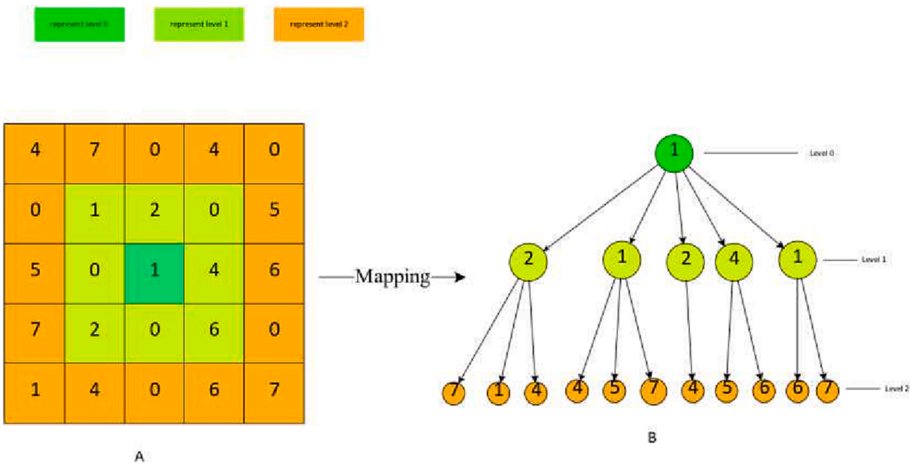
**Fig. 16.** Symbolizing sandpile simulation result to 2-dimensional tensor.

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.

### 3.2.4. *Tree-structured graph data*

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.

The root node of the tree is regarded as level 0. A sandpile lattice is given in Fig. 17A, 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 Fig. 16. The topology of associated MVD is given as Fig. 17B. 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 wt Hence, it could only be considered as 5 nodes in the level 1 as shown in Fig. 17B. Similarly, it can be easily to conclude that there would be 11 nodes in level 2 (5 nodes assigned weight 0 among 16 nodes). And all three levels, there supposed to be 17 nodes including root node as shown in Fig. 17B. 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



| | | | | |
|---|---|---|---|---|
| 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 |

**Fig. 17.** Mapping of a layout of sandpile simulation result to tree-structured graph.

universally consistent cross the users. This forms the theoretical basis of unifying MVD development then making a standard “standard”.

### 3.3. Data transformation for MVD transformation

This subsection introduces the data transformation that could convert graph data to MVD in tunnel lining IFC representation. Fig. 18 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 3.2. 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.

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 [17,33,45,50,55]. 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 Fig. 19. 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 relathion 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

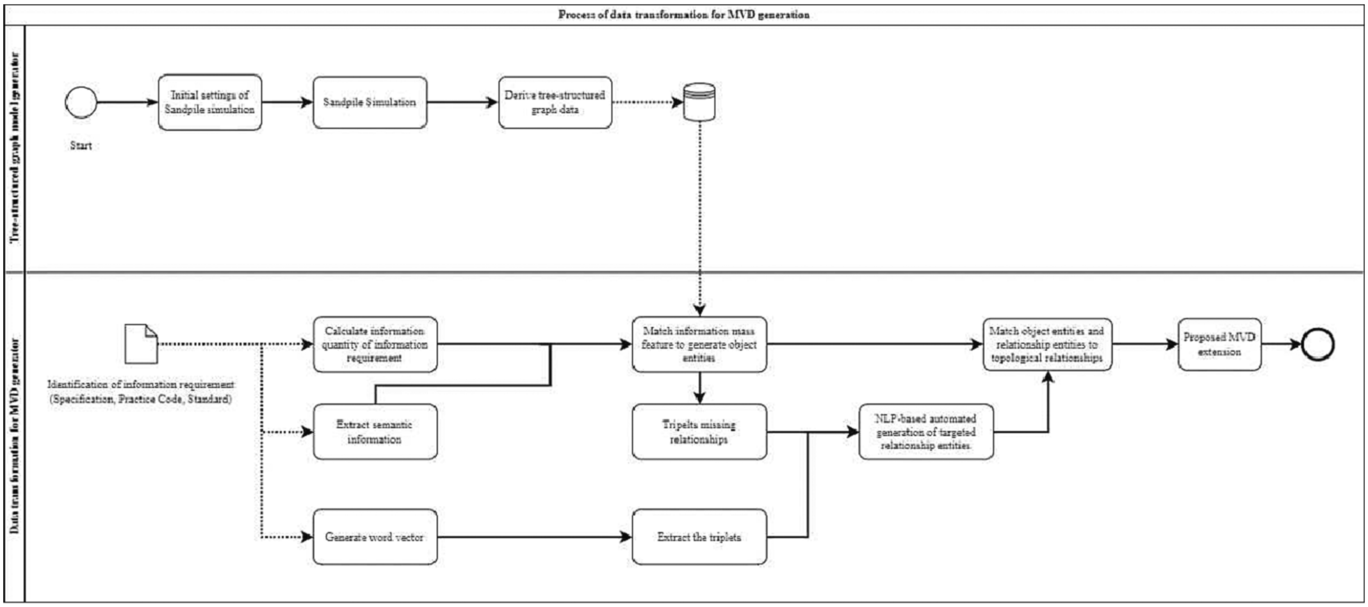


**Fig. 18.** Process of data transformation for MVD transformation.

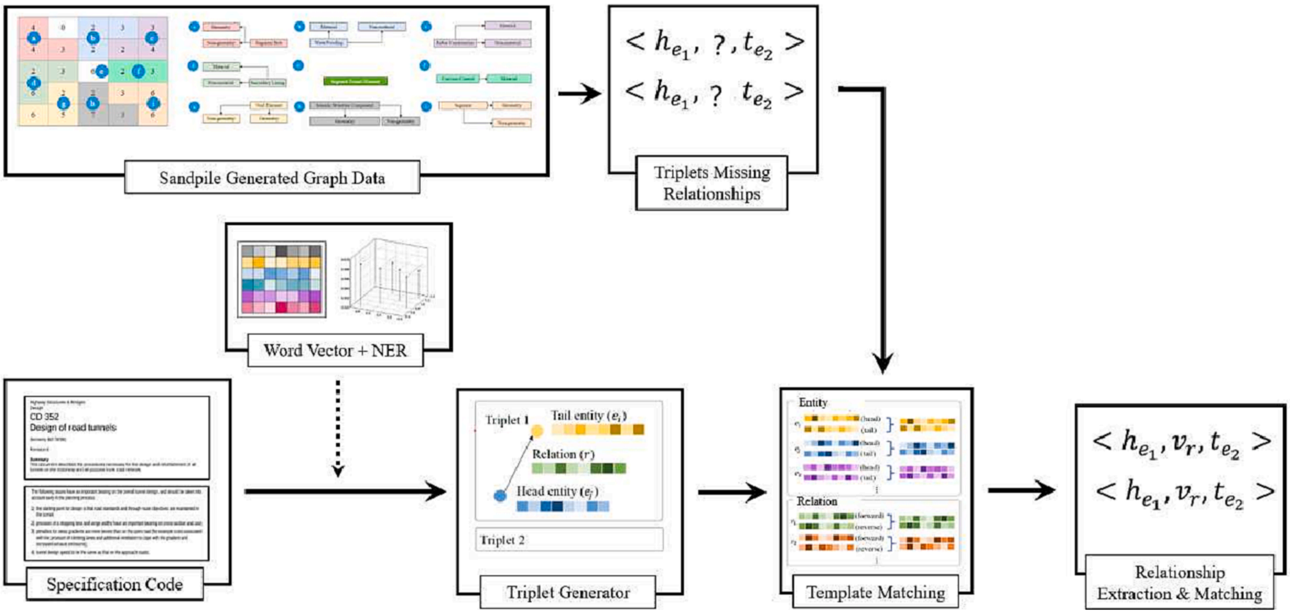


**Fig. 19.** The architecture of proposed integrated entity and relationship completion model.

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.

Take the Xiaojun's MVD (given in Fig. 7)as a simplified example, presume there is a project (Q km shield tunnel needs to be constructed) using the tree-structured graph data to interpret the proposed MVD would be given following:

- 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 3.2.1.
- 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 3.2.1.
- 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 Fig. 20.

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 the 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

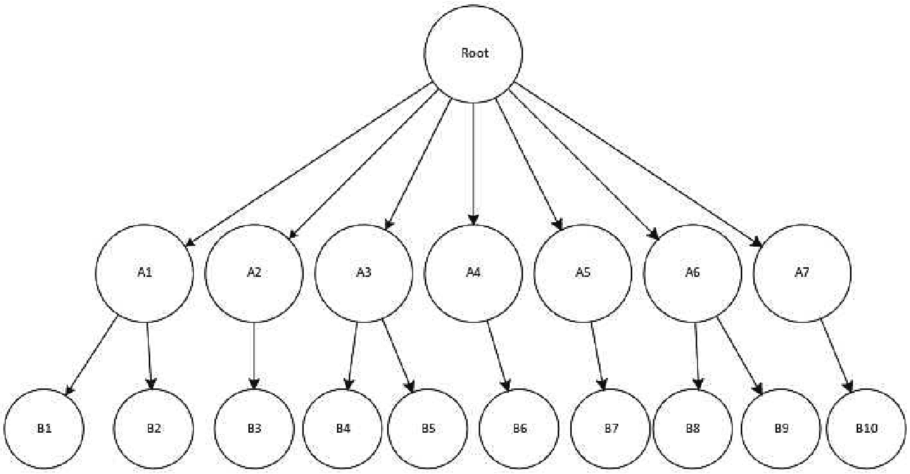


**Fig. 20.** Tree structured graph that aligns with Xiaojun's MVD.

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 $A_{1=}$ $G_1$, then the 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. The on the graph, the nodes are linked with their nearest nodes.

## 4. Illustrative examples

### 4.1. Initial settings

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 [2,46]. 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 represents the entities. The initial configuration of all the vertices is set to 0, which stands for there is need to sort information by giving the data schema, and while the information is few, 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.

(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 $L^2 = 51 \times 51$, 867 vertices toppling, the limits of standard for expressing the information mass prevails. Two considerations shape these regulations, a. grains of sand dropping is 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 km. All the tunnels are underground, considering the width of a ring is roughly 1.2 m, 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

$\times$ 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 [48]. It is a free and open-source software platform with a simplified and flexible programming language [16]. Hence, it is chosen to be used for the current study. And the sandpile model developed by Weintrop et al. in 2011 is imported to Netlogo for further developments to suit the current study [47]. And the information requirements analysis complies with “China National Design Code of Shield Tunnel Engineering”.

With aforementioned initial configurations of sandpile simulation. For each simulation, besides the generation of tree-structured graph data model, another several observations require additional attention which would be reported here, 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).

### 4.2. Illustration

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 Fig. 20. It could be seen that there are 3 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 Fig. 21. With the mapping system proposed in section 3.2, the associated tree-structure graph is given step by step. Then following the process proposed in section 3.3, 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 Fig. 21 shows, the center vertex of layout represents the root node, and 8 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 6, and the 8 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.

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 3.2, 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 12 alternatives. Hence, according to information entropy formula, the total information mass would be $\log_2 2^N = N = \log_2 12 = 3.58bit$; 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 $\log_2 2^N = N = \log_2 5 = 2.32bit$. 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 6 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 8 generated information entropy based entity pairs would be given in Table 2 Translating information entropy based entity pair to actual entity pair according to specifications as following.

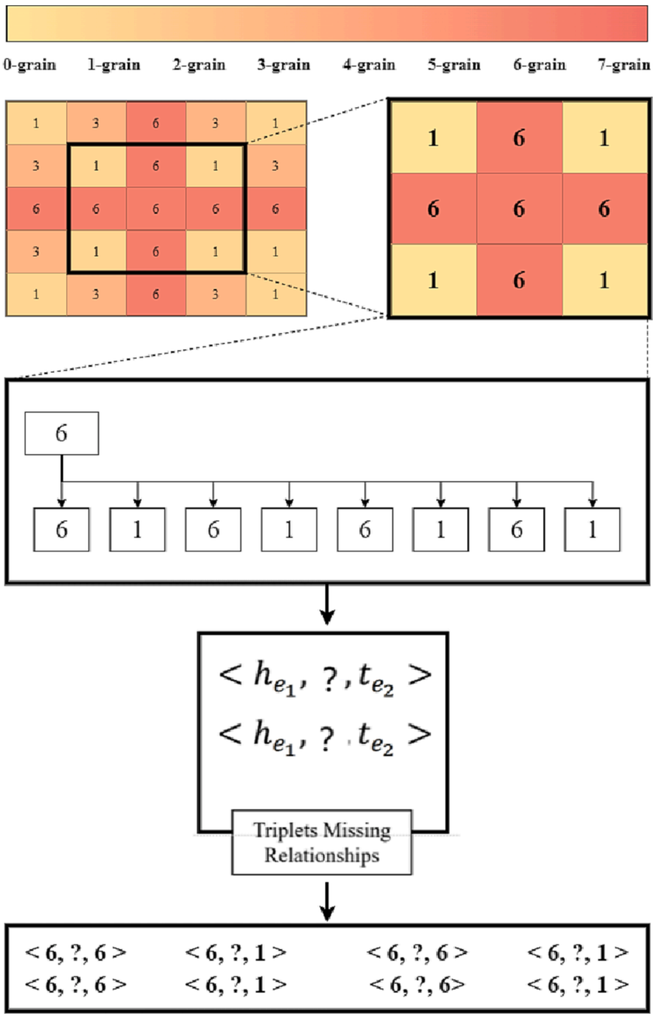


**Fig. 21.** The layout of lattice after Sandpile Simulation and transformation to tree structure taking root node and first layer as a demonstrative example.

Then according to section 3.3, employ NLP based technique to extract triplet from same specification/code of practice as the ones used to consist with entity pairs. Such as, in the China National Design code of Shield Tunnel Engineering, there is original description as “The segment tunnel element can be constructed with a single-layer or double-layer design”, to adapt to the naming convention for IFC relationships –

**Table 2**
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 > |

"RelatingEntity_RelatedEntity", the new relationship entity for the case < SegmentTunnelElement,?, TunnelSegmentEnvelope > would be defined as "SegmentTunnelElement_ConstructsWith_TunnelSegmentEnvelope". Following such manner, the NLP based technique generated triplet and match the entity pair process would be given as following (f).

The rest calculation of information mass for the entities on the second level can be done in the same manner. Till now, conclude the existing calculation and covert these nodes from Fig. 21 to MVD (given in Fig. 23).

Fig. 22 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 tos MVD structure for shield tunnel proposed by Koch (the right half started with "Element") given in Fig. 6. 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 in Fig. 7, 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

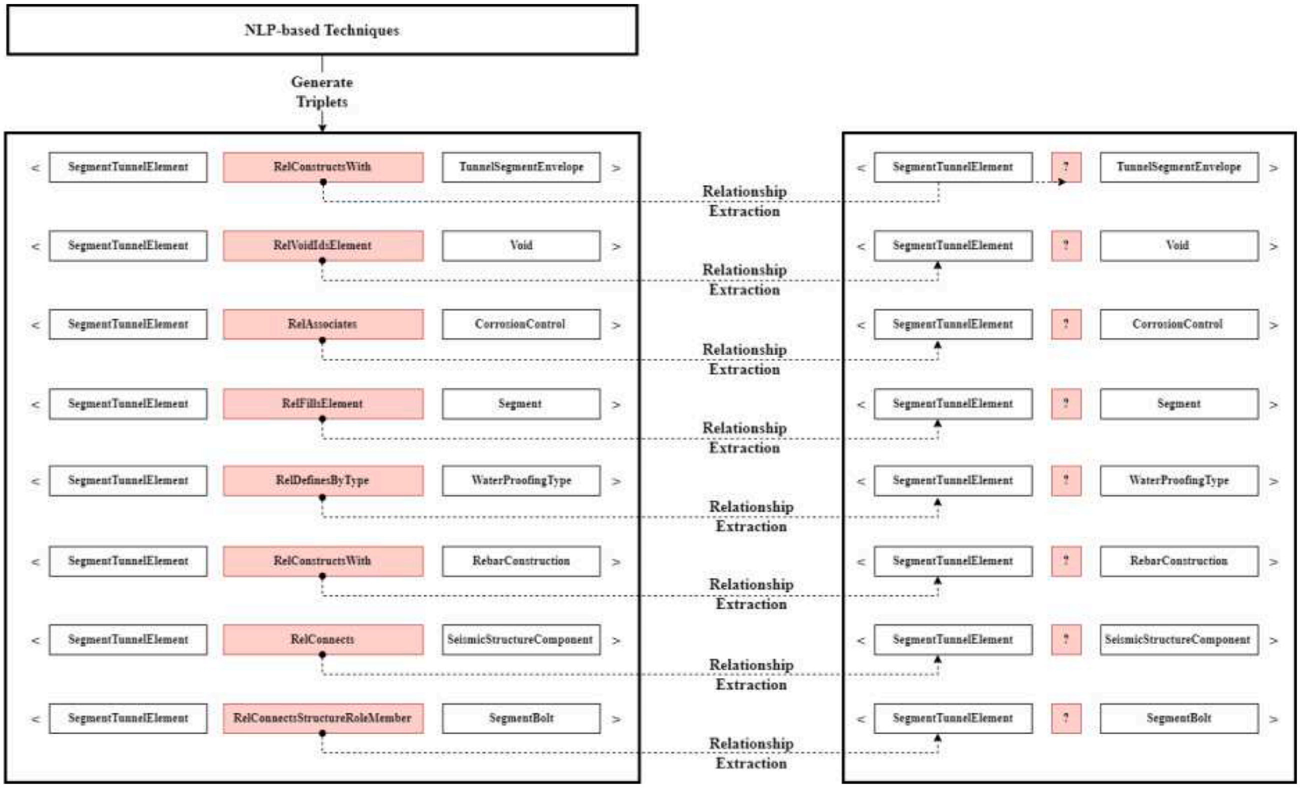


**Fig. 22.** Completing triplet by extracting relationship entity via NLP.

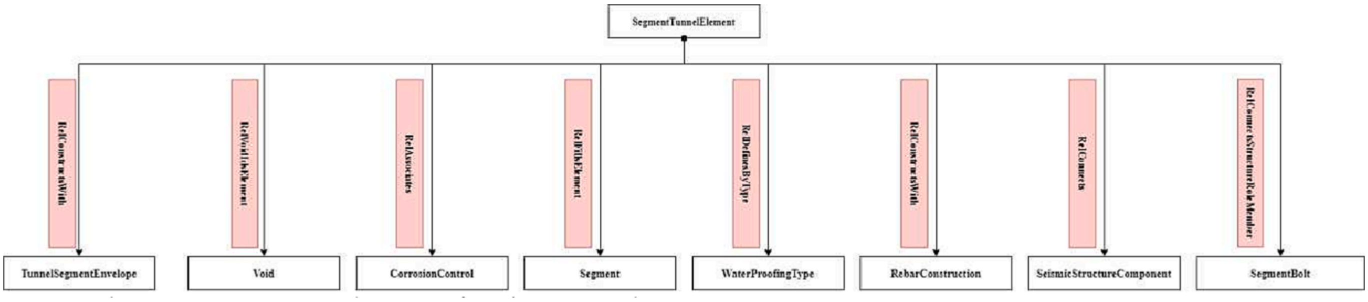


**Fig. 23.** Converted MVD for the scenario.

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.

### *4.3. Other observations*

Regardless of successfully generating object entities without causing much ambiguity by utilizing the proposed method in the MVD development, there would be another concern regarding how the entities be related. This used to be dealt with the regulated self-defined relation entities. With proposed method, this could not be directly solved. However, the advanced AI techniques like expert system and natural language processing (NLP) have huge potentials to resolve such issue. Considering a generated tree structure (MVD), all the object entities (leaves) are configured. Expert systems accumulate experience and facts in a knowledge base and integrate them with an inference or rules engine – a set of rules for applying the knowledge base to situations provided to the program. The knowledge base here is built on acquired ontologies, MVDs, code in practice. Within that, the rules of defining relation entities would be generated and stored in an inference or rules engine to guide the generation of relation entities. And NLP could be used to help identify generated object entities in the developed MVD via proposed method, then this identified information can be formulated as input for expert system to generate relation entities. Such process could automate the generation process and also reduce the ambiguity of relation entity generation to certain extent. Due to the page limit and the progression on reducing ambiguity during object entities is already a big step forward, this part of work is not covered in this study. However, integrating expert system and NLP techniques with proposed method is definitely meaningful and promising. Hence it would be discussed in future recommendation.

Apart from the demonstration of proposed methods for generating MVD, 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 4.1, on the lattice of $L^2 = 11 \times 11$:

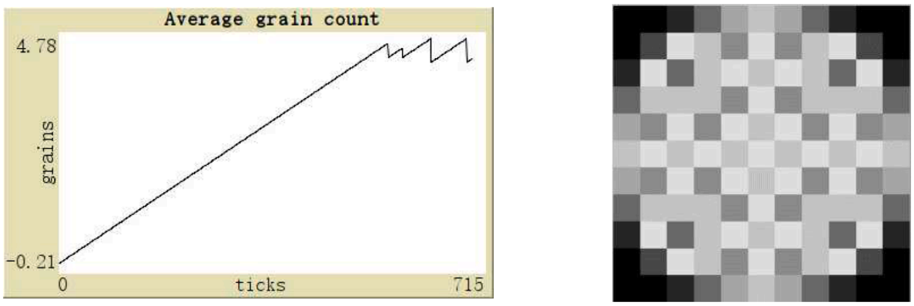
The critical state reaches while the grains of sand is 550, the average height stopped increasing and started to show the fluctuance in a level and the distribution before the critical state is given as shown in Fig. 24. The height of vertex is demonstrated with grey scale, the 0 is black ■, and the 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.

After entering the critical state, the first catastrophic avalanche happens at 575 grains of sand. The toppling vertices are given as Fig. 25a, 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 (Fig. 25b). Accordingly, avalanche sizes (the ratio between the frequency logarithm and avalanche size logarithm) descend to 0.659 (Fig. 25c). This catastrophic avalanche lasts longer than before, and during the toppling process, the vertices involved more than it appears in Fig. 26.

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 Fig. 25, hence the size of avalanche is large which results in an avalanche size close to 0.

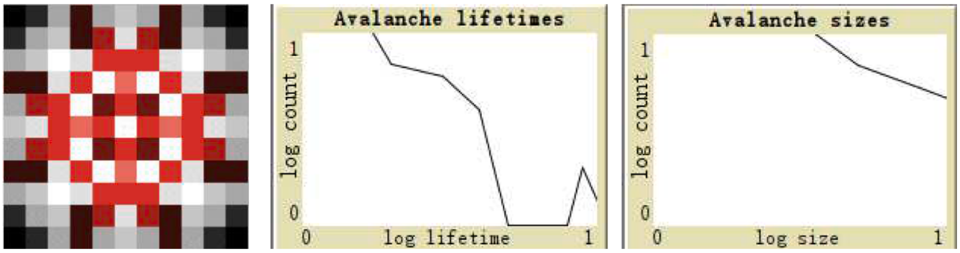
Other related coefficients of the sandpile simulation regarding both simulations are given in Table 3. The symbols a, b are the coefficients in equation (12), and c, d are the coefficients in equation (14).

Other configurations are given to examine the isomorphism since the



(a) (b)

**Fig. 24.** Simulation result of 550 grains of sand, 23-a shows the average grain count with sand dropping, and 23-b shows the layout of sandpile while enters critical state.



(a) (b) (c)

**Fig. 25.** Simulation result of first catastrophic avalanche. a - shows toppling scale and pattern; b shows the log–log plot of avalanche lifetimes and 18-c provides the log–log plot of avalanche.

(a) (b) (c)

**Fig. 26.** 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.

**Table 3**
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) |

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 Fig. 27, it can be seen the log–log plot of both avalanche lifetime and avalanche size for different lattices are similar.

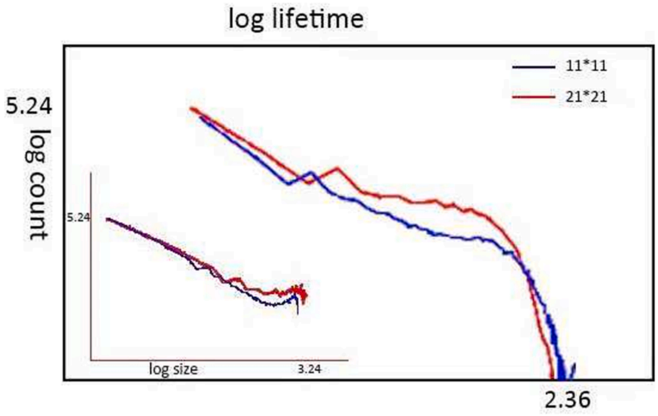


**Fig. 27.** Comparison of log–log plot (lifetime and size) between 2 simulations.

## 5. Discussions

There are a few intriguing findings based on the simulation results while sandpile model applies to the development of tunnel lining MVD. To comprehensively elaborate on the results, this section first discusses the findings, contributions, and limitations. At the end of this section, future recommendations are brought forth.

### 5.1. *Findings and contributions*

Firstly, since the lattice and grains of sand in the sandpile model, as the phase and projection of the tunnel lining BIM model, present the pattern of the topology of MVD development. For the purpose of affirming practical value, a few points need to be clearly pointed out. MVD, as a part of standardization for BIM, is required to be aligned among parties. With proposed approach to develop MVD, the proposed MVD is just the final product that parties could generate locally rather than transmitting MVD as both carrier and final product. According to section 3 and section 4, it could be seen that as long as all parties agree on several things, including the systems of measurement for information quantity calculation (bit/nat), local codes, initial settings on sandpile lattice layout, toppling rules and engineering design, which are supposed to be easily obtained and agreed on, the MVD would be generated directly on each ends uniformly. And noticeably, referring to the simulation results in section 4, a 11 × 11 lattice, can be translated to 6 layers of tree structure, the central point regarded as the IFC shield tunnel segment. With the information mass of l lining increasing, the grains of sand increasing, original expression of data is no longer effective, hence the extension is needed. The extending process is represented by the toppling of vertices when toppling expands the range of affected vertices (such as from 3 × 3 to 5 × 5). It corresponds to MVD entity branching down to the next layer in the MVD schema therefore IFC schema.

Manifestly, there are possibilities that toppling would not expand the affected vertices, in which scenario, the MVD/IFC standard does not need to change the topology but amend it slightly. Such slight amendment as an additional version can extend to a few more property sets, creating a requirement for a case-oriented ontology, updating the model view definition. In essence, it is the adjustments of semantic aspects. The avalanche size coincides with the change of existing data standard. The greater avalanche is, the more revolutionary change should be made towards existing standards.

As can be seen in Fig. 27, 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 [7]. 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 [54]. 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 Fig. 28). 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 Fig. 27and Table 3shows 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 synergetics. This, again, verifies the assumption i that expects different information mass requiring different MVD to accommodate. Moreover, another intriguing inspiration from section 4.3 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.

Utilizing the sandpile model to simulate the topology of the tunnel lining MVD is self-consistent under the provided hypotheses. 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 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 another significant addition of this paper.

### 5.2. Limitations of the research

The beginning of this paper 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 ii, 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

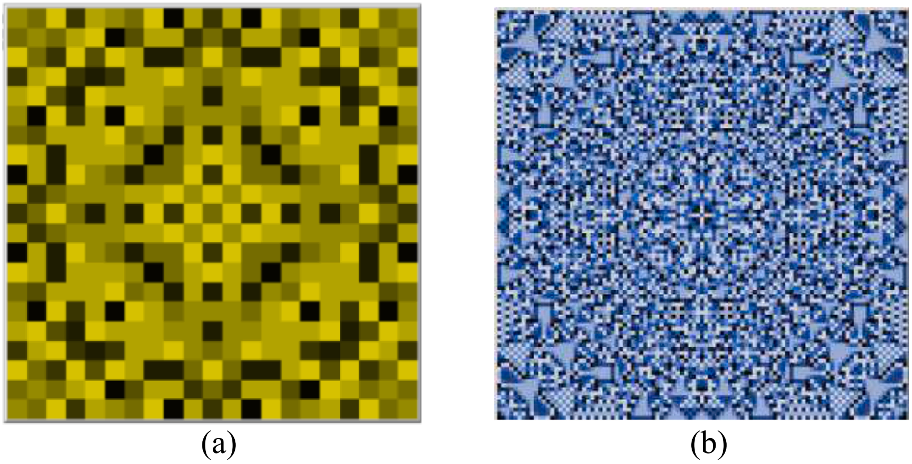


**Fig. 28.** (a). Layout of $L^2 = 21 \times 21$ with 18,280 grains and (b). Layout of $L^2 = 101 \times 101$ with 1,000,000 grains.

information varies with end-users and use case scenarios. This paper designed assumption ii for the purpose of simplicity as the authors intended to bring more methodological thinking rather than the simulation results. To fit in practice, assumption ii requires more critical revision. Regarding assumption iii, ID3 as the tendency of branching trigger (the higher entropy value node should be branched) is the choice of the writer, which is based on the writer's academic background and personal views. An excellent example of it is replacing ID3 with C4.5. Also, as aforementioned, the branching triggering tendency is tightly linked to objective collapse. In this context, it is use case sensitive. All the hypotheses made in the paper enable the simplicity and self-consistency throughout the whole research, but the viable application in practice has not been further discussed for the same reason. Thus, over pursuing the simplicity and self-consistency limits the application of the research findings in practice to some extent. Even both Koch's MVD and Xiaojun's MVD show certain alignment with the illustrative example, the differences of code, engineering design details and information mass remains.

In terms of rigorous expression: Another limitation of this paper is that, the idea of the paper is to attempt integrating information theory, sandpile model and topology of MVD in a relatively self-consistent and explanatory way. Hence all rigorous proof of the homomorphisms, such as the homomorphisms mapping from BIM model to communication, the homomorphisms mapping from topology development of MVD to sandpile model are omitted for the sake of readability as well as page limits. Beyond that, the detection of chaos in the system is regarded as an influential contribution, the missing rigorous mathematical proof of chaos (Lypunov exponent based proof) restricts the crucial claim that the system is a chaotic system. Consequently, the expression is incomplete. As a result, the incompleteness of proof limits fractal discovery even if it can be clearly identified in the case of $L^2 = 101 \times 101$ sandpile simulation. Otherwise, the system might be optimized significantly through renormalization.

### 5.3. Future directions

In summary, it is concluded that the topology development of MVD is not mature to properly integrate information theory, chaos theory for the tasks with low uncertainty, high speed, and comprehensive interpretation at the same time. Therefore, more studies need to be conducted to improve and enlarge the application of Chaos theory along with graph theory in BIM field. Based on the findings, analysis and discussion above, key research directions for the future are highlighted in the following list:

(1) Studies on applying proposed approach to develop MVD in practice should be stipulated to further verify the applicability of research. Meanwhile, the process could be automated through an advanced algorithm which would pay more attention to, especially when it comes to the relation entities of MVD, how expert system and NLP techniques could be integrated with to propose a more comprehensive one.

(2) The more unified detailed information requirement analysis is supposed to advocated, which could serve the information mass calculation consistent across borders.

(3) 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.

(4) The findings and discussions show the potential of employing chaos/fractal thinking to the BIM field. Hence, the rigorous mathematical symbolic system is expected to build as the basis of mathematical modeling for the BIM model development system for the purpose of providing strict proof of the existence of chaos. As a result, the developed mature theories of chaos theory become applicable.

(5) 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.

(6) 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).

## 6. Conclusions

In the present work, a novel approach for interpreting and developing topology of MVD for shield tunnel lining using sandpile simulation in the context of information theory is proposed, which enhance the standardization of MVD development in shield tunnel areas. The proposed approach is strictly set on the basis of 4 plausible hypotheses. For mathematical modelling purposes, a.the applied sandpile model is customized to fit the MVD development and its according priori knowledge is given in group theory language; b. a denotation system is defined to facilitate mapping BIM topology development to applied sandpile model. The simulation with an initial configuration shows that taking results of simulation as guidance to interpret topology development of MVD is feasible and reasonable. The process of modelling the system and defining new symbols for the topology of BIM is referential. This study provides added value to the standardization of design information for shield tunnels, which algins with IFC-Tunnel Project. A tree-structured graph formulation containing information entropy opens a new way to characterize building elements, which can be more easily leveraged with computational methods to explore the more optimal MVD solution followed by least action principles.

Beyond that, a profound finding of this paper is that there are some essences under the appearance are across the scalability, it reveals the existence of chaos in the system of topology development of BIM model which is universal among all types of BIM model rather than tunnel lining BIM model. Under certain circumstances, the fractal could be easily recognized. From a methodological perspective, using the sandpile model, proves that the integration of information theory and BIM model development is attainable and by doing so enables more usefulness of graph theory. This study justifies the philosophy of (inter alia) exploring potentials of applied physics for building information modelling through identifying least action principles in the sandpile simulation. This study also demonstrates that information theory, chaos theory is not fully ready to be adopted in the AEC industry, and that there remains a research and development gap.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## References

[1] J. Abualdenien, A. Borrmann, A meta-model approach for formal specification and consistent management of multi-LOD building models, Adv. Eng. Inform. 40 (2019) 135, https://doi.org/10.1016/j.aei.2019.04.003.
[2] Y. An, Geological Data Extension for Subway Tunnel BIM models using a Linked Data Approach, 2017.
[3] K.P.B.D.R. Anderson, Model Selection and Multimodel Inference–A Practical Information-Theoretic Approach, 2002. doi:10.1007/b97636.

[4] J.B. Andre Borrmann, Koch Christian, Liebich Thomas, Muhic Sergej, 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 (Eds.), Building Information Modeling. Switzerland: Springer International Publishing AG, 2018.

[5] M.K. Andre Borrmann, Christian Koch, Jakob Beetz, Springer, Building Information Modeling, 2018.

[6] J. Angulo, J.C. Angulo, J.M. Angulo, An application of information theory to stochastic classical gravitational fields, Physica A 499 (2018) 129–141, https://doi.org/10.1016/j.physa.2018.02.009.

[7] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: An explanation of 1/f noise, Phys. Rev. Lett. 59 (1987) 381–384, https://doi.org/10.1103/PhysRevLett.59.381.

[8] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: An explanation of the 1/f noise, Phys. Rev. Lett. 59 (4) (1987) 381–384, https://doi.org/10.1103/PhysRevLett.59.381.

[9] A. Ben-Naim, Information Theory and Selected Applications: Springer Cham, 2023.

[10] D. Blokh, I. Stambler, The application of information theory for the research of aging and aging-related diseases, Prog. Neurobiol. 157 (2017) 158–173, https://doi.org/10.1016/j.pneurobio.2016.03.005.

[11] A. Borrmann, M. Flurl, J.R. Jubierre, R.-P. Mundani, E. Rank, Synchronous collaborative tunnel design based on consistency-preserving multi-scale models, Adv. Eng. Inform., 28(4) (2014) 499-517. doi:https://doi.org/10.1016/j.aei.2014.07.005.

[12] 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.

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

[14] C. Eastman, P.T.R. Sacks, K. Liston, BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors. Hoboken, NJ, USA: John Wiley & sons Inc, 2011.

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

[16] J.C. Castilla-Rho, G. Mariethoz, R. Rojas, M.S. Andersen, B.F.J. Kelly, An agent-based platform for simulating complex human-aquifer interactions in managed groundwater systems, Environ. Model. Softw. 73 (2015) 305–323, https://doi.org/10.1016/j.envsoft.2015.08.018.

[17] H. Chen, X. Luo, An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing, Adv. Eng. Inf. 42 (2019), 100959, https://doi.org/10.1016/j.aei.2019.100959.

[18] M. Creutz, Abelian sandpiles, Nucl. Phys. B Proc. Suppl. 20 (1991) 758–761, https://doi.org/10.1016/0920-5632(91)91016-D.

[19] G. de Novaes Pires Leite, G.T.M. da Cunha, J.G. dos Santos Junior, A.M. Araújo, P.A.C. Rosas, T. Stosic, O.A. Rosso, Alternative fault detection and diagnostic using information theory quantifiers based on vibration time-waveforms from condition monitoring systems: Application to operational wind turbines. Renew. Energy, 164 (2021) 1183-1194. doi:https://doi.org/10.1016/j.renene.2020.10.129.

[20] A. Delgado-Bonal, F.J. Martín-Torres, Human vision is determined based on information theory, Sci. Rep. 6 (2016) 36038, https://doi.org/10.1038/srep36038.

[21] China National Design code of Shield Tunnel Engineering, 2021.

[22] J. Feder, 2013. Fractals.

[23] I. Flavio Barbosa, L.M.Z. Maciel, I.E.M.V. IV Paulo, M. de Azevedo Marques, I.J.E.J. II Valdair Francisco Muglia, Radiological reports: a comparison between the transmission efficiency of information in free text and in structured reports, Clin. Sci. 65 (1) (2010) 15–21, https://doi.org/10.1590/S1807-59322010000100004.

[24] V.J.L. Gan, BIM-based graph data model for automatic generative design of modular buildings, Autom. Constr. 134 (2022), 104062, https://doi.org/10.1016/j.autcon.2021.104062.

[25] T. Gehring, C. Lupo, A. Kordts, D. Solar Nikolic, N. Jain, T. Rydberg, U.L. Andersen, Homodyne-based quantum random number generator at 2.9 Gbps secure against quantum side-information, NatureCommunications 12 (1) (2021), https://doi.org/10.1038/s41467-020-20813-w.

[26] H. Haken, Information and Self-Organization: A Macroscopic Approach to Complex Systems, Springer, 2006.

[27] L. Hao, *IFC based IFC shield tunnel structure life-cycle information modelling research and applicaztion* (Msc), Tongji University, Shanghai, 2016.

[28] Jean-Yves Chouinard, Paul Fortier, T.A. Gulliver, (1996, May 28 - 30, 1995). Information Theory and Applications II. Paper presented at the CWIT: Canadian Workshop on Information Theory, Lac Delage, Quebec, Canada.

[29] C. Koch, A. Vonthron, M. König, A tunnel information modelling framework to support management, simulations and visualisations in mechanised tunnelling projects, Autom. Constr. 83 (2017) 78–90, https://doi.org/10.1016/j.autcon.2017.07.006.

[30] M.K. Laakso, Ao., The IFC standard: A review of History, development, and standardization, Information Technology, ITcon, 2012. Retrieved from http://usir.salford.ac.uk/id/eprint/28373.

[31] M. Lang, M. Shkolnikov, Harmonic dynamics of the abelian sandpile, Proc. Natl. Acad. Sci. 116 (8) (2019) 2821, https://doi.org/10.1073/pnas.1812015116.

[32] G. Lee, Y.H. Park, S. Ham, Extended Process to Product Modeling (xPPM) for integrated and seamless IDM and MVD development, Adv. Eng. Inf. 27 (4) (2013) 636–651, https://doi.org/10.1016/j.aei.2013.08.004.

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

[34] O. Ludwig, U. Nunes, R. Araújo, L. Schnitman, H.A. Lepikson, Applications of information theory, genetic algorithms, and neural models to predict oil flow, Commun. Nonlinear Sci. Numer. Simul. 14 (7) (2009) 2870–2885, https://doi.org/10.1016/j.cnsns.2008.12.011.

[35] Michel Rives, A. B., Abdullah Alsahly, Heiko Atzbacher, Gabriele Brino, Arianna Bucci, Jure Cesnik, Federica Danise, Nataliya Dias, Jonathan Dunn, Jan Erik Hoel, Rie Kudoh-Wada, Greta Lucibello, Shoichi Nishiyama, Reza Rangsaz Osgoui, Florent Robert, Federica Sandrone, Andrew Sheil, Hiromasa Shima, Stefania Stefanizzi, Magdalena Stelzer, Frederikke Syversen, Jan-Christian Thoren, Mirko Vendramini, Simone Villa, Götz Vollmann, André Vonthron, Jonas Weil, Hichem Zammit, Nan Zhu, Report WP2: Requirements analysis report (RAR). Retrieved from 2020.

[36] L.J. Moniz, E.G. Cooch, S.P. Ellner, J.D. Nichols, J.M. Nichols, Application of information theory methods to food web reconstruction, Ecol. Model. 208 (2) (2007) 145–158, https://doi.org/10.1016/j.ecolmodel.2007.05.016.

[37] N. Passalis, A. Tefas, Entropy Optimized Feature-Based Bag-of-Words Representation for Information Retrieval, IEEE Trans. Knowl. Data Eng. 28 (7) (2016) 1664–1677, https://doi.org/10.1109/TKDE.2016.2545657.

[38] J.R. Pierce, An Introduction to Information Theory, Dover Publications Inc, Symbols, Signals and Noise, 1980.

[39] M.S. Preston, Motivating child welfare case managers: An application and extension of feedback information theory, Child Youth Serv. Rev. 35 (4) (2013) 734–741, https://doi.org/10.1016/j.childyouth.2013.01.014.

[40] R. Rakoczy, The application of the informational theory to the analysis of the grinding process under action of transverse rotating magnetic field, Powder Technol. 201 (2) (2010) 161–170, https://doi.org/10.1016/j.powtec.2010.03.021.

[41] C.E. Shannon, A Mathematical Theory of Communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[42] S. Uda, Application of information theory in systems biology, Biophys. Rev. 12 (2) (2020) 377–384, https://doi.org/10.1007/s12551-020-00665-w.

[43] D. Vries 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, 2005.

[44] G. Wang, Improving Data Transmission in Web Applications via the Translation between XML and JSON, 2011.

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

[46] Y.Z. Wang, J. Li, IFC based building structure construction design information model presentation, Inform. Technol. Civil Eng. (2014) 30–35.

[47] D.T. Weintrop, S. Tinker, R. Head, B.U. Wilensky, NetLogo Sandpile model, 2011. Retrieved from http://ccl.northwestern.edu/netlogo/models/Sandpile.

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

[49] Working Group No.2, I. T. A. (2000). *Guidelines for the Design of Shield Tunnel Lining.* Retrieved from France.

[50] S. Wu, Q. Shen, Y. Deng, J. Cheng, Natural-language-based intelligent retrieval engine for BIM object database, Comput. Ind. 108 (2019) 73–88, https://doi.org/10.1016/j.compind.2019.02.016.

[51] R.W. Yeung, Information Theory and Network Coding, Springer, New York, NY, New York, 2008.

[52] K. Yusof, M. Man, Efficiency of JSON approach for Data Extraction and Query Retrieval, Indonesian Journal of Electrical Engineering and Computer Science 4 (2016) 203, https://doi.org/10.11591/ijeecs.v4.i1.pp203-214.

[53] F. Zhang, X. Zhang, K. Cao, D. Li, C. Li, Information theory applications in error bit rate analysis of digital image watermarking, Comput. Electr. Eng. 39 (2) (2013) 309–318, https://doi.org/10.1016/j.compeleceng.2012.11.004.

[54] X. Zhao, T. Chen, Type of self-organized criticality model based on neural networks, Phys. Rev. E 65 (2) (2002), 026114, https://doi.org/10.1103/PhysRevE.65.026114.

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