



IFC-Based Indoor Space Extraction with Topological Graph Modeling

Zheng Zhao^{1,2,3}, Weiya Chen^{1,2,3}, Zhiyuan Guo^{1,2,3}, and Wenming Jiang^{1,4}

¹ Huazhong University of Science and Technology, Wuhan, China

² National Center of Technology Innovation for Digital Construction, Wuhan, China

³ International Joint Research Laboratory of Smart Construction, Wuhan, China

⁴ China Academy of Building Research, Beijing, China.

zheng_zhao@hust.edu.cn, weiya_chen@126.com,

guozhiyuan@hust.edu.cn, jiangwenming@cabrtech.com

Abstract

The semantic enrichment of building information models (BIMs) has been widely explored, with various approaches utilizing graph neural networks (GNNs) to infer the types of indoor spaces. However, there is a gap in the intermediate process that translates the building information model into a graph model suitable for GNNs. To address this problem, we propose a structured graph model designed to represent the attributes and topological relationships of indoor spaces for space type classification. Based on the Industry Foundation Classes (IFC) file format, we define the concept of indoor space and propose an automated method for its extraction. The extracted interior subspaces are analyzed based on their geometric and topological properties, focusing on their relationship to the overall layout of the building's interior spaces. The subspaces are represented as nodes in the graph model, and the edges between the nodes are defined according to the topological relationships between the subspaces. A result was carried out to demonstrate the effect of the proposed indoor space extraction algorithm, which provided the basis for the inference of indoor spatial semantic types in building information models.

1 Introduction

BIM is a technology and a series of processes used to produce, communicate, and analyze building models, widely used in architecture, engineering, and construction (AEC) (Eastman et al., 2011). The adoption of Building Information Modeling (BIM) in the architecture, engineering, and construction industry has brought many benefits, but it has also introduced interoperability issues among different stakeholders and their BIM platforms (Olofsson et al., 2008). The Industry Foundation Classes (IFC) schema was created by Building Smart to overcome this issue (buildingSMART, 2013). However, when

IFC model files are exchanged between different platforms in practice, there are issues of information loss or distortion (Pazlar et al., 2008). In practical modeling, space-related entities and information in BIM models are often overlooked. Therefore, space extraction and type classification are gradually becoming key tasks for the semantic enrichment of BIM models.

A space can be defined in a few different ways, based on its physical separation from other spaces, or based on the users' required function for the space (Ekholm et al., 2000). Thus there is not an only way to define and classify space type. A graph is a data structure composed of nodes and edges, where the nodes are connected by edges, and each node has zero or more associated edges (West, 2001). Graph representation is a traditional approach to designing building layouts. For example, a classic graph-based method used in space syntax is the justified plan graph, which analyzes building layouts by considering the relative size, shape, location, and orientation of rooms (Ostwald & Michael, 2011). With the introduction of graph neural networks (GNNs), a new field of graph structure learning has emerged in recent years, offering novel approaches to identifying indoor space types and enriching BIM semantics. Relational graph convolutional networks (RGCN) and traditional machine learning are used to infer the usage of rooms in public buildings, with the floor plan represented as a graph (Hu et al., 2020). Two methods are proposed and compared for predicting both the topology and the categories of rooms given a partial map. The map is represented as a graph, with rooms as nodes and their connectivity as edges (Aydemir et al., 2012). An improved algorithm based on Graph-SAGE has been proposed, which takes into account the features of edges representing the topological relationships between rooms and has achieved good results in the room type classification task (Wang et al., 2021). A few other papers focus on room type classification using different GNN models and graph representations, with tests conducted on the open-source database House-GAN (Verma & Jadeja, 2023; Paudel et al., 2021; Nauata et al., 2020).

The establishment of indoor spaces can be categorized into two main approaches. The first approach involves manually modeling space entities using different software during the design phase. For example, in Autodesk Revit, this is typically achieved by creating and defining rooms with the "Room" tool, which automatically assigns spatial boundaries based on walls, floors, and ceilings. The second approach involves automatically extracting indoor spaces from BIM models using geometric, topological, and semantic attributes. This method supplements spatial information that may be missing during the design phase. These methods can be classified based on the different types of model files used. A method to analyze the geometry and semantic information of 2D vector floor plans and automatically reconstruct the corresponding 3D building models has been proposed. This approach involves preprocessing wall lines using an odd-even-based method and employs the loop search method to extract indoor spaces (Zhu et al., 2014). Several studies have focused on indoor space reconstruction and subspace partitioning based on point cloud data (Tang et al., 2024; Yue et al., 2024; Patrick et al., 2021). The ArcGIS platform is chosen to integrate BIM data, and a method for extracting indoor space information in complex buildings is proposed. The connectivity between upper and lower floor spaces is determined by recognizing the polylines of floor slab hole symbols, and the indoor multi-floor spaces are extracted accordingly (Pang et al., 2018). However, there is a lack of discussion on indoor space extraction methods based on IFC files. Our approach builds upon and improves the method proposed by Pang et al. (2018), enhancing its effectiveness in extracting indoor spaces by incorporating IFC-based geometric data and enriching the semantic information of the spaces. Similarly, the space extraction task in the field of building energy modeling is also discussed and can be broadly categorized into the following types: graph- and rule-based algorithms, geometry computing-based algorithms, domain knowledge-based algorithms, hybrid algorithms, and 2D to 3D algorithms. The building topology graph (BTG) is introduced as a graph data structure where nodes represent building elements and edges connect adjacent building elements. A space recognition algorithm based on this graph structure and local features is proposed to support performance-oriented design in the early design stage (Chen et al., 2017).

Therefore, it is worth investigating methods for indoor space analysis in BIM models using IFC as input for semantic enrichment. Similarly, the application of GNNs for identifying indoor space functions relies on pre-existing graph models. It is worth studying how to automatically extract information from building information models to form graph models that can support GNN algorithms for semantic enrichment.

In this paper, we define indoor spaces and present an algorithm for their automatic extraction from IFC files. This algorithm involves analyzing and processing the geometric representations of BIM components within the IFC schema. We then introduce a graph model designed to integrate geometric and topological features of the extracted indoor spaces. The model serves as a bridge to transform the complex geometric information within IFC-based BIM models into a spatial graph structure that can effectively be processed by GNNs.

2 Method

2.1 Descriptions of Indoor Spaces

Building components are the basic elements that constitute a building, either individually or in combination with each other to fulfill the structural and functional requirements of the building. Common examples of building components include columns, slabs, walls, stairs, and windows. A space is defined as a portion of the building enclosed by the interconnection of these components. The elements that contribute to this enclosure are termed boundary components. These components define the indoor space by separating the interior of the building from its external environment. Moreover, boundary components subdivide the indoor space into smaller, functional subspaces, such as rooms, corridors, or stairwells, to satisfy diverse requirements within the building. After specifying the definition of the indoor space, it becomes essential to predefine the boundary components utilized in the space extraction algorithm. Table 1 details the semantic descriptions of all boundary component types, along with their respective entity names as specified in the IFC file.

| Component Name | Description | IFC Entity Name |
|----------------|---|-----------------|
| Wall | A primary vertical component that serves structural, enclosing, and partitioning functions. It is a key element in forming the boundaries of indoor spaces. | IfcWall |
| Curtain Wall | A non-load-bearing facade element, used on building exteriors to form the boundary between indoor spaces and the external environment. | IfcCurtainWall |
| Column | A primary vertical component that typically does not directly serve as a space partitioning element but influences space segmentation. | IfcColumn |
| Slab | A primary horizontal component that bears vertical loads. It defines floors in vertical space and acts as the upper or lower boundary of a room. | IfcSlab |

Table 1: Semantic descriptions of boundary components

Indoor spaces within a building are defined by the enclosure formed by various boundary components, which can be expressed using Equation (1):

$$S = \text{Interior} \left(\bigcup_{B_i \in B} \text{Geometry}(B_i) \right) \tag{1}$$

Here, the indoor space is determined by the geometric representations of boundary components and an operation to extract the enclosed region. Let $B = \{B_1, B_2, \dots, B_n\}$ represent the set of building boundary components, which B_i refers to individual components such as walls, floor slabs, and columns that contribute to enclosing the indoor space. The term $\text{Geometry}(B_i)$ denotes the geometric representation of each boundary component B_i . By performing a union operation on all boundary components and extracting the internal region, the actual enclosed indoor space S is obtained.

2.2 Indoor Space Extraction Algorithm

Buildings consist of various components distributed across different areas. The IFC schema organizes these elements into a hierarchical spatial structure using aggregation relationships. An IFC model file contains exactly one `IfcProject` entity, which represents the overall context of the project and acts as the parent object for all other spatial elements, including `IfcBuilding`, `IfcBuildingStorey`, and `IfcSpace`. Entities at different hierarchical levels are connected via the `IfcRelAggregates` relationship. This hierarchical structure is acyclic, meaning that lower-level entities cannot contain higher-level entities. It is worth noting that in real projects, designers often ignore the necessity of `IfcSpace` in the modeling phase, which makes it impossible to get an indoor space directly from the IFC file. In this case, we need to try other methods to extract the indoor space.

The indoor space extraction algorithm uses an IFC-based building information model as input, and it is assumed that all boundary components are complete and continuous geometries without missing parts. Furthermore, the elements must be properly connected to eliminate geometric overlaps or gaps.

The extraction algorithm is designed to focus on one specific storey of a building. Meanwhile, the `IfcBuildingStorey` entity is used to represent individual floors in IFC schema. Consequently, the first step is to parse the IFC file and divide the building into separate storeys. As shown in Figure 1, this model contains five `IfcBuildingStorey` entities, each corresponding to one of the five floors of the actual building. It is possible to identify all boundary components related to a specific storey through the `IfcRelContainedInSpatialStructure` relationship.

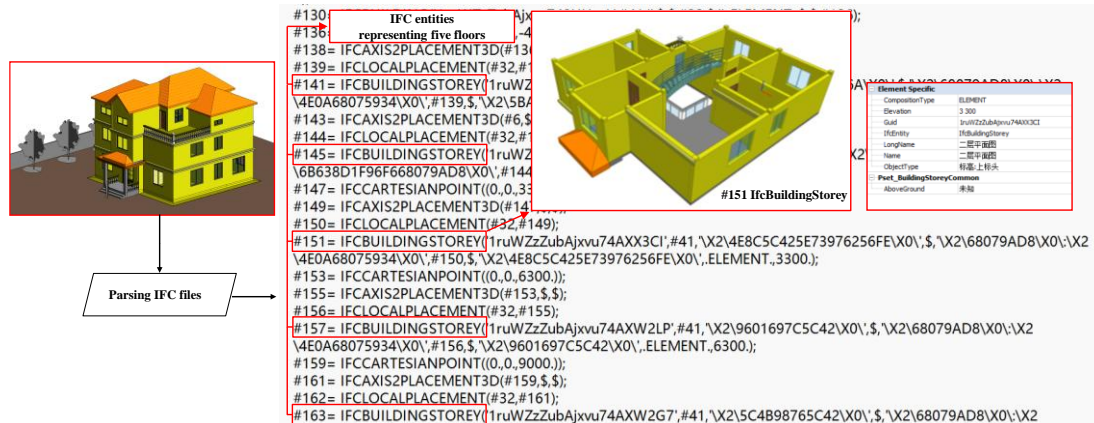


Figure 1: Extraction of all components from one storey in an IFC file, illustrating how floors are represented using the `IfcBuildingStorey` entity

Each building component in the IFC file contains detailed geometric information, which may be represented in various formats, such as Boundary Representation (BRep), Constructive Solid Geometry (CSG), or Extruded Solids. The geometric data of boundary components consists of a series of vertices and faces. Specifically, vertices are defined by their (x, y, z) coordinates, while faces are described by lists of vertex indices that define the surfaces of the geometry. Coordinate transformations and the generation of 3D meshes from this geometric data ensure that the information is appropriately prepared for subsequent processing steps.

Simplifying the model to a two-dimensional plane can effectively improve computational efficiency. Since the upper and lower boundaries of the indoor space are defined by the ceiling and the floor, a customized horizontal cutting plane is introduced between them. After obtaining the geometric information of all boundary components of the storey, these components are projected onto the cutting plane. At this stage, each boundary component B_i in the set $B = \{B_1, B_2, \dots, B_n\}$ corresponds to a projected polygon P_i . Subsequently, all the polygons P_i are merged through a union operation to form a unified geometric object P , as shown in Figure 2(a). The convex hull C of geometric shape P is computed, as shown in Figure 2(b). A Boolean difference operation is performed between the convex hull C and the shape P , where C is subtracted from P , resulting in the uncovered region set S , which represents the area not covered by the boundary components, as depicted in Figure 2(c). The set S consists of several regions represented by polygons, which can be classified into two categories based on their interaction with the boundary of the convex hull. The first category includes spaces that are separated from the outdoor environment by the outermost continuous building elements, such as walls and a few columns. These spaces are further subdivided by internal building components, resulting in distinct indoor subspaces. These subspaces represent the target spaces we need to extract, as shown by the blue areas in Figure 2(d). The other category consists of spaces located between the outermost components and the convex hull boundary, which are discarded as non-target outdoor spaces, indicated by the red areas in Figure 2(d).

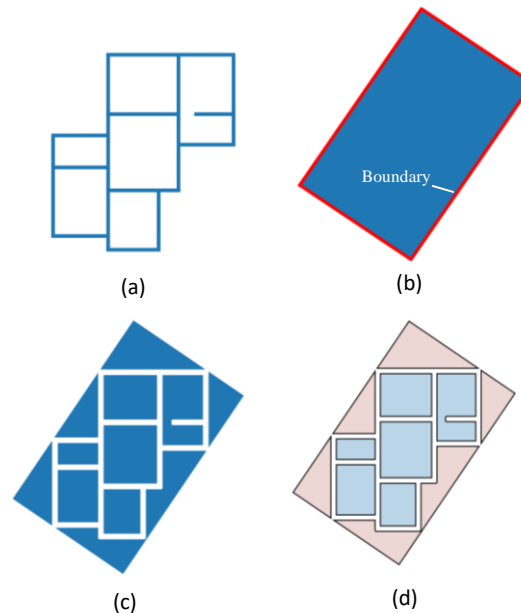


Figure 2: Visualization of the indoor space extraction process. (a) Projection of boundary components (b) Convex hull and its boundary (c) Boolean difference operation between the convex hull and boundary components (d) Classification of spaces into target indoor

2.3 The ISTAG Model

To represent the characteristics of indoor spaces within a building information model, this paper constructs the indoor space topology and attribute graph (ISTAG) model to describe the attributes and relationships of the spaces. This model is designed to support subsequent graph-based spatial analysis applications, enabling more efficient space utilization and functional analysis within the built environment.

$$G = (V, E) \quad (2)$$

The composition of the graph is represented as shown in Equation (2), where $V = \{v_1, v_2, \dots, v_n\}$ represents a finite set of nodes, with n denoting the number of nodes in the set V , each of which corresponds to an indoor subspace. Each subspace node has two types of attributes: geometric properties and topological information of the space. $E = \{e_1, e_2, \dots, e_m\}$ is the set of undirected edges, representing the connection relationships between two indoor subspaces. An edge $e = (v_i, v_j)$, where $1 \leq i, j \leq n$ and $i \neq j$, indicates a connection between nodes v_i and v_j . The edge attributes include connectivity and adjacency relations.

1. Geometric Attribute of Indoor Spaces

In this section, the geometric features of indoor spaces in building information models are extracted and processed to construct the node feature information of ISTAG models. These geometric attributes include both basic geometric information (such as aspect ratio) and processed measurement values, which are used to more effectively describe spatial characteristics. A detailed description of each geometric attribute is given below:

Proportion of room area relative to total indoor floor area. The area is one of the most fundamental geometric attributes of a room. The function of a room is closely related to its size, and the ratio of a room's area to the total indoor floor area serves as a key distinguishing feature for different room types. For example, smaller spaces are typically used for storage or utility purposes, whereas larger spaces are more likely to serve as living rooms.

Aspect Ratio. The aspect ratio, defined as the ratio of a room's length to its width, is a key indicator of a room's shape. Spaces with distinct functions tend to exhibit varying aspect ratios. For example, corridors, which serve as connecting spaces between rooms, typically have larger aspect ratios, whereas rooms with an aspect ratio close to 1:1 are more commonly associated with living spaces such as kitchens or living rooms.

Compactness. Compactness is measured as the ratio of a room's area to its perimeter, indicating the degree of geometric regularity. Spaces with high compactness often exhibit regular geometric shapes, whereas spaces with low compactness tend to have irregular or elongated forms.

Number of Boundary Lines. The number of boundary lines represents the count of edges that define a space's boundaries. Rooms with fewer boundary lines generally correspond to simpler geometric shapes, such as squares or rectangles, which are commonly found in primary functional spaces like living rooms, bedrooms, or offices. In contrast, spaces with more boundary lines typically correspond to complex or irregular shapes, or areas designed with specific functions. This attribute serves as a valuable reference for tasks such as space classification and spatial region identification.

Distance Between Subspace Center and Floor Center. This value measures the distance between the geometric centroid of a room and the center of the minimum oriented bounding box (OBB) that encompasses the entire floor space. It reflects the spatial positioning of the room within the floor layout. Rooms situated in the core areas of a building are often functionally more significant compared to those located at the edges or corners.

2. Topological Attribute of Indoor Spaces

This paper introduces two types of topological relationships for indoor spaces: connectivity and adjacency. To support the establishment of edges in the ISTAG model, we propose a method for constructing indoor topology based on the IFC building information model. By extracting topological information between spaces, the logical structure of the building layout can be more effectively characterized, laying a solid foundation for advanced spatial analysis and functional inference.

Connectivity refers to whether different enclosed spaces can connect through openings (such as doors or windows), thereby enabling the movement of people or objects between spaces. In this study, the determination of spatial connectivity is primarily based on doors as critical components. Specifically, two enclosed spaces are considered connected if they are directly linked by at least one door. To evaluate the connectivity between spaces, the geometric representations of all doors (IfcDoor) on a given floor are extracted during the parsing of the IFC file. These geometries, along with other boundary components, are projected onto a two-dimensional plane. Figure 3(a) illustrates the results of an indoor space extraction algorithm, showing several subspaces $\{S_1 \sim S_6\}$.

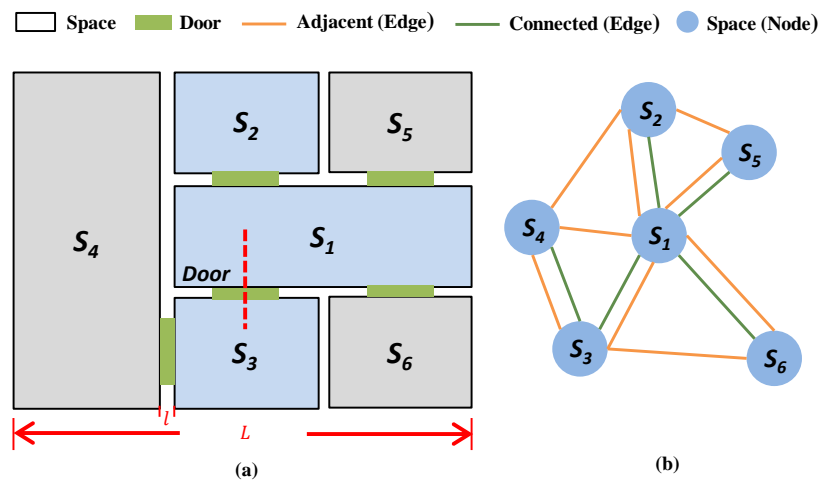


Figure 3: Topological analysis of indoor spaces for the ISTAG model

Each door (represented as green polygons) naturally connects two indoor subspaces. If a door is associated with only one subspace, it is considered as leading to the exterior. To ensure precise identification of connected spaces, a line segment perpendicular to each door element is drawn, as illustrated by the red dashed lines in Figure 3(a). In this study, the length of this segment is set equal to the door's length. This fixed length ensures that a single door does not erroneously associate with multiple spaces. For example, as shown in the blue-highlighted area in Figure 3(a), without this constraint, the door element could incorrectly connect both Space S_1 and Space S_2 . Connectivity emphasizes the accessibility between spaces, whereas adjacency focuses more on the proximity of two spaces within the layout. If two spaces share one or more boundary components (typically walls), they are considered adjacent spaces. Due to the presence of gaps between indoor subspaces, resulting from the Boolean difference operation applied to wall components during the indoor space extraction process, adjacency cannot be determined by whether spaces share a boundary. Therefore, we define a threshold to determine if two spaces are "in contact." For example, in Figure 3(a), if the distance l between two spaces is less than 5% of the floor plan length L , the two spaces are considered adjacent. In this case, an edge is created between the corresponding nodes in the ISTAG model to represent the adjacency relationship.

In the final ISTAG model, as shown in Figure 3(b), blue nodes represent indoor subspaces $\{S_1 \sim S_6\}$. The adjacency and connectivity relationships are differentiated by different edge types. Adjacency

relationships are represented by orange edges, while connectivity relationships, are represented by green edges.

3 Results

In this section, the visualization results are presented to validate the indoor space extraction algorithm based on the IFC model and the ISTAG model representation. First, individual floors were extracted from the building information model, and the results are shown in the left panel of Figure 4. The yellow components are extracted as boundary components enclosing the indoor space, so the red lines in the right picture represent the indoor space. The blue part corresponds to the outdoor space, whose boundary is defined by the outer railing (IfcRailing), which is not a predefined boundary component.

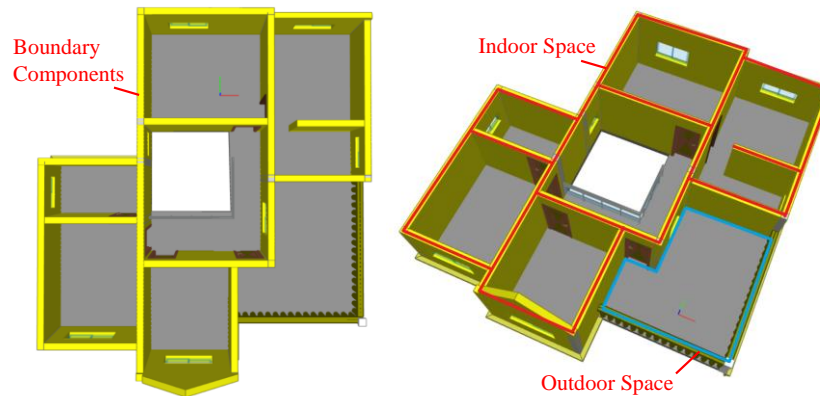


Figure 4: Boundary components extracted from IFC files

After the operation of the indoor space extraction algorithm, the indoor space is successfully extracted as shown in Figure 5. The geometric attributes are then computed and extracted for each indoor subspace, and the output is shown on the right. We take each indoor subspace as a node in the ISTAG model and the geometric attributes of the space as node features. It is worth noting that when a door is connected to the only indoor space, this door is considered to be connected to the outdoor space. Then this door is not considered when performing the topological information extraction of the indoor space.

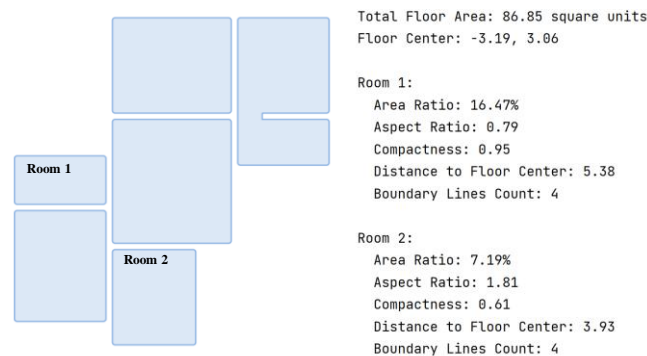


Figure 5: Geometric attributes of indoor spaces analysis results

The final results of the indoor space connectivity and adjacency visualization built according to the method are shown in Figure 6.

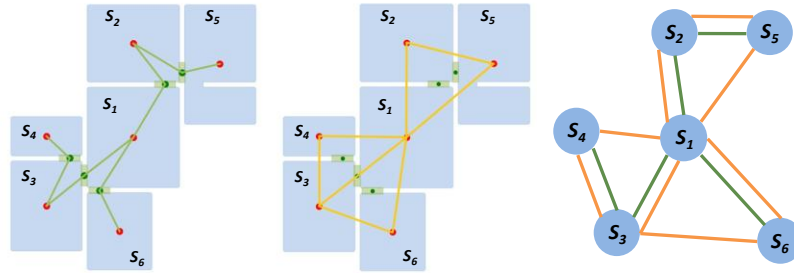


Figure 6: Visualization of ISTAG model construction. The green line indicates the connectivity established between spaces through doors. The orange line is used to indicate the adjacency of two spaces. On the right side is the ISTAG model established by the method

4 Discussion

In this paper, an indoor space extraction method is designed for models in IFC file format to supplement the lack of spatial semantic information. On this basis, the interior subspace is analyzed to construct the ISTAG model. The research gap from raw BIM model files to the input of the GNNs algorithm is effectively filled. However, the current study still has some limitations, such as the inability to automatically identify virtual boundaries, and the degree of spatial connectivity or adjacency is not discussed yet. These issues are our next step work.

5 Conclusions

In conclusion, this paper proposed an indoor space extraction algorithm based on the IFC files, focusing on accurately processing BIM models using boundary components like walls, slabs, and columns. The algorithm effectively extracts indoor spaces from building models. The ISTAG model plays a crucial role in bridging the gap between BIM data and advanced spatial analysis using GNNs. The current spatial extraction algorithm is not able to deal with the virtual boundaries of indoor spaces, and the degree of spatial connectivity is not quantified in the ISTAG model. These aspects present valuable opportunities for future research.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (No. 2022YFC3802100) and the National Natural Science Foundation of China (Grant No. 72001086).

References

- Aydemir, A., Jensfelt, P., & Folkesson, J. (2012). What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 4675–4682. <https://doi.org/10.1109/IROS.2012.6386110>
- buildingSMART. (2013). IFC4 Release Candidate 4. Retrieved from buildingSMART website: <http://www.buildingsmart-tech.org/ifc/IFC2x4/rc4/html/index.htm>
- Chen, Hongzhong & Li, Ziwei & Wang, Xiran & Lin, Borong. (2017). A graph- and feature-based building space recognition algorithm for performance simulation in the early design stage, *Building Simulation*, 11. 10.1007/s12273-017-0412-x.
- Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2011). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors* (2nd ed.). John Wiley & Sons.
- Ekholm, Anders & Fridqvist, A. (2000). Concept of space for building classification, product modelling, and design, *Automation in Construction*, 9 (3) (2000) pp. 315-328. 9. 10.1016/S0926-5805(99)00013-8.
- Hu, Xuke & Fan, Hongchao & Noskov, Alexey & Wang, Zhiyong & Zipf, Alexander & Gu, Fuqiang & Shang, Jianga. (2020). Room semantics inference using random forest and relational graph convolutional networks: A case study of research building, *Transactions in GIS*, 25. 10.1111/tgis.12664.
- Hübner, P., Weinmann, M., Wursthorn, S., & Hinz, S. (2021). Automatic voxel-based 3D indoor reconstruction and room partitioning from triangle meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 254–278.
- Nauata, Nelson & Chang, Kai-Hung & Cheng, Chin-Yi & Mori, Greg & Furukawa, Yasutaka. (2020). House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation, 10.1007/978-3-030-58452-8_10.
- Olofsson, Thomas & Lee, Ghang & Eastman, Charles. (2008). Editorial - Case studies of bim in use, 13. 244-245.
- Ostwald, Michael. (2011). The Mathematics of Spatial Configuration: Revisiting, Revising and Critiquing Justified Plan Graph Theory, *Nexus Network Journal*, 13. 445-470. 10.1007/s00004-011-0075-3.
- Pang, Yueyong & Zhang, Chi & Zhou, Liangcheng & Lin, Bingxian & Lü, Guonian. (2018). Extracting Indoor Space Information in Complex Building Environments, *ISPRS International Journal of Geo-Information*, 7. 321. 10.3390/ijgi7080321.
- Paudel, Abhishek & Dhakal, Roshan & Bhattarai, Sakshat. (2021). Room Classification on Floor Plan Graphs using Graph Neural Networks, 10.48550/arXiv.2108.05947.
- Pazlar, Tomaž & Turk, Žiga. (2008). Interoperability in practice: Geometric data exchange using the IFC standard, *Electronic Journal of Information Technology in Construction*, 13. 362-380.
- Tang, S., Huang, J., Cai, B., Du, H., Zhou, B., Zhao, Z., Li, Y., Wang, W., & Guo, R. (2024). Back to geometry: Efficient indoor space segmentation from point clouds by 2D–3D geometry constrains. *International Journal of Applied Earth Observation and Geoinformation*, 135, 104265.
- Verma, Atul Kumar & Jadeja, Mahipal. (2023). CB-SAGE: A novel centrality based graph neural network for floor plan classification. *Engineering Applications of Artificial Intelligence*, 126. 107121. 10.1016/j.engappai.2023.107121.
- Wang, Zijian & Sacks, Rafael & Yeung, Timson. (2021). Exploring Graph Neural Networks for Semantic Enrichment: Room Type Classification, *Automation in Construction*, 134. 10.1016/j.autcon.2021.104039.
- West, D.B. (2001) *Introduction to Graph Theory*. 2nd Edition, Prentice-Hall, Inc., Upper Saddle River, 82-83.

Yue, H., Wu, H., Lehtola, V., Wei, J., & Liu, C. (2024). Indoor functional subspace division from point clouds based on graph neural network. *International Journal of Applied Earth Observation and Geoinformation*, 127, 103656.

Zhu, J., Zhang, H., & Wen, Y. (2014). A New Reconstruction Method for 3D Buildings from 2D Vector Floor Plan. *Computer-Aided Design and Applications*, 11(6), 704–714.