Automation in Construction 91 (2018) 31–43

Contents lists available at ScienceDirect

Automation in Construction

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

BIM log mining: Discovering social networks

Limao Zhang[a], Baabak Ashuri[b,*]

[a] School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore
[b] Economics of Sustainable Built Environment (ESBE) Lab, School of Building Construction and School of Civil & Environmental Engineering, Georgia Institute of Technology, 280 Ferst Drive, Atlanta, GA 30332-0680, United States

ARTICLE INFO

Keywords:
Social network
BIM
Log mining
Collaborative design
Revit
Production

ABSTRACT

This research develops a systematic methodology to deeply mine tremendous volumes of design logs (that are generated from Building Information Model (BIM) design process) to discover social networks in BIM-based collaborative design practices and examine the relationship between the characteristics of the design social network and production performance of designers. Firstly, a data extraction procedure consisting of data harvesting, parsing, and cleaning is proposed to obtain BIM design logs in a Comma Separated Values (CSV) format from several designers over the course of working on multiple projects. Secondly, a metric of working together on joint cases is proposed to build up a weighted sociogram that is consisting of performers *P*, relations *R*, and weights *W*. Lastly, a number of indicators are defined to measure and analyze structural characteristics of the discovered BIM-based collaborative network at macro-, meso-, and micro- levels. A dataset of design logs that involves 51 designers working on 82 projects with 620,492 lines of commands, provided by a major international design firm, is used as a case study to demonstrate the feasibility and applicability of the developed approach in this research. Results indicate that: (i) Strong positive correlations exist across all centrality measures calculated based on the discovered social network of BIM-based collaborative design where designers located in the center of the interaction map (with the greatest degree centralities), such as designers "#2" and "#24", are generally those who provide the shortest communication channels (with highest betweenness centralities) and are most reachable for others (with highest closeness centralities); and (ii) All the node centrality measures are significantly and positively related to the production performance of designers in the BIM-based collaborative network. Particularly, the measured node degree centrality by weight is capable of explaining the greatest percentage of variations (71.13%) in the production performance of designers. This research contributes to: (a) The state of knowledge by proposing a novel methodology that is capable of capturing and modeling collaborations among designers from tremendous event logs to discover social networks; and (b) The state of practice by providing insight into a better understanding of relationships between sociological network characteristics and production performance of designers within a design firm.

1. Introduction

Given the fact that design projects become larger and more complicated [41,58], Building Information Modeling (BIM) provides a collaborative platform to facilitate information exchange and sharing among various specialty design participants for better decision-making [41]. Indeed, collaboration is a critical factor that determines the success of design firms [4]. As Architecture, Engineering and Construction (AEC) firms have become increasingly dependent on cumulative knowledge of key stakeholders for maintaining firm competitiveness, collaboration studies have attracted increasing attention in recent years [23,62]. The key challenge is to identify and develop appropriate collaboration models for improving production rate. A need exists in exploring and incorporating differences of team members and groups within a firm, where the differences come from pre-existing work patterns and/or expectations by collaborating groups and individuals [11,50]. According to Durugbo et al. [15], collaboration requires a network where: (i) Individuals/groups are interconnected [43,63]; (ii) Tasks/processes are linked [32,60]; (iii) Differences of group members are explored and integrated [49]; and (iv) Collaboration is dependent on decision making, teamwork, and coordination [33,51].

The body of knowledge on complex networks research aims to better understand collaboration features of an organization [3,37] and shed lights into its decision-making process [18]. Social Network Analysis (SNA) is the main approach adopted within the complex network research to identify and analyze relationships, social roles, and

* Corresponding author.
E-mail addresses: limao.zhang@ntu.edu.sg (L. Zhang), baabak@gatech.edu (B. Ashuri).

https://doi.org/10.1016/j.autcon.2018.03.009
Received 8 September 2017; Received in revised form 6 February 2018; Accepted 2 March 2018

social structure in collaborative organizations [5,23]. SNA uses sociological and mathematical techniques to represent and quantify an organization's information structure [24]. SNA has been applied in characterizing collaborative organizations, such as virtual design and construction (VDC) networks [54], energy saving networks [38], scientific collaboration networks [40], and complex project coordination networks [9]. A constructed social network can be analyzed to (i) Identify patterns of relations and involvements (centralized vs. decentralized) [15]; (ii) Determine most important actors and most critical connections in the network [14,29]; and (iii) Detect subgroups or subsets of the network that are especially cohesive [2,39]. In general, SNA is a powerful tool to study complex systems, but the main issue is the daunting task of data collection, as a network consisting of n actors requires the identification of the properties of $n(n\text{-}1)/2$ pairs of actors [34]. To address this issue, this research aims to investigate the possibility of extracting implicit process information and inter-organizational collaborations from design log data in BIM applications for mining social networks within a design organization.

In recent years, BIM has been widely adopted in the AEC industry all over the world. The growing utilization of BIM applications has resulted in tremendous volumes of computer-generated data, which provides a unique opportunity to extract valuable information. Particularly, design logs are rich data sources stored in Autodesk Revit journal files that record all modeling activities in design sessions [45]. The design logs are initially proposed to assist diagnosis and troubleshooting for software support staff. Most recently, several efforts have been made to use BIM design logs for other purposes. For instance, Yarmohammadi et al. [64] used a sequential pattern mining algorithm, the generalized suffix tree, to capture common command execution patterns and structures from Revit journal files. Zhang et al. [65] proposed a new metric on monitoring, measuring, and analyzing design productivity by utilizing design logs as a non-intrusive data capturing mechanism and further examined the impact of learning effects on the productivity improvement. The main advantage of mining BIM design logs lies in the rich information stored in design sessions that capture compute-user interactions, designer's preferences in executing design tasks, productivity rates of individual designers and the design team, and collaboration characteristics of the design team, all together in just one place. Process mining is an emerging machine learning methodology that aims to exploit information in event logs captured from systems and processes to explore, monitor, and improve processes [52,55]. Thus, this research attempts to develop a systematic social network mining approach that is capable of extracting implicit process information from BIM design logs for the discovery of collaborations in BIM-based design projects.

The main research questions are: (i) How can design logs data be utilized to capture collaborations among designers? (ii) How can a social network of the design team be created extracting information from BIM log files? and (iii) How can the developed social network be used to explore relationships between collaborative characteristics and designer's production within a design firm? In this research, the feasibility of using Revit journal files as a non-intrusive data collection approach to capture designers' interactions with the software is examined. A systematic methodology with detailed steps necessary to collect, process, and analyze data to discover social networks of collaborations among designers is developed. Design logs provided by a large international design firm are utilized as a case to verify the feasibility and applicability of the proposed approach. This research contributes to (i) The state of the knowledge by proposing a novel methodology that is capable of objectively capturing collaborations among actors from event logs to discover social networks; and (ii) The state of practice by providing insight into the design process through better understanding of relationships between network characteristics and production performance of designers for enhancing the likelihood of project success within organizations.

The remainder of the paper is organized as follows. Section 2 reviews the literature on BIM data reusing and social network mining. Section 3 presents a novel systematic approach to utilize information embedded in design log files for quantifying the main social network attributes of the collaborative design team. The proposed process consists of three main steps: (i) Data extraction, (ii) Social network modeling, and (iii) Social network analysis. In Section 4, the real data resources collected from a large international design firm are presented. In Section 5, the social networks of collaborations among designers are built and analyzed in detail. In Section 6, the relationship between network characteristics and production performance of designers is further investigated. In Section 7, the conclusions and future works are drawn.

2. Literature review

2.1. BIM data reusing

Several benefits are found for the wide adoption of BIM technologies in the AEC industry: (i) Improved visualization [67]; (ii) Improved productivity due to easy retrieval of information; (iii) Increased coordination of construction documents [17]; (iv) Standardized platform for collaboration among multiple participants; (v) Embedded and linked vital information; (vi) Increased speed of delivery [44]; and (vii) Reduced costs [59]. Deployment of BIM technologies results in the creation of a substantial amount of data recorded in several forms, such as industry foundation classes (IFC) files [12,31] and event logs [65]. Specifically, Autodesk Revit automatically records all design event logs, as well as executed commands in Revit journal files [45].Thus, creating, sharing, reusing, and integrating BIM data throughout the building lifecycle represents a new opportunity for the AEC industry [35].

A number of research efforts have been made to extract useful information from BIM-generated documents for a wide range of applications. For instance, Hu et al. [27] used data from historical Tekla BIM models and developed a regression model to estimate man-hour quantity for steel fabrication projects. Abdelmohsen et al. [1] developed a cost analysis model based on the elements' dimensions captured from BIM models. Du et al. [13] and Liu et al. [36] utilized a cloud-based approach for performance benchmarking of BIM users. Zhang et al. [66] developed a BIM-based Risk Identification Expert System (B-RIES) to automate the identification of safety risk in tunneling projects, where the system extracted information from BIM model components, including geometry, property and relation information of engineering parameters. Chen and Chu [8] used building geometry information extracted from BIM models and developed a time-dependent vehicle routing problem (TDVRP) model that was capable of enhancing accuracy and efficiency for in-building rescue and evacuation operations. Gerrish et al. [21] enhanced visualization for building energy performance through capturing, assembling, and linking data from several disparate BIM models.

Overall, the existing research efforts mainly emphasis on utilizing data stored in BIM models to enhance non-design practices, such as cost estimation, construction risk identification, facility management, and energy simulation. Very few studies, however, have been done to capture and analyze BIM design process data (i.e., event logs) for enhancing productivity and competitiveness within organizations. Zhang et al. [65] indicated that challenges in re-using BIM design data mainly stem from two aspects: (i) design event logs, stored in a plain text format, are basically unstructured with noise, and thus, are difficult for data processing and modeling; and (ii) the volume of the design event logs are tremendous and require computational methods that are capable of dealing with a huge amount of data. To address this gap, this research attempts to utilize BIM design event logs as a non-intrusive data collection approach to discover and analyze social networks of collaboration among the design project team.