Check for updates

Detecting Information Bottlenecks in Architecture Engineering Construction Projects for Integrative Project Management

J. D. Thekinen, Ph.D.[1]; Nishchal Pandey[2]; Sinem Mollaoglu, Ph.D.[3]; Meltem Duva[4]; Kenneth Frank, Ph.D.[5]; and Dong Zhao, Ph.D., M.ASCE[6]

Abstract: Information bottlenecks are a central issue in architecture, engineering, and construction (AEC) projects. It is crucial to detect their occurrence to mitigate their impacts. Existing approaches to detect information bottlenecks use metrics that require estimating the quantity of information published, which is a challenging task in practice. In this paper, building on the literature and recent developments in automation, we introduce a revised model that uses unresolved activities during project delivery and frequency of information publication actions to detect information bottlenecks. We validated our model by applying it to a case study. We compared bottlenecks detected by our model to ground truth obtained in the case study independently using observational data and external verification. We further study how project activities requiring different levels of interactions among project participants with different roles (e.g., designer, contractor) affect our model's performance. Results support our model and show that it is possible to attain comparable performance by considering only high-interdependency activities. Our research enables teams and project managers to keep a focus on issues that require intense communication among different experts and organizations to avoid bottleneck. Our contribution to the body of knowledge includes a revised model that detects information bottlenecks during project delivery and insights into integrative project management where the focus is on both project activities and collaboration needs across roles to streamline the whole process. **DOI: 10.1061/JCEMD4.COENG-13019.**

Introduction

Group decision-making is a complex process in architecture, engineering, and construction (AEC) project delivery, given the fragmented nature of the construction industry, the barriers set up by traditional contractual relationships, and the conflicted roles and responsibilities of project teams (Issa et al. 2006). Readiness of information is critical to ensure smooth coordination between interdisciplinary team members. The lack of readiness affects coherence between teams, thereby causing significant cost overruns and delays in project delivery (Bashir et al. 2022; Xu and Luo 2014; Tribelsky and Sacks 2011).

A bottleneck is a point of congestion in a production system that occurs when workloads arrive faster than the capacity of the production process (Leporis and Králová 2010). An information bottleneck in AEC projects is the failure to provide project-related information by team members (Tribelsky and Sacks 2010). Timely flow of information between team members, in the form of technical drawings and task specifications, is vital for project management (Eckert et al. 2001; Moreau and Back 2000) because it can lead to faster issue resolution, improved project outcomes, and lower probability for claims and disputes (Ali et al. 2023). Timely information is critical for project progress, particularly for distributed design teams in any discipline (Back and Moreau 2000). According to McKinsey and Company (2016), over 80% of all construction projects fail to meet their target schedules. Bashir et al. (2022) highlights the importance of information flow interdependencies for successful construction project management. When information is withheld, inaccurate, or erroneous, waste ensues. Love et al. (2008a) analyzed a construction project case study to identify the causal behaviors that led to rework as a result of design process problems and found the project manager's failure to recognize the criticality of information flow among teams to be a key root cause. Xu and Luo (2014) performed observational analysis on information management in construction projects to provide comprehensive statistics on time wasted due to inconsistent, dislocated, and ambiguous information. Therefore, identifying bottlenecks before they occur during project delivery can help improve team efficiency and performance outcomes.

Tribelsky and Sacks (2010) introduced development velocity (DV) and work in progress (WIP) as two explanatory variables to detect information bottlenecks in AEC projects. A drawback of their approach is that DV and WIP rely on estimating the total quantity of information, which is not feasible in practice. The total quantity of information is estimated by calculating the package sizes of all information objects. Information objects in AEC projects are building components such as ceilings or floors. There

[1]Assistant Professor, Dept. of Mechanical and Manufacturing Engineering, Univ. of Calgary, Calgary, AB, Canada T2N4L5 (corresponding author). ORCID: https://orcid.org/0000-0003-0431-9812. Email: joseph.thekinen@ucalgary.ca

[2]School of Planning, Design, and Construction, Michigan State Univ., East Lansing, MI 48824. Email: pandeyni@msu.edu

[3]Professor, School of Planning, Design, and Construction, Michigan State Univ., East Lansing, MI 48824. ORCID: https://orcid.org/0000-0001-9228-7093. Email: sinemm@msu.edu

[4]School of Planning, Design, and Construction, Michigan State Univ., East Lansing, MI 48824. Email: duvamelt@msu.edu

[5]Professor, College of Education, Michigan State Univ., East Lansing, MI 48824. Email: kenfrank@msu.edu

[6]Associate Professor, School of Planning, Design, and Construction, Michigan State Univ., East Lansing, MI 48824. Email: dzhao@msu.edu

Note. This manuscript was submitted on August 23, 2022; approved on July 26, 2023; published online on October 10, 2023. Discussion period open until March 10, 2024; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

could be thousands of such objects, with each object having several attributes such as dimensions on a drawing. It is labor intensive to calculate the total quantity of information by counting all attributes from all information objects in the project. Further, Tribelsky and Sacks (2010) do not validate the suggested model to ground truths.

In search of a model that can detect information bottlenecks across all phases of project delivery in AEC projects while accounting for the limitations listed previously, we reformulate DV and WIP. First, we use frequency of information publication action instead of quantity of information in our formulation of DV. A publication action is a transfer of any information from the private to public domain (Tribelsky and Sacks 2010). Publication actions are either using emails with attachments or file-sharing platforms. Second, our formulation of WIP depends only on the timeline of project activities. Project activities are unresolved project issues discussed and tracked during project team meetings. An earlier work by the authors showed that the timeline of project activities can be automatically tracked from project team meeting minutes documents using Jaccard similarity (Bayhan et al. 2021). For these reasons, our formulation of DV and WIP, as described in later sections of this paper, is more feasible to implement in practice than in existing literature.

We validate our model by comparing our model with ground truth on an AEC case study project. We obtained the ground truth independently from ethnographic data and verified them via an interview with an industry representative who worked as the primary project manager as the owner's representative throughout the whole project delivery process. Ethnographic data here refer to observational data collected from attending project meetings and interacting with the project team (Wimmer and Dominick 2013). We used techniques such as diary-keeping of meeting discussion; analysis of meeting documents such as agenda, minutes, and team video and voice recordings; and interactions with team members to collect ethnographic data. This is the first work that validates a model to detect information bottlenecks with ground truth data. Thus, this paper serves as a baseline to compare the accuracy of future models.

We further study how project activities with different types of collaboration needs and interdependencies among project participants for resolution (Bell and Kozlowski 2002) affect our model performance. Project activities exhibit different types of interdependencies, representing the nature of the production process, including pooled, sequential, reciprocal, and intensive (Bell and Kozlowski 2002). In pooled-type activities, team members from a single expertise area or organization independently do the work. Sequential, reciprocal, and intensive activities involve team members with multiple expertise areas. In sequential activities, work flows in a uniform direction from one team member to the next until it is resolved. In reciprocal activities, work flows back and forth between team members. Intensive activities demand the most collaborative teamwork between team members. Interdependency between team members is lowest for pooled-type and sequential-type activities and highest for reciprocal-type and intensive-type activities. We show how activities with high interdependencies have more influence on bottleneck occurrence as compared to activities with low interdependencies. We refer to pooled and sequential activities as low-complexity (LC) activities and reciprocal and intensive activities as high-complexity (HC) activities. We also show that we can achieve good performance to detect information bottlenecks considering only HC activities. Thus, we further simplify our model because our result allows excluding estimating LC activities to detect information bottlenecks.

Literature Review

The decision-making process in the AEC industry is highly fragmented among stakeholders in different organizations with varying levels of expertise (Issa et al. 2006). Coordination among stakeholders is critical for a product that meets all the owner requirements (Cavka 2010). The AEC industry is both information intensive and information dependent to meet project costs and schedule goals (Back and Moreau 2000). Such reliance on information implies that detecting information bottlenecks and taking mitigation steps are critical to project success.

However, most of the existing research relating to information in the AEC industry focuses on issues such as information management (Mak 2001), information security in construction supply chains (Hijazi et al. 2019), computing in construction (Miyatake and Kangari 1993), building information modeling (BIM) (Tang et al. 2019; Liu et al. 2021; Jung and Joo 2011; Bryde et al. 2013), and digital technology in construction (Wong et al. 2018). Although numerous works study the importance of information management and its use in the AEC domain, studies relating to information bottlenecks have been few and far between.

The overall performance of project delivery, spanning the design and construction phases, depends on the weakest link. When a designer underperforms because of additional workloads or unavailability of information, the entire project team is affected (Tribelsky and Sacks 2006). Significant portions of information flow for design are devoted to waiting, inspection, conversion, and moving (Ballard and Koskela 1998) and are highly prone to creating variability in the work processes and leading to wasteful or non–value-added episodes (Freire and Alarcón 2002; Ko and Chung 2014) during project delivery. As such, design activities and the overlap of design and production processes have long been studied through lean principles (Howell et al. 2011) and are important to address for information bottlenecks across project delivery. A significant portion of design time is idle time (Freire and Alarcón 2002); therefore, it is important to address information bottlenecks. Grønbæk et al. (1993) studied information bottlenecks in cooperative tasks in the AEC domain. However, their focus was on challenges created by such bottlenecks and the need for computer-supported systems to address those challenges. Cavka (2010) studied bottlenecks by observing ethnographic data from 27 design coordination meetings from the design development phase of a building construction project. However, their focus was on bottlenecks in meeting processes to improve meeting efficiency through new technologies. Love et al. (2008b) studied bottlenecks due to design-induced rework. Tory et al. (2008) compared various information exchange artifacts in coordination meetings to recommend computer-supported systems that minimize information bottlenecks. However, they focused only on coordination meetings while ignoring other channels of information exchange such as file-sharing platforms and emails. Chinowsky et al. (2011) looked at aligning actual information exchanged with information sharing requirements in the project network. However, none of these works addresses the problem of detecting information bottlenecks.

Existing approaches to detect information bottlenecks are infeasible to apply in practice. Metrics used to quantify information exchange have been the focus of research for detecting information bottlenecks. Staron and Meding (2011) developed a unique software platform using lean principles by conducting a case study at a software development unit and successfully identified one bottleneck in the design phase. Although this approach pioneered using software technology in identifying project bottlenecks, construction firms have not used it due to the lack of software support. Tribelsky and Sacks (2010) introduced two metrics, DV and WIP, to detect information bottlenecks. Al Hattab and Hamzeh (2018) extended the formulations of Tribelsky and Sacks (2010) to calculate DV and WIP at the project and agent levels. Al Hattab and Hamzeh (2018) do not apply those metrics to information bottlenecks. The metrics of Tribelsky and Sacks (2010), as described earlier, require estimating the total quantity of information, which is infeasible in practice. Moreover, none of the existing works validate the performance of models that detect information bottlenecks with ground truth.

With the advancement of BIM in terms of technology (Wen et al. 2021) and its integration into the life cycle of projects spanning across design, construction, and operations (Ma et al. 2018), measurement of information quantity has become easier over time, as has BIM's use for project management and interdisciplinary collaboration. BIM (Olanrewaju et al. 2021), however, is not a project process measurement and improvement tool. Additionally, its use is limited across projects with less complexity, and it is still regarded as an innovation in its diffusion stage in the construction industry (Wen et al. 2021). Therefore, BIM falls short in providing meaningful metrics in detecting information bottlenecks across all phases of project delivery and various types of projects, especially for those that are on the lower side of the complexity spectrum. There is a need for a method that applies to all types of projects regardless of complexity and party handling the information and provides common metrics for all phases, that is, planning design and construction, across project delivery, especially in those where there is an overlap between design and construction [e.g., design build, construction manager at risk, and integrated project delivery (IPD)].

Model Development

Fig. 1 represents our overall model to detect and validate information bottlenecks. Our model uses data from file-sharing platforms (e.g., Unifier and PlanGrid), email exchange data that captures email headers, and project team meeting minutes (i.e., held biweekly among owner, designer, and contractor representatives) to detect information bottlenecks. File-sharing platforms and email exchange data capture the information publication actions. Project team meeting minutes help capture project activities because these meetings discuss all project issues needing attention and collaboration across roles. We used qualitative analysis of ethnographic data and an interview with an industry representative who worked as the primary project manager representing the owner throughout the project delivery to obtain ground truth dates for information bottlenecks. Neither the researcher nor the industry representative viewed our model and detected bottleneck dates while arriving at the ground truth dates. We compared our detected dates to the ground truth dates to validate our model.

Model for Information Bottlenecks

Following the work of Tribelsky and Sacks (2010), we use two explanatory variables—DV and WIP—to detect information bottlenecks. The novelty of our method is in formulation and data to estimate DV and WIP.

DV

DV is the rate at which information details are published and made available to the team members (Tribelsky and Sacks 2010). New information in the form of drawings, submittals, estimates, and updates gets generated as the project progresses.

Existing work calculates the rate of growth of information package size to estimate DV. Package size is calculated by counting either the number of information items or the number of attributes in all the information objects associated with a package (Tribelsky and Sacks 2010). An information object is a distinct component of a building with technical attributes, such as ceilings, pipes, beams, and so on. An information item is a single information element, such as a dimension on a drawing.

However, estimating package size is often not feasible or practical. There could be thousands of information objects, each with numerous information attributes, which makes package size expensive to track. Second, there is a delay between new information generation in the design office and their publication. Package size does not consider whether the new information has been made available to the team members.

In this study, we reformulate DV as the rate of publication actions between two successive project team meetings. A publication action is any action by a team member that transfers information from a private domain to a public domain. The variable A_t denotes the rate of publication actions at time t. Publication actions include email exchanges and updates on file-sharing platforms. Emails and document-sharing platforms are two of the key channels to publish

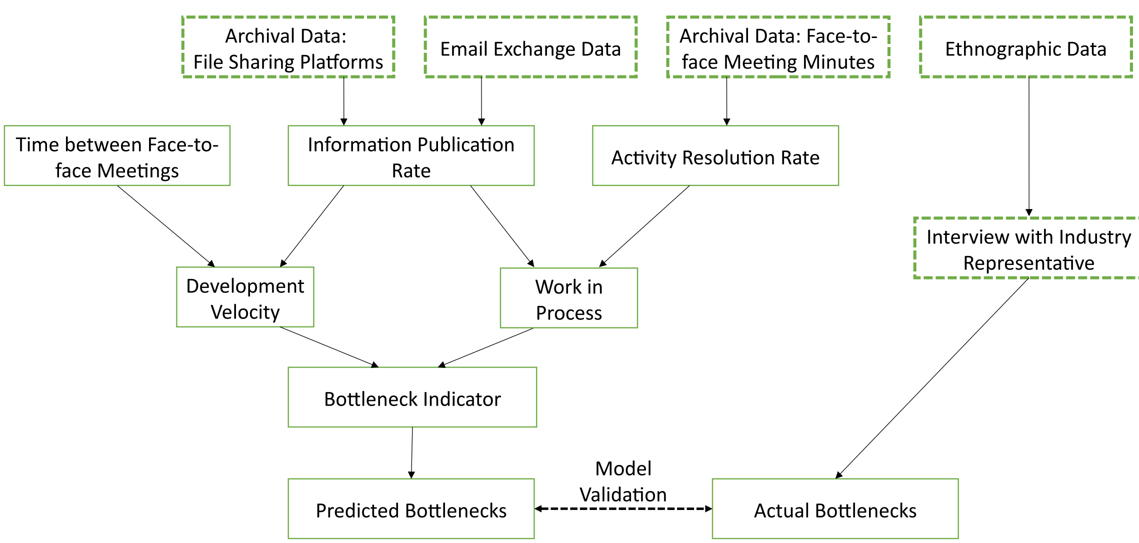

Fig. 1. Overview of research methodology. A dashed border indicates data collected, and a solid indicates a research step.

information in AEC projects (Demian and Walters 2014). Even with the advent of newer technologies for information exchange, such as Asite, SAP, PPManager, and recommendations to opt out of emails for project-related communications as a protective measure against claims (SCL Protocol 2017), email remains a popular and reliable medium to document key information exchange in construction projects (Al Hattab and Hamzeh 2018). In this study, we consider every email with an attachment (Mulugeta 2010) and all project documents uploaded to a file-sharing platform a publication action. If there were M meetings, and T_m denotes the day of the mth meeting from the beginning of the project, then the DV on day T_m is given by Eq. (1). In Eq. (1), dt denotes the differential of the time variable

$$\mathrm{DV}_m = \frac{\int_{T_{m-1}}^{T_m} A_t dt}{T_m - T_{m-1}} \qquad m = 2, 3, \ldots, M \tag{1}$$

In our formulation of DV, we ignore the quantity of information in individual publication actions. Despite this limitation, our formulation of DV is effective in explaining bottlenecks. According to Tribelsky and Sacks (2010), a high DV indicates a higher probability of bottleneck occurrence; thus, it is the variance in DV that explains bottlenecks and not the exact value of DV. The variance in DV is captured through the frequency of information publication actions because they resemble the frequency of information development. The frequency of information publication actions is more straightforward to estimate because we need to consider only the timestamp of publication action and not the quantity of information in those actions.

WIP

WIP is a measure of unused information packages that were available (Tribelsky and Sacks 2010). Tribelsky and Sacks (2010) calculate WIP at time t as the sum of the package size of all unopened information packages weighted by the time they have been in the system. Again, the difficulty in estimating the size of information packages makes estimating WIP challenging in practice. Moreover, a team member viewing or downloading a package does not necessarily imply they processed and performed the actionable items.

We define WIP in terms of project activities and publication action. We assume there are N project activities in total during the duration of the project. An activity is active on the day T_m if it is unresolved on that day. If activity n is active on day T_m, then we have $U_m^n = 1$. If activity n is active on day T_m and day T_{m-1}, then we have $V_m^n = 1$. We have

$$U_m^n = \begin{cases} 1 & \text{if activity } n \text{ is active on day } T_m \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$V_m^n = U_m^n U_{m-1}^n \tag{3}$$

The value $\sum_{n=1}^{N} U_m^n$ denotes the total number of active activities on day T_m, $\sum_{n=1}^{N} V_m^n$ denotes the total number of activities that were active on both days T_{m-1} and T_m, and $(\sum_{n=1}^{N} V_m^n)/(\sum_{n=1}^{N} U_{m-1}^n)$ denotes the fraction of activities that were unresolved in the interval (T_{m-1}, T_m) that are still unresolved on day T_m. We define WIP_m as the product of fraction of unresolved activities and the total number of publication actions in the interval (T_{m-1}, T_m), as shown in Eq. (4). In our definition, a higher fraction of unresolved activities is an indicator of a bottleneck. We can extend this definition to subteams or individual team members by considering only activities that the subteam or individual team members are responsible for. In this study, we detect information bottlenecks for the entire project team

$$\mathrm{WIP}_m = \frac{\sum_{n=1}^{N} V_m^n}{\sum_{n=1}^{N} U_{m-1}^n} \int_{T_{m-1}}^{T_m} A_t dt \qquad m = 2, 3, \ldots, M \tag{4}$$

Bottleneck Indicator

Neither WIP nor DV is independently sufficient to measure bottlenecks. WIP measures activity resolution rate. A resolved activity is useful only when the information created from its resolution is available to other team members. Similarly, DV measures the rate of information publication. Information published that is not relevant to resolving project activities is of little value in preventing bottlenecks. Therefore, both DV and WIP are used in conjunction to detect bottlenecks.

A bottleneck indicator, denoted by BN, considers both DV and WIP to indicate potential information bottlenecks. BN is a number from 0–1, with a higher value indicating a higher chance of information bottlenecks. When WIP is high, the influence of DV is weaker as compared to when WIP is low. This is because unresolved issues that pile up over time have a cascading effect because the information produced by resolving some of the earlier issues in the project affects the resolution rate of future issues. Tribelsky and Sacks (2010) mapped DV and WIP to a bottleneck indicator accounting for this imbalance in the weighting of DV and WIP. We use the same mapping function, which is given in Table 1.

The mapping function assumed that DV and WIP are normalized in the 0–1 range. In our approach, we used a max–min function for this normalization, as shown in Eq. (5). We only need data available up to time T_m to estimate WIP and DV at time T_m. BN on day T_m is denoted as BN_m

$$\mathrm{dv}_m = \frac{\mathrm{DV}_m - \min\{\mathrm{DV}_1, \ldots, \mathrm{DV}_m\}}{\max\{\mathrm{DV}_1, \ldots, \mathrm{DV}_m\} - \min\{\mathrm{DV}_1, \ldots, \mathrm{DV}_m\}} \tag{5}$$

$$\mathrm{wip}_m = \frac{\mathrm{WIP}_m - \min\{\mathrm{WIP}_1, \ldots, \mathrm{WIP}_m\}}{\max\{\mathrm{WIP}_1, \ldots, \mathrm{WIP}_m\} - \min\{\mathrm{WIP}_1, \ldots, \mathrm{WIP}_m\}} \tag{6}$$

Model Validation

To validate our model, we compared our model output to ground truths. An independent researcher in our team attended biweekly project team meetings for almost 2 years to collect ethnographic data. Dossick and Neff (2011) indicated that face-to-face conversations between team members are less about briefings and more about problem-solving. The data collected through observation

Table 1. Mapping DV and WIP to BN

WIP	DV ≤ 0.25	0.25 < DV ≤ 0.5	0.5 < DV ≤ 0.75	0.75 < DV ≤ 1
WIP ≤ 0.25	0.4	0.2	0	0
0.25 < WIP ≤ 0.5	0.6	0.4	0.2	0
0.5 < WIP ≤ 0.75	0.8	0.8	0.6	0.4
0.75 < WIP ≤ 1	1	1	1	0.8

provide flexibility and insight because the researcher can observe all interactions (Wimmer and Dominick 2013). Therefore, the ethnographic coding of the biweekly project meetings constitutes an important source to identify issues and attempts to resolve those issues and bottlenecks.

The researcher, without observing model outputs, used the ethnographic data to identify information bottlenecks. We interviewed an industry representative associated with the project to verify the dates obtained from analyzing ethnographic data. The verified dates are our ground truth dates of actual bottlenecks.

Case Study

The case study project was an institutional renovation and expansion project with a budget of over 20 million USD delivered via Construction Management-at-Risk (CMR). The project spanned 2 years starting in December 2018, going through schematic design (SD), design development (DD), construction documents (CD), and construction phases. We collected email exchanges, and archival (i.e., meeting minutes and file-sharing platform) and ethnographic data throughout the project delivery.

Email Exchange Data

Each email is an information publication action in our model. We exclude emails that are conversations because they are typically not information publication actions. To achieve this, we consider only emails with attachments in our calculations. Specifically, we count only emails that have attachments with cumulative sizes exceeding 75 kb regardless of the filetype of the attachment (Mulugeta 2010). We consider the email subject lines to filter out irrelevant emails. We have a roster containing the email IDs of team members in our case study project. The case study project team is across organizations and roles (e.g., owner, designer, and general contractor). In other words, team members mentioned here refer to individuals tasked in the case study project, working across organizations that are contracted to each other temporarily for project-related purposes. In our study, only the emails between members of the project are considered. Specifically, we use a filter such that, to include an email, the sender and at least one receiver in that email must belong to the roster of project team members.

We extracted email data from servers and personal computers of the owner, designer, and general contractor representatives. We developed a Visual Basic for Applications code in Microsoft Excel to extract project-related emails from personal systems. We extracted the following attributes for each email: (1) address and name of the sender; (2) addresses and names of each receiver, including cc and bcc; (3) subject line; (4) timestamp; (5) size and title of any attachments; and (6) network message ID. We filtered emails relevant to the project based on the sender and list of receivers, subject line, and titles of attachments. Extracting email headers from multiple project participants may result in duplicate entries. We used network message ID, timestamps, email size, and subject line to remove duplicates. Fig. 2(a) plots email rate as a function of project timeline. In Fig. 2(a), sequential, reciprocal, and intensive activities span the timeline starting from schematic design (SD phase), which is unique to integrative project delivery methods (e.g., IPD, design build (DB), and CMR such as the one used in this case) to include general contractors.

Archival Data: File-Sharing Platforms

We collected archival data from all file-sharing platforms. Our case study project used two different platforms to generate and share project documents. We imported project documents as a .csv file. We identified the author and timestamp of publication for each document from the imported files. We used software called Autovue version 21.0.2.6 to automatically calculate the modifications in the project drawings. Autovue software calculates any changes or updates made in the shared documents during the project timeline. Data shared are normalized based on the total data across the three phases of design. Fig. 2(a) plots the publication action rate through file-sharing platforms as a function of the project timeline.

Archival Data: Meeting Minutes

We observed and analyzed minutes from project team meetings spanning the schematic design, design development, construction documents, and construction phases of the AEC case study project.

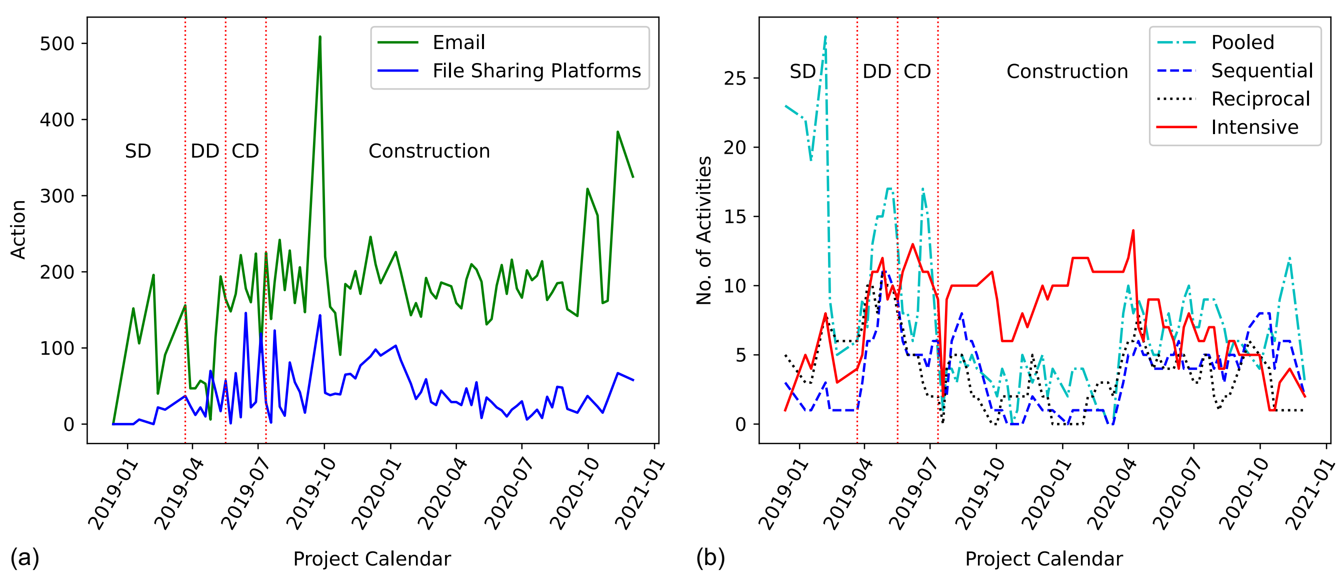

Fig. 2. Rate of publication actions and active activities as a function of project timeline.

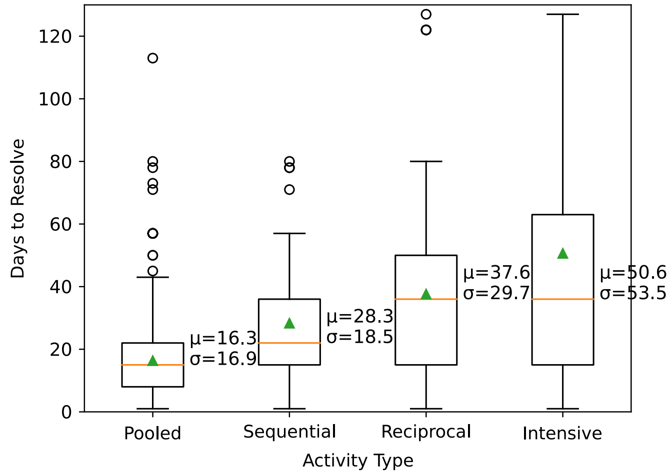

Fig. 3. Boxplot for average number of days to resolve activities for each activity type. The triangular mark indicate mean value.

We identified unresolved project activities from the analysis of meeting minutes and coded their timeline using Gantt charts. We classified each activity into pooled, sequential, reciprocal, and intensive types, analyzing their interdependencies across roles (i.e., owner, designer, contractor) from the Gantt chart (Bell and Kozlowski 2002). Fig. 2(b) shows the unresolved activities of each type as a function of the project timeline. The average number of days to resolve each activity type is shown in Fig. 3. We observe that the average number of days to resolve is the lowest for pooled and the highest for intensive activities. Pooled and sequential activities have similar resolution times, with the sequential type marginally higher than the pooled type. Reciprocal and intensive activity types have similar and higher resolution times. This is expected because interdependency is higher for intensive and reciprocal activities due to the need for coordination between multiple disciplines.

Ethnographic Data—Establishing Ground Truth

Two coders independently attended biweekly project meetings to collect observational data. Each coder prepared a document for every meeting capturing information exchanged in these meetings. Each coder produced a Microsoft Excel sheet marking every instance when a participant gave or asked for information in the meeting. The coders also indicated flags on topics and activities that demanded attention and took an unusually long time to resolve. In the same week, both coders had a meeting to merge their documents to produce a single coding file. They documented differences in their coding sheets as a reference file for future coding. We also used Spearman's rank correlation ($r = 0.89$, $p < 0.01$) to calculate the consistency of coding between the two coders, thereby ensuring the reliability of the coded file.

The researcher who also sat through all case study project team meetings made the final determination on bottlenecks using the data coded by the coders. The researchers in our team who worked on the study model and the ethnographic data were independent from each other. At this stage, 10 dates with potential bottlenecks were identified: (1) toward the end of the schematic design phase due to coordination issues between different parties; (2) toward the end of the design development phase due to unaligned expectations between the general contractor and designer; (3) at the beginning of the construction documents phase due to a higher cost estimate than what the owner expected; (4) 1 month after construction start, pending issues due to external parties such as permits; (5) 2 months after construction start, due to unexpected soil and site conditions; (6) 3 months after construction start, due to the increased need for coordination between different consultants and subcontractors; (7) 5 months after construction start, due to design complexity and application matters; (8) 6 months after construction start, due to a bidding package; (9) 9 months after construction start, due to the COVID-19 outbreak and creating new baseline designs as the owner decided to pause the project; and (10) 1 year after construction start, due to remote working modality in relation to the COVID-19 pandemic and design consultants' extended timeline in this new work environment.

After determining the dates with possible bottlenecks as shown, we conducted an interview with an industry representative closely associated with the project to verify them. This industry member was the owner's representative for the case study project. We selected this industry team member due to their uninterrupted and close engagement with the project from the beginning to the end, whereas some of the other project representatives in the core team were in and out. Fig. 4 shows the bottleneck dates identified by our researcher and the ones verified by the industry representative, revealing that our researcher was more conservative in identifying bottlenecks using ethnographic data. The owner's representative did not identify any project bottlenecks outside the ones the researcher had demarcated. Ultimately, we used the owner representative's verified dates (i.e., seven in total, excluding the first, sixth, and eighth) as the ground truth to validate our model.

Based on these findings, the research team observed that bottlenecks in this project most likely occurred due to design issues that required heavy coordination or unforeseen situations such as

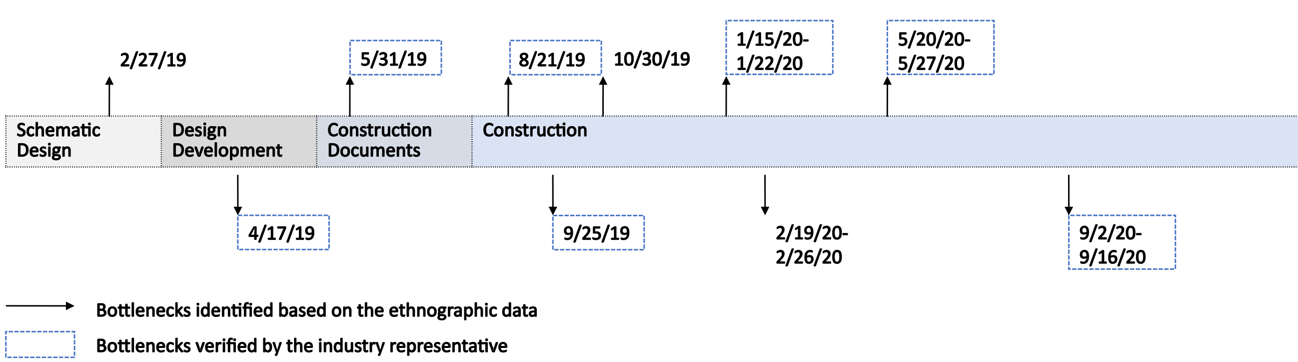

Fig. 4. List of bottleneck dates qualitatively identified and verified.

soil conditions and the COVID-19 outbreak. Moreover, the involvement of outside parties, such as obtaining state or city approval, increased the resolution time for issues, causing bottlenecks.

Results

In this section, we show the performance of our model. We compare our model to ground truth dates. We also do ablation studies. We do this by relaxing one of the variables—DV or WIP—and using only the other variable in the model to measure the relative impact of each variable on detection accuracy. The results of ablation studies are described in the section "Model Evaluation." We also evaluate the model performance of including only activities with either high or low interdependencies against all activities.

Comparing Model Outputs to Ground Truths

Our case study project ran for over 2 years. The project team meetings took place either weekly or at a biweekly frequency. There were a total of 83 meetings. The variable T_m denotes the time of the mth meeting, where $m = 1, 3, \ldots, 83$. We calculated DV and WIP for the days in which project meetings took place. We needed project activities recorded in meeting minutes to calculate WIP. Therefore, we calculated DV and WIP only on the meeting days. The values DV_m and WIP_m denote DV and WIP at time T_m, and BN_m denotes the bottleneck indicator at time T_m. The variable BN_m takes discrete values of 0, 0.2, 0.4, 0.6, 0.8, and 1, with a higher value indicating a higher probability of bottleneck occurrence (Tribelsky and Sacks 2010). We set our discriminator as 0.8; that is, when $\mathrm{BN}_m > 0.8$, our model detects a bottleneck on day T_m. We set the discriminator at a higher value to minimize the number of false negatives. A false negative detection would result in unnecessary alarm by project participants, which can be detrimental. Fig. 5 compares bottlenecks detected by our model to ground truths. Ground truths are industry-verified bottleneck dates, as shown in Fig. 4. The shaded region in Fig. 5 indicates the ground truth dates. In total, there were seven bottlenecks. Dotted vertical lines demarcate the SD, DD, CD, and construction phases of the project. There were two bottlenecks in the first half of the project consisting of the SD, DD, and CD phases. There were five bottlenecks in the second half of the project, which is the construction phase. The performance of our model was better in the second half of the project (construction phase) than in the first (design phase). For example, our model successfully detected four of the five information bottlenecks in the construction phase. The improved performance in the second half is because our model works by estimating the relative variance in the frequency of publication actions and active activities. The estimation of relative variance at time T_m uses all data available until time T_m. There are not much data available in the early phases of the project to accurately capture relative variance, which affects our model performance. The one bottleneck which our model did not detect in the construction phase was the result of the COVID outbreak. This information bottleneck resulted from the need for new baseline designs due to the pandemic. Our model does not account for abrupt unforeseen causes because it relies on the history of information exchange frequency.

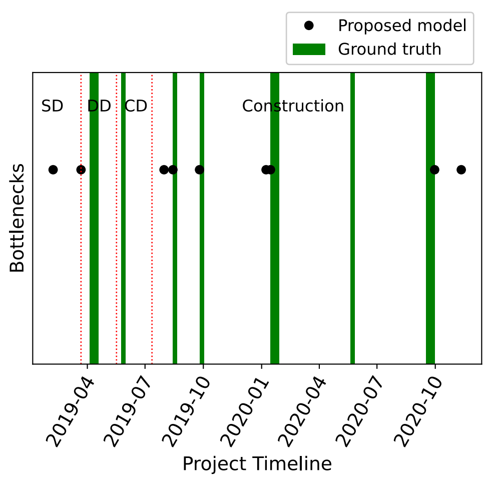

Fig. 5. Comparing bottlenecks detected using our model to industry-verified ground truth bottleneck dates. Dotted vertical lines demarcate project phases.

Model Evaluation

WIP depends on information on activities documented in the meeting minutes. Therefore, our model outputs are only on the dates when project team meetings occur. Observations from qualitative observations might not match the exact date on which these meetings took place. To account for this effect, we introduced a tolerance variable, denoted as tol. We counted a detection as successful if we detected it within tol days preceding the ground truth.

Fig. 6(a) shows precision, recall, and F-score of our model as a function of tolerance days considering the ground truth bottlenecks as positives. To understand the discriminatory capacity of our model, we used the receiver operating characteristic (ROC) curve analysis and area under curve (AUC) metrics (Adams and Hand 1999). AUC varies between 0 and 1, with a higher value indicating a model with a better discriminating capacity. An AUC of 0.5 implies the model has no discriminatory power. Generally, an AUC

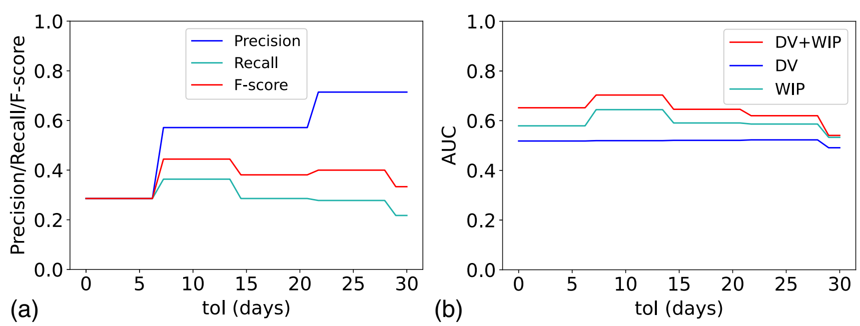

Fig. 6. Performance based on tolerance days (tol).

J. Constr. Eng. Manage., 2023, 149(12): 04023135

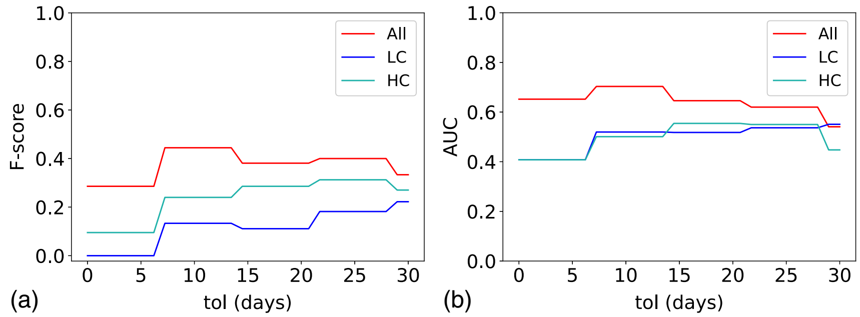

Fig. 7. Performance of our model based on the type of activities.

over 0.7 is considered good (Mandrekar 2010). Fig. 6(b) shows our model performance in terms of AUC. Within a week preceding the bottleneck, our model achieved an AUC of more than 0.7. Our model used both DV and WIP as explanatory variables. Fig. 6(b) compares our model performance considering only one of the explanatory variables. Fig. 6(b) shows that including both explanatory variables improves model performance over just one explanatory variable. Therefore, it is important to include both DV and WIP as explanatory variables in the model. We also observed that between explanatory variables, WIP had a higher effect on discriminating bottlenecks than DV. This observation is consistent with our reasoning because WIP has a higher effect on BN occurrence, as discussed in the section "Model for Information Bottlenecks."

Fig. 7 compares model performance considering all activities against considering only either the LC or HC activity type. Performance evaluated using the F-score and AUC was higher when we included all activities than when we included only a specific type. This result is encouraging in practice because it is easier to count the total number of activities across all types without needing to classify which type they belong to. Between LC and HC activity types, HC activities explained the bottlenecks better than LC, as observed from the higher F-score in Fig. 7. HC activities had higher interdependencies. The accumulation of unresolved HC activities is an indicator of information deprivation for interdependent stakeholders. Therefore, variance in frequency of unresolved HC activities explains information bottlenecks better than LC activities. This result implies that we can obtain good performance by counting only HC activities.

Limitations

The model, however, has a few limitations. Our model does not capture external factors that lead to bottlenecks because we base our model on internal information exchange. Another limitation of our model is a low detection accuracy in the early phases of the project. A lower accuracy in earlier phases is attributed to the limited availability of data. Future research such as data-driven models by comparing data on similar projects can address this limitation. A third limitation is in the logistical effort to collect email data from all contractors and designers. Our model does not use priority information or content in emails. In other words, we only use the frequency of exchange and size of emails without access to their content. However, obtaining permission to collect and analyze such data is more practical. Such collection is quite possible to achieve if and when there are agreement and contract conditions among parties to integrate this information in project management through various software that would be sensitive to the organizational concerns of all contributing parties. This can be developed into new software or incorporated into existing software such as the well-known project management tool Procore (Procore 2022). A fourth limitation is that our email exchange information is sourced from primary representatives of the owner, designer, and general contractor as the parties holding the primary contracts to deliver the project. Although the whole project network's email and other forms of communication are not reflected in this data set, the logic here is that all primary information central to the project flows through these parties and gets recorded via emails. Although this phenomenon is supported by the literature (Garcia et al. 2020), we still added a verification step in our study and verified with multiple project parties that the final communication network we developed using this data set was reflective of the overall project communications. A fifth limitation is that our model only considers information published through email and file-sharing platforms. Our model does not consider tacit knowledge that is shared via verbal conversations between the team members. However, email data are representative of team collaboration interactions and particularly appropriate in this study for three reasons: (1) email remains a primary communication tool in AEC project management; (2) extensive use of emails in project management facilitates efficient information exchange, documentation, and reliable decision making. Emails allow geographically dispersed members to collaborate remotely and share accountability; and (3) emails are a reliable data source about interactions. The project manager, construction manager, and chief architect of this AEC project in their interviews confirmed that the networks we identified from email data represented their collaboration patterns. A sixth limitation is that issues related to claims and disputes, especially the ones created forensically, are difficult to solve via emails. Our model uses a rich data set across many documents in the project in calculating bottlenecks, as shown in Fig. 1, which then is used to create publication actions and active activities across the project timeline, as displayed in Fig. 2, which in the future could be an inspiration for a software interface in bringing this model to reality for day-to-day use in project management operations. From this perspective, it is very likely that claims and disputes could be reduced because teams and project managers would have an objective and data-rich tool to visualize and detect bottlenecks and they can make decisions to intervene the project team accordingly. The model can also be used in resolving disputes and claims faster and more efficiently by reviewing the bottleneck points in project delivery and backtracking delivery activities and team correspondence accordingly. Even though email would not be an appropriate media to review for claims and disputes, the aggregate data to calculate the bottleneck model in our study, including emails, would provide a rich foundation to detect priority time periods during project delivery for things that may have gone wrong. Finally, it is important to monitor the evolution of technology and contract law alongside user behavior and industry norms to

J. Constr. Eng. Manage., 2023, 149(12): 04023135

ensure the relevance of email data in this context and consider the various applications of the fundamentals suggested in this study moving forward.

Conclusions

This work contributes to the body of knowledge by a new approach to detecting information bottlenecks in AEC projects. Unlike previous approaches, our model detects information bottlenecks without estimating the total quantity of information published. Our model uses data of active unresolved activities and frequency of information publication actions without estimating the total quantity of information. As a result, our approach is more feasible in practice than previous approaches. Comparing bottlenecks detected by our model to ground truth dates showed that our model accurately detected the occurrence of information bottlenecks. Our analysis showed that both DV and WIP are important to accurately detect information bottlenecks. Thus, both unresolved activities and frequency of information publication actions must be considered in the model. We also showed that a model that uses only data of high complexity activities achieves as much accuracy as a model that includes all activity types. Thus, even if we track only high complexity activities, they will serve as an effective measure to detect information bottlenecks.

Our model has higher accuracy in the later phases of the project because there are more data to calculate the relative frequency of information publication actions. However, for project teams using integrated project delivery for construction and BIM-enabled project delivery processes, the bottlenecks usually occur at early stages of delivery, or the DV might be slower early on in design phases due to the level of multidisciplinary coordination involved. The natural progression of IPD and BIM-enabled projects is team members going through complex decision-making processes involving many expertise areas for resolution early on in the delivery process. Our model still applies to IPD and BIM-enabled projects. Our DV considers all information publication actions, including email exchanges and coordination activities. In IPD and BIM-enabled project delivery processes, it is common to see a lower DV and higher number of bottlenecks in the earlier phases of the project delivery because this is a part of the natural progression of such projects—which is team members going through complex decision-making processes involving many expertise areas for resolution.

The study findings have implications for interorganizational and interdisciplinary project teams. First, our model detects the occurrence of information bottlenecks. The low computational expense of the model makes it straightforward to leverage our approach in industrial practice. Creating a standardized bottleneck identification model can help project managers learn from information exchanges and eliminate potential bottlenecks in the future. For example, assume that the level of information bottlenecks in a project this month is higher than the last month; the project manager then can use this information to identify issues that are taking longer than average to resolve and objectively decide if certain patterns in the project network are leading to these bottlenecks, such as the need of approvals or expertise from outside parties, prioritization for issue resolution, or additional resources to better distribute tasks to relevant personnel. Eliminating bottlenecks can smooth the production process, reduce the number of delays, and increase team efficiency while potentially helping facilitate faster dispute resolution due to the accessibility of timely information. Making such changes informed by the ongoing bottlenecks this month can help the project get back on track next month and before it is too late for the overall project performance to be adversely impacted. Second, our study evaluates the effect of activity interdependencies. Project managers can allocate resources better by taking into consideration the activity interdependencies and making sure that there are accommodations in place for high-complexity activities such as inviting necessary experts to team meetings at optimal times and adopting activity prioritization systems at the project team level as opposed to following organizational priorities and responding to project deadlines.

Future research can further refine our model to reformulate DV and WIP using activity type to improve detection accuracy. Future work can study creating separate models for information bottlenecks in various phases of the project, particularly in the early phases such as schematic design and design development. We do not propose mitigation steps in response to information bottlenecks. Building up on this work, future research should focus on developing just-in-time detection models to predict potential bottlenecks as project delivery continues. This paper is a solid foundation that uses automation techniques on information published in all project teams universally that is, meeting minutes and email exchanges, to detect information bottlenecks. With further data collection, this model can be used to predict project bottlenecks before they occur. We learned a few lessons from the nature and causes leading to information bottlenecks: (1) external team, (2) external lead to team coordination, (3) external issues leading to bottlenecks, and (4) external issues leading to team coordination issues. Insights into the bottlenecks (e.g., externally induced due to permitting issues or internally induced due to high-complexity issues not getting resolved fast) can inspire project managers to adopt mitigation strategies to reduce stress on project communication networks and streamline production, potentially resulting in improved outcomes.

Data Availability Statement

Part of the data, models, and codes that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This research was supported by the National Science Foundation (NSF) through Grants CMMI-1825678 and IIS-1928278. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the researchers and do not necessarily reflect the views of the NSF.

Notation

The following symbols are used in this paper:

A_t = rate of publication actions at time t;
BN_m = average BN on day T_m;
DV_m = DV on day T_m;
dt = differential of time variable;
dv_m = normalized DV_m;
M = total number of meetings;
m = meeting number;
N = number of project activities;
n = activity number;
T_m = time of meeting m in days from beginning of project;

tol = tolerance variable denoting days preceding ground truth bottleneck date;

U_m^n = resolution status of activity n on day T_m;

V_m^n = status indicating if activity n was active on days T_m and T_{m-1};

WIP_m = average WIP on day T_m; and

wip_m = normalized WIP_m.

References

Adams, N. M., and D. J. Hand. 1999. "Comparing classifiers when the misallocation costs are uncertain." *Pattern Recognit.* 32 (7): 1139–1147. https://doi.org/10.1016/S0031-3203(98)00154-X.

Al Hattab, M., and F. Hamzeh. 2018. "Simulating the dynamics of social agents and information flows in BIM-based design." *Autom. Constr.* 92 (Aug): 1–22. https://doi.org/10.1016/j.autcon.2018.03.024.

Ali, B., A. A. Aibinu, and V. Paton-Cole. 2023. "Unearthing the difficulties in the information management of disruption claims." *J. Leg. Aff. Dispute Resolut. Eng. Constr.* 15 (2): 04523001. https://doi.org/10.1061/JLADAH.LADR-934.

Back, W. E., and K. A. Moreau. 2000. "Cost and schedule impacts of information management on EPC process." *J. Manage. Eng.* 16 (2): 59–70. https://doi.org/10.1061/(ASCE)0742-597X(2000)16:2(59).

Ballard, G., and L. Koskela. 1998. "On the agenda of design management research." In Vol. 98 of *Proc., IGLC*, 52–69. Guarujá, Brazil: International Group for Lean Construction.

Bashir, H., U. Ojiako, A. Marshall, M. Chipulu, and A. A. Yousif. 2022. "The analysis of information flow interdependencies within projects." *Prod. Plann. Control* 33 (1): 20–36. https://doi.org/10.1080/09537287.2020.1821115.

Bayhan, H. G., Y. Ma, J. Thekinen, J. Tang, and S. Mollaoglu. 2021. "Using Jaccard similarity to identify new issues from AEC project team meeting minutes." In *Proc., Computing in Civil Engineering 2021*, 671–678. Reston, VA: ASCE.

Bell, B. S., and S. W. Kozlowski. 2002. "A typology of virtual teams: Implications for effective leadership." *Group Organ. Manage.* 27 (1): 14–49. https://doi.org/10.1177/1059601102027001003.

Bryde, D., M. Broquetas, and J. M. Volm. 2013. "The project benefits of building information modelling (BIM)." *Int. J. Project Manage.* 31 (7): 971–980. https://doi.org/10.1016/j.ijproman.2012.12.001.

Cavka, H. B. 2010. "Characterizing bottlenecks in building design coordination meetings." Ph.D. thesis, Dept. of Civil Engineering, Univ. of British Columbia.

Chinowsky, P., J. E. Taylor, and M. Di Marco. 2011. "Project network interdependency alignment: New approach to assessing project effectiveness." *J. Manage. Eng.* 27 (3): 170–178. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000048.

Demian, P., and D. Walters. 2014. "The advantages of information management through building information modelling." *Construct. Manage. Econ.* 32 (12): 1153–1165. https://doi.org/10.1080/01446193.2013.777754.

Dossick, C. S., and G. Neff. 2011. "Messy talk and clean technology: Communication, problem-solving and collaboration using building information modelling." *Eng. Project Organ. J.* 1 (2): 83–93. https://doi.org/10.1080/21573727.2011.569929.

Eckert, C., P. J. Clarkson, and M. Stacey. 2001. "Information flow in engineering companies: Problems and their causes." In Vol. 1 of *Proc., ICED*, 43–50. Glasgow, UK: Design Society.

Freire, J., and L. F. Alarcón. 2002. "Achieving lean design process: Improvement methodology." *J. Constr. Eng. Manage.* 128 (3): 248–256. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:3(248).

Garcia, A. J., M. Duva, S. Mollaoglu, D. Zhao, K. A. Frank, and J. Benitez. 2020. "Expertise flows and network structures in AEC project teams." In *Proc., Construction Research Congress 2020*. Reston, VA: ASCE.

Grønbæk, K., M. Kyng, and P. Mogensen. 1993. "CSCW challenges: Cooperative design in engineering projects." *Commun. ACM* 36 (6): 67–77. https://doi.org/10.1145/153571.163272.

Hijazi, A. A., S. Perera, A. Alashwal, and R. N. Calheiros. 2019. "Blockchain adoption in construction supply chain: A review of studies across multiple sectors." In *Proc., CIB World Building Congress*. Hong Kong: Institute of Physics Publishing.

Howell, G. A., G. Ballard, and I. Tommelein. 2011. "Construction engineering—Reinvigorating the discipline." *J. Constr. Eng. Manage.* 137 (10): 740–744. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000276.

Issa, M., J. Rankin, and A. Christian. 2006. "A framework to assess the collaborative decision making process in interactive workspaces." In *Proc., Joint Int. Conf. on Computing and Decision Making in Civil and Building Engineering*, 14. Annapolis, MD: Institute for Operations Research and the Management Sciences.

Jung, Y., and M. Joo. 2011. "Building information modelling (BIM) framework for practical implementation." *Autom. Constr.* 20 (2): 126–133. https://doi.org/10.1016/j.autcon.2010.09.010.

Ko, C.-H., and N.-F. Chung. 2014. "Lean design process." *J. Constr. Eng. Manage.* 140 (6): 04014011. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000824.

Leporis, M., and Z. Králová. 2010. "A simulation approach to production line bottleneck analysis." In *Proc., Int. Conf. Cybernetics and Informatics*, 13–22. New York: IEEE.

Liu, Z., Y. Lu, M. Shen, and L. C. Peh. 2021. "Transition from building information modeling (BIM) to integrated digital delivery (IDD) in sustainable building management: A knowledge discovery approach based review." *J. Cleaner Prod.* 291 (Apr): 125223. https://doi.org/10.1016/j.jclepro.2020.125223.

Love, P., P. Davis, K. London, and T. Jasper. 2008a. "Causal modelling of construction disputes." In *Proc., 24th Annual Conf. of ARCOM (Association of Researchers in Construction Management)*. London: Association of Researchers in Construction Management.

Love, P. E., D. J. Edwards, and Z. Irani. 2008b. "Forensic project management: An exploratory examination of the causal behavior of design-induced rework." *IEEE Trans. Eng. Manage.* 55 (2): 234–247. https://doi.org/10.1109/TEM.2008.919677.

Ma, X., F. Xiong, T. O. Olawumi, N. Dong, and A. P. Chan. 2018. "Conceptual framework and roadmap approach for integrating BIM into lifecycle project management." *J. Manage. Eng.* 34 (6): 05018011. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000647.

Mak, S. 2001. "A model of information management for construction using information technology." *Autom. Constr.* 10 (2): 257–263. https://doi.org/10.1016/S0926-5805(99)00035-7.

Mandrekar, J. N. 2010. "Receiver operating characteristic curve in diagnostic test assessment." *J. Thoracic Oncol.* 5 (9): 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.

McKinsey and Company. 2016. "Imagining construction's digital future." Accessed May 1, 2021. https://www.mckinsey.com/business-functions/operations/our-insights/imagining-constructions-digital-future.

Miyatake, Y., and R. Kangari. 1993. "Experiencing computer integrated construction." *J. Constr. Eng. Manage.* 119 (2): 307–322. https://doi.org/10.1061/(ASCE)0733-9364(1993)119:2(307).

Moreau, K. A., and W. E. Back. 2000. "Improving the design process with information management." *Autom. Constr.* 10 (1): 127–140. https://doi.org/10.1016/S0926-5805(99)00057-6.

Mulugeta, T. D. 2010. "Effects of different IMAP clients on mailservers performance." M.S. thesis, Dept. of Informatics, Univ. of Oslo.

Olanrewaju, O. I., A. F. Kineber, N. Chileshe, and D. J. Edwards. 2021. "Modelling the impact of building information modelling (BIM) implementation drivers and awareness on project lifecycle." *Sustainability* 13 (16): 8887. https://doi.org/10.3390/su13168887.

Procore. 2022. "Procore user guide." Accessed December 25, 2022. https://support.procore.com/products/online/user-guide/project-level/emails.

SCL (Society of Construction Law) Protocol. 2017. *Society of construction law delay and disruption protocol*. London: SCL.

Staron, M., and W. Meding. 2011. "Monitoring bottlenecks in agile and lean software development projects—A method and its industrial use." In *Proc., Int. Conf. on Product Focused Software Process Improvement*, 3–16. New York: Springer.

Tang, S., D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, and X. Gao. 2019. "A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends." *Autom. Constr.* 101 (May): 127–139. https://doi.org/10.1016/j.autcon.2019.01.020.

Tory, M., S. Staub-French, B. A. Po, and F. Wu. 2008. "Physical and digital artifact-mediated coordination in building design." *Comput. Supported Cooperative Work* 17 (4): 311–351. https://doi.org/10.1007/s10606-008-9077-4.

Tribelsky, E., and R. Sacks. 2006. "Measures of information flow for lean design in civil engineering." In *Proc., CME 25 Conf. Construction Management and Economics*, 1493. London: Taylor & Francis.

Tribelsky, E., and R. Sacks. 2010. "Measuring information flow in the detailed design of construction projects." *Res. Eng. Des.* 21 (3): 189–206. https://doi.org/10.1007/s00163-009-0084-3.

Tribelsky, E., and R. Sacks. 2011. "An empirical study of information flows in multidisciplinary civil engineering design teams using lean measures." *Archit. Eng. Des. Manage.* 7 (2): 85–101. https://doi.org/10.1080/17452007.2011.582332.

Wen, Q.-J., Z.-J. Ren, H. Lu, and J.-F. Wu. 2021. "The progress and trend of BIM research: A bibliometrics-based visualization analysis." *Autom. Constr.* 124 (Apr): 103558. https://doi.org/10.1016/j.autcon.2021.103558.

Wimmer, R. D., and J. R. Dominick. 2013. *Mass media research*. Boston, MA: Cengage Learning.

Wong, J. K. W., J. Ge, and S. X. He. 2018. "Digitisation in facilities management: A literature review and future research directions." *Autom. Constr.* 92 (Aug): 312–326. https://doi.org/10.1016/j.autcon.2018.04.006.

Xu, S., and H. Luo. 2014. "The information-related time loss on construction sites: A case study on two sites." *Int. J. Adv. Rob. Syst.* 11 (8): 128. https://doi.org/10.5772/58444.