CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

POTENTIAL OF BAYESIAN NETWORKS FOR FORECASTING THE RIPPLE EFFECT OF PROGRESS EVENTS

Kareem Mostafa.[1,2] and Tarek Hegazy [1]
[1] University of Waterloo, Canada
[2] ktmostafa@uwaterloo.ca

Abstract: As an important component of project control, project managers frequently update the progress of the various tasks in an effort to forecast project completion time. Current methods for progress analysis and forecasting, however, rely on well-known performance indices such as the Schedule Performance Index (SPI) and the Estimate at Completion (EAC) of the Earned-Value (EV) analysis. These indices, however, were developed with the assumption that the remaining part of the project will follow the latest progress trend, without regard for how current events, changes, and interruptions may shape future ones (i.e., the ripple effect of progress events). While simulation and uncertainty analysis could improve forecasting, there is still a need for more accurate forecasting methods. This paper investigates the benefits of developing Bayesian networks to predict project completion time based on a certain event(s) affecting the ongoing and/or the upcoming project tasks. To facilitate accurate forecasting, a register of possible events that trigger changes in future forecasts is identified, and includes events such as productivity loss in similar tasks, payment delays, potential site congestion, etc. A hypothetical seven-activity project is then used to demonstrate the steps in Bayesian network modeling and to highlight the difference between schedule updates with and without Bayesian relationships among task durations. The paper then discusses the potential advantages as well as the challenges of developing Bayesian models. The proposed model assists in decision-making regarding proactive remedial actions on construction projects.

1. INTRODUCTION

Project managers need to constantly forecast future project progress and completion time based on current project data. However, construction progress is of highly uncertain and dynamic nature due to the multiple stakeholders and processes involved, each with its own uncertainties. This makes the traditional forecasting approaches inaccurate and, therefore, unreliable. The Project Controls report developed by the Construction Industry Institute (CII) views forecasting as an area of weakness for both owners and contractors in construction projects (CII 2011). Hence, there is a crucial need for better forecasting tools and techniques as large resource variances can affect the viability of the project and can even jeopardize its completion. A good forecasting system should be simple in nature as well as in its data requirements, and generates forecasts that are accurate, timely, unbiased, and stable.

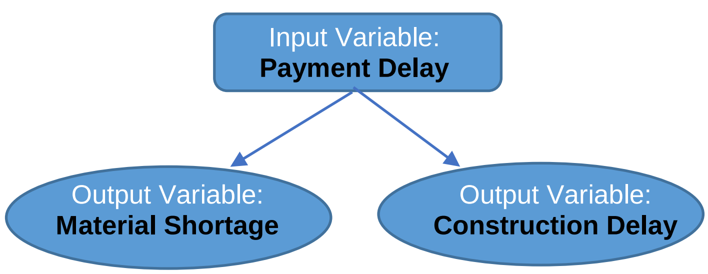
Most project updates rely on an assumption about future productivity, with the most common assumptions being either the originally planned or the real-time measured productivity. This does not account for any means of interaction between the activities and external factors or among the activities themselves (ripple effect). Changes and disruptions can affect both changed and unchanged work since the workflow and planned progress have been interrupted (Jones 2001). Appeal boards and courts have gradually begun to accept the premise of cumulative impact, if causal links can be demonstrated (Ibbs 2005).

Many professional entities in the industry tried to quantify this ripple effect of changes. For example, The Mechanical Contractors Association of American (MCAA) published a guide that discretizes 16 factors (e.g., morale, stacking of trades, and crew size inefficiency) and their projected impact on labor productivity (MCAA 2011). For instance, in this guide, minor site access affects productivity by 5%, while severe site issues can have a 20% effect, without explaining how these values were determined. These percentages are added to the expected man hours allocated to the related change order, in an additive manner in case more than one factor is involved. Similarly, The National Electrical Contractors Association (NECA) has a manual with 25 variables and five degrees of severity (NECA 1976). Both the MCAA and NECA manuals were developed by contractor groups. From the owner side, the Army Corps of Engineers also used to have their own evaluation guide (US Army Corps of Engineers 1979) until it was rescinded in 1996 (Geneni 1996). With no insight regarding how these manuals were created, a question of objectivity arises and whether these manuals were developed to maximize the gains of one party.

As an effort to address the cumulative or ripple effect on productivity, Bayesian networks are introduced in the paper as a probability-based approach that can be applied at the activity level to update predictions based on additional observations related to other progress events. Using Bayesian networks (BNs) on the activity level to provide a more flexible, reliable and quantitative forecasts that help in performing accurate project updates in view of the current project state, and taking into account how current events could shape future ones. Next section provides an introduction to the theory behind Bayesian networks and its relevant practical applications. This is followed by a description of the proposed model and its application on a hypothetical schedule at different progress stages, to highlight its versatility and strengths. Afterwards, the pros and cons of the proposed model are highlighted along with future enhancements.

2. BAYESIAN NETWORKS

A Bayesian network consists of two parts (Pearl 2014): (1) a directed acyclic graph (DAG) which represents a visual component of the relationships among a group of dynamically changing input variables, and a set of impacted outputs; and (2) a set of conditional probability tables (CPT) of the values of the observed and/or measured conditional probabilities about the occurrence of events that cause the impact to happen (e.g., Figure 1). For example, in Figure 1, the probability of material shortage to take place is 75% if there was payment delay, and 15% if payment delays did not occur. Bayesian networks can be constructed manually through expert opinions (e.g., Fan and Yu 2014; Nadkarni and Shenoy 2004), generated automatically through machine learning and observed data (e.g. Lee et al. 2009); or a combination of both (e.g., Flores et al. 2011; Procaccino et al. 2005).



Directed Acyclic Graph (DAG)

Conditional Probability Table for Material Shortage

	Material Shortage	
Payment Delay	Yes	No
Yes	0.75	0.15
No	0.25	0.85

Figure 1: Sample Bayesian Network Diagram

As shown in Figure 1, the DAG is composed of nodes that represents the different variables; and the uni-directional arcs represent the causal relationships between parent nodes and child nodes. The probability of occurrence of a child node is dependent on whether the event represented by the parent node occurred or not. The conditional probabilities in the CPT, thus, follow Bayes' theorem (Charles River Analytics 2008), which can be easily expressed using equation (1):

[1] $$P(b|a) = \frac{P(a|b) * P(b)}{P(a)}$$

Where, *P(a)* is the probability of *a*, and *P(a|b)* is the probability of *a* given that *b* has occurred. As opposed to static networks, dynamic Bayesian networks (DBNs) incorporate temporal analysis, and utilize equation (1) but with events *(a)* and *(b)* referring to an event occurring in the past and the future, respectively.

Bayesian networks' feature of combining analytical and visual components, as well as their simple underlying theory give them the following advantages (Hu et. al. 2013; Wang and Wang 2016):

(1) Visually modelling cause-effect relationships to help identify risk sources;
(2) Ability to calculate the independent impact, or lack thereof, of a subset of the variables;
(3) Applicability to model "what-if" scenarios; and
(4) Ability to be continuously updated as new information and observations are collected.

As such, Bayesian networks can be used to calculate the probability of unknown events using variables that are now known. This makes Bayesian networks applicable for both prediction and diagnosis, as well as sensitivity analysis. These advantages have made Bayesian networks under the spotlight in a variety of research fields. Applications include industrial engineering and plant modelling (e.g. Khakzad and Van Gelder 2018; Zhu et al. 2018); logistics (Guo 2015); traffic safety (Sun et al. 2015); software risk management (Fan and Yu 2004); maintenance (Bortolini and Forcada 2017); medicine (Constantinou et al. 2016); banking (Tavana et al. 2018); and others.

Within the construction industry, Bayesian networks have been utilized for various site safety applications. Xu et al. (2015) used Bayesian networks to build a construction safety pre-warning system. Gerassis et al. (2016) used it to analyze the causes of accidents related to construction of embankments while Martin et al. (2009) analyzed workplace accidents related to falls from heights. Zhang et al. (2014) and Wu et al. (2015) developed decision support tools for tunnel construction safety analysis using fuzzy Bayesian networks and dynamic Bayesian networks, respectively. Other areas of prominent Bayesian networks applications are pipeline risk assessment (Zhang et al. 2012); infrastructure integrity (Straub 2009); project delivery method selection (Bypaneni et al. 2018); project cost benefit analysis (Yet et al. 2016); cost risk analysis (Khodakarami and Abdi 2014); and estimating overall project delay (Luu et al. 2008).

3. PROPOSED MODEL FOR ACTIVITY PROGRESS UPDATES

A well-designed network shall properly represent the true system and may result in a smaller number of required parameters as well as less processing time. For the proposed progress updates model, the first step was to identify the significant factors that, when observed, may suggest the potential delay to the construction activity being analyzed. Since the current model is in its early stage of development, detailed steps of constructing Bayesian network for progress updates were applied to a sample activity within a hypothetical construction project (Figure 2). The same analysis can be extended to the whole project network to develop a complete progress update model. The small project of seven activities in Figure 2 serves to demonstrate how Bayesian networks can be applied in real life.

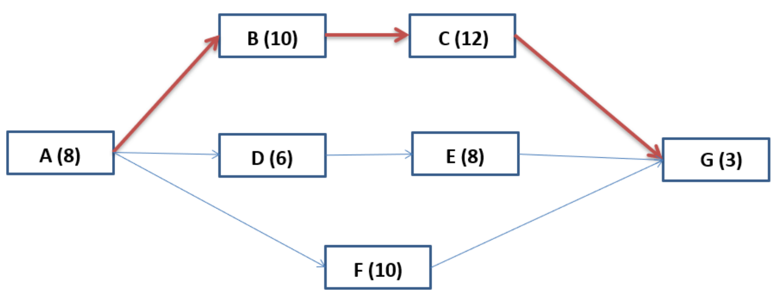
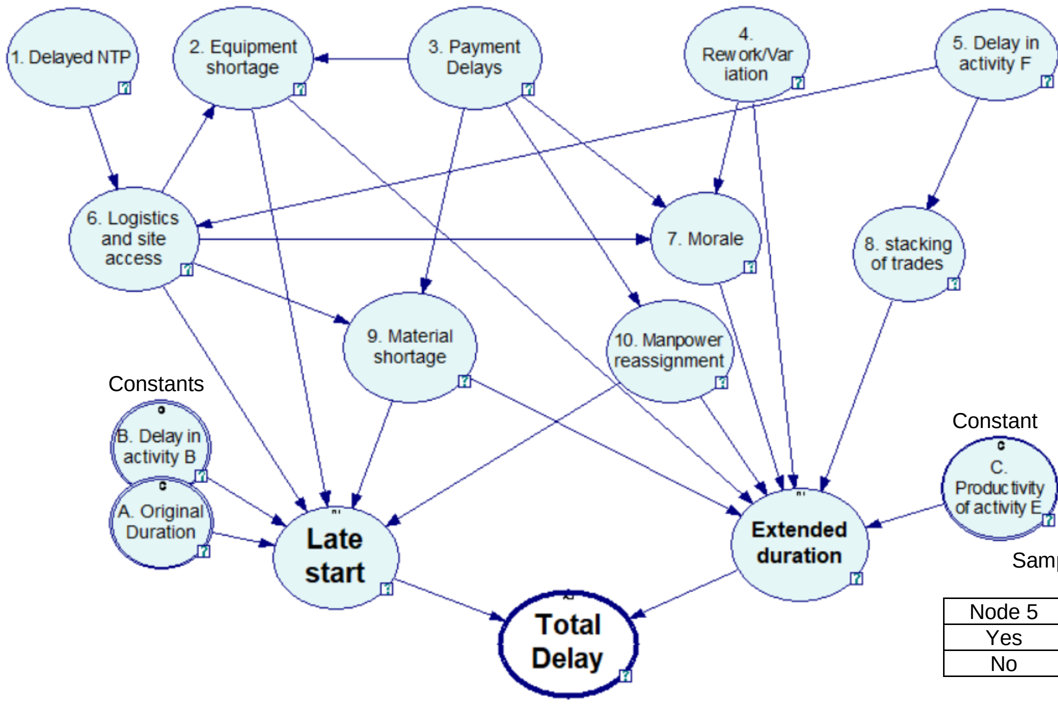


Figure 2: CPM Activity Network for the Sample Project

The first step was to define the factors that affect progress updates. For this simple project, Activity C is the one being analyzed, where Activity B is the immediate predecessor; activity F is being performed within the same working space; and activity E is a similar activity that is taking place at a different location in the project. Having this information about activity C, a visual DAG network was constructed from the factors that affect its progress updates (identified based on brainstorming and expert opinion), as illustrated in Figure 3. This was developed using a commercial software, GeNIe Ver.2.2 (http://genie.sis.pitt.edu/), for modeling Bayesian belief networks. The figure shows the cause and effect relationships represented by the uni-directional arcs among ten factors that affect two primary causes of activity delay: late start; and extended duration. Three constants (A, B, and C) are also shown in the figure: A (original duration); B (Delay in Activity B); and C (Productivity of Activity E). Also sample CPT for node 8 is shown in the figure.



Sample CPT

	Node 8	
Node 5	Yes	No
Yes	0.5	0
No	0.5	1

Figure 3: Proposed Bayesian Network for Activity C

Full conditional probabilities were defined for each of the nodes, relating them to the probabilities of their parent nodes, if any. While ensuring that the sum of all probabilities for each node equals to 1.0, the conditional probability table can be built using statistical data or expert advice, thus creating what is known as the prior distributions of the nodes.

For simplicity, the states for each variable limited to two (yes/no), and conditional probability values being 25%, 50%, 67%, and 75% only. The probabilities for extended duration and late start of the activity (both expressed as a percentage of the activity's original duration) are assumed to follow a gamma distribution. Gamma distributions are commonly used in Bayesian analysis due to its analytical convenience in terms of creating the posterior distributions for the outputs. For each of those nodes, the distribution's shape factor is the sum of the expected values of the parent nodes' states (Yes=1, No=0) and while the scale factor was assumed to be constant (0.1). This means that the higher the occurrences (or expectations) of delay events, the more likely it is for the expected activity delay to assume a greater value, all while fixing the scale parameter to ensure the obtained values are within a reasonable domain. Parent nodes that do not contribute to the gamma distributions of their respective child nodes are the three constants (A, B, and C). Node A represents the original duration of the activity, and is implemented in the network as a constant so late start can be represented as a percentage of the original activity duration in order to be able to add this effect to the extended duration effect, and node B which represents the delay in the predecessor activity directly leads to a late start of the activity under investigation. Node C represents the productivity of activity E, a similar activity taking place at a different location, it is assumed that if actual productivity of activity E was different than planned, then the expected duration of activity C would be affected by the same amount. Finally, total delay is defined as the sum of the expected duration extension and the expected late start that the activity might experience.

Figure 4 shows all the DAG nodes along with their probability of occurrence. For example, stacking of trades (node 8) is affected by only one factor (i.e. has only one parent node: Delay in activity F, node 5). Since node 5 was assigned a probability of occurrence ($P(5) = 0.5$) and the probability of occurrence assigned to node 8 is 50% if node 8 happened ($P(8)|P(5) = 0.5$) and zero if node 5 did not happen. Thus, the expected value at node 8 is (0.5 x 0.5 + 0.5 x 0 = 0.25, as shown on node 8 in Figure 4), indicating that the stacking of trade has a 25% chance of occurrence.

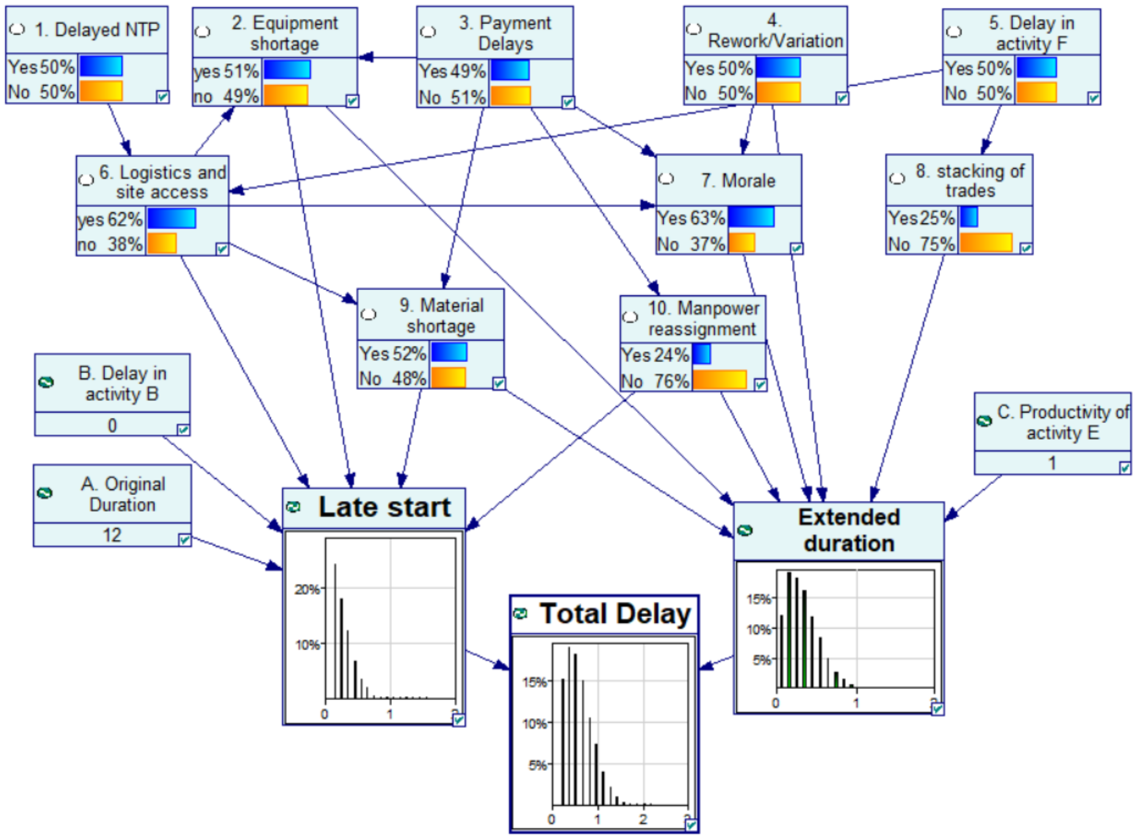


Figure 4: Initial Bayesian Network (priori) for Activity C. Nodes in bold font enter the model as constants

3.1. Case 1: Planning Phase

The model as shown in figure 4, can be used in the planning phase where no observations are collected to update the model at this stage. In this scenario, extended duration shall follow a gamma distribution of shape parameters 2.5 (the sum of the expected values of its parent nodes: nodes 7, 9, 11, 12, 13, and 15) and 0.1, plus a constant value of the expected time extension due to the reduced observed productivity of activity E following equation 2. Similarly, late start (node 4) will follow a gamma distribution of parameters 1.93 (sum of nodes 9 through 12) and 0.1, plus the delay in the predecessor activity (B) expressed as a percentage of the original duration of activity C under investigation.

[2] *Extended Duration due to reduced productivity* = $\dfrac{1}{1-\dfrac{actual\ productvity}{planned\ productivity}} - 1$

Ultimately, adding the late start and extended duration nodes yields the total expected delay for activity C. The probability density and the cumulative distribution functions (PDF and CDF, respectively) are featured in Figure 5. From the CDF function, there is a 95% confidence level that the total delay in activity C shall not exceed 120% of the activity duration.

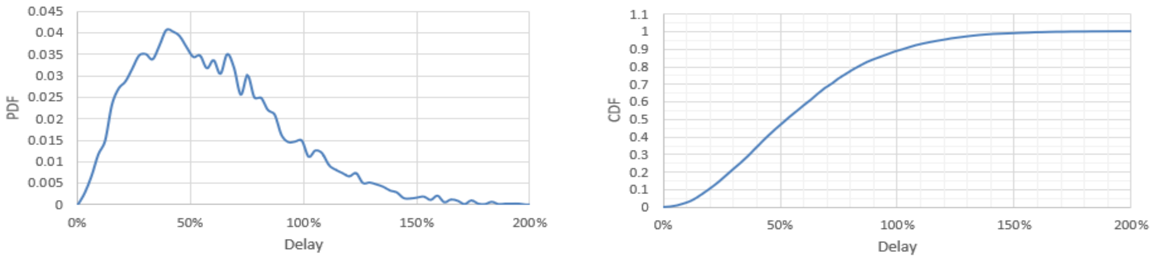


Figure 5: Probability Density function (PDF) and Cumulative Distribution Function (CDF) for the expected delay in activity C based on the baseline network

3.2. Case 2: Project Updates

During construction, progress updates and, accordingly, it automatically calculates the delay probabilities as well as the probabilities of other events that were yet to be observed. As more real progress updates are collected, the Bayesian network updates the probability distributions, creating new distributions (posterior distributions) that are better representations of the real situation and can be used in future delay calculations and predictions.

In this scenario, it is assumed that there was a delay in the Notice to Proceed (NTP); activity F experienced a late start; and actual productivity of activity E was 75% of the planned. No payment delays took place and there is no evidence to confirm or deny the presence of rework/variations relevant to that activity (Activity C). The model allows for easy updates simply by clicking on the node to be updated and specifying the value. The updated network is shown in Figure 6 (the manually updated nodes have a darker color). Looking at nodes 5 and 8 again, node 5 now has a value of 1 since its occurrence has been confirmed. Hence, the probability for node 8 occurrence has now changed to be 50% (0.5*1+0*0).

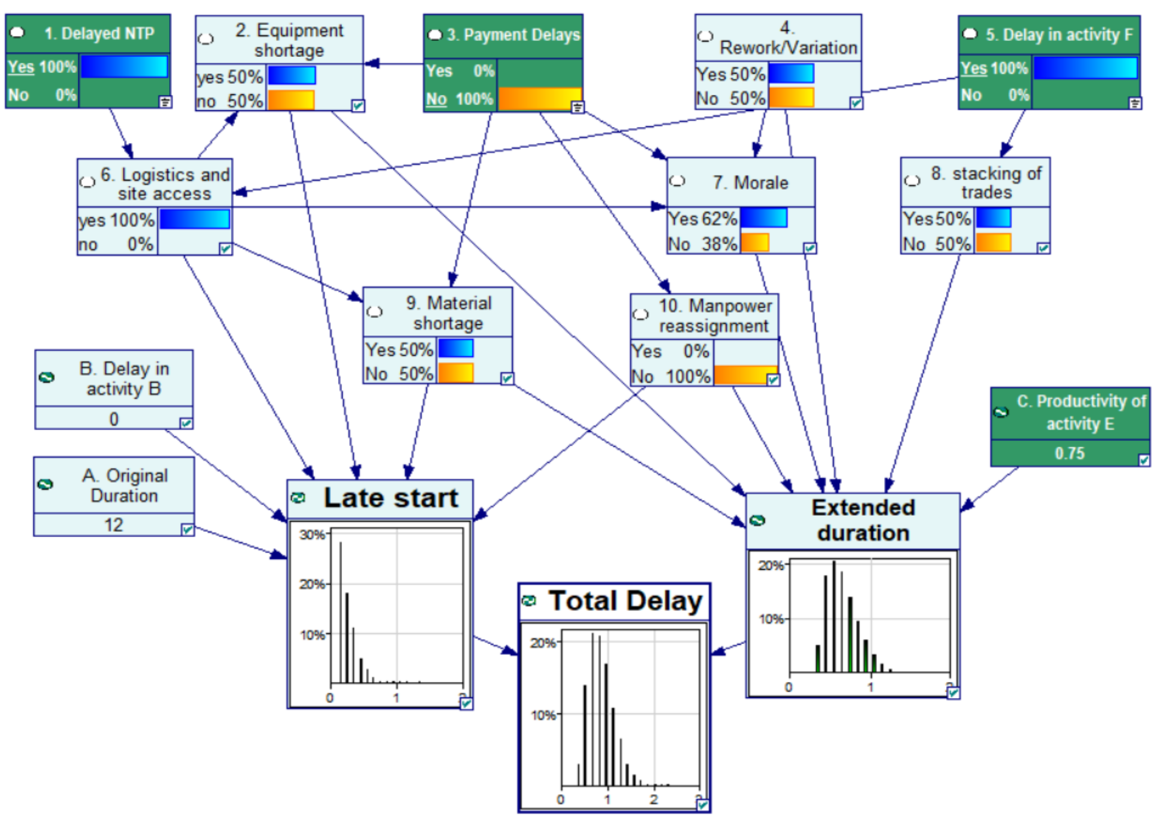


Figure 6: Updated Bayesian Network for Activity C after Progress Updates

The CDF for the total delay for this case is shown in Figure 7. Since the time extension due to the reduced productivity in activity E is added as a constant to the model, the graph indicates that the minimum amount of delay expected is the one caused by said effect as shown by the flat line at the beginning of the graph. An interesting observation is after confirming the occurrence of 3 delay events (delayed NTP, reduced productivity in activity E, and delay in activity F) the expected total delay in activity C with 95% confidence level is only 140% or less. This shows the significance of the payment delays and the fact that simply confirming the non-occurrence of that event greatly counteracted the adverse effects of the other events in the network. Similarly, stakeholders can investigate the effects of one event at a time to determine the best course of action. Another example is the case when the delay in activity F was the result of a change order, the contractor could use the same procedure to investigate how such change order might affect other portions of the project.

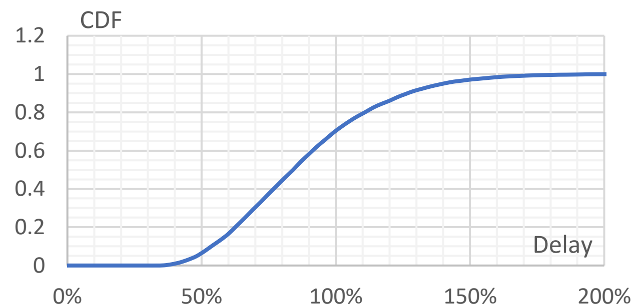


Figure 7: Cumulative Distribution Function (CDF) for the Expected Delay in activity C based on the Updated Network

4. DISCUSSION AND ONGOING WORK

Ultimately, activity-based Bayesian networks similar to the one presented in this paper could be aggregated to present work packages that make up the entire project. Original project duration is determined by a suitable deterministic method (e.g., CPM, LOB), and information regarding delays in various project work packages are updated as the project progresses and more observations enter the Bayesian network. Thus, providing more certainty about the potential overall project delay.

This study presented, in a conceptual form, a project planning and updates procedure utilizing Bayesian networks. The procedure was applied at the activity level to a hypothetical schedule to demonstrate the flexibility of the model and the various project stages in which applying it may prove useful. Bayesian networks are mathematically sound yet simple enough and include a visual component which facilitates interaction with decision makers and experts. Moreover, Bayesian networks allow to build a knowledge database that grows stronger and more reliable as new data are collected (Martin et al. 2006). This proves to be helpful at the beginning stages of the project when less project-specific data is available. The approach proves to be more practical than PERT due to its dynamic nature, allowing for duration expectations and critical path activities to be updated as the project moves forward as well as targeting specific risks and potential delay events with corrective actions to ensure the project stays on track.

The utilized software has many features that were not implemented in building the model featured in this study. For example, sub-models can be nested within the main model to properly calculate the probability states of a parent node instead of it having it as a set value and to ease the network presentation and communication. Other node types that can be used are decision nodes, which can represent potential corrective actions.

One main limitation to the method is its high reliance on records to build the conditional probability tables, which can prove to be a laborious process especially in more complex projects. While the network allows for integrating expert opinions into the model to make up for the missing data, performing expert surveys is not that easy of a task and it introduces subjectivity to the model. Bayesian networks also tend to discretize data to function properly, which might lead to data loss or increased model complexity.

5. CONCLUSION

This paper discussed the utilization of Bayes theorem in project updates and presented a conceptual Bayesian network to predict activity durations based on a certain event(s) affecting the activity. The model was applied in two different scenarios: during the planning phase to assign schedule contingencies, and during the project progress to consider remedial actions. The model could be expanded to address work packages or the entire project. Bayesian networks allow for dynamic project updates and its graphical nature allows for easy representation of its findings. In addition, Bayesian networks have the ability to "learn" as more observations are added into the model, thus constructing a powerful knowledge base that could prove helpful for future projects where project-specific data are not yet available.

6. REFERENCES

Bortolini, R., and Forcada, N. 2017. Discussion About the Use of Bayesian Networks Models for Making Predictive Maintenance Decisions. *Lean and Computing in Construction Congress,* IGLC, Heraklion, Greece, **1**: 975-982.

Bypaneni, S., Tran, D., and Nguyen L. D. 2018. A Bayesian Network Approach to Selecting Project Delivery Methods. *Construction Research Congress*, ASCE, New Orleans, LA, USA, **2**: 418-427

Charles River Analytics (2004). *About Bayesian Belief Networks.* Cambridge, MA.

Constantinou, A. C., Fenton, N., Marsh, W., and Radlinski, L. 2016. From Complex Questionnaire and Interviewing Data to Intelligent Bayesian Network Models for Medical Decision Support. Artificial Intelligence in Medicine, **67:** 75–93.

Construction Industry Institute 2011. Global Project Controls and Management Systems. Research Summary 244-1. Austin, TX.

Geneni, Jr., A. J. 1996. *Circular No. 25-1-244*, Dept. of the Army, U.S.Army, Corps of Engineers.

Fan, C.-F., and Yu, Y.-C. 2004. BBN-Based Software Project Risk Management. Journal of Systems and Software. **73**(2): 193–203.

Flores, M. J., Nicholson, A. E., Brunskill, A., Korb, K. B., and Mascaro, S. 2011. Incorporating expert knowledge when learning Bayesian network structure: A medical case study. Artificial Intelligence in Medicine, **53**(3): 181–204.

Gerassis, S., Martín, J. E., García, J. T., Saavedra, A., and Taboada, J. 2017. Bayesian Decision Tool for the Analysis of Occupational Accidents in the Construction of Embankments. Journal of Construction Engineering and Management,**143**(2): 04016093.

Guo, Q., Huang, Y., and Zhao, H. 2015. Reliability Analysis of a Cold Chain Logistics System Based on the Fault Bayesian Network. *Int. Conf. on Transportation Engineering*, ASCE, Dalian, China, **1**: 778-787

Hu, Y., Zhang, X., Ngai, E., Cai, R., and Liu, M. 2013. Software project risk analysis using Bayesian networks with causality constraints. Decision Support Systems, **56:** 439–449.

Ibbs, C. W. 2005. Impact of change's timing on labor productivity. Journal of Construction Engineering and Management, **131**(11): 219–1223.

Jones, R. M. 2001. Lost productivity: Claims for the cumulative impact of multiple change orders. Public Contract Law Journal, **31**(1), 1–46.

Khakzad, N., and Van Gelder, P. 2018. Vulnerability of industrial plants to flood-induced natechs: A Bayesian network approach. Reliability Engineering and System safety, **169**: 403–411.

Khodakarami, V., and Abdi, A. 2014. Project cost risk analysis: A Bayesian networks approach for modeling dependencies between cost items. International Journal of Project Management. **32**(7), 1233–1245.

Lee, E., Park, Y., and Shin, J. G. 2009. Large engineering project risk management using a Bayesian belief network. Expert Systems with Applications. **36**(3), 5880–5887

Luu, V. T., Kim, S.-Y., Tuan, N. V., and Ogunlana, S. O. 2009. Quantifying schedule risk in construction projects using Bayesian belief networks. *International Journal of Project Managment.* 27(1), 39–50.

Martín, J., Rivas, T., Matías, J., Taboada, J., and Argüelles, A. 2009. A Bayesian network analysis of workplace accidents caused by falls from a height. Safety Science, **47**(2): 206–214.

Mechanical Contractors Association of America (MCAA). 2011. *Factors Affecting Labor Productivity*, Rockville, Md.

Nadkarni, S., and Shenoy, P. P. 2004. A causal mapping approach to constructing Bayesian networks. Decision Support Systems, **38**(2): 259–281

National Electrical Contractors Association (NECA). 1976. *Manual of labor units*, Bethesda, Md.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann

Procaccino, J. D., Verner, J. M., Darter, M. E., and Amadio, W. J. 2005. Toward predicting software development success from the perspective of practitioners: an exploratory Bayesian model. Journal of Information Technology, **20**(3): 187–200.

Straub, D. 2009. Stochastic Modeling of Deterioration Processes through Dynamic Bayesian Networks. Journal of Engineering Mechanics, **135**(10): 1089–1099.

Sun, B., Dong, D., and Liu, S. 2015. Bayesian Belief Net Model-Based Traffic Safety Analysis on the Freeway Environment. *Int. Conf. on Transportation Engineering*, ASCE, Dalian, China, **1**: 2754-2760

Tavana, M., Abtahi, A.-R., Caprio, D. D., and Poortarigh, M. (2018). An Artificial Neural Network and Bayesian Network model for Liquidity Risk Assessment in Banking. Neurocomputing, **275**: 2525–2554.

US Army Corps of Engineers 1979. Modification impact evaluation guide, EP 415-1_3,

Wang, T., and Wang, X. 2016. A Bayesian Network-Based Risk Assessment Framework for the Impact of Climate Change on Infrastructure. *Construction Research Congress*, ASCE, San Juan, Puerto Rico, **1**: 1353-1361

Wu, X., Liu, H., Zhang, L., Skibniewski, M. J., Deng, Q., and Teng, J. 2015. A Dynamic Bayesian Network Based Approach to Safety Decision Support in Tunnel Construction. Reliability Engineering and System safety **134**: 157–168.

Xu, B., Xu, C., and Lin, X. 2015. Methods Based on the Bayesian Network for a Construction Accident Pre-Warning System. *Int. Conf. on Construction and Real Estate Management,* ASCE, Lulea, Sweden **1**: 741-748

Yet, B., Constantinou, A., Fenton, N., Neil, M., Luedeling, E., and Shepherd, K. (2016). A Bayesian Network Framework for Project Cost, Benefit and Risk Analysis with an Agricultural Development Case Study. Expert Systems with Applications. **60**:141–155.

Zhang, L., Wu, X., Skibniewski, M. J., Zhong, J., and Lu, Y. 2014. Bayesian-Network-Based Safety Risk Analysis in Construction Projects. Reliability Engineering and System safety **131**: 29–39.

Zhang, P., Long, X., Duan, Y., Yu, J., and Peng, H. 2012. Analysis of Third-Party Damage Weights in Pipeline Risk Assessment System Based on Bayesian Networks. *International Conference on Pipelines and Trenchless Technology,* ASCE, Wuhan, China, **1**: 1400-1408.

Zhu, J., Ge, Z., Song, Z., Zhou, L., and Chen, G. 2018. Large-scale plant-wide process modeling and hierarchical monitoring: A distributed Bayesian network approach. Journal of Process Control, **65**: 91–106.