# GRAPHICAL ANALYSIS ON NON-CONFORMANCES OF CONSTRUCTION PRODUCTION PROCESSES: ONGOING RESEARCH AND CASE STUDY

Jianfeng Wu[1] **and Lucio Soibelman[2]**

## ABSTRACTS

The Last Planner™ system of production control, based on lean construction principles, has been broadly and successfully implemented in several projects over the past years. Its focus on work plan realization is useful in helping project management teams identify main problems that constrain the timely completion of individual activities, and decrease possible variability. However, potential problems in designs of production processes, which also contribute to risks and variability during their implementations, have been seldom studied and could be only learned and used by human planners in a subjective and implicit manner.

In this paper, a research approach is detailed to address this problem by creating a generic and concise data representation for networks of construction production processes in support of graphical analysis and pattern recognition. As a part of this ongoing research, a case study is presented with preliminary results, which were obtained by applying the research approach on a Last Planner™ database of production control from a large capital facility project. Networks of production processes were analyzed by comparing type descriptions of the original plans and their actual performance. Interesting and statistically valid patterns were recognized in this study, such as correlations between the topology of a work plan and its probability of having non-completions during implementation. Such objective and explicit patterns could help project managers better understand potential problems in original designs of construction processes, and make informed decisions to decrease corresponding variability and increase reliability in planning and control.

## KEY WORDS

Graphical analysis, production processes, pattern recognition, knowledge discovery

[1] PhD Student and Research Assistant, Civil and Env. Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, Phone +1 412/268-6126, jianfen1@andrew.cmu.edu

[2] Associate Professor, Civil and Env. Engineering., Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, Phone +1 412/268-2952, FAX +1 412/268-7813, lucio@andrew.cmu.edu

## INTRODUCTION

Reliability of work flows in production processes is critical for successful or unsuccessful completion of construction projects (Ballard, 2000). Even though perfectly reliable work flows are unlikely to occur due to inherent variability of construction production, proper control measures should be taken to monitor, record, and analyze such processes, so that possible causes for non-completions could be investigated to mitigate negative impacts of variability in future implementations (Oliveira, Soibelman and Choo, 2004).

The Last Planner™ system of production control (Ballard, 2000), developed on lean construction principles, has been broadly and successfully applied in several projects over the past years. Instead of focusing on discrepancies on cost and schedule objectives as in general project management, the system is concentrated on assignment generation, work plan realization, and identification of problems that have constrained completions of previous activities as planned. After the sources for variability are located, corrective actions can be taken accordingly to diminish the adverse influences of known risks. Such preventive actions thus reduce potential variability and improve reliability of work flows in construction production processes (Choo et al, 1999).

However, current solutions are mainly focused on the discovery of most common problems for non-completions in individual activities, while other causes for variability of work plans remain undisclosed to project management teams who want to further enhance work flow reliability. For example, how does a particular failure of an activity influence its downstream work assignments, i.e., are there any meaningful patterns among the sequences of non-completion events? Or in general, how does the network structure of a work plan influence its overall reliability? Are non-completions of activities correlated with their positions within the work plans?

All such questions could be relevant to potential correlations between original designs of work plans, and variability of the overall plans and their consisting activities during executions. If the existence of such correlations could be validated and well understood, many undesirable outcomes could be prevented by project management teams through the selection of appropriate and proactive measures, such as revision of original work plans with high risks caused by its complicated graphical architectures; provision of additional managerial efforts to activities with potential problems due to its interrelations with other activities; and special watches for following activities when an upstream assignment fails to complete for specific reasons. Moreover, further integration of such relationships with other project related data, such as certain characteristics of production processes and/or individual activities, could enable more informative and effective reuse and analysis of records for designs and executions of previous work plans. Such an extra functionality will also help promote usefulness and applicability of project control systems like Last Planner™ by highlighting potential problems from new perspectives (in addition to those for individual activities), which can be used to further reduce variability of construction production processes and improve the overall project flow.

This paper is organized as follow. Major intentions and objective of this research are first introduced. Related work and an introduction to available techniques for graphical analysis are provided in the following section. Then a case study presents our ongoing research on new data representation for network-based work plans, with preliminary data analysis results evaluating the feasibility and validity of the developed solutions. Finally the paper summaries its findings, limitations in this work, future directions to extend the current solution, and importance of the research approach to improve current construction project control.

## OBJECTIVES

The major objective of this paper is to present our ongoing research effort to represent network-based work plans into novel and concise type descriptions that support graphical analysis and pattern recognition. A case study is presented to show how data preparation, representation, and analysis tools were applied on project control data from a large capital facility project to:

- Create generic and concise descriptions for network-based work plans with varied sizes and complicated architectures in support of further data analysis;

- Identify relations between original designs of work plans and their reliability, i.e., the probabilities that non-completions would occur during their implementations;

- Analyze possible associations between the position of a particular activity in a work plan and its probability of not being completed as originally planned.

## GRAPHICAL ANALYSIS

Work plans in production control systems could be represented in the format of Activity-On-Node networks with activities as nodes and precedent relationships among them as edges. Such nodes and edges are combined together into a complicated and integrated data structure, so that certain characteristics of these nodes or subsections within the graph are not only determined by their individual features, but also dependent on their interdependencies between each other. In this section, related research in graphical data analysis are first reviewed to show how existing tools have been applied on different domain problems and pattern recognition tasks to identify correlations between architectures of graphs and performance of their components such as individual nodes and subsections. Assumptions and focuses of such tools are then compared with the ongoing research for graphical analysis in work plans to show the necessity of a novel research approach (which will be detailed in the case study part) in order to achieve our objective as described above.

## PREVIOUS WORK IN SEQUENTIAL ANALYSIS

The first question presented in the introduction has been studied in a previous research effort (Oliveira, Soibelman and Choo, 2004) to recognize a special type of graphical pattern, ordered sequences of non-completions among work plans. The idea for this research on sequential pattern analysis was based on the observation that delays due to a specific cause sometimes result in non-completions of succeeding activities due to a same, or different but related reason. The research was intended to validate the hypothesis on correlations of problems occurring to sequences of activities, so that when a problem is identified, project managers can focus their production control efforts on eliminating a more limited range of problems for variability in the following activities (Oliveira, Soibelman and Choo, 2004).

In this research project, eight reasons for non-completions of individual assignments were defined in the Last Planner™ system employed by Strategic Project Solutions, Inc.: 1) information; 2) material; 3) labor; 4) plant/equipment; 5) weather; 6) directive; 7) pre-requisites; and 8) site access. Following the lean construction principles, the system has recorded the codes of reasons for failures to complete individual assignments in a large capital facility project on a continuous base. Applying sequential analysis tools developed in related research (e.g., Aggrawal and Srikant, 1995), the previous work by Oliveira, Soibelman and Choo (2004) was focused on discovering frequent patterns about problem codes among sequences of failures to complete activities as planned. Although no strong support were found in particular sequences of problem codes, such sequences provided good indications of some common and repetitive patterns that managerial personnel could be aware of and prepared for during production control. For example, same error codes were more frequently seen to occur in sequences of non-completions, which might imply that once an error code is identified, it should be corrected as soon as possible so that it will not contribute to further non-completions of downstream work. Also, it was argued that even though fewer instances for sequences of different error codes were discovered, explication and understanding of such potential connections might improve the overall control of project production processes as well (Oliveira, Soibelman and Choo, 2004).

## Related Research in Graph Mining

A limitation in applying sequential analysis tools on graphs like network-based work plans is its restriction in pattern representation, because sequences are only one type of specific graphical patterns existing in such complex graphical data. Some problems, e.g., duplicated counts of sub-paths shared by multiple sequences, may also be introduced due to misrepresentations of more general graphical patterns such as trees or sub-graphs (Oliveira, Soibelman and Choo, 2004).

To overcome this limitation, researchers in the data mining community have put a lot of efforts on discovering other generic and complicated patterns among graphical data in problem domains such as bioinformatics, social networks, and web links, among others. For example, supervised/unsupervised concept learning and anomaly detection techniques have been applied to extract general patterns in CAD graphs of electronic circuits, and to highlight potential defects in such graphs (Cook and Holder 2000; Noble and Cook 2003); graphical search and indexing tools were developed to discover frequent sub-sections in chemical compounds or protein molecules, with the objectives to associate such sections with specific characteristics of the compounds or proteins (Yan and Han 2003); and data modeling and analysis tools for matrix-based graphical data, e.g., social networks and co-authors of research papers, were used to identify closely related groups such as terrorist units or collaborating researchers (Chakrabarti, Zhan and Faloutsos, 2004). Interested readers can refer to their publications for further technical details.

Here we are more interested in introducing two research efforts in graph mining on work flow management systems, which are directly relevant to our research because they were both specifically focused on Directed Acyclic Graphs (DAG) such as work plans. Compared with other general graphical data as objects for analysis in the research efforts mentioned above, DAG have two distinct characteristics that require specific graph mining approaches for better data analysis performance: 1) all edges in such graphs are directed from a predecessor to a successor; and 2) there are no cycles within the graph, i.e., no activity could be its own predecessors or successor (otherwise processes based on such graphs would be implemented within the cycles for ever).

Agrawal, Gunopulos and Leymann (1998) proposed an approach in mining process models from workflow logs. It was assumed in their research that an unknown process model, consisting of multiple activities and precedent relations among them, has been implicitly implemented by many work flow instances. This situation is not unusual in practice because many procedures may be repeatedly executed without being explicated. These instances had been recorded in the format of ordered lists of implemented activities that satisfy the implicit precedent constrains, so that an activity can be only executed if *any* of its immediate predecessor has been finished. And the major objective of this approach is to recover the underlined process model as shown in Fig 1.

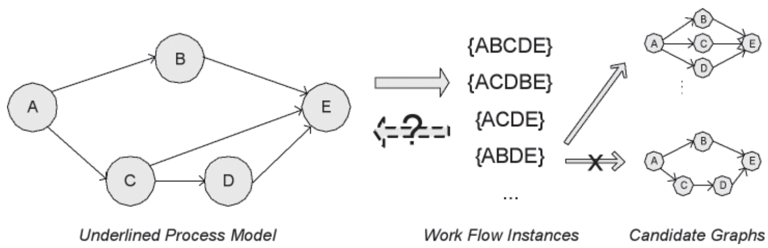


*Figure 1: An example of process model learning from work flow instances (Adapted from Agrawal, Gunopulous and Leymann, 1998)*

By analyzing each of the work flow logs and eliminating candidate graphs that are not consistent with the current instance as illustrated in Fig 1, the proposed approach could eventually identify a set of conformal graphs for all work flow logs. In this way, work flow experts may work on a very limited range of candidate graphs in order to develop a final process model from previous logs, instead of wasting a lot of time and efforts in manually generating such a model from scratch.

Greco et al (2005) initiated another research effort in graphical analysis on workflow management systems. Its basic assumption is that there exists a predefined work flow with mutually exclusive branches, uncertain results for specific activities, plus multiple ending points that are defined as successful or unsuccessful terminations. The objects for analysis are again work flow instances with their final results, i.e., successes or failures. By reviewing and analyzing a lot of work flow instances based on the same work flow graph, the major objective is to characterize crucial factors for successful or unsuccessful implementation results; to predict potential influences of choices among branches at each decision point; and to make optimal decisions when a new process is implemented. Fig 2 shows a simplified example of a work flow graph and its work flow instances.

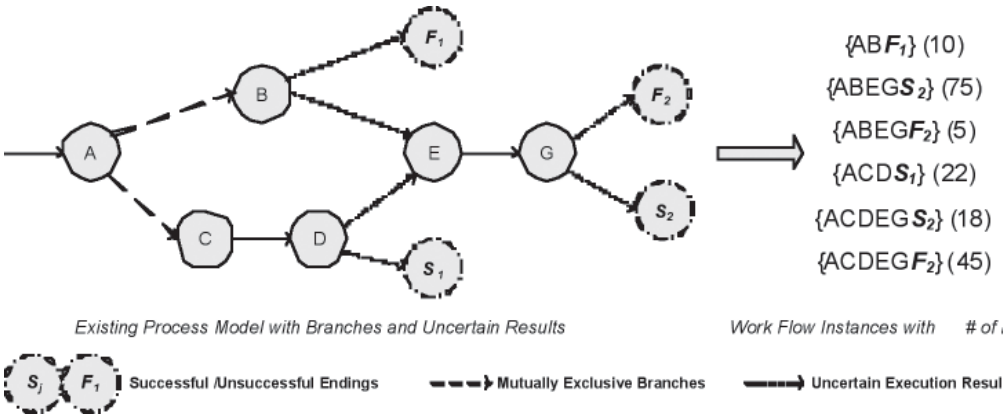


*Figure 2: An example of work flow pattern recognition (Envisioned Based on Greco et al, 2005)*

A possible pattern that could be identified in such instances, as envisioned in Fig 2, might be that among the mutually exclusive branches AàB and AàC at decision point A, the former branch should be less risky because instances following this branch yielded significantly higher percent of successful terminations, compared with instances with the second branch chosen during executions.

### Proposed Research on Graphical Analysis in Work Plans

The discussions in this section so far present many interesting research approaches for graphical analysis in various types of graphs with different pattern recognition tasks. However, such efforts described above had different assumptions and focuses compared with our research on graphical analysis in network-based work plans. Sequential analysis tools are restricted by their pattern representations in mining complex graphical networks; many graph-based data mining tools work on identifying frequent structures in generic graphical data without employing specific features of DAG ; and the two major research efforts on graphical analysis in DAG like work flows share similar assumptions that all work flow records are instances of a same underlined process model, with the objective to discover or analyze this process model based on the work flow instances.

In this research, however, we are intended to discover more generic patterns in network-based work plans employing their specific DAG features. Assumed that the work plans could have varied sizes and architectures without sharing a same underlined process model, the objective of this research is to test the hypothesis whether there exist certain correlations between graphical architectures of work plans and performance of the overall plans or their consisting activities, as suggested by the questions in the first section. Such differences require that a new research approach be developed for this particular graph type and corresponding pattern recognition tasks, which is under study in our ongoing research. The following section will detail our current solution with a preliminary case study, as a part of this research effort, on a same production control database used in the previous work (Oliveira, Soibelman and Choo, 2004).

## CASE STUDY

In addition to variability of individual activities, potential problems in original designs of production processes could also contribute to risks and variability during their executions: for example, complicated interdependencies among activities could introduce additional difficulties in coordination, while overly parallel implementations might result in more conflicts for explicit or implicit resources between activities. When generating and monitoring work plans, project managers need to take such factors into considerations in order to further decrease potential variability and improve work flow reliability. Related questions may include: which work plans, among all that will be implemented by various subcontractors or crews, are more likely to have delayed activities somewhere during their implementations, i.e., *non-conformances* with the original plans? Which activities could be more troublesome, i.e., more liable to non-completions due to their complex interdependencies with others? Answers to such questions could help project managers be better prepared for the problems, and distribute their limited managerial time and efforts to work plans and activities with higher risks to ensure their reliable executions.

Currently, such patterns have been seldom studied and could be only learned and used by individuals in a subjective and implicit way. Last Planner™ systems, by recording original designs of work plans and their implementation results, could provide useful data sources for learning more objective and explicit patterns for the above problems. The case study, as a part of our ongoing research, provides initial results for recognizing general patterns among work plans by just comparing their performance and data representations for their network structures developed in this research. Our further studies will integrate information of individual activities and production processes with data representations of their work plans for more specific graphical analysis and pattern recognition tasks.

Following some general processes for data mining, this case study was implemented in several steps including data understanding and preparation, data representation, data analysis, and pattern evaluation. Specific operations for network-based work plans in these steps were also identified and presented in each of the sub-sections as below.

## DATA UNDERSTANDING AND PREPARATION

The data for this study were collected from the same construction project as studied in the previous work (Oliveira, Soibelman and Choo, 2004). Strategic Project Solutions, Inc was working with the main contractor of the project under study to implement a Last Planner™ system for production control (Ballard, 2000) based on lean construction principles. The data spanned four months of excavation and foundation work. Roughly 38,000 activities were recorded with interdependencies between them, among which 18,000 activities were planned into about 2,600 work plans composed of 3~80 activities. Data regarding whether individual activities were delayed during implementation, as well as their corresponding reasons were also collected.

Besides the necessary work to separate the activities into independent networks based on the data table regarding predecessors and successors of all dependencies, a specific data preparation operation for graphical work plans was identified in this case study as the removal of redundant dependencies in each network. According to Kolisch, Sprecher, and Drexel (1995), a direct dependency from activities X to Y, say  xcavation and  aying pipes is needed if no other activities are involved. However, if there is another activity Z, e.g.,  leaning the bottoms which is a successor of X and a predecessor of Z (directly or indirectly), the dependency from X to Y becomes redundant because Y can not be started immediately after X anyway. Fig 3 illustrates an example for such a dependency.

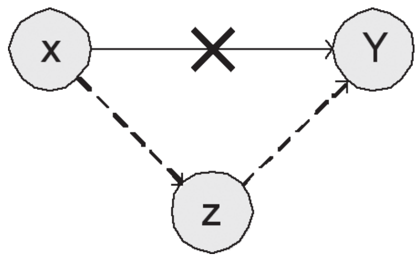


*Figure 3: An Example for a Redundant Dependency*

### Data Representation

It would be difficult, if not impossible, to analyze or identify common patterns among all 2,600 work plans with varied size and complicated architectures by just reviewing their original dependency tables or graphical networks. A generic and concise representation is thus necessary to transform such graph-based information into appropriate formats for data analysis while preserving the original architectural information of work plans.

Intuitively, how a network is split into branches or converged from branches could be a good indicator of its reliability during implementation. A possible explanation is that such splits or converges generate or eliminate parallelisms, if we assume that parallel implementation is a major cause for resource conflicts, and thus variability during actual implementation. Based on such an intuition, we developed our new data representation to describe each network in an abstract way focusing on how a work flow goes through it by splitting/converging, and finally arriving at its end. In this study, we established a novel abstraction process comprising of two major stages:

 Recursive Divisions in a top-down manner: the network in Fig 4 could be divided into two sequential sub-networks, N1/N2, so that N1 is preceding to N2 only through activity B, and N2 could be further divided into two parallel sub-networks, N21/N22, which only have common starting activity B and ending activity C; such divisions could continue until a sub-network eventually consists of only  tomic sequences of activities without any branches (e.g., all the three sequences from A to B).

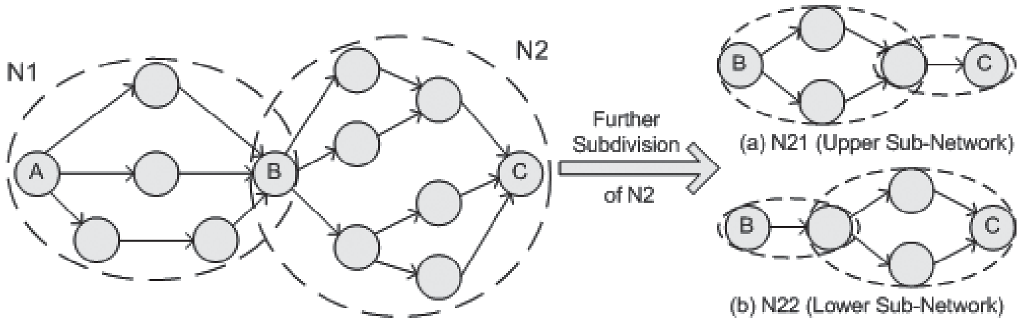


*Figure 4: An example for recursive division of a scheduling network*

Type description in a bottom-up manner: in this stage, activity identifications were removed since we only concerned the abstract description of the network architecture. In a reverse direction to recursive divisions, the type description for a network/sub-network could be obtained by following the rules illustrated in Fig 5. For example, the final type description for the network in Fig 4 will be $1 \rightarrow 3 \rightarrow 1 \rightarrow (2 \rightarrow 1, 1 \rightarrow 2) \rightarrow 1$.

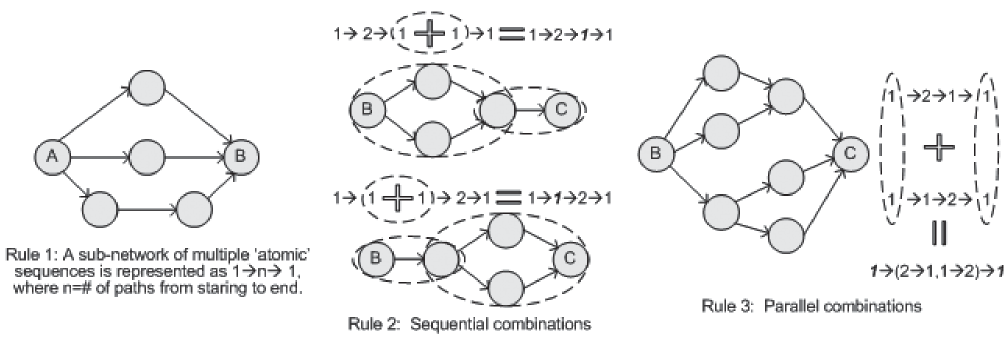


*Figure 5: Examples for type descriptions rules*

## DATA ANALYSIS

Type descriptions for all 2,600 work plans in this study were generated using a computer program developed on a Matlab® 7.0 platform. When the abstract type descriptions of these graphical networks were compared with their probabilities of non-conformances during implementation, some interesting patterns were identified as follow.

First of all, among the identified 63 distinct types for all 2,600 networks (except ~5% of them with exceptional network architectures on which the current solution was not applicable while we are still studying), some types of networks were apparently less likely to not conform with the original plans (i.e., to have at least one non-completions during implementation) than others. In an attempt to simplify the analysis, these types were categorized into 4 groups based on their common graphical representations: single sequence, single start to multiple branches, multiple starts to single sequence, and multi-hierarchical starts. The following table (Table 1) shows the general description, the total number of networks, and the *non-conformance rate* (measured by the number of work plans with non-completions over the total number of work plans in the same group) of each group, together with examples of graphical architectures and type descriptions for networks in those groups (Fig 6).

*Table 1: Groups of Network Types and their Non-Conformance Rates*

| Group No. | Group | Examples of Type Descriptions | General Description | # of Networks | # of Non-Conf. | Non-Conf. Rate (%) |
|---|---|---|---|---|---|---|
| 1 | Single Sequence | 1 | 1 | 1851 | 493 | 26.6 |
| 2 | Single Start to Branches | 1®2, 1®3®1, 1®2®1®2®1, … | 1®n®… | 51 | 12 | 23.5 |
| 3 | Multiple Starts to Single Sequence | 2®1,3®1®2®1, 4®1… | n®1®… | 365 | 159 | 43.6 |
| 4 | Multi-Hierarchical Starts | (1® 6,1)®1, (2®1,2®1)®1, … | Others | 167 | 98 | 58.7 |

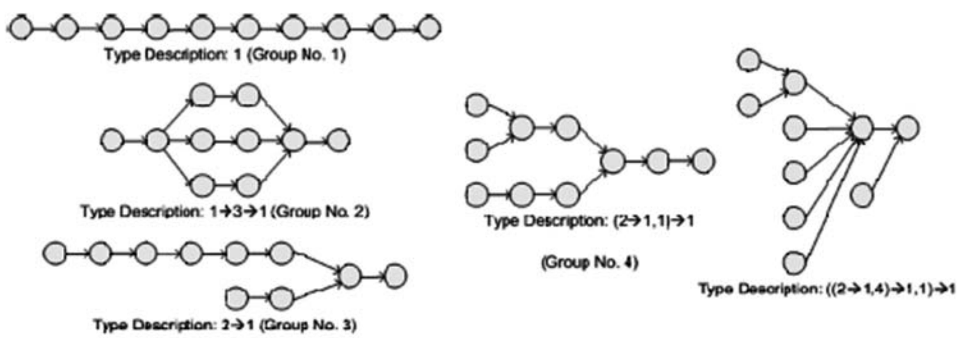


*Figure 6: Examples of Networks in Different Groups*

These non-conformances rates were then compared in a pair-wise manner using a statistical hypothesis testing equation below (Hogg and Tanis, 2001):

$$z_{\alpha/2} = \frac{|p_a - p_b|}{\sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}}} \quad (1)$$

in which 2 groups, A and B, with different non-conformances rates, were tested with a null hypothesis $p_a=p_b$ vs. an alternative hypothesis $p_a$‘«$p_b$. In such tests, $z_{±/2}$e»1.960 means that the null hypothesis was rejected with a 95% confidence, i.e., that groups A and B had dissimilar non-conformance rates by statistics. In comparisons using equation (1), groups 1 and 2 had significantly lower non-conformance rates than group 3 (with $z_{\pm/2}$=6.06 for the comparison between group 1 and 3 and 3.09 for group 2 and 3), and non-conformance rate of group 3 was also significantly lower than that of group 4 (with $z_{\pm/2}$=3.28).

Also, comparisons between non-completion rates of activities at different positions turned out some meaningful results. The non-completion rate of tasks connecting sequential or parallel sub-networks, e.g. those with more than one preceding and/or succeeding tasks, was significantly larger than that of tasks among tomic sequences (15.6% vs. 8.4% with a significant value $z_{\pm/2}$=8.94). When looked into more closely, non-completion rates of tasks in 'atomic' sequences varied with their positions in their corresponding sequences as well: the closer that a task is to the starting/ending tasks of the sequence, the higher non-completion rate it would have during implementation.

## Pattern Evaluation and Potential Usages

With these results, we may conclude that in this construction project, a production control network could be more reliable if it was designed to be a single sequence, or to start with just one activity. This makes sense in practice because when a work plan is started, the management team may not have been well prepared with all required resources; but after the plans are started, usually more necessary resources will be ready to support parallel implementations. In response to such identified patterns, if network structures of some work plans could not be changed to lower their potential risks at the design stage, project managers should focus their managerial efforts to control networks with type descriptions from group 3 and 4 to mitigate as much variability as possible, in addition to what they have been enabled to do for individual activities by Last Planner™ systems.

Also, in this project, activities at the splitting or converging points within a network, and those close to these activities, seemed to have higher probabilities of non-completions. This is a reasonable discovery too, considering that conflicts between activities, and thus non-completions of related activities, could be most likely to occur when parallel jobs are started or ended. Project managers should be better prepared for such activities by double-checking available resources, prerequisites, and other conditions.

## Discussions and future Work

The developed data representation for network-based work plans provided a concise and precise abstraction for graphical and complicated production processes supporting data analysis and pattern recognition. As shown in the case study, the current solution is both feasible and valid since it enabled automated abstraction of production control networks with similar architectures and performance into same or similar type descriptions. Moreover, even though these patterns are quite simple in their statements, they may be much harder to discover and describe if the complicated architectures of these networks were not automatically represented into such novel type descriptions.

It should be also noted that several assumptions had been made to simplify this case study as an initial step of our ongoing research on potential correlations between original designs of production processes and their variability, with the eventual goal to improve overall reliability of construction project control based on lean construction principles. For example, only topologies/

architectures of work plans were concerned in this case study, with no further data analysis that integrates independent features of activities, as well as other contexts of project production processes. Also, only non-completions of planned activities were used for comparisons of performance between different network architectures, while structural changes between as-designed and as-built work plans were not considered. In the next steps of this ongoing research, we will try to address such problems in order to extend the applicability of our current solution to broader types of production processes and more general pattern recognition tasks.

Moreover, the developed data representation and analysis approach could be related with other research areas in project scheduling and control. Potential research directions may include, but are not limited to: the data representation approach for network-based work plans could be applied on other CPM-based schedules with larger scales and more complex structures, so that schedules with thousands of activities could be decomposed and represented at multiple levels of details for user-friendly data exploration; the patterns recognized through graphical analysis could be integrated into knowledge-based systems developed in previous research (e.g., Dzeng and Tommelain, 1997) complementing other knowledge collected manually from books or experts for decision support; and further applications of data and/or graph mining tools on DAG-like project schedules could be developed based on primary insights and lessons obtained from this work, in order to identify and deal with more in-depth challenges and opportunities.

### Conclusions

This paper presents a case study in which a novel approach was developed to preprocess and represent network-based work plans from a project control system into abstraction type descriptions in support of graphical analysis and pattern recognition. Although the statistically valid and practically meaningful patterns identified are still simple and common senses at this early stage, they verified validity and feasibility of the current data representation approach in enabling knowledge discovery and lessons learning from production control processes for more thorough comprehension of potential variability in lean construction management.

Last Planner™ system of production control has been successfully applied in project control due to its focuses on plan realization and collection of problems that constrain the completions of work assignments. As discussed in this paper, further research on records from production processes should be one of the most important next steps, with the objectives to expose other potential problems related to non-completions of activities, e.g., influences from previous failures or variability due to complicated interdependences. Successful achievement of such objectives with related data representation and analysis tools could help project management teams understand and use such patterns in a more objective and explicit way, which will thus further improve their overall quality of project control.

### Acknowledgements

The authors would like to thank the U.S. National Science Foundation for its support under Grants 0201299 (CAREER) and 0093841, and Strategic Project Solutions, Inc. for providing the datasets for case study.

## REFERENCES

Agrawal, R., Gunopulos, D., Leymann, F. (1998). “Mining Process Models from Workflow

Logs." *Proceedings of EDBT'98*, Valencia, Spain Agrawal, R. and Srikant, R. (1995). "Mining Sequential Patterns." *Proceedings of 11th International Conference on Data Engineering*, Taipei, TaiwanBallard, G. (2000). "The Last Planner™ System of Production Control." PhD thesis, School of Civil Engineering, The University of Birmingham, Birmingham, UKChakrabarti, D., Zhan. Y. and Faloutsos, C. (2004) -MAT: A Recursive Model for Graph Mining *Proceedings of SIAM 004 International Conference on Data Mining*, Lake Buena Vista, Florida, USAChoo, H.J., Tommelein, I.D., Ballard, G., Zabelle, T.R. (1999). "WorkPlan: Constraint-based Database for Work Package Scheduling." ASCE, *J. of Constr. Engrg. and Mgmt.*, 125(3), 151-160Cook, D. and Holder, L. (2000). "Graph-based Data Mining." IEEE, *Intelligent System*, 15(2), 32-41Dxeng, R. and Tommelain, I.D. (1997). oiler Erection Scheduling Using Product Models and Case-Based Reasoning. ASCE, *\pard lang2052 J. of Constr. Engrg. and Mgmt.*, 123(3), 338-347Greco, G., Guzzo, A., Manco, G., Sacca, D. (2005). "Mining and Reasoning on Workflows." IEEE, *Transactions on Data and Knowledge Engrg.*, 17(4), 519-534Hogg, R.V.and Tanis, E.A. (2001). robability and Statistical Inference. (6th Ed.) Prentice Hall, Upper Saddle River, NJ, USAKolisch R., Sprecher A., Drexl A. (1995). haracterization and Generation of a General Class of Resource-Constrained Project Scheduling Problems. *Mgmt. Science*, 41(10), 1693-1703Nobel, C. and Cook, D. (2003). raph-based Anomaly Detection. *ACM SIGKDD 3*, Washington, DC, USAOliveira F., J.N., Soibelman, L., Choo, J. (2004). "Sequential Analysis for Reasons of Non-Completions of Activities: Case Study and Future Directions." *Proceedings of 12th Annual Conference of the International Group for Lean Construction,* Copenhagen, DenmarkYan, X.and Han, J. (2003). "CloseGraph: Mining Closed Frequent Graph Patterns." *Proceedings of ACM SIGKDD'03*, Washington, DC, USA,