Cascading Effects – Assessing the Performance Impacts of Fragile Tasks

Alicia Curran[1], Erhan Guven[1], Joshua Hamilton[1], Mark Hodgins[1], Troy Miller[2]

Abstract

The growing popularity of Joint Cost & Schedule Analysis has highlighted the need for quality Schedule Risk Assessments (SRAs). Modeling schedule risk and uncertainty requires an understanding of the behavior of schedule networks. Network Analytics (NA) has been furthered in recent years due to research in fields such as social networks, IT networks, and transportation networks. Key aspects of these advancements can be used in SRAs to improve our understanding of schedule risk and mature our modeling techniques. This paper integrates classical concerns in schedule analytics, principally Merge Bias, with NA processes, such as node centrality measures and edge properties, to uniquely identify fragile tasks and illustrate how delays in these tasks cascade through a schedule and ultimately affect program execution.

Section 1: Introduction

Mature standards and best practices exist to monitor and enforce adequate contract execution for the Department of Defense (DoD). Earned value management (EVM), for example, seeks to accomplish the technical scope of a project, subject to minimizing schedule delays and cost overruns, by creating an integrated baseline of program planning and performance (Alexander, 2002). Despite these best practices, the acquisition process occasionally fails, as evidenced by programs incurring Nunn-McCurdy breaches that necessitate a re-baseline or, worse yet, program cancellation. Johnson (2018) asserts that schedule delays and cost growth are the norm for acquisition programs rather than the exception. As one empirical example, the Government Accountability Office (GAO) reported that 98 major development acquisition programs (MDAPs) cumulatively overran their budget baseline by $402 billion and were an average of 22 months delayed in their schedules in 2010 (Hofbauer et al., 2011: 3). Industry seems to be experiencing similar challenges with nearly 90% of large construction projects overrunning (Santolini et al., 2021: 1).

As a result, research is being conducted to further our understanding of the various program management tools and to improve our techniques. One tool that has experienced a level of success is Joint Cost/Schedule Analysis. This technique attempts to combine Cost Risk Analysis with Schedule Risk Analysis (SRA) in order to identify a Joint Confidence Level (JCL). While JCL has experienced a level of success, our understanding of schedule risk is still in its infancy. As such, we have a great deal to learn about schedule variability and how task delays materialize and propagate through a schedule. This paper attempts to further our understanding of schedule risk and uncertainty by adapting techniques developed in Network Analysis and other academic research to a DoD MDAP program schedule to assess the applicability of these techniques in a DoD environment.

[1] Johns Hopkins University Applied Physics Laboratory
[2] Tecolote Research Inc.

An immediate problem that confronts schedule risk analysis is that, in some respects, programs seem to perform well compared to their baseline. Figure 1, below, illustrates the distribution of task performance relative to its baseline. This histogram represents performance of a typical team on the DoD MDAP program. It will be noticed that there is a significant peak at 1.0 (Actual Duration = Baseline Duration) with many tasks coming in with a ratio less than 1 (Actual Duration < Baseline Duration). This performance seems to be prevalent in other DoD programs as well.

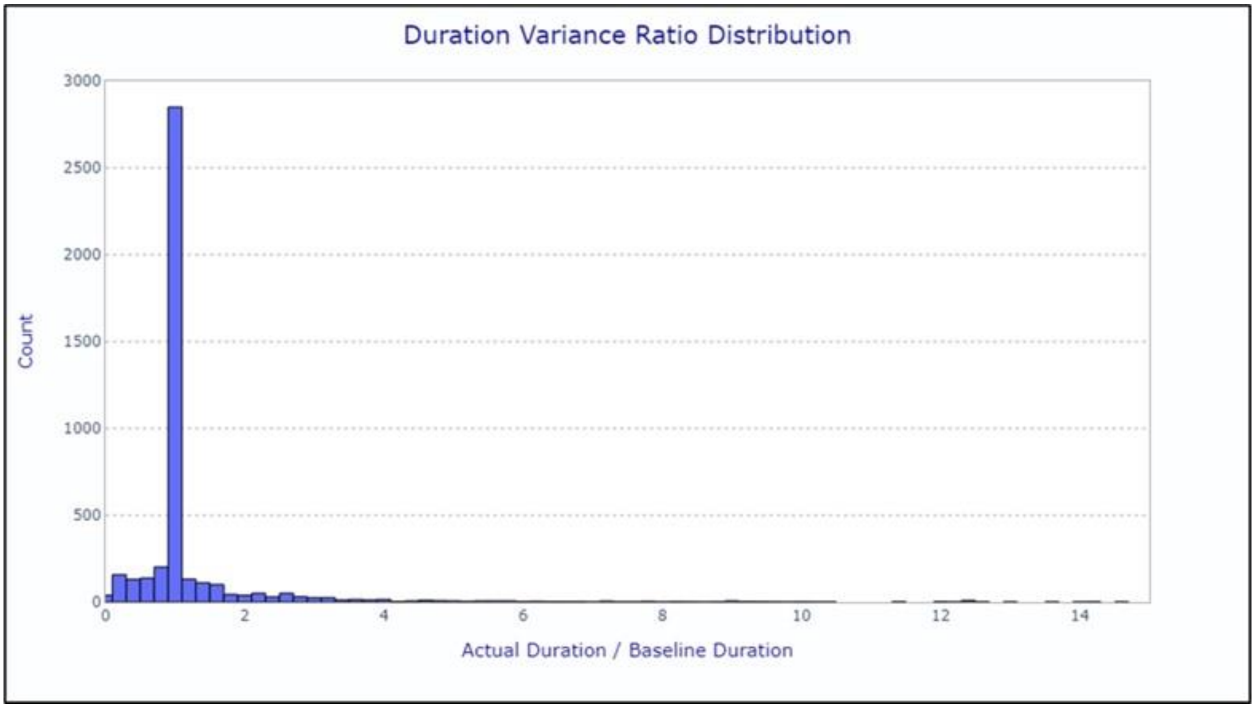

Figure 1: Duration Variance Distribution

With an actual duration less than or equal to the baselined duration in the vast majority of tasks, why is it that programs consistently slip their schedules? There are, perhaps, many explanations for this phenomenon, but it seems that there are three primary candidates: One, data is reported improperly and tasks are marked as being completed when, in reality, they have not. As a result, slips (delays) are being realized, but not reported until it is too late to mitigate. While this explanation may be correct, we have little ability to test the theory and, therefore, it will not be addressed in this paper.

The second possibility is that program teams are overly optimistic during the baselining processes. It seems likely that excessive optimism about execution and/or program requirements, leads to missed scope that must be added to the program after it has already been baselined. Baseline Change Requests (BCR) are the mechanism used to manage the added complexities and derived requirements. BCRs frequently result in additional tasks and logic being incorporated into the schedule. Generally, the BCR process considers only the expected task durations which leads to the conclusion that the added tasks and logic can be incorporated without materially affecting the milestone of interest. Because the analysis does not account for the associated uncertainty, the baseline changes may show little-to-no change to the milestone while the probability of achieving said milestone has dramatically decreased. While this explanation seems likely and is a good candidate for further research, we will not pursue it in this paper.

A third explanation, and the one that will be pursued in this paper, focuses on the complexities of schedule networks. Academic and industry experts have conducted root cause analysis to understand why project management, in general, is repeatedly unable to ensure timely and financially affordable results. System complexity is often cited as a reason. Geraldi et al. (2011) executed a systematic review of the project complexity literature. The speed of execution, misaligned incentives from organizational hierarchy, and dynamic changes in personnel or system requirements are three commonly observed reasons, but they are not the number one reason. Structural complexity, in fact, was identified as the most significant cause of project execution issues, with large project size, task variety, and high interdependencies noted as evidence (Geraldi et al., 2011, pp. 976-977).

Current schedule estimation methods often generate a critical path and use the idea of free schedule float, which is the amount of time a task can delay before successor tasks are subsequently delayed, to estimate if a delay will occur. This type of analysis treats the delay relationship between predecessors and successors linearly. However, interdependencies and task uniqueness may impose nonlinear relationships, where a delay in a predecessor task could lead to larger delays in successor tasks.

This non-linear growth in a delay is also referred to as the cascading effects of a delay. Ellinas et al. (2023: 4) evaluated a multi-billion-dollar infrastructure project with more than 65,000 tasks and found that using the free float method underestimated the probability of a delay propagating and cascading from predecessors to successors. Similarly, Vazquez et al. (2023) postulate that the focus on a critical path narrows the focus to an increasingly small number of tasks as modern-day projects grow in complexity and does not address how project complexity impacts the critical path. The authors model the schedule as a complex network to show that the relative size of a notional project’s critical path shrinks as the network increases in size.

Given the strong evidence that schedule complexity can exasperate delays as they cascade through a schedule network, it was determined that we need to understand if and how these patterns manifest in DoD programs and the extent to which the programs are affected. Specifically, what tasks should leadership be most concerned with, and can we identify those tasks by their potential to yield a catastrophic cascade? Ultimately, the question “boils down to” 1) which tasks are most likely to slip and 2) which tasks are most likely to cause a catastrophic failure to the schedule. During the literature review process, we began to refer to these tasks as “Fragile Tasks.” In the following sections we define several NA metrics (section 2) and then discuss their application to schedule analytics (sections 3-5).

Section 2: Network Analysis (NA)

Section 2.1: NA Applications to Project Schedules

Networking and graphing analytic techniques may enable the acquisition community to explore today's highly complex schedules so that programs can more effectively monitor and enforce contractor's schedule performance. Graphs and networks are mathematical structures used to model relationships between objects, represented in the form of nodes and the edges that join and connect pairs of nodes. Node relationship types include cycles and feedback loops, the direction of a property relationship, and relationship strength. Networks may be either cyclic or acyclic and either directed or undirected. The relationships among the nodes can further be defined by weighting the strength between two nodes or directing property relationships among the nodes. As an example, the Internet is a weighted undirected graph with routers and hosts as nodes. Graph edges represent the volume of connections, such as using fiber cables (strong) versus WIFI (lighter) that connect the nodes together (Brandes and Erlebach, 2005, p. 8).

Many real-world structures and relationships can be modeled using the foundations of network analytics (Jacob et al., 2017). As a result, graphs provide a versatile framework for analyzing complex systems and relationships across a large number of academic and industry fields, such as computer science, biology, epidemiology, social sciences, and operations research. As an example, this analysis is used in the social sciences to identify corporate relationships that maximize profitability and the construction of social structures in education that maximize a person's creativity.

Similarly, large project schedules, such as DoD MDAPs, may be treated as a network. Project tasks and their dependencies can be modeled as a directed graph, where edges have directions specifying property relationships, such as the order of tasks, completion dependencies, and parent and child relationships. Such a data model facilitates the application of a large set of tools to examine characteristics of the project, such as project execution within a portion of the weapon system (e.g., specific work breakdown structure efforts), within a subgraph (e.g., a project milestone), or within the global network (e.g., the complete network of project tasks).

Recent graph and network analytic efforts, in fact, have emerged as powerful tools to diagnose and forecast schedule risks for large-scale engineering and construction projects (Santolini et al., 2021; Pozzana et al., 2021; Ellinas et al., 2023; Vazquez et al., 2023). These methodologies enable the examination of complex interdependencies among various project activities, providing a detailed understanding of how different components work together or influence each other. By representing projects as networks of singular or connected entities, analysts can identify critical nodes and paths that significantly impact the project's acquisition health. This approach is essential for understanding the propagation of risks, delays, and cost overruns within a project, which can have substantial financial and schedule implications.

Graph and network analysis may offer several specific analytic advantages relative to other methodological approaches. These techniques enable visual observations of the intricate relationships among different nodes. As a social network example, a node's location in a network serves a significant role in the node's outcome by defining the node's opportunities and constraints (Borgatti et al., 2009, p. 894). This capability may be useful for project schedule assessments as well. Network diagrams can illustrate how various tasks are linked, highlighting potential bottlenecks or critical paths that may affect the project's overall timeline and budget. By analyzing these connections, project managers can make informed decisions on resource allocation, scheduling adjustments, and contingency planning, ultimately leading to more efficient and cost-effective project execution.

Moreover, quantitative measurements potentially provide unique insights into the value of specific nodes within the network. The most prominent investigation at the node level is the concept of centrality, which is a "family of node-level properties relating to the structural importance or prominence in the network" (Borgatti et al., 2009, p. 894). Centrality measures are often used to identify critical nodes, such as key influencers on social media platforms. In terms of acquisition scheduling, centrality measures can help identify critical tasks within the project network. These nodes, often representing key activities or milestones, can have a disproportionate impact on the project's cost or schedule outcomes. By analyzing the network structure, project managers can prioritize these critical nodes, ensuring that sufficient resources and attention are allocated to them. This prioritization may help mitigate risks and avoid costly delays or overruns that could negatively affect the project's performance (Pozzana, 2021).

Importantly, network analysis may help overcome existing EVM limitations. The application of network analytics in acquisition scheduling may allow for a more holistic view of project risks and opportunities. As stated above, traditional schedule analysis methods often assess cost and schedule risk linearly. In addition, standard methods often focus on individual elements of a project, such as cost estimates or timelines for specific tasks. On the contrary, network analytics considers the project as an interconnected system, enabling the identification of interdependencies. These cascaded dependencies can create a cumulative effect where issues in one part of the project can ripple and surge through the entire network.

A comprehensive, unified perspective is particularly valuable in large-scale projects where the complexity and interdependencies of tasks make it challenging to predict outcomes based on isolated factors. The graph and network analytics also facilitate scenario analysis and stress testing in DoD project planning. By simulating different scenarios and their potential impacts on the network, analysts can assess the project's resilience to various risks and uncertainties. This approach allows for the development of robust acquisition plans that account for a range of possible outcomes, enhancing the project's ability to withstand unexpected challenges and changes in the external environment. The next section discusses the adapted methodology that is applied to a contractor's integrated master schedule (IMS) for a DoD MDAP.

Section 2.2: NA Methodological Approach

This study primarily uses the methodological approaches from Pozzana et al. (2021) and Santolini et al. (2021) to explore real-world, completed projects as complex[3] networks and identify the project characteristics that cause the most significant schedule risks and issues. Both papers emphasize the potential for a delay in an activity to manifest in larger schedule deviations as a project continues. A dynamic, complex network can lead to "spreading," where a localized schedule issue on a single activity disrupts the entire project's schedule performance (Pozzana et al., 2021, p. 2). Santolini et al. (2021, p. 3), likewise, focus on how small, isolated task delays can lead to large scale project wide delays because they "cascade" across the schedule network.

These papers attempt to statistically measure the relationship between activity perturbations, which are deviations between actual and planned events, and a few centrality measures, which help identify tasks of interest. This study defines perturbations as delays and assesses several types, which include start delay, finish delay, and duration delay. The centrality measurements incorporated in the papers and this study, are node reachability and node degree. Node reach is generally defined as the number of nodes (e.g., successor tasks) reachable downstream for a given task. Pozzana et al. (2021, pp. 6-7) created a reachability – heterogeneity (RH) measure to test a hypothesis that the combination of a task's reach and the complexity of a project leads to elevated network fragility and schedule risk[4]. This study replicates the RH metric and adopts it as a measure of reach.

Node degree is a second type of centrality measure pertinent to schedule risk evaluation. Node degree is measured by the number of nodes that can be reached from the reference node (Jacob et al., 2017). In acquisition efforts, tasks with significant dependencies are commonly cited as a source of schedule risk. Network Analysis further stratifies the definition of degree to include In-degree and Out-degree.

- In-degree is defined as the number of nodes that directly feed into the node of interest. In schedule analytics, these are known as direct predecessors. A task with a high in-degree has many direct predecessors, and, therefore, a greater possibility of incurring a delay.
- Out-degree is defined as the number of nodes that directly follow the task of interest. In schedule analytics, these tasks are known as direct successors. So, a task with high out-degree has lots of successors and, therefore, greater opportunity to propagate a delay.

While not a focal point of the study, additional network centrality measures may help score the importance of a node, so they are included in forecasting future task fragility. Examples include: (1.) Betweenness centrality, which measures a node's importance in a network based on its

[3] The academic community refers to a network as complex when it contains a large spectrum of nodes with irregularities in the number of edges that connect the nodes, while a non-complex network possesses nodes that mostly possess the same number of edges, thereby visually representing a lattice structure (Jacob 2017: 1-2).

[4] The RH metric is a modified heterogeneity measure (Estrada 2010), which seeks to measure the complexity as the uniqueness in the number of node degrees and structure of the network (e.g., schedule for our purposes). The RH metric compares the overall quantity of predecessors and successors for all tasks across the entire schedule against the same measure, but with a specific node and its lineage of predecessors and successors removed.

presence on the shortest paths between other nodes; (2.) Closeness centrality quantifies how close a node is to all other nodes in a network based on the average length of the shortest paths; and (3.) PageRank algorithm, which is used to rank nodes by assessing their importance through the quantity and quality of links to them.

The centrality measures of interest share similarities with project scheduling concepts. In-degree is particularly important in schedule analytics because, under the right circumstances, a high in-degree is almost certain to result in a task slip. This concept is commonly known as Merge Bias. The following section discusses the merge bias concept and how the properties of tasks possessing this bias are typically analyzed.

Section 3: Merge Bias

To answer the question which tasks are most likely to slip, we turn to a topic that is not terribly new to schedule analytics – Merge Bias. Merge Bias occurs when a task or milestone has a high in-degree. In scheduling terminology, merge bias occurs when a task or milestone has many direct predecessors. Under certain circumstances these predecessors almost certainly yield a schedule slip to the successor task, or milestone, and virtually ensures that it occurs on its worst possible date.

Section 3.1: Simple Illustration

To illustrate this phenomenon, consider a common situation where a milestone can only be completed once all of its predecessors have been completed. These are known as Finish-to-Start relationships and can be illustrated by a test that cannot occur until all the component hardware has been delivered, or a Critical Design Review that cannot take place until the system design has been completely documented.

Let's pursue the example of a test and relate it to the tossing of a fair coin. Tossing a single coin might be analogous to a situation where only one prototype needs to be available before the test can be conducted. For this illustration let's consider that a "Head" represents delivering the prototype on-or-ahead of schedule and a "Tail" to represent a late delivery and thus a slip to the testing schedule. In this situation, we would consider that the probability of successfully meeting the testing milestone is 50 percent.

If, however, we require that two, prototypes be present for the test and that the delivery of both components are independent and scheduled for the same day. The probability that the test is completed on time is reduced to 25 percent. The analogy to the flipping of coins is quite simple since we have two coins and there are four, possible outcomes to the flipping of the coins. The result could be two heads (both prototypes are delivered on time), a head and a tail (prototype 1 is delivered on time and prototype 2 is delivered late), a tail and a head (prototype 1 is delivered late and a prototype 2 is delivered on time) or two tails (both prototypes are delivered late). Again, since a head represents on-time delivery and tail represents late delivery and both components must be received prior to execution of the test, only one of the four outcomes yields on-time testing (head for coin one and another head for coin two). Thus, the probability of successfully completing the test on time reduces to 25 percent.

Without pursuing the mathematics (since that is not the purpose of this paper), it is known that as the number of independent components increases the probability of on-time delivery of every component decreases and the probability of a delay to the test also decreases. Because all of the deliveries are planned for the same date, the efforts to design and build them converges (or merge) to the test activity and practically ensure a slip to the test (or are biased toward a slip).

Section 3.2: Application to Project Schedules

In an SRA, a task duration has a probability distribution associated with it. In our example, the design/build process has a start date and a duration. The finish date, or delivery date, can be calculated as the start date plus the duration. So, while we may think this process takes, for example, 60 days, we accept that it could possibly take as few as 50 days or take as long as 100 days. The distribution of delivery dates, then can be calculated as illustrated in the cumulative distribution function, or S-Curve, in Figure 2, below.

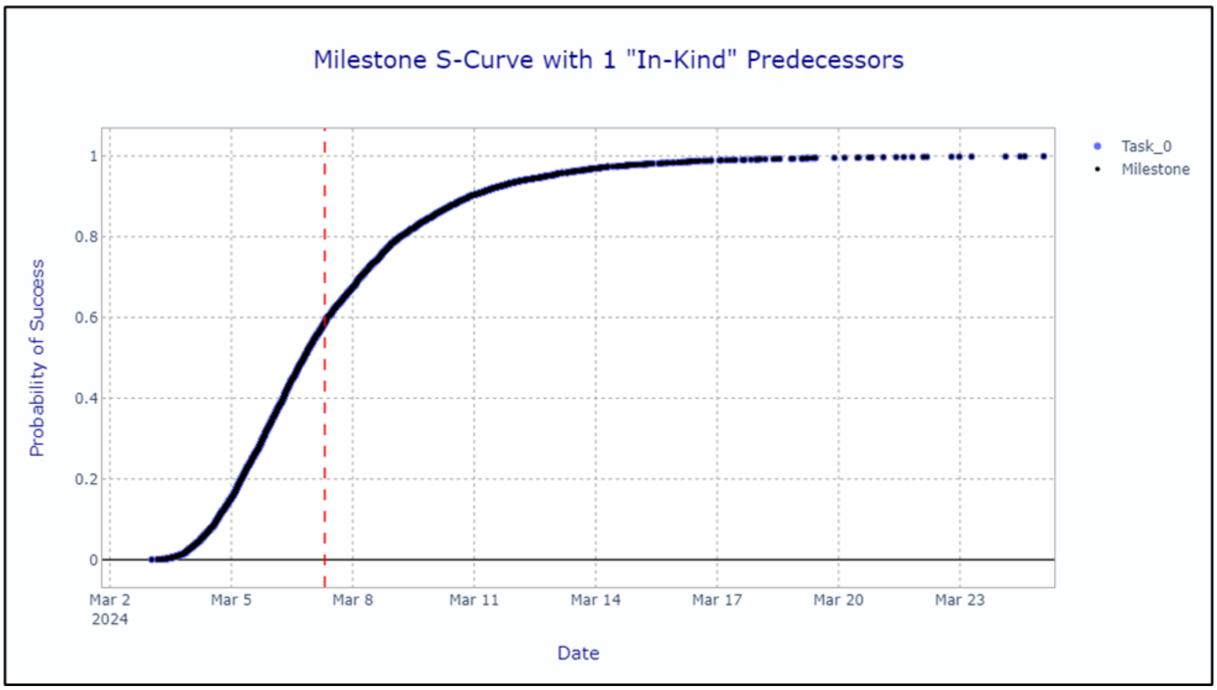

Figure 2: Milestone S-Curve with 1 "In-Kind" Predecessors

In this example, testing occurs as soon as the prototype is delivered. Therefore, the distribution of the testing milestone mirrors the distribution of prototype delivery. Here the baseline schedule has a 60 percent probability of on time execution.

If we add a second prototype that shares the same delivery schedule as the first but who's delivery is independent, then the distribution of the testing milestone changes – as illustrated in Figure 3 – and the probability of on-time test execution reduces to less than 40 percent. So, while the probably of on-time delivery for each individual prototype is 60 percent, the probability that they both are delivered on time – and thus that testing occurs on time – reduces by nearly 20 percentage points.

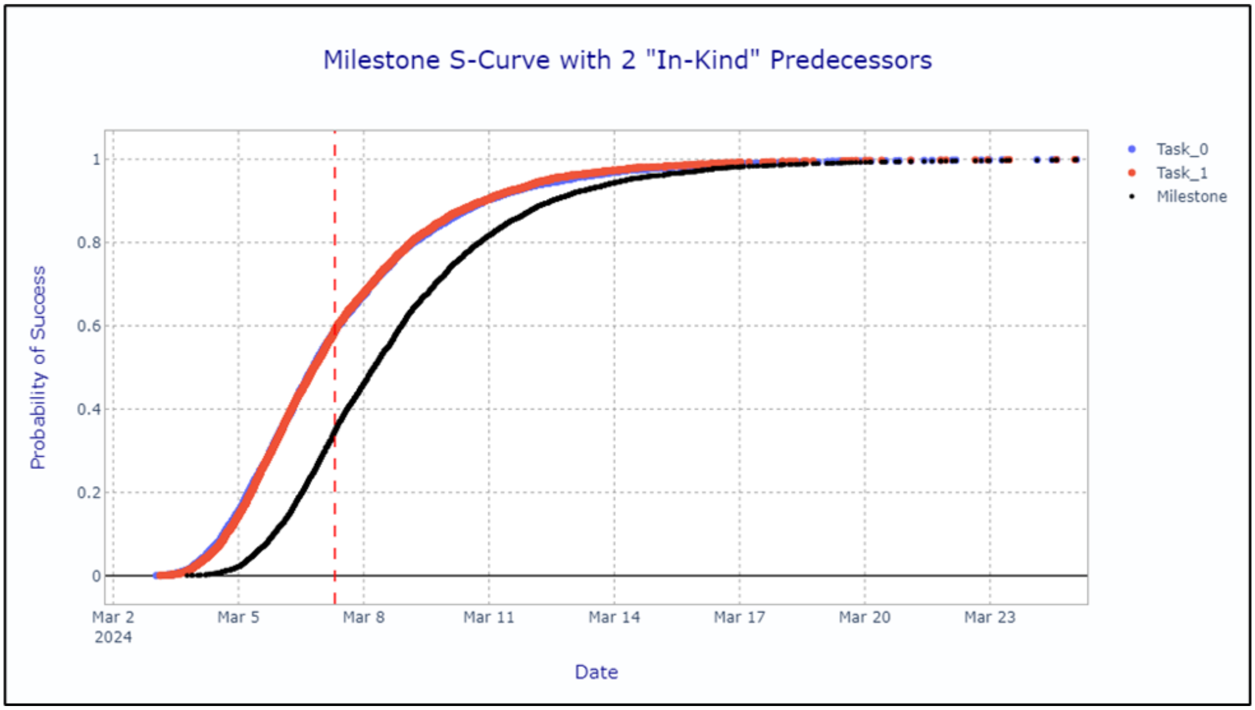

Figure 3: Milestone S-Curve with 2 "In-Kind" Predecessors

Figure 4, illustrates the effect of five prototypes. Again, each component has the same delivery date with a 60 percent probability of on-time delivery. In this case the probability of conducting the test on time, decreases to less than 10 percent.

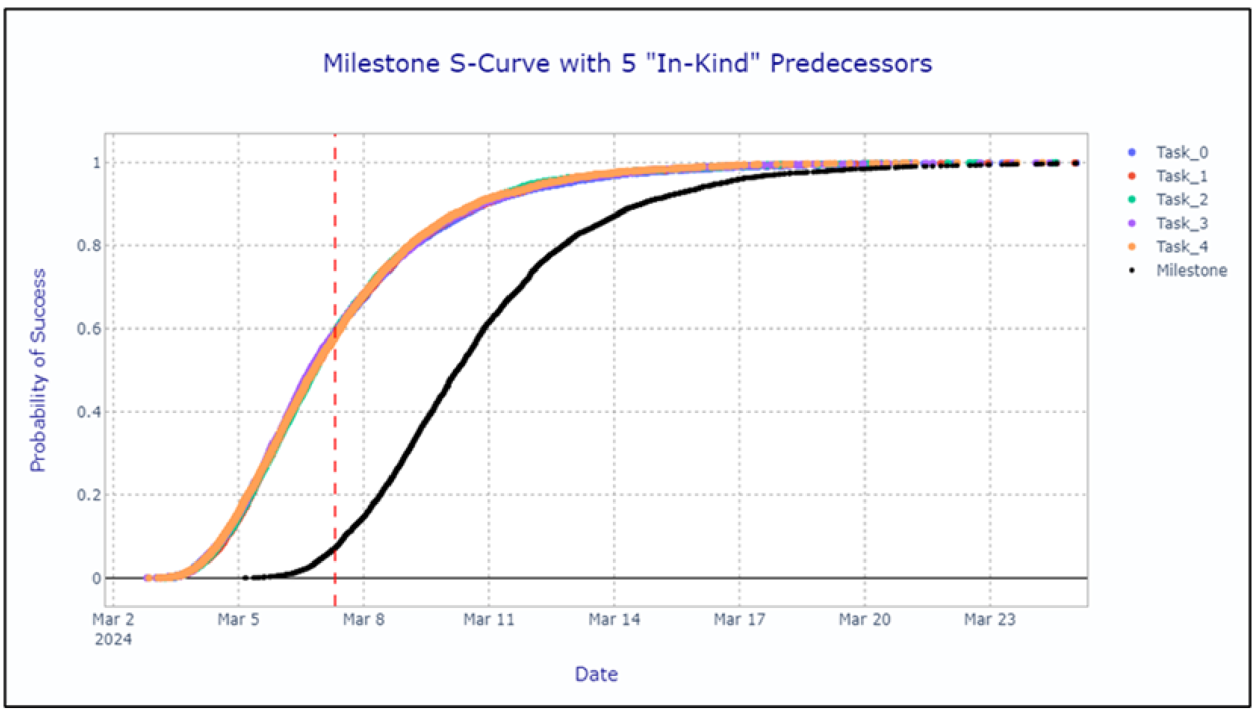

Figure 4: Milestone S-Curve with 5 "In-Kind" Predecessors

Finally, as illustrated in Figure 5, with only 15 predecessors the probability of on-time testing reduces to practically zero.

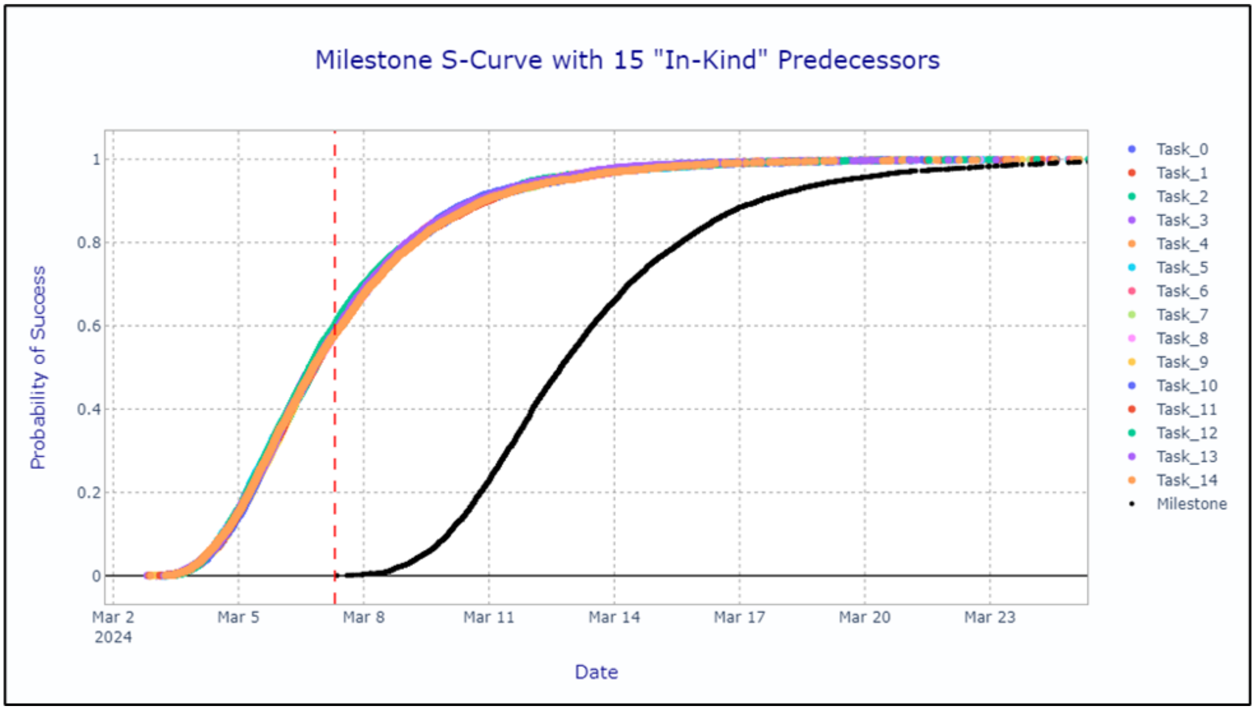

Figure 5: Milestone S-Curve with 15 "In-Kind" Predecessors

Section 4: Methodological Execution and Results

A few details about the data sources and analytic environment are warranted before describing how the study team employed this methodological approach to a DoD MDAP. The primary data source is Integrated Program Management Data Analysis Report (IPMDAR), which is a monthly electronic data deliverable of cost and schedule performance provided by the contractor. The two main data sets provided in each report are the contract performance dataset (CPD), which includes earned value and budget baseline data, and the schedule performance dataset (SPD), which contains current information regarding the contractor's IMS. This research predominantly used the Tasks, Task Schedule Data, and Task Relationships tables in the SPD.

The study team executed the analysis in SQL and Python based environments. Specifically, the team utilized SQL, Pandas and NumPy to support data frame structures, data preprocessing, and normalization. We implemented igraph to represent schedules as network graphs, assess those networks, and generate visualizations. Plotly was also used in conjunction with igraph to provide more interactive visualizations. Lastly, we used SciPy for statistical testing, such as Spearman correlation.

Section 4.1: Merge Bias Execution and Results

As described in earlier sections, for merge bias to be meaningful some conditions must be met. These conditions include:

1. Multiple schedule paths must converge on a single task/milestone.
2. The tasks must have a "Finish-to-Start" relationship with the task/milestone of interest.
3. The baseline finish date for all predecessors must be roughly the same.
4. And any lag between the predecessor tasks and the milestone must be minimal.

For the MDAP program, approximately 397 tasks were found to have predecessors that meet the criteria for merge bias.

Figure 6, below, plots the number of predecessors against the Days of slip. It may be noted that with only one predecessor there are a number of tasks that slipped as well as a number that came in on-time. However, of the tasks with between two and ten predecessors, very few came in on-time. It will also be noted that there are very few tasks that meet the criteria and have more than 15 predecessors. With a sample size that is so small, it's difficult to make a definitive statement about the effect of the number of predecessors on the on-time execution of the task-on-interest.

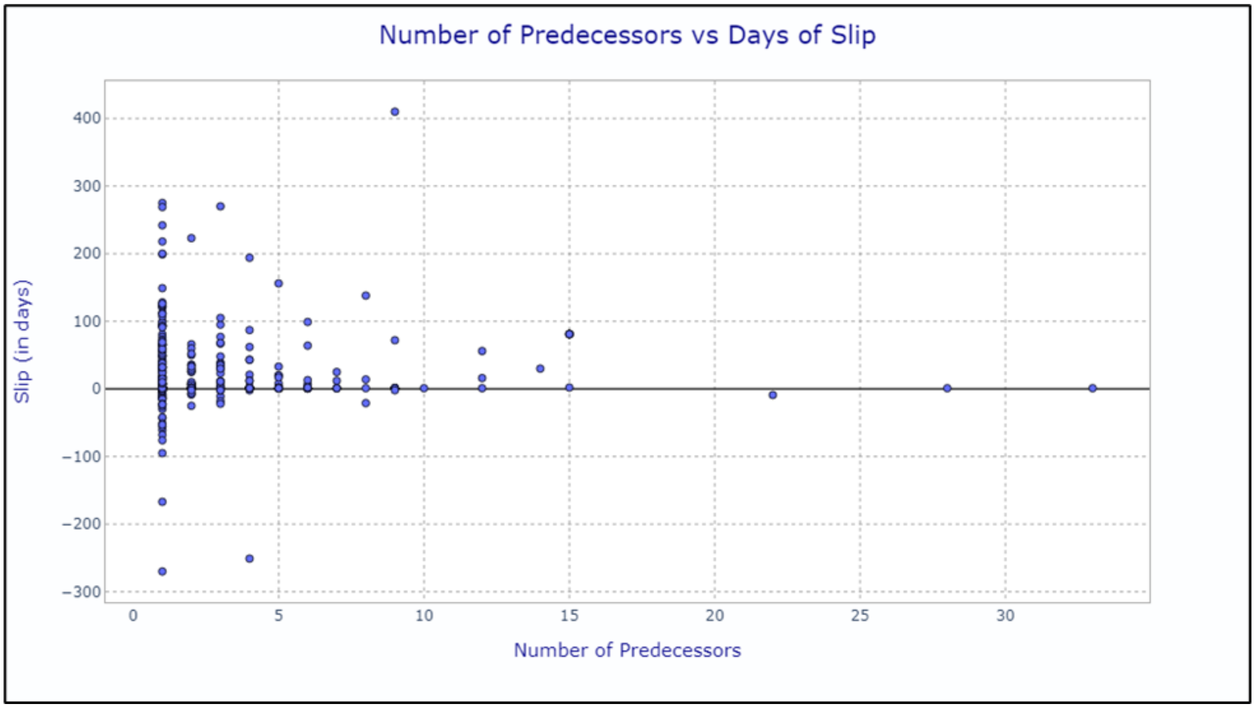

Figure 6: Number of Predecessors vs Days of Slip

Section 4.2: Network Analysis Execution and Results

The approach executed by Santolini et al. (2021) and Pozzana et al. (2021) served as the basis for this study; however, several modifications were necessary due to the constraints encountered by the research team. The primary constraint is data accessibility. Santolini et al. (2021) conducted their exploratory and explanatory analysis on approximately 15 large scale construction and engineering projects. On the contrary, the study team has access to one MDAP project schedule. Moreover, the project is less than 50% complete, which limits the extent and significance of the explanatory analysis. An ongoing project entails fewer completed tasks and thereby limits the ability to statistically evaluate whether certain task characteristics led to cascading failures in the project.

Therefore, the study team examined whether the major milestone activities within the overall MDAP project were representative of a complex directed acyclic graph (DAG) with individual tasks that could be fragile to cascading failures. The MDAP project adheres to rolling wave planning, where tasks are defined for a limited number of future time periods in order to accommodate planning uncertainty. The study team examined activities designated as sub-milestones, which are the next immediate children to major milestones within the current planning horizon, as a mitigation to incomplete future schedule planning.

In total, the IMS includes 12 major milestones, and 13,000 sub-milestones. The study team focused on one major milestone with approximately 5,500 tasks and 2,250 sub-milestones. The evaluation of sub-milestones introduced the possibility of task interdependency among the sub-milestones as an unintended consequence. A lack of independence could bias the centrality measurement of high-risk tasks upward. Thus, a correlation matrix of tasks among sub-milestones was executed, which found a near zero relationship of tasks among the sub-milestones. The tasks across the sub-milestones are consequently assumed to be independent, which allowed for the methodological execution to proceed.

Next, the study team conducted an exploratory analysis to confirm that the sub-milestone networks possessed properties of a relatively complex network with potentially fragile tasks. Santolini et al. (2021) measures were adopted in this analysis and their valuations served as a benchmark for determining the network complexity. The sub-milestones task count varies considerably, ranging between 1 predecessor task and roughly 725 predecessor tasks. Smaller node (e.g., task) counts are less likely to possess characteristics of a network where delays could propagate across the schedule network, so sub-milestones with a predecessor count less than or equal to 25 tasks are excluded from the analysis, leaving less than 750 sub-milestones. Another constraint with the study's MDAP IMS pertains to the overall length of sub-milestone networks. The global structure of the sample networks (e.g., project sub-milestones) possesses network diameters (e.g., the longest path of successive tasks) ranging between 5 and 54 activities, while Santolini et al. (2021: 3) size ranges between 31 and 191 activities. Longer diameters are associated with a greater likelihood of complexity; however, approximately 35% of the sub-milestones fell within the exemplar paper's range and no other data was available, so the study deemed the diameters an acceptable limitation.

For the remaining network complexity metrics (density, cycling, clustering, spreading distance, and cascade size), the measurements are comparable to the previous related work Santolini et al. (2021). A brief discussion follows for each:

- The density[5] of the sub-milestones ranges between 0.003622 and 0.07381, which exhibits characteristics of a sparse network, similar to Santolini (page 3).
- The number of cycles for each remaining sub-milestone was calculated by computing the relations among each task in each network to ensure a linear rather than circuitous

[5] The ratio between the total number of sub-milestones edges and the total edges in the network.

relationship. All sub-graphs possessed zero cycling, which ensures that the schedules are directed and acyclic.

- Clustering is calculated as the number of completed tasks with a schedule delay and the percentage of their predecessor tasks that also have a delay. The clustering calculations for the MDAP program align with Santolini (page 5). The findings suggest that predecessor tasks with a schedule delay are more likely to cascade into successors.
- Spreading distance measures the correlation of the average schedule delay over varying distances for each task in the network. The schedules in this study, in fact, possess a positive correlation, which indicates that delays are propagating across multiple successors in the networks. This finding aligns with Santolini (pages 10-11).
- According to Santolini (page 6), cascade size is calculated as the number of downstream nodes from an initial perturbation that also experienced a delay. Cascade size can be categorized by a perturbation in the Finish Dates, Start Dates, and Duration.
    - Finish Date Cascade: The study team finds (Figure 7) that for finish delay (i.e., a positive difference in actual and baseline finish dates), the impact of the initial delay declined after the first downstream task, with significant positive values up to six activities downstream, which is indicative of delay clustering in local neighborhoods. The correlation values then become analogous to those obtained when delays are randomly assigned to nodes within each network.
    - Start Date Cascade: We find similar patterns for start delay where the duration of the impact lasts for eight downstream activities (Figure 8).
    - Duration Cascade: Lastly, in accordance with Santolini et al. (2021), we find that correlation values related to duration delay closely follow a distribution with randomly dispersed delay (Figure 9). Thus, the cascades for start and finish delay resemble a distribution that departs from a simulated model with random task delay durations and is subject to compounding schedule issues.

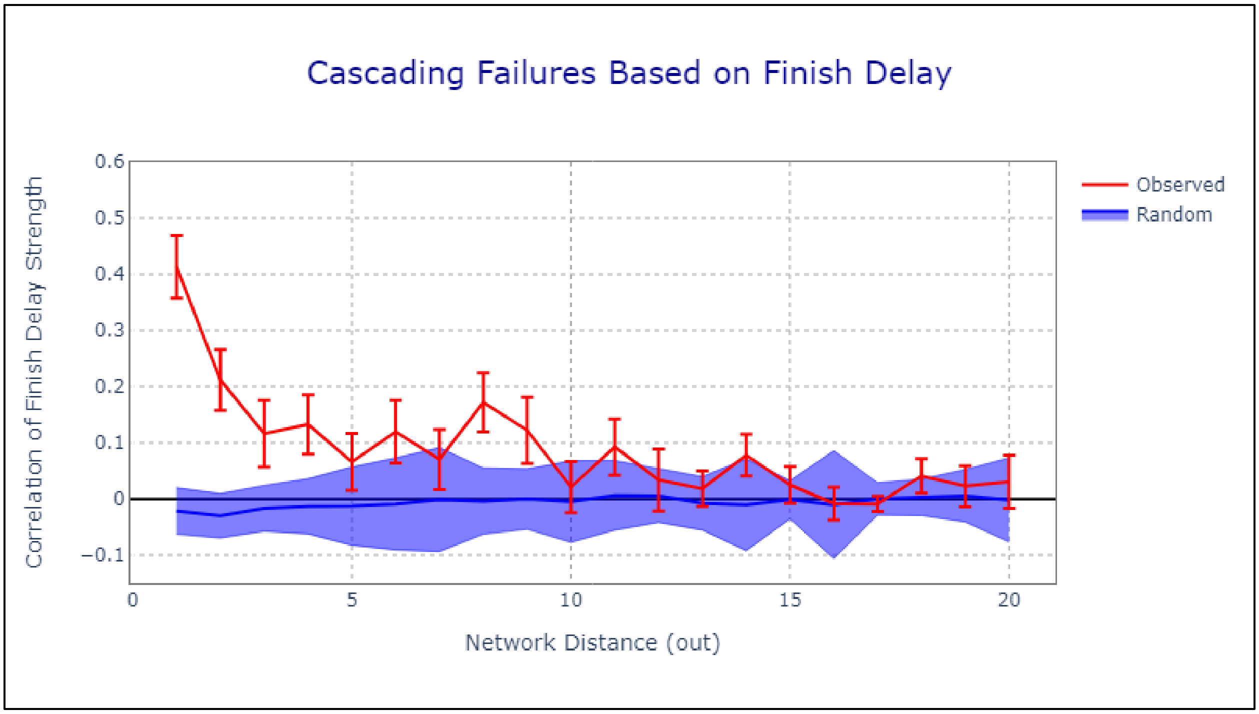

Figure 7: Cascading Failures Based on Finish Delay

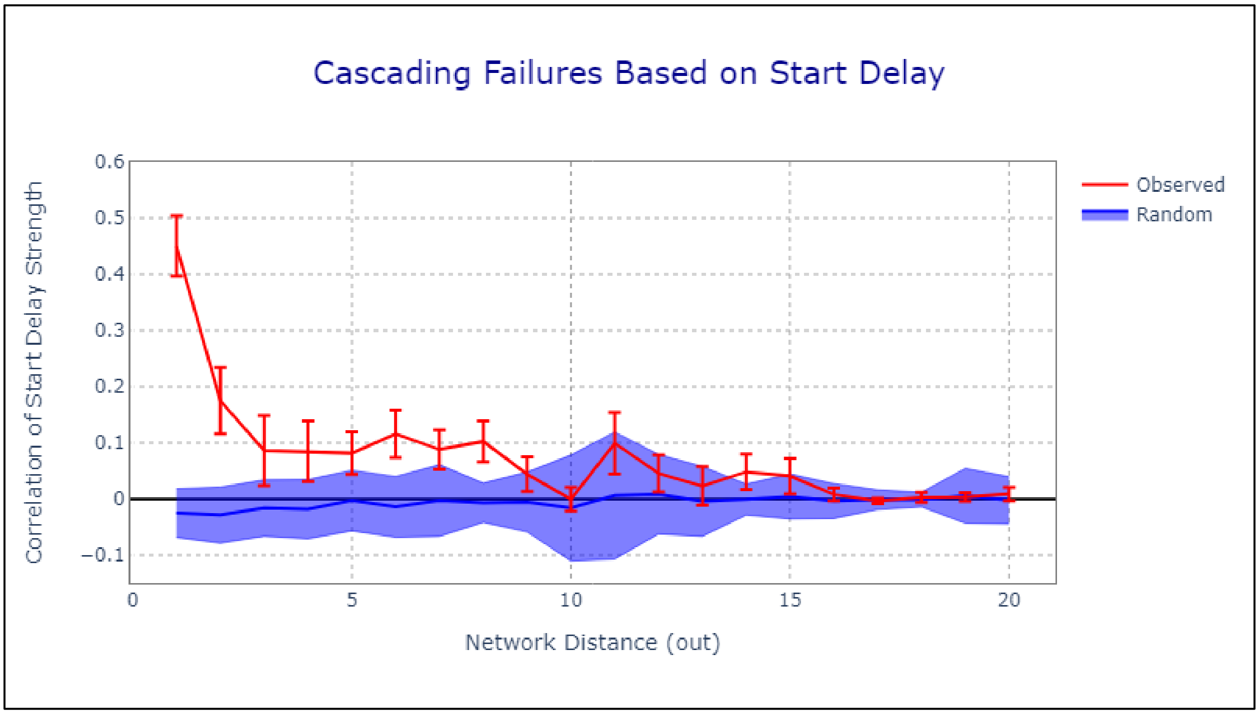

Figure 8: Cascading Failures Based on Start Delay

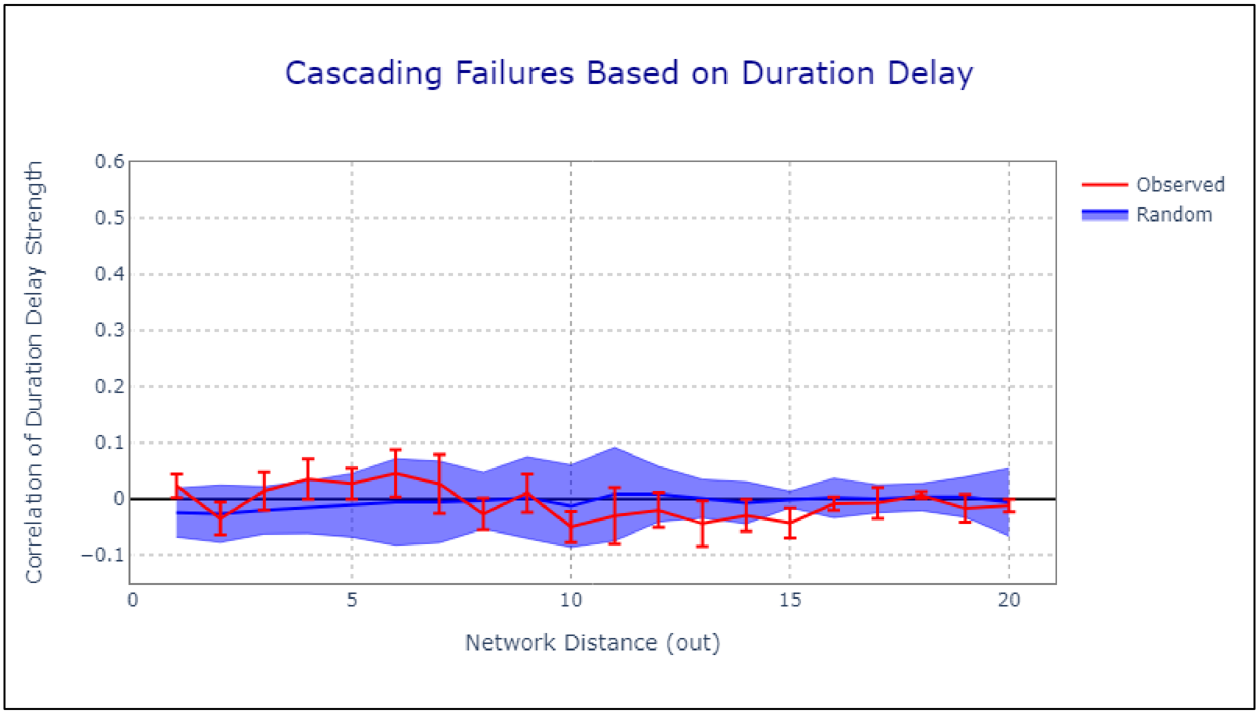

Figure 9: Cascading Failures Based on Duration Delay

The exploratory analysis verified that the MDAP milestone schedule hold complex network characteristics and that the centrality measures, node reach and node degree, are appropriate for examining the potential cascading impacts of fragile tasks. To conduct the formal analysis, we determined to employ cross-correlation and Spearman's rank correlation.

- A cross-correlation was executed between the aforementioned centrality measures and various measures of delay duration. Santolini et al. (2021, p. 7) used a delay rate, which is measured as the proportion of activities where the actual task duration exceeded the baseline task duration. Schedule analysts for DoD acquisition programs, however, are typically more interested in potential time delay of a project overall, than the number of tasks that were delayed. A project, for example, may experience a 75% delay rate, but only finish a few days behind schedule. Similarly, a project may have a 10% delay rate, but the 10% of the tasks that slipped were at high risk of delay propagation, leading to a significant duration overrun. Therefore, the study team used start delay and finish delay as schedule setback measures.
- The Spearman's rank correlation coefficient, which is a nonparametric measurement, was executed due to the non-normality of the data. Coefficients were calculated for approximately 25 of the network representatives, completed sub-milestones (Table 1). The results in the table are ordered by the severity of schedule slip as measured by finish delay from least to greatest delay. The left three columns are for the model with real data, while the right three columns represent a null model where task delays are randomly assigned via a simulation with a uniform distribution. While the null model correlation results are near zero and do not possess a pattern, the model with real data possesses larger values with patterns.

Table 1: Correlation of Finish Delay and Centrality Measurements

Relationship Between Finish Delay Strength and Centrality Metric Sorted by Finish Delay Rate

Delay Rate	Observed				Null Model		
	reach	in-degree	out-degree		reach	in-degree	out-degree
8.6%	-0.226	0.181	0.019	Sub-Milestone 7	-0.003	-0.001	0.010
20.0%	-0.053	0.016	0.122	Sub-Milestone 19	0.016	0.009	0.036
27.6%	0.007	0.006	-0.098	Sub-Milestone 8	-0.006	0.003	0.001
37.5%	0.075	-0.094	-0.051	Sub-Milestone 10	-0.005	0.031	-0.050
40.7%	-0.360	0.026	-0.232	Sub-Milestone 16	-0.019	0.016	-0.001
45.5%	-0.264	-0.028	-0.264	Sub-Milestone 15	0.028	-0.016	0.030
47.7%	-0.294	0.082	-0.043	Sub-Milestone 18	0.009	0.009	0.003
50.0%	-0.279	-0.161	-0.092	Sub-Milestone 25	0.031	-0.005	-0.004
50.0%	-0.348	-0.067	-0.039	Sub-Milestone 20	0.008	0.006	0.012
54.5%	-0.149	0.181	-0.021	Sub-Milestone 5	0.003	0.006	0.004
57.7%	-0.373	0.042	-0.563	Sub-Milestone 13	-0.002	-0.011	0.005
58.8%	-0.452	0.201	-0.120	Sub-Milestone 1	0.043	-0.010	-0.002
60.0%	-0.543	0.219	-0.141	Sub-Milestone 23	-0.006	-0.016	-0.017
60.3%	-0.503	-0.133	-0.319	Sub-Milestone 3	-0.001	0.017	-0.005
60.8%	-0.233	-0.082	-0.250	Sub-Milestone 2	-0.019	0.010	-0.016
61.9%	-0.701	0.288	-0.231	Sub-Milestone 17	0.024	0.008	-0.014
62.1%	-0.456	-0.016	-0.327	Sub-Milestone 4	0.000	-0.032	0.000
63.3%	-0.187	0.114	-0.328	Sub-Milestone 22	-0.005	0.011	-0.020
66.7%	-0.143	0.335	-0.445	Sub-Milestone 12	-0.010	0.014	-0.049
72.0%	-0.549	0.096	-0.338	Sub-Milestone 9	0.010	-0.008	0.049
78.4%	-0.388	0.145	-0.504	Sub-Milestone 21	0.017	0.044	-0.014
78.6%	-0.709	0.202	-0.202	Sub-Milestone 11	-0.001	0.004	0.005
78.6%	-0.542	0.272	-0.220	Sub-Milestone 24	-0.041	0.013	-0.021
83.3%	-0.763	0.376	-0.081	Sub-Milestone 6	0.012	0.001	0.008
95.2%	-0.683	0.530	-0.124	Sub-Milestone 14	0.019	-0.043	-0.003

1 0.5 0 -0.5 -1

The RH metric is increasingly negatively correlated with finish delay. These results are the opposite of the Santolini et al. (2021) observations, as their reachability metric was positively associated with schedule delay. Here, the lower network diameters for the sub-milestones relative to the exemplar study may explain these results. The longest diameter, which is the longest path, for the MDAP IMS sub-milestone networks is approximately 50 tasks, while the minimum longest path for Santolini et al. (2021) is roughly 30 tasks. If the MDAP had more completed milestones, the node reach directionality and significance may change.

Similarly, the results for the relationship between node degree and schedule delays are inconsistent for the two efforts. Santolini et al. (2021) did not find a clear pattern; however, node degree is increasingly and positively associated with finish delays in this case.

The findings for node degree are important because they conform with our schedule analysis discussion earlier in this study. The existence of strong task convergence in a project, where multiple, parallel activities must be completed before a successor task may start, leads to increased merge bias. Merge points can include important program activities, like critical design reviews or the start of production (GAO 2015, pp. 100-102). High in-degree valuation is the networking equivalent of strong task convergence. The analysis conducted by this study team corroborates the perils of merge bias because sub-milestones with higher in-degree (e.g., higher merge bias) are, in fact, associated with longer schedule delays.

The networking analysis is ultimately used to forecast schedule risk given the positive and increasing correlation between degree and finish delay. The study team calculated centrality measures for future tasks that were in the 18-month planning time horizon. Though degree is the most statistically significant measure for the program, multiple centrality measures are estimated for completeness. The table below (Table 2) lists future sub-milestones ordered by an aggregate rank, which is based on the selected centrality measures and the average ranking of the task. The aggregate rank is defined as the combination of multiple rank indices via averaging. The rank indices are computed using all available measures, such as network properties, and centrality measures; specifically, in-degree, out-degree, PageRank, closeness, betweenness, and reachability metrics. The highest rankings are displayed at the top of the table. Simply put, the tasks are ordered in descending schedule risk.

Table 2 – Future Sub-Milestone Fragility Forecasts

Name	Contractor Critical	WBS ID	Aggregate Rank	In-Degree Rank	Out-Degree Rank
Future Sub-Milestone 11	True	1.02.XX.XX.XX.XX	1	1	1
Future Sub-Milestone 2	True	1.02.XX.XX.XX	4	2	6
Future Sub-Milestone 16	True	1.02.XX.XX.XX	6.5	6	7
Future Sub-Milestone 10	True	1.02.XX.XX.XX	9	4	14
Future Sub-Milestone 14	True	1.04.XX.XX.XX.XX	18.5	33	4
Future Sub-Milestone 7	False	1.02.XX.XX.XX.XX.XX	28	28	28
Future Sub-Milestone 4	True	1.02.XX.XX.XX	29.5	11	48
Future Sub-Milestone 3	True	1.02.XX.XX.XX	30	13	47
Future Sub-Milestone 9	True	1.02.XX.XX.XX	34.5	20	49
Future Sub-Milestone 12	True	1.04.XX.XX.XX.XX	39	45	33
Future Sub-Milestone 5	True	1.02.XX.XX.XX.XX.XX	40	53	27
Future Sub-Milestone 6	False	1.02.XX.XX.XX	40.5	7	74
Future Sub-Milestone 15	True	1.08.XX.XX.XX.XX	42	17	67
Future Sub-Milestone 13	True	1.02.XX.XX.XX.XX.XX	42	79	5
Future Sub-Milestone 17	True	1.02.XX.XX.XX	45	18	72
Future Sub-Milestone 1	True	1.07.XX.XX.XX.XX.XX	47.5	22	73
Future Sub-Milestone 8	True	1.02.XX.XX.XX.XX.XX	47.5	78	17

In this case, the table displays the aggregate ranking based on in-degree and out-degree. Future Sub-Milestone 2, for example, possesses the second highest score based on the in-degree measure and the sixth highest score based on the out-degree measure. Additional features can help inform the analyst when reviewing the networking forecasts. Data pertaining to whether the task is on the contractor's critical path or the contractor's driving path, both available in IPMDAR's TaskScheduleData table (i.e., OnCriticalPath and OnDrivingPath), can be appended and displayed. Cost and schedule analysts may be particularly interested in high centrality ranking tasks that are also on the critical path because they are identified as significant risks along multiple margins. Likewise, high centrality ranking tasks that are not currently identified as critical may warrant further investigation to determine if the task necessitates monitoring. Overall, the intent of the forecasts is to support the identification of potentially high-risk activities for the execution of root cause analyses so that preventive action may buy down the risk.

Section 5: Discussion and Future Work

The study team encountered several challenges and limitations while conducting this research effort, one of which is data availability. The analysis was conducted on a single, in-progress project. Future efforts should seek to apply comparable cascading network methods to completed or near-finalized projects to independently verify that node degree, and possibly node reach, are associated with schedule fragility.

The forecasts are also currently limited in both their analytic rigor and value. The methodological approach can identify future tasks that are potentially fragile based on their node degree values. However, the forecasts do not currently account for certain edge characteristics that are pertinent to EVM analysts. As an example, the type of task relationship, like start-to-start and finish-to-finish, is not accounted for because most tasks possess a finish-to-start relationship. Nonetheless, the added granularity could improve the forecast by modifying the strength of node relationships in the network.

Perhaps of greater importance, the forecasts do not quantify what the potential impact of these fragile tasks are. In other words, the current method predicts the likelihood of a schedule risk, but does not capture the schedule consequences. Future research may adopt mature analytic approaches from other industries to improve upon these limitations, as demonstrated here with this schedule network analysis pathfinder. The financial industry, for example, has developed frameworks to stress test the resilience of the financial system (Amini et al. 2012). The DoD acquisition community may look to adapt these methods to examine how and if a shock, such as a time delay, to a fragile task causes a "contagion" across the entirety of a project (Levy-Carciente 2015: 164).

Despite the limitations, this effort is potentially groundbreaking because it applies emergent research on project schedule networking analysis to a real IMS for a DoD MDAP. In addition, the approach, adapted to an ongoing program at the sub-milestone level, replicates some of the existing research hypotheses, specifically the importance of node degree on task delays. Moreover, the analysis corroborates current project schedule postulates regarding merge biases. Tasks with a greater in-degree are leading to cascading failures for the completed portions of this MDAP's IMS. The incorporation of both standard and adapted centrality measurements that correlate with schedule fragility may be readily used to investigate high cascade probability tasks on other MDAPs.

Furthermore, this study modified several of Pozzana et al. (2021) and Santolini et al. (2021) exemplar analysis to accommodate the realities of DoD acquisition. In particular, the use of start and finish delay rather than delay rate and duration delays may serve as more applicable variables of interest to the EVM community because they capture the manifestation of cascading delays, which can thwart DoD MDAP success. In addition, the analysis is executed using predominant tools and libraries in the Python language, which provides strong analytic rigor and high interoperability with many other computing tools through application program interfaces (APIs). Finally, this effort is a joint collaboration of multiple organizations, which culminated in substantial synergies across several sub-disciplines of expertise.

In closing, information technology growth has led to an ever-increasing supply of digital data, otherwise known as the lifeblood of data sciences, for our costing community to request, collect, normalize, store, and report upon. The proliferation of data will likely continue to increase the complexity of acquisition integrated program baselines (IPBs) and IMSs. It is our hope that this effort incentivizes others researchers and practitioners to build-upon the networking analysis research conducted here, so that we can help programs to develop and field critical capabilities on time and at cost.

References

Alexander, S. (2002). Earned Value Management Systems (EVMS).

Amini, H., Cont, R., & Minca, A. (2012). Stress testing the resilience of financial networks. *International Journal of Theoretical and applied finance*, *15*(01), 1250006.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science*, *323*(5916), 892-895.

Brandes, U. (2005). *Network analysis: methodological foundations* (Vol. 3418). Springer Science & Business Media.

Ellinas, C., Avraam, D., & Nicolaides, C. (2023). Neglecting complex network structures underestimates delays in a large-capital project. *Journal of Physics: Complexity*, *4*(2), 02LT01.

Estrada E (2010) Quantifying network heterogeneity. Phys Rev E 82(6):066102

GAO - United States Governance Accountability Office. (2015). *GAO-16-89G Schedule Assessment Guide*.

Geraldi, J., Maylor, H., & Williams, T. (2011). Now, let's make it really complex (complicated) A systematic review of the complexities of projects. *International journal of operations & production management*, *31*(9), 966-990.

Hofbauer, J., Sanders, G., Ellman, J., Morrow, D., Berteau, D., Ben-Ari, G., & Lombardo, N. (2011). *Cost and time overruns for major defense acquisition programs*. Acquisition Research Program.

Jacob, R., Harikrishnan, K. P., Misra, R., & Ambika, G. (2017). Measure for degree heterogeneity in complex networks and its application to recurrence network analysis. *Royal Society open science*, *4*(1), 160757. https://doi.org/10.1098/rsos.160757

Johnson, J. C. (2018, May 1). Cost overruns schedule slips the norm not the exception. https://www.nationaldefensemagazine.org/articles/2018/5/1/cost-overruns-schedule-slips-the-norm-not-the-exception

Levy-Carciente, S., Kenett, D. Y., Avakian, A., Stanley, H. E., & Havlin, S. (2015). Dynamical macroprudential stress testing using network theory. *Journal of Banking & Finance*, 59, 164-181.

Pozzana, Iacopo, et al. "Spreading of performance fluctuations on real-world project networks." *Applied Network Science* 6.1 (2021): 25.

Santolini, M., Ellinas, C., & Nicolaides, C. (2021). Uncovering the fragility of large-scale engineering projects. *EPJ Data Science*, *10*(1), 36.

Vazquez, A., Pozzana, I., Kalogridis, G., & Ellinas, C. (2023). Activity networks determine project performance. Scientific Reports, 13(1), 509.

Zello, N., Duvall, K., & ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY ABERDEEN PROVING GROUND MD. (2018). AMSAA Schedule Risk Guidebook.