

Modeling Evolutionary Architectural Growth in Major Defense Acquisition Programs

Matthew Dabkowski

Arthur Middlebrooks

Ricardo Valerdi

Abstract: One of the critical artifacts in the system design process is the architecture. Represented in a variety of ways, architectures evolve over time as user needs evolve and design limits are reached. The emergent requirement to submit Department of Defense Architecture Framework (DoDAF) models prior to Milestone A has created opportunities for the early life cycle cost estimation of major defense acquisition programs. Although this development could be seen as positive, any cost estimation procedure ultimately needs to be informed and validated by real-world data. In this paper, we examine the assumptions underlying one such method – an algorithm to estimate the cost of unanticipated, evolutionary architectural growth via the SV-3 (or Systems-Systems Matrix) and the Constructive Systems Engineering Cost Model. Specifically, using SV-3s from 24 different defense programs, we found that while the data generally comport with the method, alternative growth and connection mechanisms are needed. As such, we propose a modified, data-driven approach using number theory, network science, simulation, and statistical analysis – one which not only improves the algorithm’s fidelity but also reinforces the value of viewing DoDAF models as computational objects.

Introduction and Context

The Budget Control Act (BCA) initiated substantial budget cuts to reduce spending within the federal government (2011). Inside the Department of Defense (DoD), sequestration – the unofficial title given to the BCA – drastically reduced spending for military operations and defense acquisitions. In the years that followed, various pieces of legislation were passed to modernize and streamline the defense acquisition process to facilitate the reduction in defense spending. Much of this legislation focused on increasing the useable lifecycle of defense systems and reducing development costs; however, these changes alone were not substantial enough to create meaningful and lasting change. To address this gap and other modernization efforts, the National Defense

Strategy (NDS) was updated for the first time in over ten years.

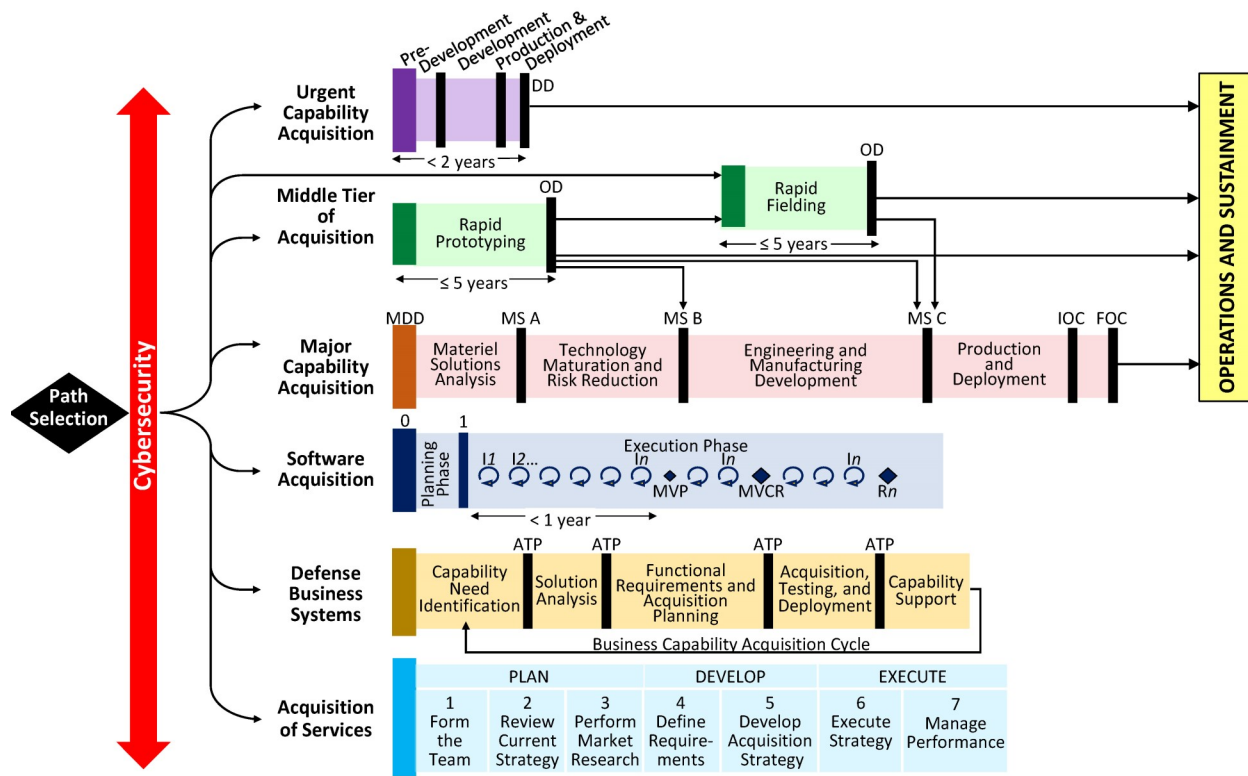
As a strategic document, the NDS “provides a clear road map for the [DoD] to meet the challenges posed by a re-emergence of long-term strategic competition with China and Russia” (DoD, n.d.). To this end, the NDS outlines several strategic objectives, including “sustaining Joint Force military advantages, both globally and in key regions, . . . defending allies from military aggression and bolstering partners against coercion, fairly sharing responsibilities for common defense; . . . [and] continuously delivering performance with affordability and speed” (DoD, 2018, p. 4). With respect to the equipment service members require to execute their global mission, the latter objective cannot be achieved without changes to how the DoD approaches the design, development, and

acquisition of military technology. In a budget-constrained environment, this imperative becomes increasingly more important.

Notably, even prior to the initiation of sequestration, researchers found that most of a program’s life cycle costs are often committed early in the system design process without knowledge of future requirements (e.g., Blanchard & Fabrycky, 1998, p. 37; Dowlatshahi, 1992, p. 1803). Put simply, the architectural choices made by systems engineers today have profound implications for a system’s total cost at retirement. The inability to forecast this structural dependence reduces the DoD’s ability to be responsive and flexible to future capability gaps. Unfortunately, despite the opportunity to affect substantive change on the resulting cost of the system early in its life cycle, it is difficult to

anticipate how requirements will change over time. For example, researchers found that more than 10% of a system’s baseline requirements will change during the development phase of a system’s life cycle (Peña & Valerdi, 2015, pp. 63-65). Recognizing this uncertainty and its implications, the NDS identified “reforming the Department’s business practices for greater performance and affordability” as one of its key lines of effort (DoD, 2018, p. 5).

Within the performance and affordability line of effort, the NDS outlines five competitive approaches to improving the DoD’s practices, namely: “[1] deliver performance at the speed of relevance . . . [2] organize for innovation . . . [3] drive budget discipline and affordability to achieve solvency . . . [4] streamline rapid, iterative approaches from development to fielding . . . [5]



Legend:

ATP: Authority to Proceed	DD: Disposition Decision	FOC: Full Operational Capability
I: Iteration	IOC: Initial Operational Capability	MDD: Materiel Development Decision
MS: Milestone	MVCR: Minimum Viable Capability Release	MVP: Minimum Viable Product
OD: Outcome Determination	R: Release	

Figure 1. AAF Pathways (USD(A&S), 2020b, p. 5.)

harness and protect the Nation Security Innovation Base” (DoD, 2018, pp. 10-11). While each of these approaches is worthy of intense pursuit, particularly streamlining development and fielding processes, the appropriate mechanisms must be in place for the right capabilities to reach service members at the right time. Recognizing this imperative, two years following the 2018 update to the NDS, the DoD issued *DoD Instruction (DoDI) 5000.02 – Operation of the Adaptive Acquisition Framework (AAF)*.

With the NDS’s priorities and focus on improving the DoD’s ability to deliver improved technical solutions at a reduced cost, “[t]he AAF supports the [Defense Acquisition System] with the objective of delivering effective, suitable, survivable, sustainable, and affordable solutions to the end user in a timely manner” (USD(A&S), 2020a, p. 3). Figure 1 depicts the AAF’s six pathways, each of which is tailored to the desired capability gap. Within the scope of this research, we focus specifically on Major Capability Acquisition efforts, which consist of an initial Material Development Decision (MDD), followed by a series of critical milestones (MS).

Following the MDD, but prior to MS A, there are several key considerations with respect to the system itself, including “technical, cost and schedule risks, and the plans and funding to offset them during the [Technology Maturation and Risk Reduction] TMRR phase” (USD(A&S), 2020b, p. 12). This involves the designated PM conducting an analysis of the “Should Cost” targets, which are directly tied to the requirements specified in the Initial Capabilities Document. These targets are critical, as they establish the foundation for executing the final Request for Proposals and the cost for incorrectly doing so are high. For example, a RAND Project Air Force study found that across 35 mature programs unstable requirements accounted for a 12.9% increase in

total costs (Bolten, Leonard, Arena, Younossi, & Sollinger, 2008, p. 72). In FY2005 dollars, this translated to a \$23.7 billion increase. Additionally, between 1997 and 2009, the majority of Nunn-McCurdy Breaches (significant cost overruns that must be statutorily reported to Congress) cited engineering/design issues and requirement changes as significant factors contributing to their programs’ unexcepted, excessive cost growth (GAO, 2011, p. 5).

Taken together, these findings support the need to improve two dimensions of the system design and acquisition process with respect to cost reduction: requirements development and extending the useable life of systems. The AAF seeks to address the first issue; however, as specified by the NDS, these solutions must not only address today’s capability gaps but also be capable of mitigating those in future environments. Additionally, they must do so in a cost-effective manner. In recognition of this necessity, the recently published *DoDI 5000.88 – Engineering of Defense Systems* mandates that all Major Defense Acquisition Programs (MDAP) require a formalized Systems Engineering Plan (SEP) (USD (R&E), 2020b, p. 12). Moreover, within the SEP and specific to extending the useable life of systems, DoDI 5000.88 directed the use a Modular Open Systems Approach (MOSA).

As a framework for addressing capability gaps, MOSA “provides an integrated business and technical strategy for competitive and affordable

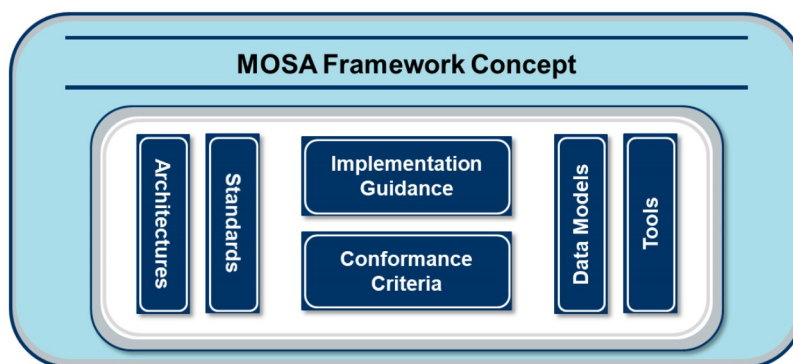


Figure 2. MOSA Framework (USD(R&E), 2020a, p. 4).

acquisition and sustainment of a new or legacy system (or a component within a new or legacy system) over the system life cycle” (USD(R&E), 2020a, p. 2). As seen in Figure 2, this framework includes six elements. Specific to this research, architecture is a representation of the “fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution” (ISO, 2011, p. 2).

Within the AAF, during Material Solutions Analysis and prior to MS A, MOSA forces system designers to consider not only the first instantiation of the system but also how it will evolve over time. In doing so, they must consider and communicate the system’s architecture, specifying how components interface with one another to better facilitate future changes or upgrades. In support of MOSA, the DoD utilizes the DoD Architecture Framework (DoDAF), a series of viewpoints and models, to communicate information about the system. The requirement to develop these models Pre-MS A not only supports the NDS and AAF, but more practically, provides analysts with rich data regarding the components and interfaces of the system.

For example, the DoDAF’s Systems Viewpoint 3 (SV-3 or Systems-Systems Matrix) “provides a tabular summary of the system interactions” (DoD DCIO, 2010, p. 209). Within the context of system requirements and extending the useable life of the system, the SV-3 depicts relationships between the system components that execute the system’s functions, providing the analyst with critical information that can be used to explore options for evolving the system over time. This ensures that while we may not be capable of predicting the future operational environment, at the earliest point in the systems engineering process, we can be more reasonably assured that the system is designed to remain responsive to external changes in a budget-constrained environment.

Given the NDS’s focus on improving the defense acquisition process for a rapidly changing

operational environment, the dire implications of incorrectly estimating and managing system costs, and the adoption of the AAF and MOSA, there exists a tremendous responsibility and opportunity to improve early life-cycle cost estimates.

Exploiting an Untapped Source of Early Life Cycle Information – Mapping DoDAF’s Models to COSYSMO’s Drivers

In a recent article, Valerdi, Dabkowski, and Dixit (2015) exploited the Pre-MS A availability of DoDAF’s models by mapping them to the drivers of the Constructive Systems Engineering Cost Model (COSYSMO), a parametric, open academic cost model with the following cost estimating relationship (CER):

$$PM_{NS} = A \cdot \left(\underbrace{\sum_{i \in \{e,n,d\}} \sum_{k=1}^4 w_{ik} \Phi_{ik}}_{\text{size}} \right)^E \cdot \underbrace{\prod_{j=1}^{14} EM_j}_{\text{effort}} \quad (1)$$

where:

PM_{NS} = systems engineering effort in person months (nominal schedule),

A = calibration constant derived from historical project data (assume as 0.25),

w_{ik} = weight for the i^{th} complexity level of the k^{th} size driver ($i \in \{e \text{ (easy)}, n \text{ (nominal)}, d \text{ (difficult)}\}$),

Φ_{ik} = quantity of the k^{th} size driver with complexity level i ($k \in \{1 \text{ (requirements)}, 2 \text{ (interfaces)}, 3 \text{ (algorithms)}, 4 \text{ (operational scenarios)}\}$),

E = diseconomies of scale constant (assume as 1.06), and

EM_j = systems engineering effort multiplier for the j^{th} cost driver (assume $\prod_{j=1}^{14} EM_j = 0.89$) (Valerdi, 2008, p. 34).

Specifically, using a blend of text mining and social network analysis techniques, they analyzed

DoDAF’s manual to identify subsets of models that, on a basis of doctrine, should contain useful information for populating COSYSMO’s 18 drivers. The results of this analysis are summarized graphically in Figure 3, where: (a) a link between DoDAF model X and COSYSMO driver Y indicates model X should be relevant for rating driver Y, (b) DoDAF models shaded yellow were required Pre-MS A between 2012 and 2015 (CJCS, 2012a; CJCS, 2012b), (c) DoDAF models highlighted with a black star (*) were required Pre-MS A between 2015 and 2018 (CJCS, 2015), (d) DoDAF models with their names highlighted in green are currently required pre-MS A (CJCS, 2018), and (e) COSYSMO drivers shaded red are not linked to any of the DoDAF models required by the Capability Development Document (CDD).

As Figure 3 shows, the collection of DoDAF models required Pre-MS A nearly spans COSYSMO’s parameters, as they cover 14 of the 18 drivers. Additionally, several DoDAF models that are not explicitly required by the CDD are derivatives of data contained within the mandatory models, suggesting even greater coverage is possible. For example, the SV-3 is simply a more compact, summary representation of the interfaces described in the SV-1 (DoD DCIO, 2010, p. 209).

Exploiting this untapped, relevant source of early life cycle information, Valerdi et al. (2015) provide an algorithm for organizations to improve the measurement reliability of their MDAP cost estimates using COSYSMO (pp. 543-545).

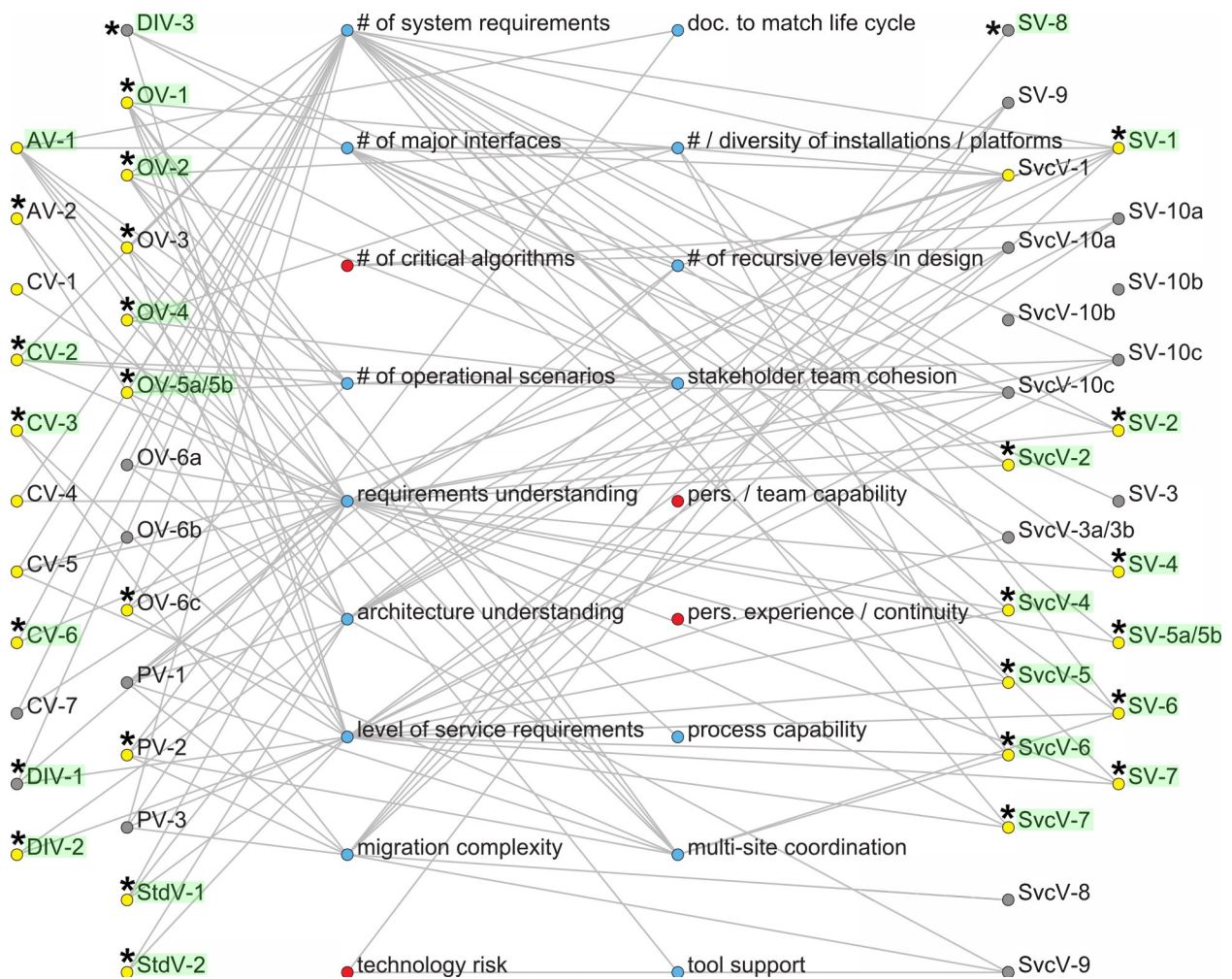


Figure 3. Mapping of DoDAF’s models to COSYSMO’s drivers.

Additionally, there are other tangential benefits stemming from their work, including opportunities for “data mining techniques . . . to extract knowledge from existing systems engineering concepts” (Valerdi et al., 2015, pp. 546). Given the rapid proliferation of such techniques and their application to new domains, algorithms employing these methods should be accompanied by a detailed explanation and assessment of any underlying assumptions. Consistent with this assertion, in this paper we methodically examine one such algorithm, namely, a method for estimating the cost of unanticipated, evolutionary architectural growth via the SV-3, COSYSMO, and network science (Dabkowski, Valerdi, & Farr, 2014).

Estimating the Cost of Architectural Growth Early in the Life Cycle

In “Exploiting Architectural Communities in Early Life Cycle Cost Estimation,” Dabkowski et al. (2014) estimate the cost of adding a future subsystem (\mathbf{X}) to an existing system architecture when the purpose and function of \mathbf{X} are unknown. Consisting of 12 steps and seen in Algorithm 1 below, their approach leverages the current SV-3’s structural properties to iteratively and randomly attach \mathbf{X} to architectural modules (or communities), generate interfaces of reasonable complexity, and estimate the marginal cost of attachment.

Algorithm 1 (Dabkowski et al., 2014, p. 101)

For a specified, suitably large number of iterations . . .

Preprocessing

- (1) Initialize the system as the current system,
- (2) Use the Girvan-Newman (2002) community detection heuristic to identify architectural communities,
- (3) Randomly assign \mathbf{X} to community k ,

Intracommunity Growth

- (4) Generate a realization for $M_{\mathbf{X},\text{intra}}$ given \mathbf{X} is assigned to community k (m_{intra}),
- (5) Connect \mathbf{X} to m_{intra} subsystems inside community k using the Barabási-Albert (BA) model (1999),
- (6) For each interface established in (5), assign complexity ($w_{i\mathbf{X},\text{intra}}$),

Intercommunity Growth

- (7) Generate a realization for $M_{\mathbf{X},\text{inter}}$ given \mathbf{U} is assigned to community k (m_{inter}),
- (8) Connect \mathbf{U} to m_{inter} communities using the BA model,
- (9) For each interface established in (8), assign complexity ($w_{i\mathbf{X},\text{inter}}$),

Cost Estimation

- (10) Estimate the cost for the augmented system using COSYSMO (PM_{NS^*}),
- (11) Calculate the additional cost of adding subsystem \mathbf{U} ($PM_{NS^*} - PM_{NS}$), and
- (12) Store results and return to (3).

For example and without loss of generality, imagine the hypothetical SV-3 in Panel (a) of Figure 4 serves as the **current system** in Step (1) of Algorithm 1. With 8 subsystems and 12 interfaces, this SV-3 graphically summarizes the relationships between existing subsystems, where shading in row i and column j implies subsystem i interfaces with subsystem j , and darker shades indicate greater interface complexity. Moving to Step (2), given the emphasis on MOSA within the DoD (USD(R&E), 2020b), we suspect that the current system’s architecture might contain **communities of subsystems** where the density of interfaces within communities is high relative to the density of interfaces between them. To identify and exploit this structure, we apply the popular Girvan-Newman (2002) community detection heuristic, which yields the permuted, isomorphic SV-3 seen in Panel (b) of Figure 4. In Step (3), subsystem \mathbf{X} is randomly assigned to one the current system’s two architectural communities, namely C1 or C2.

Continuing with the example, suppose subsystem **X** is assigned to C1. Using a rich-by-birth effect, the number of subsystems **X** interfaces with in C1 is modeled as a discrete random variable ($M_{X,intra}$) with a probability mass function (PMF) equal to the observed intracommunity degree distribution of C1's subsystems. By inspection, C1's PMF for $M_{X,intra}$ is $P(M_{X,intra} = 2) = 0.5$ and $P(M_{X,intra} = 3) = 0.5$. As per Step (4), we generate a realization (m_{intra}) from this PMF to represent the number of **X**'s intracommunity interfaces.

Assuming $m_{intra} = 2$, we connect **X** to 2 subsystems in C1 using the BA model from network science, which mimics a rich-get-richer effect. Observed in a variety of real-world networks (e.g., Redner, 1998; Banavar, Maritan, & Rinaldo, 1999; Newman, 2001; Cancho, Janssen, & Solé, 2001), the BA model specifies that the probability of **X** attaching to an existing subsystem *i* in C1 is proportional to *i*'s number of intracommunity interfaces (Barabási & Albert, 1999). In Panel (b) of Figure 4, C1's subsystems **A**, **D**, **F**, and **H** have degrees 2, 3, 3, and 2, respectively. Accordingly, their corresponding attachment probabilities are 0.2, 0.3, 0.3, and 0.2, and in Step (5) we use these probabilities to determine **X**'s adjacency.

If Step (5) determines that **X** connects to subsystem **A**, then in Step (6) we use the observed interface complexity distribution of subsystem **A** as the PMF for the complexity of the interface between **X** and subsystem **A** (w_{AX}). Specifically, one of **A**'s existing interfaces is rated **easy**, and the other is rated **nominal**. Accordingly, we apply the relative weights for COSYSMO's **number of major interfaces** size driver (Valerdi, 2008, p. 86) to obtain the following PMF for w_{AX} : $P(w_{AX} = 1.1) = 0.5$ and $P(w_{AX} = 2.8) = 0.5$, and a Bernoulli trial determines the outcome.

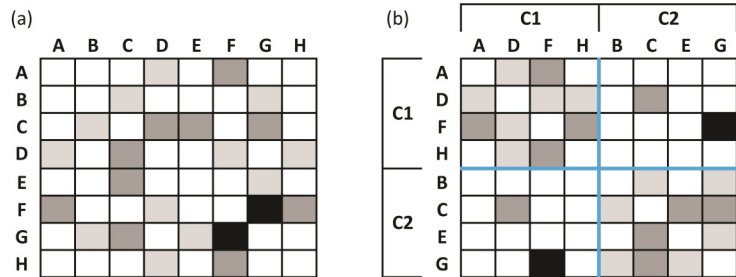


Figure 4. Hypothetical SV-3 with 8 subsystems and 12 interfaces in its original (Panel (a)) and permuted (Panel (b)) forms, where shading indicates interface complexity such that light gray ⇒ easy, medium gray ⇒ nominal, and black ⇒ difficult.

Steps (7) through (9) model intercommunity growth in a manner similar to Steps (4) through (6), and Steps (10) and (11) apply COSYSMO's CER to calculate the marginal cost of attachment. Finally, Step (12) completes the loop and captures the result. Subsequent iteration generates the data necessary to estimate the cumulative distribution functions and statistics for the estimated cost of adding subsystem **X** to C1 and C2.

Algorithm 1's Underlying Assumptions and Associated Hypotheses

Although Algorithm 1 has intuitive appeal, the validity of its underlying assumptions must be assessed. Starting with its principal input, we note that an SV-3 is nothing more than the adjacency matrix representation of a network, where nodes symbolize subsystems and edges denote interfaces. As such, if Dabkowski et al. (2014) have designed the right algorithm (i.e., Algorithm 1 is valid), the networks represented by real-world SV-3s will be similar to the hypothetical example in Figure 4.

Methodologically, network similarity is typically assessed by comparing a list of structural statistics (Shore & Lubin, 2015), but selecting which statistics to compare is a matter of debate. With this in mind, study objectives typically drive the selection (e.g., Hunter, Goodreau, & Handcock, 2008), and our primary goal is to determine

whether Algorithm 1 can accommodate real-world SV-3s. Therefore, we restrict our focus to three characteristics, namely: type, density, and community structure.

With respect to type, we note that the hypothetical example reflects the interfaces that exist between a single set of subsystems. Specifically, the subsystems represented by the rows match those identified in the columns. In the parlance of network analysis, this is known as a **one-mode network** (Wasserman & Faust, 2009, p. 36-37). Moreover, upon closer examination, we note that the hypothetical example is symmetric across its main diagonal, implying the interfaces are bidirectional; the network is **undirected**. Finally, as mentioned earlier, the hypothetical example's cells are shaded according to interface complexities, which ultimately map to numerical weights in COSYSMO's CER. Accordingly, the network is **valued**. Taken together, Algorithm 1 accepts **one-mode, undirected, valued SV-3s** as input, and our first hypothesis is as follows: (H1) – Real-world SV-3s are one-mode, undirected, valued networks.

Moving on to density, Algorithm 1 makes use of the Girvan-Newman community detection heuristic, which “was designed with sparse networks in mind, . . . [and] may not perform as well on dense networks” (2002, p. 7826). Formulaically, the sparsity of a one-mode, undirected network with N nodes can be assessed by its density, which is simply the ratio of its observed number of edges (E) to the maximum possible number of edges or $2E / (N(N - 1))$ (Wasserman & Faust, 2009, p. 101). Although there is no definitive standard for characterizing a network as sparse, informally, the adjacency matrix of a sparse network consists primarily of zeros. Adopting this as our standard, this implies that a sparse network will have a density less than 0.5. For example, with $N = 8$ subsystems and $E = 12$ interfaces the hypothetical example in Figure 4 has a density of 0.428. As such, it is sparse, and, as expected, the Girvan-Newman

community detection heuristic performs well, correctly identifying its two architectural communities. In sum, by using the Girvan-Newman community detection heuristic in Algorithm 1, Dabkowski et al. (2014) implicitly assume real-world SV-3s are sparse, and this leads us to our second hypothesis: (H2) – The densities of real-world SV-3s are less than 0.5.

As for community structure, recall that the hypothetical example in Figure 4 consists of two, well-defined communities, and Algorithm 1 conforms to this structure. The questions are: “Does the community structure in the hypothetical example warrant this approach,” and “If so, do real-world SV-3s exhibit similar behavior?” To address these issues, we use Girvan and Newman's well-known modularity metric, which ranges from its minimum of 0, when “the number of within-community edges is no better than random,” to its theoretical maximum of 1, indicating very strong community structure (Newman & Girvan, 2004, p. 7). In practice, however, Girvan and Newman note modularities above 0.7 are uncommon, and values greater than 0.3 suggest strong community structure (Newman & Girvan, 2004, p. 7). Adopting this standard, we used the `edge.betweenness.community` function from the statistical software R's `igraph` package (Csardi & Nepusz, 2006) to calculate the modularity for the hypothetical example. With a value of 0.333, its modularity exceeds the 0.3 threshold, and Algorithm 1 rightly abides its strong community structure. Whether or not this also applies to real-world SV-3s needs to be assessed, yielding our third hypothesis: (H3) – The modularities of real-world SV-3s are greater than 0.3.

Beyond validity related to input, the validity of Algorithm 1's growth and connection mechanisms must also be assessed. Looking at the growth mechanism, Dabkowski et al. (2014) modeled an incoming subsystem's number of interfaces (or degree) as a random variable with a PMF equal to the observed degree distribution of the current system. In fact, their approach

differs substantially from the growth step in the standard BA model, where a fixed versus random number of edges (m) are added each time a new node enters the network (Barabási & Albert, 1999). That said, when m is constant the total number of edges in a network of N nodes is fixed, reducing the flexibility of the standard BA model. For example, if a network starts as the smallest possible connected graph prior to the addition of new nodes, it will have m nodes and $m - 1$ edges. In this case, following the network's eventual growth to N nodes, it will have $E = m(N - m) + (m - 1) = m(N - m + 1) - 1$ total edges. For the hypothetical example in Figure 4, this is problematic. Specifically, if $m = 1$, once the SV-3 grows to its final size of $N = 8$ subsystems it will have $E = 7$ interfaces. This is 5 too few; therefore m must be greater than 1. However, if $m = 2$, it will ultimately have $E = 13$ interfaces, which is 1 too many.

Accordingly, holding an incoming subsystem's number of interfaces constant is overly restrictive, and using the observed degree distribution in Algorithm 1 seems reasonable. As Dorogovtsev and Mendes (2003) note:

How can we account for the influence of the network on the properties on a newborn vertex? This, our baby, has only one characteristic, namely the number of its connections. Then, let this number not be fixed by God but be distributed with some distribution function which depends on the current state of the network . . . let the degree distribution of the newborn vertex be dependent on the degree distribution of the network (p. 42).

Nonetheless, as with the standard BA model, the validity of this approach must be assessed. Ultimately, we want a stochastic model capable of growing SV-3s with N subsystems and

approximately E interfaces. More precisely, if the mean absolute percentage error (MAPE) in interfaces is no more than 20%, we feel the model is sufficient, and our fourth hypothesis is as follows: (H4) – Using the observed degree distribution of a real-world SV-3 to model its incoming subsystems' number of interfaces grows SV-3s with a MAPE in interfaces less than or equal to 20%.

Finally, for Dabkowski et al.'s (2014) connection mechanism, in Steps (5) and (8) of Algorithm 1 they used linear preferential attachment to connect an incoming subsystem to subsystems already in the architecture. As such, if real-world SV-3s utilize linear preferential attachment, their connection mechanism is valid; otherwise, it is not. Unfortunately, proving the presence of linear preferential attachment in a network's evolution requires longitudinal data or snapshots of its growth over time (Newman, 2001; Jeong, Néda, & Barabási, 2002), and such data is often unavailable.

Nonetheless, even without longitudinal data, we can use a network's degree distribution to examine whether linear preferential attachment played a role in its evolution, as networks grown via linear preferential attachment have a characteristic statistical marker. Specifically, the probability that a node has degree d , $p(d)$, is proportional to $d^{-\omega}$ where $\omega = 3 - p$ and p represents the fraction of edges that are directed (Barabási & Albert, 1999). Moreover, while our earlier discussion demonstrated that m is not fixed and the standard BA model does not apply, this marker still holds. As Barabási and Albert note in their landmark 1999 paper: "For most networks, the connectivity m of the newly added vertices is not constant. However, choosing m randomly will not change the exponent" (p. 512). In short, if real-world SV-3s have been grown via linear preferential attachment, ω should be 3. Formally known as a power law, this scale-free behavior typically presents itself in the heavy, right-hand tail of the network's degree distribution, and following normalization, its

discrete distribution function is:

$$p(d) = \frac{d^{-\omega}}{\sum_{n=0}^{\infty} (n + d_{min})^{-\omega}} \quad (2)$$

where an approximate maximum likelihood estimator for ω is given by:

$$\hat{\omega} \approx 1 + n \left[\sum_{i=1}^n \ln \frac{d_i}{d_{min} - 0.5} \right]^{-1} \quad (3)$$

and:

d_{min} = the lower bound for the power law behavior (Clauset, Shalizi, & Newman, 2009).

Using these facts, we formulate our fifth and final hypothesis as follows: (H5) – The observed degree distributions of real-world SV-3s follow a power-law with $\hat{\omega} = 3$.

Results and Analysis of Hypothesis Testing

In order to test the hypotheses identified in the previous section, we used the SV-3s from 24 different defense programs. Consisting of weapons systems; combat and transportation vehicles; command, control, and communications suites; and intelligence, surveillance, and reconnaissance platforms, the programs vary considerably in both size and scope, providing a broad sample and facilitating generalization (AIMD, 2014). The results and analysis of this testing are given below.

(H1) – Real-world SV-3s are one-mode, undirected, valued networks.

As seen in the first several columns of Table 1, real-world SV-3s do not match Algorithm 1's expected input; therefore, (H1) is refuted. First, with respect to mode, 6 of the 24 SV-3s only reflect the interfaces that exist between two different sets of subsystems (e.g., internal and external) [Endnote 1]. Known as **two-mode networks** (or bipartite graphs), these SV-3s do

not comport with the standard BA model [Endnote 2]. Next, of the 18 remaining SV-3s, 4 are asymmetric across their main diagonal or **directed**. While this does not agree with Algorithm 1's expected input, the standard BA model is extensible to directed networks by using each node's in-degree ($d_{i,in}$, the number of directional edges pointing to node i) when calculating attachment probabilities. Last, none of the 24 SV-3s are valued, let alone weighted according to interface complexity. Without interface complexities, the validity of using subsystem i 's observed interface complexity distribution to estimate future interface complexity cannot be assessed. In sum, although (H1) is refuted, 14 of the 24 real-world SV-3s are one-mode, undirected networks, and these will form the basis for our subsequent analysis.

(H2) – The densities of real-world SV-3s are less than 0.5.

Of the 14 one-mode, undirected SV-3s in Table 1, only one (i.e., System 6) has a density greater than 0.5. Accordingly, in general, (H2) is affirmed. Moreover, the average and median density among these 14 SV-3s are 0.245 and 0.214, respectively, which compare favorably with the hypothetical example's density of 0.428. As such, our sample suggests real-world SV-3s are sparse, and using the Girvan-Newman community detection heuristic in Algorithm 1 is appropriate.

(H3) – The modularities of real-world SV-3s are greater than 0.3.

While Table 1 provides definitive evidence for refuting (H1) and affirming (H2), the data for (H3) is less clear. In particular, 6 of the 14 one-mode, undirected SV-3s have modularities greater than 0.3. Put another way, roughly 50% of the real-world SV-3s in our sample have strong community structure worth exploiting, and Algorithm 1 accommodates this. On the other

System	Type			Size			Community Structure	
	Modes	Undirected?	Valued?	<i>N</i>	<i>E</i>	Density	Communities	Modularity
1	2	–	–	–	–	–	–	–
2	2	–	–	–	–	–	–	–
3	1	Y	N	19	26	0.152	4	0.365
4	1	Y	N	14	19	0.209	4	0.256
5	1	Y	N	21	46	0.219	2	0.072
6	1	Y	N	4	4	0.667	1	0
7	1	Y	N	15	16	0.152	3	0.498
8	1,2	Y	N	9	12	0.333	2	0.247
9	2	–	–	–	–	–	–	–
10	1	N	N	–	–	–	–	–
11	1	Y	N	24	19	0.069	8	0.611
12	1	N	N	–	–	–	–	–
13	1	N	N	–	–	–	–	–
14	1	Y	N	10	16	0.356	3	0.271
15	1	Y	N	28	23	0.061	8	0.749
16	1	Y	N	22	67	0.290	11	0.057
17	1	Y	N	19	18	0.105	6	0.532
18	1	Y	N	18	30	0.196	6	0.150
19	1	N	N	–	–	–	–	–
20	1,2	Y	N	9	8	0.222	3	0.414
21	2	–	–	–	–	–	–	–
22	2	–	–	–	–	–	–	–
23	2	–	–	–	–	–	–	–
24	1,2	Y	N	6	6	0.400	4	0.042

Table 1. Summary of the type, density, and community structure of real-world SV-3s. Bold entries highlight modularity values which indicate strong community structure. Systems with both one- and two-mode SV-3s are denoted by “1,2” in the “Modes” column. For these systems, values in the size and community structure panels correspond to the one-mode SV-3.

hand, when an SV-3’s modularity is less than 0.3, the identified community structure may be spurious, and exploiting it in Algorithm 1 is not recommended. As such, a non-community version of the connection mechanism is needed, and Algorithm 1 should be modified accordingly.

(H4) – Using the observed degree distribution of a real-world SV-3 to model its incoming subsystems’ number of interfaces grows SV-3s with a MAPE in interfaces less than or equal to 20%.

To test (H4), we simulated the growth of the 14 one-mode, undirected SV-3s in Table 1 using Algorithm 1’s growth and connection mechanisms. In particular, at $t = 0$ we start with an SV-3 consisting of two subsystems connected by a single interface. Accordingly, the PMF for the first incoming subsystem’s number of interfaces is simply $P(M = 1) = 1$, and it generates one interface, which is attached to either of the two existing subsystems with equal probability. At $t = 1$, the SV-3 consists of 3 subsystems, where two subsystems have one interface and one subsystem has two interfaces. As such, the PMF for the second incoming subsystem’s number of interfaces is

$$P(M = 1) = 2/3 \text{ and } P(M = 2) = 1/3.$$

Drawing a random variate from this PMF determines the number of interfaces the second incoming subsystem generates, and these interfaces are connected to the three existing subsystems using linear preferential attachment. At $t = 3$ the process repeats itself, and it continues until the SV-3 consists of N subsystems, at which point the total number of interfaces is recorded. The results of this simulation for 1,000 trials at each value of N are given in Table 2.

System	3	4	5	6	7	8	11	14	15	16	17	18	20	24
<i>N</i>	19	14	21	4	15	9	24	10	28	22	19	18	9	6
<i>E</i>	26	19	46	4	16	12	19	16	23	67	18	30	8	6
Density	0.15	0.21	0.22	0.67	0.15	0.33	0.07	0.36	0.06	0.29	0.11	0.2	0.22	0.4
Rank (by density)	10	8	7	1	10	4	13	3	14	5	12	9	6	2
95% CI on μ (LB)	63	35	76	3	40	15	99	18	131	83	63	56	15	7
95% CI on μ (UB)	65	37	79	3	41	15	103	19	137	86	65	58	15	7
MAPE	146	92	74	17	153	35	431	29	485	40	255	93	86	26
Rank (by MAPE)	10	8	6	1	11	4	13	3	14	5	12	9	7	2

Table 2. Results of growing SV-3s when the observed degree distribution is used to generate an incoming subsystem’s number of interfaces. The 95% confidence intervals (CIs) reflect a plausible range of values for the mean number of interfaces (μ), where LB and UB denote the lower and upper bounds, respectively.

As Table 2 shows, (H4) is refuted. Specifically, with the exception of System 6, using the observed degree distribution generated too many interfaces on average, and this is reflected in the magnitude of the lower bounds of the 95% confidence intervals on the mean number of interfaces as well as the MAPEs. Furthermore, as highlighted by the system ranks according to density and MAPE, sparse SV-3s performed worse than dense SV-3s, with the MAPE reaching 485% for the least dense SV-3 (System 15). Accordingly, the observed degree distribution is an inadequate model for an incoming subsystem’s number of interfaces, and an alternative growth mechanism is needed.

(H5) – The observed degree distributions of real-world SV-3s follow a power-law with $\hat{\omega} = 3$

As late as 2008, the preferred method for testing (H5) would have been to plot the histogram of the SV-3’s degrees on a log-log graph and subsequently perform least-squares regression to estimate ω . However, this method has several statistical shortcomings, and in 2009 Clauset et al. formalized hypothesis tests for assessing power law behavior, along with procedures for estimating ω and d_{min} in Equation (3) [Endnote 3]. These procedures have been implemented in R’s igraph package via the power.law.fit function (Csardi & Nepusz, 2006), and the results of this estimation for the 14 one-mode, undirected SV-3s are given in Table 3.

Power law distribution fitting		H_0 : Data follow a power law distribution with the fitted parameters													
		H_1 : Data do not follow a power law distribution with the fitted parameters													
System		3	4	5	6	7	8	11	14	15	16	17	18	20	24
$\hat{\omega}$		3.06	2.27	2.78	∞	2.8	2.69	2.52	2.26	2.33	2.57	2.06	2.22	2.01	∞
\hat{d}_{min}		3	2	3	–	2	2	1	2	1	4	1	2	1	–
<i>p</i> -value		0.99	0.99	0.58	–	0.99	0.99	1	0.67	1	0.5	0.28	0.99	0.83	–
95% CI on $\hat{\omega}$ (LB)		2.13	1.68	2.09	–	1.95	1.86	1.98	1.67	1.9	2	1.66	1.72	1.51	–
95% CI on $\hat{\omega}$ (UB)		5.11	3.64	4.09	–	4.82	4.76	3.46	3.62	3.05	3.57	2.77	3.16	3.19	–
N_{total}		19	14	21	4	15	9	24	10	28	22	19	18	9	6
N_{fit}		10	9	15	–	9	8	24	9	28	18	19	14	9	–

Table 3. Results of fitting discrete power law distributions to the observed degree distributions of real-world SV-3s. Systems 6 and 24 do not have sufficient data to estimate ω .

Based on these results, the power law distribution is a statistically plausible model for the observed degree distributions of the SV-3s. After all, among the 12 cases with sufficient data to estimate ω and d_{min} , the smallest p -value is 0.28 (System 17), suggesting we unanimously fail to reject the null hypothesis that the data follow a power law distribution with the fitted parameters. Moreover, in all but one case (System 17), the 95% confidence intervals for ω include 3, suggesting there is statistical support for the presence of linear preferential attachment in the SV-3s' evolution. Nevertheless, the high p -values for several of the systems are concerning, and they are likely due to the small sizes of the data sets used to fit the parameters (N_{fit}) [Endnote 4]. As Clauset et al. (2009) caution, "It is possible for small values of n that the empirical distribution will follow a power law closely, and hence that the p -value will be large, even when the power law is the wrong model for the data" (p. 678).

With this in mind, we can test whether the data follows another heavy-tailed distribution more closely, and the discrete exponential distribution is an appropriate choice, as it represents the limiting distribution of the subsystems' degrees when the p_i are uniform and preferential attachment is not present (Barabási, Albert, & Jeong, 1999). Accordingly, we performed a likelihood ratio test to determine whether a discrete power law or a discrete exponential distribution better fits the data (Clauset et al., 2009). Implemented in R's `powerLaw` package (Gillespie, 2015), the results of this test are summarized in Table 4, where p -values less than 0.05 indicate statistically significant results and, if significant, a positive (negative) test statistic

implies the power law (exponential) distribution is a better fit.

As seen in Table 4, only System 5 produced a statistically significant result, with the power law being favored over the exponential distribution. Additionally, Systems 14 and 20 were nearly significant, and, in both cases, the exponential distribution better fits the data. In sum, our likelihood ratio testing cannot refute the presence of linear preferential attachment in real-world SV-3s; however, due to small n , our conclusion lacks statistical power. Accordingly, (H5) is affirmed, but further investigation is warranted.

Accommodating Reality – A Modified, Data-Driven Approach

As the hypothesis testing in the previous section demonstrated, Algorithm 1 cannot accommodate all real-world SV-3s in its current form, and it must be modified. In particular, evaluating (H1) identified that SV-3s may be two-mode and directed, requiring different growth and connection mechanisms. Moreover, none of the 24 real-world SV-3s we examined are valued; therefore, the validity of using the observed interface complexity distribution to estimate future interface complexity cannot be assessed. Although these shortcomings must be addressed, they are beyond the scope of this work, and we restricted our focus to the 14 one-mode, undirected SV-3s.

For these SV-3s, evaluating (H2) confirmed the appropriateness of using the Girvan-Newman community detection heuristic, and analyzing

Power law versus exponential		H_0 : Both distributions are equally far from the true distribution H_1 : One of the test distributions is closer to the true distribution													
System	3	4	5	6	7	8	11	14	15	16	17	18	20	24	
p -value	0.33	0.56	0.01	–	0.86	0.28	0.31	0.06	0.65	0.24	0.21	0.53	0.08	–	
Test Statistic	-0.9	-0.6	2.8	–	-0.2	-1.1	1	-1.9	-0.5	-1.2	-1.2	-0.6	-1.8	–	

Table 4. Likelihood ratio test results reflecting whether a discrete power law or a discrete exponential distribution better fits the observed degree distributions of real-world SV-3s. Bold entries indicate statistically significant results.

	$E_{(1)}$	$E_{(2)}$	$E_{(3)}$	$E_{(4)}$	$E_{(5)}$	$E_{(6)}$	$E_{(7)}$	$E_{(8)}$	$E_{(9)}$	$E_{(10)}$	$E_{(11)}$	$E_{(12)}$	$E_{(13)}$	$E_{(14)}$	$E_{(15)}$	$E_{(16)}$	$E_{(17)}$	$E_{(18)}$	Sum
1	1	2	2	2	1	1	1	1	1	1	2	1	1	4	2	1	1	1	26
2	1	1	1	1	1	3	1	6	1	1	1	1	1	1	2	1	1	1	26
3	1	1	2	2	1	1	2	2	2	1	1	1	1	2	1	1	2	2	26

Figure 5. Three possible solutions to Equation (4).

(H3) indicated strong community structure for 6 of the 14 SV-3s. To accommodate SV-3s without strong community structure, Algorithm 1 needs an additional, non-community connection mechanism. Furthermore, investigating (H4) clearly indicated that using the observed degree distribution to model an incoming subsystem’s number of interfaces generates too many interfaces on average, and an alternative growth mechanism is needed. Finally, although examining (H5) provided statistical support for the presence of linear preferential attachment in the SV-3s’ evolution, the results lack statistical power.

While these shortcomings may seem damning, we see them as an opportunity to develop a modified, data-driven approach. Specifically, real-world SV-3s hold the key to finding models of evolutionary architectural growth, starting with the PMF for an incoming subsystem’s number of interfaces.

Establishing Feasible PMFs for an Incoming Subsystem’s Number of Interfaces ($P(M = m)$)

Consider System 3 from Table 1. With 19 subsystems and 26 interfaces, if m is modeled by the observed degree distribution, Table 2 indicates that the resulting SV-3s will have 38 too many interfaces on average. Therefore, our first objective is to find a distribution for m ($P(M = m)$) that grows SV-3s of size ($N = 19, E = 26$).

To do this, we return to System 3 and make the following observation. If the SV-3 starts as a single unconnected subsystem and each incoming subsystem must generate at least one interface, the remaining $N - 1 = 18$ subsystems must produce exactly $E = 26$ interfaces, and the following relation holds:

$$E_{(1)} + E_{(2)} + \dots + E_{(N-1)} = E \tag{4}$$

Where $E_{(i)}$ represents the number of interfaces the i^{th} oldest subsystem generates and

$1 \leq E_{(i)} \leq \min\{i, E - (N - 2)\}$ [Endnote 5]. For example, consider Figure 5, which gives three possible solutions to Equation (4) for System 3.

Examining the first solution, we note that $1 \leq E_{(i)} \leq \min\{i, 9\}$ for each subsystem and $\sum_{i=1}^{18} E_{(i)} = 26$; it is feasible. However, given the distribution of the $E_{(i)}$ ’s, the first solution has multiple feasible permutations. To show this directly, note that there are 12 subsystems with 1 interface, and one of these subsystems must fill the first position ($E_{(1)} = 1$). Following this assignment, 16 subsystems with less than 3 interfaces are available to fill the second and third positions, which yields four cases, namely:

- (a) $\{E_{(1)} = 1, E_{(2)} = 1, E_{(3)} = 1\}$,
- (b) $\{E_{(1)} = 1, E_{(2)} = 1, E_{(3)} = 2\}$,
- (c) $\{E_{(1)} = 1, E_{(2)} = 2, E_{(3)} = 1\}$, and
- (d) $\{E_{(1)} = 1, E_{(2)} = 2, E_{(3)} = 2\}$.

At this point, the single subsystem with four interfaces becomes available for assignment, and any of the 15 remaining subsystems can fill positions 4 through 18, where the distribution of the remaining E_i ’s is dependent on the case. In particular, if we denote the number of remaining subsystems with 1, 2, or 4 interfaces as n_1, n_2 , and n_4 , respectively, we note:

- (a) $\{n_1 = 9, n_2 = 5, n_4 = 1\}$,
- (b) $\{n_1 = 10, n_2 = 4, n_4 = 1\}$,
- (c) $\{n_1 = 10, n_2 = 4, n_4 = 1\}$, and
- (d) $\{n_1 = 11, n_2 = 4, n_4 = 1\}$.

Armed with this information and applying the well-known formula for the permutation of indistinguishable objects with repetition, the number of ways to feasibly permute the first solution in Figure 5 is:

$$\frac{15!}{(9! 5! 1!)} + \frac{15!}{(10! 4! 1!)} + \frac{15!}{(10! 4! 1!)} + \frac{15!}{(11! 3! 1!)} = 65,520 \quad (5)$$

Of course, similar logic applies to the second and third solutions in Figure 5, as well as every other non-permutationally equivalent solution. Simply put, the total number of feasible solutions is large.

Fortunately, while the sequencing of the subsystems affects feasibility and must be checked, it does not affect the distribution of m . Specifically, each of the 65,520 feasible permutations of the first solution has 12 subsystems with 1 interface, 5 subsystems with 2 interfaces, and 1 subsystem with 4 interfaces, implying $P(M = m) = \{0.667, m = 1; 0.278, m = 2; 0.55, m = 4\}$. Recalling our objective is to find a PMF for m that consistently grows SV-3s of size (N, E) , we do not need to find every feasible sequence for m . In the context of Equation (4), we need to find the unordered sets of $E_{(i)}$, $\{E_1, E_2, \dots, E_{N-1}\}$, such that $\sum_{i=1}^{N-1} E_i = E$ where $1 \leq E_i$ for all i .

In fact, this is equivalent to solving a restricted partition problem from number theory, where E (a positive integer) is decomposed into the sum of exactly $N - 1$ positive integers. Using the generating function approach (Gupta, 1970), the total number of restricted partitions, $p(E, N - 1)$, is equivalent to the coefficient of the x^E term after expanding the polynomial:

$$x^{N-1} \prod_{r=1}^{N-1} (1 + x^r + x^{2r} + \dots) \quad (6)$$

where each of the infinite order polynomials of the form $(1 + x^r + x^{2r} + \dots)$ can be truncated just prior to its order exceeding $N - 1$. Applying this result to System 3 yields:

$$x^{18} \prod_{r=1}^{18} (1 + x^r + x^{2r} + \dots) = \dots 30x^{27} + 22x^{26} + 15x^{25} + \dots$$

Therefore, $p(26, 18) = 22$, and these 22 restricted partitions (r) capture the possible PMFs for m , which can be enumerated using the restrictedparts function from R's partitions package (Hankin, 2006). Calling this function for $E = 26$ and $N - 1 = 18$ returns the result in Figure 6.

As Figure 6 shows, although each of the 22 restricted partitions sum to 26 and have a minimum value of 1, the maximum value ranges from a high of 9 to a low of 2. Accordingly, while

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}	E_{13}	E_{14}	E_{15}	E_{16}	E_{17}	E_{18}	Sum
1	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
2	8	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
3	7	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
4	6	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
5	5	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
6	7	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
7	6	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
8	5	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
9	5	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
10	4	4	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
11	6	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
12	5	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
13	4	4	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
14	4	3	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
15	3	3	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26
16	5	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	26
17	4	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	26
18	3	3	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	26
19	4	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	26
20	3	3	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	26
21	3	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	26
22	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	26

Figure 6. Restricted partitions of System 3's interfaces.

these PMFs for m are all feasible, they are different, and some may better replicate the degree distribution of System 3 when we use them to grow SV-3s of size ($N = 19, E = 26$).

Parameterizing the Strength of Preferential Attachment (β)

Armed with feasible PMFs for m , our next task is to determine the role linear preferential attachment plays in the formation of System 3's SV-3. Looking at Table 3, we failed to reject the null hypothesis that System 3's degree distribution followed a discrete power law distribution with $\omega = 3$, implying linear preferential attachment is present. However, as seen in Table 4, subsequent likelihood ratio testing failed to reject that a discrete exponential distribution fits the data equally well, suggesting linear preferential attachment might be absent.

These two cases represent poles along a continuum, and another possibility is that preferential attachment is present but it is not linear. For example, the probability of an incoming subsystem attaching to an existing subsystem i could be given by $p_i = d_i^\beta / \sum_{j=1}^N d_j^\beta$ where $0 < \beta < 1$ (Barabási, 2015). In this case, the incoming subsystem is more likely to connect with highly connected subsystems but the attachment probability is a sublinear function of degree – it is more muted. Moreover, even if likelihood ratio testing concludes that a discrete power law distribution fits the data better, it does not address the possibility that the preferential attachment could be a superlinear function of degree ($\beta < 1$). As with the feasible PMFs for m , settling on a best value for β necessitates that we grow SV-3s of size ($N = 19, E = 26$) for various values of β and measure how close their degree distributions are to System 3's. The question is:

“How should the difference between the degree distributions of two SV-3s of size (N, E) be measured?”

Assessing the Difference Between Degree Distributions (z^*)

Although there are several well-known methods for calculating the difference between probability distributions (e.g., the two-sample Kolmogorov-Smirnov test statistic (Darling, 1957) and the 1st Wasserstein metric or Earth Mover's Distance (Rubner, Tomasi, & Guibas, 1998)), the SV-3s' equivalent size invites an intuitive approach based on their observed degrees. Drawing on a frequently used example from combinatorial optimization, imagine we have N skiers with heights (h_i) and N sets of skis with lengths (l_j), and we want to match the skis to the skiers such that the sum of the absolute pairwise differences between the heights of the skiers and the lengths of the skis ($z = \sum |h_i - l_j|$) is minimized. As simple as it sounds, a globally optimal solution to this problem can be obtained by ordering the skis/skiers by length/height and assigning skis to skiers on a basis of their relative size (i.e., $z^* = \sum |h_{(k)} - l_{(k)}|$, where (k) denotes the k^{th} height/length of smallest skier/ski) (Lawler, 1976, p. 208).

By analogy, if we consider the N subsystems of SV-3s A and B as the skis and skiers, respectively, and match their subsystems on a basis of relative degree, the sum of the absolute pairwise differences between their degrees ($z = \sum |d_i - d_j|$) is minimized. If the degrees of the two SV-3s are identical, their degree distributions are the same, and the value of the minimum (z^*) equals zero. On the other hand, as the degrees become dissimilar, their degree distributions begin to diverge, and z^* increases. Consequently, z^* provides an intuitive measure of the difference between the degree distributions. With this in mind, when growing SV-3s of size (N, E), with the goal of replicating the degree distribution of a real-world SV-3, smaller values of z^* are preferred.

Simulating Growth to Identify Optimal $P(M = m)$ and Values for β

At this point, we have generated a set of feasible PMFs for an incoming subsystem’s number of interfaces $\{P(M = m)\}$, and we have parameterized the attachment probabilities to vary the strength of preferential attachment (β). Moreover, we have developed an intuitive measure z^* to assess the difference between the degree distributions of two SV-3s of size (N, E) . Ultimately, for a real-world SV-3, we would like to find the member(s) of $\{P(M = m)\}$ and value(s) of β that minimize z^* . Given the stochastic nature of the connection mechanism, as well as the need to feasibly permute the restricted partitions indicated by $\{P(M = m)\}$, z^* is a random variable (Z^*). Accordingly, we designed an experiment to

simulate the growth of SV-3s of size (N, E) for all restricted partitions and various values of β to identify which, if any, combination(s) of the parameters produces the smallest population mean for Z^* (μ_{z^*}). The pseudocode for this procedure is as follows:

- (1) Calculate, sort, and store the degrees of a real-world SV-3 of size (N, E)
- (2) Generate and store the restricted partitions for SV-3s of size $(N - 1, E)$,
- (3) Determine a representative set of values for β (i.e., $\beta = \{0, 0.1, 0.2, \dots, 1\}$), and
- (4) For each restricted partition (r) and value of β , for a specified, suitably large number of iterations, on each iteration $l \dots$

		β										Mean	
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9		1
Restricted Partition Number (r)	1	11.68	11.70	11.49	11.43	11.20	11.65	11.52	12.00	11.92	12.59	13.21	11.85
	2	10.34	10.24	10.02	10.09	10.05	10.07	10.22	10.41	10.92	11.39	11.90	10.51
	3	8.61	8.41	8.56	8.45	8.42	8.60	8.81	9.04	9.65	10.10	10.62	9.02
	4	7.45	7.37	7.29	7.32	7.27	7.47	7.90	8.31	8.67	9.34	10.06	8.04
	5	7.22	7.07	7.046	7.13	7.22	7.41	7.69	8.02	8.59	9.22	9.67	7.84
	6	9.08	8.88	8.93	8.74	8.62	8.88	9.12	9.45	9.77	10.35	10.66	9.32
	7	7.56	7.50	7.38	7.50	7.52	7.72	7.94	8.13	8.61	9.19	9.76	8.07
	8	7.72	7.60	7.36	7.45	7.58	7.51	7.72	8.03	8.46	9.09	9.64	8.02
	9	7.42	7.23	7.23	7.18	7.054	7.24	7.55	7.73	8.18	8.58	9.31	7.70
	10	7.93	7.84	7.68	7.55	7.61	7.71	7.80	8.20	8.42	8.95	9.51	8.11
	11	8.06	8.04	7.84	7.97	7.80	8.16	8.20	8.37	8.80	9.38	9.75	8.40
	12	7.95	7.71	7.59	7.52	7.58	7.55	7.81	8.16	8.43	8.74	9.37	8.04
	13	8.38	8.15	7.90	7.97	7.92	7.98	8.07	8.25	8.64	8.97	9.74	8.36
	14	7.98	7.91	7.63	7.57	7.70	7.54	7.74	8.08	8.29	8.65	9.19	8.02
	15	8.06	7.73	7.55	7.20	7.32	7.20	7.36	7.67	7.97	8.22	8.85	7.74
	16	8.26	8.29	8.00	8.17	8.06	7.97	8.31	8.57	8.72	9.15	9.69	8.47
	17	8.45	8.34	8.22	8.14	8.07	8.10	8.35	8.53	8.53	8.90	9.40	8.46
	18	8.44	8.14	8.13	7.77	7.72	7.85	7.87	7.96	8.43	8.60	9.25	8.20
	19	9.13	8.94	8.88	8.59	8.56	8.59	8.76	9.04	9.16	9.46	10.06	9.02
	20	8.96	8.73	8.52	8.60	8.44	8.43	8.36	8.65	9.01	9.08	9.65	8.77
	21	9.60	9.38	9.15	8.92	9.06	9.02	9.07	9.31	9.53	9.92	10.38	9.39
	22	10.09	9.87	9.72	9.64	9.64	9.73	10.07	10.05	10.46	10.77	11.30	10.12
Mean	8.56	8.41	8.28	8.22	8.20	8.29	8.47	8.73	9.05	9.48	10.04		

Figure 7. Sample means of the minimum absolute pairwise differences between the degrees of System 3’s SV-3 and the degrees of 1,000 simulated SV-3s for each (r, β) pair. Yellow shading denotes the five smallest sample means, and blue shading reflects the relative magnitude of the sample means across the 242 (r, β) pairs, where darker shades reflect larger means.

- (a) Randomly permute the restricted partition until feasible $(E_{(i)} \leq i \text{ for } i \in \{1, 2, \dots, N - 1\})$,
- (b) Initialize the simulated SV-3 as a single subsystem,
- (c) Sequentially add $N - 1$ subsystems to the simulated SV-3, where the i^{th} incoming subsystem generates $E_{(i)}$ interfaces and attaches to existing subsystem k with probability $p_k \propto d_k^\beta$,
- (d) Calculate and sort the degrees of the simulated SV-3, and
- (e) Sum the absolute pairwise differences between the sorted degrees of the simulated SV-3 and the real-world SV-3; store the result as $Z^*_{r,\beta,l}$.

Increasing β from 0 to 1 in increments of 0.1 and running 1,000 iterations for each (r, β) pair produce 242,000 realizations of Z^* for System 3. The sample means of this experiment are summarized in Figure 7.

As seen in Figure 7, the sample means range from a minimum of 7.046 for the $(r = 5, \beta = 0.2)$ pair to a maximum of 13.21 for the $(r = 1, \beta = 0.1)$ pair. Additionally, the poorest performing (r, β) pairs seem to be concentrated in $r = \{1, 2, 21, 22\}$ and $\beta = \{0.9, 1\}$. Based on these observations, parameter settings affect the fit, and this is highlighted in Figure 8, where the sample means have been converted to ascending ranks by row and column and shaded accordingly.

In particular, on the left side of Figure 8 it appears that $\beta = \{0.2, 0.3, 0.4, 0.5\}$ produce SV-3s with degrees that more closely match the degrees of System 3's SV-3, immaterial of r . Similarly, regardless of the value of β , on the right side of Figure 8 it seems that $r = \{5, 9, 15\}$ perform relatively well. Based on these observations, we conclude that the underlying population means depend on the restricted partition and the value of β and that a best fitting (or better fitting) (r, β) pair(s) may exist.

To test this conclusion formally, consider (a) we have two factors (the restricted partitions and

		β										β													
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1		
Restricted Partition Number (r)	1	6	7	3	2	1	5	4	9	8	10	11	22	22	22	22	22	22	22	22	22	22	22	1	
	2	7	6	1	4	2	3	5	8	9	10	11	21	21	21	21	21	21	21	21	21	21	21	2	
	3	6	1	4	3	2	5	7	8	9	10	11	15	15	16	15	15	17	16	18	18	18	18	3	
	4	5	4	2	3	1	6	7	8	9	10	11	3	3	3	4	3	4	9	11	12	14	16	4	
	5	4	2	1	3	5	6	7	8	9	10	11	1	1	1	1	2	3	3	4	9	13	10	5	
	6	6	3	5	2	1	3	7	8	9	10	11	17	17	18	18	18	18	19	19	19	19	19	6	
	7	5	3	1	2	4	6	7	8	9	10	11	4	4	5	6	5	9	10	7	10	12	14	7	
	8	6	5	1	2	4	3	7	8	9	10	11	5	5	4	5	6	5	4	5	7	10	8	8	
	9	6	4	3	2	1	5	7	8	9	10	11	2	2	2	2	1	2	2	2	2	2	4	9	
	10	7	6	3	1	2	4	5	8	9	10	11	6	8	9	8	8	8	6	9	4	7	7	10	
	11	5	4	2	3	1	6	7	8	9	10	11	9	10	10	12	11	14	12	12	14	15	13	11	
	12	7	5	4	1	3	2	6	8	9	10	11	7	6	7	7	7	7	7	8	5	5	5	12	
	13	8	6	1	3	2	4	5	7	9	10	11	12	12	11	11	12	12	11	10	11	8	12	13	
	14	7	6	3	2	4	1	5	8	9	10	11	8	9	8	9	9	6	5	6	3	4	2	14	
	15	9	7	5	1	3	1	4	6	8	10	11	9	7	6	3	4	1	1	1	1	1	1	15	
	16	5	6	2	4	3	1	7	8	9	10	11	11	13	12	14	13	11	13	14	13	11	11	16	
	17	7	5	4	3	1	2	6	9	8	10	11	14	14	14	13	14	13	14	13	8	6	6	17	
	18	9	7	6	2	1	3	4	5	8	10	11	13	11	13	10	10	10	8	3	6	3	3	18	
	19	8	6	5	3	1	2	4	7	9	10	11	18	18	17	16	17	16	16	17	16	16	15	19	
	20	8	7	4	5	3	2	1	6	9	10	11	16	16	15	17	16	15	15	15	15	9	9	20	
	21	9	7	5	1	3	2	4	6	8	10	11	19	19	19	19	19	19	18	18	17	17	17	21	
	22	8	5	3	1	2	4	7	6	9	10	11	20	20	20	20	20	20	20	20	20	20	20	22	

Figure 8. Ascending ranks of the sample means from Figure 7 by restricted partition number (left side) and β (right side). Yellow shading denotes the seemingly best performing levels of β and r by rank, and blue shading reflects the relative magnitude of the ranks across the rows (left side) and columns (right side), where darker shades reflect higher ranks.

the set of β s); (b) we are evaluating each factor at every level of interest; (c) we are interested in which, if any, combination(s) of the factors produce the minimum mean; and (d) the number of iterations at each combination of the factors is the same. Accordingly, our simulation experiment is a balanced, fixed effects two-factor factorial design, and the corresponding effects model for our simulation experiment is given by:

$$z_{r,\beta,l}^* = \mu + \tau_r + \gamma_\beta + (\tau\gamma)_{r,\beta} + \epsilon_{r,\beta,l} \begin{cases} r = 1, 2, \dots, p(E, N - 1) \\ \beta = 0, 0.1, \dots, 1 \\ l = 1, 2, \dots, 1,000 \end{cases} \quad (7)$$

where μ is the overall mean effect; τ_r is the effect of the r^{th} restricted partition; γ_β is the effect of β at its indicated level; $(\tau\gamma)_{r,\beta}$ is the effect of the interaction between τ_r and γ_β ; and $\epsilon_{r,\beta,i}$ is a random error component (Montgomery, 2005, p. 165). Furthermore, if we assume the $\epsilon_{r,\beta,i}$ are independently and normally distributed with mean 0 and variance σ^2 for each (r, β) pair, we can use two-way analysis of variance (ANOVA) to perform an omnibus F test of the following hypotheses (Montgomery, 2005, p. 166):

$$H_0: \tau_1 = \tau_2 = \dots = \tau_{p(E,N-1)} = 0 \quad (8)$$

$$H_1: \text{at least one } \tau_r \neq 0$$

$$H_0: \gamma_0 = \gamma_{0.1} = \dots = \gamma_1 = 0 \quad (9)$$

$$H_1: \text{at least one } \gamma_\beta \neq 0$$

$$H_0: (\tau\gamma)_{r,\beta} = 0 \text{ for all } r, \beta \quad (10)$$

$$H_1: \text{at least one } (\tau\gamma)_{r,\beta} \neq 0.$$

Unfortunately, the $Z_{r,\beta,l}^*$ constitute a finite set of positive integers; therefore, the $\epsilon_{r,\beta,l}$ will not be normally distributed. Nonetheless, with respect to the Type I error rate, the F test is robust against violations of the normality assumption (Donaldson, 1968), even when the dependent variable assumes a very small number of discrete values (Bevan, Denton, & Myers, 1974). Additionally, although the F test assumes equality of variance in the $\epsilon_{r,\beta,l}$, when this assumption is violated in balanced designs numerous studies have shown that the actual probability of

committing a Type I error closely matches the nominal level of significance (e.g., Glass, Peckham, & Sanders, 1972). That said, these studies typically employ experimentation versus analytical derivation, and generalizing their conclusions to our specific situation seems questionable. With this in mind, we can alleviate any issues by halving the nominal level of significance (i.e., from $\alpha = 0.05$ to $\alpha = 0.025$), as Keppel and Wickens (2004) note this is “the fastest and simplest way to eliminate concerns about heterogeneity” (p. 152). Using this approach, we performed a two-way ANOVA on the output of our simulation experiment, and the results are summarized in Table 5.

As seen in Table 5, both of the main effects and the interaction effect are significant at $\alpha = 0.025$,

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	p -value
r	243083	21	11575	1607.38	0
β	77610	10	7761	1077.71	0
Interaction	9043	210	43	5.98	0
Error	1740994	241758	7		

Table 5. Results of two-way ANOVA.

and we reject the null hypotheses captured in Equations (8), (9), and (10) in favor of their alternatives [Endnote 6]. Additionally, given the significant interaction effect, the focus of our post-hoc testing is on the individual cell means in Figure 7 versus the row or column means.

With this in mind, we can treat each of the 242 (r, β) pairs as separate levels of a single factor, thereby reducing our two-way ANOVA problem to a one-way problem. Moreover, as we are interested in whether there is an (r, β) pair with the smallest population mean, Hsu’s multiple comparisons with the best (MCB) procedure (Hsu, 1984) is an appropriate post-hoc test. In particular, if we denote the population mean of the i^{th} (r, β) pair as μ_i , Hsu’s MCB constructs simultaneous, two-sided confidence intervals for $\mu_i - \min_{j \neq i} \mu_j, i, j = 1, 2, \dots, 242$ such that the family

-wise error rate (FWER) is controlled at a specified level (typically 0.05). If a confidence interval contains 0, then the population means of the corresponding (r, β) pairs are deemed equivalent and optimal. Otherwise, a statistically significant difference exists, and the sign of the confidence interval's bounds determines which (r, β) pair is best. With this in mind, we set the FWER at 0.025 to account for the known heterogeneity, and we ran Hsu's MCB in the statistical software Minitab (2015) [Endnote 7]. The results of this post-hoc testing are given Figure 9.

As Figure 9 indicates, although the $(r = 5, \beta = 0.02)$ pair has the smallest sample mean, Hsu's MCB suggests that 23 additional (r, β) pairs are also optimal. Interestingly, the optimal values of β

range from 0 to 0.6, with the majority falling on 0.2 and 0.3. This suggests that the preferential attachment mechanism is sublinear, and this fits with our previous likelihood ratio testing (see Table 4). Specifically, although we failed to reject the null hypothesis that the discrete power law and exponential distributions are equally far from System 3's true degree distribution, the test statistic is slightly negative. As such, the evidence (albeit not statistically significant) favors the exponential distribution, and it hints that the preferential attachment mechanism is more uniform than linear. Nonetheless, in the absence of additional evidence, the optimal $P(M = m)$ and values for β in Table 6 constitute a set of equally compelling conditions for generating an incoming subsystem's interfaces and preferentially attaching them to System 3.

		β											Mean
		1	0.9	0.8	0.7	0	0.6	0.1	0.5	0.2	0.3	0.4	
Restricted Partition Number (r)	1	13.21	12.59	11.92	12.00	11.68	11.52	11.70	11.65	11.49	11.43	11.20	11.85
	2	11.90	11.39	10.92	10.41	10.34	10.22	10.24	10.07	10.02	10.09	10.05	10.51
	22	11.30	10.77	10.46	10.05	10.09	10.07	9.87	9.73	9.72	9.64	9.64	10.12
	21	10.38	9.92	9.53	9.31	9.60	9.07	9.38	9.02	9.15	8.92	9.06	9.39
	6	10.66	10.35	9.77	9.45	9.08	9.12	8.88	8.88	8.93	8.74	8.62	9.32
	3	10.62	10.10	9.65	9.04	8.61	8.81	8.41	8.60	8.56	8.45	8.42	9.02
	19	10.06	9.46	9.16	9.04	9.13	8.76	8.94	8.59	8.88	8.59	8.56	9.02
	20	9.65	9.08	9.01	8.65	8.96	8.36	8.73	8.43	8.52	8.60	8.44	8.77
	16	9.69	9.15	8.72	8.57	8.26	8.31	8.29	7.97	8.00	8.17	8.06	8.47
	17	9.40	8.90	8.53	8.53	8.45	8.35	8.34	8.10	8.22	8.14	8.07	8.46
	11	9.75	9.38	8.80	8.37	8.06	8.20	8.04	8.16	7.84	7.97	7.80	8.40
	13	9.74	8.97	8.64	8.25	8.38	8.07	8.15	7.98	7.90	7.97	7.92	8.36
	18	9.25	8.60	8.43	7.96	8.44	7.87	8.14	7.85	8.13	7.77	7.72	8.20
	10	9.51	8.95	8.42	8.20	7.93	7.80	7.84	7.71	7.68	7.55	7.61	8.11
	7	9.76	9.19	8.61	8.13	7.56	7.94	7.50	7.72	7.38	7.50	7.52	8.07
	4	10.06	9.34	8.67	8.31	7.45	7.90	7.37	7.47	7.29	7.32	7.27	8.04
	12	9.37	8.74	8.43	8.16	7.95	7.81	7.71	7.55	7.59	7.52	7.58	8.04
	14	9.19	8.65	8.29	8.08	7.98	7.74	7.91	7.54	7.63	7.57	7.70	8.03
	8	9.64	9.09	8.46	8.03	7.72	7.72	7.60	7.51	7.36	7.45	7.58	8.02
	5	9.67	9.22	8.59	8.02	7.22	7.69	7.07	7.41	7.046	7.13	7.22	7.85
15	8.85	8.22	7.97	7.67	8.06	7.36	7.73	7.20	7.55	7.20	7.32	7.74	
9	9.31	8.58	8.18	7.73	7.42	7.55	7.23	7.24	7.23	7.18	7.054	7.70	
Mean	10.04	9.49	9.05	8.73	8.56	8.47	8.41	8.29	8.28	8.22	8.20		

Figure 9. Results of Hsu's MCB with FWER 0.025, where the rows and columns of Figure 7 have been permuted by descending sample means. Yellow shading denotes the (r, β) pair with the smallest (optimal) sample / population mean ($\mu_{r=5, \beta=0.2}$); green shading indicates additional (r, β) pairs with population means equal to ($\mu_{r=5, \beta=0.2}$); and blue shading reflects the relative magnitude of the sample means across the remaining, suboptimal (r, β) pairs, where darker shades reflect larger means.

<i>r</i>	Optimal $P(M = m)$						Optimal β						
	<i>m</i>						β						
	1	2	3	4	5	6	0	0.1	0.2	0.3	0.4	0.5	0.6
4	0.889			0.056		0.056	1	1	1	1	1		
5	0.889				0.111		1	1	1	1	1	1	
7	0.833	0.056	0.056						1				
8	0.833	0.056		0.056	0.056				1	1			
9	0.833		0.111		0.056		1	1	1	1	1	1	
15	0.778		0.222							1	1	1	1

Table 6. Optimal $P(M = m)$ and values for β . The restricted partitions (*r*) listed in the leftmost column appear at least once in an optimal (*r*, β) pair (i.e., the corresponding rows of yellow or green-shaded cells in Figure 9), and these have been transformed into their corresponding PMFs in the panel titled “Optimal $P(M = m)$.” Similarly, in the panel titled “Optimal β ,” 1s indicate that the corresponding values of β are part of an associated optimal (*r*, β) pair (i.e., the corresponding columns of yellow or green-shaded cells in Figure 9).

Incorporating Findings into a Modified, Data-Driven Approach

As seen above, using linear preferential attachment to connect an incoming subsystem to System 3’s existing architecture is ill-advised. Moreover, as seen in Table 7, ANOVA and Hsu’s MCB analysis of Systems 4, 6, 7, 8, 11, 14, 15, 17, 18, 20, and 24 reinforce the appropriateness and importance of using a data-driven approach.

In particular, while Brown-Forsythe testing found statistically significant heterogeneity in each system’s data except System 24, the maximum to minimum variance ratios were less than 4 to 1, allowing us to safely proceed with ANOVA using half the nominal level of significance. As with System 3, ANOVA identified significant main effects in every system except System 24, and for systems with multiple restricted partitions, the interaction effect was also significant. Subsequent Hsu’s MCB analysis for the ten systems with significant effects revealed optimal β ranging from 0 to 1, with some systems favoring uniform attachment (i.e., Systems 14, 15, and 17) and others leaning towards linear attachment (i.e., Systems 4, 6, and 11). Additionally, with the exception of System 15, every system had multiple optimal (*r*, β) pairs, yet the number of optimal pairs was a fraction of the total number of pairs, especially for systems with multiple restricted partitions.

Simply put, real-world SV-3s suggest a one-size-fits-all approach is overly simplistic, and this also applies to the statistical methods we employed. For example, Systems 5 and 16 are not represented in Table 7, and this is a deliberate omission. Specifically, Systems 5 and 16 have 2,417 and 98,222 restricted partitions, respectively, and, assuming a significant interaction effect and with 11 levels of β , this implies the simultaneous testing of 26,587 and 1,080,442 hypotheses via Hsu’s MCB. This is well over 20 times the number of hypotheses tested in the next closest system in Table 7, and it falls into the domain of large-scale simultaneous inference, where minor deviations from the theoretical null hypothesis can substantially affect the results (Efron, 2012). In short, more advanced methods are necessary to analyze these systems.

Beyond statistical methods, our approach is limited to simultaneously finding PMFs for an incoming subsystem’s number of interfaces and estimating the strength of preferential attachment. It does not address architectural communities. For example, during the growth of System 3 using the (*r* = 5, β = 0.02) pair, we applied the Girvan-Newman community detection heuristic to each of the 1,000 simulated SV-3s, producing the histogram and kernel density plot seen in Figure 10.

Preprocessing					ANOVA Source of Variation <i>p</i> -value ($\alpha = 0.025$)			Hsu's MCB Analysis (FWER = 0.025)		
System #	# of restricted partitions	# of feasible (r, β) pairs	Brown-Forsythe <i>p</i> -value	Max to Min Variance Ratio	<i>r</i>	β	Interaction	# of optimal (r, β) pairs	<i>r</i>	β
4	11	121	0	2.2	0	0	0	6	1	{0.5,...,1}
6	1	11	0	1.89	NA	0	NA	5	1	{0.6,...,1}
7	2	22	0	2.08	0	0	0.002	6	2	{0,...,0.5}
8	5	55	0	1.47	0	0	0	5	5	{0,...,0.4}
11	1	11	0	1.27	NA	0	NA	2	1	{0.9, 1}
14	15	165	0	2.45	0	0	0	10	7	{0, 0.1, 0.2}
									10	{0.2}
									12	{0,...,0.3}
									14	{0}
15	1	11	0	1.44	NA	0	NA	1	1	{0}
17	1	11	0	1.61	NA	0	NA	3	1	{0, 0.1, 0.2}
18	101	1111	0	3.62	0	0	0	31	5	{0.3,...,0.8}
									6	{0.3, 0.5}
									10	{0.5,...,0.8}
									11	{0.4, 0.5, 0.7}
									14	{0.7}
									23	{0.6, 0.7, 0.8}
									40	{0.6, 0.7, 0.8}
									41	{0.7}
									58	{0.6,...,0.9}
									72	{0.7, 0.8, 0.9}
									95	{1}
20	1	11	0	1.72	NA	0	NA	3	1	{0, 0.1, 0.2}
24	1	11	0.8716	1.15	NA	0.25	NA	NA	NA	NA

Table 7. ANOVA and Hsu's MCB analysis of remaining one-mode, undirected SV-3s. Systems with a single restricted partition only have effects due to β ; thus, their interaction effects are undefined.

Based on Figure 10 and recalling Table 1, these results are encouraging. After all, System 3's SV-3 displayed strong community structure with four communities and a modularity of 0.365. In Figure 10, the plurality of the 1,000 simulated SV-3's contained four communities, and, when this occurred, the mean modularity was 0.361 – a difference of just 1%. When we consider that our simulation does not control for community structure, this miniscule difference in modularity

is remarkable. Nonetheless, a majority of the 1,000 simulated SV-3s did not have four communities; therefore, we cannot claim our growth mechanism reliably replicates System 3's community structure.

With this in mind, we see the PMF for an incoming subsystem's number of interfaces as a system-level property. Thus, unlike Algorithm 1, we no longer recommend calculating separate PMFs for an incoming subsystem's number of

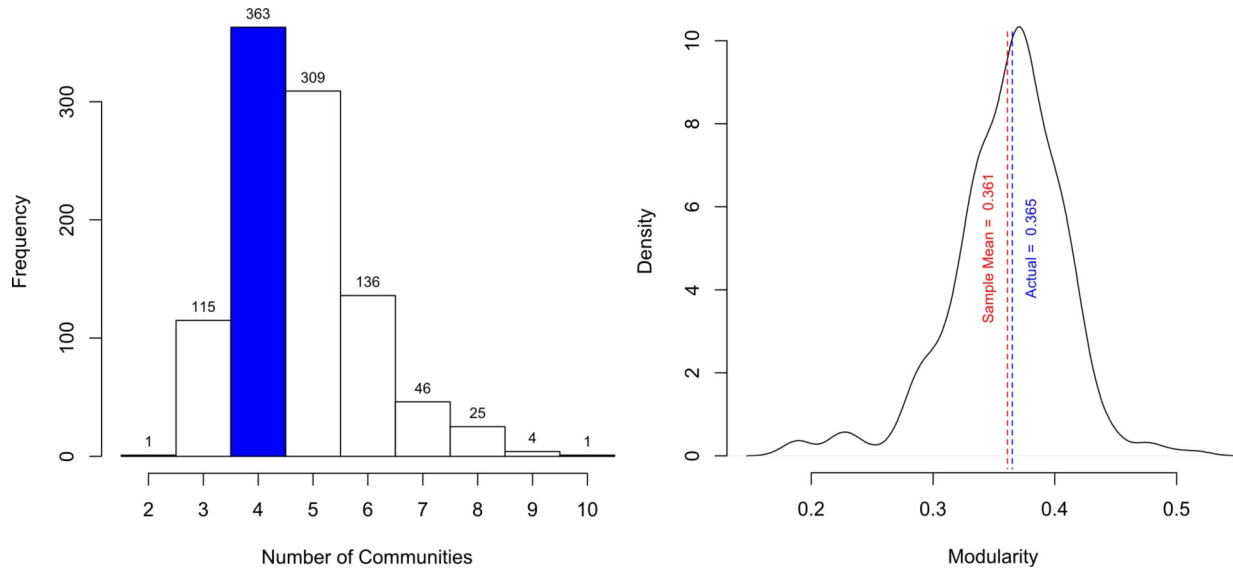


Figure 10. Histogram and kernel density plot reflecting the community structure of System 3’s simulated growth using the $r=5, \beta=0.02$ pair.

intra and intercommunity interfaces, and we adopt a different approach. Specifically, if an incoming subsystem generates m interfaces and it is subsequently assigned to community k , we can view the intracommunity designation of the m interfaces as a sequence of m Bernoulli trials, where the probability of success (p) is given by:

$$p = \frac{\text{number of community } k\text{'s intracommunity interfaces}}{\text{number of community } k\text{'s intra and intercommunity interfaces}}$$

When viewed in this way, an incoming subsystem’s number of intracommunity interfaces, M_{intra} , is simply a Binomial (m, p) random variable. Of course, it is possible that M_{intra} could generate a realization (m_{intra}) that exceeds the size of community k (N_k). Accordingly, we can set the number of intra and intercommunity interfaces as $m'_{intra} = \min\{m_{intra}, N_k\}$ and $m_{inter} = m - m'$ respectively.

Using this approach and in light of our previous findings, we recommend the following, modified version of Algorithm 1:

Algorithm 2

For a specified, suitably large number of iterations ...

Preprocessing

- (1) Initialize the system as the current system,
- (2) Build an optimal set of $\{P(M = m), \beta\}$ pairs,
- (3) Use Girvan-Newman to identify architectural communities and calculate modularity,

Growth

- (4) Randomly select a member from the optimal set of $\{P(M = m), \beta\}$ pairs,
- (5) Generate a realization for the incoming subsystem’s (X ’s) number of interfaces using $P(M = m)$; if the modularity suggests strong community structure, use **Connection Option A**; otherwise, use **Connection Option B**,

Connection Option A

- (6a) Randomly assign \mathbf{X} to community k ,
- (6b) Model M_{intra} as a Binomial (m, p) random variable; generate a realization for M_{intra} ; and set the number of intra and intercommunity interfaces as $m'_{\text{intra}} = \min\{m_{\text{intra}}, N_k\}$ and $m'_{\text{inter}} = m - \{m'_{\text{intra}}\}$ respectively,
- (6c) Attach \mathbf{X} to m'_{intra} subsystems inside community k and m'_{inter} subsystems outside community k using attachment probabilities $p_i = d_i^\beta / \sum_{j=1}^N d_j^\beta$
- (6d) For each intracommunity interface established in (6c), assign complexity $(w_{i\mathbf{X},\text{intra}})$,
- (6e) For each intercommunity interface established in (6c), assign complexity $(w_{i\mathbf{X},\text{inter}})$,

Connection Option B

- (6a) Attach \mathbf{X} to m subsystems using attachment probabilities,
- $$p_i = d_i^\beta / \sum_{j=1}^N d_j^\beta$$
- (6b) For each interface established in (6a), assign complexity $(w_{i\mathbf{X}})$,

Cost Estimation

- (7) Estimate the cost for the augmented system using COSYSMO (PM_{NS^*}),
- (8) Calculate the additional cost of adding subsystem \mathbf{X} ($PM_{NS^*} - PM_{NS}$), and
- (9) Store results and return to (4).

Limitations and Directions for Future Research

Although Algorithm 2 addresses several of Algorithm 1's shortcomings, it still has limitations. First, as indicated earlier, SV-3s are

not currently weighted by interface complexity. Thus, the validity of using the observed interface complexity distribution to estimate future interface complexity in Steps (6b), (6d), and (6e) above cannot be assessed, and further research is necessary. Second, while Algorithm 2 accepts one-mode, undirected SV-3s as input, real-world SV-3s can be two-mode or directed. With this in mind, methods that accommodate these SV-3 types could yield additional, valuable information and should be explored. Third, Algorithm 2 employs the Girvan-Newman community detection heuristic, and, despite its appropriateness, better performing heuristics exist (see Danon, Diaz-Guilera, Duch, & Arenas, 2005). Nonetheless, any community detection method, regardless of its performance, may ignore other, more compelling macrostructures within the architecture. For example, subsystems may partition into a hierarchy of clusters, where subsystems in lower ranking clusters not only have a high density of interfaces with subsystems inside their clusters but also have a high density of interfaces with subsystems inside higher ranking clusters. To identify this and other hidden macrostructure, one can apply the network analysis technique known as blockmodeling, and this represents an intriguing way to generalize the current approach.

Conclusions


The requirement to submit DoD component-approved DoDAF models prior to MS A has created interesting, new possibilities for the early life cycle cost estimation of MDAPs. In particular, Valerdi et al. (2015) demonstrate that the DoDAF models required Pre-MS A nearly span COSYSMO's parameters, and Dabkowski et al. (2014) exploited this mapping in Algorithm 1 by estimating the cost of unanticipated, evolutionary architectural growth via the SV-3 and COSYSMO. Although this development could be seen as positive, any cost estimation procedure ultimately needs to be informed and validated by real-world data. Accordingly, in this paper, we

examined the assumptions underlying Algorithm 1 using the SV-3s from 24 different defense programs.

The results were mixed. Specifically, while the type, density, and community structure of real-world SV-3s generally comport with Algorithm 1, modifications and extensions are necessary to accommodate two-mode or directed SV-3s. Moreover, using the observed degree distribution to model an incoming subsystem's interfaces generates too many interfaces, and an alternative growth mechanism is needed. Finally, although there is statistical support for linear preferential attachment as a method of connecting subsystems, the SV-3s are small, and our statistical results inevitably lack power.

With this in mind, we developed a modified, data-driven approach that addresses several of these concerns. In particular, using number theory, network science, simulation, and statistical analysis, we were able to find optimal sets of PMFs and strengths of preferential attachment for 12 of the 14 one-mode, undirected SV-3s. Integrated into Algorithm 2, these optimal sets better represent a system's evolutionary growth, and they improve the fidelity of the algorithm. Nonetheless, as noted earlier, our approach has several limitations, and these represent

opportunities for future research.

Aside from developing Algorithm 2, this paper also makes several tangential contributions to network science. First, to the best of our knowledge, this is the first attempt at simultaneously estimating a growing network's incoming edge distribution and detecting the strength of preferential attachment. To date, these efforts have been disconnected, and linking them is not only natural but also necessary in light of their significant interaction effect. Second, assessing the presence of linear preferential attachment has traditionally been confined to large networks with longitudinal data. By modeling the set of feasible edge sequences via restricted partitions, we have provided researchers with a way to accommodate small networks with a single realization. Last, while other metrics exist, our use of the minimum sum of the absolute pairwise differences between the degrees of two identically-sized networks provides analysts with an intuitive measure of the similarity between their degree distributions. Taken together, these contributions highlight the benefits of applying network science to new domains, and they reinforce the value of viewing DoDAF models as computational objects. 

Acknowledgements:

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-13-1-0032 and the Office of the Secretary of Defense.

Disclaimer:

The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Contributions:

Author contributions are as follows: M.D. and R.V. designed research; M.D. analyzed data; M.D. and A.M. performed research and wrote the paper.

Endnotes

1. For those familiar with DoDAF, this seems to violate the generally prescribed structure of the SV-3. However, during DoDAF's most recent, major revision, models shifted from a template-driven paradigm to a "fit for purpose" construct (DoD DCIO, 2010, p. 3).
2. In practice, two-mode networks are often transformed into one-mode networks to facilitate analysis (Borgatti, 2009). For example, imagine a two-mode network \mathbf{X} , where rows represent professors, columns denote institutional committees, and a 1 in cell (i, j) implies professor i is a member of committee j . Under the assumption that co-memberships in committees imply meaningful connections between professors, $\mathbf{X}\mathbf{X}^T$ yields a useful one-mode, valued network \mathbf{A} , where rows and columns represent the professors and values provide the number of co-memberships for each pair of professors. Further binarizing \mathbf{A} (such that cells greater than 1 are set to 1) yields a simple one-mode network. That said, using this approach for a two-mode SV-3 is ill-advised, as internal subsystems that interface with common external subsystem(s) do not necessarily interface with one another.
3. While the details of estimating d_{\min} are beyond the scope of this work, Clauset et al. (2009) note that \hat{d}_{\min} is selected such that its value "makes the probability distributions of the measured data and the best-fit power-law model as similar as possible above \hat{d}_{\min} " (p. 671).
4. Data to the left of d_{\min} is discarded prior to estimating ω ; therefore, $N_{\text{fit}} \leq N_{\text{total}}$.
5. To ascertain the validity of the final inequality, note: (a) prior to the i^{th} oldest subsystem entering the SV-3 there are i subsystems in it, which implies the i^{th} oldest subsystem can connect to at most i subsystems and (b) if $E_i \geq 1$, E_j is maximized when $E_i = 1$ for all $i \neq j$, which implies $E_j = E - (N - 2)$.
6. As stated previously, halving the nominal level of significance provides a reasonable hedge against unequal variance in the $\epsilon_{r, \beta, I}$; however, if the variances are equal, it is unnecessary. Accordingly, given the non-normality of the data, we evaluated the homogeneity of the residuals using the Brown-Forsythe test, and this returned a p -value of 0, firmly rejecting the variances of the $\epsilon_{r, \beta, I}$ are equal. That said, unequal variances "typically cause Type I error rates to be slightly inflated . . . less than 0.02 at the 0.05 level . . . provided the ratio of the largest to the smallest variance is no more than 4 to 1, and n is at least 5" (Myers, Well, & Lorch, 2010, p. 191). In our case, the maximum and minimum variances are 14.02 (for $(r = 22, \beta = 0.9)$) and 4.35 (for $(r = 5, \beta = 0.1)$), yielding a ratio of 3.22. Accordingly, halving the nominal level of significance is appropriate.
7. In fact, Hsu's MCB is equivalent to Dunnett's multiple comparisons with a control (MCC) procedure, where the control condition is selected as the factor setting (i.e., (r, β) pair) with the minimum observed mean (Lawson, 2010, p. 47). This has favorable implications for the robustness of Hsu's MCB. Specifically, when the design is balanced, Dunnett's MCC is known to be robust against non-normality and unequal variance, provided the maximum to minimum variance ratio is less than 4:1 (Toothaker, 1993). As Toothaker (1993) notes, you can use Dunnett's MCC at $\alpha = 0.05$ "with little consequence of unequal variances if the maximum true α you would tolerate would be 0.075" (p. 61). Thus, as with our earlier two-way ANOVA, we halved our nominal FWER to 0.025.

References:

- Architecture Integration and Management Directorate (AIMD), Army Capabilities Integration Center (2014). *SV-3s from select defense programs 2011-2014*. [Data file]. Retrieved from the Marine Collaborative Architecture Environment (MCAE).
- Banavar, J. R., Maritan, A., & Rinaldo, A. (1999). Size and form in efficient transportation networks. *Nature*, 399(6732), 130-132. doi:10.1038/20144
- Barabási, A. L. (2015). Non-linear preferential attachment. In *Network Science* (Chapter 5, Section 7). Retrieved from http://barabasi.com/networksciencebook/content/book_chapter_5.pdf
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512. doi:10.1126/science.286.5439.509
- Barabási, A. L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1), 173-187. doi:10.1016/S0378-4371(99)00291-5
- Bevan, M. F., Denton, J. Q., & Myers, J. L. (1974). The robustness of the *F*test to violations of continuity and form of treatment population. *British Journal of Mathematical and Statistical Psychology*, 27(2), 199-204. doi:10.1111/j.2044-8317.1974.tb00540.x
- Blanchard, B., & Fabrycky, W. (1998). *Systems engineering and analysis* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Bolten, J. G., Leonard, R. S., Arena M. V., Younossi, O., & Sollinger, J. M. (2008). *Sources of Weapon System Cost Growth - Analysis of 35 Major Defense Acquisition Programs*. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG670.pdf
- Borgatti, S. P. (2009). Two-Mode Concepts in Social Network Analysis. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 8281-8291). New York, NY: Springer.
- Budget Control Act of 2011, Pub. L. No. 112–25. 125 Stat. 240 (BCA). (2011). Retrieved from <http://www.gpo.gov/fdsys/pkg/PLAW-112publ25/html/PLAW-112publ25.htm>
- Cancho, R. F., Janssen, C., & Solé, R. V. (2001). Topology of technology graphs: Small world patterns in electronic circuits. *Physical Review E*, 64(4), 046119. doi:0.1103/PhysRevE.64.046119
- Chairman of the Joint Chiefs of Staff (CJCS) (2012a, January 19). *Manual for the Operation of the Joint Capabilities Integration and Development System*. Retrieved from <https://dap.dau.mil/policy/Documents/2012/JCIDS%20Manual%2019%20Jan%202012.pdf>
- Chairman of the Joint Chiefs of Staff (CJCS) (2012b, September 20). *JCIDS Manual Errata*. Retrieved from <https://dap.dau.mil/policy/Documents/2012/JCIDS%20Manual%20Errata%20-%2020%20Sept%202012.pdf>
- Chairman of the Joint Chiefs of Staff (CJCS) (2015, February 12). *Manual for the Operation of the Joint Capabilities Integration and Development System*. Retrieved from https://dap.dau.mil/policy/Documents/2015/JCIDS_Manual_-_Release_version_20150212.pdf
- Chairman of the Joint Chiefs of Staff (CJCS) (2018, August 31). *Manual for the Operation of the Joint Capabilities Integration and Development System*.

- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4), 661-703. doi:10.1137/070710111
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal - Complex Systems*, 1695. Retrieved from <http://igraph.org>
- Dabkowski, M., Valerdi, R., & Farr, J. (2014). Exploiting Architectural Communities in Early Life Cycle Cost Estimation. *Procedia Computer Science*, 28, 95-102. doi:10.1016/j.procs.2014.03.013
- Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008. doi:10.1088/1742-5468/2005/09/P09008
- Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises Tests. *The Annals of Mathematical Statistics*, 28(4), 823-838. Retrieved from <http://www.jstor.org/stable/2237048>
- Department of Defense (DoD). (n.d.). *Spotlight - National Defense Strategy*. Retrieved October 18, 2021, from <https://www.defense.gov/Spotlights/National-Defense-Strategy/>
- Department of Defense (DoD). (2018). *Summary of the 2018 National Defense Strategy of the United States of America*. <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>
- Department of Defense Deputy Chief Information Officer (DoD DCIO) (2010, August). *The DoDAF Architecture Framework Version 2.02*. Retrieved from http://dodcio.defense.gov/Portals/0/Documents/DODAF/DoDAF_v2-02_web.pdf
- Donaldson, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *Journal of the American Statistical Association*, 63(322), 660-676. doi:10.1080/01621459.1968.11009285
- Dorogovtsev, S. N., & Mendes, J. F. (2003). Non-equilibrium networks. In *Evolution of networks: From biological nets to the Internet and WWW* (Chapter 5). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780198515906.003.0006
- Dowlatshahi, S. (1992). Product design in a concurrent engineering environment: an optimization approach. *International Journal of Production Research*, 30(8), 1803-1818. doi:10.1080/00207549208948123
- Efron, B. (2012). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association*, 99(465), 96-104. doi:10.1198/016214504000000089
- Gillespie, C. S. (2015). Fitting Heavy Tailed Distributions: The powerLaw Package. *Journal of Statistical Software*, 64(2), 1-16. Retrieved from <http://www.jstatsoft.org/v64/i02/>
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821-7826. doi:10.1073/pnas.122653799
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, 237-288. Retrieved from <http://www.jstor.org/stable/1169991>

- Government Accountability Office (GAO) (2011). *DoD Cost Overruns: Trends in Nunn-McCurdy Breaches and Tools to Manage Weapon Systems Acquisition Costs*. Retrieved from <http://www.gao.gov/assets/130/125861.pdf>
- Gupta, H. (1970). Partitions—a survey. *Journal of Research of the National Bureau of Standards*, 74B(1), 1-29.
- Hankin, R. K. S. (2006). Additive integer partitions in R. *Journal of Statistical Software - Code Snippets*, 16(1), 1-3.
- Hsu, J. C. (1984). Constrained simultaneous confidence intervals for multiple comparisons with the best. *The Annals of Statistics*, 12(3), 1136-1144. doi:10.1214/aos/1176346732
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481), 248-258. doi:10.1198/016214507000000446
- International Organization for Standardization (ISO). (2011). *Systems and software engineering – Architecture description* (ISO/IEC/IEEE Standard No. 42010). Retrieved from <https://www.iso.org/standard/50508.html>
- Jeong, H., Néda, Z., & Barabási, A. L. (2002). Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4), 567–572.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Lawler, E. L. (1976). *Combinatorial optimization: networks and matroids*. Mineola, NY: Courier Corporation.
- Lawson, J. (2010). *Design and Analysis of Experiments with SAS*. New York, NY: CRC Press.
- Minitab 17.2.1 Statistical Software (2015). [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com)
- Montgomery, D. C. (2005). *Design and Analysis of Experiments* (6th ed.). New York, NY: John Wiley and Sons.
- Myers, J. L., Well, A., & Lorch, R. F. (2010). *Research design and statistical analysis*. New York, NY: Routledge.
- Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2), 025102. doi:10.1103/PhysRevE.64.025102
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113. doi:10.1103/PhysRevE.69.026113
- Peña, M., & Valerdi, R. (2015). Characterizing the Impact of Requirements Volatility on Systems Engineering Effort. *Systems Engineering*, 18(1), 59-70. doi:10.1111/sys.21288
- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2), 131-134. <http://dx.doi.org/10.1007/s100510050359>.
- Rubner, Y., Tomasi, C., & Guibas, L. J. (1998, January). A metric for distributions with applications to image databases. In *Proceedings from Sixth International Conference on Computer Vision*, 59-66. Bombay, India: IEEE.

- Shore, J., & Lubin, B. (2015). Spectral goodness of fit for network models. *Social Networks*, 43, 16-27. <http://dx.doi.org/10.1016/j.socnet.2015.04.004>
- Toothaker, L. E. (1993). *Multiple comparison procedures* (No. 89). Newbury Park, CA: Sage.
- Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)) (2020, January 23). *Operation of the Adaptive Acquisition Framework* (DoD Instruction 5000.02). Retrieved from <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf>
- Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)) (2020, August 6). *Major Capability Acquisition* (DoD Instruction 5000.85). Retrieved from <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500085p.pdf>
- Under Secretary of Defense for Research and Engineering (USD(R&E)) (2020, May). *Modular Open Systems Approach (MOSA) Reference Frameworks in Defense Acquisition Programs*. Retrieved from <https://ac.cto.mil/wp-content/uploads/2020/06/MOSA-Ref-Frame-May2020.pdf>
- Under Secretary of Defense for Research and Engineering (USD(R&E)) (2020, November 18). *Engineering of Defense Systems* (DoD Instruction 5000.88). Retrieved from <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500088p.pdf>
- Valerdi, R. (2008). *The Constructive Systems Engineering Cost Model (COSYSMO): Quantifying the Costs of Systems Engineering Effort in Complex Systems*. Saarbrücken, Germany: VDM Verlag.
- Valerdi, R., Dabkowski, M., & Dixit, I. (2015). Reliability Improvement of Major Defense Acquisition Program Cost Estimates – Mapping DoDAF to COSYSMO. *Systems Engineering*. 18(5), 530-547. doi:10.1002/sys.21327
- Wasserman, S., & Faust, K. (2009). *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.
-

Matthew Dabkowski is a Colonel in the US Army and an Associate Professor in the Department of Systems Engineering at the U.S. Military Academy. He has served in the United States Army for 24 years as an Infantry Officer, Operations Research Analyst, and Academy Professor. He is a graduate of West Point (B.S. in Operations Research) and the University of Arizona (M.S. in Systems Engineering and Ph.D. in Systems and Industrial Engineering). His research interests include applied statistics and simulation modeling.

Arthur Middlebrooks is a Major in the US Army and an Assistant Professor in the Department of Systems Engineering at the U.S. Military Academy. He has served in the United States Army for 11 years as an Armor Officer and Operations Research Analyst. He is a graduate of West Point (B.S. in Systems Engineering) and the Massachusetts Institute of Technology (M.S. in Engineering and Management). His research interests include value-based decision analysis and system dynamics simulation.

Ricardo Valerdi is a Distinguished Outreach Professor at the University of Arizona in the Department of Systems and Industrial Engineering. His research focuses on systems engineering, cost estimation, and sports analytics. He was the Founder and Co-Editor-in-Chief of the *Journal of Enterprise Transformation* and was the Editor-in-Chief of the *Journal of Cost Analysis and Parametrics*. He served on the Board of Directors of the International Council on Systems Engineering (INCOSE) and is a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE) and a Fellow of INCOSE. He won the Frank Freiman Award from ICEAA for his contributions to cost model development, is a recipient of the USC Center for Systems & Software Engineering Lifetime Achievement Award, and was elected as a foreign member to the Mexican National Academy of Engineering. He received a Ph.D. in Industrial and Systems Engineering from the University of Southern California.



The International Cost Estimating and Analysis Association is a 501(c)(6) international non-profit organization dedicated to advancing, encouraging, promoting and enhancing the profession of cost estimating and analysis, through the use of parametrics and other data-driven techniques.

www.iceaaonline.com

Submissions:

Prior to writing or sending your manuscripts to us, please reference the JCAP submission guidelines found at

www.iceaaonline.com/publications/jcap-submission

Kindly send your submissions and/or any correspondence to
JCAP.Editor@gmail.com

International Cost Estimating & Analysis Association

4115 Annandale Road, Suite 306 | Annandale, VA 22003

703-642-3090 | iceaa@iceaaonline.org