Erica Gralla[1]
Associate Professor
Department of Engineering Management
and Systems Engineering,
The George Washington University,
Washington, DC 20052
e-mail: egralla@gwu.edu

Nikolai Joseph
Department of Engineering Management
and Systems Engineering,
The George Washington University,
Washington, DC 20052
e-mail: njoseph37@gwu.edu

Zoe Szajnfarber
Professor
Department of Engineering Management
and Systems Engineering,
The George Washington University,
Washington, DC 20052
e-mail: zszajnfa@gwu.edu

Fit As a Diagnostic Tool: An Analytic Review of Approaches to Measure Correspondence Between Technical and Organizational Architectures

Standardized design approaches such as those embodied by concurrent design facilities have many benefits, such as increased efficiency of the design process, but may also have hidden costs. Specifically, when their standardized organizational decomposition is a poor fit for the particular design problem, important design trades might be missed or poor decisions made. Before we can understand how this lack of fit impacts the design process, we must be able to empirically observe and measure it. To that end, this paper identifies measures of "fit" from the literature along with attributes likely to impact design process performance, then evaluates the measures to determine how well the measures can detect and diagnose potential issues. The results provide comparative insights into the capabilities of existing fit measures, and also build guidance for how the systems engineering and design community can use insights from the "fit" literature to inform process improvement. [DOI: 10.1115/1.4055150]

Keywords: conceptual design, design process, systems design, systems engineering, systems architecture

1 Introduction

The increasing complexity of engineered systems makes it critical to integrate expertise and knowledge from multiple disciplines throughout the design process. To that end, concurrent design approaches bring together multiple disciplines and speed up design iterations [1]. These approaches are popular, since when they work well, they support a fast and efficient design process [1–3]. However, there are some potential hidden costs which have not received enough attention. Specifically, this efficiency is enabled by decomposing the design problem in a "standard" way and routinizing the interactions across the decomposed subproblems [4,5]. When this imposed standard structure is a good fit for the particular design problem, efficiency improves because designers can focus on the most difficult challenges without considering irrelevant information or unnecessary interdependencies [6,7]. However, as we argued in Ref. [8], "even minor mismatches between the organizational decomposition (people and tasks) and product decomposition (the problem being solved) can cause designers to miss important trades and make poor choices" [7,9,10]. With the popularity of efficiency shortcuts like those embodied by concurrent design facilities, it is important to understand these potential hidden costs. This will enable us to enjoy the advantages without the penalties.

Before we can understand how a lack of fit between the technical product and the imposed organization impacts the design process, we must be able to empirically observe and measure it. To that end, this paper takes a critical look at how the concept of "fit" has been defined and operationalized in the literature. Here we draw extensively on work related to the so-called *mirroring hypothesis*, which states that the structures of a technical product and the organization designing it should mirror each other [9] (see Ref. [9] for a detailed overview). Motivated by the well-documented failures that arise when a lack of fit is not identified quickly enough [7,11], the mirroring literature has focused on characterizing the phenomenon and collecting evidence to substantiate it. However, there is not yet a systematic approach to formalizing the measurement of fit. As a result, there are a wide variety of measures that all purport to measure fit, but take very different approaches to doing so [11–13]. There is a need to systematically study existing measures to assess whether, and under what conditions, each can be appropriately adopted for our proposed purpose: using fit measures to diagnose issues in the design process and to provide guidance for how to improve it.

To accomplish this, we identified several representative fit measures from the literature, as well as a list of attributes of fit that the literature expects to be important to design process performance. We then tested how well the measures can detect and diagnose these potential issues. Specifically, we developed a series of generic product and organizational architecture pairs and systematically perturbed them to represent the predefined fit attributes (or problems). We then evaluated each pair with each of the fit measures and studied the measures' ability to "see" different kinds of fit problems. The results provide comparative insights into the capabilities of existing fit measures, and also build guidance for how the systems engineering and design community can use insights from the "fit" literature to inform process improvement.

2 Background and Literature Review

This section frames the practical and theoretical basis for studying the implications of fit (or lack thereof), reviews dominant conceptual definitions and operational formulations of fit, and synthesizes the attributes of fit that a good measure needs to be able to detect.

In practice, complex systems are designed by many specialized engineers working together in a coordinated way. These engineers need to communicate to transfer information about shared design

---

variables and resolve design conflicts when they arise. The structure of this communication is influenced by the technical system being designed. For example, the power subsystem is sized based on its anticipated utilization by every other powered subsystem. If one subsystem needs more power than initially planned, design trades must be made with other subsystems, either moderating power usage or increasing the availability of power. This technical dependency creates a need for an organizational dependency, whereby engineers responsible for each subsystem design have a path for communication with one another. On a small design team, this is straightforward and happens naturally, but as the size of the team increases (often reaching hundreds or thousands of engineers on large projects), particular communication paths and liaising functions need to be defined in advance. If every path is enabled, teams can spend all their time talking (and none working), but if too few are enabled, costly problems may not be identified until later. For example, continuing the above example, if two subsystems have both blown their power budgets and do not communicate regularly, they may each (erroneously) assume that they can steal margin from the other. The sections below formalize the theory that defines these interactions.

2.1 Mirroring As a Strategy for the Design of Complex Systems.
The question of how closely the technical and organizational systems need to be matched to support effective design has been studied across multiple disciplines. In the management literature, for example, this question was motivated by a number of well-documented firm failures resulting from mismatches between established organizational structures and the architecture of a new technology [7,11]. From this effort emerged the realization that "the formal structure of an organization will (or should) 'mirror' the design of the underlying technical system" (quoted from Ref. [9], which cites Refs. [4,7,14–16]). Colfer and Baldwin [9] labeled this the "Mirroring Hypothesis." In its descriptive form it states that: "In a complex technical system, organizational ties are more likely to exist in places where technical interdependencies are present (or dense). Organizational ties are likely to be absent where technical interdependencies are absent (or sparse)" [9]. In its normative form, it states that: "Mirrored systems achieve good performance outcomes, and unmirrored systems achieve poor performance outcomes" [9].

From a design process perspective, the key idea is that the core organizational function of coordinating interdependent tasks is fundamental to designing complex systems [17,18]. To that end, Baldwin et al. [4,9] view design as an organizational

Fig. 1 Representation of product and organizational architectures as design structure matrices, with potential "fit" problems highlighted

problem-solving process that aims to conserve scarce cognitive resources. A dominant strategy for achieving this is by partitioning a complex problem into loosely coupled subproblems, which Simon [19,20] noted reduces complexity. From an organizational perspective, Thompson [17] argued that "reciprocally interdependent" tasks should be placed within a common organizational group to achieve efficiency in the face of underlying complex interdependencies. From a product perspective, Parnas [6] explained that it is easier for development work to be split across groups when their tasks are independent, enabling them to work in parallel.

So far, the work on mirroring has been largely observational, documenting the prevalence of mirroring in the field [4,9,21,22] or measuring the effects of mirroring (or lack thereof) on organizational performance [23–25]. However, the underlying theory also provides a basis for intentionally designing technical dependencies to match organizational unit boundaries [6,26]. If technical modules are isolated from other modules within a framework of "design rules," complex systems can be built efficiently without complex coordination [4].

2.2 Conceptual Definition of "Fit".
Colfer and Baldwin [9] documented an extensive body of literature that nominally compares the structure of a technical product to the associated organization that created it.

In formulating the mirroring hypothesis, Baldwin and colleagues [4,9] define each of the conceptual components of a measure of fit. Their definition of the *technical architecture* is largely consistent with the literature on design structure matrices (DSMs) [27–29], namely a representation of the system in terms of "what depends on what." By convention, the system is drawn as an $N \times N$ matrix where rows and columns correspond to system components and off-diagonal entries denote dependencies; see Fig. 1. Below the diagonal, dependencies are feedforward, and above the diagonal, they are feedback. Two elements are considered interdependent when there is both a feedforward and feedback relationship. Per Baldwin's theory of mirroring [30], the basic units of the technical system are *tasks* (these are the components in the DSM) and *transfers* (the off-diagonal entries). Transfers specify how material, energy, and information [31] in the system must be exchanged by dependent tasks to achieve a technological goal [30]. For example, in an observing satellite, the productive resolution of the telescope is limited by the capacity to downlink data to a ground station.

They define the *organizational architecture* as the scheme by which tasks in the technical architecture are assigned to organizational resources (e.g., people or teams) [9,30]. This representation is an application of task contingency theory to complex technical systems [17,18,32–35]. Adopting a DSM framework, each element now corresponds to the unit of assignment (e.g., a person, team, or business unit), with off-diagonal entries denoting organizational ties (see Fig. 1). (Here we will focus on organizational in the context of the firm, but the same logic extends to enterprises and supply chains as well, defining the elements and ties at the appropriate unit of analysis.) The function of an *organizational tie* is to enable conflict resolution among interdependent tasks. Organizational ties can take multiple forms depending on the context. Common ones include *co-location* to facilitate direct interaction [36], *communication links* including both the mechanics and understanding required to exchange information across boundaries [30], *social ties* between actors to foster collaboration, which can include past working relationships or employment contracts, and *conflict-resolution processes* within an organization [30].

Mirroring is then a measure of *fit* between the technical and organizational architectures. Perfect mirroring would be seen as an exact overlap between the technical and organizational DSMs. In other words, every technical transfer would by matched by an organizational tie. Imperfect mirroring might include organizational ties without corresponding product transfers (which we term "overfit") and/or product transfers without corresponding organizational ties (which we term "underfit"). Both cases are illustrated in
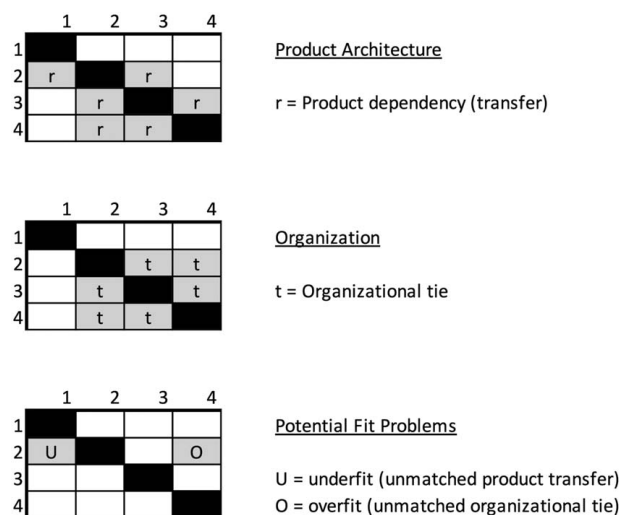
Fig. 1. In practice, perfect mirroring rarely exists, therefore in assessing the evidence for mirroring, Colfer and Baldwin [9] applied a standard of strong "correlation" between transfers and ties, classifying studies as demonstrating high mirroring, partial mirroring and no correlation. Here, partial mirroring usually indicated that part of the system exhibited mirroring, while the rest did not, such as in a core-periphery structure.

2.3 Existing Operationalizations of Fit. As we described in the previous section, the literature provides a clear theoretical definition of mirroring and "fit." However, there is wide variation across the literature in how mirroring is measured—i.e., in how the concept of "fit" is operationalized.

Colfer and Baldwin's [9] comprehensive analysis of the literature focused on the extent to which each study offered support for the mirroring hypothesis. As a result, they only noted whether the approach to measurement was qualitative or quantitative and whether the assessment tested the descriptive or normative version of the hypothesis. They explicitly did not examine the methodology used to compare product and organizational structures as it was not relevant to their research question. However, if the goal is to diagnose the extent of observed mirroring, taking stock of existing measures is critical.

In the comprehensive sample developed by Colfer and Baldwin [9], and the associated unpublished technical appendix, we observed three main approaches to measurement of mirroring, which we illustrate below with examples of each.

The first approach to measuring mirroring assesses the correspondence between all of the technical transfers and their associated organizational ties across the system. Typically, each of the technical and organizational architectures is represented as a DSM (or the corresponding network), and the cells (edges) of the DSMs are compared to assess whether each technical dependency is associated with an organizational tie and vice versa. A quantitative measure of the extent of correspondence is typically computed. Morelli et al. [37] is an example of this type. They conducted interviews and weekly surveys at a computer hardware company. Based on an *ex ante* analysis of the technical architecture, they were able to predict coordination-oriented communication among team members with 81% accuracy. This suggested that technical dependencies were strongly associated with organizational ties. Comparing the locations of technical interdependencies and organizational ties has been a popular approach among descriptive within-firm studies (e.g., Refs. [25,37–40]).

The second approach assesses fit by comparing summary measures of the technical and organizational architectures, such as modularity or network centralization, rather than the correspondence of particular technical transfers and their associated ties (as in the previous approach). For example, Fixson and Park [11] studied the bicycle industry and found that changes to the modularity of the technical architecture prompted restructuring of the industry. To make this determination, they assessed the modularity of the product and value chain separately before comparing them. Across studies, the specific choice of summary measure varies (e.g., Zhou [12] examined a product's task complexity and decomposability and an organization's divisionalization and hierarchy; and Parraguez [41] compared the centralization and clustering of information networks). Nonetheless, the common approach across all of these papers is to compute a summary measure of each network and assess the correspondence of those summary values.

The third approach focuses more narrowly on unpacking particular instances of technical-organizational correspondence or lack thereof, rather than assessing correspondence across the entire architecture (as in the previous two approaches). Several qualitative studies elaborate key aspects of the system where rich interactions were needed (or lacking). Gulati and Puranam [42] is an example of this type. They performed a qualitative case study of major organizational restructuring at Cisco, a networking firm. They observed that when the new organization left key technical interdependencies unaddressed, employees informally maintained legacy information practices. Eventually, parts of the old structure were integrated into the new one to maintain desired cross-functional collaboration (enabled by mirroring). In this type of study, the full system of technical and organizational architectures may not be explicitly represented; instead, the focus is on an in-depth explanation of what the match (or mismatch) of a specific set of transfers-and-ties enables (or limits).

While the last approach has been important in understanding the phenomenon of mirroring, it does not constitute a systematic approach to measuring fit; therefore, in the remainder of the discussion we will focus on measures from the first two families of fit measurement.

2.4 Attributes of Fit That Should Be Detected to Diagnose Mirroring. While pluralism of measurement approach is appropriate, and even preferred, when the goal is to establish the phenomenon of mirroring, we contend that it can be detrimental for studies that seek to use the mirroring construct to diagnose potential operational issues [43] or identify a need to transform an organization to match a new product [7]. In such cases, there is a need for a more careful consideration of what a measure captures or "sees."

Our goal with this work is to identify a fit measure capable of diagnosing fit in a given product–organizational pair as a basis for improving the design process. Such a measure should be able to satisfy the following conditions:

(1) Distinguish categorically between perfect fit and imperfect fit. Baldwin [30] defines perfect fit as a one-to-one match between each technical transfer and organizational tie. Therefore, a valid measure must consistently measure perfect fit (i.e., as zero difference) and distinguish it from every other arrangement.

(2) Distinguish categorically between overfit, underfit, and mixed. In explaining instances where the mirror "broke," Colfer and Baldwin [9] note that "over" and "under" fit have different performance implications. Overfit, the situation where organizational ties exist in the absence of a technical transfer, can be a valid organizational choice, since it is often efficient to have a co-located interdisciplinary team, even if not every interaction is expected. On the other hand, underfit, where technical transfers are not supported by organizational ties, generally leads to negative design outcomes since design conflicts might arise unobserved. Finally, Camuffo and others [44,45] have explored more complex hybrid conditions of "misted" mirrors, explaining the implications of partial mirroring, related to the nature of the dependency. Therefore, it is important to know which kind of mismatch is at play, because the kind of mismatch impacts the desired corrective action.

(3) Report the extent of fit: for example, whether one situation shows more or less fit than another (in the same direction). Recognizing that perfect fit rarely manifests in practice, Baldwin [9,30] proposed a standard of "extent" of fit based on correlation among technical and organizational dependency structures. In terms of using fit as a design heuristic, it is important to know what level of mismatch becomes problematic. Therefore, a useful measure of fit must show scale, with sufficient spread to interpret meaningful differences. This means that if zero is a perfect fit, for most normalized measures, an opposite fit should show up as one.

(4) Weight the sources of lack of fit. Within the design literature, there is an understanding that certain types of dependencies are more likely to create both more, and more disruptive, re-work iterations [43]. This is the basis of many of the DSM reordering algorithms that focus on maximizing lower-triangular or diagonal-ness (see Ref. [29] for a comprehensive review). As a result, it is desirable for a good measure of fit to identify mismatches near the diagonal as less problematic than those far from it.

Table 1 Summary of measures, with the systems to which they were applied in the original references

Measure	Description	Reference(s)	System(s)
Alignment (α_2, α_3)	Compare DSMs for each of the product and organization cell by cell; report the percent of unmatched product transfers (α_2) and unmatched organizational ties (α_3) out of all cells in the DSM	[39,47]	Aircraft engine
Coordination deficit (β)	Compare DSMs for each of the product and organization cell by cell; report the percent of needed coordination that is missing (unmatched product transfers out of all transfers)	[25]	Automobiles
Network: centralization (γ_n) & clustering (γ_s)	Reports the difference between summary measures of the product and organization DSMs. Centralization is the extent to which some nodes in the DSM-defined network are more central than others. Clustering reflects the extent to which the neighbors of a node are connected to one another	[41]	Renewable energy plant
Modularity (δ)	Reports the difference between the modularity scores of the product and organizational DSMs	As a measure of fit: [11,12]. Modularity metric: [48–51]	U.S. equipment manufacturers, bicycles

(5) Provide a measure that is easily interpretable, such that its result has an intuitive meaning and its scale is sufficient to distinguish different fit conditions.

3 Methods

To assess whether, and the extent to which, existing measures in the literature are able to diagnose fit to inform design, we developed a DSM-based testbed as suggested in Ref. [46]. As elaborated in the sections below, we selected four measures that represent the two "families" of measures described in Sec. 2.3 and applied them to a series of synthetically generated product and organizational DSMs. The DSMs were generated to embody systematic variation along each of the types of mismatch enumerated in Sec. 2.4. The resultant fit scores allow us to explore the advantages and disadvantages of each family and each measure for diagnosing potential mirroring-related challenges.

In the following subsections, all four measures are first introduced, then the generation of test DSMs and their variations is described, and finally five criteria are advanced for evaluating what kinds of fit problems each measure can diagnose, within the DSM-based testbed.

3.1 Measures. This section presents the four measures that will be tested in the subsequent analysis. To select these measures, we relied on Colfer and Baldwin's [9] systematic review of the literature, which included papers that used a variety of approaches to measuring fit. We identified three main families of related approaches, described in Sec. 2.3, then chose two measures from each of the first two families that represented different approaches within each family. (The third family involved deeper qualitative comparisons rather than comprehensive measures of fit, so it was not relevant to our goals.) The first two measures are from the family that assesses the correspondence between technical transfers and their associated organizational ties by comparing the cells (edges) in a DSM (or network), with a DSM representing each of the product and organizational architectures. The third and fourth measures are from the family that compares summary measures of the overall architectures rather than characteristics of specific interdependencies or components. The intent with this selection was not to be comprehensive but rather to select contrasting measures in each category to support a rich comparison.

The measures are explained in more detail in the sections below and summarized in Table 1. In all cases, they are defined to make comparisons between a product DSM and an organization DSM and to produce a score (or two scores) that measure the mirroring or fit between these two matrices. The entries in the product DSM indicate the presence or extent of transfers (dependencies) among product components or subsystems, and the entries in the organization DSM indicate the presence or extent of organizational ties. For simplicity, we assume that the matrices are symmetric and that they are the same size. Let the product DSM be an $N \times N$ matrix $P = [P_{ij}]$ and the organization DSM be an $N \times N$ matrix $G = [G_{ij}]$.

Note that implementing these measures in real systems might require further work to either transform data into this simplified format or to adjust the measures to work on less constrained types of DSMs. For example, product and organization DSMs may not be equally detailed, may not be the same size, and/or may not be symmetric [46,52]. To use the simplified measures described in this paper, one DSM may be expanded to match the other (with little impact on the results), and both may be made symmetric (which could indeed impact the results by assuming two-way dependencies or none where there are one-way dependencies—so this must be done carefully). On the other hand, future work could refine these measures to avoid requiring these transformations. For this paper, however, the value of a simplified testbed is that it "controls for" these complex variations and simplifies the interpretation of the results, to support our purpose of evaluating the measures' ability to "see" fit problems.

3.1.1 Alignment. The *alignment* measure is adapted from the work of Sosa and colleagues [39,47]. Although the approach varies slightly in their different papers, the general idea is to develop DSMs for each of the product and the organization, in which the entry in each cell indicates the presence or absence of a transfer or tie, respectively. These two matrices are then overlaid and compared. Many different analyses can be performed on these results, but the most straightforward is to report the number of matched and unmatched dependencies in four categories: matched present dependencies (transfers and ties), matched missing dependencies, dependencies present in the organization but not the product, and dependencies present in the product but not the organization. We use this concept to develop our *alignment* measure of fit.

Because this measure operates on unweighted matrices that indicate only the presence or absence of dependencies, the DSMs are first converted to unweighted DSMs P_u and G_u by including an arc only if the entry is greater than a threshold. Then, the alignment α can be computed in a few ways.

The first and simplest, α_1, is, roughly, the chance that the product DSM and the organizational DSM will not align when they should, or the number of mismatched cells divided by the total number of cells (not including the diagonal). Thus, we first compute the difference D between P_u and G_u, $D = |P_u - G_u|$, then normalize to find α_1

$$\alpha_1 = \frac{\sum_{i,j:i \neq j} D_{ij}}{N \times N - N} \tag{1}$$

The second way to measure alignment produces two numbers: α_2, the chance that there will be unmatched product transfers, and α_3, the chance that there will be unmatched organizational ties. We first compute the differences D^P and D^G

$$D^p = \max(P_u - G_u, 0), \quad D^G = \max(G_u - P_u, 0) \quad (2)$$

then the measures α_2 and α_3

$$\alpha_2 = \frac{\sum_{i,j:i \neq j} D_{ij}^P}{N \times N - N}, \quad \alpha_3 = \frac{\sum_{i,j:i \neq j} D_{ij}^G}{N \times N - N} \quad (3)$$

These measures each range between 0 and 1.

3.1.2 Coordination Deficit.
The *coordination deficit* measure is adapted from Gokpinar et al. [25]. They create weighted networks to represent the product and the organization, normalize them, and compare the extent to which product transfers are unmatched by organizational ties, for each subsystem. We adapt this measure to generate a network-wide value rather than a value for each subsystem.

Specifically, the product and organizational DSMs P and G are first normalized to obtain P^n and G^n, respectively, by dividing each link by the total weight of all the links in its network. (This is intended to enable the two networks' weights to be measured on very different scales)

$$P_{ij}^n = \frac{P_{ij}}{\sum_{i,j} P_{ij}/2}, \quad G_{ij}^n = \frac{G_{ij}}{\sum_{i,j} G_{ij}/2} \quad (4)$$

Then, the coordination deficit β_i is defined for each subsystem i (each row of the DSM) to measure the extent to which product transfers are unmatched by organizational ties

$$\beta_i = \sum_j \max(P_{ij}^n - G_{ij}^n, 0) \quad (5)$$

Note that this measure focuses only on the potentially more problematic situation in which organizational ties do not exist to coordinate product transfers (dependencies), but does not "count" the less problematic cases in which there are unneeded organizational ties (i.e., ties without product transfers to coordinate).

We adapt this subsystem-level measure into a network-level measure by defining the overall network-level coordination deficit β as the sum of all the deficits network-wide, dividing by two to avoid counting each deficit twice

$$\beta = \frac{1}{2} \sum_i \beta_i = \frac{1}{2} \sum_{i,j} \max(P_{ij}^n - G_{ij}^n, 0) \quad (6)$$

This measure ranges from 0 (no deficit) to 1 (high deficit).

3.1.3 Centralization and Clustering.
The *network* measures, *centralization* and *clustering*, were adapted from the work of Parraguez et al. [41]. They use measures of clustering and centralization in combination to characterize information flows in a communication network. Here, we use the same measures to compare whether the organization's planned information flows match the product's needed information flows.

First, consider *centralization*. Eigenvector centralization is a network-level measure that indicates the extent to which some nodes are more central than others. A high value means "only one or a few nodes intermediate most information exchanges" [41]. Each node's centrality, in turn, is computed from the eigenvector centrality, which measures how central or influential a node is; connections to other influential nodes elevate a node's centrality.

The eigenvector centralization is computed in the following manner [41,53]. Let C_A be the network centralization value of network A. Then,

$$C_A = \frac{\sum_i [C_A(n*) - C_A(n_i)]}{\max \sum_i [C_X(n*) - C_X(n_i)]} \quad (7)$$

$C_A(n_i)$ is the node-level eigenvector centrality value of node i, and $C_A(n*)$ is the maximum of all the node-level eigenvector centrality measures in network A. Then, the numerator is a sum of the difference between the maximum and observed node-level centrality measures. The denominator is that same quantity but for a benchmark network X, a maximally centralized network of the same size as A. For eigenvector centralization, this theoretical maximum is a network of the same size with only one edge between two of the nodes (for other centrality measures, it is a star). The resulting value for the denominator is $\sqrt{2}/2(N-2)$.

The eigenvector centralization scores are computed for both the product DSM, C_P, and the organization DSM, C_G. These values range between 0 and 1. Then, the centralization measure γ_n is the difference between the centralization scores for each network

$$\gamma_n = C_P - C_G \quad (8)$$

This value ranges from -1 to 1. The magnitude of the difference reflects how different the centralization is across the networks. The sign of the difference indicates which network is more centralized.

The weighted clustering coefficient is first computed at the node level. It reflects the extent to which the neighbors of a node are connected to one another [54,55]. Next, the network-level measure of clustering is found as the average of the clustering coefficients of all the nodes in the network [55]. The weighted clustering coefficient WCC is found separately for P and G, then the degree of fit γ_s is the difference between them

$$\gamma_s = WCC(P) - WCC(G) \quad (9)$$

The value ranges from -1 to 1. The magnitude of the difference reflects how different the clustering is across the networks. The sign of the difference indicates which network is more clustered. (The measure is computed with available MATLAB code [56].)

3.1.4 Modularity.
The *modularity* measure is inspired by several mirroring studies that consider the relative modularity of the product and organizational DSMs, respectively [11,12]. However, the modularity measures in these studies are not easily compatible with the DSM-based format of our test cases, so we instead implemented a modularity metric that is the basis of much modularity work in the design literature [48–51].

Specifically, modularity is computed from a metric described by Guo and Gershenson [48]. Modularity is computed for each matrix, P and G, separately, then the difference is reported as the measure of fit. We describe how to compute the modularity score for P, but the calculations are performed in exactly the same way for G.

Computing the modularity score requires as input both the DSM P and a description of the locations of modules, including M, the number of modules in the DSM, p_k, the index of the first element in the kth module, and q_k, the index of the last element in the kth module. Recall that N signifies the number of rows (and columns) in the DSM.

This measure is defined for unweighted matrices, so a weighted matrix P is first converted to an unweighted matrix P^u by including an arc only if the entry is greater than a threshold. We also assume that the entries along the diagonal are "1."

The modularity measure is as follows:

$$d_{P^u} = \frac{1}{M} \times \left(\sum_{k=1}^{M} \frac{\sum_{i=p_k}^{q_k} \sum_{j=p_k}^{q_k} P_{ij}^u}{(q_k - p_k + 1)^2} \right.$$
$$\left. - \sum_{k=1}^{M} \frac{\sum_{i=p_k}^{q_k} \left(\sum_{j=1}^{p_k-1} P_{ij}^u + \sum_{j=q_k+1}^{N} P_{ij}^u \right)}{(q_k - p_k + 1)(N - q_k + p_k - 1)} \right) \quad (10)$$

The measure d evaluates, roughly, the extent to which each module is completely filled in and the extent to which the areas outside the module (in the same rows) are *not* filled in. It subtracts the proportion filled outside the modules from the proportion filled inside the modules. The measure's value is 1 for perfect modularity,

0 for integral, and negative when it is denser outside the modules than inside.

To measure the fit δ between two DSMs P and G, we compute the modularity d_{P^u} and d_{G^u} for each, then subtract

$$\delta = d_{P^u} - d_{G^u} \qquad (11)$$

This measure ranges from -2 to 2, because the measure d ranges from -1 to 1. However, in most realistic cases, the value of d should be between 0 and 1 (since modules would typically be more densely connected within than outside the modules), and therefore δ should typically fall between -1 and 1.

3.2 Generating Test Design Structure Matrices and Variations.
To test how each of the measures reports fit, we designed test cases to have different degrees of fit.

First, we designed baseline test matrices to represent generic structures. We designed a 9×9 test matrix of a simple product with modules and a few off-module dependencies. Then, we designed two 30×30 test matrices, roughly modeled on the loosely coupled and tightly coupled DSMs from open-source and traditional software companies, described in Ref. [57].

Each baseline test matrix has a weighted and an unweighted version. The weighted versions, for simplicity, contain dependencies of magnitude 1, 3, 6, and 9. Their approximate distribution in the matrix is (1: 20%, 3: 50%, 6: 20%, and 9: 10%), based on the notion that most dependencies require a moderate amount of communication (3), some require a little more or less (6 or 1), while only a few require a lot (9). All six baseline test matrices are shown in Fig. 2.

We design four variations on these baseline test matrices, to represent organizations with different kinds of fit problems compared to the baseline product. They are illustrated in Fig. 3. Consider first an unweighted network. The *opposite* variation has edges where there are none in the baseline, and no edges where there are edges in the baseline. The *overfit* variation has the same edges as the baseline, plus additional edges where there are none in the baseline. The *underfit* variation has some of the baseline edges removed, but no additional edges. The *mixed* variation is a mix of over- and underfit, in that some baseline edges are missing and some new edges are added. For weighted networks, rather than adding or removing edges, a value m (for magnitude) is added to or subtracted from existing or potential edges. In the *overfit* variation, this could result in larger weights on existing edges (if m is added to an existing edge), or in new edges (if m is added to a previously nonexistent edge).

We generate these variations by randomly choosing the locations to add or remove edges (as applicable). The magnitude m is fixed for each set of test cases, but is varied as part of an experiment to learn its effect on the results.

We also vary the extent by which the modified test matrix is varied from the baseline, i.e., the number of edges that are changed (removed or added in the case of unweighted networks, and increased or decreased in the case of weighted networks). For the "low" case, we make modifications to n_l edges, where n_l is computed as the floor of 20% of the existing edges in the network. For the medium case, n_m is the floor of 50% of the existing edges in the network. Note that n depends on the baseline extent of communication in the network, so that more-connected networks have more edges modified.

Finally, we design an additional specialized variation in which all of the added edges are located in the most off-diagonal parts of the matrix (upper right and lower left corners).

Note that we did not reorder any of the DSMs after perturbing them, since the intention was to artificially create lack of fit in systematic ways.

3.3 Evaluating Measures of Fit.
Next we evaluated how well each of the four measures reports the extent of fit between the baseline and modified test matrices (representing the product and organizational DSMs). Each baseline test matrix (from Fig. 2) was compared to 100 randomly generated modified matrices for each of the variations described in Sec. 3.2 (*opposite, overfit, underfit,* etc.). Recall that the goal is to understand how well each of the proposed measures meets the five goals or criteria laid out in Sec. 2.4: the attributes that a measure should detect to diagnose mirroring or fit problems. To accomplish this, we formulated the following five evaluation criteria:

(1) To meet the first criterion, distinguishing categorically between mirrored or not mirrored, the measure should find a different value for a perfect fit than for any imperfect fit. To determine whether this criterion is met, we compare the result for the *perfect* case to that of all the others.

(2) To meet the second criterion, distinguishing categorically between overfit, underfit, and mixed, we should be able to define mutually exclusive ranges of the measure value that correspond to overfit, underfit, and mixed fit. To determine whether this criterion is met, we compare the results for the *overfit* test cases, the *underfit* test cases, and the *mixed* test cases, to determine whether they fall within mutually exclusive ranges.

(3) To meet the third criterion, reporting the extent of fit, the measure should find a higher (or more extreme) value when the fit is worse than when it is better. To determine whether this criterion is met, we compare the results for all of the *low* test cases to all of the *medium* test cases; the returned value(s) for the *medium* cases should be more extreme than those for the *low* cases. (Recall that the *low* case varied 20% of the existing edges and the *medium* case varied 50% of the existing edges.) In addition, we compare the results for different magnitudes m of change in weighted networks. The results for cases with higher-magnitude changes should be more extreme than for those with lower-magnitude changes.

(4) To meet the fourth criterion, reflecting differences due to the sources of the fit problems, the measure should find a higher (or more extreme) value when mismatches are located in the off-diagonal locations than when they are located in less problematic locations. To determine whether this criterion is met, we compare the results for the *off-diagonal* test case to those for the *overfit medium* case. In both cases, the same number of dependencies were added, but in the off-diagonal case, they were added specifically to the off-diagonal corners. The results for the *off-diagonal* test case should be more extreme than those of the *overfit medium* case.

(5) To meet the fifth criterion, a measure must be easily interpretable. This requires a more subjective evaluation. We consider whether the measure's value has an intuitive meaning and how much it reveals about the (lack of) fit. We also consider whether the scale is meaningful: for example, if perfect fit yields a zero, then opposite fit (or some other extremely poor fit condition) should yield an extreme value (such as a 1), depending on the measure.

4 Results and Discussion

The results from this computer experiment are summarized in Table 2. For each measure, we show the results for one illustrative test case in Fig. 4: 30×30 with loose coupling, unweighted. The results for this and three additional test cases are provided in Figures S1–S4, available in the Supplemental Materials (two additional unweighted matrices—30×30 with tight coupling and 9×9—and one weighted matrix—30×30 with loose coupling with change magnitude $m = 3$). Further additional variations described in Sec. 3 were also examined but are not shown, for brevity.
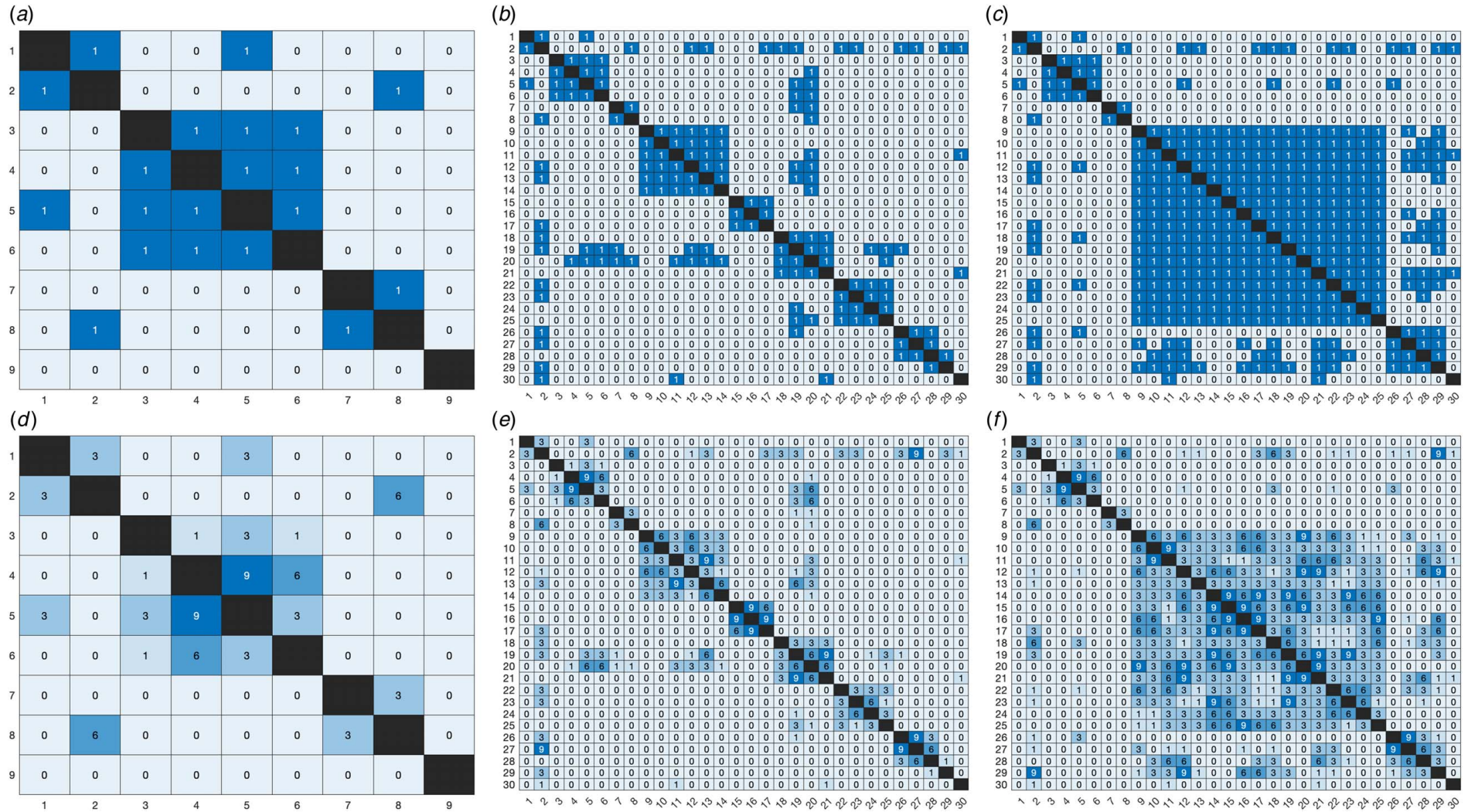
Fig. 2 Baseline test matrices. For weighted matrices, darker colors indicate larger weights: (*a*) test matrix 9 × 9, unweighted; (*b*) test matrix 30 × 30, loosely coupled, unweighted; (*c*) test matrix 30 × 30, tightly coupled, unweighted; (*d*) test matrix 9 × 9, weighted; (*e*) test matrix 30 × 30, loosely coupled, weighted; and (*f*) test matrix 30 × 30, tightly coupled, weighted.

(a)

	1	2	3	4	5	6	7	8	9
1	-	1	0	0	1	0	0	0	0
2	1	-	0	0	0	0	0	1	0
3	0	0	-	1	1	1	0	0	0
4	0	0	1	-	1	1	0	0	0
5	1	0	1	1	-	1	0	0	0
6	0	0	1	1	1	-	0	0	0
7	0	0	0	0	0	0	-	1	0
8	0	1	0	0	0	0	1	-	0
9	0	0	0	0	0	0	0	0	-

(b)

	1	2	3	4	5	6	7	8	9
1	-	0	1	1	0	1	1	1	1
2	0	-	1	1	1	1	1	0	1
3	1	1	-	0	0	0	1	1	1
4	1	1	0	-	0	0	1	1	1
5	0	1	0	0	-	0	1	1	1
6	1	1	0	0	0	-	1	1	1
7	1	1	1	1	1	1	-	0	1
8	1	0	1	1	1	1	0	-	1
9	1	1	1	1	1	1	1	1	-

(d)

	1	2	3	4	5	6	7	8	9
1	-	1	0	0	1	0	0	0	0
2	1	-	0	0	0	0	0	0	0
3	0	0	-	0	0	1	0	0	0
4	0	0	1	-	1	1	0	0	0
5	1	0	1	1	-	1	0	0	0
6	0	0	0	1	1	-	0	0	0
7	0	0	0	0	0	0	-	1	0
8	0	1	0	0	0	0	1	-	0
9	0	0	0	0	0	0	0	0	-

(c)

	1	2	3	4	5	6	7	8	9
1	-	1	0	0	1	0	0	0	0
2	1	-	0	0	0	0	0	1	1
3	0	0	-	1	1	1	0	0	0
4	0	0	1	-	1	1	0	0	0
5	1	0	1	1	-	1	0	0	0
6	0	0	1	1	1	-	0	1	0
7	0	0	0	0	1	0	-	1	0
8	0	1	0	0	0	0	1	-	0
9	0	0	0	1	0	0	0	0	-

(e)

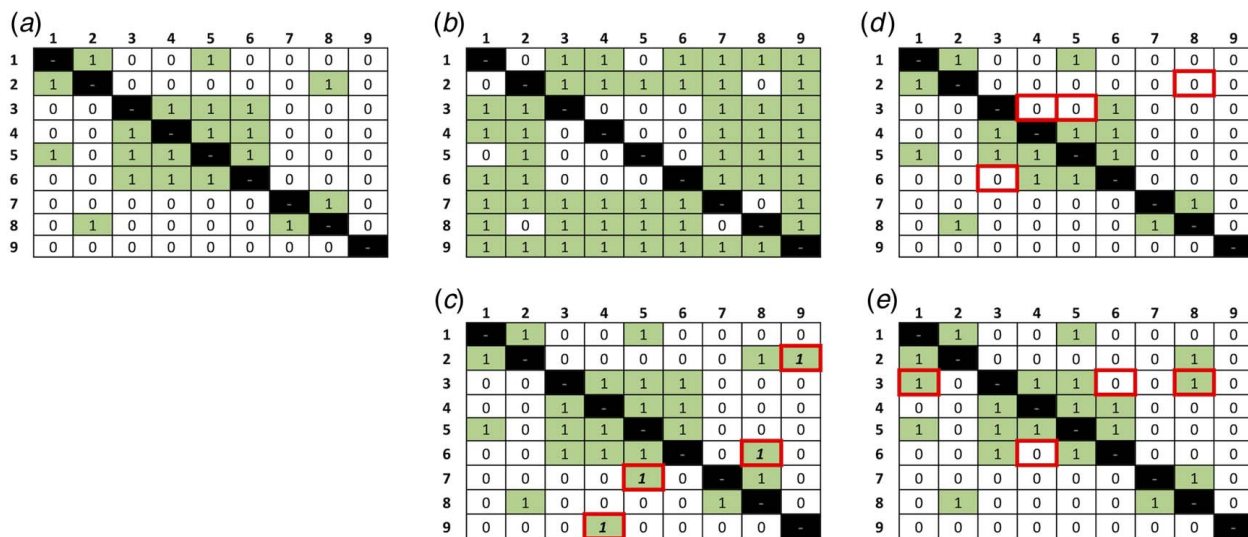|	1	2	3	4	5	6	7	8	9
1	-	1	0	0	1	0	0	0	0
2	1	-	0	0	0	0	0	1	0
3	1	0	-	1	1	0	0	1	0
4	0	0	1	-	1	1	0	0	0
5	1	0	1	1	-	1	0	0	0
6	0	0	0	0	1	-	0	0	0
7	0	0	0	0	0	0	-	1	0
8	0	1	0	0	0	0	1	-	0
9	0	0	0	0	0	0	0	0	-

Fig. 3 Baseline test matrixsu (a) with four variations that represent different kinds of fit problems: (b) opposite, (c) overfit, (d) underfit, and (e) mix of over and underfit

Consider first the *alignment* measure. Results are shown in Fig. 4(a) and Supplemental Figure S1 (available in the Supplemental Materials). On the first criterion (1), the measure distinguishes categorically between perfect and imperfect fit in that perfect fit is measured as zero for both α_2 and α_3, and any lack of mirroring will measure $\alpha_2 > 0$ and/or $\alpha_3 > 0$. On the second criterion (2), the measure distinguishes overfit, underfit, and mixed fit. Specifically, the *overfit* cases show $\alpha_2 = 0$ and $\alpha_3 > 0$, the *underfit* cases show $\alpha_2 > 0$ and $\alpha_3 = 0$, and the *mixed fit* cases show both greater than zero. On the third criterion (3), the measure reflects the extent of fit in that a higher value is reported for the *medium* cases than for the *low* cases, where more dependencies are mismatched. However, this measure distinguishes only dependencies that are missing, but does not detect mirroring problems where a dependency is present in both matrices but is not of the same magnitude. Therefore, the lack of fit is "under-counted" in the weighted test cases (e.g., Figure S1(b), available in the Supplemental Materials). On the fourth criterion (4), the measure does not reflect worse fit based on the location of fit problems. The *off-diagonal* case reports exactly the same value as the *overfit medium* case, when it

Table 2 Summary of results

Criterion	Evaluation	Alignment	C.D.	Cl. & Cent.	Modularity
1: Distinguish fit from not-fit	Compare *perfect* to all other cases; *perfect* should have a value distinguishable from all others	**Yes**[a] *Perfect* fit makes $\alpha_1 = \alpha_2 = 0$; any lack of mirroring makes α_1 or $\alpha_2 > 0$	**Yes.** For *perfect* fit, $\beta = 0$, while any lack of fit gives $\beta > 0$	**Yes.** Value is zero for *perfect* and nonzero for all other cases	**Yes**[a] Value is zero for *perfect* and nonzero for all other cases
2: Distinguish between overfit, underfit, and mixed	Compare *overfit*, *underfit*, and *mixed* test cases; values should fall within mutually exclusive ranges	**Yes**[a] Overfit shows $\alpha_2 = 0$, $\alpha_3 > 0$; underfit shows $\alpha_2 > 0$, $\alpha_3 = 0$; mixed show $\alpha_2 > 0$, $\alpha_3 > 0$	**No.** Positive values are returned for *overfit*, *underfit,* and *mixed* when they should be distinguishable	**No.** Wide variation in values depending on random variability of perturbed test matrices	**No.** The *overfit, underfit,* and *mixed* cases return indistinguishable positive values
3: Show relative extent of fit	Compare *low* and *medium* test cases; values for *medium* should be more extreme than for *low*	**Partial**[a] Higher values for *medium* than *low* test cases, but only detects missing, not inadequate, fit problems	**Yes.** Higher values for *medium* than *low* test cases	**No.** Wide variation in values depending on random variability of the perturbed test matrices	**Partial**[a] Higher values for *medium* than *low* test cases, but only detects missing, not inadequate, fit problems
4: Reflect worse fit for harder locations	Compare *off-diagonal* to *overfit medium*; values for *off-diagonal* should be more extreme	**No.** Value for *off-diagonal* is equal to *overfit medium* but should be higher	**No.** In the unweighted cases, value for *off-diagonal* is equal to *overfit medium* but should be higher	**No.** Registers different fit between *off-diagonal* and other cases, but not always worse fit	**No.** In the unweighted cases, value for *off-diagonal* is equal to *overfit medium* but should be higher
5: Easily interpretable	Subjective judgment of intuitive meaning, and how much value and scale reveal about differences between test cases	**Easiest.** The value is, roughly, the chance that a given cell will be "unmirrored," though values may be misleadingly small	**Only for underfit.** For *underfit*, value reflects proportion of needed communication that is missing; but misleading values for other cases	**No.** No intuitive link between values of γ_n and γ_s and the implications for mirroring	**Somewhat.** Value reflects difference in how modular the matrices are, but there is no clear link to implications for mirroring

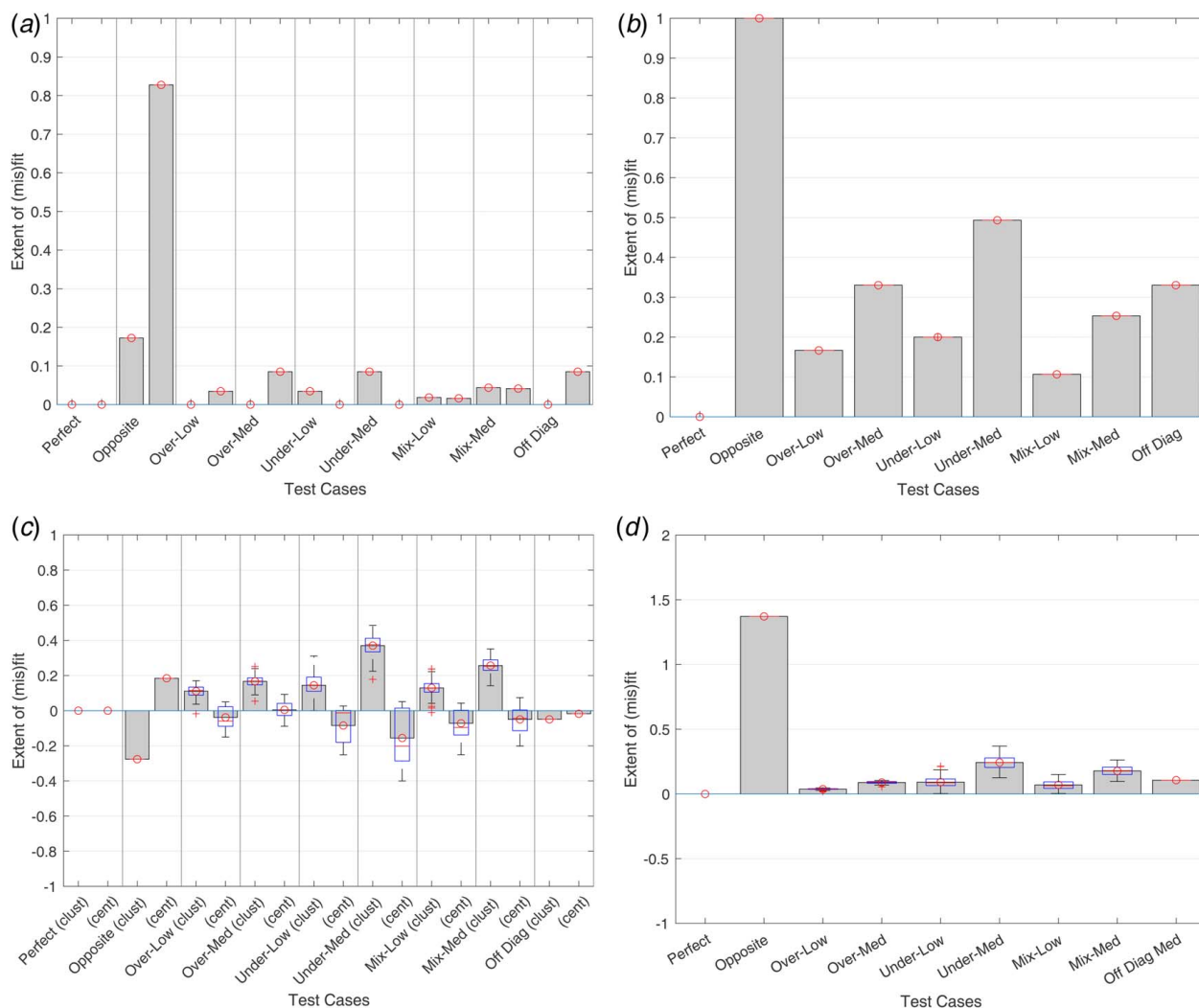[a] Detects mismatched edges but not edge weight.

Fig. 4 Results for all four measures, in the 30 × 30 loosely coupled unweighted test case: (a) results for the *alignment* measure (α_2 on left, α_3 on right), (b) results for the *coordination deficit* measure, (c) results for the *centralization and clustering* measures, and (d) results for the *modularity* measure

should report a higher value to meet this criterion. In general, these results are not sensitive to the specific baseline structure: the same trends were obtained for all three baseline matrices—small (Figure S1(*d*), available in the Supplemental Materials), large loosely coupled (Figures S1(*a*) and S1(*b*), available in the Supplemental Materials), and large, tightly coupled (Figure S1(*c*), available in the Supplemental Materials).

For the fifth criterion (5), the measure is fairly easy to interpret. The value is, roughly, the chance that a given cell will be "unmirrored," i.e., the percent of all pairs of nodes that are not mirrored. The scale is also easy to interpret: perfect fit returns zero, and opposite fit returns $\alpha_1 + \alpha_2 = 1$. On the other hand, the values may be misleadingly small. Because the denominator is all pairs of nodes rather than only those with existing dependencies, the measure gives "credit" for matched noninteractions (node pairs with no organizational ties and no product transfers), which can lead to fairly small values of the measure even when there are a large number of mismatches compared to the number of existing dependencies.

Consider next the *coordination deficit* measure. Results are shown in Fig. 4(*b*) and Supplemental Figure S2 (available in the Supplemental Materials). On the first criterion (1), the measure distinguishes categorically between perfect and imperfect fit in that perfect fit is measured as zero and any lack of fit is measured as a positive value. (Because the measure is defined to measure only coordination deficits, one might expect it to "miss" fit problems in

the *overfit* case. However, extra organizational ties still result in coordination deficits, because the values are normalized to the total amount of communication in each network. Adding an organizational tie increases the denominator slightly, and therefore decreases the value in each cell in the organizational network, resulting in small coordination deficits throughout the network.) On the second criterion (2), the measure does not distinguish overfit, underfit, and mixed fit. In all test cases, *overfit*, *underfit*, and *mixed fit* return positive values. (See the explanation above, under criterion 1.) On the third criterion (3), the measure reflects the extent of fit in that a higher value is reported for the *medium* cases than for the *low* cases, where more dependencies are mismatched. It also reflects the extent of fit where there is a difference in the dependency magnitude m (not shown in the figure). On the fourth criterion (4), the measure does not reflect worse fit based on the location of fit proble. In the unweighted cases (Fig. 4(*b*), Supplemental Figures S2(*c*) and S2(*d*), available in the Supplemental Materials), the *off-diagonal* variation reports exactly the same value as the *overfit medium* variation, when it should report a higher value.

For the fifth criterion (5), the measure is easy to interpret when the problems result from underfit, but can be misleading when the problems result from overfit or mixed fit. For the *underfit* cases, the measure is very intuitive: it reflects the proportion of needed communication that is missing. For example, in the unweighted

cases (such as Fig. 4(b)): the *underfit low* and *medium* cases removed 20% and 50% of the required organizational ties, respectively, and the measure returns precisely that percentage; the *opposite* variation returns 1 (100%). On the other hand, the measure is not intuitive for cases of *overfit*, because a positive value is returned even though there are excess organizational ties in the network (see the explanation in the previous paragraph).

Consider next the *centralization* and *clustering* measures. Results are shown in Fig. 4(c) and Supplemental Figure S3 (available in the Supplemental Materials). The box-and-whisker results show the variability across the 100 randomly generated instances (there was no variability for the previous two measures). On the first criterion (1), the measure distinguishes categorically between perfect and imperfect fit in that perfect fit is measured as zero and any lack of fit is measured as either a positive or a negative value. On the second criterion (2), the measure does not distinguish overfit, underfit, and mixed fit. There is wide variation in the measure's value based on the random variability of the modified test matrices, with many of the distributions overlapping zero, so it would be difficult to interpret the measure for any given instance. Furthermore, even in the average, there is no pattern that could distinguish these cases: for example, in Fig. 4(c), the *overfit*, *underfit*, and *mixed* fit cannot be distinguished because they all show positive clustering and negative or zero centrality. On the third criterion (3), the measure does not reliably reflect the extent of fit. Again, there is wide variability based on the randomly constructed matrices, making the measure unreliable. Furthermore, even in the average, sometimes the *medium* cases show higher-magnitude values than the *low* cases (e.g., Fig. 4(c) and Supplemental Figure S3(d), available in the Supplemental Materials), but sometimes this is not the case (in Figure S3(c), available in the Supplemental Materials, the *overfit low* and *overfit medium* do not have this pattern). Indeed, unlike the other measures considered so far, the *opposite* test case does not always show the worst value of the measure. On the fourth criterion (4), the measure does not clearly show worse fit based on the location of fit problems. Unlike the other measures, it *does* clearly register that the fit is different between the *off-diagonal* and *overfit medium* test cases, but it does not show a *worse* fit in these cases: there is no pattern to the results for the *off-diagonal* test cases that could indicate worse fit.

On the fifth criterion (5), this measure does not appear to be easily interpretable. Specifically, there is no intuitive link between a "more/less clustered" or "more/less centralized" network and the implications for mirroring. Moreover, given the wide variability in the value of the measure for different randomly generated matrices, and without distinguishing overfit from underfit and more mirroring from less, it is hard to interpret the meaning of the values.

Consider finally the *modularity* measure. Results are shown in Fig. 4(d) and Supplemental Figure S4 (available in the Supplemental Materials). On the first criterion (1), the measure distinguishes categorically between perfect and imperfect fit in that perfect fit is measured as zero, and imperfect mirroring shows as nonzero. (In these test cases, it is always positive, but it could also be negative if the organization were more modular than the product.) On the second criterion (2), the measure does not distinguish overfit, underfit, and mixed fit. Supplemental Figure S4 shows that in all the test cases, the *overfit*, *underfit*, and *mixed fit* return positive values that are not distinguishable from one another. On the third criterion (3), the measure reflects the extent of fit in that a higher value is reported for the *medium* cases than for the *low* cases, where more dependencies are mismatched. However, this measure distinguishes only dependencies that are missing, but does not detect mirroring problems where a dependency is present in both matrices but is not of the same magnitude (see Figure S4(b), available in the Supplemental Materials). On the fourth criterion (4), the measure does not detect worse fit based on the location of fit problems. In all the unweighted cases (Fig. 4(d), Supplemental Figures S4(c) and S4(d), available in the Supplemental Materials), the *off-diagonal* case reports the same value as the *overfit medium* case, when it should report a higher value to meet this criterion.

For the fifth criterion (5), the measure is somewhat interpretable. The value is, roughly, the difference in how modular (versus integral) each of the two matrices are. Smaller values mean the two matrices are more similar in modularity. However, there is no easy interpretation of the value itself in terms of how well mirrored two matrices are. Moreover, this measure requires that modules be defined in advance of measuring. This may not always be feasible, and/or could lead to misleading results if the module definitions themselves are incorrect.

Among all of the measures considered here, none meet all of the criteria laid out in Sec. 2.4 (see Table 2). The *alignment* measure comes closest: it can distinguish overfit from underfit and mixed fit, shows the relative extent of fit in cases where ties or transfers are missing, and it is easy to interpret. Minor modifications might alleviate one potential interpretation challenge—that the values may be misleadingly small—which could be considered in future work. The main weakness of this measure is that it cannot detect fit problems where dependencies exist in both cases but are mismatched in magnitude. The *coordination deficit* measure, on the other hand, detects differences in the magnitude of associated dependencies, but is only intuitive when measuring underfit, and cannot distinguish underfit from overfit and mixed fit. (Note that this problem only applies when a network-level measure is computed, not when the measure is used for its original purpose of measuring deficits at the subsystem level [25]).

The *clustering* and *centrality* measures, surprisingly, only detect perfect fit but otherwise do not meet any of the criteria. We had hoped they would detect complex structural problems such as the *off-diagonal* case, but the results were inconsistent across the test cases. The *modularity* metric is also limited in its ability to detect fit problems, only identifying the relative extent of fit with the same weaknesses as *alignment*. These measures are powerful in assessing network characteristics but seem less appropriate for detecting mirroring problems. We believe this is due to the wide variety of ways in which fit problems might manifest. These measures are sensitive to minor variations in the network's structure and do not focus on specific features as required to satisfy the criteria we defined.

5 Conclusions

This study set out to assess and characterize the ability of existing measures of fit to inform improvements in the context of system design. A first essential step was to understand what such a fit measure must detect in order to support this purpose. Based on a review of the literature on product–organizational mirroring, we extracted four specific types of mismatches that a measure must be able to identify. Not all mismatches are equally problematic, and an actionable result needs to accommodate this phenomenon. In other words, to be able to correct "fit" problems, we need to know their nature, their extent, and where they are.

Next, we selected four measures that are representative of the current perspectives in the literature, and evaluated their ability to identify these types of fit problems. Specifically, we considered two types of measures: those that make an element-to-element comparison between a product and organizational DSM, and those that abstract the structure of each DSM first before comparing.

We found that none of the measures satisfied all of the defined criteria, and some were able to detect surprisingly few types of mismatches. Specifically, the second family of measures, those which assess summary measures on the product and organization separately before comparing, showed notable weaknesses in assessing "fit" for improving system design. The first family of measures showed more promise; however, even the best-performing measure, *alignment*, has some relevant "blind spots" for informing improvements for system design. Specifically, these include (1) the inability to detect differences in the magnitude of associated dependencies, (2) the potentially misleading small values returned, and (3) an inability to detect the severity of off-diagonal mismatches. The first and second issues could be alleviated relatively easily—

by normalizing appropriately and/or by integrating aspects of the coordination deficit measure. The third issue is much harder to address, as will be elaborated below.

In summary, this work made two important contributions. First, it clarified and provided a theoretical basis for how "fit" issues impact the design process, formalized in terms of the test conditions. Second, it performed a neutral comparison of existing fit measures and found that, while one comes close, none of the measures can identify all of the relevant fit-driven problems the literature identifies. Our literature-derived criteria and testbed-based evaluation approach outline a path for developing a useful measure for future studies aiming to diagnose fit in the context of design improvement.

These contributions also lay the foundation for future work. A clear first step is to build on these results to develop an appropriate metric for measuring fit in system design processes. One weakness we identified was in distinguishing different sources of lack of fit. Understanding whether misfits occur near or far from the diagonal, for example, could be very important in environments where the work is more distributed and separated by higher organizational boundaries than is typical in concurrent design. An improved measure might draw on some less algorithmic measurement approaches like that of Ref. [11]. This class of measure relies on additional pre-processing of the input data to apply designer knowledge of the importance of different features and their hierarchy within the system, as in Guo and Gershenson's [48] modularity metric which defines a module structure in advance. The same could be done with, for example, the off-diagonal distance. An additional step for future work is to adapt these measures and test them on more realistic systems, such as those that are differently structured (e.g., buses rather than modules), those with asymmetric transfers and ties, and those for which the product and organizational DSMs are not the same size and structure.

More broadly, measures of fit enable many further investigations of the role of organizational and product structure in the design process. While the literature provides ample evidence for the premise that fit should lead to improved design outcomes, further empirical investigation in the specific context of system design would be worthwhile. The measures developed in this paper enable such investigations by providing a way to observe and measure fit. Future work can thereby explore empirically whether and how different fit problems impact real design processes, and how fit interacts with other drivers of design performance. Based on such investigations, approaches can be developed for mitigating such problems, along with thresholds for when a technical problem is so novel that an alternative organizational structure should be employed. As systems engineers and designers continue to engage in more rigid upfront decomposition necessary to adopt efficiency-oriented design practices (such as concurrent design) and further distribute work across supply chains, we hope to enable more careful study of both the costs and benefits of these practices. Furthermore, measures of fit could even enable studies of how agile or flexible organizations restructure to meet novel challenges. Agreed-upon and trusted measures of core constructs like fit play an important role in enabling many streams of future research.

Acknowledgment

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Sherwood, B., and McCleese, D., 2013, "JPL Innovation Foundry," Acta Astron., 89, pp. 236–247.

[2] Ivanov, A., Noca, M., Borgeaud, M., Belloni, F., Guzman, E., Fueglistaler, A., Gallay, S., and Triguero-Baptista, A., 2013, "Concurrent Design Facility at the Space Center EPFL," SECESA 2010, Lausanne, Switzerland, Oct. 13–15.

[3] Bandecchi, M., Melton, B., Gardini, B., and Ongaro, F., 2000, "The ESA/ESTEC Concurrent Design Facility," EuSEC 2000, Munich, Germany, Sept. 13–15, p. 8.

[4] Baldwin, C., and Clark, K., 2000, Design Rules, Vol. 1: The Power of Modularity, MIT Press, Cambridge, MA.

[5] Doyle, J. C., and Csete, M., 2011, "Architecture, Constraints, and Behavior," Proc. Natl. Acad. Sci. U. S. A., 108(Suppl._3), pp. 15624–15630.

[6] Parnas, D. L., 1972, "On the Criteria to Be Used in Decomposing Systems Into Modules," Commun. ACM, 15(12), pp. 1053–1058.

[7] Henderson, R. M., and Clark, K. B., 1990, "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," Adm. Sci. Q., 35(1), pp. 9–30.

[8] Szajnfarber, Z., and Gralla, E., 2017, "A Framework for Measuring the "Fit" Between Product and Organizational Architectures," Conference on Systems Engineering Research, Redondo Beach, CA, Mar. 23–25.

[9] Colfer, L. J., and Baldwin, C. Y., 2016, "The Mirroring Hypothesis: Theory, Evidence, and Exceptions," Indus. Corporate Change, 25(5), pp. 709–738.

[10] Ethiraj, S. S. K., and Levinthal, D., 2004, "Modularity and Innovation in Complex Systems," Manage. Sci., 50(2), pp. 159–173.

[11] Fixson, S. K., and Park, J. -K., 2008, "The Power of Integrality: Linkages Between Product Architecture, Innovation, and Industry Structure," Res. Policy, 37(8), pp. 1296–1316.

[12] Zhou, Y. M., 2013, "Designing for Complexity: Using Divisions and Hierarchy to Manage Complex Tasks," Organ. Sci., 24(2), pp. 339–355.

[13] Zirpoli, F., and Camuffo, A., 2009, "Product Architecture, Inter-Firm Vertical Coordination and Knowledge Partitioning in the Auto Industry," Eur. Manage. Rev., 6(4), pp. 250–264.

[14] Conway, M. E., 1968, "How Do Committees Invent?," Datamation, 14(4), pp. 28–31.

[15] von Hippel, E., 1990, "Task Partitioning: An Innovation Process Variable," Res. Policy, 19(5), pp. 407–418.

[16] Sanchez, R., and Mahoney, J. T., 1996, "Modularity, Flexibility, and Knowledge Management in Product and Organization Design," Strat. Manage. J., 17(S2), pp. 63–76.

[17] Thompson, J. D., 1967, Organizations in Action: Social Science Bases of Administrative Theory, McGraw-Hill, New York.

[18] Galbraith, J., 1974, "Organization Design: An Information Processing View," Interfaces, 4(3), pp. 28–36.

[19] Simon, H. A., 1962, "The Architecture of Complexity," Proc. Am. Philos. Soc., 106(6), pp. 467–482.

[20] Simon, H. A., 1981, The Sciences of the Artificial, 2nd ed., MIT Press, Cambridge, MA.

[21] King, A., 1999, "Retrieving and Transferring Embodied Data: Implications for the Management of Interdependence Within Organizations," Manage. Sci., 45(7), pp. 918–935.

[22] Puranam, P., Singh, H., and Chaudhuri, S., 2009, "Integrating Acquired Capabilities: When Structural Integration Is (Un)necessary," Organ. Sci., 20(2), pp. 313–328.

[23] Amrit, C., and van Hillegersberg, J., 2008, "Detecting Coordination Problems in Collaborative Software Development Environments," Inform. Syst. Manage., 25(1), pp. 57–70.

[24] Cataldo, M., Herbsleb, J. D., and Carley, K. M., 2008, "Socio-Technical Congruence: A Framework for Assessing the Impact of Technical and Work Dependencies on Software Development Productivity," Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement – ESEM '08, Kaiserslautern, Germany, Oct. 9–10, ACM Press, p. 2.

[25] Gokpinar, B., Hopp, W. J., and Iravani, S. M. R., 2010, "The Impact of Misalignment of Organizational Structure and Product Architecture on Quality in Complex Product Development," Manage. Sci., 56(3), pp. 468–484.

[26] Ulrich, K., 1995, "The Role of Product Architecture in the Manufacturing Firm," Res. Policy, 24(3), pp. 419–440.

[27] Browning, T. R., 2001, "Applying the Design Structure Matrix to System Decomposition and Integration Problems: A Review and New Directions," IEEE Trans. Eng. Manage., 48(3), pp. 292–306.

[28] Eppinger, S., and Browning, T. R., 2012, Design Structure Matrix Methods and Applications, MIT Press, Cambridge, MA.

[29] Browning, T. R., 2016, "Design Structure Matrix Extensions and Innovations: A Survey and New Opportunities," IEEE Trans. Eng. Manage., **63**(1), pp. 27–52.

[30] Baldwin, C. Y., 2021, Design Rules, Vol. 2.

[31] De Weck, O. L., Roos, D., and Magee, C. L., 2016, *Engineering Systems: Meeting Human Needs in a Complex Technological World*, MIT Press, Cambridge, MA.

[32] Lawrence, P. R., and Lorsch, J. W., 1967, "Differentiation and Integration in Complex Organizations," Adm. Sci. Q., **12**(1), p. 1.

[33] Tushman, M. L., 1979, "Work Characteristics and Subunit Communication Structure: A Contingency Analysis," Adm. Sci. Q., **24**(1), p. 82.

[34] Drazin, R., and de Ven, A. H. V., 1985, "Alternative Forms of Fit in Contingency Theory," Adm. Sci. Q., **30**(4), p. 514.

[35] Tushman, M. L., and Nadler, D. A., 1978, "Information Processing As an Integrating Concept in Organizational Design.," Acad. Manage. Rev., **3**(3), pp. 613–624.

[36] Allen, T. J., 1984, Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization, Vol. 1. MIT Press Books.

[37] Morelli, M., Eppinger, S., and Gulati, R., 1995, "Predicting Technical Communication in Product Development Organizations," IEEE Trans. Eng. Manage., **42**(3), pp. 215–222.

[38] Sosa, M., Eppinger, S., Pich, M., McKendrick, D., and Stout, S., 2002, "Factors that Influence Technical Communication in Distributed Product Development: An Empirical Study in the Telecommunications Industry," IEEE Trans. Eng. Manage., **49**(1), pp. 45–58.

[39] Sosa, M. M. E., Eppinger, S. D. S., and Rowles, C. M. C., 2004, "The Misalignment of Product Architecture and Organizational Structure in Complex Product Development," Manage. Sci., **50**(12), pp. 1674–1689.

[40] Sosa, M. E., 2008, "A Structured Approach to Predicting and Managing Technical Interactions in Software Development," Res. Eng. Design, **19**(1), pp. 47–70.

[41] Parraguez, P., Eppinger, S. D., and Maier, A. M., 2015, "Information Flow Through Stages of Complex Engineering Design Projects: A Dynamic Network Analysis Approach," IEEE Trans. Eng. Manage., **62**(4), pp. 604–617.

[42] Gulati, R., and Puranam, P., 2009, "Renewal Through Reorganization: The Value of Inconsistencies Between Formal and Informal Organization," Organ. Sci., **20**(2), pp. 422–440.

[43] Sosa, M. E., Gargiulo, M., and Rowles, C., 2015, "Can Informal Communication Networks Disrupt Coordination in New Product Development Projects?," Organ. Sci., **26**(4), pp. 1059–1078.

[44] Cabigiosu, A., and Camuffo, A., 2012, "Beyond the "Mirroring" Hypothesis: Product Modularity and Interorganizational Relations in the Air Conditioning Industry," Organ. Sci., **23**(3), pp. 686–703.

[45] Furlan, A., Cabigiosu, A., and Camuffo, A., 2014, "When the Mirror Gets Misted Up: Modularity and Technological Change: Modularity and Technological Change," Strateg. Manage. J., **35**(6), pp. 789–807.

[46] Hennig, A., Topcu, T. G., and Szajnfarber, Z., 2022, "So You Think Your System Is Complex?: Why and How Existing Complexity Measures Rarely Agree," ASME J. Mech. Des., **144**(4), p. 041401.

[47] Sosa, M. E., Eppinger, S. D., and Rowles, C. M., 2003, "Identifying Modular and Integrative Systems and Their Impact on Design Team Interactions," ASME J. Mech. Des., **125**(2), pp. 240–252.

[48] Guo, F., and Gershenson, J. K., 2004, "A Comparison of Modular Product Design Methods Based on Improvement and Iteration," 16th International Conference on Design Theory and Methodology, Vol. 3a, Salt Lake City, UT, Sept. 28–Oct. 2, ASME, pp. 261–269.

[49] Guo, F., and Gershenson, J. K., 2007, "Discovering Relationships Between Modularity and Cost," J. Intell. Manufact., **18**(1), pp. 143–157.

[50] Hölttä-Otto, K., Chiriac, N. A., Lysy, D., and Suk Suh, E., 2012, "Comparative Analysis of Coupling Modularity Metrics," J. Eng. Design, **23**(10–11), pp. 790–806.

[51] Hölttä, K., Suh, E. S., and de Weck, O., 2005, "Tradeoff Between Modularity and Performance for Engineered Systems and Products," Proceedings of the 15th International Conference on Engineering Design, Melbourne, Australia, Aug. 15–18, p. 14.

[52] Hölttä-Otto, K., Chiriac, N., Lysy, D., and Suh, E. S., 2014, "Architectural Decomposition: The Role of Granularity and Decomposition Viewpoint," *Advances in Product Family and Product Platform Design: Methods & Applications*, T. W. Simpson, J. R. Jiao, Z. Siddique, and K. Hölttä-Otto, eds., Springer New York, New York, pp. 221–243.

[53] Wasserman, S., and Faust, K., 1994, *Social Network Analysis: Methods and Applications.* No. 8 in Structural Analysis in the Social Sciences, Cambridge University Press, Cambridge, New York.

[54] Watts, D. J., and Strogatz, S. H., 1998, "Collective Dynamics of 'Small-World' Networks," Nature, **393**(6684), pp. 440–442.

[55] Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A., 2004, "The Architecture of Complex Weighted Networks," Proc. Natl. Acad. Sci. U. S. A., **101**(11), pp. 3747–3752.

[56] Bounova, G. and de Weck, O. L., 2011, "Matlab Tools for Network Analysis," Technical Report, Massachusetts Institute of Technology.

[57] MacCormack, A., Baldwin, C., and Rusnak, J., 2012, "Exploring the Duality Between Product and Organizational Architectures: A Test of the "Mirroring" Hypothesis," Res. Policy, **41**(8), pp. 1309–1324.