Clustering Product Development Project Organization From the Perspective of Social Network Analysis

Qing Yang, Na Yang, Tyson R. Browning, Bin Jiang, and Tao Yao

***Abstract*—In product development (PD) organizations, coordinating technical dependencies among teams with different expertise in overlapping processes is a fundamental challenge. This article takes a more sophisticated approach than prior methodologies to improve coordination via organizational clustering, by accounting for both team structural and attribute similarity from the perspective of social network analysis. We built models to quantify the impact of the overlapping processes on the interaction strength among PD teams, which we then used to construct structural similarity by combining tie strength and social cohesion among teams via the design structure matrix. To evaluate the organization network, we propose social embeddedness-related centrality indices within (intracluster) and across (intercluster) team groupings. To facilitate knowledge sharing, we base team attribute similarity on product- and process-related expertise among teams. We integrate the modularity index and an improved silhouette index to find an optimal number of clusters, which we then incorporate with team similarity measures as inputs to a spectral clustering algorithm. An industrial example illustrates the proposed model. The clustering results reinforce several managerial practices but also yield new insights, such as how to measure similarity among teams based on organizational network characteristics and how structural and attribute similarities impact the optimal organizational structure.**

***Index Terms*—Clustering algorithm, design structure matrix (DSM), organization design, product development (PD), project management, social network analysis (SNA).**

I. Introduction

A KEY managerial issue in product development (PD) is how to establish an effective organizational architecture to help coordinate hundreds or even thousands of specialists, because the complexity of their interactions may reduce efficiency and introduce risks [1], [2]. An effective organizational architecture can reduce management complexity by facilitating communication, coordination, and innovation [3]. Therefore, many researchers have explored ways to facilitate communication and coordination by improving the organizational architecture [4], [5].

A challenge in developing an organizational architecture concerns modularization—parsing the set of organizational elements (e.g., teams) into subsets, groups, or modules, such that the elements' intragroup relationships outweigh those across groups [5], [7]. Assigning elements to groups is also known as finding communities [8], partitioning [9], and clustering [10], [11], [40], [44]. Although nontrivial, this assignment problem can provide an effective way to develop an organizational architecture for improving team communication and coordination, thereby reducing management complexity [12].

However, existing organizational clustering approaches make the grouping decisions based almost exclusively on the relationships among the elements—without accounting for important properties of the elements themselves, such as their similarity. Meanwhile, social network analysis (SNA) techniques provide an effective approach for developing a similarity matrix model of the organizational architecture [13], [41]. In this article, we enhance our understanding of organizational clustering by incorporating the perspective of SNA, which helps uncover important properties in the PD organization. For example, communications are more likely to occur among teams that are similar in the organizational network [14], and increased communications potentially result in the emergence of expertise and knowledge [15]. High similarity among teams may increase interaction and enable more intensive communication and coordination, which can lead to higher organizational performance. Thus, team similarity should be taken into account when optimizing an organizational network.

The social network contains two important dimensions: a structural dimension representing different kinds of relationships among the elements (or nodes) and an attributes dimension representing features of the nodes [16]. However, graph clustering algorithms have traditionally focused on the structural dimension (i.e., edge weights) without accounting for the attribute dimension [12], so they provide only a partial representation of the real social system. Therefore, our research questions are two-fold: (i) how can we quantify structural similarity and the attribute similarity among PD teams from the perspective of SNA, and (ii) how can we identify clusters based on these similarities?

Manuscript received 3 February 2019; revised 9 June 2019 and 16 August 2019; accepted 22 August 2019. Date of publication 17 October 2019; date of current version 1 November 2022. This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 71872011 and Grant 71929101. Review of this manuscript as arranged by Department Editor L. Santiago. (*Corresponding author: Qing Yang*.)

Q. Yang and N. Yang are with the School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China (e-mail: yangqing@manage.ustb.edu.cn; yangna_ustb@163.com).

T. R. Browning is with the Neeley School of Business, Texas Christian University, Fort Worth, TX 76129 USA (e-mail: t.browning@tcu.edu).

B. Jiang is with the Driehaus College of Business, DePaul University, Chicago, IL 60604 USA (e-mail: bjiang@depaul.edu).

T. Yao is with The Harold & Inge Marcus Department of Industrial & Manufacturing Engineering, Pennsylvania State University, University Park, PA 16802 USA (e-mail: taoyao2005@gmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEM.2019.2939398

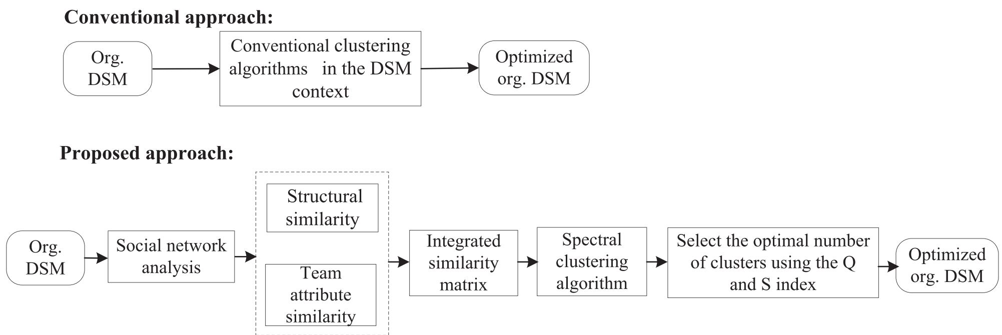

Fig. 1. Comparing the conventional and proposed approaches to optimize an org DSM.

In this article, we present an improved PD organizational clustering approach, based on the spectral clustering algorithm [17], [18], that synthesizes SNA with a design structure matrix (DSM) model. The organization DSM (org DSM) is a powerful network modeling tool for displaying and analyzing the coordination dependency relationships between teams, thereby highlighting the organizational architecture in PD [10], [11], [42]. As we discuss in the next section, an org DSM can be used to identify clusters, but current DSM clustering methods focus on the direct relationships among teams without accounting for their similarity. Fig. 1 compares the conventional DSM clustering approach with the new approach presented in this article. As we discuss in the next section, many prior studies have applied some kind of clustering algorithm to optimize a model of the organization architecture, such as an org DSM. This article takes a more sophisticated view to clustering a PD organization by incorporating structural similarity and team attribute similarity from the perspective of SNA.

Our proposed PD organizational clustering method uses an integrated similarity matrix from the team attributes and relationships among teams as inputs to a spectral clustering algorithm. The integrated similarity matrix is a network-based similarity measure. A typical characteristic of PD projects is for two or more teams to work simultaneously on an overlapping process—i.e., concurrent activities, starting a downstream process before completing an upstream one, which can accelerate the schedule [1], [19]. Therefore, we model the impact of an overlapping process on a PD team's dyadic interactions, using these to measure tie strength (TS) and social cohesion (SC) among teams. In the social network context, teams that share social embeddedness (i.e., strong TS and SC) are motivated to work closely together and thus are willing to devote their time and energy to communicate and interact. Social embeddedness influences the level of integration between teams [20], and the most productive organizations are internally cohesive [29]. Network centrality—e.g., *degree centrality* (DC) or *betweenness centrality* (BC)—can be used to measure an individual node's position in an organization network [6], [22]. Therefore, in this article, we construct the structural similarity measure by combining TS and SC, and we propose the centrality indices of intacluster and intercluster to analyze the organization network.

This article also constructs measures of team attribute similarity using product- and process-related expertise, which can facilitate knowledge sharing between teams [35]. Next, we incorporate teams' attributes and relationships (structural similarity) to develop an integrated similarity matrix, more comprehensively and accurately than other similarity measurement methods. This article also presents an approach integrating the modularity index Q and an improved Silhouette index S using node attribute differences (ADs) (intacluster and interclister) to determine the optimal number of clusters, which is a particular challenge for spectral clustering algorithms [7]. Finally, we use centrality indices of intracluster and intercluster, with Q and S indices, to evaluate the clustering results.

Therefore, this article makes three key contributions.

1) It takes a more sophisticated view than prior approaches to cluster in PD organizations by accounting for structural and team attribute similarities from the perspective of SNA.
2) It extends existing studies on overlapping processes to predict the dyadic interaction strength (social embeddedness) in the PD organizational network, which is then used to build the structural similarity measure by combining the direct and indirect TSs. It also uses the product- and process-related expertise overlap (EO) among PD teams to measure the team attribute similarity.
3) It proposes the social embeddedness-related centrality indices of intracluster and intercluster, which are integrated with the Q and S indices to evaluate the clustering results.

II. Literature Review

A. *Organization DSM Modeling and Analysis*

A PD project presents not only the engineering design challenge of creating product concepts and configurations but also the organizational design challenge of managing team interactions and coordination [1]. The technical communication and coordination required for effective PD require much of the designers' time and energy. General organizational design theory [38] emphasizes the importance of enabling effective coordination through careful choices about team structures and relationships. Therefore, an important managerial challenge is to facilitate this communication and coordination with effective and efficient organizational structures that make the best use of designers' time and effort and help ensure that critical technical issues are addressed. Good organizational architecture makes it easier for teams to give and receive important information by collocating and establishing other integrative and coordination mechanisms among the most highly interactive teams [4]. Hence, it is essential to cluster-related teams to facilitate coordination in the PD organizational structure.

An org DSM can be used to identify clusters—i.e., potentially advantageous modules/groups—and their interfaces in PD projects [10], [11], [32]. A DSM is a square matrix, with diagonal entries representing system elements, and off-diagonal entries (i, j) representing directional dependencies from elements j to i. DSMs have been used to model a variety of system architectures, especially those of products, processes, and organizations [10], [11]. So, how to measure the dependencies among elements and how to apply clustering approaches to DSM are decisive factors for modularization [42]–[44].

Org DSMs are useful models because they represent the information flows in a PD organization and can be manipulated to reveal alternative organizational architectures, such as ones with improved modularity [42]. This manipulation is typically achieved with clustering algorithms, which seek to assign the system elements (here, teams) into various groups, modules, or clusters of highly interactive teams, thereby localizing relationships within groups and minimizing relationships across groups [5], [11], [19], thus reducing managerial complexity and coordination costs. Numerical DSMs can have various measures and attributes attached to the elements (e.g., size or importance) and relationships (e.g., quantity, importance, or frequency)—although this makes the design of the clustering objective function more challenging [10], [43].

Although existing org DSM models have focused on the dependency strength among teams, they have not analyzed the embedded influence of organizational networks and cannot characterize the attributes of each team in the network. By doing so, we would be able to measure team similarity in the organizational network for clustering the PD organization.

B. SNA and Spectral Clustering

A PD project's organizational structure is like a complex social network. Many researchers have combined the org DSM with SNA techniques to analyze the structural characteristics of the PD organization [20], [21]. Sosa [23] found that strong ties serve as effective catalysts for the generation of creative ideas when they link actors who are intrinsically motivated to work closely together. Sosa [24] also analyzed rework in PD projects from the perspective of social networks. Grewal *et al.* [6] found that centrality indices (e.g., DC, BC, and eigenvector centrality) have strong effects on technical success. However, existing research has not used SNA characteristics (e.g., such as TS and centrality indices) to measure structural similarity and evaluate clustering results.

Spectral clustering algorithms based on similarity provide a stronger and more stable approach for finding the global optimum [18], [25], especially for nonconvex datasets [9], and

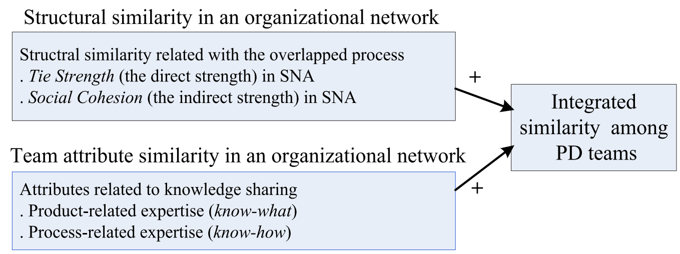

Fig. 2. Measuring integrated similarity among PD teams.

are well suited for application to real problems [7]. The spectral clustering algorithm maximizes intercluster similarity and minimizes intercluster similarity [18]. The similarity matrix is thus a critical input to a spectral clustering algorithm [9]. Many researchers have developed methods to measure similarity [8], [34], which can obtain modularity internally with a high similarity.

Several researchers have analyzed the influence of team attributes—such as interaction, coordination, knowledge sharing, and team functioning [14], [28]—on PD organizations. In social networks, the attributes of teams and their patterns of interaction have an important impact on network structure [8]. Similar interests and expertise may lead two teams to be close to each other in a group. Socially similar individuals or teams are more likely to be connected and to interact more frequently [30]. However, most spectral clustering algorithms depend solely on the relationships among nodes (structural similarity) without accounting for their attribute similarity.

In summary, even though previous research has brought considerable insight into clustering teams to reduce their coordination complexity, several improvements are possible through synthesizing spectral clustering and DSM methods by incorporating structural similarity and team attribute similarity from the perspective of SNA.

III. Organizational Networks in PD Projects

A. Measuring Team Similarity Based on the Organizational Network Characteristics

In PD projects, the main reason for clustering teams can be their similarity, the degree to which teams or members "view themselves as having few differences" in terms of their interaction, relationships, or specific attributes, such their way of working or their expertise [14]. The network structure (i.e., relationships among teams) and team attributes are the basic characteristics of an organizational network. Hence, this article proposes an innovative approach to measuring integrated similarity among teams based on structural similarity and team attribute similarity in the organizational network (see Fig. 2).

The relationships among teams determine structural similarity. Individuals resolving interdependent activities are more likely to exchange technical information, so activity interdependence is a main factor determining structural similarity. According to the overlapping process and the embedded influence of the organizational network, social embeddedness among teams

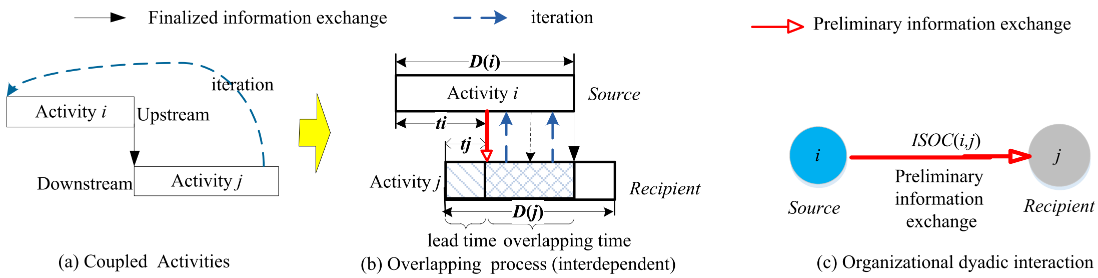

Fig. 3. Organizational dyadic interaction based on the overlapping process.

can be used to construct a structural similarity matrix, in which the direct strength among teams can be captured by the TS, and indirect strength can be captured by SC [24]. Team attribute similarity is a fundamental factor driving knowledge sharing and technical communication. The EO leads to the recipient and partner teams being more inclined to cooperate and better understanding the linkages between one another's knowledge, hence providing more favorable conditions for communication [14]. For PD teams, the main types of expertise include product- and process-related expertise. Thus, we use expertise-related factors to identify team attribute similarity.

B. PD Teams' Dyadic Interactions Due to Overlapping Processes

In the PD organizational network, teams with dyadic interaction exchange information to carry out their PD activities. Two fundamental features of PD processes, overlapping and iteration, both stimulate the exchange of information among teams [19]. To establish the dyadic interaction, each team can be viewed as either an information source or recipient in the organizational network [24]. Hence, we identify the dyadic interaction between the source and recipient based on the dependencies of the overlapped activities they perform.

The representative dependencies between activities can normally be divided into sequential and coupled [10]. As the coordination among teams can be negligible in sequential activities, we focus on coupled activities [31], where activities need information inputs from each other, so information exchanges continue until the activities converge on a mutually satisfactory solution [see Fig. 3(a)]. Overlapping, where a downstream activity begins earlier by using preliminary information from an upstream activity can accelerate coupled activities [19], [33]. Fig. 3(b) and (c) illustrates the organizational dyadic interaction among overlapping activities, in which the source (i.e., the upstream team) sends preliminary technical information to the recipient (i.e., the downstream team), and then the recipient feeds back information to the source (iteration). In Fig. 3(b), the overlapping time is the period in which coupled information exchanges take place. The overlapping necessitates increased communication (two-way information exchanges between teams performing coupled activities). The lead time, t_j, refers to work the downstream team can start before it receives input from the upstream team, and the durations of activities *i* and *j* are *D*(*i*) and *D*(*j*), respectively.

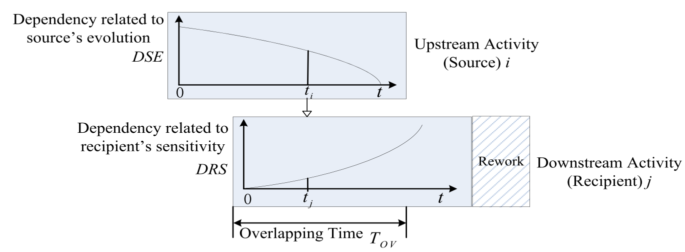

Fig. 4. Impact of the source's evolution and recipient's sensitivity on the interaction strength.

The influence of overlapping time on the dyadic interaction between teams can be measured using the concepts of *evolution degree (Evol)* and *sensitivity degree (Sens)* [19], [33]. Evol is the percentage of an upstream activity completed before its first information outputs, and Sens is the percentage of a downstream activity completed before its first information inputs. If t_i units of time are required to release the initial information from activity *i*, then $\mathrm{Evol}(t_i) = t_i \,/\, D(i)$. If activity *j* has already worked t_j units of time before it receives information from activity *i*, then $\mathrm{Sens}(t_j) = t_j \,/\, D(j)$. Fig. 4 illustrates the impact of Evol and Sens on the interaction strength between two teams [33]. To measure dyadic interaction strength, we build models of the dependency related to source evolution and recipient sensitivity.

Evolution refers to the gradual refinement of the upstream activity from its preliminary form to a final form over the period of its output information [33]. The dependency related to a source's evolution is a function decreasing with time (see Fig. 4). For instance, the later the upstream activity releases information (i.e, the larger $\mathrm{Evol}(t_i)$), the closer the upstream information to its final value, and thus the less coordination and communication needed. The dependency related to the source's evolution (DSE) can be approximated as a linear function [19]

$$\mathrm{DSE}(t_i) = 1 - \varepsilon(2\mathrm{Evol}(t_i) - 1) \tag{1}$$

where the parameter $\varepsilon \in [-1, 1]$.

Sensitivity refers to the communication and coordination required for a downstream activity to accommodate information received. The dependency related to recipient's sensitivity (DRS) is a function increasing with overlapping time (see Fig. 4). For instance, the later the downstream activity receives information (or changes) from an upstream activity (i.e, the larger $\mathrm{Sens}(t_j)$), the greater the coordination and communication needed. Because the finish time of the upstream activity

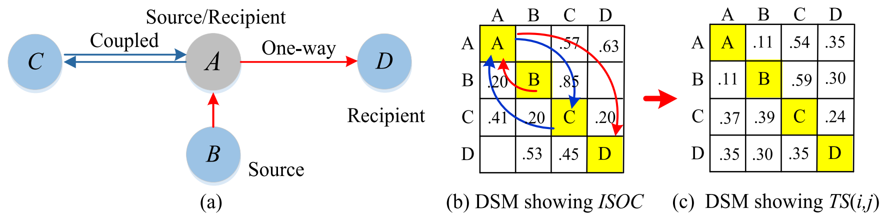

Fig. 5. Example of measuring TS among teams with a DSM model.

can be either later or earlier than that of the downstream activity [19], the overlapping time can be calculated as follows:

$$T_{OV}(i,j) = \begin{cases} (1-\mathrm{Evol}(t_i)) \times D(i) & \text{if } (1-\mathrm{Evol}(t_i)) \times D(i) \\ & < (1-\mathrm{Sens}(t_j)) \times D(j) \\ (1-\mathrm{Sens}(t_j)) \times D(j) & \text{else} \end{cases}. \tag{2}$$

The DRS is determined by the degree of overlapping (i.e., the ratio of T_{OV} to $D(i)+D(j)$), calculated as follows [19]:

$$\mathrm{DRS}\,(t_j) = \mu \ln\left(\gamma \frac{T_{OV}}{D(i)+D(j)} \times \mathrm{Sens}\,(t_j) + 1\right) \tag{3}$$

where μ represents the uncertainty (or technical complexity) of a design activity, and γ represents the capability of both teams to reduce uncertainty in the design process. If a team has a strong ability to cope with uncertainty, γ will be small. In this article, $\mu = 1$ and $\gamma = 1$.

Based on (1)–(3), we measure a PD team's dyadic *interaction strength based on overlapping and coordination* (ISOC) with (4), which reflects the communication frequency associated with the overlapping time.

$$\mathrm{ISOC}(i,j) = \mathrm{DSE}(t_i) \cdot \mathrm{DRS}\,(t_j)\,. \tag{4}$$

C. Measuring TS and SC in the Organizational Network

From a behavioral perspective, the willingness or motivation of teams to communicate has an important impact on the efficiency of information exchanges [13], [29]. Such willingness is a function of how deeply the recipient is embedded in the social relationship with the source [13]. Thus, we apply *dyadic (social) embeddedness* to analyze how dyadic interaction (i.e., ISOC) impacts the willingness and propensity of technical communication among PD teams.

In the social network context, *dyadic* (*social*) *embeddedness* is a combination of two modules: TS and SC [23], [24], [29]. TS is a function of the amount of time and effort that both the source and the recipient spend in their direct dyadic interaction (i.e., within direct two-way interactions), measured by the network proportions relative to the aggregate level across the network. SC refers to the extent to which a relationship is surrounded by strong third-party connections (i.e., common contacts) [29]. Hence, TS reflects the direct strength, and SC the indirect strength, between teams.

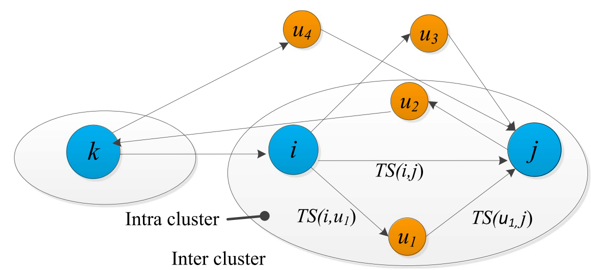

Fig. 6. Examples of SC and intraclustrer/intercluster coordinators.

1) Modeling the TS Among Teams With a DSM: Factors impacting TS between teams include communication frequency and emotional closeness [29]. We assume that teams are emotionally close to teams with whom they coordinate frequently, so we estimate TS as a function of interaction strength. In this article, we use a DSM model [e.g., Fig. 5(b)] as a basis for calculating TS, where each element's inputs appear in its corresponding column, and its outputs appear in its corresponding row (c.f. [11]). The TS between teams *i* and *j* represents how team *i* allocates its time or effort (interaction strength) to team *j* relative to all of *i*'s direct relationships to other teams [24]. Using DSM, tie strength $\mathrm{TS}(i,j)$ can be measured as the proportion of the interaction strength between teams *i* and *j* relative to the total interaction strength with all of their adjacent teams

$$\mathrm{TS}(i,j) = \frac{\mathrm{ISOC}(i,j) + \mathrm{ISOC}(j,i)}{\sum_{q=1}^{N} (\mathrm{ISOC}(i,q) + \mathrm{ISOC}(q,i))} \quad \text{for} \quad i \neq j \tag{5}$$

where $\mathrm{ISOC}(i,j)$ is the interaction strength from *i* to *j*, *N* is the total number of teams, and $\sum_{q=1}^{N} (\mathrm{ISOC}(i,q) + \mathrm{ISOC}(q,i))$ represents the total interaction strength between team *i* and all of its adjacent nodes (i.e., teams) *q* in the organizational network.

2) Modeling the SC Among Teams With a DSM: SC, an indirect structural link between two teams due to their common third-party connections, is a function of the time or effort that both the source and the recipient spend in their relationships with common contacts. Fig. 6 illustrates SC with teams *i, j*, and *k* having common contacts.

The SC from team *i* to team *j* can be calculated as

$$\mathrm{SC}(i,j) = \sum_{u \in \Gamma(i) \cap \Gamma(j)} \mathrm{TS}(i,u) \cdot \mathrm{TS}(u,j) \quad \text{for} \quad u \neq i,j \tag{6}$$

where $\Gamma(i)$ represents the neighborhood set of team *i* containing its adjacent teams. The common contact set is represented by *u*, the intersection of $\Gamma(i)$ and $\Gamma(j)$. Thus, we can calculate *social embeddedness* (*SE*) as a combination of TS and SC

$$\mathrm{SE}(i,j) = \mathrm{TS}(i,j) + \mathrm{SC}(i,j) \quad \text{for} \quad u \neq i,j. \tag{7}$$

D. Modeling SE-Related Centrality: Intra and Intercluster

Fig. 6 helps illustrate the concept of intra and intercluster coordinators. A coordinator or broker is a player that lies between two others who do not have direct communication, acting as a channel by which they can relate [21]. We extend the traditional notion of a coordinator—which includes internal coordinator, external coordinator, gatekeeper, and liaison [21]—to distinguish intra and intercluster coordinators. An intracluster coordinator (e.g., u_1 in Fig. 6) mediates the relationship between two teams where both the mediated teams and the coordinator reside in the same cluster, whereas an intercluster coordinator (e.g., u_2, u_3, and u_4) mediates the relationship between two teams where the mediated teams and/or the coordinator reside in different clusters. Note that these intercluster coordinators can themselves be members of particular clusters or not: u_2 is internal; u_3 and u_4 are both external.

Network centrality, the existence of a number of connections or a proportion of redundant ties between teams, is often associated with team performance [22]. Thus, a high network centrality—e.g., DC or BC—implies that the complex activities in PD can be spread over more teams, resulting in better performance [6]. DC is an indicator of a team's connectivity with other teams in a network, based only on direct connections. A team's DC indicates its "power" within a group [22]. BC provides an index of the potential to control or facilitate information exchange among teams. From an information flow perspective, nodes with high BC will have access to a relatively large portion of the information flowing among other nodes [21]. Greater BC in a network has been positively related to individual performance ratings [22], [26]. Internal communication is positively related to PD outcomes [39]. Previous BC and DC research has focused on individuals, whereas, in this article, we analyze these metrics at the level of clusters of teams.

In this article, we explore the problem of finding an optimal organization with intracluster centrality $>>$ intercluster centrality, meaning more direct and indirect communication among teams within clusters ("tightly knit" groups) than across clusters.

1) Intra and Intercluster DC: We measure intra and intercluster DC based on direct TS. Team *i*'s intracluster DC is defined as its in- and out-degree in cluster *K* divided by the standardized DC of intracluster connected at maximum with all other teams

$$\mathrm{DC}_{\mathrm{intra}}(K_i) = \left(\sum_{j=n_k}^{m_k} TS(i,j) + \sum_{j=n_k}^{m_k} TS(j,i) \right) \Big/ 2(cl_k - 1) \tag{8}$$

where cl_k is the size of cluster *K*, n_k, and m_k are the indices of first and last elements in cluster *K*, respectively. Similarly, given team *i* within cluster *K*, the team's standardized intercluster DC with respect to *K* is

$$\mathrm{DC}_{\mathrm{inter}}(K_i) = \left(\sum_{j=1}^{N} \mathrm{TS}(i,j) + \sum_{j=1}^{N} \mathrm{TS}(j,i) - \left(\sum_{j=n_k}^{m_k} \mathrm{TS}(i,j) + \sum_{j=n_k}^{m_k} \mathrm{TS}(j,i) \right) \right) \Big/ 2(N - cl_k). \tag{9}$$

Then, a cluster's intra and intercluster DCs are the sums of each of its team's intra and intercluster DCs, respectively.

2) Intra and Intercluster BC: We measure intra and intercluster BC based on indirect SC. BC is based on the shortest paths linking pairs of nodes. A high BC score indicates that a node can reach many others via relatively shortest paths or that it lies on a considerable proportion of the shortest paths connecting other nodes. Given team *u* in cluster *K*, its BC in the cluster is

$$\mathrm{BC}_{\mathrm{intra}}(K_u) = \sum_{i,j,u \in K, i<j} \frac{\sigma_{i,j}(K_u)}{\sigma_{i,j}} \Big/ (cl_k - 1)(cl_k - 2) \tag{10}$$

where $\sigma_{i,j}$ is the shortest indirect connection between teams *i* and *j*. If the indirect communication occurs within a cluster, it acts as a bridge between teams, which can facilitate the interactions among teams in a cluster. An approximate BC can be obtained from the connections between the neighbors of each team (common contacts) [27]. In this article, SC can reflect the shortest indirect connection between teams via a common contact. Hence, we propose the concept of intra and intercluster BC based on SC.

Given team *u* in cluster *K*, the team's standardized, intracluster BC is the ratio of the total SC (6) between each pair of teams, via common team *u* as an intracluster coordinator within cluster *K*, to the overall SC of the entire organizational network

$$\mathrm{BC}_{\mathrm{intra}}(K) = \left(\mathrm{SC}(K) \Big/ \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{SC}(i,j) \right) \Big/ (cl_k - 1)(cl_k - 2) \tag{11}$$

where $\mathrm{SC}(K) = \sum_{i=n_k}^{m_k} \sum_{j=n_k}^{m_k} \mathrm{SC}(i,j)$ represents the SC of the coordinator teams and the connected teams belonging to the same cluster *K* (e.g., u_1 in Fig. 6). A large $\mathrm{BC}_{\mathrm{intra}}$ represents strong indirect connections in a cluster, which can facilitate communication among teams.

Intercluster BC is the ratio of the SC of the coordinator teams acting as intercluster coordinators (e.g., u_2, u_3, and u_4 in Fig. 6) to the overall SC of the entire organizational network

$$\mathrm{BC}_{\mathrm{inter}} = \frac{\left(\sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{SC}(i,j) - \sum_{K=1}^{N_C} \mathrm{SC}(K) \right) \Big/ \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{SC}(i,j)}{(N-1)(N-2)} \tag{12}$$

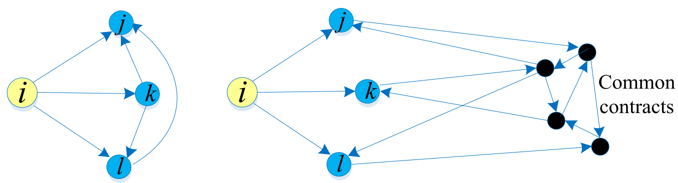

Fig. 7. Impact of SE on structural similarity. (a) Strong tie strength. (b) Strong social cohesion.

where Nc is the total number of clusters.

High intracluster BC and DC imply groups of teams that can intercept a large portion of information and quickly disseminate it among group members, respectively.

IV. Modeling the Similarity Matrix of the Organizational Network

A. Modeling Structural Similarity

We constructed a structural similarity measure based on SE, which is a combination of TS and SC [(7), and see Fig. 7]. PD teams directly connected via strong TS, or by many common contacts via strong SC, may have a high similarity level. For example, in Fig. 7(a), teams *i*, *j*, *k*, and *l* are directly connected. The strength of these teams' relationships depends on the possibility of information moving from one team to another [13], so these teams have high similarity due to strong TS.

If two teams are surrounded by strong, common partners, they are more willing to share knowledge, thus facilitating technical communication. Thus, structural similarity based on SC (i.e., shared neighbors) can effectually characterize the local connectivity density of any two adjacent nodes in an organizational network [34]. For example, in Fig. 7(b), although teams *i, j, k,* and *l* are not directly connected, they have many common contacts, and thus high similarity due to strong SC. The presence of common, third parties may reinforce the predisposition of interdependent teams to share information [3].

Based on SE, we construct a structural similarity metric. Structural similarity has been measured as the cosine similarity [8], [34] of each pair of nodes

$$S_{\text{cosine}}(i,j) = \frac{|\Omega(i) \cap \Omega(j)|}{\sqrt{|\Omega(i)| \cdot |\Omega(j)|}} \tag{13}$$

where $\Omega(i)$ is the neighbor set of node i, including itself, and $|\Omega(i) \cap \Omega(j)|$ represents the common neighbors of nodes *i* and *j*.

Cosine similarity was originally designed for unweighted, undirected networks [34]. However, the PD organizational network more accurately as a weighted, directed network. Let $G(V,E,\text{TS})$ be a weighted graph where V is the nodes in graph G, E is the edges (links) between nodes, and $\text{TS}(i,j)$ is the edge weight (5). To measure structural similarity of the weighted, directed network, we first specify the structural similarity matrix$\text{Sim}_{\text{struc}}$ as an extension of the cosine similarity measure to indicate the local connectivity density of any two adjacent nodes in the weighted, undirected network [34]

$$\text{Sim}_{\text{stru}}(i,j) = \frac{\sum\limits_{u \in \Omega(i) \cap \Omega(j)} \text{TS}(i,u) \cdot \text{TS}(u,j)}{\sqrt{\sum\limits_{u \in \Omega(i)} \text{TS}^2(i,u)} \cdot \sqrt{\sum\limits_{u \in \Omega(j)} \text{TS}^2(j,u)}}. \tag{14}$$

In (14), the structural neighborhood of team i is the set $\Omega(i)$ containing i and its adjacent teams: $\Omega(i) = \{u \in V | (i,u) \in E\} \cup \{i\}$. However, $\Omega(i)$ cannot distinguish whether i is a source or a recipient, so (14) cannot be used for the directed network. As mentioned above, strong TS and SC can lead to high similarity. We assume $\text{TS}(i,i) = 1$, so if i is a source and $i = u$, then $\text{TS}(i,u) \cdot \text{TS}(u,j) = \text{TS}(i,i) \cdot \text{TS}(i,j) = \text{TS}(i,j)$. Hence, in the weighted, directed network

$$\begin{aligned} &\sum_{u \in \Omega_S(i) \cap \Omega_S(j)} \text{TS}(i,u) \cdot \text{TS}(u,j) \\ &= \text{TS}(i,j) + \sum_{u \in \Gamma_S(i) \cap \Gamma_S(j)} \text{TS}(i,u) \cdot \text{TS}(u,j) \end{aligned} \tag{15}$$

where $\Omega_S(i) = \{u | (i,u) \in E\} \cup \{i\}$ is the set containing *i* and *i*'s adjacent nodes, when *i* is a source.$\Gamma_S(i) = \{u | (u,i) \in E\}$is *i*'s adjacent nodes when *i* is a source. The difference between $\Omega(i)$ and $\Gamma(i)$ is that $\Omega(i)$ contains *i* itself.

Using an org DSM, we can determine the structural similarity of an organizational network as in (16) shown at the bottom of the next page, where $\Gamma_R(i) = \{u | (i,u) \in E\}$ is *i*'s adjacent nodes when *i* is a recipient.

According to (7), (16) simplifies to (17) is shown at the bottom of the next page.

B. Modeling Team Attribute Similarity

In addition to node relationships, node attributes should also affect clustering. Teams' required knowledge and expertise are the key attributes influencing communication and knowledge sharing. Expertise includes two key components: *know-what* and *know-how*. *Know-what* stands for an appreciation of the kinds of phenomena worth pursuing; *know-how* represents an understanding of reproductive processes that constitute phenomena [35]. In this article, we extend *know-what* and *know-how*, as proposed by Garud [35], to the PD context. *Know-what* represents product-related expertise (e.g., appropriate component technologies), whereas *know-how* represents process-related expertise (e.g., designing, manufacturing, and testing activities). Product- and process-related expertise can be viewed as team attributes. Know-what is connected with the particular functional and architectural attributes of the product under development [24]. It represents an understanding of specific technologies and configurations that satisfy requirements. Modularizing the PD organization may help to preserve know-what knowledge. Know-how is associated with the activities in a PD project: knowledge about how to perform them at a particular stage of the project can be accumulated with experience. In PD projects, process-related expertise represents an understanding of the appropriate activities and procedures.

Teams with a high similarity of process- and product-related expertise likely work on similar activities and components in a PD project, so they should be clustered to facilitate knowledge sharing and interaction. We measure team attribute similarity with two factors: a team's level of involvement (L) in allocating its communication effort in different areas of expertise, and the EO in various areas.

The level of involvement of team i in expertise area k is defined as the ratio of team i's amount of work effort required in area k to its entire amount of work effort in all expertise areas

$$L(i,k) = z(i,k) \Bigg/ \sum_{k=1}^{N_K} z(i,k) \tag{18}$$

where $z(i,k)$ is the amount of work effort or time spent by team i in the area k and N_K is the number of expertise areas. Fig. 8 gives an example of measuring matrix $\boldsymbol{L}$. The numerical values of each row in the "team-expertise area" matrix [see Fig. 8(a)] represent the amount of work effort or time spent by each of five teams in each of three areas of expertise.

EO refers to the ratio of the same expertise area between teams. It reflects the degree of common knowledge between teams, which can facilitate communication and knowledge transfer inside a group. Reagans and McEvily [29] modeled the EO between teams i and j as

$$\mathrm{EO}_{ij} = \sum_{k=1}^{N_K} a_{ik} a_{jk} \Bigg/ N_K \tag{19}$$

where $a_{ik} = 1$ if team i is an expert in area k; otherwise, $a_{ik} = 0$. However, (19) is a binary model. Using L, we propose a matching coefficient method to quantify EO

$$\mathrm{EO}_k(i,j) = \begin{cases} \frac{\min\{L(i,k),L(j,k)\}}{\max\{L(i,k),L(j,k)\}} & \text{if } L(i,k)+L(j,k) \neq 0 \\ 1 & \text{else} \end{cases} \tag{20}$$

where $L(i,k)+L(j,k) \neq 0$ signifies that at least one team is involved in area k.

Using (20), we can measure the product- and process-related EO between teams. EO ranges from [0, 1]. For example, in Fig. 8, a project involves three areas of expertise: materials engineering (k_1), electronics (k_2), and dynamics (k_3). Team A needs expertise k_1 and k_2, and the degree of involvement in the two areas is 30% and 70%, respectively. If team B needs k_1, k_2, and k_3, and the degree of involvement in these areas are 50%, 35%, and 15%, respectively, then $\mathrm{EO}_1(i,j) = 0.3/0.5 = 0.6$, $\mathrm{EO}_2(i,j) = 0.35/0.7 = 0.5$, and $\mathrm{EO}_3(i,j) = 0$.

	1	2	3
Team A	6	14	
Team B	20	14	6
Team C		12	
Team D	6		24
Team E			12

(a)

	1	2	3
Team A	.3	.7	
Team B	.5	.35	.15
Team C		1	
Team D	.2		.8
Team E			1

(b)

	A	B	C	D	E
Team A	A	.31	.35	.23	.18
Team B	.31	B	.21	.24	.19
Team C	.35	.21	C	.18	.24
Team D	.23	.24	.18	D	.36
Team E	.18	.19	.24	.36	E

(c)

Fig. 8. Example of measuring team attributes similarity via DSM. (a) "team-expertise area" matrix. (b) Matrix L. (c) Attribute similarity DSM.

To measure attribute similarity, we first use Euclidean distance to calculate the total difference of attributes between teams i and j

$$M(i,j) = \sqrt{\sum\nolimits_{k=1}^{N_K} \left(1 - \mathrm{EO}_k(i,j)\right)^2}. \tag{21}$$

A large $M(i,j)$ indicates that teams i and j have small EO and low attribute similarity. Li *et al.* [8] used $1/(M(i,j)+1)$ to measure node attribute similarity. However, even if two teams have small expertise overlaps (especially if $\mathrm{EO}_k(i,j) = 0$), the node similarity is large and cannot reflect the real situation. Hence, we measure the attribute similarity between teams i and j as

$$\mathrm{Sim}_{\mathrm{attr}}(i,j) = e^{-M(i,j)}. \tag{22}$$

Similar expertise between teams can help to integrate acquired capabilities, make the assimilation of knowledge proceed more easily, and develop a consensus among the interactive teams, which will have a constructive effect on knowledge sharing and increase the frequency of interaction and coordination.

Using (17) and (22), we calculate the integrated similarity among teams as a linear combination of the structural similarity and team attribute similarity matrices [16]

$$\mathrm{Sim}(i,j) = \alpha \times \mathrm{Sim}_{\mathrm{stru}}(i,j) + (1-\alpha) \times \mathrm{Sim}_{\mathrm{attr}}(i,j) \tag{23}$$

where α is a similarity coefficient ranging from [0, 1], depending on the actual situation. If we emphasize the structural impact on the similarity, $\alpha > 0.5$.

$$\mathrm{Sim}_{\mathrm{stru}}(i,j) = \frac{\mathrm{TS}(i,j) + \sum\limits_{u \in \Gamma_S(i)\cap\Gamma_S(j)} \mathrm{TS}(i,u)\mathrm{TS}(u,j) + \mathrm{TS}(j,i) + \sum\limits_{u\in\Gamma_R(i)\cap\Gamma_R(j)} \mathrm{TS}(j,u)\mathrm{TS}(u,i)}{\sqrt{\sum\limits_{u\in\Gamma_S(i)\cup\Gamma_R(i)} \left(\mathrm{TS}(i,u)+\mathrm{TS}(u,i)\right)^2} \cdot \sqrt{\sum\limits_{u\in\Gamma_S(j)\cup\Gamma_R(j)} \left(\mathrm{TS}(j,u)+\mathrm{TS}(u,j)\right)^2}} \tag{16}$$

$$\mathrm{Sim}_{\mathrm{stru}}(i,j) = \frac{\mathrm{SE}(i,j) + \mathrm{SE}(j,i)}{\sqrt{\sum\limits_{u\in\Gamma_S(i)\cup\Gamma_R(i)} \left(\mathrm{TS}(i,u)+\mathrm{TS}(u,i)\right)^2} \cdot \sqrt{\sum\limits_{u\in\Gamma_S(j)\cup\Gamma_R(j)} \left(\mathrm{TS}(j,u)+\mathrm{TS}(u,j)\right)^2}} \tag{17}$$

C. Spectral Clustering Method and Determining the Optimal Number of Clusters

In multivariate statistics and data clustering, spectral clustering aims to make good use of the spectrum (i.e., eigenvalues) of the data's similarity matrix (an input) to perform dimensionality reduction before clustering. The similarity matrix provides a quantitative assessment of the relative similarity of nodes in the dataset. The optimal partition maximizes the similarity of elements in a cluster (or subgraph) while minimizing the similarity between elements in different clusters. Although spectral clustering has been applied to social networks, it has not previously been used on data that accounts for both the attributes and relationships of nodes (i.e., both attribute and structural similarity matrices) [8], which we do here. We apply the Ng-Jordan-Weiss algorithm-based, normalized spectral clustering procedure [36] because of its more robust performance [9].

In addition to using appropriate input, the spectral clustering method usually faces another challenge in determining the optimal number of clusters [18]. In many cases, we will not have any *a priori* information on the appropriate number of clusters k^*. The modularity index *Q* [7] and the Silhouette index *S* [17] can be used to evaluate the quality of a clustered structure related to the network's structure and node attributes, respectively. In this article, we integrate them to find the optimal k^*.

First, we use the modularity index to evaluate the results of clustering related to the network structure. The *Q* index of a directed, weighted network is defined as the sum of the weights of all of the edges included within subgraphs (after clustering), less the expected edge weight sum under the condition that edges were placed at random [16]

$$Q = \frac{1}{m}\sum_{ij}\left(\mathrm{TS}(i,j) - \frac{d_i^{(\mathrm{out})} d_j^{(\mathrm{in})}}{m}\right)\delta(G_i, G_j) \tag{24}$$

where $d_i^{(\mathrm{out})} = \sum_{j=1}^{N}\mathrm{TS}(i,j)$ and $d_j^{(\mathrm{in})} = \sum_{i=1}^{N}\mathrm{TS}(i,j)$ represent node *i*'s out-degree and in-degree, respectively, *m* is the sum of all of the edge weights, $m = \sum_{i=1}^{N} d_i^{(\mathrm{out})} = \sum_{j=1}^{N} d_j^{(\mathrm{in})}$, and $\delta(G_i, G_j)$ returns 1 when nodes *i* and *j* belong to the same cluster and 0 otherwise.

Next, we present an improved *S* index to evaluate the clustering results based on intra and intercluster ADs. The *S* index proposed by Arbelaitz [37] measures clustering quality by calculating the separation between clusters and the compactness among teams in each cluster [17]. A team's AD of intracluster is defined as the average differences between team *i* and other teams in cluster *K*

$$\mathrm{AD}_{\mathrm{intra}}(K_i) = \frac{1}{cl_k - 1}\sum_{j=n_k}^{m_k} M(i,j) \tag{25}$$

where cl_k is the size of cluster *K* and *M*(*i*, *j*) is the total difference of attributes involving all areas of expertise between teams *i* and *j*. A team's AD of intercluster is defined as the minimum average differences between team *i* in cluster *K* and all the teams in another cluster *H*

$$\mathrm{AD}_{\mathrm{inter}}(K_i) = \min_{1\le h\le N_c, h\ne k}\left[\frac{1}{cl_h}\sum_{j\in H} M(i,j)\right]. \tag{26}$$

Next, the *AD* of both intra and intercluster for cluster *K* can be calculated with $\mathrm{AD}_{\mathrm{intra}}(K) = \frac{1}{cl_k}\sum_{i=n_k}^{m_k}\mathrm{AD}_{\mathrm{intra}}(K_i)$ and $\mathrm{AD}_{\mathrm{inter}}(K) = \frac{1}{cl_k}\sum_{i=n_k}^{m_k}\mathrm{AD}_{\mathrm{inter}}(K_i)$, respectively. Hence, the improved *S* index can be calculated as follows:

$$S = \frac{1}{N_C}\sum_{K=1}^{N_C}\left\{\frac{1}{cl_k}\sum_{i=n_k}^{m_k}\frac{\mathrm{AD}_{\mathrm{inter}}(K_i) - \mathrm{AD}_{\mathrm{intra}}(K_i)}{\max\left[\mathrm{AD}_{\mathrm{inter}}(K_i), \mathrm{AD}_{\mathrm{intra}}(K_i)\right]}\right\}. \tag{27}$$

The range of *S* is $[-1, 1]$, and its maximum indicates the best clustering result.

V. Industrial Application

To validate the proposed concepts and models, we applied them to an industrial example, a commercial aircraft system development (CASD) project in China. We interviewed 28 individuals, including the project manager and other core project members from the firm's R&D, production, and human resource departments. We raised the following questions in the interviews.

1) How can we optimize the project's organization regarding communication between teams?
2) How much does one design activity influence other, overlapping activities that teams perform?
3) What is the minimum information required to start downstream work?
4) How many product- and process-related expertise areas are required of each PD team?
5) To what extent is each team involved in each area of expertise? And so on.

Based on the responses and other information provided, we found that the main purpose of communication is work coordination and knowledge sharing. Work coordination is associated with the overlapping process, and knowledge sharing is associated with the product- and process-related expertise of teams. Although activity concurrency can reduce duration, it increases the risks of rework. We can overcome the weaknesses caused by initially incomplete information exchange among overlapping activities by strengthening the communication between teams in a group. Hence, it is necessary to modularize teams and let them work in the same group. The CASD project involves a large number of activities and teams. Building a modular organization by increasing intracluster dependencies while reducing intercluster dependencies can improve organizational efficiency.

A. Data on Team Time Allocation to Product- and Process-Related Areas of Expertise

To facilitate knowledge sharing, we should consider both the product- and process-related expertise of each team. Because we focus on the impact of the overlapping process and knowledge sharing on the communication and coordination among PD teams, we selected 20 teams (each performing a unique

TABLE I
TEAM TIME ALLOCATION ACROSS DIFFERENT AREAS OF EXPERTISE

ID#	Teams	Activity Durations (hours)	Product-related expertise				Process-related expertise		
			PS	SS	ECS	A	RA	D	IV
A	Market analysis	16	4	5	3	4	12	4	0
B	Requirements identification	13	3	4	3	3	10	3	0
C	Perform function design	9	2	3	2	2	7	2	0
D	System-level design	20	1	15	2	2	2	18	0
E	Product configuration design	21	5	6	5	5	0	16	5
F	Develop Verification scheme	18	4	5	4	5	0	14	4
G	Perform shape design	17	0	15	2	0	2	15	0
H	Perform concept analysis	16	4	5	4	3	12	4	0
I	Confirm scheme	20	5	6	4	5	15	5	0
J	Perform load analysis	9	1	8	0	0	1	8	0
K	Die design	33	2	26	3	2	3	30	0
L	Airfoil design	30	2	25	0	3	3	27	0
M	Perform aerodynamic design	16	1	13	2	0	2	14	0
N	Perform structural design	11	1	8	1	1	1	10	0
O	Cabin design	48	3	38	5	2	5	43	0
P	Propulsion system design	16	0	12	4	0	2	14	0
Q	System assembly	10	2	3	2	3	0	2	8
R	Perform wind tests	19	1	14	2	2	0	10	9
S	Test & Verification	36	9	9	9	9	0	9	27
T	Airworthiness certification	51	12	13	13	13	0	12	39

activity) as summarized in Table I. To measure the expertise overlap among teams, we captured the teams' areas of technical expertise. In this case, the product-related expertise includes four areas: power system (PS), structural system (SS), environmental control systems (ECS), and avionics (A). We classified the process-related expertise of these teams as requirements analysis (RA), design (D), and integration/verification (IV). Table I shows each team's allocation of time across these areas of expertise.

B. Cluster Generation and Selection: Finding the Optimal Number of Clusters

Based on the interview responses and other information provided, we built an org DSM in which DSM(*i,j*) represents ISOC using (1)–(4) [see Fig. 9(a)]. The large number of marks in this DSM signifies a management challenge to enable appropriate coordination within this complex network. Fig. 9(b) will be discussed later.

Next, we calculated the similarity matrix with (5)–(7) and (14)–(23) and implemented the spectral clustering algorithm in MATLAB 18 software. The algorithm selects k clusters, although the optimal number of clusters is generally unknown. A simple approach to finding the optimal k^* involves obtaining a set of data partitions with different values of k and then selecting the best result according to cluster validation indices. Hence, we select different k values depending on the size of the organizational network and compare the Q and S indices of different k clusters using (24) and (27).

Fig. 10 shows the Q and S indices as a function of k for $\alpha = \{0.2, 0.4, 0.6, 0.8\}$ in (23). Both indices achieve their maxima ($Q = 0.58$, $S = 0.35$) for all values of α when $k^* = 3$. So, we select three clusters as the optimal number of clusters (groups), where the modularization of the organizational structure from the perspective of the network's structure and node attributes reaches the highest level. In this case, we set base value of α to 0.6.

C. Clustering Results: Challenge and Insights

Fig. 9(b) shows a dashed line across the cluster tree to distinguish three groups of teams: G1 [G, N, J, M, D, P, K, O, L, and R], G2 [A, B, H, C, I]), and G3 [E, Q, T, S, and F]. Fig. 11(a) shows the corresponding DSM, where the greater intensity of intracluster dependencies relative to intercluster dependencies is clearly visible. For example, G3 is a group of teams with strong, iterative relationships in the RA phase of PD.

There are several important considerations in these results.

1) TS provides a better predictor of organizational clustering and relationship similarity than interaction strength (see Section III-B). The clustered TS DSM [see Fig. 11(b)] has more marks in the group (e.g., group one) than the interaction strength DSM [see Fig. 11(a)]. For instance, even though teams G, P, and L have no or weak interaction

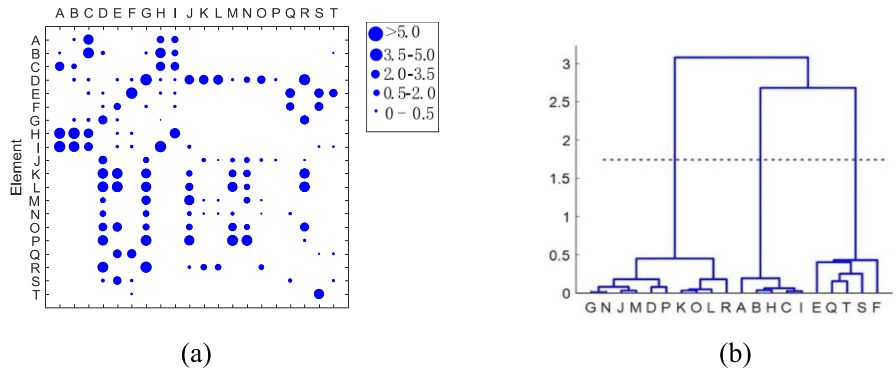

Fig. 9. Original org DSM and cluster tree. (a) Original org DSM. (b) Cluster tree.

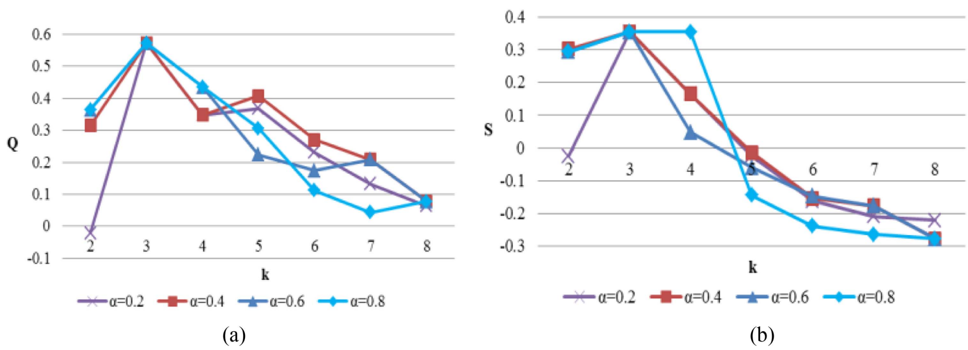

Fig. 10. Experimental results from adjusting parameter α for different values of k. (a) Modularity index Q. (b) Silhouette index S.

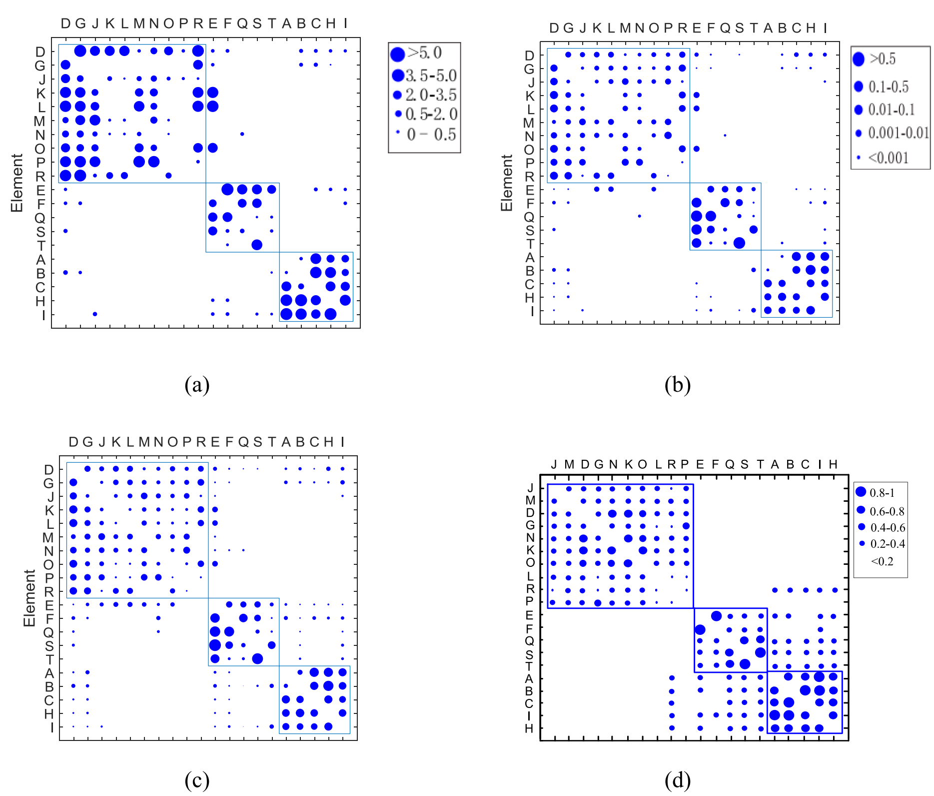

Fig. 11. Clustering results. (a) DSM showing interaction strength. (b) DSM showing TS. (c) DSM showing social embeddedness. (d) DSM showing team attributes similarity.

with most of the other teams in group one [see row G, columns P and L in Fig. 11(a)], they have strong TS with other teams in the group. Strong TS in one group can strengthen technical communication and reduce the possibility of rework in the module. On the other hand, some intercluster TS is essential to support integrative information exchange across groups.

2) SC can reinforce the similarity of interdependent teams and support the exchange of information, so teams with strong SC are clustered in one group. Fig. 11(c) shows social embeddedness, which can strengthen the similarity among teams due to SC. For example, teams K, L, O, and P in group one have limited interaction strength with each other, but they have common intracluster coordinator team R and common intercluster contact team E, which forms SC. Two teams connected by common teams can facilitate the sharing of technical information.

3) Team attribute similarity also has a significant influence on clustering. Fig. 11(d) illustrates the attribute similarity of the clustered DSM, in which teams with high attribute similarity are brought together in groups. Compared to Fig. 11(b), almost all teams have strong attribute similarity with other teams in one group. For instance, team R has attribute similarity with teams M and N, even though they have no TS [see Fig. 11(b)]. Usually, teams with a high attribute similarity will lead to strong communication interaction, so their structural similarity is high as well. This empirical evidence suggests that high interaction strength is positively associated with similar expertise across teams (e.g., teams in group three). However, for some nodes, a high attribute similarity and structural similarity may not coincide. For example, in group three, even though teams E and F have small attribute similarity with teams Q, S, and T [when $\alpha = 0.6$ in (23)], they are still clustered in the same group because of strong TS. When we increase the weight of attribute similarity (e.g., $\alpha = 0.05$), teams E and F are removed from group two and added to group [E, F, J, P].

Further, we take $\alpha = 0.1$ and 1 in (23), respectively, which represent the two extreme cases of the similarity matrix. When $\alpha = 1$, which means the largest structural similarity weight, the optimal number of clusters is three. When $\alpha = 0.1$, which means a large attribute similarity weight, the optimal number of clusters becomes four. Thus, a large attribute similarity weight will affect the optimal clustering result.

4) Some teams act as coordinators not only within clusters but also across clusters. For example, teams D, E, G, and I not only have strong interaction with teams within their cluster but also interact with teams in other clusters. Thus, these teams act as communication bridges across clusters.

D. Sensitivity Analysis

Because team E has strong interaction strengths with other teams, we select team 5E as an example to analyze sensitivity to the Q index (see Fig. 12). Three levels of Evol(i, 5) and Sens(i, 5) in cluster two are evaluated, which are 0.9, 0.5, and 0.1. We fix Evol(i, 5) and Sens(i, 5) for $i = \{F, Q, S, T\}$ at three levels, respectively, while allowing the values of Evol(5, j) and Sens(5, j) for $j = \{F, Q, S, T\}$ to vary at different levels to observe how Q index changes. Because increasing the value of Evol(i, j) or Sens(i, j) decreases overlapping, when we increase the value of Evol(5, j) and Sens(5, j), the interaction strength between team E and the other teams is decreased. This leads to team E being removed from the cluster two and added to the cluster one as well as decreasing Q. Hence, the variable of Evol(i, j) and Sens(i, j) will lead to changed clustering results.

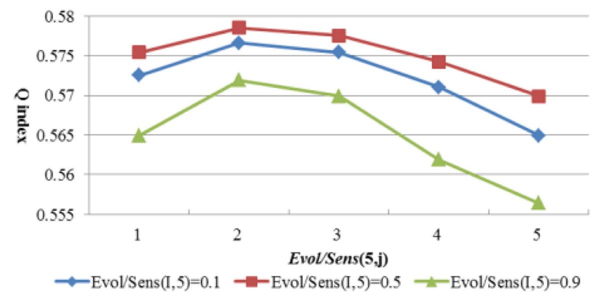

Fig. 12. Sensitivity analysis of Q index.

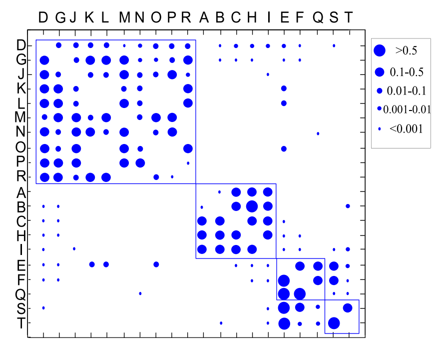

Fig. 13. Results of alternative 1 showing TS.

E. Comparison Tests

Comparing our proposed clustering approach to other existing methods is also necessary. In Table II, the first alternative is a conventional DSM metaheuristic algorithm [32] with two stages of clustering criteria. Because this algorithm yields stochastic results, we selected the most frequent result (four clusters) from 100 runs as its proposed optimum (see Fig 13). The second alternative is a DSM spectral clustering algorithm using a cosine similarity for the undirected network without node attributes [25], in which two clusters are selected as optimal due to the large gap between the second and third eigenvalues.

The experimental results shown in Table II indicate that the clustering result from our proposed method is superior to these two alternatives. Compared to alternative one, our method increases Q by 330.8%, S by 178.7%, and the ratios of intra/intercluster DC and BC by 161% and 167%, respectively, while decreasing the ratio of intra/intercluster AD by 29.9%. Our approach also results in significant improvements compared to alternative two due to the selection of an optimal number of groups and assignments of teams to those groups. These

TABLE II
EXPERIMENTAL RESULTS

	Our proposed method	**Alternative 1**	**Alternative 2**
Number of clusters (k)	3	4	2
Q index (Eq. (24))	0.573	0.133	0.157
S index (Eq. (27))	0.354	0.127	0.323
DC of intra-cluster (D1) (Eq. (8))	3.420	4.193	1.911
Group 1	1.055	1.058	0.991
Group 2	1.220	1.220	0.920
Group 3	1.145	1.057	-
Group 4	-	0.857	-
DC of inter-cluster (D2) (Eq. (9))	0.085	0.270	0.367
D1/D2 ratio	40.235	15.530	5.208
BC of intra-cluster (B1) (Eq. (11))	0.382	0.390	0.144
Group 1	0.067	0.060	0.111
Group 2	0.180	0.150	0.032
Group 3	0.135	0.180	-
Group 4	-	0	-
BC of inter-cluster (B2) (Eq. (12))	0.004	0.011	0.008
B1/B2 ratio	96.637	36.239	17.906
AD of intra-cluster (A1) (Eq. (25))	0.249	0.339	0.220
Group 1	0.207	0.207	0.192
Group 2	0.277	0.264	0.248
Group 3	0.264	0.345	-
Group 4	-	0.538	-
AD of inter-cluster (A2) (Eq. (26))	0.399	0.381	0.323
A1/A2 ratio	0.624	0.890	0.680

comparisons indicate our proposed method can obtain improved results in terms of: Q index, the ratio of intra/intercluster DC and BC (the teams within groups have much stronger social embeddedness than outside groups, which facilitates the direct and indirect communication within groups), S index, and ratio of intra/intercluster AD (the teams within groups have a higher similarity of expertise than outside groups, which facilitates knowledge sharing within groups). Hence, our proposed method can improve clustering results from two perspectives, that of strong, cohesive groups (when TSs among teams are intense, which is positively related to PD effectiveness and efficiency [39]) and that of groups with similar expertise, which can enhance organizational learning and the accumulation of project expertise. The optimal organizational structure can reduce coordination complexity and improve group performance.

VI. Conclusion

This article presented an innovative method for solving two critical issues in PD organizational design: how to quantify the structural and attribute similarities among PD teams from the perspective of social networks and how to identify clusters based on those similarities. We provided a framework that enables managers to optimize the PD organizational architecture more effectively. First, we leveraged existing studies on overlapping activities to predict the interaction strengths among PD teams. Then, we analyzed the embedded influence of the organizational network in terms of both TS and SC to expand on structural similarity among teams. We proposed social embeddedness-based centrality indices and ADs of intra and intercluster. To facilitate knowledge sharing, we build team attribute similarity based on product- and process-related expertise. This article was the first research to redesign an actual PD organizational architecture by incorporating structural and team attribute similarities from the perspective of SNA. We also integrated the modularity index (Q) and an improved S index to find the optimal number of clusters, which we then use with the integrated similarity matrix as inputs to a spectral clustering algorithm to identify clusters.

Several aspects of the model presented in this article warrant further examination. First, apart from overlapping activities, other dependency measurement methods deserve future investigation for constructing the structural similarity measure, and the accuracy of input information related to overlapping would benefit from further analysis. Second, apart from shared expertise, other factors influencing team attribute similarity could be studied. Third, apart from direct, dyadic interactions, a triadic (or more complex) structure comprising multiple dyadic relationships might merit exploration. Fourth, further synthesis and analysis of org DSM and SNA techniques could be a productive area of DSM research. We expect that our findings will provide the appropriate seeds for further, related developments in PD practice.

References

[1] C. H. Loch and C. Terwiesch, "Communication and uncertainty in concurrent engineering," *Manage. Sci.*, vol. 44, no. 8, pp. 1032–1048, 1998.

[2] R. V. Ramasesh and T. R. Browning, "A conceptual framework for tackling knowable unknown unknowns in project management," *J. Oper. Manage.*, vol. 32, no. 4, pp. 190–204, 2014.

[3] M. E. Sosa, M. Gargiulo, and C. Rowles, "Can informal communication networks disrupt coordination in new product development projects?" *Organ. Sci.*, vol. 26, no. 4, pp. 1059–1078, 2015.
[4] T. R. Browning, "Using the design structure matrix to design program organizations," in *Handbook of Systems Engineering and Management*. Hoboken, NJ, USA: Wiley, pp. 1401–1424, 2009.
[5] A. Tripathy and S. D. Eppinger, "Structuring work distribution for global product development organizations," *Prod. Oper. Manage.*, vol. 22, no. 6, pp. 1557–1575, 2013.
[6] R. Grewal, G. L. Lilien, and G. Mallapragada, "Location, location, location: How network embeddedness affects project success in open source systems," *Manage. Sci.*, vol. 52, no. 7, pp. 1043–1056, 2006.
[7] C. H. Lee, M. N. Hoehn-Weiss, and S. Karim, "Grouping interdependent tasks: Using spectral graph partitioning to study complex systems," *Strategic Manage. J.*, vol. 37, no. 1, pp. 177–191, 2016.
[8] W. Li, J. Xie, M. Xin, and J. Mo, "An overlapping network community partition algorithm based on semi-supervised matrix factorization and random walk," *Expert. Syst. Appl.*, vol. 91, pp. 277–285, 2018.
[9] U. Von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
[10] S. D. Eppinger and T. R. Browning, *Design Structure Matrix Methods and Applications*. Cambridge, MA, USA: MIT Press, 2012.
[11] T. R. Browning, "Design structure matrix extensions and innovations: A survey and new opportunities," *IEEE Trans. Eng. Manage.*, vol. 63, no. 1, pp. 27–52, Feb. 2016.
[12] J. Alcácer and M. Zhao, "Zooming in: A practical manual for identifying geographic clusters," *Strategic Manage. J.*, vol. 37, no. 1, pp. 10–21, 2016.
[13] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Cambridge, MA, USA: Harvard Univ. Press, 2009.
[14] J. Backmann, M. Hoegl, and J. L. Cordery, "Soaking it up: Absorptive capacity in interorganizational new PD teams," *J. Prod. Innov. Manage.*, vol. 32, no. 6, pp. 861–877, 2015.
[15] S. Faraj and L. Sproull, "Coordinating expertise in software development teams," *Manage. Sci.*, vol. 46, no. 12, pp. 1554–1568, 2000.
[16] C. Bothorel, J. D. Cruz, M. Magnani, and B. Micenkova, "Clustering attributed graphs: Models, measures and methods," *Netw. Sci.*, vol. 3, no. 3, pp. 408–444, 2015.
[17] A. Mur, R. Dormido, N. Duro, S. Dormido-Canto, and J. Vega, "Determination of the optimal number of clusters using a spectral clustering optimization," *Expert. Syst. Appl.*, vol. 65, pp. 304–314, 2016.
[18] M. C. Nascimento and A. C. Carvalho, "Spectral methods for graph clustering–a survey," *Eur. J. Oper. Res.*, vol. 211, no. 2, pp. 221–231, 2011.
[19] Q. Yang, T. Yao, T. Lu, and B. Zhang, "An overlapping-based design structure matrix for measuring interaction strength and clustering analysis in product development project," *IEEE Trans. Eng. Manage.*, vol. 61, no. 1, pp. 159–170, Feb. 2014.
[20] D. Y. Kim, "Understanding supplier structural embeddedness: A social network perspective," *J. Oper. Manage.*, vol. 32, no. 5, pp. 219–231, 2014.
[21] D. A. Batallas and A. A. Yassine, "Information leaders in product development organizational networks: Social network analysis of the design structure matrix," *IEEE Trans. Eng. Manage.*, vol. 53, no. 4, pp. 570–582, Nov. 2006.
[22] S. Pryke, S. Badi, H. Almadhoob, B. Soundararaj, and S. Addyman, "Self-organizing networks in complex infrastructure projects," *Project Manage. J.*, vol. 49, no. 2, pp. 18–41, 2018.
[23] M. E. Sosa, "Where do creative interactions come from? The role of tie content and social networks," *Organ. Sci.*, vol. 22, no. 1, pp. 1–21, 2011.
[24] M. E. Sosa, "Realizing the need for rework: From task interdependence to social networks," *Prod. Oper. Manage.*, vol. 23, no. 8, pp. 1312–1331, 2014.
[25] S. Sarkar, A. Dong, J. A. Henderson, and P. A. Robinson, "Spectral characterization of hierarchical modularity in product architectures," *J. Mech. Des.*, vol. 136, no. 1, p. 011006, 2014.
[26] R. Cross and J. N. Cummings, "Tie and network correlates of individual performance in knowledge-intensive work," *Acad. Manage. J.*, vol. 47, no. 6, pp. 928–937, 2004.
[27] U. Brandes, "A faster algorithm for betweenness centrality," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001.
[28] P. Hong, W. J. Doll, E. Revilla, and A. Y. Nahm, "Knowledge sharing and strategic fit in integrated product development projects: An empirical study," *Int. J. Prod. Econ.*, vol. 132, no. 2, pp. 186–196, 2011.
[29] R. Reagans and B. McEvily, "Network structure and knowledge transfer: The effects of cohesion and range," *Admin. Sci. Quart.*, vol. 48, no. 2, pp. 240–267, 2003.
[30] R. Reagans, "Preferences, identity, and competition: Predicting tie strength from demographic data," *Manage. Sci.*, vol. 51, no. 9, pp. 1374–1383, 2005.
[31] Q. Yang, X. Zhang, and T. Yao, "An overlapping-based process model for managing schedule and cost risk in product development," *Concurrent Eng., Res. Appl.*, vol. 20, no. 1, pp. 3–16, 2012.
[32] Q. Yang, S. Kherbachi, Y. S. Hong, and C. Shan, "Identifying and managing coordination complexity in global product development project," *Int. J. Project Manage.*, vol. 33, no. 7, pp. 1464–1475, 2015.
[33] V. Krishnan, S. D. Eppinger, and D. E. Whitney, "A model-based framework to overlap product development activities," *Manage. Sci.*, vol. 43, no. 4, pp. 437–451, 1997.
[34] J. Huang, H. Sun, J. Han, and B. Feng, "Density-based shrinkage for revealing hierarchical and overlapping community structure in networks," *Physica A*, vol. 390, no. 11, pp. 2160–2171, 2011.
[35] R. Garud, "On the distinction between know-how, know-what, and know-why," *Adv. Strategic Manage.*, vol. 14, pp. 81–102, 1997.
[36] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst.*, 2002, pp. 849–856.
[37] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, 2013.
[38] J. R. Galbraith, *Organization Design*. Reading, MA, USA: Addison-Wesley, 1977.
[39] N. Sivasubramaniam, S. J. Liebowitz, and C. L. Lackman, "Determinants of new product development team performance: A meta-analytic review," *J. Prod. Innov. Manage.*, vol. 29, no. 5, pp. 803–820, 2012.
[40] H. Jaber, F. Marle, and M. Jankovic, "Improving collaborative decision making in new product development projects using clustering algorithms," *IEEE Trans. Eng. Manage.*, vol. 62, no. 4, pp. 475–483, Nov. 2015.
[41] R. Reagans and E. W. Zuckerman, "Networks, diversity, and productivity: The social capital of corporate R&D teams," *Organ. Sci.*, vol. 12, no. 4, pp. 502–517, 2001.
[42] S. D. Eppinger, "Innovation at the speed of information," *Harvard Bus. Rev*, vol. 79, no. 1, pp. 149–158, 2001.
[43] M. E. Sosa and F. Marle, "Assembling creative teams in new product development using creative team familiarity," *J. Mech. Des.*, vol. 135, no. 8, 2013, Art. no. 081009.
[44] F. Marle and L. A. Vidal, "Project risk management processes: Improving coordination using a clustering approach," *Res. Eng. Des.*, vol. 22, no. 3, pp. 189–206, 2011.

Qing Yang received the B.E., M.E., and Ph.D. degrees in aeronautics and astronautics technology from Northwestern Polytechnical University, Xi'an, China, in 1991, 1996, and 2003 respectively.

He is a Professor of Project Management with the School of Economics & Management, the University of Science & Technology Beijing, Beijing, China. He has prior work experience with an aerospace company and served as a Visiting Scholar with the Department of Industrial Engineering, Pennsylvania State University, State College, PA, USA, in 2010. He currently serves as the International Editorial Board Member of *International journal of Project Management*, and other journals. His research interests include managing the development process of complex engineering project, the design structure matrix, and design process modeling.

Na Yang received the master's degree in logistics engineering from Beijing Wuzi University, Beijing, China. She is currently working toward the Ph.D. degree in project management with the School of Economics & Management, University of Science & Technology Beijing, Beijing, China.

Her current research interests include project management, optimization and clustering.

Tyson R. Browning received the B.S. degree in engineering physics from Abilene Christian University, Abilene, TX, USA, in 1993, two S.M. degrees in aeronautics and astronautics and technology and policy, both in 1996, and the Ph.D. degree in technology, management and policy from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1999.

He is currently a Professor of Operations Management with the Neeley School of Business, Texas Christian University, Fort Worth, TX, USA, where he conducts research on managing complex projects and teaches MBA courses on project management, operations management, risk management, and process improvement. He has served as a Consultant for several organizations and has prior work experience with Lockheed Martin and Honeywell. His research papers have appeared in the IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, the *Journal of Operations Management*, *Manufacturing and Service Operations Management*, *Production and Operations Management*, *Systems Engineering*, and other journals. He has coauthored a book on the design structure matrix.

Dr. Browning is a former Department Editor for the IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT and currently serves as Co-Editor-in-Chief of the *Journal of Operations Management*.

Bin Jiang received the B.S. degree in telecoms engineering from the Nanjing Institute of Communication Engineering, Nanjing, China, in 1986, the MBA degree in international business from the University of Southern California, Los Angeles, USA, in 2001, and the Ph.D. degree in operations management from the University of Texas at Arlington, USA, in 2004. He is a Professor of Operations Management and the Driehaus Fellow with DePaul University, Chicago, IL, USA. His research focuses on outsourcing and supply chain management. He has authored more than 50 research articles.

Dr. Jiang's some of the articles were awarded the Best Paper in Journal of Operations Management, the Stan Hardy Award, the American Production and Inventory Control Society's Best Paper Award, and the Second Place of Production and Operations Management Society's Wickham Skinner Paper Award.

Tao Yao received the Ph.D. degree in management science and engineering from Stanford University, Stanford, CA, USA.

He is an Associate Professor with the Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, State College, PA, USA. His research interests include optimization, stochastic modeling, analytics, statistics, machine learning, and game theory.

Dr. Yao was the recipient of an honorable mention in the INFORMS George B. Dantzig Dissertation Award, several best paper awards from IERC Supply Chain and Logistics track and CIS track.