**20TH INTERNATIONAL DEPENDENCY AND STRUCTURE MODELING CONFERENCE, DSM 2018**

TRIESTE, ITALY, OCTOBER 15 – 17, 2018

# Clustering Organization Structure in Product Development Projects Using Similarity

Na Yang[1], Qing Yang[1], Tao Yao[2]

[1]School of Economics and Management, University of Science & Technology Beijing, China
[2]The Harold & Inge Marcus Department of Industrial & Manufacturing Engineering, Pennsylvania State University, USA

**Abstract:** A product development (PD) project is a complex social network, in which teams have their own attributes and are related by information flow. Similar team attributes and the complex patterns of technical dependency among teams both affect organization modularity. This paper provides an innovative spectral clustering approach that merge team attributes and relationship of teams. To measure the similarity of PD teams, we analyze the similar attributes of team and build structural models to capture the technical communication dependency among teams via the product-organization multi-domain matrix (MDM). We use two metrics to evaluate the clustering solutions and confirm that the proposed approach provides effective reduction of PD coordination complexity.

*Keywords: product development, organization design, design structure matrix (DSM), similarity, spectral clustering*

## 1 Introduction

A key managerial issue in product development (PD) is how to establish an effective organization architecture, because the complexity of interactions among which may reduce efficiency and introduce additional risks (Yang et al., 2014). A common but challenging objective in organization architecting concerns modularity—i.e., parsing the set of organizational elements (e.g., teams or individuals) into subsets, groups, or modules, such that the elements' relationships within each group are much stronger than those across groups (Tripathy and Eppinger, 2013). Many prior studies have applied some kind of clustering algorithm to optimize a model of the organization architecture, such as an organization design structure matrix (org DSM) (Tripathy and Eppinger, 2013; Yang et al., 2014).

Classical clustering algorithms are popular. For example, k-means (Ahmad & Hashmi, 2016) are based on the node attributes, while Fast-Newman algorithm (Newman, 2004) focuses on relationship. Most of clustering algorithms separate the attribute and relationship of nodes while clustering a complex graph. In fact, they both affect the results of modularity. For example, similar interest and friendship make two users close to each other in social network. Therefore, we aimed to formulate the DSM clustering problem combined the attributes and relationship of teams. Team attributes are based on social similarity with respect to significant background characteristics, such as race, sex and level of education et al. Teams who share important social characteristics are presumed to have common experiences, leading to shared knowledge.

Spectral clustering algorithm based on graph theory provides a stronger and more stable approach for finding the global optimum (Schaeffer, 2007; Sarkar et al., 2014), especially for non-convex datasets, and are well suited for application to real problems (Sarkar et al., 2014). The spectral clustering algorithm maximizes intra-cluster similarity and minimizes inter-cluster similarity. The similarity matrix is thus a critical input to a spectral clustering algorithm (Schaeffer, 2007). Many researchers have developed methods to measure similarity (Schaeffer, 2007).

Amount of research highlights the importance of similarity between teams or members for team process, such as team functioning and knowledge exchange. Larzarsfeld and Merton(1954) believes that interactions are more likely to occur between members or teams that are similar to each other. The similarity of knowledge bases inherent results in the recipient and partner team being more inclined to interact with one another and being able to understand the linkages between one another's knowledge stocks, which provides more favorable conditions for knowledge sharing. Therefore, the more similar team attributes are, the more intensive communication and interaction will be.

In this paper, we present an improved optimization approach, based on spectral clustering, that accounts for the similarity of teams in the PD organization.

# 2 An Improved spectral clustering for measuring modularity

It is important to take both attribute of teams and relationships between them into consideration. Thus we define a similarity matrix which merges team attribute and relationship. First, we analyze the similar characteristics of team, such as product-related expertise, process-related expertise and so on, which enhance the formation of relationships and interactions among them. Then, we infer technical communication strength among teams which reflect each team's role toward the design of components. Finally, we establish the cluster model of the graph containing both attribute of teams and relationship between them.

## 2.1 The attributes of the organization team

There are a lot of similar characteristics when selecting a cooperative team. For example, social-category similarity, work-style similarity, similar work habits and ethics (Zellmer-Bruhn et al., 2008) and so on. This paper examines two types of similarity attributes between teams− product-related expertise and process-related expertise.

### 2.1.1 Product-related expertise

Sosa (2011) defined product-related expertise that is associated with the specific functional and architectural attributes of the product under development. To collect data on areas of expertise, we ask them to indicate "the areas in which they considered themselves experts" based on what component they complete. Teams could select from $n$ areas of product-related expertise which provided a more granular description of each area of expertise, was assembled by a technical product manager. The score of team's product-related expertise that is between $0$ and $1$ is ascertained by the project manager, design engineers, and other subject matter experts, according to their knowledge and experience, which reflects team members had expertise relevant to area of product-related expertise. The product-related expertise differential between team *i* and *j* can be calculated with equation (1):

$$M_{ij}^{PE} = \sqrt{\sum_{k=1}^{n} (P_{ik} - P_{jk})^2} \qquad (1)$$

where $P$ captures the team's product-related expertise technologies. Then we devised the expertise differential $M_{ij}^{PE}$ based on the Euclidean distance between *i* and *j*.

2.1.2 Process-related expertise

Sosa (2011) defined process-related expertise that is associated with the procedures and activities associated with product development generally. For example, "process and product management," "product conception," "system design" and so on. The score of team's process-related expertise that is between *0* and *1* is the same as product-related expertise. The process-related expertise differential between team *i* and *j* can be calculated with equation (2):

$$M_{ij}^{TE} = \sqrt{\sum_{k=1}^{n} (T_{ik} - T_{jk})^2} \qquad (2)$$

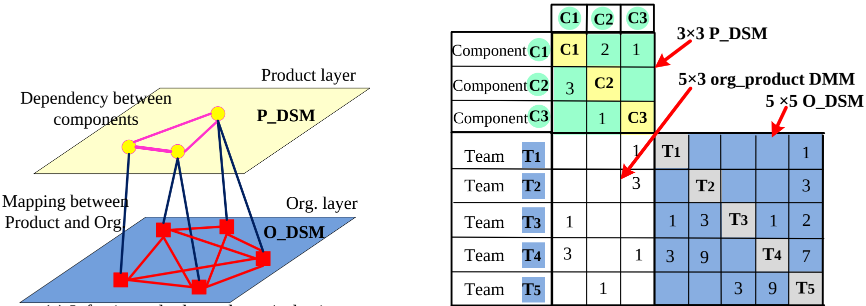
where $T$ captures the team's product-related expertise categories. Then we devised the expertise differential $M_{ij}^{TE}$ based on the Euclidean distance between *i* and *j*.

So, the total differences between team *i* and *j* can be calculated with Eq. (3), where $\omega_1, \omega_2$ are weight coefficients, $\omega_1 + \omega_2 = 1$. In this paper, we discuss only the case when $w_1 = w_2 = 0.5$.

$$M_{ij}^{TD} = \omega_1 M_{ij}^{PE} + \omega_2 M_{ij}^{TE} \qquad (3)$$

**2.2 Modeling the relationship between organization team via product-organization MDM**

We adopt an approach, recently proposed by (Yang et al., 2014), to derive the technical dependency between teams in org DSM from an MDM model inclusive of a product DSM and an organization-product DMM, as shown in Fig. 1(b). In the upper-left of the MDM, product DSM *P_DSM* models the technical communication among teams at the component level, which reflects the roles of teams in the design process of components containing some functions and allows teams to maintain control over all the functions that perform related tasks. And in the lower-left of the MDM, $DMM_{OP}(i, I)$ models the degree of involvement (e.g., the consumed time) of team *i* in the design of component *I*. For example, the *P_DSM* (3, 2) is nonzero in the product DSM, which means the design of component C2 will directly impact C3. Further, from the column of $DMM_{OP}$, we find that teams T5 and T4 responsible for developing product components C2 and C3 respectively. Then, we can infer a dependency of T4 on T5 which reflects the direct role relationship between these teams in the designing process of components C2 and C3.




(a) Inferring role dependency/cohesion between teams from the product layer

(b) Org_product MDM

**Fig.1.** Modeling technical dependency among teams via MDM

So, we derive the technical dependency between teams in org DSM from the product DSM via the DMM (which are all part of the MDM in Fig. 1(b)). The org DSM, $O\_DSM(i,j)$, to the right of Fig. 1(b), reflects the integrated effects of the dependency relationships among the product components and the teams' degrees of involvement in the components' design. Hence, using $P\_DSM$ and $DMM_{OP}$, the technical communication strength between teams $i$ and $j$ is modeled as:

$$O\_DSM(i,j) = \sum_{I=1}^{p} (DMM_{OP}(i,I) \times \sum_{J=1, J\neq I}^{p} (DMM_{OP}(j,J) \times (P\_DSM(I,J) + P\_DSM(J,I))) \tag{4}$$

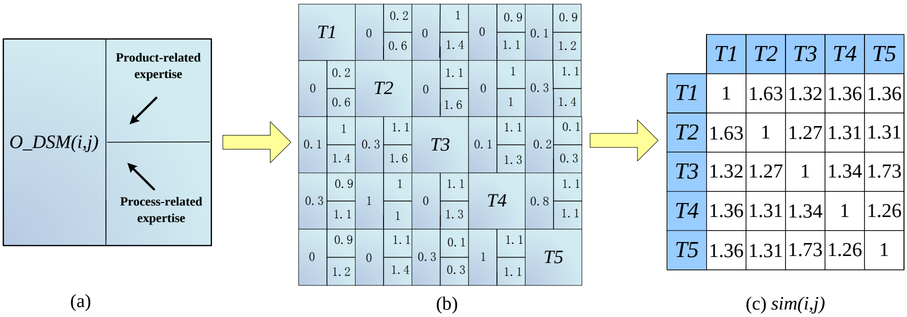
In this paper, the value of $P\_DSM$ and $DMM_{OP}$ are evaluated by analyzing the functional dependency relationships among components and the team's involvement degree in the component's design, respectively, as ascertained by the project manager, design engineers, and other subject matter experts, according to their knowledge and experience. $P\_DSM(I, J)$ and $DMM_{OP}(i, I)$ model the relationship at four levels: 0 = none, 1 = weak/low, 2 = medium, and 3 = strong/high. We normalize $O\_DSM$ by dividing all cells by the maximum cell value, thereby bounding all values in $O\_DSM\ (i,j)$ in [0, 1].

### 2.3 Building the Similarity Matrix of PD Teams

The differences of attributes between team $i$ and $j$ is defined as $M_{ij}^{TD}$.All the relationship can be denoted by $O\_DSM(i,j)$.In order to merge the attribute and relationship of teams, we define the similarity matrix containing both information (attribute and relationship) of the entire graph. Thus, for each pair of team $i$ and $j$, $S_{ij} = sim(i,j)$, in which S represents the ultimate similarity matrix. In this experiment, on the base of data density (Yi Xu et al., 2018), the functions are defined as follows:

$$sim(i,j)=\alpha\times e^{-\frac{O\_DSM(i,j)+O\_DSM(j,i)}{\sum_{u\in\Gamma(i)}(O\_DSM(i,u)+O\_DSM(u,i))+\sum_{u\in\Gamma(j)}(O\_DSM(j,u)+O\_DSM(u,j))}}+(1-\alpha)\times e^{-\frac{1}{M_{ij}+1}} \quad (5)$$

where $\Gamma(i)$ means the set of adjacent teams of *i*, and $\alpha$ means the similarity coefficient which is usually set as 0.4(Yi Xu et al., 2018). Data Density methods discover dense regions in space, where objects are adjacent to each other and separate them from sparse regions.




| | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|
| T1 | 1 | 1.63 | 1.32 | 1.36 | 1.36 |
| T2 | 1.63 | 1 | 1.27 | 1.31 | 1.31 |
| T3 | 1.32 | 1.27 | 1 | 1.34 | 1.73 |
| T4 | 1.36 | 1.31 | 1.34 | 1 | 1.26 |
| T5 | 1.36 | 1.31 | 1.73 | 1.26 | 1 |

Fig.2. An example of calculating the sim matrix

Fig. 2 provides an example of calculating the similarity between teams. Fig. 2(a) can be captured with Eqs. (1)-(4).

### 2.4 Spectral Clustering Approach

Spectral clustering techniques make use of the spectrum of the data's similarity matrix to perform dimensionality reduction before clustering the data in fewer dimensions. The similarity matrix is an input to spectral clustering and the optimal partition maximizes the similarity of elements in the cluster (or subgraph) while minimizing the similarity between elements in different clusters. Ng-Jordan-Weiss (NJW) algorithm (Ng et al., 2002), which utilizes the Laplacian matrix, a simple normalization of the similarity matrix to optimize the normalized cut criterion according to the eigenvectors associated with the largest eigenvalues. We apply the following NJW algorithm-based, normalized spectral clustering procedure (Ng et al., 2002) because of its more robust performance.
We use two metrics to evaluate the clustering solutions. First, we adapt the *numerical dependency density* (NDd) measure (Chen and Lin 2003), the ratio of the total interaction strength (TIS) of all (non-zero) elements outside the clusters to the total number of cells outside the clusters:

$$NDd=\frac{TIS}{cell\_out} \quad (6)$$

Second, we use the global Silhouette index of the clustering (Slobodan Petrovi´c, 2006), which measures the quality of clustering by calculating the distance between each cluster and the distance between each team in the cluster. The definition of Silhouette index is as follows:

$$S(k)=\frac{1}{k}\sum_{i=1}^{k}\left\{\frac{1}{m_i}\sum_{i=1}^{mi}\frac{b_j^i-a_j^i}{\max\left[a_j^i,b_j^i\right]}\right\} \tag{7}$$

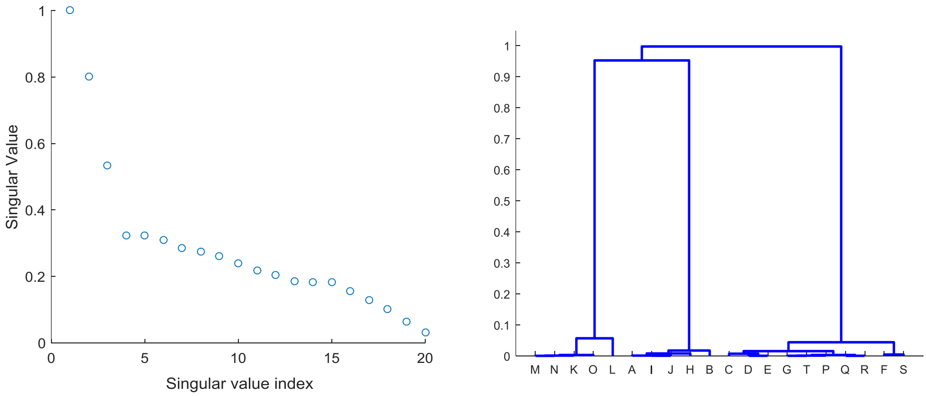
*where* $$a_j^i=\frac{1}{m_i-1}\sum_{k=1,k\neq j}^{mi}d(T_j^i,T_k^i),\ j=1,...,mi.$$

$$b_j^i=\min_{n=1,...,k;n\neq i}\left[\frac{1}{m_n}\sum_{k=1}^{m_n}d(T_j^i,T_k^n)\right],j=1,...,mi.$$

where $O=\{C_1,C_2,...,C_k\}$ is its clustering into k clusters, $d(T_k,T_l)$ is the distance between $T_k$ and $T_l$, $C_i=\{T_1^i,...,T_{mi}^i\}$ is the $i$-th cluster, $i=1,...,k$ and $m_i=|C_i|$. The global silhouette take values between -1 and 1, the maximum value of which indicate the best clustering result.

## 3 Case Studies

We applied the proposed concepts and models to a PD project in an IT company involving 20 teams and 18 components. Based on the responses and other information provided, we built the product DSM, and the product-organization DMM. *P_DSM*(*I*, *J*) are measured by the added cost on component *I* when component *J* is designed or redesigned and *DMM*(*i*, *I*) are measured by the time required of team *i* in the design of component *I*.
First, using equation (4), we derived the technical communication/dependency strength among the teams. Next, we calculated the similarity matrix with equations (1)-(5) and applied the spectral clustering procedure in the Matlab® 15 software.




(a) (b)

Fig.3. Results of singular value and cluster tree using spectral clustering

Fig. 3(a) shows the sigular values for the similarity matrix, which is composed by the attributes and relationship of organizaion teams. 3 large singular values appear, which signals the appearance of 3 modules in the organization. Sarkar(2014) found that the number of outlying eigen or singular values, separated from the bulk of the spectrum, provides a good estimate of the actual number of modules in the system.
Fig. 3(b) shows the results of the modularity analysis: group 1 from teams 13 to 12(i.e., G1 [M, N, K, O, L]), group 2 from teams 1 to 2(i.e., G2 [A, I, J, H, B]), group 3 from teams 3 to 19(i.e., G3 [C, D, E, G, T, P, Q, R, F, S]).

| | *M* | *N* | *K* | *O* | *L* | *A* | *I* | *J* | *H* | *B* | *C* | *D* | *E* | *G* | *T* | *P* | *Q* | *R* | *F* | *S* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *M* | *M* | .6 | .6 | .5 | .1 | | .5 | | | | | | | | | | | | | |
| *N* | .6 | *N* | .5 | .3 | .6 | | | | | | | | | | | | | | | |
| *K* | .5 | .5 | *K* | .3 | .6 | | | | | .3 | | .2 | | | | | | | | |
| *O* | .7 | .3 | .7 | *O* | .5 | | | | | | .1 | | | | | | | | | |
| *L* | .5 | .4 | .5 | .4 | *L* | .1 | | .2 | | | | | | | | | | | | |
| *A* | | | | | .1 | A | .1 | .5 | .1 | .7 | .1 | .1 | | .1 | | | | | .1 | |
| *I* | .5 | | | | | .9 | I | .2 | .5 | .1 | | | | | | | | | | |
| *J* | | | | | .4 | .5 | .2 | J | .6 | | | | | | | | | | | |
| *H* | | | | | | .1 | .5 | .6 | H | .1 | | | | | | | | | | |
| *B* | | | .3 | | | .7 | .1 | | | B | .1 | .9 | | .1 | | | | | | |
| *C* | | .3 | | | | .1 | | | | .5 | C | .3 | .5 | .6 | .3 | .3 | | | .1 | |
| *D* | | | | | | .1 | | | | | .3 | D | .4 | .9 | | | | | .9 | |
| *E* | | | | | | | | | | | .1 | .4 | *E* | .3 | | | | | .2 | |
| *G* | | | | | | .1 | | | | .1 | .1 | .9 | .3 | *G* | | | | | 1 | |
| *T* | | | | | | | | | | | .5 | | | | *T* | .5 | .8 | .3 | | |
| *P* | | | | | | | | | | | .3 | | | | | *P* | | | | |
| *Q* | | | | | | | | | | | .5 | | | .1 | .8 | .5 | *Q* | .5 | | .8 |
| *R* | | | | | | | | | | | .5 | | | | .3 | .6 | .5 | *R* | | .3 |
| *F* | | | | | | .1 | | | | | .1 | .9 | .2 | 1 | | | | | *F* | |
| *S* | | | | | | | | | | | .6 | | | | .8 | .8 | .3 | | | *S* |

Fig.4. Clustered O_DSM

Fig. 4 shows the resulting, clustered O_DSM that teams with high similarity (i.e., strong information exchange) are brought together in groups while connections between groups become weaker, thereby reducing the coordination challenges.
The Ndd of our proposed spectral clustering method is 0.022 and the Silhouette index of our method is 0.5323, which indicate the clustering result is well.

## 4 Conclusions

This paper provides a framework that enables managers to design a PD organization that can be coordinated more efficiently and effectively. The proposed approach of constructing the similarity avoids the use of Radial Basis Function, imports similar team attributes and the directed relationship into the similarity matrix.
The main limitations of this research are: how to quantify team attributes is very difficult; benchmarking our method against other clustering methods when it is very difficult to judge which one is the best (e.g., the applied situation may vary) and obtain (or reproduce) their programs.
Several aspects of the model presented in this paper merit further examination in future research. First, from the experiments, the attributes of teams can greatly affect the clustering results. There probably exist more factors we have not considered. Second, other data collection methods and dependency measurement methods theory that reduce the ambiguity of respondents’ judgments.

## Acknowledgments

This study was supported by the National Science Foundation of China (No. 71472013, No. 71528005 and No. 71872011)

## Reference

Ahmad, A., & Hashmi, S. (2016). K-harmonic means type clustering algorithm for mixed datasets. *Applied Soft Computing, 48*, 39-49.

Chen S. J., L. Lin., 2003. Decomposition of interdependent activity group for concurrent engineering. Computers & Industrial Engineering, 44(3):435-459.

Lazarsfeld, P. F., & Merton, R. K., 1954. Friendship as a social process: A substantive and methodological analysis. Freedom and control in modern society, 18(1), 18-66.

Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys, 69*(6 Pt 2), 066133.

Ng, A. Y., M.I. Jordan, Y. Weiss., 2001. On spectral clustering: analysis and an algorithm. Proceedings of Advances in Neural Information Processing Systems14, pp. 849-856. MIT Press, Cambridge

Petrovic, S., 2006, October. A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. In Proceedings of the 11th Nordic Workshop of Secure IT Systems (pp. 53-64).

Sarkar, S., A. Dong, J. A. Henderson, P. A. Robisnon., 2014. Spectral characterization of hierarchical modularity in product architectures. Journal of Mechanical Design, 136(1):249-256.

Schaeffer, S. E., 2007. Survey: graph clustering. Computer Science Review, 1(1): 27-64.

Sosa, M. E. (2011). Where do creative interactions come from? The role of tie content and social networks. *Organization Science, 22*(1), 1-21.

Tripathy, A., S.D.Eppinger., 2013. Structuringbwork distribution for product development organizations. Production & Operations Management, 22(6): 1557-1575.

Xu, Y., Zhuang, Z., Li, W., & Zhou, X., 2017. Effective community division based on improved spectral clustering. Neurocomputing.

Yang, Q., T. Lu, T. Yao, B. Zhang., 2014. The impact of uncertainty and ambiguity related to iteration and overlapping on schedule of product development projects. International Journal of Project Management, 32(5):827–837.

Zellmer-Bruhn, M. E., Maloney, M. M., Bhappu, A. D., & Salvador, R. B., 2008. When and how do differences matter? An exploration of perceived similarity in teams. *Organizational Behavior and Human Decision Processes*, *107*(1), 41-59.

**Contact: Qing Yang,** University of Science & Technology Beijing, School of Economics and Management, No.30 Xueyuan Road, Haidian District, 100083 Beijing, China, 13911463353, yangqing@manage.ustb.edu.cn.