Discovering Cross-disciplinary Talent Teams Based on Nonnegative Matrix Tri-Factorization

Yu Pan[1], Pei Wang[1], Jingbo Huang[1], Zheng Zeng[2], FengYao[1*]

[1] National University of Defense Technology, Changsha, CHN
[2] Academy of Military Sciences, Beijing, CHN

*panyu0511@nudt.edu.cn

Abstract. The importance of interdisciplinary talent teams is becoming increasingly prominent. Talents of these teams often have multiple identities and belong to several teams simultaneously. In this paper, we employ the methods and theories of complex networks to discover cross-disciplinary talent teams. Firstly, we proposed a unified nonnegative matrix tri-factorization framework, in which the talent association information and research direction information are jointly utilized for collaborative learning and effective integration. Moreover, the talent team relationship matrix is introduced to capture and quantify the tightness of the association between the talent teams, and regularization constraints are applied to ensure the consistency of relationships among talent teams. Finally, a strategy for dividing multidisciplinary and intersecting talent teams is designed, enabling the precise discovery of multidisciplinary talent teams.

Keywords: Cross-disciplinary talent team discovery, Nonnegative matrix tri-factorization, Human resource management.

1 Introduction

With the relentless advancement of technology, the boundaries of the combat domain are becoming increasingly blurred, and traditional knowledge within a single discipline is no longer sufficient to meet the demands of modern combat forces. Consequently, the construction and maintenance of new types of combat forces rely on a team of talents who can span multiple disciplinary fields and possess the ability to integrate diverse and intersecting competencies. These individuals must not only have a profound knowledge base within their respective areas of expertise but also be capable of understanding and applying theories and practices from other disciplines to form an integrated combat capability [1] [2].

The field of new combat forces encompasses multiple disciplinary directions, creating a complex demand for interdisciplinary and multi-domain talent. Individual talents can belong to multiple teams, and different talent teams often have overlapping parts. Against this backdrop, the discovery and cultivation of interdisciplinary talent teams

S. M. Ghadiri et al. (eds.), *Proceedings of the 2025 4th International Conference on Engineering Management and Information Science (EMIS 2025)*, Advances in Computer Science Research 121,
https://doi.org/10.2991/978-94-6463-736-6_20

have become key to the pre-positioning and replenishment of new combat forces. Members of these teams often belong to multiple teams, with their knowledge and skills overlapping across different teams. This structure provides combat forces with flexible adaptability and innovation capabilities.

The discovery and construction of interdisciplinary talent teams not only promote the exchange of knowledge and collaborative innovation between different disciplines but also provide more effective solutions when facing complex and changing combat environments.

2 The Framework of DCTT

2.1 Method

Firstly, we construct the talent association matrix $A \in R^{n\times n}$, where n is the number of talents. If there exists an academic association between talent i is j, then $A_{ij}=1$ otherwise, $A_{ij}=0$. At the same time, we construct the talent research direction attribute matrix $X \in R^{n\times m}$, where m is the number of research directions, if the research direction of talent i is j, $X_{ij}=1$, otherwise $X_{ij}=0$. If there is an academic connection between the two talents, such as an academic citation relationship, we believe that the research direction of the two talents is similar.

For the discovery of cross-disciplinary talent teams, the closeness of association between different teams holds significant practical importance. In this paper, the non-negative matrix tri-factorization framework is introduced, which can obtain accurate results of talent team discovery and the closeness of the relationship between talent teams. First, we introduce the talent indicator matrix H, where H_{ij} represents the probability that talent i belongs to team j. Additionally, we introduce the team relationship matrix S, where S_{ij} indicates the degree of relationship association between team i and team j. The expected association between talents i and j is denoted by $\sum_{r=1}^{k}\sum_{q=1}^{k} H_{ir}S_{rq}H_{jq}$. The probability of association between two individuals in the same team is higher, and the expected association between talents should be as consistent as possible with the talent association matrix A. Therefore, the optimization objective function can be constructed as follows:

$$\min_{H\in R^{n\times k}, S\in R^{k\times k}} \| A - HSH^T \|_F^2 \quad s.t.\ H_{ij} \geq 0, S_{ij} \geq 0, \forall i,j \tag{1}$$

Similarly, we introduce the research direction indicator matrix U, where U_{ij} represents the probability that team i possesses research direction j. If an individual's research direction is similar to that of a team, then the probability that the talent belongs to that team is relatively high. Since the team information implied by the talent association relationships should be consistent with that implied by the research directions, we construct the following objective function:

$$\min_{U\in R^{n\times k},H\in R^{m\times k}} \| H - XU \|_F^2 \quad s.t.\ U_{ij} \ge 0, H_{ij} \ge 0, \forall i,j \tag{2}$$

We believe that not only the team structure implied by the talent association information and the research direction information is consistent, but also the implied relationship between the teams should be consistent. Therefore, the similarity between any two teams i and j($\sum_{r=1}^{m} U_{ri}U_{rj}$) should be as consistent as possible with the team relationship matrix S. Consequently, we construct the following regularization term:

$$\min_{H\in R^{n\times k},S\in R^{k\times k}} \| S - U^T U \|_F^2 \tag{3}$$

The above formula is the Frobenius norm of the difference between the team relationship matrix S and the element-wise product of H and its transpose H^T, ensuring that the inferred team relationships from talent associations are consistent with the actual team relationships.

In summary, the final objective function can be obtained as follows:

$$\begin{aligned} &\min_{U\in R^{n\times k},H\in R^{m\times k},S\in R^{k\times k}} \| A - HSH^T \|_F^2 + \lambda_1 \| H - XU \|_F^2 + \lambda_2 \| S - U^T U \|_F^2 \\ &s.t.\ U_{ij} \ge 0, S_{ij} \ge 0, H_{ij} \ge 0, \forall i,j \end{aligned} \tag{4}$$

where λ_1 and λ_2 are regularization parameters that control the weight of this consistency constraint in the overall objective function.

2.2 Optimization

We adopt the iterative optimization method to solve the objective function (4). Let α, β, ω be the Lagrange multipliers, constraining equations $U \ge 0, S \ge 0, H \ge 0$ respectively. Then, the Lagrangian function for equation (4) is:

$$\begin{aligned} l &= \| A - HSH^T \| + \lambda_1 \| H - XU \|_F^2 + \lambda_2 \| S - U^T U \|_F^2 - (\alpha H + \beta S + \omega U) \\ &= tr[(A - HSH^T)(A - HSH^T)^T] + \lambda_1 tr[(H - XU)(H - XU)^T] \\ &\quad + \lambda_2 tr[(S - U^T U)(S - U^T U)^T] - (\alpha H + \beta S + \omega U) \\ &= tr(AA^T - AHS^T H^T - HSH^T A + HSH^T HS^T H^T) + \lambda_1 tr(HH^T - HU^T X^T - XUH^T \\ &\quad + XUU^T X^T) + \lambda_2 tr(SS^T - SU^T U - S^T U^T U + U^T UU^T U) - (\alpha H + \beta S + \omega U) \end{aligned} \tag{5}$$

We can obtain the following update rules:

$$H_{ij} \leftarrow H_{ij} \sqrt{\frac{(AHS^T + AHS + \lambda_1 XU)_{ij}}{(HSH^T (HS + HS^T) + \lambda_1 H)_{ij}}} \tag{6}$$

$$S_{ij} \leftarrow S_{ij} \sqrt{\frac{(H^T AH + \lambda_2 U^T U)_{ij}}{(H^T HSH^T H + \lambda_2 S)_{ij}}} \tag{7}$$

$$U_{ij} \leftarrow U_{ij} \sqrt{\frac{(\lambda_1 X^T H + \lambda_2 (US + US^T))_{ij}}{(\lambda_1 X^T XU + 2\lambda_2 UU^T U)_{ij}}} \tag{8}$$

2.3 Discovery Strategy of Cross-disciplinary Talent Teams

Through iterative optimization, we can obtain the talent indicator matrix H and the talent team relationship matrix S. For the division of cross-disciplinary teams, a talent can be affiliated with multiple teams. In this paper, we propose an overlapping partitioning strategy to identify the labels of the teams to which talents belong, thereby achieving the discovery of cross-disciplinary talent teams. Recall that, in the above section, H is the talent indicator matrix which H_{ij} represents the probability that the talent i belongs to the talent team j. For the cross-disciplinary talent team discovery, a talent may participate in more than one talent team. Therefore, in addition to dividing the talent into a talent team corresponding to the maximum of the indicator matrix, we also need to consider whether it belongs to other talent teams. We propose a partition principle to find talent team labels, whose probability of belonging to the team is greater than a threshold value. The discovery strategy of cross-disciplinary talent teams is as follows:

$$C_i = \left(\frac{H_{ic} - \overline{H_{i\cdot}}}{\max H_{i\cdot} - \overline{H_{i\cdot}}} \geq \omega \right) \tag{9}$$

3 Experiment

3.1 Evaluation Metrics and Baselines

In the experiment, we select two baseline algorithms to evaluate the performance of the proposed method. The SCI [3] (Semantic Community Identification method) method, which can mine the group structure and semantic description simultaneously. The other algorithm is SNMF [4] (Symmetric Nonnegative Matrix Factorization), which only focuses on the connectivity structure of the network and disregards node attributes. The two baseline algorithms develop a flexible membership distribution matrix that encompasses various groups, enabling the identification of overlapping groups.

To evaluate the accuracy of the cross-disciplinary talent team discovery, we employ the Overlapping Normalized Mutual Information (ONMI) [5] metric, which quantifies the congruence between the predicted and actual group structures.

$$ONMI\left(A,B\right)=\frac{-2\sum_{i=1}^{k_A}\sum_{j=1}^{k_B}N_{ij}\log\left(\frac{nN_{ij}}{N_{i.}N_{.j}}\right)}{\sum_{i=1}^{k_A}N_{i.}\log\left(\frac{N_{i.}}{n}\right)+\sum_{j=1}^{k_B}N_{.j}\log\left(\frac{N_{.j}}{n}\right)} \tag{10}$$

where k_A and k_B present the number of groups in A and B respectively. N_{ij} denote the number of nodes in the i-th group of A that appear in the j-th group of B. The higher value indicates better performance and accurate division.

3.2 Evaluation of Real-world Datasets

We first verify the performance of the DCTT algorithm using four real-world datasets: Citeseer, Cora, Engineering, and Politics-ie. The Cora and Citeseer are citation networks. The Politics-ie dataset is a network of Irish politicians, containing information on 348 Irish politicians on Twitter. The Engineering dataset is a co-authorship network, which is constructed by analyzing author cooperation information in Microsoft Academic Graph. The experimental datasets are described as follows, and the details are listed in Table 1.

Table 1. Statistics of real-world datasets

Dataset	Nodes	Edges	Attributes
Citeseer	3312	4660	3703
Cora	2708	5278	1433
Engineering	14927	98610	4800
Politics-ie	419	27340	3614

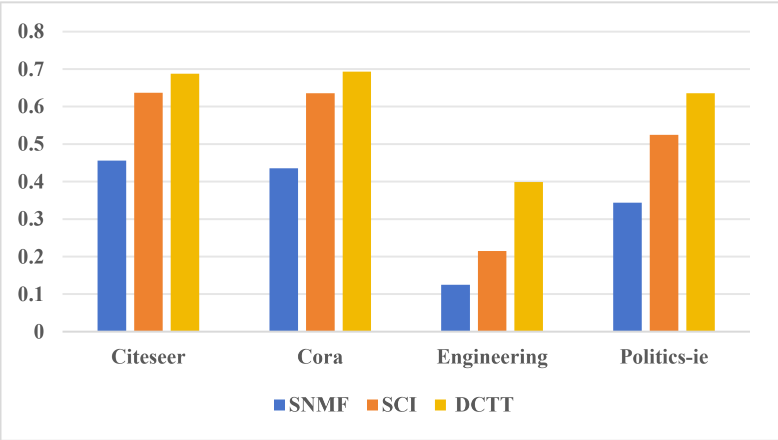

Fig. 1. The ONMI score of the algorithm on the four real-world datasets.

The performance of the algorithms on the four real-world datasets is presented in Figure 1. As shown in Figure 1, the proposed DCTT achieved a higher OMNI score, indicating its strong ability to divide nodes into accurate groups. The superior results achieved by our algorithm can be attributed to its sophisticated approach to handling the intricacies of group structures and its ability to leverage the rich information present in the co-authorship networks. Real-world networks are often characterized by non-linear structures. Since our framework is based on shallow models, it has limited capability in discovering the organizational structure of large-scale real-world networks.

3.3 Evaluation of Synthetic Datasets

In this section, we verify the DCTT algorithm on the synthetic dataset generated by the Lancichinetti-Fortunato-Radicch (LFR) [6] benchmark, which can simulate a complex network's power-law distribution characteristics and small-world characteristics. It allows users to control the size and degree distribution of the entire network through parameters, which is widely used to simulate the community network. To evaluate the algorithm in attributed networks, we extended the LFR synthetic networks into attributed networks based on the model proposed in Reference [7]. The detailed parameter settings of the LFR benchmark are shown in Table 2.

Table 2. The parameters of the networks generated by the LFR benchmark.

Datasets	*n*	*k*	*minc*	*maxc*	*avgk*	*maxk*	τ_1	τ_2	μ
Network1	500	10	15	30	20	40	2	1	0.2
Network2	1000	22	15	50	20	40	2	1	0.2

Table 3. The ONMI score of the algorithm on the LFR benchmark.

Datasets	SNMF	SCI	DCTT
Network1	0.5936	0.61254	0.63254
Network2	0.4684	0.5124	0.5687

As shown in Table 3, our algorithm exhibited unparalleled efficacy, achieving the highest ONMI scores, indicating its superior capability to detect talents in overlapping talent teams. Because the SNMF algorithm only considers the structure information and ignores the attribute information, it can not get the best performance. Our proposed method combines the talent association information and research direction information in a unified framework and constrains the consistency of the relationship between talent teams through regularization. Therefore, accurate overlapping group partitioning can be obtained.

4 Conclusion

In this paper, we presented a novel method for discovering cross-disciplinary talent teams based on nonnegative matrix tri-factorization. Our approach integrates talent association information and research direction information to effectively identify and mine interdisciplinary talent teams, which is crucial for the pre-positioning and replenishment of modern combat forces. Through the introduction of the talent affiliation indicator matrix and the team relationship matrix, we can quantify and capture the closeness between different talent teams, leading to a more accurate and nuanced understanding of team dynamics. In this paper, we employ nonnegative matrix factorization to discover talent teams. However, this model is a shallow model and is insufficient in capturing non-linear relationships. In future research, deep learning models could be applied to talent team discovery.

References

1. Wei Y Y , Chen W F , Xie T ,et al.Cross-disciplinary curriculum integration spaces for emergency management engineering talentcultivation in higher education [J]. Computer Applications in Engineering Education, 2022, 30:1175 - 1189.
2. Zhao B.The Practical Path to Cultivating Innovative Talents: Implementing Cross-Disciplinary Learning Communities in Environmental Design Education [J]. Journal of Contemporary Educational Research, 2023,(11): 33-37.
3. Deng C, Lv Z, Liu W, et al. Multi-view matrix decomposition: a new scheme for exploring discriminative information[C]. International Conference on Artificial Intelligence. AAAI Press, 2015:3438-3444.
4. Ding C, He X, Simon H D, et al. On the Equivalence of Nonnegative Matrix Factorization and K-means Spectral Clustering[J]. Factorization, 2005.
5. Shen H, Cheng X, Cai K, et al. Detect Overlapping and Hierarchical Community Structure in Networks[J]. Physica A Statistical Mechanics & Its Applications, 2008, 388(8):1706-1712.
6. Lancichinetti A, Fortunato S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2009, 80(1):016118.
7. Huang Z, Zhong X, Wang Q, et al. Detecting community in attributed networks by dynamically exploring node attributes and topological structure[J]. Knowledge-Based Systems, 2020:105760.