Unleashing the Power of SVD and Louvain Community Detection for Enhanced Recommendations

TOKALA, Srilatha, ENDURI, Murali Krishna and TANGIRALA, Jaya Lakshmi <http://orcid.org/0000-0003-0183-4093>

Unleashing the power of SVD and Louvain Community Detection for Enhanced Recommendations

Srilatha Tokala
ACT Lab, Department of CSE,
SRM University-AP, India
srilatha_tokala@srmap.edu.in

Murali Krishna Enduri
ACT Lab, Department of CSE,
SRM University-AP, India
muralikrishna.e@srmap.edu.in

T. Jaya Lakshmi
ACT Lab, Department of CSE,
SRM University-AP, India
jayalakshmi.t@srmap.edu.in

***Abstract*—Recommendation systems play a vital role in delivering personalized content to users, thereby enhancing their overall experiences across diverse applications. Collaborative filtering based recommendation systems have demonstrated success through the application of matrix factorization techniques. However, the incessant growth in dataset size and complexity presents challenges regarding the scalability of recommendation algorithms. Consequently, efficiently addressing these scalability concerns becomes imperative to ensure the seamless functioning of recommendation systems in handling increasingly large and diverse datasets. This research introduces an innovative method that seamlessly integrates matrix factorization techniques and community detection algorithms to effectively tackle the scalability issue in recommendation systems. Through numerous experiments utilizing real-world datasets, the proposed method's efficiency is thoroughly assessed. These compelling findings underscore the method's potential as a promising solution for constructing robust and scalable recommendation systems, effectively mitigating the challenges posed by burgeoning datasets. Ultimately, the overarching objective is to enhance user experiences by providing personalized and relevant content recommendations that cater to the evolving needs of modern recommendation systems. By optimizing scalability and recommendation accuracy, this innovative approach seeks to elevate the efficacy and user satisfaction of recommendation systems across various domains.**

***Index Terms*—Recommendation, Collaborative filtering, Matrix factorization, Community detection**

I. INTRODUCTION

Matrix Factorization techniques play a pivotal role across a diverse array of fields and applications. The notable advantage lies in their ability to perform dimensionality reduction, transforming high-dimensional data into lower-dimensional representations [1]. By decomposing matrices into latent features, these methods effectively capture essential information while mitigating noise and redundancy. Such dimensionality reduction is particularly critical for optimizing storage, computation, and analysis of vast and complex datasets, especially those characterized by large-scale dimensions [2].

Within the domains of data analysis and network science, analyzing intricate datasets presents significant challenges in achieving computational efficiency and scalability. Such datasets may encompass large networks, user-item interaction matrices, or diverse information sources, underscoring the need for efficient analytical approaches. In our investigation, we leverage the power of Singular Value Decomposition (SVD), a foundational technique widely acclaimed for extracting dominant latent features and revealing underlying data structures. Additionally, we incorporate the Louvain method, a prominent community detection algorithm, to further augment the analysis of complex networks by uncovering meaningful community structures and interactions [3].

Through the integration of SVD and the Louvain method, our research enables the extraction of essential patterns from complex datasets while judiciously managing computational resources. Leveraging the strengths of these techniques, we address computational challenges in data analysis and network science, offering valuable insights into the relationships, interactions, and modular structures inherent in large-scale datasets [4]. The application of these advanced methods contributes to the development of scalable and potent data analysis tools, with broad applicability across various domains, including social networks, biological systems, recommendation engines, and information retrieval systems.

As complex networks grow in size and intricacy, their underlying structures pose increasingly challenging analytical tasks. To address this, community detection has emerged as a fundamental method in network analysis, becoming an indispensable tool for revealing the intricate organization and modular nature of such networks [5]. The core objective of community detection is to identify densely connected groups of nodes, referred to as communities or clusters, within the network. These communities exhibit a higher density of connections among their internal nodes compared to nodes outside the community. By unveiling these coherent and functionally related subgroups, community detection illuminates the network's organizational principles, revealing patterns of interaction, information flow, and influence [6]. Considering the immense complexity exhibited by real-world networks, community detection algorithms play a pivotal role in gaining deeper insights into the underlying structure and dynamics of these systems. Leveraging advancements in network science and computational techniques, community detection continues to evolve, offering new avenues for comprehending and harnessing the potential of complex networks in diverse domains [7].

With their wide-ranging applications, matrix factorization and community detection techniques continue to push the boundaries of data analysis, propelling research, and innovation in countless domains [8]. As data complexity continues to grow, this powerful synergy remains at the forefront of modern data science, empowering researchers, and practitioners to unlock hidden patterns and optimize decision-making processes [9]. Several notable areas that gained importance in SVD and Louvain are like

- **Recommender Systems:** In recommendation systems, the Louvain method discovers unified social groupings or communities, optimizing user experiences and content curation, whereas SVD uses collaborative filtering to forecast user preferences and offer personalized item suggestions [10].

- **Network Science:** The Louvain approach in a network assist in identifying functional modules in biological networks, and SVD-based latent semantic analysis identifies latent topics and relationships in textual data, which is helpful for biological and natural language processing tasks [11].
- **Data Compression and Storage:** Both methods improve data compression and storage effectiveness by reducing dataset dimensionality using SVD and identifying useful data clusters by the Louvain approach, which optimizes storage and transmission [12].
- **E-Commerce:** While the Louvain approach can divide clients into discrete groups to better personalized recommendations, SVD is used in e-commerce platforms to suggest products to customers based on their past purchases and interests [13].
- **Network Visualization:** The combination of SVD and Louvain facilitates a deeper understanding of network relationships by lowering the dimensionality of complicated networks through SVD and emphasizing community patterns through the Louvain approach [14].

The Louvain method and SVD combination continue to motivate creative applications in a variety of domains, enabling effective data analysis, better decision-making, and improved user experiences in the ever-expanding field of data-driven technology.

II. LITERATURE REVIEW

The singular value decomposition is one of the fundamental and crucial tools of contemporary numerical analysis, particularly numerical linear algebra. SVD was introduced by the French mathematician Augustin-Louis Cauchy in the early 1800s. The study by Lieven De Lathauwer *et al.* in 2000 explored a multilinear extension of the SVD that extends its application to higher-order tensors [15]. A robust analogy between the properties of matrices and the higher-order tensor decomposition, including uniqueness, the connection with the matrix eigenvalue decomposition, and the effects of first-order perturbations. Additionally, the impact of tensor symmetries on the decomposition process and propose a multilinear generalization of the symmetric eigenvalue decomposition, particularly tailored for pair-wise symmetric tensors. Arkadiusz Paterek *et al.* in the year 2007 introduced several extensions to regularized SVD in their research [16]. These extensions involve the addition of biases to the predictor set, post-processing SVD using kernel ridge regression, utilizing separate linear models for each movie, and applying methods like regularized SVD but with a reduced number of parameters. In 2015, Musrrat Ali *et al.* proposed a watermark embedding technique using SVD [17]. This method involves embedding watermark bits into the target blocks by modifying the first column coefficients of the left singular vector matrix, utilizing a threshold. To mitigate visible distortion caused by the watermark embedding, the coefficients of the right singular vector matrix are adjusted with compensation parameters. This approach ensures effective watermark embedding while minimizing perceptible distortions in the watermarked data. In 2021, Neeraj *et al.* introduced a novel method for detecting image forgery using Biorthogonal Wavelet Transform with Singular Value Decomposition (BWT-SVD) based feature extraction [18]. The proposed approach commences with dividing the test images into overlapping blocks. This technique offers a powerful tool for image forgery detection, enabling the identification of manipulated regions within images through effective feature extraction and analysis.

In 2020, Claudio D. G. Linhares *et al.* presented a novel methodology to evaluate the performance of community detection algorithms through network visualization [19]. Through comprehensive statistical and visual analytics, the study showcases how users can effectively identify the most suitable network community detection algorithm for specific network analysis tasks. This approach provides valuable insights into the strengths and weaknesses of various algorithms, facilitating informed decision-making and enhancing the accuracy and efficiency of community detection in complex networks. In 2021 Maryam Mohammadi *et al.* present an innovative approach called the ACLM algorithm, which harnesses the power of GPUs [20]. The algorithm optimizes thread allocation by dynamically assigning threads to each block based on the number of streaming multiprocessors (SMs) and warps required on the GPU. This adaptive allocation strategy leverages the GPU's parallel processing capabilities, leading to enhanced efficiency and accelerated computation in community detection tasks using the Louvain method. In 2022, Hang Qie, Shijie Li *et al.* introduce an innovative graph partition algorithm designed specifically for the parallel Louvain method [21]. Unlike existing graph partition algorithms, this novel approach divides the graph into isolated sets, where vertices are relatively decoupled from one another. The method efficiently computes and synchronizes information without any delay, allowing for seamless and real-time community swapping within these isolated sets.

III. METHODOLOGY

In the following section, you will find a concise explanation of the SVD and Louvain community detection methods.

A. Singular Value Decomposition (SVD)

The SVD method is crucial in recommender systems, especially in methods that use collaborative filtering. Through the capture of latent user preferences and item features, SVD enables recommender systems to generate personalized recommendations [22]. It reveals obscure patterns in user-item interaction data that enable the system to make recommendations based on each user's particular preferences. The user-item interaction matrix can be made into a lower-dimensional space with the help of SVD while still retaining the most crucial data. This procedure decreases data noise and increases calculation performance.

SVD method has been recognized as a foundational approach in recommender systems, with its initial application dating back to. In the context of matrix factorization (MF) using SVD, the rating matrix R undergoes a decomposition into three latent feature matrices: P, s, and Q, as illustrated in (1). The rating matrix R possesses dimensions m × n, while the latent factor matrices P, s, and Q have sizes m × m, m × n, and n × n, respectively. Notably, matrices P and Q are orthogonal, and matrix s is singular [23].

$$R \approx PsQ^T \qquad (1)$$

Upon computation, the latent feature matrices X and Y are derived from P and Q, respectively, in the following manner: X = P.s and Y = Q.

After updating the values for X and Y, the dot product of these updated latent feature matrices is considered as the predicted rating matrix (XY^T) as $\tilde{R}$. The discrepancy between the original rating matrix R and the predicted rating matrix R̃ is measured as the RMSE value, as illustrated in (2).

$$\text{RMSE} = \sqrt{\frac{1}{P}\Sigma\left(r_{ij} - \tilde{r}_{ij}\right)^2} \qquad (2)$$

where P is the quantity of predictions,

r_{ij} is the original rating,

$\tilde{r}_{ij}$ is the predicted rating.

B. Louvain Community Detection Method

The Louvain community detection method is a distinguished and influential technique, devised in 2008 by Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre [24]. This approach has been highly effective in discerning clusters or communities within complex networks, providing invaluable insights into the organizational structure of intricate systems, and revealing intricate relationships, interactions, and functional modules. At its essence, the Louvain method is grounded in the concept of modularity, a measure of the effectiveness of dividing a network into communities. By iteratively merging or relocating nodes between communities, the method maximizes the overall modularity score, thereby enhancing the accuracy of community detection. This comprehensive approach involves first optimizing modularity within each community and subsequently constructing a new network wherein communities are treated as distinct nodes. The iterations persist until further improvements in modularity are no longer feasible, guaranteeing a thorough exploration of the community structure.

One of the key strengths of the Louvain method lies in its remarkable efficiency and scalability, making it ideal for analyzing large-scale networks and handling intricate and extensive datasets. This efficiency has been instrumental in analyzing real-world complex systems and has contributed to its growing popularity. The Louvain community detection method's ability to accurately identify meaningful communities while maintaining relatively fast computational times has rendered it an indispensable tool in network analysis, graph mining, and diverse fields that necessitate in-depth community exploration and understanding. Its wide-ranging applications span domains such as social network analysis, biological networks, and information retrieval systems, further underscoring its broad significance in a multitude of research and practical contexts.

IV. Proposed Method

To cater to user's personalized recommendations, the incorporation of Matrix Factorization (MF) has emerged as a highly effective technique. Given the massive scale of user-item interactions and the resultant rating matrices, evaluating such extensive datasets can be computationally intensive. In response to this challenge and to optimize the recommendation process, we propose a novel parallel approach that synergizes Singular Value Decomposition (SVD) with the Louvain community detection method.

Our proposed method for personalized recommendations incorporates a parallel approach that synergizes the Singular Value Decomposition (SVD) method with the Louvain community detection technique. The first step involves assembling a comprehensive rating matrix (R) by collating user-item interactions and their corresponding ratings. Subsequently, a bipartite graph is constructed from the dataset, representing users and items as nodes, and their interactions as edges denoted by the ratings. Utilizing the Louvain community detection method, we identify meaningful communities within this bipartite graph, which allows for the efficient partitioning of the rating matrix. After dividing the bipartite graph into distinct communities with appropriate structures, individual rating matrices are obtained for each community. In parallel, the SVD method is applied to each rating matrix, generating predicted rating matrices for every community. These predicted rating matrices are then combined to create a comprehensive predicted rating matrix (R̃). Finally, we assess the accuracy of the recommendations by calculating the Root Mean Squared Error (RMSE), measuring the differences between the original rating matrix (R) and the predicted rating matrix (R̃). This innovative parallel approach not only enhances the accuracy of personalized recommendations but also offers the potential for scalable and efficient recommendation systems, particularly in handling large-scale user-item interactions and extensive datasets.

In this context, SVD serves as a foundational tool for MF, and any suitable MF method can be employed. By leveraging the strengths of MF, our approach fosters the creation of robust community structures, enabling efficient partitioning of the rating matrix.

By combining SVD and community detection in this parallel approach, we can significantly reduce the computational burden and expedite the recommendation process. This innovative method not only improves the accuracy of personalized recommendations but also opens up opportunities for scalable and efficient recommendation systems in the era of big data and extensive user-item interactions.

V. EXPERIMENTAL RESULTS

For implementing our novel and innovative approach, we have meticulously selected two diverse datasets, the restaurant recommendation dataset and the anime recommendation dataset, both obtained from the reputable data repository, Kaggle. These datasets were chosen due to their relevance and suitability for evaluating the efficacy of our proposed method.

TABLE I

DATASET STATISTICS.

Dataset	Users	Items	Ratings	Rating Range	Sparsity
Restaurant Recommendation	268	130	1161	0-2	0.983
Anime Recommendation	1295	17384	62656	1-10	0.997

Table I presents an intricate compilation of comprehensive statistics for these datasets, encompassing vital attributes like size, characteristics, and data point distribution. This detailed information lays a critical groundwork for comprehending the inherent nature of the datasets and the potential implications it may have on our algorithm's performance. By thoroughly analyzing these statistics, we can gain valuable insights into the data's complexity and variations, enabling us to make informed decisions about refining and optimizing our approach for optimal outcomes.

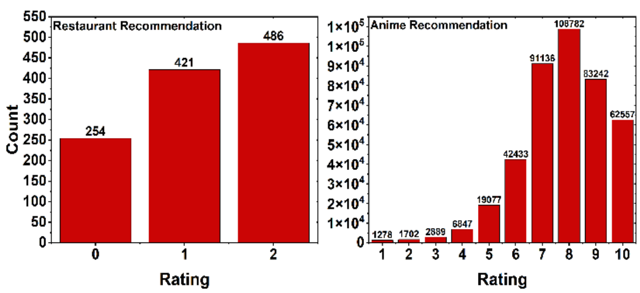

Fig.1. Rating distribution plots for restaurant recommendation and anime recommendation datasets.

Figure 1 presents the rating distribution of two essential datasets, the restaurant recommendation, and the anime recommendation datasets, which play a pivotal role in our research. The figure provides a visual representation of the distribution of ratings across the datasets, shedding light on the frequency and abundance of different rating values. The X-axis in Figure 1 signifies the distinct rating values, while the Y-axis displays the corresponding count of each rating within the datasets. This visual aid enables us to grasp the spread of ratings and identify potential trends or patterns in user preferences.

Analyzing the restaurant recommendation dataset, we discern that the rating "2" dominates the dataset, with a count of 486, suggesting that it is the most prevalent rating among users for restaurant recommendations. Conversely, the rating "0" appears to be less common, with a minimal count of 254, indicating it is the least favored rating. Similarly, in the case of the anime recommendation dataset, the rating "8" emerges as the most prominent, exhibiting a substantial count of 108782. On the other hand, the rating "1" appears relatively infrequently, with a count of 1278, suggesting that it is the least assigned rating by users for anime recommendations.

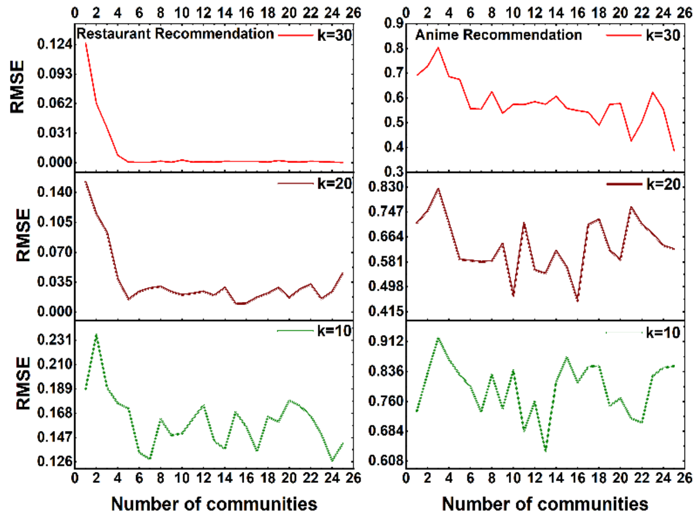

Fig. 2. Comparison of RMSE values for SVD method on Restaurant recommendation and Anime recommendation datasets across different latent features and communities.

Figure 2 presents a compelling visualization of the stacked representation of RMSE values across two significant datasets, the restaurant recommendation dataset and the anime recommendation dataset. The figure highlights the impact of utilizing community detection with varying latent features k = 10, 20, and 30 on 25 distinct communities.

Analyzing the results for the restaurant recommendation dataset, it is observed that when employing the SVD method combined with Louvain community detection, the RMSE value is comparatively high for k = 20 and 30 without community detection at c = 1. However, for k = 10, the RMSE value is notably lower without community detection, whereas it rises when applying the community detection approach. Interestingly, as the number of communities increases, a consistent trend of decreasing RMSE values is observed for each k value. This suggests that the introduction of community detection enhances the accuracy of the recommendation algorithm, particularly when enough communities are incorporated. Similarly, within the anime recommendation of the dataset, the RMSE values are lower when the community detection algorithm is not utilized, but significantly higher when it is applied. As the number of communities increases, the RMSE value exhibits a consistent decline, indicating that the community detection algorithm effectively improves the recommendation accuracy, albeit after reaching a certain threshold of communities.

Overall, the findings suggest that the integration of community detection algorithms contributes to reducing the RMSE values, signifying enhanced recommendation performance, particularly when an optimal number of communities is employed. These insights underscore the significance of community-based approaches in refining recommendation systems and offer valuable guidance for optimizing such algorithms in real-world applications.

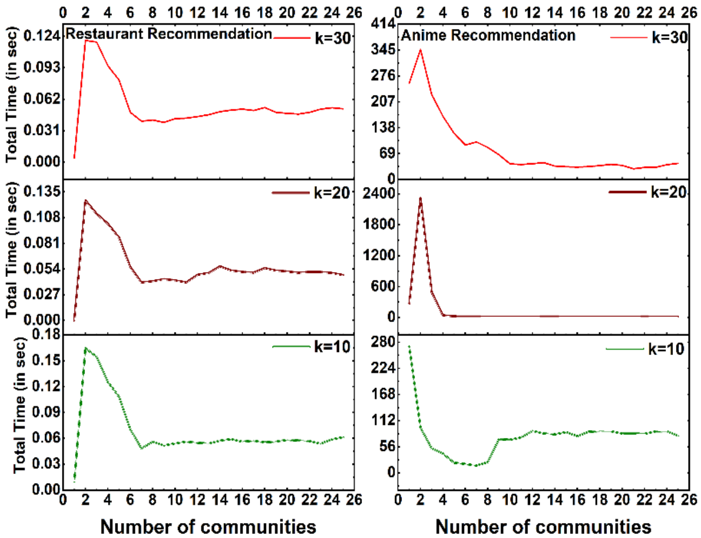

Fig. 3. Time Taken in seconds assessment of SVD Method across different communities and latent features on Restaurant recommendation and Anime recommendation datasets.

Figure 3 presents a comprehensive depiction of the total time required to evaluate the SVD method in conjunction with the community detection method, measured in seconds. This total time encompasses both the time taken for community division and the time taken for computing the RMSE value. Examining the results from the figure, a notable trend emerges across both datasets. At the onset, without utilizing community division (c = 1 community), the evaluation time is relatively lower. However, as the number of communities increases, there is a consistent decline in the time taken for all distinct latent features. This reduction in evaluation time suggests that community division significantly contributes to streamlining the computational process, effectively reducing the overall evaluation time. Moreover, for both datasets, once a certain number of community divisions is reached, the evaluation time stabilizes and remains relatively constant. This observation indicates that beyond a certain point, the benefits of additional community divisions in terms of time reduction may remain constant. In the context of the anime recommendation dataset and for k = 10, it is noteworthy that the evaluation time is considerably higher without utilizing community division. However, as the number of communities increases, there is a continuous decline in evaluation time, ultimately reaching a point of stability.

VI. Conclusion And Future Work

The convergence of SVD and the Louvain community detection method has emerged as a potent and effective strategy, reshaping the landscape of recommendation systems. This harmonious fusion has not only surmounted the hurdles of scalability and accuracy but has also considerably elevated the standard of personalized recommendations. Leveraging the power of SVD, we proficiently extracted latent features and underlying patterns from user-item interactions, resulting in a repertoire of precise and pertinent recommendations. Additionally, the Louvain community detection method has played a pivotal role, enriching our analyses with the revelation of significant user clusters sharing common interests and preferences. As a result, recommendations have been finely honed, ensuring a more satisfying and tailored experience for users. Our research findings undeniably showcase the comprehensive solution that emerges from the combined use of SVD and Louvain community detection in building robust and scalable recommendation systems. The resulting heightened computational efficiency and improved recommendation accuracy constitute invaluable assets, adeptly meeting the demands of modern recommendation engines. These outcomes emphasize the significance of adopting this powerful integration, which holds great promise for reshaping the landscape of recommendation systems and delivering tailored and high-quality user experiences.

In the future, an exciting avenue for exploration lies in extending the applicability of the SVD and Louvain combination to cross-domain recommendation systems. This would involve leveraging information from multiple domains to provide users with more comprehensive and diverse recommendations, catering to their varied interests and preferences. Moreover, we can focus on optimizing the implementation of the combined approach for large-scale recommendation systems. By harnessing distributed computing frameworks and parallel processing techniques, we can efficiently handle massive datasets, ensuring seamless scalability and improved performance in handling the increasing demands of modern recommendation engines.

References

[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[2] Gohberg, I., Kaashoek, M. A., & Spitkovsky, I. M. (2003). An overview of matrix factorization theory and operator applications. *Factorization and Integrable Systems: Summer School in Faro, Portugal, September 2000*, 1-102.

[3] Albano, A. M., Muench, J., Schwartz, C., Mees, A. I., & Rapp, P. E. (1988). Singular-value decomposition and the Grassberger-Procaccia algorithm. *Physical review A*, *38*(6), 3017.

[4] Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics reports*, *659*, 1-44.

[5] Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical review E*, *80*(5), 056117.

[6] Plantié, M., & Crampes, M. (2012). Survey on social community detection. In *Social media retrieval* (pp. 65-85). London: Springer London.

[7] Leung, I. X., Hui, P., Lio, P., & Crowcroft, J. (2009). Towards real-time community detection in large networks. *Physical Review E*, *79*(6), 066107.

[8] Isinkaye, F. O. (2021). Matrix factorization in recommender systems: algorithms, applications, and peculiar challenges. *IETE Journal of Research*, 1-14.

[9] Garcia, J. O., Ashourvan, A., Muldoon, S., Vettel, J. M., & Bassett, D. S. (2018). Applications of community detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proceedings of the IEEE*, *106*(5), 846-867.

[10] Paleti, L., Radha Krishna, P., & Murthy, J. V. R. (2021). Approaching the cold-start problem using community detection based alternating least square factorization in recommendation systems. *Evolutionary Intelligence*, *14*, 835-849.

[11] Liu, Z., & Barahona, M. (2020). Graph-based data clustering via multiscale community detection. *Applied Network Science*, *5*, 1-20.

[12] Lu, H., Halappanavar, M., & Kalyanaraman, A. (2015). Parallel heuristics for

scalable community detection. *Parallel Computing*, *47*, 19-37.

[13] Paleti, L., Radha Krishna, P., & Murthy, J. V. R. (2021). Approaching the cold-start problem using community detection based alternating least square factorization in recommendation systems. *Evolutionary Intelligence*, *14*, 835-849.

[14] Surian, D., Nguyen, D. Q., Kennedy, G., Johnson, M., Coiera, E., & Dunn, A. G. (2016). Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *Journal of medical Internet research*, *18*(8), e232.

[15] De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, *21*(4), 1253-1278.

[16] Paterek, A. (2007, August). Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of KDD cup and workshop* (Vol. 2007, pp. 5-8).

[17] Ali, M., Ahn, C. W., Pant, M., & Siarry, P. (2015). An image watermarking scheme in wavelet domain with optimized compensation of singular value decomposition via artificial bee colony. *Information Sciences*, *301*, 44-60.

[18] Rathore, N. K., Jain, N. K., Shukla, P. K., Rawat, U., & Dubey, R. (2021). Image forgery detection using singular value decomposition with some attacks. *National Academy Science Letters*, *44*, 331-338.

[19] Linhares, C. D., Ponciano, J. R., Pereira, F. S., Rocha, L. E., Paiva, J. G. S., & Travençolo, B. A. (2020). Visual analysis for evaluation of community detection algorithms. *Multimedia Tools and Applications*, *79*, 17645-17667.

[20] Mohammadi, M., Fazlali, M., & Hosseinzadeh, M. (2021). Accelerating Louvain community detection algorithm on graphic processing unit. *The Journal of Supercomputing*, *77*, 6056-6077.

[21] Qie, H., Li, S., Dou, Y., Xu, J., Xiong, Y., & Gao, Z. (2022). Isolate sets partition benefits community detection of parallel Louvain method. *Scientific Reports*, *12*(1), 8248.

[22] Baker, K. (2005). Singular value decomposition tutorial. *The Ohio State University*, *24*, 22.

[23] Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis* (pp. 91-109). Boston, MA: Springer US.

[24] Blondel Vincent, D., Jean-Loup, G., Renaud, L., & Etienne, L. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *10*(2008), P10008.