

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
COMPUTER SCIENCE

doi: 10.17586/2226-1494-2022-22-5-941-950

**An enforced non-negative matrix factorization
based approach towards community detection in dynamic networks**Bashir Shafia¹✉, Ahmad Manzoor Chachoo²^{1,2} Department of Computer Science, University of Kashmir, Srinagar, 190006, India¹ imshafia@gmail.com✉, <https://orcid.org/0000-0002-5570-967X>² manzoor@kashmiruniversity.ac.in, <https://orcid.org/0000-0001-6702-6633>**Abstract**

Identifying community structures within network dynamics is important for analysing the latent structure of the network, understanding the functions of the network, predicting the evolution of the network as well as detecting unusual events of the network. From various perspectives, a diversity of approaches towards dynamic community detection has been advised. However, owing to the difficulty in parameter adjustment, high temporal complexity and detection accuracy is diminishing as time slice rises; and recognizing the community composition in dynamic networks gets extremely complex. The basic models, principles, qualities, and techniques of latent factor models, as well as their various modifications, generalizations and extensions, are summed up systematically in this study which focuses on both theoretical and experimental research into latent factor models across the latest ten years. Latent factor model like non-negative matrix factorization is considered one of the most successful models for community identification which aims to uncover distributed lower dimension representation so as to reveal community node membership. These models are mostly centred on reconstructing the network from node representations while requiring the representation to have special desirable qualities (non-negativity). The purpose of this work is to provide an experimental as well as theoretical comparative analysis of the latent factor approaches employed to detect communities within dynamic networks. Parallely we have devised the generic and improved non-negative matrix factorization-based model which will help in producing robust community detection results in dynamic networks. The results have been calculated from the experiments done in Python. Moreover our models methodology focuses on information dynamics so as to quantify the information propagation among the involved nodes unlike existing methods that considers networks first-order topological information described by its adjacency matrix without considering the information propagation between the nodes. In addition, this paper intends to create a unified, state of the art framework meant for non-negative matrix factorization conception which could be useful for future study.

Keywords

community detection, principal component analysis, orthogonality, non-negative matrix factorization, singular value decomposition, social network analysis

For citation: Bashir S., Chachoo M.A. An enforced non-negative matrix factorization based approach towards community detection in dynamic networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 5, pp. 941–950. doi: 10.17586/2226-1494-2022-22-5-941-950

УДК 51-78

**Подход к обнаружению сообщества в динамических сетях,
основанный на принудительной неотрицательной матричной факторизации**Башир Шафия¹✉, Ахмад Мансур Чачу²^{1,2} Университет Кашмира, Сринагар, 190006, Индия¹ imshafia@gmail.com✉, <https://orcid.org/0000-0002-5570-967X>² manzoor@kashmiruniversity.ac.in, <https://orcid.org/0000-0001-6702-6633>**Аннотация**

Выявление структур сообщества в сетевой динамике важно для анализа сети относительно: скрытой структуры, понимания функций, прогнозирования развития, обнаружения необычных событий. В рассмотренных научных

© Bashir S., Chachoo M.A., 2022

исследованиях рекомендуется использовать различные подходы к динамическому обнаружению сообщества. Однако из-за сложности настройки параметров, высокой временной сложности и снижения точности обнаружения по мере увеличения временного интервала распознавание состава сообщества в динамических сетях усложняется. Рассмотрены основные схемы, принципы, свойства и методы моделей латентных факторов, а также их системные модификации, обобщения и расширения. Основное внимание уделено теоретическим и экспериментальным исследованиям моделей латентных факторов за последние десять лет. Скрытая факторная модель — неотрицательная матричная факторизация, считается одной из наиболее успешных для идентификации сообщества и направлена на раскрытие распределенного представления более низкого измерения с целью определения членства в узле сообщества. Модели основаны на реконструкции сети из представлений узлов при условии, чтобы представление обладало особыми желательными качествами (например, не отрицательностью). Цель работы — получить экспериментальный и теоретический сравнительные анализы подходов со скрытым фактором, используемых для обнаружения сообществ в динамических сетях. Разработана общая и улучшенная неотрицательные матричные модели, основанные на факторизации для получения надежных результатов обнаружения сообщества в динамических сетях. Полученные результаты рассчитаны на основе экспериментов, проведенных на языке программирования Python. Предложенная методология моделей сфокусирована на динамике информации, для количественной оценки распространения информации между задействованными узлами. Отличие предложенной модели от существующих состоит в получении топологической информации сети первого порядка, описываемой ее матрицей смежности, без учета распространения информации между узлами. Предложено создание единой современной структуры, предназначенной для концепции неотрицательной матричной факторизации, которая может быть полезна для будущих исследований.

Ключевые слова

обнаружение сообщества, анализ главных компонент, ортогональность, неотрицательная матричная факторизация, разложение по сингулярным числам, анализ социальных сетей

Ссылка для цитирования: Башир Ш., Чачу М.А. Подход к обнаружению сообщества в динамических сетях, основанный на принудительной неотрицательной матричной факторизации // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 5. С. 941–950 (на англ. яз.). doi: 10.17586/2226-1494-2022-22-5-941-950

Introduction

Community structure regarded as an essential characteristic of dynamic networks reveals as well as exposes the underlying connections. There were more and more studies recently in order to develop methods for community detection. These studies, on the other hand, were mostly focused on static networks; hence they were unable to detect dynamic communities in complex networks. The construction of the network in real-world is however dynamic. The structure of community for most networks, on the other hand, is constantly changing throughout time. Existing static approaches for community detection split the community depend on the network static topology and ignore the interaction between network structures over numerous snapshots. Owing to their high-level potentiality for comprehending social phenomena across time, dynamic community identification in complex networks has recently gotten lots of attention and turn out to be a popular study area. Social networks, protein-protein connection networks and person to person communication networks are all examples of temporal networks in the real world. Within network dynamics, temporal network analysis can uncover potential rules and essential properties. As a result, developing ways to identify communities in dynamic networks has become progressively more crucial. The uncertainty of the solutions is one of the most significant challenges in recognizing temporal communities. As a result, we can't tell if the change in community discovery is due to the community's evolution or the algorithm instability. A variety of strategies have been offered to address this issue, with the ultimate goal of smoothing the community evolution.

Our goal to develop a more robust Dynamic Community Detection (DCD) methodology was achieved by focusing on information dynamics so as to quantify the information propagation among the involved nodes. This study also focussed on the quantitative (statistical) analysis of the latent factor model based approaches (Non-Negative Matrix Factorization (NMF)) related to community detection in dynamic networks so that the user may obtain a general idea of how the network is organized and a few underlying experiences of the network structure. Moreover, this study is organized in the way where introductory section addresses the fundamentals required for reviewing the DCD followed by literary section which addresses the findings and recent research in SNA (social network analysis) related to detecting dynamic communities. Next section represents the generic and improved NMF based methodology to uncover the dynamic communities. Empirical results have been shown under the result section. Evaluation section evaluates the framework against the two important existing NMF based approaches: Community Detection with Community Structure and Node Attribute (CDCN) and DPNM (NMF incorporated density peak clustering). Next section (Applications) provides the utilization of detecting communities. Lastly, we summarize and conclude the study with some potential future guidelines.

Literature Review

The recognition of a networks community division reflects nodes tendency to form clusters based on their resemblance and hence create communities. Various real-world networks show that community structure exists [1]. Community analysis is necessary to understand the

network structural and functional features. The structural feature have been applied to a variety of fields, like viral marketing, epidemic modelling, and detecting important vertices in power systems wherein their failure could create a cascading collapse.

In order to cite only and not to overemphasise the significance of community identification and it's working in the perspective of this study, community identification enables the recognition of few node groupings bearing the intra-group connections denser, implying that it works by recognizing those groups of people whose interaction among them takes place frequently. The detection of communities will aid in identifying the influential nodes as well as their subordinates in real-time. This is one of the rationales wherefore community identification techniques are being developed, tested, and enhanced.

A large number of algorithms in the literature focused on vertices characteristics rather than the link characteristics. For the challenge of dynamic community discovery in evolving networks, there aren't many favourable approaches. However, latent factor models, like NMF models, are one of the most favourable approaches. By rebuilding the original data, matrix factorization is an excellent tool in favour of representation learning. The constraints forced on the base matrix and the coding matrix differs among a variety of matrix factorization techniques. The usage of non-negativity restrictions (constraints) distinguishes NMF from other types of matrix factorization. Additional restrictions can be imposed on NMF such as orthogonality or sparseness that preserves the topological features [2]. Despite these additional constraints, NMF differs from the other techniques of matrix factorization like PCA (principal component analysis) and SVD (singular value decomposition) because to its non-negativity feature. Because they only allow additive combinations but not subtractive combinations, these constraints result in a parts-based representation. NMF approaches use every node as a network dimension, and consider community detection to be the challenge of identifying a low-dimensional network representation. However, they do not guarantee that the representation obtained matches with communities, necessitating some heuristic control to increase interpretability.

NMF, which is popularly utilised in pattern recognition and information retrieval, have recently been employed to resolve the challenge of community discovery [3, 4]. In terms of DCD several research, workers [5–7] had included NMF within a temporal framework, resulting in a plethora of useful DCD models. Wang et al. [4] proposed a model, namely DCD based on NMF, where the transition matrix of community membership was included with the notion of temporal cost so as to smooth out the community structure alterations. Whereas Yang et al. [7] developed a DCD approach based on NMF that takes node strength into account. Their primary premise was that the node pairs bearing stronger connections have a greater chance of belonging to the same community. FacetNet [8] is a well-known method for community analysis together with their evolution considering network dynamics. It combines communities altogether with their evolutions in a unified manner using NMF which differs from standard methods

wherein two-stage processes are treated asynchronously. The FacetNet technique, on the other hand, necessitates prior information and the specification of the number of the community partitions. In many cases, the number of community partitions is difficult to predict in advance. To get over this matter, Lu et al. [9] employed singular value decomposition so as to automatically uncover the number of community partitions. Though the k is representing the number of communities may be automatically collected and a higher grade of community detection can be achieved, the time complexity is considerable. Despite the fact that studies [10, 11] suggest that NMF is the best method for learning object components, it fails to capture the geometrical data structure space that is crucial for clustering. Graph regularisation approach may extract the geometric data structure, according to manifold learning theory and spectral graph theory [12]. Furthermore, the geometric network structure will not get changed significantly in a short period of time if temporal smoothness is assumed. As a result, we suggest that graph regularisation is to be used to simulate the temporal cost function which will be intuitively beneficial. Yu et al. [13] followed a two-stage approach where community detection and community evolution were studied separately, hence was not optimal. Though the authors [14] suggested the efficient approach for community detection yet their approach was very specific to criminal activities.

Ma & Dong [15] presented 2 frameworks for Evolutionary Non-negative Matrix Factorization (ENMF) and demonstrated that evolutionary spectral clustering, ENMF and evolutionary modularity density are all comparable. They also introduced sE-NMF, the semi-supervised technique that adds a priori knowledge to ENMF. Sun et al. [16] developed a model-DCD that was comparable to Wang's [5]. The distinction between them, though, is the cost of a snapshot. As a cost function of the snapshot, the earlier adopted standard NMF, while latter used the symmetric NMF (SNMF). In order to achieve better interpretability, Yuan & Liu [17] suggested a model-DCD with node weight matrix and the triple NMF.

Internal connections were used as the graph regularisation requirement by Wang et al. [18] who used the triple NMF with the purpose of increasing the functioning of community identification into networks of bipartite. Matrix factorization accurately depicts the network community structure and guarantees the significant interpretation of the community regardless of the topology of the network. NMF avoids the limitations of modularity optimization approaches [19], like resolution of limit [20], in addition to quantifying how robustly every node contributes to its community. Tokala et al. [21] employed NMF along with the cost function namely I-divergence to present 2 techniques for undirected and directed networks, respectively. Jin et al. [22] employed NMF to construct the model which is generative, treating this as the problem of optimization to identify the linkage formation of communities, grounded upon the relevance of every node while establishing linkage to every community. Even though the study [23] presented a comparative analysis of the already existing approaches and algorithms employed to detect online communities in social networks, yet the

community extraction and community evolution were treated asynchronously in most of the approaches which doesn't make sense in the real environment.

The studies [24, 25] also approached to model community detection with a significant improvement from the past studies. Wang et al. [24], in contrast to the above research studies, employed graph regularisation to capture classical geometric network structure information. They created an efficient iterative algorithm grounded on multiplicative updating principles as well as the proof method. Despite the fact that their method exceeds the competition still in respect of overall functioning, the enhancement on several networks is insignificant. As a result of this vast research review, it is clear that additional research is required and expected so as to advance this field of study.

Methodology

Past works followed a two-stage approach where community detection and community evolution were studied separately. However, our framework follows a more advanced approach by studying the community detection and evolution simultaneously as described below.

By creating a distinctive network where all snapshot nodes are all instances of all nodes, and edges can be regular edges within snapshots or a special sort of edge that connects nodes between snapshots. Then the NMF is made to run on this large network to uncover communities as illustrated in Fig. 1.

Our contribution towards this study mainly consists of the following steps:

- an effective approach to calculate information dynamics among the nodes in a targeted network;
- an integrated framework is proposed where number of communities adopted by NMF is calculated automatically based on the information flow score among nodes;
- a model that merely depends on the targeted network topology to identify true communities is aimed and formalized.

Our framework that uses social networks to recognize active communities has four main interconnected phases (stages) as described below.

Stage 1. We use BFS (Breadth First Search) first to sample a graph around an initial seed node; then the adjacency matrix A is constructed based on the targeted subnetwork having ones depicting links among network nodes and zeroes depicting unknowns like:

$$A_{ij} = \{1 \text{ if } i \text{ and } j \text{ are adjacent to each other, otherwise } 0.\} \quad (1)$$

Stage 2. Here the similarity matrix B based on contact strength among nodes of matrix A is computed first, where contact strength represents the degree of closeness among nodes of a network. Because the triangle structure may better characterise the tightness among the nodes, we employed triangles to formalise the contact strength definition. It is computed using the following formula:

$$CS_{uv} = \frac{N(u) \cap N(v)}{Tu}, \quad (2)$$

where Tu is the number of triangles of vertex u , and the intersection between $N(u)$ and $N(v)$ is the number of triangles common to node u and node v .

So the strategy is based on the fact that strong ties play a major part in community formation and information diffusion. In the network having n nodes, the similarity of every node pair according to the above described contact strength formula is computed first and therefore $n \times n$ form of similarity matrix as $B = \{s_{ij}\}$ is obtained, wherein an element s_{ij} indicate the contact strength closeness between i and j nodes.

Stage 3. Here we made an attempt to train our framework to calculate the number of communities based on the node information flow. The information about a v node over t time is calculated as:

$$I_{v(t+1)} = I_{v(t)} + \sum_{u \in N(v)} (CS_{uv}). \quad (3)$$

Wherein $I_{v(t)}$ describes the information of a v node at t time, and second part of the expression represents the information which is gained from its neighbours. As may be seen, information of v node at time $t + 1$ involves the information over time t plus the information gained from its neighbours at time $t + 1$. With the time evolution, the propagation of information inclines to zero. Ultimately, the network information will achieve a state of equilibrium which blocks the further information interaction.

Stage 4. Then NMF is exploited using the similarity matrix B together with the calculated information flow thus uncovering the communities. NMF uncovers the inherent network community composition and improves interpretability and compression because of its "sparse" and "parts based" representation with the solely additive constraint or non-negativity. NMF reduces the matrix $B \in R^{m \times n}$ into two nonnegative matrices like $V \in R^{m \times k}$ and $U \in R^{k \times n}$, such that $B \cong VU$. Given an information node B matrix, we intend to acquire U the matrix of node membership using NMF as follows:

$$\begin{aligned} \min_{U, V} \|B - VU\|_F^2 \\ \text{s.t. } U \geq 0. \end{aligned} \quad (4)$$

After obtaining the clusters around the seed nodes, the quality scores of each node are computed and the node having highest score in every cluster is selected as the high quality seed. We illustrate the quality-score of a node $v(QS_V)$ as:

$$\begin{aligned} QS_V = Sim(YG(v), YG(vs)) + \\ + \frac{\sum_{u \in N(v)} Sim(YG(u), YG(v))}{|N(v)|}, \end{aligned} \quad (5)$$

where $N(v)$ is the neighbour set of v ; $YG(v)$ is the vector embedding of v node on G ; and Sim represents the cosine similarity.

Moreover we summarize our frameworks work flow in a flow chart (Fig. 1). Four main tasks employed by our methodology discussed above as well are further elaborated as follows:

1. Initialization — where adjacency matrix A is constructed based on targeted network having ones depicting links among network nodes and zeroes depicting unknowns. Then initialization of secondary data structures takes place accordingly.
2. Pre-processing — where matrix B is constructed based on the information statistics of nodes considering matrix A as a base matrix.
3. Training — where our framework is made to learn the no. of communities based on the node information flow.
4. Detecting — where finally the associations between the nodes and the communities are attained via NMF over learnt latent factor representations of the nodes of the network.

The process is repeated until all the snapshots are visited. Once it is done with all the snapshots a sequence of multiple communities are obtained.

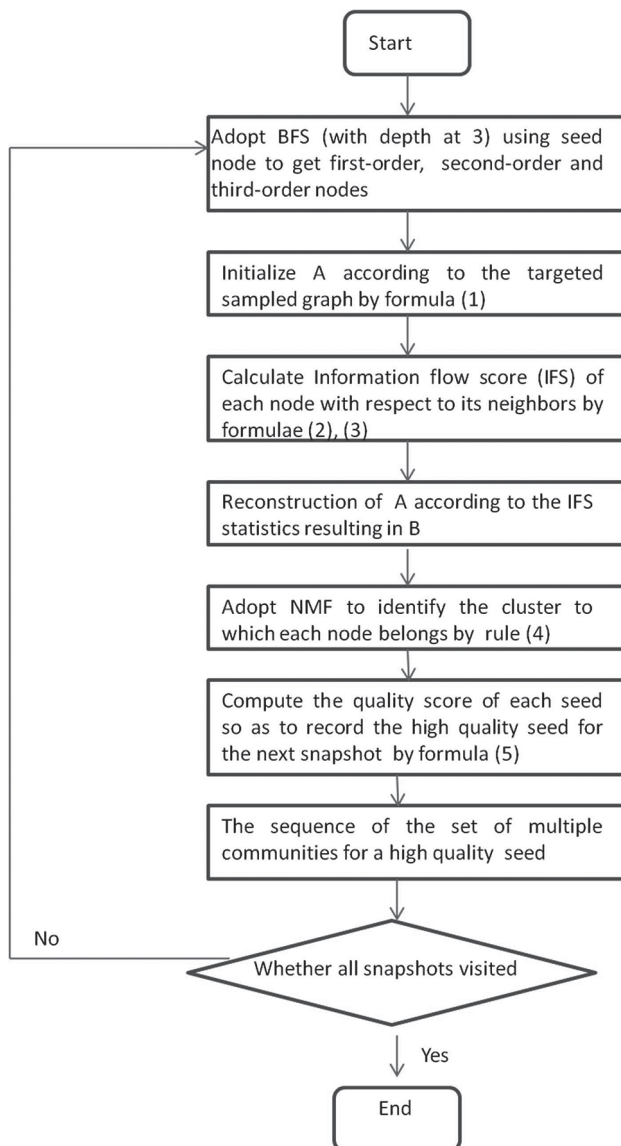


Fig. 1. Work flow of our proposed framework

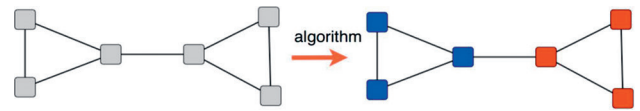


Fig. 2. A simple illustration of the community detection in our framework

Our framework is capable of uncovering the connected components in a graph without explicitly specifying the number and size of communities as was required in previous studies [8, 15]. Moreover the approach is scalable enough to accommodate very large networks as often observed in real data. Fig. 2. shows how communities are identified in our framework — blue colored nodes represent community 1 and red colored nodes represent community 2.

Empirical Results

Using three real-world networks as discussed below, the effectiveness of the existing community detection methods based NMF was examined in this study.

Simulated Data

We used the combination of small and large 4 real-world networks as discussed and demonstrated below in order to validate the effectiveness of our model.

In 1970s, 34 fellows from a club-karate at an American university formed a friendship social network known as the Zachary karate club [26]. The network was broken into two sets of friendship due to a disagreement involving the club manager along with the trainer about the charge of training karate.

In the year 2000, American Football College became a games network among Division-I Colleges (teams). Conferences are made up of colleges, with each conference serving as a ground-truth community. The nodes (115) represent Colleges (teams) whereas edges (616) denote games played among teams [1].

Dolphin is a group of 62 dolphins that were spotted in New Zealand Doubtful Sound between 1994 and 2001. Two dolphins that were observed together “more often than the predicted probability” are connected by an edge. The network was separated into two primary communities (partition (a) contains 1-20 nodes and partition (b) contains 21–62 nodes), both of which can be subdivided into three sub communities [27].

Facebook is a network of friendships among Facebook users, wherein the vertex denotes a user and the edge depicts that the users represented with the end points are friends. This is a friendship network (available in SNAP library also) with 347 vertices and 5038 edges [28].

Visualizing Networks Result by our Model

In this study an attempt has been made to produce robust community detection results in dynamic networks. At first our framework was applied on the two fundamental and important networks (Fig. 3) and fortunately we obtained satisfactory results compared with the baseline approaches (Fig. 8).

As shown in Fig. 4 the Dolphin network can be treated as a network having 4 communities, Wherein the main two communities are node 1 — node 20 and node 21 —

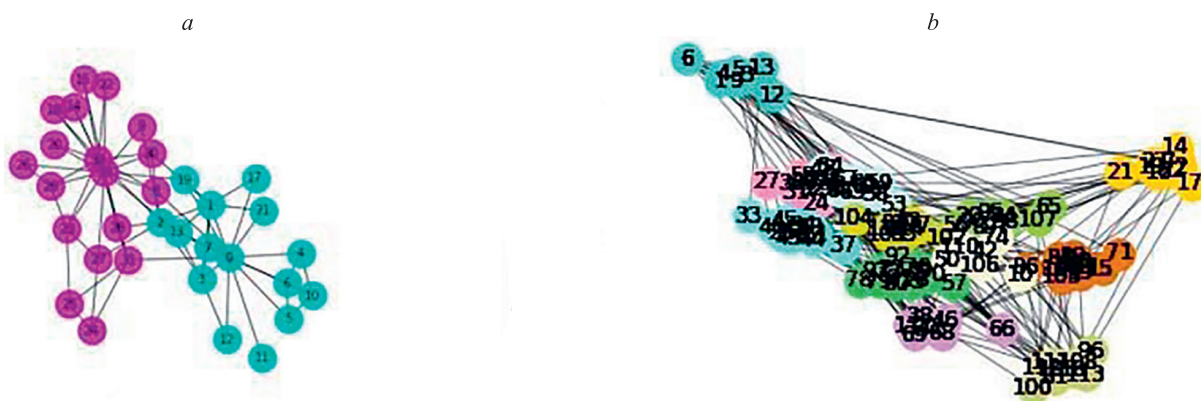


Fig. 3. Community structures of the Karate network (a) and American football network (b) as detected by our model where each color represents distinct community

node 62. Furthermore the community comprising of nodes from node 21 — node 62 are further split into 3 sub-communities.

Next our model was applied on Facebook network. For the sake of expositional clarity, we skip the labels of the nodes so that the different community structures can be seen clearly as shown in Fig. 5.

To visualize the relationship strength and the corresponding relationship score in a better way, we have plotted the Heatmap wherein nodes are represented by columns and rows showing which pair of nodes are most closely related as shown in Fig. 6. Each square shows the relationship between the nodes on each axis where maroon color indicates the positive relationship and blue color indicates the negative relationship. The varying intensity of color represents the measure of relationship. The stronger the color is the larger is the relationship magnitude.

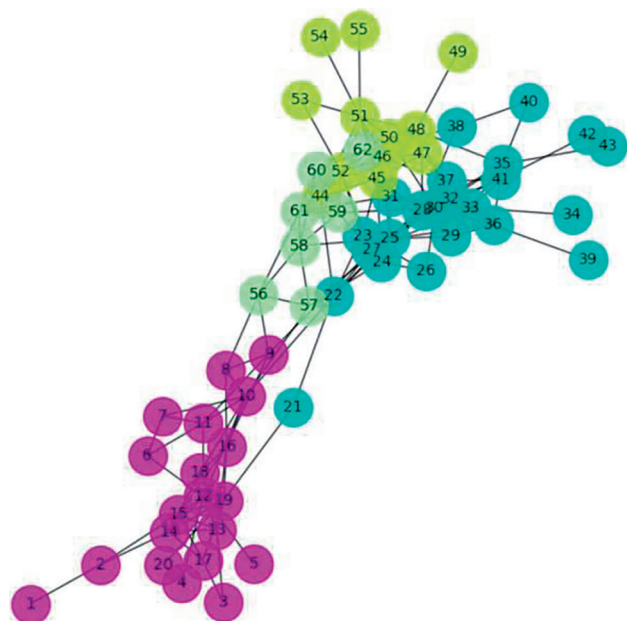


Fig. 4. Community structure of the Dolphin network clearly showing further splitting of 2nd community represented by green, light green and blue color

Finally a brief internal working has been shown in Fig. 7 for better understanding of the model. A sampled graph obtained after performing BFS sampling on a simple graph having 15 nodes by taking seed node as 6, since BFS is fast in exploring the neighbourhood of a node so was chosen.

Evaluation Metrics and Performance Comparison

To validate the performance of our model we have employed NMI (normalized mutual information) and ARI (adjusted random index), conductance, F1 score to measure the similarity between the ground truth and detected community structures, and the results are shown in Fig. 8. Both the NMI and the ARI are commonly used similarity assessment standard metrics which are based on information theory and have been proven reliable. Given two community network divisions **A** and **B**, NMI (**A**; **B**) is computed as:

$$NMI(A; B) = \frac{2I(A; B)}{H(A) + H(B)} \tag{6}$$

Wherein $I(A; B)$ represents mutual information of **A** with **B** and $H(B)$ denotes entropy from **B**. The NMI value ranges between 0 through 1, wherein 0 means that the communities detected are totally independent of the ground



Fig. 5. Community structure of Facebook network, each color represents a separate community

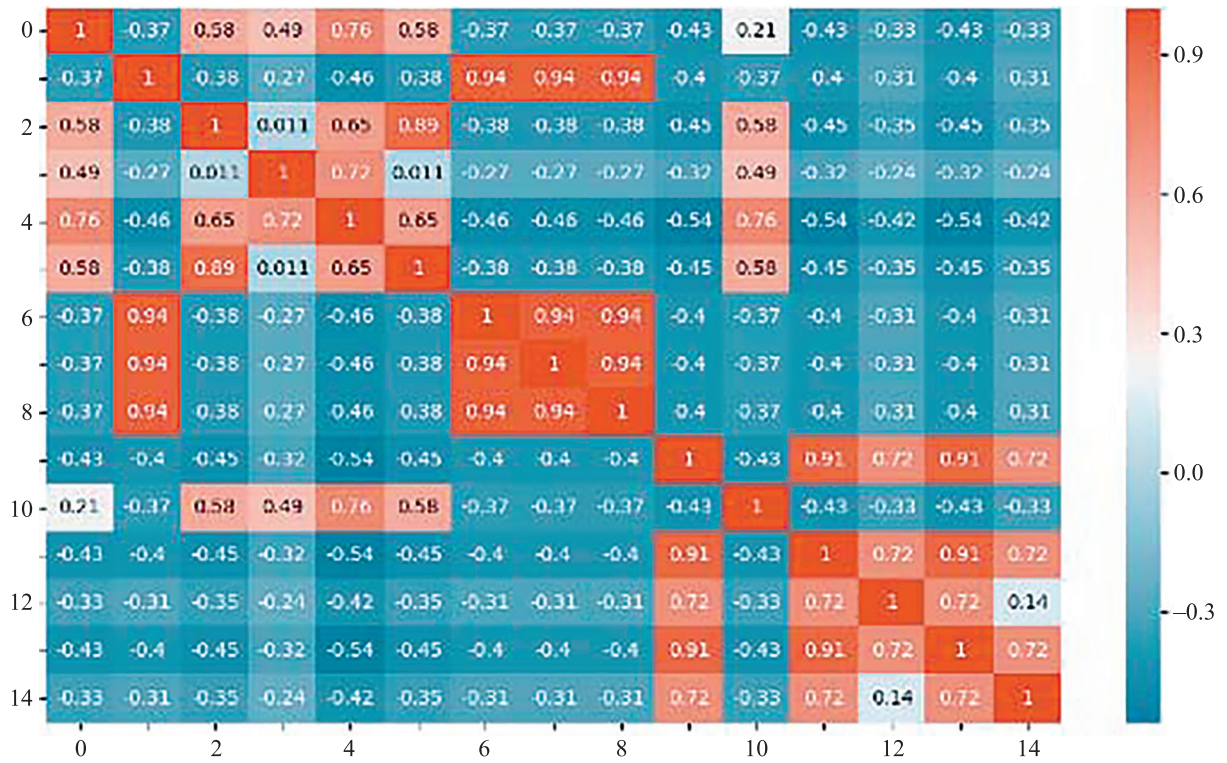


Fig. 6. Correlation Heatmap for the sample dataset (used in Fig. 7)

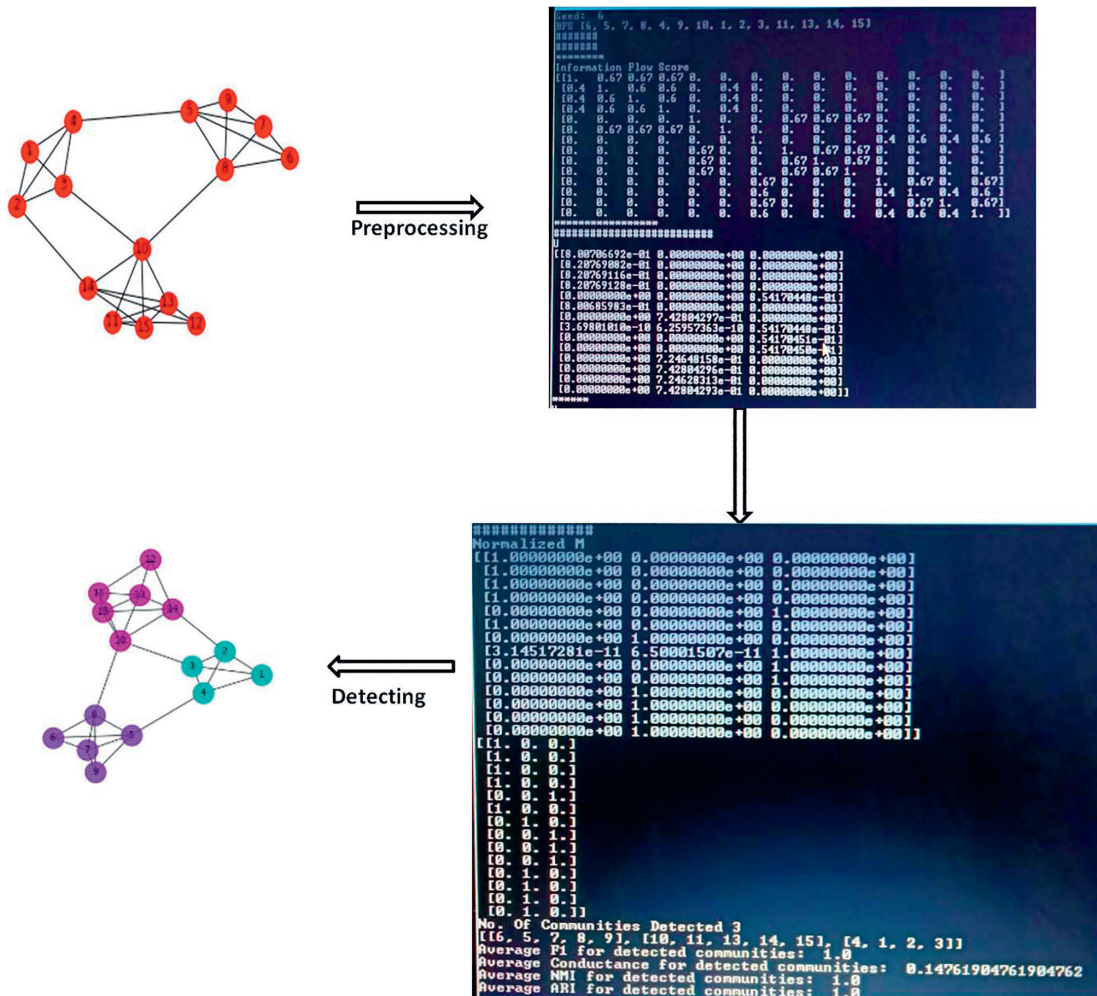


Fig. 7. Outline of our research framework

truth communities, while 1 means a complete match with the actual communities.

$$ARI = \frac{RI - Expected\ RI}{\max RI - Expected\ RI} \quad (7)$$

Wherein RI represents the similarity between two networks, partitions comprise all sample pairs. Subsequently, it computes the number of pairs in the expected and actual partitions of the network which are allotted to the different or the same network partitions.

Conductance measures the fraction of total edge volume that points outside the partition. The lower the conductance value the better is the partitioning. Conductance is computed as:

$$f(c) = sc / (2mc + sc) \quad (8)$$

Wherein $f(c)$ is the quality-measure function which estimates the quality of a given community c ; mc is the number of edges inside the community c ; and sc is the number of edges leaving the community c , i.e. edges that connect the members of c to the other communities within that network.

F1 Score is the harmonic mean of precision and recall whose value ranges from 0 to 1 and it is calculated as:

$$F1\ Score = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

Fig. 8 reports the results of our framework on various networks using the above mentioned metrics, computed from the formulas (6)–(9).

In order to further compare the performance of our model with other baseline approaches, such as CDCN NMF proposed by Ye et. Al [6] and DPNMF proposed by Lu et al. [25], we used the NMI metric computed from equation (6) and the results are shown below in Fig. 9.

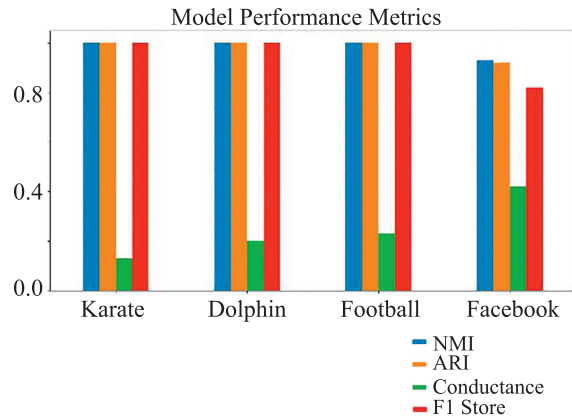


Fig. 8. Comparison of the different metric values obtained for the results of our model for the 4 public datasets

Applications

NMF has become an essential technique in multivariate data study because to the better semantic interpretability and resulting sparsity in accordance with the non-negativity. It has a long history of application to the domains of optimization, maths, neural computing, machine learning, pattern recognition, data mining, computer vision and image engineering, spectral analysis of data, chemo metrics, bioinformatics, criminology, geophysics, economics and finance. More peculiarly, such applications cover digital watermark, data mining of text, denoising of image, restoration of image, segmentation of image, image fusion, classification of image, image retrieval, hallucination of face, recognition of face, recognition of facial expression, audio pattern separation, speech recognition, music genre classification, microarray analysis, spectroscopy, blind source separation, classification of gene expression, cell analysis, processing of EEG signal, pathologic diagnosis, online discussion and prediction, email surveillance, network security, stock market pricing, earthquake prediction, and all that.

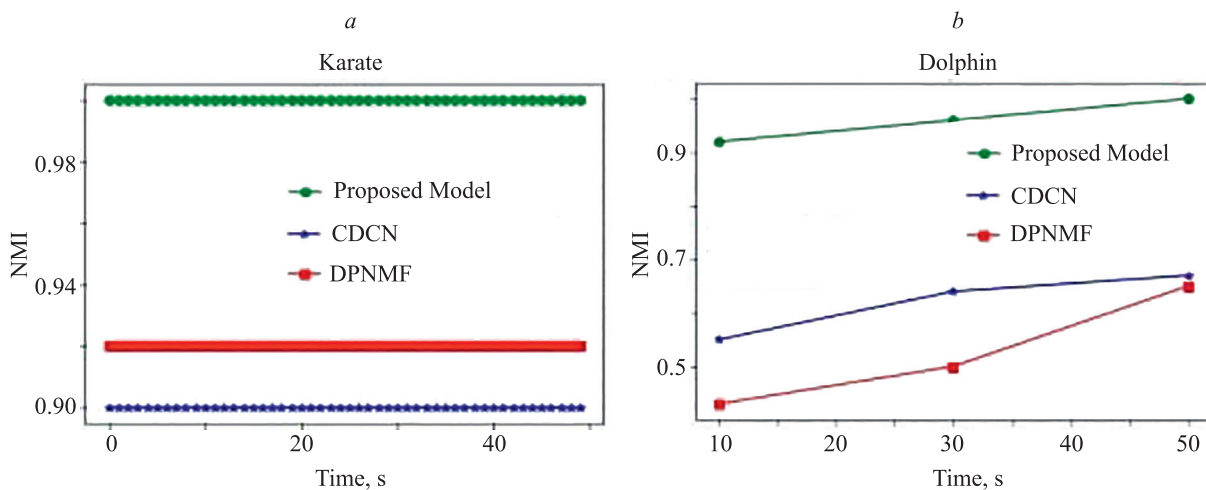


Fig. 9. Performance comparison in term of NMI on Karate (a) and Dolphin networks (b)

Conclusion

This work presented an improved non-negative matrix factorization model as well as comprehensive review of existing NMF methods employed in support of dynamic community detection. Moreover, this work highlighted the strengths and limitations of the existing methods in order to determine the extent to which work in this field is being done and to identify the significant research gap that exists as well as to investigate if significant improvement can be achieved on these existing NMF-based models. Fortunately, our model produced satisfactory results compared with the baseline approaches such as CDCN and DPNMF. Simultaneously there are some issues discussed below

References

1. Girvan M., Newman M.E.J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, vol. 99, no. 12, pp. 7821–7826. <https://doi.org/10.1073/pnas.122653799>
2. Wang Y., Zhang Y. Nonnegative Matrix Factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 2013, vol. 25, no. 6, pp. 1336–1353. <https://doi.org/10.1109/TKDE.2012.51>
3. Yang J., Leskovec J. Overlapping community detection at scale: A nonnegative matrix factorization approach. *Proc. of the 6th ACM International Conference on Web Search and Data Mining (WSDM)*, 2013, pp. 587–596. <https://doi.org/10.1145/2433396.2433471>
4. Wang F., Li T., Wang X., Zhu S., Ding C. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 2011, vol. 22, no. 3, pp. 493–521. <https://doi.org/10.1007/s10618-010-0181-y>
5. Gao F., Yuan L., Wang W. Dynamic community detection using nonnegative matrix factorization. *Proc. of the International Conference on Computing Intelligence and Information System (CIIS)*, 2017, pp. 39–45. <https://doi.org/10.1109/CIIS.2017.56>
6. Ye Z., Zhang H., Feng L., Shan Z. CDCN: A new NMF-based community detection method with community structures and node attributes. *Wireless Communications and Mobile Computing*, 2021, pp. 5517204. <https://doi.org/10.1155/2021/5517204>
7. Yang K., Guo Q., Liu J.Q. Community detection via measuring the strength between nodes for dynamic networks. *Physica A: Statistical Mechanics and its Applications*, 2018, vol. 509, pp. 256–264. <https://doi.org/10.1016/j.physa.2018.06.038>
8. Lin Y.R., Chi Y., Zhu S., Sundaram H., Tseng B.L. Facetnet: A framework for analyzing communities and their evolutions in dynamic networks. *Proc. of the 17th International Conference on World Wide Web*, 2008, pp. 685–694. <https://doi.org/10.1145/1367497.1367590>
9. Lu H., Sang X., Zhao Q., Lu J. Community detection algorithm based on nonnegative matrix factorization and pairwise constraints. *Physica A: Statistical Mechanics and its Applications*, 2020, vol. 545, pp. 123491. <https://doi.org/10.1016/j.physa.2019.123491>
10. Lee D., Seung H.S. Learning the parts of objects with non-negative matrix factorization. *Nature*, 1999, vol. 401, no. 6755, pp. 788–791. <https://doi.org/10.1038/44565>
11. Cai D., He X., Han J., Huang T.S. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, vol. 33, no. 8, pp. 1548–1560. <https://doi.org/10.1109/TPAMI.2010.231>
12. Chung F.R.K. *Spectral Graph Theory*. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997, 108 p.
13. Yu W., Wu H., Jiao P., Wu H., Sun Y., Tang M. Modeling the local and global evolution pattern of community structures for dynamic networks analysis. *IEEE Access*, 2019, vol. 7, pp. 71350–71360. <https://doi.org/10.1109/ACCESS.2019.2920237>
14. Shafia, Chachoo M.A. Social network analysis based criminal community identification model with community structures and node attributes. *Proc. of the 4th International Conference on Smart Systems*

that need to be considered in the future and we hope to make our approach more effective by resolving these issues and investigating other forms of matrix norm in order to develop the robust objective function.

Although NMF is competitive with well-known community detection approaches, on a variety of well-known data sets networks, however in systems such as telephone or email networks, asymmetric communication rates pose problems because negative links in the graph are not allowed in NMF. Furthermore, because the NMF is totally a marked adjacency matrix adopter, it is irrelevant in numerous real-world applications where there are issues with data collection. Thus, the model to be used at any given time is determined by a variety of factors.

Литература

1. Girvan M., Newman M.E.J. Community structure in social and biological networks // *Proceedings of the National Academy of Sciences of the United States of America*. 2002. V. 99. N 12. P. 7821–7826. <https://doi.org/10.1073/pnas.122653799>
2. Wang Y., Zhang Y. Nonnegative Matrix Factorization: A comprehensive review // *IEEE Transactions on Knowledge and Data Engineering*. 2013. V. 25. N 6. P. 1336–1353. <https://doi.org/10.1109/TKDE.2012.51>
3. Yang J., Leskovec J. Overlapping community detection at scale: A nonnegative matrix factorization approach // *Proc. of the 6th ACM International Conference on Web Search and Data Mining, WSDM*. 2013. P. 587–596. <https://doi.org/10.1145/2433396.2433471>
4. Wang F., Li T., Wang X., Zhu S., Ding C. Community discovery using nonnegative matrix factorization // *Data Mining and Knowledge Discovery*. 2011. V. 22. N 3. P. 493–521. <https://doi.org/10.1007/s10618-010-0181-y>
5. Gao F., Yuan L., Wang W. Dynamic community detection using nonnegative matrix factorization // *Proc. of the International Conference on Computing Intelligence and Information System (CIIS)*. 2017. P. 39–45. <https://doi.org/10.1109/CIIS.2017.56>
6. Ye Z., Zhang H., Feng L., Shan Z. CDCN: A new NMF-based community detection method with community structures and node attributes // *Wireless Communications and Mobile Computing*. 2021. P. 5517204. <https://doi.org/10.1155/2021/5517204>
7. Yang K., Guo Q., Liu J.Q. Community detection via measuring the strength between nodes for dynamic networks // *Physica A: Statistical Mechanics and its Applications*. 2018. V. 509. P. 256–264. <https://doi.org/10.1016/j.physa.2018.06.038>
8. Lin Y.R., Chi Y., Zhu S., Sundaram H., Tseng B.L. Facetnet: A framework for analyzing communities and their evolutions in dynamic networks // *Proc. of the 17th International Conference on World Wide Web*. 2008. P. 685–694. <https://doi.org/10.1145/1367497.1367590>
9. Lu H., Sang X., Zhao Q., Lu J. Community detection algorithm based on nonnegative matrix factorization and pairwise constraints // *Physica A: Statistical Mechanics and its Applications*. 2020. V. 545. P. 123491. <https://doi.org/10.1016/j.physa.2019.123491>
10. Lee D., Seung H.S. Learning the parts of objects with non-negative matrix factorization // *Nature*. 1999. V. 401. N 6755. P. 788–791. <https://doi.org/10.1038/44565>
11. Cai D., He X., Han J., Huang T.S. Graph regularized nonnegative matrix factorization for data representation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011. V. 33. N 8. P. 1548–1560. <https://doi.org/10.1109/TPAMI.2010.231>
12. Chung F.R.K. *Spectral Graph Theory*. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997. 108 p.
13. Yu W., Wu H., Jiao P., Wu H., Sun Y., Tang M. Modeling the local and global evolution pattern of community structures for dynamic networks analysis // *IEEE Access*. 2019. V. 7. P. 71350–71360. <https://doi.org/10.1109/ACCESS.2019.2920237>
14. Shafia, Chachoo M.A. Social network analysis based criminal community identification model with community structures and node attributes // *Proc. of the 4th International Conference on Smart*

- and Inventive Technology (ICSSIT), 2022, pp. 334–339. <https://doi.org/10.1109/ICSSIT53264.2022.9716286>
15. Ma X., Dong D. Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks. *IEEE Transactions on Knowledge & Data Engineering*, 2017, vol. 29, no. 5, pp. 1045–1058. <https://doi.org/10.1109/TKDE.2017.2657752>
 16. Jiao P., Yu W., Wang W., Li X., Sun Y. Exploring temporal community structure and constant evolutionary pattern hiding in dynamic networks. *Neurocomputing*, 2018, vol. 314, pp. 224–233. <https://doi.org/10.1016/j.neucom.2018.03.065>
 17. Liu H.-F., Yuan L.-M.-Z. Community detection in temporal networks using triple nonnegative matrix factorization. *DEStech Transactions on Computer Science and Engineering*, 2017. <https://doi.org/10.12783/dtce/mmsta2017/19682>
 18. Wang T., Liu Y., Xi Y.-Y. Identifying community in bipartite networks using graph regularized-based non-negative matrix factorization. *Journal of Electronics and Information Technology*, 2015, vol. 37, no. 9, pp. 2238–2245. (in Chinese). <https://doi.org/10.11999/JEIT141649>
 19. Newman M.E.J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, vol. 103, no. 23, pp. 8577–8582. <https://doi.org/10.1073/pnas.0601602103>
 20. Blondel V.D., Guillaume J.L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, pp. P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
 21. Nguyen N.P., Dinh T.N., Tokala S., Thai M.T. Overlapping communities in dynamic networks: their detection and mobile applications. *Proc. of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom'11 and Co-Located Workshops*, 2011, pp. 85–95. <https://doi.org/10.1145/2030613.2030624>
 22. He D., Jin D., Baquero C., Liu D. Link community detection using generative model and nonnegative matrix factorization. *PLoS ONE*, 2014, vol. 9, no. 1, pp. e86899. <https://doi.org/10.1371/journal.pone.0086899>
 23. Bashir S., Chachoo M.A. Community detection in online social networks: models and methods, a survey. *Gedrag & Organisatie Review*, 2020, vol. 33, pp. 1164–1175. <https://doi.org/10.37896/gor33.03/498>
 24. Wang S., Li G., Hu G., Wei H., Pan Y., Pan Z. Community detection in dynamic networks using constraint non-negative matrix factorization. *Intelligent Data Analysis*, 2020, vol. 24, no. 1, pp. 119–139. <https://doi.org/10.3233/IDA-184432>
 25. Lu H., Zhao Q., Sang X., Lu J. Community detection in complex networks using nonnegative matrix factorization and density-based clustering algorithm. *Neural Processing Letters*, 2020, vol. 51, no. 2, pp. 1731–1748. <https://doi.org/10.1007/s11063-019-10170-1>
 26. Zachary W.W. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 1977, vol. 33, no. 4, pp. 452–473. <https://doi.org/10.1086/jar.33.4.3629752>
 27. Lusseau D., Schneider K., Boisseau O.J., Haase P., Slooten E., Dawson S.M. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 2003, vol. 54, no. 4, pp. 396–405. <https://doi.org/10.1007/s00265-003-0651-y>
 28. McAuley J., Leskovec J. Learning to discover social circles in ego networks. *Advances in Neural Information Processing Systems*, 2012, vol. 1, pp. 539–547.
 - Systems and Inventive Technology (ICSSIT). 2022. P. 334–339. <https://doi.org/10.1109/ICSSIT53264.2022.9716286>
 15. Ma X., Dong D. Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks // *IEEE Transactions on Knowledge & Data Engineering*. 2017. V. 29. N 5. P. 1045–1058. <https://doi.org/10.1109/TKDE.2017.2657752>
 16. Jiao P., Yu W., Wang W., Li X., Sun Y. Exploring temporal community structure and constant evolutionary pattern hiding in dynamic networks // *Neurocomputing*. 2018. V. 314. P. 224–233. <https://doi.org/10.1016/j.neucom.2018.03.065>
 17. Liu H.-F., Yuan L.-M.-Z. Community detection in temporal networks using triple nonnegative matrix factorization // *DEStech Transactions on Computer Science and Engineering*. 2017. <https://doi.org/10.12783/dtce/mmsta2017/19682>
 18. Wang T., Liu Y., Xi Y.-Y. Identifying community in bipartite networks using graph regularized-based non-negative matrix factorization // *Journal of Electronics and Information Technology*. 2015. V. 37. N 9. P. 2238–2245. (in Chinese). <https://doi.org/10.11999/JEIT141649>
 19. Newman M.E.J. Modularity and community structure in networks // *Proceedings of the National Academy of Sciences of the United States of America*. 2006. V. 103. N 23. P. 8577–8582. <https://doi.org/10.1073/pnas.0601602103>
 20. Blondel V.D., Guillaume J.L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment*. 2008. P. P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
 21. Nguyen N.P., Dinh T.N., Tokala S., Thai M.T. Overlapping communities in dynamic networks: their detection and mobile applications // *Proc. of the 17th Annual International Conference on Mobile Computing and Networking, MobiCom'11 and Co-Located Workshops*. 2011. P. 85–95. <https://doi.org/10.1145/2030613.2030624>
 22. He D., Jin D., Baquero C., Liu D. Link community detection using generative model and nonnegative matrix factorization // *PLoS ONE*. 2014. V. 9. N 1. P. e86899. <https://doi.org/10.1371/journal.pone.0086899>
 23. Bashir S., Chachoo M.A. Community detection in online social networks: models and methods, a survey // *Gedrag & Organisatie Review*. 2020. V. 33. P. 1164–1175. <https://doi.org/10.37896/gor33.03/498>
 24. Wang S., Li G., Hu G., Wei H., Pan Y., Pan Z. Community detection in dynamic networks using constraint non-negative matrix factorization // *Intelligent Data Analysis*. 2020. V. 24. N 1. P. 119–139. <https://doi.org/10.3233/IDA-184432>
 25. Lu H., Zhao Q., Sang X., Lu J. Community detection in complex networks using nonnegative matrix factorization and density-based clustering algorithm // *Neural Processing Letters*. 2020. V. 51. N 2. P. 1731–1748. <https://doi.org/10.1007/s11063-019-10170-1>
 26. Zachary W.W. An information flow model for conflict and fission in small groups // *Journal of Anthropological Research*. 1977. V. 33. N 4. P. 452–473. <https://doi.org/10.1086/jar.33.4.3629752>
 27. Lusseau D., Schneider K., Boisseau O.J., Haase P., Slooten E., Dawson S.M. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations // *Behavioral Ecology and Sociobiology*. 2003. V. 54. N 4. P. 396–405. <https://doi.org/10.1007/s00265-003-0651-y>
 28. McAuley J., Leskovec J. Learning to discover social circles in ego networks // *Advances in Neural Information Processing Systems*. 2012. V. 1. P. 539–547.

Authors

Shafia Bashir — Research Scholar, University of Kashmir, Srinagar, 190006, India, 57210047446, <https://orcid.org/0000-0002-5570-967X>, imshafia@gmail.com

Manzoor Ahmad Chachoo — PhD, Scientist-D, University of Kashmir, Srinagar, 190006, India, 56252797100, <https://orcid.org/0000-0001-6702-6633>, manzoor@kashmiruniversity.ac.in

Авторы

Шафия Башир — научный сотрудник, Университет Кашмира, Сринагар, 190006, Индия, 57210047446, <https://orcid.org/0000-0002-5570-967X>, imshafia@gmail.com

Мансур Ахмад Чачу — PhD, научный работник, Университет Кашмира, Сринагар, 190006, Индия, 56252797100, <https://orcid.org/0000-0001-6702-6633>, manzoor@kashmiruniversity.ac.in

Received 17.04.2022

Approved after reviewing 13.07.2022

Accepted 18.09.2022

Статья поступила в редакцию 17.04.2022

Одобрена после рецензирования 13.07.2022

Принята к печати 18.09.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»