*Review Article*

# Complex Networks, Communities, and Clustering: A survey

*Biswajit Saha[1], Amitabha Mandal[2], Soumendu Bikas Tripathy[3] & Debaprasad Mukherjee[4]*

*[1]Department of Computer Science & Engineering and Department of Information Technology,*

*Dr. B. C. Roy Engineering College, Durgapur1, West Bengal*

*Email- sahabiswaji@gmail.com*

*[2]Department of Computer Science & Engineering and Department of Information Technology,*

*Dr. B. C. Roy Engineering College, Durgapur1, West Bengal*

*[3]Camellia Institute of Engineering & Technology & Camellia Institute of Polytecnic, BudBud, West Bengal,*

*[4]Department of Computer Science & Engineering and Department of Information Technology,*

*Dr. B. C. Roy Engineering College, Durgapur1, West Bengal*

*Email-mdebaprasad@gmail.com*

## Abstract

*This paper is an extensive survey of literature on complex network communities and clustering. Complex networks describe a wide variety of systems in nature and society, especially systems composed of a large number of highly interconnected dynamical entities. Complex networks, like real networks, can also have community structure. There are several types of methods and algorithms for the detection and identification of communities in complex networks. Several complex networks have the property of clustering or network transitivity. Some of the important concepts in the field of complex networks are small-world and scale-free networks, evolving networks, the relationship between topology and the network's robustness, degree distributions, clustering, network correlations, random graph models, models of network growth, dynamical processes on networks, etc. Some current areas of research on complex network communities are those on community evolution, overlapping communities, communities in directed networks, community characterization, and interpretation, etc. Many of the algorithms or methods proposed for network community detection through clustering are modified versions of or inspired by the concepts of minimum-cut-based algorithms, hierarchical connectivity-based algorithms, the original Girvan–Newman algorithm, concepts of modularity maximization, algorithms utilizing metrics from information and coding theory, and clique based algorithms.*

**Keywords**: Networks, Clustering, Communities, Hierarchy, Information, Modularity

## Introduction

### Complex Networks

Complex networks describe a wide variety of systems in nature and society esp., systems composed of a large number of highly interconnected dynamical entities. The Internet, social networks, business networks, large circuits, networks of chemicals linked by chemical reactions, transportation networks, power networks, networks of citations of documents/ web pages,

etc, are some of the popularly cited examples of complex networks. Two of the many important questions in this field are: a) are there any unifying principles underlying the topology complex networks, and b) from the perspective of nonlinearity, how colossal networks of interacting and/or communicating dynamical systems will behave collectively, given their individual dynamics and coupling architecture.

One of the intuitive approaches to capture the global properties of such complex systems is to model them as graphs, where nodes represent the dynamical units, and links/edges represent the interactions between the nodes. It has now been widely recognized that the topology and evolution of real-world complex networks are controlled by various organizing principles of topology and dynamics. Researchers have been addressing structural and topological issues of complex networks, e.g., a) characterization of the complex interconnection architectures to comprehend the unifying principles that are the foundations of real-world complex networks, and b) constructing models to simulate the growth and replicate the structural properties of these complex networks. Researchers have also addressed the complex networks' dynamics, e.g., characterization of the collective behavior of large ensembles of dynamical systems that interact through complex interconnections topology. Some of the important concepts in the field of complex networks are small-world and scale-free networks, evolving networks, the relationship between topology and the network's robustness, degree distributions, clustering, network correlations, random graph models, models of network growth, dynamical processes on networks, etc. [1-4]Network Communities

Community structures are quite common in real networks. Complex networks have a community structure if the nodes of the networks can be grouped into sets of nodes where each set of nodes is densely connected, at least internally. For non-overlapping network communities, the complex network should split into groups of nodes with dense connections internally but sparser connections between the groups. Overlapping network communities are also possible, and they are found in many real-world complex networks, e.g., social networks. An alternative way of expressing the network community concept, inclusive of the overlapping and non-overlapping issues, is that pairs of nodes are more likely to be connected if they are both members of the same community(ies) and less likely to be connected if they are not members of the same communities. The communities can themselves also join together to form meta-communities, and those meta-communities can join together, and so on, in a hierarchical fashion. It is now well accepted that the identification of the community structure of complex networks provides insight into the relationships between network function and topology.

There are several types of methods and algorithms for the detection and identification of communities in complex networks. But, roughly, these methods can be divided into 4 categories (not exclusive):

a) node-centric (algorithms utilizing the information that each node in a group satisfies certain properties), b) group-centric (algorithms utilizing the connections within a group as a whole, i.e., the group as a single entity satisfies certain properties without zooming into node-level), c) network-centric (algorithms utilizing the process of the partitioning of the whole network into several disjoint sets), and d) hierarchy-

centric (algorithms utilizing the processes of construction of hierarchical structures of communities). There are other categories of algorithms/methods of community detection in complex networks, but they are not being discussed here. The optimal methods to detect network communities vary depending on applications, complex networks, computational resources, etc. Some current lines of research on complex network communities are those on community evolution, overlapping communities, communities in directed networks, community characterization, and interpretation, etc.[5-8].

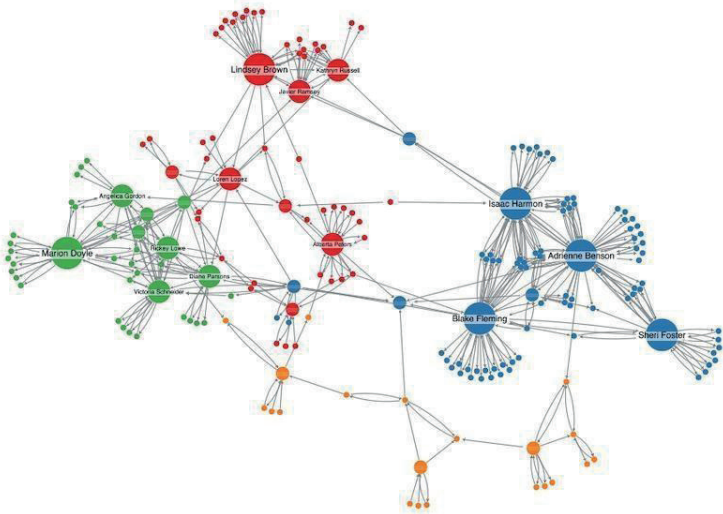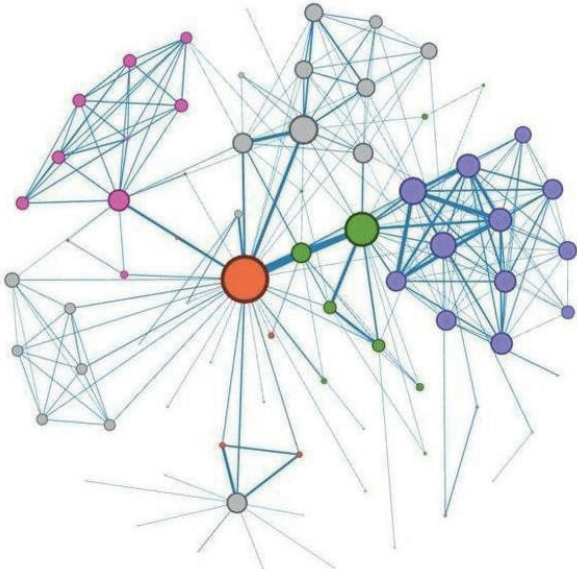**Figure-1: Communities in Complex Networks [46], 2022**



**Figure-2: Communities in Complex Networks [47] 2018**

## Clustering

Several complex networks have the property of clustering or network transitivity. It is the property that two vertices of the network that are both adjacent to the same third vertex have an increased probability of also being adjacent to one another. Clustering of network communities is the task of grouping a set of nodes in such a way that nodes in the same group (cluster) are more similar, in some measure(s), to each other than to those in other groups (clusters), based on the above-mentioned adjacency property. The choice or design of clustering algorithms and their corresponding controlling functions, e.g., distance measure/ norm, threshold/cut-offs for similarity/ dissimilarity measures, number of final expected stable clusters, degrees of randomization, criteria for cluster partitioning or merging, etc., depending on the complex networks being clustered and the nature and function of the communities to be identified in them. The clustering of network communities can also be formulated as a multi-objective optimization problem.

The differences in the cluster/ community models and their properties reflect the differences between the various clustering and network community detection algorithms. Some popular cluster/community models are a) connectivity models, e.g. hierarchical clustering algorithms create models utilizing distance-based interconnections between nodes or already existing communities, b) centroid models, e.g. models those characterize clusters/ network communities by mean vectors computed from set of properties of sets of nodes , c) distribution models based on statistical distributions of the values features of the nodes in multidimensional space, d) density models, based on density or sparseness of connections/ regions in data space,

and a very important category e) subspace models e.g. in bi clustering/ co-clustering where clusters/network communities are modeled with respect to both cluster members and pertinent set of features of the nodes, g) graph-based models e.g models/ algorithms based on quasi-cliques- subset of nodes in a graph/complex network, where a significant fraction of node pairs in the subset are connected by edges, considered as cluster/ community seeds for building up larger clusters or communities.

Many of the algorithms or methods proposed for network community detection through clustering are modified versions of or inspired by the concepts of minimum-cut-based algorithms, hierarchical connectivity-based algorithms, the original Girvan–Newman algorithm, concepts of modularity maximization, algorithms utilizing metrics from information and coding theory, and clique based algorithms [9-12].

## Research questions and topics in complex network clustering and community detection

The contemporary research questions and topics in this field are generative models, communities in time-evolving networks, characterization, modeling and the analysis of communities, impact of community structure and its dynamics on networked systems, modularity and the use of modularity maximization as the basis for community detection, stochastic strategies (that require little or no information about the network topology at the expense of their performance), recovery of stochastic models (finds the latent partition of networks nodes into the communities which are equal to or correlate with the truth communities used for generation of the given network), overlapping community structures, multilayer community benchmarks, incremental

clustering, online/real time community finding, predicting community evolution, measuring the quality of the obtained evolving communities, finding provably optimal algorithms, designing well-controlled benchmark systems on which the proposed algorithms can be tested and compared, modular centrality using local and global strategies, taxonomy of various community detection algorithms, and real-world applications of networked systems.

## Classification of Community Detection Clustering Algorithms in Complex Networks

Community detection clustering algorithms in complex networks can be classified as hierarchy-based algorithms, information-theoretic algorithms, modularity-based algorithms, and some other special categories.

### Hierarchy based algorithms

A hierarchical clustering method works by means of grouping data into a tree of clusters. It starts by considering every data point as a distinct cluster. Then, it executes the subsequent steps repeatedly. At first, the two clusters which can be closest together are identified. Then, the two maximum comparable clusters are merged. This continues till all the clusters are merged. In hierarchical clustering, the goal is to produce a hierarchical series of nested clusters. A diagram called a Dendrogram (a tree-like diagram that statistics the sequences of merges or splits) graphically represents the hierarchy. It is an inverted tree that depicts the order in which factors are merged (bottom-up view), or clusters are broken up (top-down view).Hierarchy-based algorithms are one of the main types of community detection clustering algorithms in complex networks.

Silva et al. [13] 2007 utilized the concept of topological orders among input data represented as graph and developed an algorithm to obtain clusters in different scales. The algorithm consisted of initially the network construction from input data and, secondly, the hierarchical partitioning of the formed network. The algorithm, although being completely free of the computation of the physical distances among input data, was found to constantly produce a connected graph with heavily linked nodes within a community and sporadically linked nodes among different communities. The authors applied their algorithm to the problem of pixel clustering. Zhang et al.[14] 2007, proposed a hierarchical clustering approach, based on the graph diffusion kernels of networks, to reveal the community organization of different levels of complex networks. They verified the method on some networks with well-known community structures and found that the algorithm is an effective one.

Guang Xu et al.[15] 2012, proposed a new algorithm called Latent Community Discovery, for community detection in complex social networks. Specifically, their algorithm divides the core actors based on a hierarchical probabilistic model and a statistical topic model, which are normalized by the network arrangement in data. Their algorithm is inspired by the Pareto Principle, which accounts for the uneven existence of two different types of network actors, esp. the core actors who typically occupy only a small share of the nodes but have a large influence on the complex network. They tested their algorithm on three large social networks and found its performance competitive to the existing popular algorithms for this category of problems.

de Oliveira et al. [16], 2008 propose a clustering algorithm based on graph theoretic representations and community discovery in complex networks. Initially, they represent

the input data as a network and then divide the network into sub-networks to create data groups. In the first stage, each of the nodes has a randomly assigned initial angle. This initial angle is gradually modified according to agreement with the neighbors' angle. Ultimately the network reaches a steady state. In this state, the nodes in the same cluster have comparable angles. This process is repeated and results in a hierarchical and graded divisive clustering. Simulations by the authors demonstrate that this algorithm has the potential to find clusters in different forms, compactness, and proportions. The algorithm also has the capability to generate clusters with diverse refinement grades. Furthermore, the proposed algorithm is also robust and efficient.

Malzer, C., & Baum, M. [17] 2020, in their research, demonstrated the combination of DBSCAN* and HDBSCAN clusters due to the application of an extra threshold value. They also showed the potential benefits of this hybrid approach when clustering data having various types of densities are used. Their approach is beneficial in cases where a low minimum cluster size is required, and avoiding an abundance of micro-clusters in high-density regions is required. Campello et al. [18] 2020, in their review article, discussed the statistical notion of density-based clusters, classic algorithms for deriving a flat partitioning of density-based clusters, methods for hierarchical density-based clustering, and methods for semi-supervised clustering. They concluded with various open challenges associated with density-based clustering.

## Information theoretic algorithms

Information Theory involves the quantification of information in a dataset with the help of several statistical measures. The most used information-theoretic measures are Entropy and its variation. Entropy is a measure of uncertainty about a stochastic event or, in other words, it measures the amount of missing information related to an event. The idea of information gave rise to other measures of information like Mutual information, conditional mutual information, divergency, entropy, cross-entropy, joint entropy, and conditional entropy. There are two important issues in clustering: a) how the similarity (or dissimilarity) between objects or clusters in the dataset can be measured, and b) finding out the criterion function that is to be optimized. In both these cases, information theoretic measures are used in information theoretic clustering. Using the essentials of information theory as a clustering criterion takes advantage of the underlying statistical information that the data carries.

Information theoretic algorithms are another major type of community detection clustering algorithms in complex networks. Cravino et al. [19] 2012, employed the overlapping community arrangement of a linkage of tags/labels to improve text clustering. Based on a small data set of news clips/ excerpts, the authors construct a network of co-occurrence of user-defined labels of metadata fields in news excerpts. They describe a weighted cosine similarity closeness measure, which takes into account both the excerpt vectors and the tag vectors. Thereafter, they compute the tag weight using the correlated tags that exist in the discovered community and then use the ensuing vectors, together with a novel distance metric to identify socially biased document clusters.

Yang et al. [20] 2006, proposed an unsupervised graphical clustering algorithm for finding community in complex networks by determining the dissimilarity between nodes and incorporating them into a dissimilarity

distance matrix, which when sorted according to the scores, becomes equivalent to an intensity image. The clusters are indicated by dark blocks of pixels along the main diagonal.

It is now acknowledged that trust in electronic commerce has become one of the most significant concerns in online applications, with consumers searching for the best trustworthy of goods and service providers and looking for ways of confirming which service providers are the most trusted. Zhang et al. [21] 2012, have studied the critical problem of trust network and trust community clustering for the analysis of the user's most trusted relationship for electronic commerce applications. In their model, the nodes represent the various subjects involved in the trust, the connections denote relationships, and the weight of the links indicates the strength of the relationships. Initially, the algorithm constructs a trust network having the weight value of links. Subsequently, the clustering properties of the relationship according to the weights and the path lengths are analyzed. Finally, the algorithm categorizes the most trusted subjects for a user to the same cluster. Two metrics, i.e., direct trust information degree and global trust information degree, are utilized to assess trust relationships among the subjects. This principle also gives an efficient shortest-path algorithm to construct trust networks. All the above information generated is incorporated into the clustering algorithm based on the coefficient and path length for the e-commerce trust network community.

Zhang and Zhong [22] 2013, dealt with this problem by considering the aspects of the small-world nature of trust communities and the metric local trust recommendation degree. Piccardi et al. [23] 2011, analyzed the issue of clustering of financial time series based on the network community analysis

methodology, i.e., the partitioning of the nodes of a network. A network with n nodes is associated with the set of n time series, and the weight of the link, which quantifies the similarity between the two corresponding time series, is defined according to a metric based on symbolic time series analysis. Thereafter, probing for network clusters leads to the identification of groups of nodes (i.e., Time series) with strong similarity. The authors verify the algorithm on US and Italian stock exchange time series data, and the steadiness of the clusters over time is seen to be satisfactory and better than those achieved using the minimal spanning tree and the hierarchical tree-based algorithms. Piccardi and Calatroni [24] 2010 studied the same problem in its full generality, according to the same methodology, in previous work and found satisfactory results.

Though community detection in social networks is usually based on graph clustering employing the structural information, i.e., linkage structure or node topology, for group identification, but Huang and Yang [25] 2012, used the semantic information present in the posts of social media to find hidden communities in these media. They incorporated the assumption that content issued by users may express relations between users/ entities. This method is suitable for detecting communities in networks that continuously evolve, e.g., social networks.

Liu et al. [26] 2008 applied the concept of network community clustering to an important biomedical problem known as functional analysis of protein cavities. It is known that the functions of a protein are chiefly determined by its structure, and surface cavities i.e. pockets or clefts, are generally considered as possibly active sites where the protein carries out its functions. The authors proposed

a feasible solution to the problem of functional assignment by protein cavity clustering, i.e., geometrical clustering based on the geometrical community structure of pocket similarity networks. Firstly, they introduce a pocket similarity network to methodically describe structural correspondence among pockets. The pockets are connected if they have structural similarity beyond a certain threshold. Thereafter, the surface pockets are clustered into structurally related pocket groups via a graded process. The authors then reference these small pocket groups as structural patterns which represent similar functions in different proteins. Their experimental results show that identified pocket groups are biologically meaningful in terms of their functional features.

Rui and FengMing [27] 2011 proposed an algorithm to construct a distributed trust network in information sharing and exchange channels e.g., instant messengers, file-sharing tools, etc. But, it is well known that the formation of the communities is a self-organizing and evolving autonomous phenomenon, being regulated primarily through internal and member dynamics of the community. For example, users join communities based on their interests. These kinds of network communities are highly vulnerable to the spreading of malicious software, pilfering of users' information, and attack by malicious users. Thus, to form safe and reliable communities, the authors present an algorithm by incorporating the trust value of nodes computed based on their past behaviors and constructing the communities based on similarity in trust values.

XIE et al. [28] 2011 propose a new algorithm for detecting community structures in weighted complex networks. The method constructs a weighted complex network with respect to the similarity between document pairs calculated by the cosine function. Then the algorithm searches for the dense sets and applies it to cluster text documents represented by the vector space model. ZHANG et al. [29] 2013, dealt with the issue of trust in e-commerce based on social networks using the metric of trust information degree based on mutual information between subjects. They incorporated their previously developed metrics, direct trust information degree, and global trust information degree, to build trust relations among subjects. Clustering coefficients and global trust information degree were adopted to construct trust communities. Guan-yu [30] 2011 has developed an algorithm (named Mapping Vertex into Vector algorithm) which converts all vertices in a network into vectors and finds the communities in large-scale complex networks through clustering, based on the similarity between these vertex vectors.

You-yuan et al. [31] 2009, addressed the problem of web services clustering with the help of the detection of community structure in complex networks. They did this by proposing an algorithm in which the words are denoted by nodes, and the edge weights of the network are computed from the words' co-appearances. The authors applied the Newmann's algorithm to this network and extracted the clusters of words. Thus service clustering was achieved by using the relationship between the words and the services, and that too with acceptable levels of precision and accuracy, as claimed by the authors.

**Modularity based community detection**

A network that has been effectively divided into communities has fewer edges between communities than one might anticipate. Simply some careful observers would assert that this proves

there is substantial community structure if the number of edges between two groups is only what can be predicted based on accidental chance. On the other hand, it is reasonable to assume that something interesting is occurring if the number of edges between groups is significantly less than what is predicted by chance, or equivalently if the number within groups is significantly larger. Utilizing the concept of modularity, it is possible to quantify the idea that actual community structure in a network relates to a statistically notable arrangement of edges. Modularity is the number of edges lying within groups minus the predicted number in an analogous network with edges distributed at random, up to a multiplicative constant. Positive values of modularity indicate the likelihood of a community structure and are either positive or negative. As a result, one can specifically search for community structure by looking for network divisions that have positive, and preferably high, values of modularity.

Modularity-based community detection is another major type of community detection clustering algorithm in complex networks. Liu and Li [32] 2011, developed a novel metric representation, the co-neighbor modularity matrix, to assess the quality of community/clustering identification, by which the problem of community detection is transformed into that of a problem of clustering of eigenvectors in Euclidean space. Thereafter, the network community architecture is identified with a spectral clustering algorithm. One major advantage of this algorithm is that it is free from the noise generated by the initial mean points of clustering e.g., in the k-means category of algorithms.

Scibetta et al. [33] 2013, addressed the strategy of a division of the network into clusters or district-metered areas for the detection of water losses from water distribution networks since the measurement of incoming/outgoing flows for each cluster or district-metered area allows for a quantification of water losses. The authors use the community detection approach developed in the complex network theory to identify clusters or district-metered areas in water distribution systems. The method aims to find solutions satisfying the constraints of maximization of modularity and the reduction of the number of communities. The authors claim that the method is adequately scalable.

Sharma and Purohit [34] 2013, have also applied the spectral clustering algorithm for tracking community formation in complex social networks. Zhuhadar et al. [35] 2012, have proposed the design of a visual recommender system to recommend learning resources to cyber learners within the same community, by using a community detection algorithm on the large-scale complex networks of cyber learners and learning resources, based on Web Usage data of the subjects. Their algorithm uses a heuristic that initially accomplishes clustering by force-based visualization algorithm. Subsequently, the algorithm utilizes the information on network modularity to choose good decompositions from those found using visualization algorithms.

Yu and Ding [36] 2010 applied modularity clustering objective function for network community discovery. They have shown that a normalized form modularity clustering is equivalent to the prevalent normalized cut spectral clustering. They then use this information to interpret and solve the modularity clustering problem and further corroborate the algorithm on some data collections.

Singh et al. [37] 2020 presented a comprehensive analysis of the dynamic community detection algorithms in terms of computation time and accuracy. In order to provide detailed

and extensive analysis, they tested dynamic algorithms on small, medium, and large real-world network datasets. They provided some guidelines that may help to choose the best dynamic community detection algorithms for the given complex dynamic networks based on the analysis results and network properties. Hu et al. [38] 2020 presented a new algorithm based on spectral clustering to detect the communities. Experiments demonstrated that the proposed algorithm exceeded other state-of-the-art community detection algorithms among several real-world networks from diverse domains and synthetic networks. The proposed algorithm provided a high-quality and accurate performance in a wide range of data sets. Al-Andoli et al. [39] 2021 proposed an original deep auto encoder with Particle Swarm Optimization (PSO) and continuation algorithms to reveal community structures in complex networks. Their work also proposed various other methods to work in the absence of continuation and to enable premature convergence.

**Other algorithms**

It is popularly accepted that social networks offer a dominant abstraction of the organization and dynamics of varied kinds of people or people-to-technology interaction and also endorse the use of collaborative technologies for partnerships among different groups. It is also acknowledged that finding subgroups within social networks is important for understanding and possibly influencing the formation and evolution of online communities. In this context, Sharma and Joshi [40] address the issues of tracking online communities in large-scale complex social networks. They infer the dynamics of the communities to a significant extent from the online interactions of the nodes by tracking the evolution of known sub-communities over time.

Verma and Butenko [41] 2013, use the clique relaxation concept of k-community for network community identification. The clique relaxation of k-community is a connected subgraph such that endpoints of each edge have, at the minimum, k shared neighbors within the subgraph. An important aspect of this method is that it does not use any previous information about the organization of the network. By defining a cluster as a k-community, the proposed algorithm aims to provide a clustering of a network into k-communities with varying values of k.

Lu et al. [42] 2013 proposed a novel Network Community Structure Clustering Algorithm Based on Genetic Theory. Their work puts forward the idea of applying a clustering ensemble-based genetic algorithm in the domain of complex social network mining. Their procedure introduces a clustering ensemble into the crossover operator and then utilizes the clustering information of the parents to generate new individuals. This seems to avoid the problems that are triggered by merely swapping strings between crossover operators without consideration of their contents. In population generation, Markov random walk approach is used to sustain the diversity of the entities as well as the clustering accuracy. The algorithm also uses a local searching mechanism in crossover operators to reduce the searching space and improve the speed of convergence.

Eagle et al. [43] 2012, demonstrate in what way using network community identifying methods can be used to recognize sub-goals in problems in a logic tutor. Then those community structures can be utilized to produce high-level hints among sub-goals. The authors do this by presenting a new data structure, the Interaction Network, for representing interaction data from open problem-solving environment tutors.

## Complex networks in education

In their research article, Jacobson and Wilensky [44] 2006 stated the rationale for students to learn new scientific perspectives which are related to the study of complex systems. They discussed various research carried out on issues related to learning these ideas. They also proposed a set of research issues that are related to general design principles to create environments and tools which will be helpful for the students to learn scientific ideas about complex physical and social systems. They discussed five design principles that have the potential to give promising learning sciences research. The five design principles are a) to experience complex systems phenomena; b) to make explicit the complex systems framework; c) to encourage collaboration, discussion, and reflection; d) to construct theories, models and conduct experiments; e) to learn trajectories towards deep understandings and explorations. In their research, they mentioned the importance of mainstream students to actively learning about complex systems in the physical and social sciences. They also stated that research discovering student learning of complex systems, which is conceptually challenging knowledge, may actually help in pushing the boundaries of knowing the kinds of advanced knowledge that students are capable of learning. Additionally, they pointed out that conceptual perspectives and methodologies from complex systems have the ability to influence theory and research problems of fundamental importance to the domain of learning sciences itself.

In their research article, Kunihiko et al. [45] 2016, examined the improvement in collaborative work between faculty and staff in university settings and also identified the strengths and weaknesses of the university by using both SWOT analysis and complex network analysis. Calculating multiple indices of a complex network and analyzing characteristics of the network may be further carried out in the future. This research work is an example of the application of complex networks in the domain of learning more particularly in university education.

### Table-1: Classification of the algorithms

| Sl no. | Classification | Reference of the paper |
|---|---|---|
| 1 | Hierarchy Based | 13. Silva, T. C., & Zhao, L. (2007, October). **Pixel clustering by using complex network community detection technique.** In Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007) (pp. 925-932). IEEE. |
| | | 14. Zhang, S., Ning, X. M., & Zhang, X. S. (2007). **Graph kernels, hierarchical clustering, and network community structure: experiments and comparative analysis.** The European Physical Journal B, 57(1), 67-74. |
| | | 15. Xun, G., Yang, Y., Wang, L., & Liu, W. (2012, December). **Latent community discovery with network regularization for core actors clustering.** In Proceedings of COLING 2012: Posters (pp. 1351-1360). |

| | | 16. de Oliveira, T. B., Zhao, L., Faceli, K., & de Carvalho, A. C. (2008, June). **Data clustering based on complex network community detection.** In 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence) (pp. 2121-2126). IEEE. |
|---|---|---|
| | | 17. Malzer, C., & Baum, M. (2020, September). **A hybrid approach to hierarchical density based cluster selection.** In 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (pp. 223-228). IEEE. |
| | | 18. Campello, R. J., Kröger, P., Sander, J., & Zimek, A. (2020). **Density-based clustering**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(2), e1343. |
| 2 | Information Theoretic | 19. Cravino, N., Devezas, J., & Figueira, Á. (2012, June). **Using the overlapping community structure of a network of tags to improve text clustering.** In Proceedings of the 23rd ACM conference on Hypertext and social media (pp. 239-244). |
| | | 20. Yang, S., Luo, S., & Li, J. (2006, August). **A novel visual clustering algorithm for finding community in complex network.** In International Conference on Advanced Data Mining and Applications (pp. 396-403). Springer, Berlin, Heidelberg. |
| | | 21. Zhang, S., Chen, J., Zhong, H., Fang, Z., & Shi, J. (2012). **Trust network and trust community clustering based on shortest path analysis for e-commerce.** International Journal of u-and e-Service, Science and Technology, 5(2), 31-42. |
| | | 22. Zhang, S., & Zhong, H. (2013). **Trust network and small world trust community clustering for E-Commerce**. Int. J. Hybrid Inf. Technol, 6, 1-14. |
| | | 23. Piccardi, C., Calatroni, L., & Bertoni, F. (2011). **Clustering financial time series by network community analysis.** International Journal of Modern Physics C, 22(01), 35-50. |

| | | 24. Piccardi, C., & Calatroni, L. (2010, February). Clustering time series by network community analysis. In 2010 Complexity in Engineering (pp. 94-96). IEEE. |
|---|---|---|
| | | 25. Huang, H. H., & Yang, H. C. (2012, August). **Semantic clustering-based community detection in an evolving social network.** In 2012 Sixth International Conference on Genetic and Evolutionary Computing (pp. 91-94). IEEE. |
| | | 26. Liu, Z. P., Wu, L. Y., Wang, Y., Zhang, X. S., & Chen, L. (2008). P**rotein cavity clustering based on community structure of pocket similarity network.** International journal of bioinformatics research and applications, 4(4), 445-460. |
| | | 27. Zhu, R., & Liu, F. (2011, July). **A clustering algorithm of community in distributed network based on trust.** In 2011 Eighth international conference on fuzzy systems and knowledge discovery (FSKD) (Vol. 2, pp. 1070-1073). IEEE. |
| | | 28. Xie, J., & Szymanski, B. K. (2011, June). **Community detection using a neighborhood strength driven label propagation algorithm.** In 2011 IEEE Network Science Workshop (pp. 188-195). IEEE. |
| | | 29. Zhang, S. Z., Fang, Z. X., Chen, J. G., & Shi, J. (2013). **Community clustering model for E-commerce trust based on social network.** Journal of ZheJiang University (Engineering Science), 47(4), 656-661. |
| | | 30. Wang, G. Y. (2011). **Algorithm for Detecting Community of Complex Network Based on Clustering.** Jisuanji Gongcheng/ Computer Engineering, 37(10). |
| | | 31. Ou, Y. Y., Zhang, H. S., & Men, H. (2009). **Web Services clustering based on detecting community structure in complex network**. Journal of Application Research of Computers, 6. |
| 3 | Modularity Based | 32. Liu, J., & Li, L. (2011). **Network community detection based on co-neighbor modularity matrix with spectral clustering.** In Applied Mechanics and Materials (Vol. 55, pp. 1237-1241). Trans Tech Publications Ltd. |

| | | |
|---|---|---|
| | | 33. Scibetta, M., Boano, F., Revelli, R., & Ridolfi, L. (2013). **Community detection as a tool for complex pipe network clustering.** EPL (Europhysics Letters), 103(4), 48001. |
| | | 34. Sharma, S., & Purohit, G. (2013). **A Novel Framework For Tracking Online Community Interaction In Social Network.** International Journal of Information Acquisition, 9(02), 1350011. |
| | | 35. Zhuhadar, L., Yang, R., & Nasraoui, O. (2012, December).<br><br>**Toward the design of a recommender system: visual clustering and detecting community structure in a web usage network.** In 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (Vol. 1, pp. 354-361). IEEE. |
| | | 36. Yu, L., & Ding, C. (2010, July). **Network community discovery: Solving modularity clustering via normalized cut.** In Proceedings of the Eighth Workshop on Mining and Learning with Graphs (pp. 34-36). |
| | | 37. Singh, D. K., Haraty, R. A., Debnath, N. C., & Choudhury, P. (2020, February). **An analysis of the dynamic community detection algorithms in complex networks**. In 2020 IEEE International Conference on Industrial Technology (ICIT) (pp. 989-994). IEEE. |
| | | 38. Hu, F., Liu, J., Li, L., & Liang, J. (2020). **Community detection in complex networks using Node2vec with spectral clustering. Physica A: Statistical Mechanics and its Applications,** 545, 123633. |
| | | 39. Al-Andoli, M., Cheah, W. P., & Tan, S. C. (2021). **Deep autoencoder-based community detection in complex networks with particle swarm optimization and continuation algorithms.** Journal of Intelligent & Fuzzy Systems, 40(3), 4517-4533. |
| 4 | Other | 40. Sharma, S., & Joshi, N. K. **Enhancement of existing clustering algorithm for tracking online community in social network.** |

| | | 41. Verma, A., & Butenko, S. (2013). **Network clustering via clique relaxations: A community based. Graph Partitioning and Graph Clustering,** 588, 129. |
|---|---|---|
| | | 42. Lu, N., Jin, Y., & Qin, L. (2013). **Network Community Structure Clustering Algorithm Based on the Genetic Theory. Journal of Advances in Computer Networks,** 1(2). |
| | | 43. Eagle, M., Johnson, M., & Barnes, T. (2012). **Interaction Networks: Generating High Level Hints Based on Network Community Clustering.** International Educational Data Mining Society. |

## Summary and Conclusion

The topology and evolution of complex real-world networks are controlled by various organizing principles of topology and dynamics and happen to be a major research area. Researchers have been addressing structural and topological issues of complex networks as well as complex networks' dynamics for quite some time. Some of the important concepts that have evolved in the field of complex networks are small-world and scale-free networks, evolving networks, the relationship between topology and the network's robustness, degree distributions, clustering, network correlations, random graph models, models of network growth, dynamical processes on networks, etc. Researchers apply several types of methods and algorithms for detecting and identifying communities in complex networks, which can be generally divided into 4 categories. The optimal methods to detect network communities vary depending on applications, complex networks, computational resources, etc. Some current lines of research on complex network communities are those on community evolution, overlapping communities, communities in directed networks, community characterization, and interpretation, etc.

Several complex networks have the property of clustering or network transitivity. Many of the algorithms or methods proposed for network community detection through clustering are modified versions of or inspired by the concepts of minimum-cut-based algorithms, hierarchical connectivity-based algorithms, the original Girvan–Newman algorithm, concepts of modularity maximization, algorithms utilizing metrics from information and coding theory, and clique based algorithms. Consequently, Community Detection Clustering Algorithms in Complex Networks can be broadly classified into Hierarchy Based Algorithms and Theoretic Information Algorithms. Moreover, there has been a lot of research work carried out by eminent researchers in the fields of both Hierarchy Based Algorithms and Theoretic Information Algorithms, as well as Modularity based community detection. Also, in the case of social networks finding subgroups is important for understanding and possibly influencing the formation and evolution of online communities.

Researchers have inferred the dynamics of the communities to a significant extent from the online interactions of the nodes by tracking the evolution of known sub-communities over time. Researchers have also used the clique relaxation concept of k-community for network community identification wherein by defining a cluster as a

k-community, the aim is to provide a clustering of a network into k-communities with varying values of k. Work has also been carried out based on genetic theory, and a novel Network Community Structure Clustering Algorithm has been proposed based on it. Apart from these, significant research work has also been carried out based on network community identifying methods that can be used to recognize sub-goals in problems in a logic tutor. Then those community structures can be utilized to produce high-level hints among sub-goals.

As can be seen, community detection in complex networks is an active research area. It has got real-life applications in the fields of large-scale engineering, social media analysis, biomedical data analysis, online education, business and economics, etc. It is hoped that this survey will be of great help to the scientific community and senior researchers.

## References

Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attack tolerance of complex networks. *nature*, *406*(6794), 378-382.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, *298*(5594), 824-827.

Strogatz, S. H. (2001). Exploring complex networks. *nature*, *410*(6825), 268-276.

Stutzman, F. (2006). An evaluation of identity-sharing behavior in social network communities. *Journal of the International Digital Media and Arts Association*, *3*(1), 10-18.

Donetti, L., & Munoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, *2004*(10), P10012.

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior*, *25*(1), 153-160.

Hattori, F., Ohguro, T., Yokoo, M., Matsubara, S., & Yoshida, S. (1999). Socialware: Multiagent systems for supporting network communities. *Communications of the ACM*, *42*(3), 55-ff.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc..

Hartigan, J. A., & Hartigan, J. A. (1975). *Clustering algorithms* (Vol. 209). New York: Wiley

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, *27*(2_Part_1), 209-220.

Steinbach, M., Karypis, G., & Kumar, V. (2000). A comparison of document clustering techniques KDD workshop on text mining. *Department of Computer Science/Army HPC Research Center*.

Silva, T. C., & Zhao, L. (2007, October). Pixel clustering by using complex network community detection technique. In *Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)* (pp. 925-932). IEEE.

Zhang, S., Ning, X. M., & Zhang, X. S. (2007). Graph kernels, hierarchical clustering, and network community structure: experiments and comparative analysis. *The European Physical Journal B*, *57*(1), 67-74.

Xun, G., Yang, Y., Wang, L., & Liu, W. (2012, December). Latent community discovery with

network regularization for core actors clustering. In *Proceedings of COLING 2012: Posters* (pp. 1351-1360).

de Oliveira, T. B., Zhao, L., Faceli, K., & de Carvalho, A. C. (2008, June). Data clustering based on complex network community detection. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)* (pp. 2121-2126). IEEE.

Malzer, C., & Baum, M. (2020, September). A hybrid approach to hierarchical density based cluster selection. In *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)* (pp. 223-228). IEEE.

Campello, R. J., Kröger, P., Sander, J., & Zimek, A. (2020). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(2), e1343.

Cravino, N., Devezas, J., & Figueira, Á. (2012, June). Using the overlapping community structure of a network of tags to improve text clustering. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 239-244).

Yang, S., Luo, S., & Li, J. (2006, August). A novel visual clustering algorithm for finding community in complex network. In *International Conference on Advanced Data Mining and Applications* (pp. 396-403). Springer, Berlin, Heidelberg.

Zhang, S., Chen, J., Zhong, H., Fang, Z., & Shi, J. (2012). Trust network and trust community clustering based on shortest path analysis for e-commerce. *International Journal of u-and e-Service, Science and Technology*, *5*(2), 31-42.

Zhang, S., & Zhong, H. (2013). Trust network and small world trust community clustering for E-Commerce. *Int. J. Hybrid Inf. Technol*, *6*, 1-14.

Piccardi, C., Calatroni, L., & Bertoni, F. (2011). Clustering financial time series by network community analysis. *International Journal of Modern Physics C*, *22*(01), 35-50.

Piccardi, C., & Calatroni, L. (2010, February). Clustering time series by network community analysis. In *2010 Complexity in Engineering* (pp. 94-96). IEEE.

Huang, H. H., & Yang, H. C. (2012, August). Semantic clustering-based community detection in an evolving social network. In *2012 Sixth International Conference on Genetic and Evolutionary Computing* (pp. 91-94). IEEE.

Liu, Z. P., Wu, L. Y., Wang, Y., Zhang, X. S., & Chen, L. (2008). Protein cavity clustering based on community structure of pocket similarity network. *International journal of bioinformatics research and applications*, *4*(4), 445-460.

Zhu, R., & Liu, F. (2011, July). A clustering algorithm of community in distributed network based on trust. In *2011 Eighth international conference on fuzzy systems and knowledge discovery (FSKD)* (Vol. 2, pp. 1070-1073). IEEE.

Xie, J., & Szymanski, B. K. (2011, June). Community detection using a neighborhood strength driven label propagation algorithm. In *2011 IEEE Network Science Workshop* (pp. 188-195). IEEE.

Zhang, S. Z., Fang, Z. X., Chen, J. G., & Shi, J. (2013). Community clustering model for E-commerce trust based on social network. *Journal of ZheJiang University (Engineering Science)*, *47*(4), 656-661.

Wang, G. Y. (2011). Algorithm for Detecting Community of Complex Network Based on Clustering. *Jisuanji Gongcheng/ Computer Engineering*, *37*(10).

Ou, Y. Y., Zhang, H. S., & Men, H. (2009). Web Services clustering based on detecting community structure in complex network. *Journal of Application Research of Computers*, *6*.

Liu, J., & Li, L. (2011). Network community detection based on co-neighbor modularity matrix with spectral clustering. In *Applied Mechanics and Materials* (Vol. 55, pp. 1237-1241). Trans Tech Publications Ltd.

Scibetta, M., Boano, F., Revelli, R., & Ridolfi, L. (2013). Community detection as a tool for complex pipe network clustering. *EPL (Europhysics Letters)*, *103*(4), 48001.

Sharma, S., & Purohit, G. (2013). A Novel Framework For Tracking Online Community Interaction In Social Network. *International Journal of Information Acquisition*, *9*(02), 1350011.

Zhuhadar, L., Yang, R., & Nasraoui, O. (2012, December). Toward the design of a recommender system: visual clustering and detecting community structure in a web usage network. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 354-361). IEEE.

Yu, L., & Ding, C. (2010, July). Network community discovery: Solving modularity clustering via normalized cut. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs* (pp. 34-36).

Singh, D. K., Haraty, R. A., Debnath, N. C., & Choudhury, P. (2020, February). An analysis of the dynamic community detection algorithms in complex networks. In *2020 IEEE International Conference on Industrial Technology (ICIT)* (pp. 989-994). IEEE.

Hu, F., Liu, J., Li, L., & Liang, J. (2020). Community detection in complex networks using Node2vec with spectral clustering. *Physica A: Statistical Mechanics and its Applications*, *545*, 123633.

Al-Andoli, M., Cheah, W. P., & Tan, S. C. (2021). Deep autoencoder-based community detection in complex networks with particle swarm optimization and continuation algorithms. *Journal of Intelligent & Fuzzy Systems*, *40*(3), 4517-4533.

Sharma, S., & Joshi, N. K. *Enhancement of existing clustering algorithm for tracking online community in social network.*

Verma, A., & Butenko, S. (2013). Network clustering via clique relaxations: A community based. *Graph Partitioning and Graph Clustering*, *588*, 129.

Lu, N., Jin, Y., & Qin, L. (2013). Network Community Structure Clustering Algorithm Based on the Genetic Theory. *Journal of Advances in Computer Networks*, *1*(2)

Eagle, M., Johnson, M., & Barnes, T. (2012). Interaction Networks: Generating High Level Hints Based on Network Community Clustering. *International Educational Data Mining Society*.

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the learning sciences*, *15*(1), 11-34.

Takamatsu, K., Kirimura, T., Bannaka, K., Noda, I., Mitsunari, K., Omori, M., ... & Nakata, Y. (2016, July). SWOT analysis and Complex Network analysis to enhance governance in universities by collaboration between academic and administrative faculty. In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 1188-1189). IEEE.

Disney, A. (2022, October 27). *Network Clustering*. Cambridge Intelligence. https://cambridge-intelligence.com/keylines-network-clustering/

Pirker, J. (2018). Social Network Analysis in Games User Research. In *Games User Research* (pp. 491-499). Oxford University Press.