Mathematical Problems of Computer Science 47, 50--61, 2017.

# Information-Theoretic Approach to Community Detection Problem

Mariam E. Haroutunian and Karen K. Mkhitaryan

Institute for Informatics and Automation Problems of NAS RA
e-mail: armar@ipia.sci.am, karenmkhitaryan@gmail.com

**Abstract**

Real world complex networks possess hidden information called communities or clusters, which are composed of nodes that are tightly connected within communities and weakly connected between communities. Investigation of communities proved to have countless applications in different sciences such as computer science and machine learning, biology, economics, and social networks. Parallel to the development of various detection algorithms, probabilistic network models also gained more attention, particularly stochastic block model which is a generative model for random graphs generating networks with community structure. This paper explores the state of the art on the connections of stochastic block model with information theory.

**Keywords:** Community detection, Stochastic block model, Network theory, Clustering, Information theory.

## 1. Introduction

In recent times, the computer revolution has provided specialists with massive data and sufficient computational resources to process and analyze these data. The size of real networks has also grown considerably, reaching millions or even billions of vertices and edges. The need to deal with such a large number of units has produced a deep change in the way graphs are approached. In a random graph, the distribution of edges among the vertices is highly homogeneous. Real networks are not random graphs, the distribution of edges is not only globally, but also locally inhomogeneous, with high concentrations of edges within special groups of vertices, and low concentrations between these groups. This feature of real networks is called a community structure [1, 2]. The aim of community detection in graphs is to identify the modules and, possibly, their hierarchical organization, by only using the information encoded in the graph topology.

Network is a collection of entities called nodes or vertices which are connected through edges or links. Complex network is a group of interacting entities with some nontrivial dynamical

behavior. There are many types of complex networks such as social networks, technological networks, informational networks, and biological networks. The study of complex networks is a young area of scientific research stimulated largely by the study of real-world networks like computer networks and social networks.

Identifying graph communities is a popular topic in computer science. In parallel computing, for instance, it is critical to know what the best way is to allocate tasks to processors so as to minimize the communications between them and enable a fast performance of the calculation. This can be accomplished by splitting the computer cluster into groups with roughly the same number of processors, so that the number of physical connections between processors of different groups is minimal. The mathematical formalization of this problem is called graph partitioning. The basic goal of community detection is similar to that of a graph partition: we want to separate the network into groups of vertices that have few connections between them. The main difference is that the size of the groups is not fixed.

One of the most important features representing real networks is the community structure, i.e. distribution of vertices in communities with higher internal connectivity (nodes joined by edges inside the community) than external connectivity (nodes joined by edges between communities). Detecting communities in networks proved to have many applications in various fields of science such as in protein-to-protein interactions from biology, recommendation systems from online product purchasing, social network analysis from network science, problems related to big data, etc. Although it was introduced long ago and developed for decades it is hard to evaluate detection algorithms as network types may vary. This is the reason that this field still has many open problems.

Community detection is one of the central problems in network and data sciences. Recent research has focused on developing community detection methods using various approaches. A recent review of existing approaches can be found in [3].

Evaluating the performance of algorithms on models is non-trivial. In some cases, most algorithms may succeed, while in others, algorithms may fail due to computational barriers. Thus, an important question is to characterize the situations where the clustering tasks can be solved efficiently or information-theoretically. In particular, models may benefit from asymptotic phase transition phenomena, which, in addition to being mathematically interesting, allow the location of hard cases to benchmark algorithms.

Probabilistic network models can be used to model real networks, to study the average-case complexity of NP-hard problems on graphs, or to set benchmarks for clustering algorithms with well-defined ground truth. The latter is needed to show how the model fits the data sets, and is very important in community detection as a vast majority of algorithms are based on heuristics and no ground truth is available in applications. This is, in particular, a well-known challenge for Big Data problems where one cannot manually determine the quality of the clusters.

Parallel to community detection, probabilistic network models were also theoretically developed. The stochastic block model is one of the most popular network models exhibiting community structures. The model was first proposed in the 80s [4] and received significant attention in the mathematics and computer science literature, as well as in the statistics and machine learning literature [5 -9].

Theoretical investigation brought new insights about connections of stochastic block model with information theory where decoding an information sent through channel is considered somewhat similar to community detection [5, 10, 11].

In this paper we focus on the connections between information theory and community detection. In the next section we consider some community detection algorithms. We focus on the methods based on statistical inference, mainly on the stochastic block model in the section 3.We define the weak, partial and exact recovery requirements. The last section analyzes the role of

information theory in the investigation of open problems, such as fundamental limits, comparing partitions, etc.

## 2. Community Detection

Communities or clusters are groups of vertices that play an important role in the network. In real world, networks can be towns, friendship circles, virtual groups on the web, etc. Revealing communities and investigating how particular communities behave is an attractive problem resulting in countless practical applications.

Solving community detection problems on modern real world network datasets can sometimes be very much complicated because of computational complexity as graphs may contain billions of nodes and edges and complexity of exact detection is NP-hard. Even nowadays distinguishing between algorithms and questioning the fact which algorithm works best for specific network is impossible. Moreover, there are several types of networks that make the process more challenging. These types of networks include directed and weighted networks, including those which may have overlapping communities.

Detection of such community structures in complex networks is not an easy task. In recent years many community detection algorithms are developed. Although it's still tricky to distinguish between "good" and "bad" algorithms as one algorithm working well on one network can fail for another, there are traditional approaches used broadly. Algorithms are classified into different categories. Some of them try to maximize the given quality function such as modularity, some hierarchical clustering methods introduce a similarity measure such as cosine similarity and group similar nodes into communities, etc. In this section we give a short description of some of them. We are interested in the investigation of the stochastic block model, which is from the list of methods based on statistical inference. Some advanced methods include dynamic algorithms, methods to find overlapping communities, multi-resolution methods and cluster hierarchy. These topics are not included in this paper.

### *Modularity Optimization*

Modularity is a community quality measure which measures the fraction of the edges in the network that connect vertices of the same type minus the expected value of the same quantity in a random network where community divisions are the same but connections between vertices are random. Particular algorithms that maximize the modularity for finding the community structure in networks include the Louvain and Infomap algorithms, as well as the fast greedy modularity optimization algorithm.

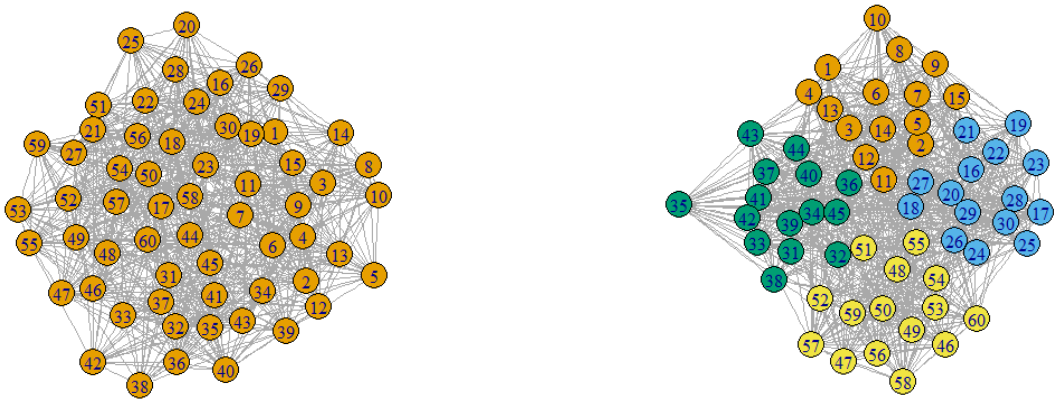

Fig. 1. In the left stochastic block model is given with 60 vertices, within communities and between communities edge probabilities 0.9 and 0.3, respectively. In the right Louvain algorithm is implemented partitioning the network into four communities with a modularity score of 0.23.

### *Minimum-cut Method*

One of the commonly used algorithms is the minimum-cut method, which tries to find communities with a predefined size by separating groups of nodes that have minimum inter group connectivity. Some of the popular algorithms related to minimum cut method are the Kernighan-Lin and Stoer-Wagner algorithms.

Kernighan-Lin algorithm inputs the undirected graph $G = (V, E)$ and partitions the vertex set $V$ into two disjoint subsets $A$ and $B$ of equal size in a way that minimizes the number of edges crossing between $A$ and $B$. The algorithm works also for weighted graphs, and the task becomes to minimize the sum of edge weights. This algorithm is used in the layout of digital circuits where minimum connections are vital for increasing performance.

For undirected graphs the Stoer-Wagner algorithm is used to calculate the minimum cut. The idea of the algorithm is to recursively merge the vertices which are tightly connected until the graph contains two vertex sets. After each step the weight of cut is listed and at the end the minimum cut will be the minimum of the graph.

### *Hierarchical Clustering*

Hierarchical clustering displays different levels of grouping vertices. The algorithms of hierarchical clustering are renowned for their applications in real world networks such as social network analysis, biology, engineering, etc., as these networks probably have a hierarchical structure. Note that communities in graphs may not have a hierarchical structure at all but with its weaknesses it's still one of the popular methods for community detection.

Hierarchical clustering starts with the definition of similarity measure of nodes. After calculating similarities between each pair of nodes in the graph one will end up with the similarity matrix and group nodes in communities.

Hierarchical clustering algorithms are divided into agglomerative algorithms in which clusters are iteratively merged if their similarity is high, and divisive algorithms where clusters are iteratively split by removing edges connecting vertices with low similarity.

***Girvan-Newman Algorithm***

Another popular method used for community detection and clustering is the Girvan-Newman algorithm. The algorithm is implemented by calculating betweenness values of all edges then the edge with the highest betweenness score is removed and after the iterative process when no edges remain the original network is split into communities.

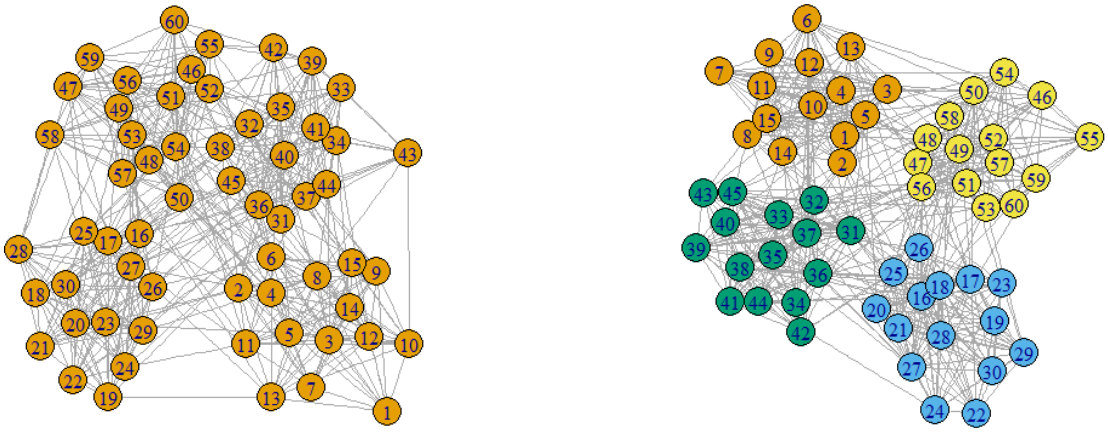

Fig. 2. In the left the stochastic block model is given with 60 vertices, within communities and between communities edge probabilities 0.9 and 0.2, respectively. In the right Girvan-Newman algorithm is implemented partitioning the network into four communities with a modularity score of 0.488

## 3. Stochastic Block Model

Block modeling is a common approach in statistics and social network analysis to decompose a graph in classes of vertices with common properties. In this way, a simpler description of the graph is attained. Vertices are usually grouped in classes of equivalence. There are two main definitions of topological equivalence for vertices: *structural equivalence* in which vertices are equivalent if they have the same neighbors; *regular equivalence*, in which vertices of a class have similar connection patterns to vertices of the other classes. Regular equivalence is a more general concept than structural equivalence. Indeed, vertices which are structurally equivalent are also regularly equivalent, but the inverse is not true. The concept of structural equivalence can be generalized to probabilistic models, in which one compares classes of graphs, not single graphs, characterized by a set of linking probabilities between the vertices. In this case, vertices are organized in classes in such a way that the linking probabilities of a vertex with all other vertices of the graph are the same for vertices in the same class, which are called *stochastically equivalent* [4].

The model appeared independently in multiple scientific communities: the terminology stochastic block model comes from the machine learning and statistics literature, while the model is called a planted partition model in theoretical computer science, and an inhomogeneous random graphs model in the mathematics literature. The stochastic block model has recently come back to the center of attention at both the practical level, due to extensions allowing overlapping communities that have proved to fit well real data sets in massive networks, and at the theoretical level due to new phase transition phenomena discovered for the two-community case.

The goal of community detection is to recover communities up to some level of accuracy.

1. *Weak recovery (also called detection).* This only requires the algorithm to output a partition of the nodes which is positively correlated with the true partition.

2. *Partial recovery*. One may ask the question of how much can be recovered about the communities.
3. *Exact recovery (also called recovery or strong consistency.)* Finally, one may ask for the regimes for which an algorithm can recover the entire clusters.

One can also study *partial-exact-recovery*, namely which communities can be exactly recovered.

The investigation of these levels is interesting not only from the mathematical point of view, but they are also relevant for applications.

**Definition 1 ([6]):** *Let a positive integer $n$ indicates the number of vertices, $m$ be the number of communities, $p = (p_1, p_2 \dots, p_m)$ be the probability vector on $\{1, \dots, m\}$ and $P$ be a symmetric $m$ x $m$ matrix of connectivity probabilities. Stochastic block model $SBM(n, p, P)$ is defined by the pair $(X, G)$, where $X$ is an $n$-dimensional random vector with components valued in $\{1, \dots, m\}$ in proportions $p$ and $G$ is an $n$-vertex undirected graph where vertices $i$ and $j$ are connected with probability $P_{X_i, X_j}$ independently of other pairs of vertices.*

| Communities | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| $C_1$ | $P_{11}$ | $P_{21}$ | $P_{31}$ | $P_{41}$ |
| $C_2$ | $P_{21}$ | $P_{22}$ | $P_{32}$ | $P_{42}$ |
| $C_3$ | $P_{31}$ | $P_{23}$ | $P_{33}$ | $P_{43}$ |
| $C_4$ | $P_{41}$ | $P_{24}$ | $P_{34}$ | $P_{44}$ |

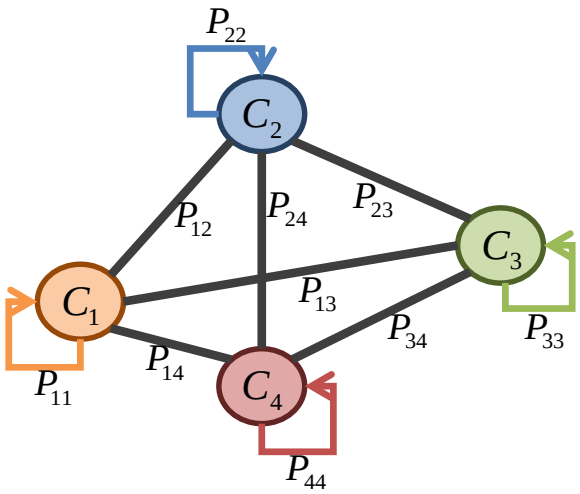


Fig. 3. Stochastic block matrix (left) of stochastic block model (right) is given, where probabilities of edges between communities is $P_{i,j}$ and inside communities $P_{i,i}$

Particularly if all $P_{i,j}$ elements of matrix $P$ are the same, the model is equivalent to Erdos-Renyi random graph model which does not have a community structure. Planted partition model is a special case where diagonal elements of matrix $P$ are constant $p$ and off diagonal elements are constant $q$. If $p > q$ the model is called assortative and if $p < q$ dissortative.

**Definition 2 ([6]):** *An algorithm detects communities with accuracy $\alpha \in [0,1]$, if it takes $G$ drawn from $SBM(n, p, P)$ and outputs a reconstruction $X'$ of $X$ that has agreement $\alpha$ with probability* $1 - o_n(1)$.

**Definition 3 ([6]):** *In stochastic block model*

**Exact recovery** is possible, if there exists an algorithm with accuracy $\alpha = 1$.

**Strong recovery** is possible, if there exists an algorithm with accuracy $\alpha = 1 - o_n(1)$.

**Weak recovery** is possible, if the algorithm detects communities with accuracy $\alpha = \frac{1}{k} + \varepsilon$ for some $\varepsilon > 0$.

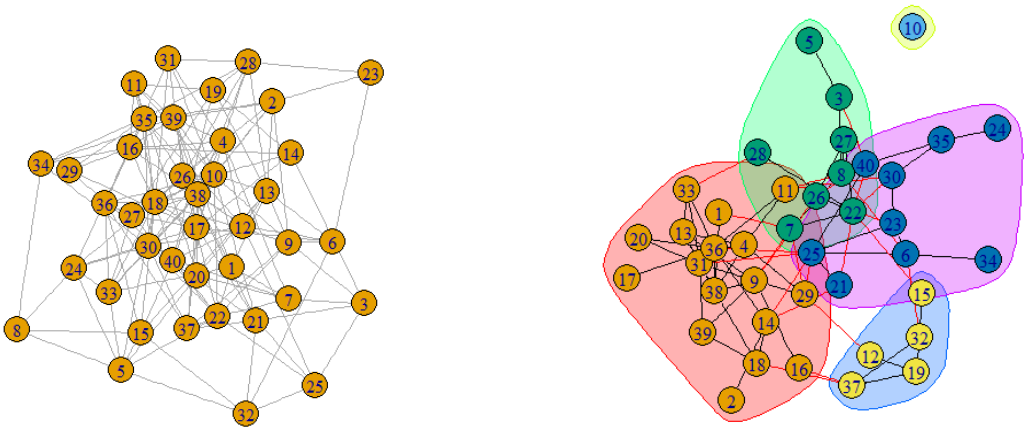

Fig. 4. In the left Erdos-Renyi random graph with 40 vertices is shown, where probability of edges between any pairs is constant. In the right , stochastic block model is given where probability of edges inside the communities is greater than probabilities of edges between communities

Majority of work in this field is done by establishing fundamental thresholds illustrating when it is possible to detect communities.

## 4. The Role of Information Theory

While the sphere of community detection is developing for many years with the construction of various algorithms to solve detecting tasks, the biggest portion of it is still unsolved. Main questions are how accurately a particular algorithm detects communities or which communities can be revealed. These issues need thorough analysis. The Information theory plays an important role in solving different problems. Community detection has natural connections with information theory at various levels. Theory behind Stochastic block model is similar to graph-based codes, f-divergences, broadcasting problems on trees which are renowned topics from information theory.

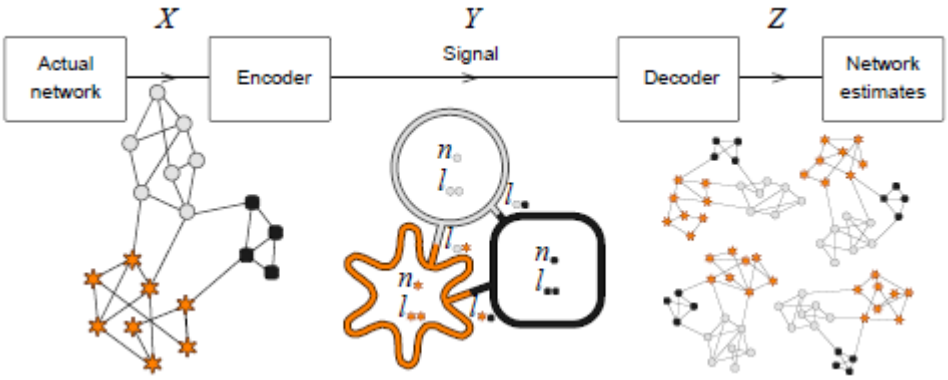


Fig. 5. An encoder sends a compressed information to a decoder about the topology of the graph on the left. The information gives a coarse description of the graph, which is used by the decoder to deduce the original graph structure. Figure reproduced from [3, 12].

The modular structure of a graph can be considered as a compressed description of the graph to approximate the whole information contained in its matrix. Based on this idea, Rosvall and Bergstrom [12] envisioned a communication process in which a partition of a graph in

communities represents a synthesis Y of the full structure that a signaler sends to a receiver, which tries to infer the original graph topology X from it.

The best partition corresponds to the signal Y that contains the most information about $X$. This can be quantitatively assessed by the minimization of the conditional entropy $H(X|Y)$ of $X$ given by $Y$. One has to look for the ideal tradeoff between a good compression and a small enough conditional entropy $H(X|Y)$.

Rosvall and Bergstrom used the same idea introducing an information-theoretic flow-based method to examine the multipartite organization of biological and social systems [13]. The method reveals the community structure in weighted and directed graphs. They used the probability flow of random walks on a network as a proxy for information flows and by compressing the description of the probability flow they decomposed the network into communities. The idea is in expressing the Shannon entropy of the random walk within and between clusters. If clusters are well separated from each other, transitions of the random walker between clusters will be infrequent, so it is advantageous to use the map, with the clusters as regions, because in the description of the random walk the codewords of the clusters will not be repeated many times, while there is a considerable saving in the description due to the limited length of the codewords used to denote the vertices. Instead, if there are no well-defined clusters and/or if the partition is not representative of the actual community structure of the graph, transitions between the clusters of the partition will be very frequent and there will be little or no gain by using the two-level description of the map.

Information theory has also been used to detect communities in graphs. Ziv et al. [14] have designed a method in which the information contained in the graph topology is compressed in such a way as to preserve some predefined information. As a criterion the mutual information $I(X;Y)$ of two random variables $X$ and $Y$ is used [15]. If $X$ is the input variable, $Z$ is the variable specifying the partition and $Y$ is the variable encoding the information we want to keep relevant variable, the goal is to minimize the mutual information between $X$ and $Z$ (to achieve the largest possible data compression), under the constraint that the information on $Y$ extractable from $Z$ be accurate. The optimal tradeoff between the values of $I(X;Z)$ and $I(Y;Z)$ (i. e., compression versus accuracy) is expressed by the minimization of a functional. In the case of graph clustering, the question is what to choose as a relevant information variable. Ziv et al. proposed to adopt the structural information encoded in the process of diffusion on the graph.

Information theory is also useful when comparing different partitions of the network. When a particular algorithm is implemented, to assess the quality of the partition, it must be compared with other partitions or with available ground truth. This can be done using several evaluation measures.

Most similarity measures can be divided into three categories: measures based on pair counting, cluster matching and information theory.

The third class of similarity measures is based on reformulating the problem of comparing partitions as a problem of message decoding within the framework of information theory [15]. The idea is that, if two partitions are similar, one needs very little information to infer one partition given by the other. This extra information can be used as a measure of dissimilarity.

The mutual information is not ideal as a similarity measure: in fact, given a partition $X$, all partitions derived from $X$ by further partitioning (some of) its clusters would all have the same mutual information with $X$, even though they could be very different from each other. In this case the mutual information would simply equal the entropy $H(X)$, because the conditional entropy would systematically be zero. To avoid that, Danon et al. [16] adopted the normalized mutual information

$$I_{norm}(X;Y) = \frac{2I(X;Y)}{H(X) + H(Y)},$$

which is currently very often used in tests of graph clustering algorithms. The normalized mutual information equals 1 if the partitions are identical, whereas it has an expected value of 0 if the partitions are independent.

Meila [17] introduced the variation of information $V(X;Y) = H(X|Y) + H(Y|X)$ which has some desirable properties with respect to the normalized mutual information and other measures. In particular, it defines a metric in the space of partitions as it has the properties of distance. It is also a local measure, i.e., the similarity of partitions differing only in a small portion of a graph depends on the differences of the clusters in that region, and not on the partition of the rest of the graph. The maximum value of the variation of information is log n, so the similarity values for partitions of graphs with different sizes cannot be compared with each other. For meaningful comparisons one could divide $V(X;Y)$ by log n, as suggested by Karrer et al. [18].

In recent years fundamental limits were obtained using the new f-divergence function, which is called the CH-divergence in [5] as it generalizes both the Chernoff and Hellinger divergences. The definite characterization of the recovery threshold in the general stochastic block models provides an operational meaning to a divergence function analog to the Kullback – Leibler divergence (KL-divergence) in the channel coding theorem.

At a high level, clustering the stochastic block model is similar to reliably decoding a codeword on a channel which is non-conventional in information theory. The channel inputs are the nodes' community assignments and the channel outputs are the network edges. It is shown [6] that this analogy is more than just high-level: reliable communication on this channel is equivalent to exact recovery, shows that the “clustering capacity" is obtained from the CH-divergence of channel-kernel PQ, which is an f-divergence like the KL-divergence governing the communication capacity. Interestingly recovering communities in stochastic block model have parallels with sending a codeword through discrete memoryless channel in coding theory.

More generally, community detection pairs well with information theory as it can be viewed as a decoding problem on a noisy channel: the community labels are the input to a black-box channel that provides local and noisy interactions of the inputs. This view point was further developed in [19], with the notion of graphical channels.

**Definition 4 ([6]):** *Let $V = \{1, \dots, n\}$ and $G = \big(V, E(G)\big)$ be a hypergraph with $N = |E(G)|$. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets called the input and output alphabets respectively and $Q$ be a channel from $\mathcal{X}^k$ to $\mathcal{Y}$ called the kernel. To each vertex in $V$, assign a vertex variable in $\mathcal{X}$, and to each edge in $E(G)$, assign an edge variable in $\mathcal{Y}$. Let $y_I$ denote the edge variable attached to edge $I$, and $x[I]$ denote the $k$ node variable adjacent to $I$. A graphical channel with graph $G$ and kernel $Q$ is defined as the channel $P$ given by $P(y|x) = \prod_{I \in E(G)} Q(y_I|x[I])$, where $x \in \mathcal{X}^V$, $y \in \mathcal{Y}^{E(G)}$.*

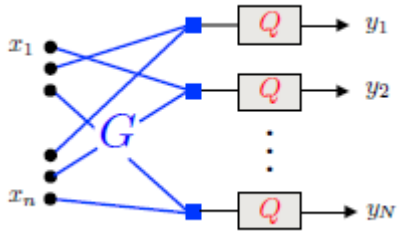


Figure reproduced from [6]

Definition 4 strongly connects community detection in stochastic block model and information sent through channel in coding theory.

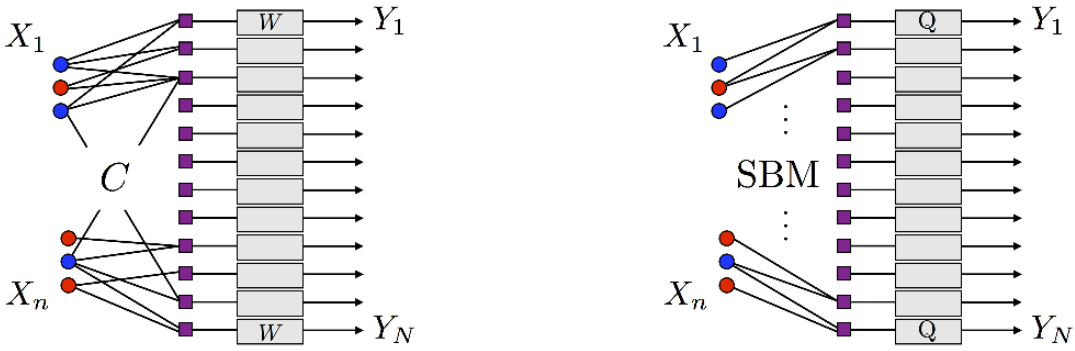


Figure reproduced from [5]

Reliable communication through noisy channel is possible if $R < 1 - H(\varepsilon)$ and this is close to exact recovery of communities in stochastic block model.

Community detection has a strong connection with information theory also because X is typically discrete.

## 5. Conclusion.

Recent analysis of community detection problems and stochastic block model proved to have a strong perspective with information theory. We hope that our experience in solving coding problems for various noisy channels [20] will lead us to new results by investigating both community detection problems and stochastic block model by applying information-theoretical approach.

## References

[1] M. Newman, “Networks: an introduction”, Oxford University Press, Oxford, 2010.
[2] M. Girvan and M. Newman, "Finding and evaluating community structure in networks", *Physical Review E*, vol. 69, 026113, 2004.
[3] S. Fortunato, "Community detection in graphs", *Physics Reports* 486, pp. 75-174, 2010.
[4] P. Holland, K. Laskey and S. Leinhardt, "Stochastic block models: First steps ", *Social Networks*, vol. 5, no. 2, pp. 109-137, 1983.
[5] E. Abbe and C. Sandon, "Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms", *IEEE 56th Annual Symp. on Foundations of Computer science, USA*, pp. 670 – 688, 2015.
[6] E. Abbe, "Community detection and the stochastic block model", *IEEE Information Theory Society Newsletter*, vol. 66, no. 1, pp. 3- 12, 2016.
[7] E. Abbe and C. Sandon, "Recovering communities in the general stochastic block model without knowing the parameters”, *Advances in Neural Information Processing Systems*, vol. 28, pp. 676 – 684, 2015.
[8] A. Decelle, F. Krzakala, C. Moore and L. Zdeborova, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications", *Physical Review E*, vol. 84, 066106, 2011.

[9] V. Kanade, E. Mossel and T. Schramm, “Global and local information in clustering labeled block models”, *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5906– 5917, 2016.
[10] E. Abbe and M. Wainwright, “Information theory meets machine learning”, *tutorial, Intern. Symp. on Information Theory*, 2015.
[11] E. Abbe, A. Bandeira and G. Hall, "Exact recovery in the stochastic block model", *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 471 – 487, 2016.
[12] M. Rosvall and C. T. Bergstrom, “An information-theoretic framework for resolving community structure in complex networks”, *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 18, pp. 7327 – 7331, 2007.
[13] M. Rosvall and C. T. Bergstrom, “Maps of random walks on complex networks reveal community structure”, *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 4, pp. 1118 - 1123, 2008.
[14] E. Ziv, M. Middendorf and C. H. Wiggins, “Information theoretic approach to network modularity”, *Physical Review E*, vol. 71, no. 4, 046117, 2005.
[15] D. J. C. Mackay, “Information theory, inference and learning algorithms”, *Cambridge University Press, fourth printing*, 2005.
[16] L. Danon, A. Diaz-Guilera, J. Duch and A. Arenas, “Comparing community stucture identification”, *Journal of Statistical Mechanics: Theory and Experiment*, P09008, 2005.
[17] M. Meila, “Comparing clusterings – an information based distance”, *Journal of Multivar. Anal.*, vol. 98, no. 5, pp. 873 – 895, 2007.
[18] B. Karrer, E. Levina and M. Newman, “Robustness of community structure in networks”, *Physical Review E*, vol. 77, 046119, 2008.

**Submitted 08.10.2016, accepted 20.02.2017.**

# Ինֆորմացիոն տեսական մոտեցում համայնքների հայտնաբերման խնդրին

Մ. Հարությունյան և Կ. Մխիթարյան

## Ամփոփում

Իրական աշխարհի բարդ ցանցերը օժտված են թաքնված ինֆորմացիայով, այսպես կոչված համայնքներով կամ խմբերով, որոնց ներսում հանգույցները ավելի խիստ են կապված, քան համայնքների միջև: Համայնքների հետազոտությունը հիմնավորված է մեծ թվով կիրառություններով տարբեր գիտություններում, ինչպիսիք են՝ կոմպյուտերագիտությունը և մեքենայական ուսուցումը, կենսաբանությունը, տնտեսագիտությունը և սոցիալական ցանցերը: Համայնքների հայտնաբերման տարբեր ալգորիթմների մշակման հետ զուգահեռ հետազոտողների ուշադրությունն են գրավում նաև հավանականային ցանցերի մոդելները, մասնավորապես, ստոխաստիկ

բլոկ մոդելը, որը պատահական գրաֆների կառուցման մոդել է և ստեղծում է համայնքների կառուցվածքով ցանցեր: Այս հոդվածում ուսումնասիրվում է ստոխաստիկ բլոկ մոդելի և ինֆորմացիայի տեսության միջև կապի վերաբերյալ գիտական արդյունքների վիճակը:

# Информационно-теоретический подход к задаче обнаружения сообществ

М. Арутюнян и К. Мхитарян

**Аннотация**

Реальные сложные сети обладают скрытой информацией под названием сообщества или кластеры, состоящих из узлов, тесно связаных в кластере и слабо связанных между сообществами. Исследование сообществ подтвердило бесчисленное множество применений в различных науках, таких как компьютерные науки и машинное обучение, биология, экономика и социальные сети. Параллельно с развитием различных алгоритмов обнаружения сообществ, модели вероятностных сетей также привлекают больше внимания, в частности стохастическая блочная модель, которая создает сети со структурой сообщества. В данной статье исследуется современное состояние науки о связях стохастической блок модели с теорией информации.