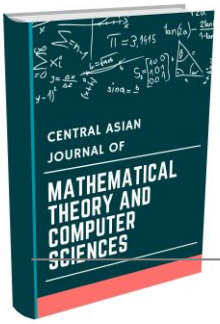


CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

https://cajmtcs.centralasianstudies.org

Volume: 04 Issue: 5 | May 2023 **ISSN: 2660-5309**

THE MAXIMUM REALIZATION METHOD OF COMMUNITY GROUPING IN SOCIAL NETWORKS

Zakhidov Dilshodbek
PhD student of Andijan State University

Shukrulloev Bektosh
PhD student of Samarkand State University
shukrulloyev92@mail.ru

Abstract

Identifying communities in social networks is one of the most important tasks nowadays. Social networks are especially relevant for networks represented by large-scale graphics. At the same time, it is important to use approximate methods that lead to near-optimal rather than optimal results within a given time. In this article, we propose to extract the communities based on the method of maximum likelihood similarity by representing them by letters. The community structure search algorithm is described and the performance of the algorithm is illustrated using examples of different views of the octagonal network.. Calculations were carried out using the Maple program.

A R T I C L E I N F O

Article history:
Received 3 Mar 2023
Revised form 5 Apr 2023
Accepted 19 May 2023

Keywords: *maximum likelihood. Graphics. Communication between teams. Teams section. Maple.*

***

Introduction

He intensity of today's time shows that our life cannot be imagined without the Internet or social networks. Most people's activities are related to being aware of the daily news happening in the world through social networks and transmitting them to each other. Formally, they are members of a social network and their interactions can be represented in the form of a graph of connections—their correspondence, common interests, and mutual friends. In this case, they can be analyzed using mathematical methods. Social networks change rapidly, so random graphs are one of the tools to study them. Mathematical analysis of graphs of social networks is carried out using different measures, using the center of the vertices and the edges of the graph.

Real social networks are characterized by considerable disorder depending on the organization of the networks[1,2]. This is because the connectivity graph structure contains groups of vertices characterized by a greater distribution of intra-group connections than vertices in other groups. Given these characteristics, it is possible to distinguish between communities in the network. Identifying communities in a social network helps to identify abnormal behavior of its members. In general, there are two approaches to identifying communities in a network: when communities may not intersect and when they may overlap. If we are interested in the professional, family, friendly relations of the participants, the latter can happen.

The article [4] closest to the present work, in which it was proposed to apply the maximum likelihood method for graph clustering. The difference in approaches is determined by the fact that in [4] the analyzed graph assumes. The real graph is generated randomly with given parameters for internal and external links.

A probabilistic approach called the maximum likelihood method, widely used in mathematical statistics, can be used to identify communities in the network. Following the approach described in [1], we write a mathematical model for community detection based on the maximum likelihood method.

Let us assume that the network is randomly generated. The number of teams is fixed. It is clear that the closeness of relationships within the community is higher than outside the community. We consider the following parameters: 1) - the probability of connection between any two vertices in the community; 2) - probability of connection between two vertices from different communities. By maximizing the most likely structure of the partitioning into teams over all possible network configurations, we obtain a partition that matches the real data.

Consider a network $G=(N,E)$ in which the set of vertices has $N=\{1,2,...,n\}$ appearances. Let the number of edges of the network be $m=m(E)$, and the connection between vertices $E(i,j)$ i and j be

$$E(i,j)=\begin{cases}1, \mathit{If}\ \ \mathit{there\ is\ a\ connection\ between\ i\ and\ j\ teams}\\ 0, \mathit{If}\ \ \mathit{there\ is\ no\ connection\ between\ teams\ i\ and\ j}\end{cases}.$$

By community S we mean a non-empty subset of network vertices, and by partition $\Pi(N)$ we mean a set of disjoint communities whose union is exactly N sets:

$N:\Pi(N)=\{S_1,S_2,...,S_k\}$ here $\bigcup_{k=1}^{K} S_k = N$

Suppose that the real part of the network is $\Pi=\{S_1,S_2,...,S_k\}$. Let the variables $n_k=n(S_k)$ and $m_k=m(S_k)$ denote the number of vertices and edges in the community $S_k, k=1,...,K$, respectively. Then $n=\sum_{k=1}^{K} n_k$ and $\sum_{k=1}^{K} m_k \le m$.

Let us express the conditions under which the division into teams is optimal.

Simple graph. Check out the $S_k \in \Pi$ community. The probability of making m_k connections between n_k vertices in community S_k is equal to

$$p_{in}^{m_k}\left(1-p_{in}\right)^{\frac{n_k(n_k-1)}{2}-m_k}.$$

Each vertex i in a community S_k may have $n-n_k$ connections with vertices from other communities, but in fact it has $\sum_{j\notin S_k} E(i,j)$ connections with vertices from other communities have

The probability of realizing a network with a given structure is equal to

$$L_{\Pi}=\prod_{k=1}^{K} p_{in}^{m_k}\left(1-p_{in}\right)^{\frac{n_k(n_k-1)}{2}-m_k}\prod_{i\in S_k} p_{out}^{\frac{1}{2}\sum_{j\notin S_k} E(i,j)}\left(1-p_{out}\right)^{\frac{1}{2}\left(n-n_k-\sum_{j\notin S_k} E(i,j)\right)} \quad (1).$$

Taking the logarithm of the probability function L_p in (1) and simplifying it, we get

$$l_{\Pi} = \log L_{\Pi} = \sum_{k=1}^{K} m_k \log p_{in} + \sum_{k=1}^{K} \left(\frac{n_k (n_k - 1)}{2} - m_k \right) \log (1 - p_{in}) +$$

$$+ \left(m - \sum_{k=1}^{K} m_k \right) \log p_{out} + \left(\frac{1}{2} \sum_{k=1}^{K} n_k (n - n_k) - \left(m - \sum_{k=1}^{K} m_k \right) \right) \log (1 - p_{out}) \quad (2).$$

The partition Π^* in which the function l_{Π} reaches its maximum over all possible partitions is called optimal. Note that there is still uncertainty in the choice of probabilities p_{in} and p_{out}. The function $l_{\Pi} = l_{\Pi}(p_{in}, p_{out})$ depends on the arguments of p_{in}, p_{out}. By maximizing l with respect to p_{in} and p_{out}, these values can then be used in numerical calculations.

Statement of the problem

Statement 1. For a fixed partition Π, the function $l_{\Pi}(p_{in}, p_{out})$ reaches a maximum in

$$p_{in} = \frac{2\sum_{k=1}^{K} m_k}{\sum_{k=1}^{K} n_k^2 - n}, \quad p_{out} = \frac{2\left(m - \sum_{k=1}^{K} m_k \right)}{n^2 - \sum_{k=1}^{K} n_k^2}. \quad (3).$$

We create a model in Maple to calculate the function l_{Π} at each division: since we have calculated several divisions, the program "restart;" Let's start with the command.

*Step 1.*We enter the number of vertices in the partitions (that is, n[k]s).

n[1] := ;

n[2] := ;

n[3] := ;

....

Step 2. We enter the number of edges in the partitions (that is, m[k]s).

m[1] := ;

m[2] := ;

m[3] := ;

…….

*Step 3.*We enter the total number of vertices and edges in the graph (that is, y and x).

y := ;

x := ;

Step 4. We introduce the formula for calculating the function l_{Π}.lp := sum(m[k]*ln(p[i]), k = 1 .. 3)+sum(((1/2)*n[k]*(n[k]-1)-m[k])*ln(1-p[i]), k = 1 .. 3)+(x-sum(m[k], k = 1 .. 3))*ln(p[o])+((1/2)*(sum(n[k]*(y-n[k]), k = 1 .. 3))-x+sum(m[k], k = 1 .. 3))*ln(1-p[o]);

Step 5. We set the function l_{Π} to 0 by differentiating it with respect to p_{in} and p_{out}.

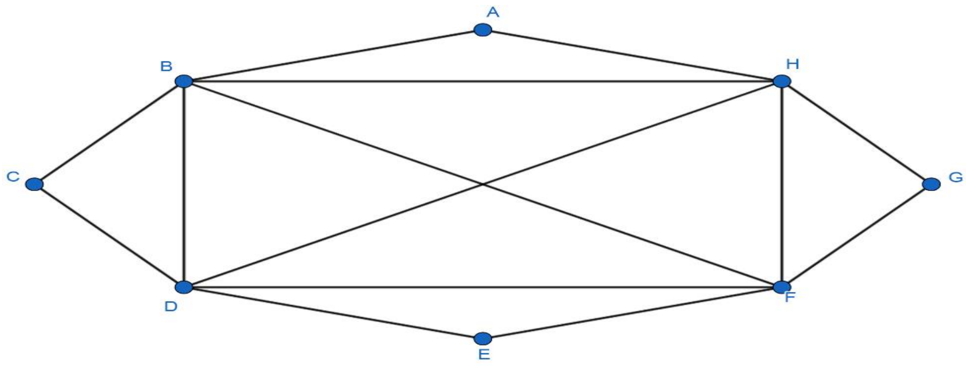
p[i] := solve(diff(lp, p[i]) = 0);

p[o] := solve(diff(lp, p[o]) = 0);

Step 6. We calculate the value of the function l_{Π} based on p_{in} and p_{out} found in step 5.

evalf(lp);

Sample: Consider the simple octagonal network shown in Figure 1



Drawing 1. octagonal network

Let's calculate the value of l_{Π} for different partitions.

We obtain from the probability function (2) for the distribution

$\Pi = \{A, B, C, D, E, F, G, H\}$.

The total number of vertices is $n=8$ and the total number of edges is $m=14$, where $n_1=8$ and $m_1=14$ since there is one group.

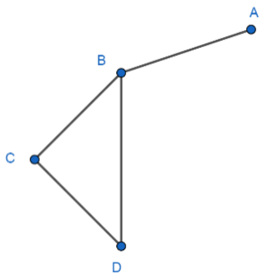
$$l_{\Pi} = 14\log p_{in} + 14\log(1-p_{in}).$$

We differentiate the function l_{Π} and set its derivative to 0:

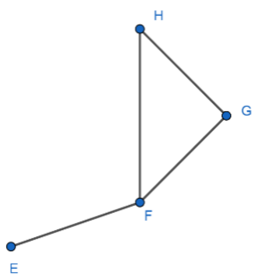
$$\frac{14}{p_{in}} - \frac{14}{1-p_{in}} = 0$$

the maximum value of the function is reached at $p_{in} = \frac{1}{2}$. Its value is -19.408.

$$\Pi = \{A, B, C, D\} \cup \{E, F, G, H\}$$



Drawing 2. $\{A, B, C, D\}$ section



Drawing 3. $\{E, F, G, H\}$ section

Let's calculate the value of l_{Π} for division. The total number of vertices is $n=8$ and the total number of edges is $m=14$, where since there are two groups, the number of vertices is $n_1=4$ and $n_2=4$ and the edges are $m_1=4$ and $m_2=4$ will be.

$$l_{\Pi} = 8\log p_{in} + 4\log(1-p_{in}) + 6\log p_{out} + 10\log(1-p_{out})$$

Here we also differentiate the function l_{Π} with respect to p_{in} and p_{out} and set the derivative to 0:

$$\begin{cases} \dfrac{8}{p_{in}} - \dfrac{4}{1-p_{in}} = 0 \\ \dfrac{6}{p_{out}} - \dfrac{10}{p_{out}} = 0 \end{cases}$$

the maximum value of the function is reached at $p_{in} = \frac{2}{3}$ and $p_{out} = \frac{3}{8}$. Its value is -18.2.

Conclusions

The value of the remaining divisions is given in the following table:

Divisions	n_k , m_k	l_{Π}	p_{in} , p_{out}	$l_{\Pi}(p_{in}, p_{out})$
$\{A,B,C,D,E,F,G,H\}$	$n_1 = 8$ $m_1 = 14$	$14\log p_{in} + 14\log(1-p_{in})$	$p_{in} = \frac{1}{2}$	-19.4081
$\{A,B,C,D\} \cup \{E,F,G,H\}$	$n_k = (4,4)$ $m_k = (4,4)$	$8\log p_{in} + 4\log(1-p_{in}) +$ $6\log p_{out} + 10\log(1-p_{out})$	$p_{in} = \frac{2}{3}$ $p_{out} = \frac{3}{8}$	-18.2231
$\{A,B\} \cup \{C,D,E,F\} \cup \{G,H\}$	$n_k = (2,4,2)$ $m_k = (1,4,1)$	$6\log p_{in} + 2\log(1-p_{in}) +$ $8\log p_{out} + 12\log(1-p_{out})$	$p_{in} = \frac{3}{4}$ $p_{out} = \frac{2}{5}$	-17.9589
$\{A,B\} \cup \{C,D,E,F,G,H\}$	$n_k = (2,6)$ $n_k = (1,8)$	$9\log p_{in} + 7\log(1-p_{in}) +$ $5\log p_{out} + 7\log(1-p_{out})$	$p_{in} = \frac{9}{16}$ $p_{out} = \frac{5}{12}$	-19.1153
$\{A,B,C\} \cup \{D,E,F,G,H\}$	$n_k = (3,5)$ $m_k = (2,7)$	$9\log p_{in} + 4\log(1-p_{in}) +$ $5\log p_{out} + 10\log(1-p_{out})$	$p_{in} = \frac{9}{13}$ $p_{out} = \frac{1}{3}$	-17.5718
$\{A,B,C\} \cup \{D,E,F\} \cup \{G,H\}$	$n_k = (3,3,2)$ $m_k = (2,3,1)$	$6\log p_{in} + \log(1-p_{in}) +$ $8\log p_{out} + 13\log(1-p_{out})$	$p_{in} = \frac{6}{7}$ $p_{out} = \frac{8}{21}$	-16.8259
$\{A,B\} \cup \{C,D\} \cup \{E,F,G,H\}$	$n_k = (2,2,4)$ $m_k = (1,1,4)$	$6\log p_{in} + \log(1-p_{in}) +$ $8\log p_{out} + 13\log(1-p_{out})$	$p_{in} = \frac{3}{4}$ $p_{out} = \frac{2}{5}$	-17.9589
$\{A,C,F\} \cup \{B,D\} \cup \{E,G,H\}$	$n_k = (3,2,3)$ $m_k = (2,1,3)$	$6\log p_{in} + \log(1-p_{in}) +$ $7\log p_{out} + 14\log(1-p_{out})$	$p_{in} = \frac{6}{7}$ $p_{out} = \frac{8}{21}$	-16.8259
$\{A,C\} \cup \{B,D\} \cup \{E,G\} \cup \{F,H\}$	$n_k = (2,2,2,2)$ $m_k = (1,1,1,1)$	$4\log p_{in} +$ $10\log p_{out} + 14\log(1-p_{out})$	$p_{in} = 1$ $p_{out} = \frac{5}{12}$	-16.3006

It can be seen that partition $\Pi = \{A, B\} \cup \{C, D\} \cup \{E, F\} \cup \{G, H\}$ gives the most probable community structure for this network.

References

1. Мазалов В. В., Никитина Н. Н. Метод максимального правдоподобия для выделения сообществ в коммуникационных сетях // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2018. Т. 14. Вып. 3. С. 200–214. https://doi.org/10.21638/11702/spbu10.2018.302.
2. Fortunato S., Barthelemy M. Resolution limit in community detection. Proceedings of the National Academy of Sciences USA, 2007, vol. 104(1), pp. 36–41.
3. Girvan M., Newman M.E. J. Community structure in social and biological networks. Proceedings of the National Academy of Sciences USA, 2002, vol. 99(12), pp. 7821–7826.
4. Copic J., Jackson M., Kirman A. Identifying community structures from network data via maximum likelihood methods // The B. E. Journal of Theoretical Economics. 2009. Vol. 9, iss. 1. P. 1635–1704.