Horizon | Journal of Humanity and Artificial Intelligence

DIVISION OF HEPTAGONAL SOCIAL NETWORKS INTO TWO COMMUNITIES BY THE MAXIMUM LIKELIHOOD METHOD

Dilshodbek Zakhidov
Senior Lecturer of TMC Institute

Shukrulloev Bektosh
PhD student of Samarkand State University

Abstract

Our work is devoted to the study of the maximum likelihood method for dividing heptagonal social networks into two groups with opposite mentalities and views of the world. We estimated the probability of connection between users and built an adjacency matrix to apply the maximum likelihood method. The results showed that the maximum likelihood method can be an effective tool for separating heptagonal social networks into two groups. We have identified conservative and liberal groups that have opposite worldviews and mentalities. Our research can help to better understand how the mentality of these user groups affects the social dynamics in general.

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

I. Введение. Social networks are complex systems that consist of many nodes and links between them. Social network analysis is an important tool for understanding social interactions and identifying communities in networks. One of the methods for dividing social networks into communities is the maximum likelihood method.

In this article, we will consider the application of the maximum likelihood method for dividing heptagonal social networks into two communities. The maximum likelihood method is a statistical method that allows you to determine the most likely network structure that divides it into two communities.

The purpose of the article is to analyze and apply the maximum likelihood method for dividing heptagonal social networks into two communities. We will consider the adjacency matrix for a social network and determine the probabilities of links between nodes in the network. We then determine the maximum likelihood for a network split into two communities and identify the most likely network structure that splits it into two communities.

The results obtained can be used to define communities in social networks and understand the social interactions between nodes within and between communities. A probabilistic approach, called the maximum likelihood method, widely used in mathematical statistics, can be used to identify communities in a network. Following the approach described in [1], we will write a mathematical model for community detection based on the maximum likelihood method.

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

It is clear that the tightness of relations within society is higher than outside society. We consider the following parameters: The probability of a connection between any two vertices within a team is p_{in}, the probability of a connection between two vertices from different teams is p_{out}.

Consider a network $G=(X,Y)$ in which the set of vertices has the form $X=\{1,2,\ldots,7\}$. The number of edges of the network is $m=m(Y)$. Let the connection between vertices i and j be as follows:

$$E(i,j)=\begin{cases}1, \text{If there are connections between vertices i and j,}\\ 0, \text{If there is no connection between vertices i and j.}\end{cases}$$

By a community S we mean a non-empty subset of network vertices, and by a partition $\Pi(X)$ we mean a set of non-overlapping communities whose union is exactly the set X: $N:\Pi(X)=\{S_1,S_2\}$, where $\bigcup_{k=1}^{2} S_k = S_1+S_2$.

Assume that the real partition of the network is $\Pi(X)=\{S_1,S_2\}$. Let the variables $n_k=n(S_k)$ and $m_k=m(S_k)$ denote the number of vertices and edges in the community $S_k, k=1,\ldots,K$, respectively. Then $n=7$ and $\sum_{k=1}^{K} m_k \le m$.

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

Let's look at the community $S_k \in \Pi$. The probability of creating m_k connections between n_k vertices in the S_k community is p_{in}.

Each vertex i in the community S_k can have $n-n_k$ connections to the vertices of other communities, but in fact it is connected to the vertices of other communities $\sum_{j\notin S_k} E(i,j)$ has connections.

The probability of realizing a network with a given structure is

$$L_\Pi=\prod_{k=1}^{2} p_{in}^{m_k}(1-p_{in})^{\frac{n_k(n_k-1)}{2}-m_k}\prod_{i\in S_k} p_{out}^{\frac{1}{2}\sum_{j\notin S_k}E(i,j)}(1-p_{out})^{\frac{1}{2}\left(7-n_k-\sum_{j\notin S_k}E(i,j)\right)} \quad (1)$$

Taking the logarithm of the likelihood function L_Π (1) and simplifying it, we get

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

$$+\left(m-\sum_{k=1}^{2}m_k\right)log p_{out}+\left(\frac{1}{2}\sum_{k=1}^{2}n_k(7-n_k)-\left(m-\sum_{k=1}^{2}m_k\right)\right)log(1-p_{out}) \quad (2)$$

The partition Π^* for 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 p_{in}, p_{out}. Maximizing l_Π with respect to p_{in}, p_{out}, one can then use these values in numerical calculations.

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

$$p_{in}=\frac{2(m_1+m_2)}{(n_1^2+n_2^2)-7}, p_{out}=\frac{2(m-(m_1+m_2))}{49-(n_1^2+n_2^2)}. \quad (3)$$

Substituting (3) into (2), we obtain an expression that depends only on the structure of the network. The given values of the parameters also maximize the likelihood function (1).

If we divide social networks into two groups, we will study which one is more realistic.

II. Main part.

Numerical experiments. Let's look at different types of heptagonal social network. We divide the given network into 2 groups and find the one that is most similar to the truth. When the heptagonal divides the network into 2 teams, the following situations occur:

№	1st team	2nd team
1	2 edges	5 edges
2	3 edges	4 edges

First, let's look at the following graph:

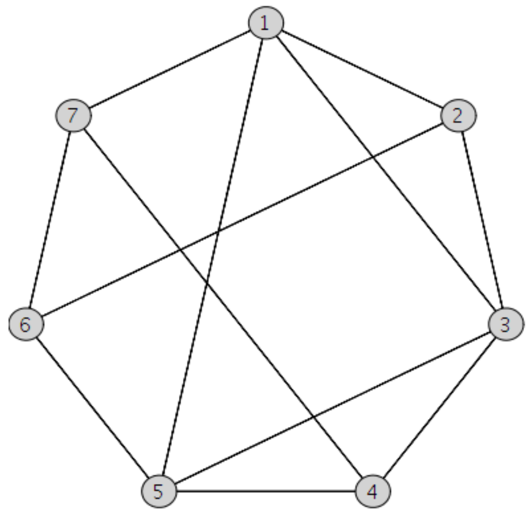

Fig.1 is a network with 7 vertices (12 edges).

This network has 7 vertices and 12 edges. Let's calculate the value l_{Π} for division in the first case.

We obtain the probability function (2) for the partition

$\Pi = \{1,2\} \cup \{3,4,5,6,7\}$

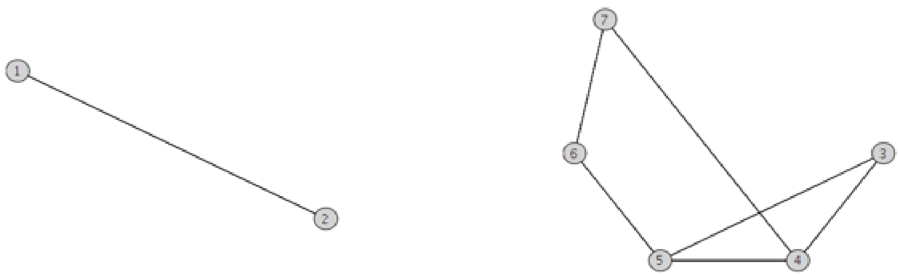

Fig. 2 division of the network into two parts

The total number of vertices is $n = 7$, and the total number of edges is $m = 12$, since there are 2 teams, the first team has 2 vertices and 1 edge, and the second team has 4 vertices and 6 edges.

$$l_{\Pi} = 7logp_{in} + 4log(1 - p_{in}) + 5logp_{out} + 5log(1 - p_{out})$$

We differentiate the function l_{Π} with respect to p_{in} and p_{out} and set the derivative to 0.

$$\begin{cases} \dfrac{7}{p_{in}} - \dfrac{4}{1-p_{in}} = 0 \\ \dfrac{5}{p_{out}} - \dfrac{5}{1-p_{out}} = 0 \end{cases}$$

Its maximum is reached at $p_{in} = \frac{7}{11}$ and $p_{out} = \frac{1}{2}$ and is equal to -14.14177132.

Let's calculate the value of l_{Π} for dividing $\Pi = \{1,2,3\} \cup \{4,5,6,7\}$ in the second case.

Fig.3. dividing the network into two parts

The total number of vertices is $n=7$, and the total number of edges is $m=12$, since there are 2 teams, the first team has 3 vertices and 3 edge, and the second team has 4 vertices and 4 edges.

$$l_{\Pi} = 7logp_{in} + 2log(1-p_{in}) + 5logp_{out} + 7log(1-p_{out})$$

We differentiate the function l_{Π} with respect to p_{in} and p_{out} and set the derivative to 0.

$$\begin{cases} \dfrac{7}{p_{in}} - \dfrac{2}{1-p_{in}} = 0 \\ \dfrac{5}{p_{out}} - \dfrac{7}{1-p_{out}} = 0 \end{cases}$$

Its maximum is reached at $p_{in} = \frac{7}{9}$ and $p_{out} = \frac{5}{12}$ and is equal to -12.91767498.

Now imagine all the divisions in the first case:

№	Разделение	l_{Π}	p_{in}	p_{out}
1	$\Pi = \{1,2\} \cup \{3,4,5,6,7\}$	-14.14177132	$\frac{7}{11}$	$\frac{1}{2}$
2	$\Pi = \{2,3\} \cup \{4,5,6,7,1\}$	-14.14177132	$\frac{7}{11}$	$\frac{1}{2}$
3	$\Pi = \{3,4\} \cup \{5,6,7,1,2\}$	-14.14177132	$\frac{7}{11}$	$\frac{1}{2}$
4	$\Pi = \{4,5\} \cup \{6,7,1,2,3\}$	-14.14177132	$\frac{7}{11}$	$\frac{1}{2}$
5	$\Pi = \{5,6\} \cup \{7,1,2,3,4\}$	-14.14177132	$\frac{7}{11}$	$\frac{1}{2}$

6	$\Pi=\{6,7\}\cup\{1,2,3,4,5\}$	-13.17559547	$\frac{8}{11}$	$\frac{2}{5}$
7	$\Pi=\{7,1\}\cup\{2,3,4,5,6\}$	-14.14177132	$\frac{7}{11}$	$\frac{1}{2}$
max		-13.17559547		

This table shows that the maximum value of l_Π is -13.17559547.

When l_Π reaches its maximum value, its p_{in} is large compared to others and its p_{out} is small compared to others.

When dividing 3 by 4, i.e. when the 1st team has 3 vertices and the 2nd team has 4 vertices, the maximum likelihood function, l_Π , reaches its maximum when dividing $\Pi=\{1,2,3\}\cup\{4,5,6,7\}$, which has a value of -12.91767498.

we can do these calculations using the algorithm in the "maple" program presented in [2].

CONCLUSION. If we compare the maximum value of l_Π in each state, we see that $l_\Pi = -12.91767498$ in division of 3 by 4 is greater than the rest, the p_{in} it gets is greater than the p_{in} in the other state, and p_{out} we see that it is small compared to other cases.

To conclude, if the above network with 7 vertices and 12 edges is divided by two, its likelihood function l_Π reaches its maximum value when dividing 3 by 4, that is, in the following case:

Fig.4. dividing the network into two parts

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. Dilshodbek, Z., & Bektosh, S. (2023). THE MAXIMUM REALIZATION METHOD OF COMMUNITY GROUPING IN SOCIAL NETWORKS. CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES, 4(5), 56-61. https://doi.org/10.17605/OSF.IO/5RQ2S
3. Fortunato S., Barthelemy M. Resolution limit in community detection. Proceedings of the National Academy of Sciences USA, 2007, vol. 104(1), pp. 36–41.
4. 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.
5. 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.