**Community detection with given confidence.**

Koldanov P.A., Semenov D.P.

National Research University Higher School of Economics, Nizhny Novgorod, Laboratory of Algorithms and Technologies for Networks Analysis

**Problem statement.** The problem of community detection frequently arise in network analysis. In the report such problem is considered in the framework of random variables network. As the definition of community we consider a set of pairwise strongly connected nodes (cliques) of the network model. We are interested in community detection by observations with given confidence. Our aim is to divide the set of conclusions on elements of community by two parts: part of confident conclusions and part of conclusions without confidence.

**Model.** Random variables network is a pair $(X,γ)$, where $X=(X\_{1},…,X\_{N})$ is a random vector and $γ=|γ\_{ij}|$ is a measure of dependence between $X\_{i}$ and $X\_{j}$. Threshold graph is a pair $(V,E)$, where $V=(V\_{1},…,V\_{N})$ is a set of vertices and $E=(\left(i,j\right):γ\_{ij}>γ\_{0})$ is a set of edges with weights greater than threshold $γ\_{0}$. Clique of size k of the threshold graph is the complete subgraph on k vertices of the threshold graph.

**Methodology.**

Different approaches to problem of cliques of threshold graph identification by observations are discussed. In the present report the multiple hypotheses testing methodology is used. It is assumed that tests for individual hypotheses $h\_{ij}:γ\_{ij}\geq γ\_{0}$ and $h'\_{ij}:γ\_{ij}\leq γ\_{0}$ are known.

**Results.**

1. Methodology for community identification which allows to devide the set of conclusions by two part is proposed. At the first step the network model is constructed by the observations on complex system. At the second step upper and low sets of edges with given confidence $P^{\*}$ are constructed. To construct the upper and low sets of edges well known multiple hypotheses testing Holm procedure is used. At the third step cliques of upper and low sets of edges are identified. It is proved that cliques in the upper and low sets of edges are upper and low bounds of true cliques respectively. This mean that edges in cliques in low set of edges are edges with given confidence and edges in difference between upper and low sets of edges are edges without given confidence.

2. Improved procedure for upper bound construction to true cliques with size greater 2 is proposed. The procedure is based on additional upper set of edges construction for edges of clique from upper set of edges.

3. Proposed methodology is applied to stock market data of several countries. The upper and low sets of edges with confidence $P^{\*}=0.9,0.95$ and communities of the network models of the stock markets are obtained.

**Acknowledgments.** The article was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE).

**References:**

1. Kalyagin V. A., Koldanov A. P., Koldanov P., Pardalos
P. M. Statistical Analysis of Graph Structures in Random Variable
Networks. Springer, 2020.
2. H. Finner a, G. Giani Closed subset selection procedures
for selecting good populations. Journal of Statistical Planning and
Inference 38 (1994) 179-200.
3. [Колданов А. П.](https://www.hse.ru/org/persons/13428988), Колданов П. А., [Семенов Д. П.](https://www.hse.ru/org/persons/46739833) [Построение доверительного множества связанных акций фондового рынка](https://publications.hse.ru/view/478712545) // Журнал Новой экономической ассоциации. 2021. Т. 2. № 50. С. 12-34.