GRAPH APPROXIMATION PROBLEMS (CORRELATION CLUSTERING)

A.V. Kononov[1], V.P. Il'ev[2], S.D. Il'eva[2]

[1]*Sobolev Mathematical Institut SB RAS, Novosibirsk*
e-mail: alvenko@math.nsc.ru
[2]*Omsk State University, Omsk*
e-mail: iljev@mail.ru

The graph approximation problems arise in analysis of systems consisting of interrelated objects, in particular, in some classification and clusterization problems. In these problems the goal is to minimize the number of links between given classes and the number of missing links within classes. The formulations and various interpretations of the graph approximation problems can be found in [1-5].

In the paper we consider only *simple* graphs, i.e., the graphs without loops and multiple edges. If $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are labelled graphs on the same vertex set V, then the *distance* $d(G_1, G_2)$ between them is defined as follows: $d(G_1, G_2) = |E_1 \setminus E_2| + |E_2 \setminus E_1|$, i.e., $d(G_1, G_2)$ is the number of noncoinciding edges in G_1 and G_2. A simlpe graph is called an M*-graph* or a *cluster graph* if each of its connected components is the complete graph.

Graph Approximation Problem. Given an arbitrary graph $G = (V, E)$, find a nearest to G M-graph M^*.

In the different versions of the problem the number of connected components of graph M^* must be equal to a positive integer $k \geq 2$ or bounded from above by k. Moreover, the problem may involve constraints on the size of the connected components. The weighted versions of the graph approximation problems were also studied. In these problems the weight function $w : V \times V \to Z_+$ is defined and $d(G_1, G_2)$ equals to the total weight of noncoinciding edges in the graphs G_1 and G_2. In general, all versions of the graph approximation problems are NP-hard.

The graph approximation problem was formulated at first as the problem of approximating symmetric relations by equivalence relations [5]. It was intensively studied in seventies of XX century. In XXI century the graph approximation problem was repeatedly rediscovered and independently studied under different names (Correlation Clustering [3], Cluster Editing [4]).

In this survey the number of known and new results on the graph approximation problems including the results on computational complexity and performance guarantees of approximate algorithms for the graph approximation problems are presented.

REFERENCES

1. A.A. Ageev, V.P. Il'ev, A.V. Kononov, A.S. Talevnin. *Computational complexity of the graph approximation problem.* — J. of Applied and Industrial Mathematics, 2007, vol.1, №1, pp. 1-8.
2. V.P. Il'ev, S.D. Il'eva, A.A. Navrotskaya. *Approximation algorithms for graph approximation problems.* — J. of Applied and Industrial Mathematics, 2011, vol.5, №4, pp. 569-581.
3. N. Bansal, A. Blum, S. Chawla. *Correlation clustering.* — Machine Learning, 2004, vol.56, pp. 89-113.
4. R. Shamir, R. Sharan, D. Tsur. *Cluster graph modification problems.* — Discrete Applied Mathematics, 2004, vol.144, №1-2, pp. 173-182.
5. C.T. Zahn. *Approximating symmetric relations by equivalence relations.* — J. of the Society for Industrial and Applied Mathematics, 1964, vol.12, №4, pp. 840-847.