Software Support for Team Engineering: Educational Example for IT Students

Olga M.Zvereva, S.M.Riad Shams
Engineering School of Information Technologies, Telecommunications and Control Systems
Ural Federal University
Ekaterinburg, Russia
OM-Zvereva2008@yandex.ru, ShamsRiad@gmail.com

***Abstract*—Every scientific project is a result of efforts of professionals from different working areas on the basis of their combined knowledge implementation. In the view of the team work significance for IT professional career, it was proposed as the main active learning method in the course from IT educational university programme. To engineer effective teams, SNA-methodology was implemented. In this study, the comprehensive information technology and software instruments for its support have been used for this problem solution. Comparative analysis of two different frameworks was carried out which has given one more educational effect. These joint studies which involve students into professional and scientific activity are very useful for educating real specialists.**

Keywords—SNA-methodology, network, team work, software, educational programme

I. INTRODUCTION

Nowadays, every scientific project is a result of efforts of professionals from different working fields on the theoretical basis of their combined knowledge implementation. While project realizing, different skills are in demand, and only various kind of knowledge in technological and social fields will yield in a project success.

Almost every Internet site with information for successful job hiring recommends to mention the skill of team work in your resume as employers expect employees to be team players. Scan any job listing and you will see that even ads that seek “self-starters” also inevitably drop the phrase “team player” [1].

Meredith Belbin, the author of the most popular theory of team work, postulated, “I think we've always been striving to find a way of helping people to work together. When people work in effective combinations they achieve so much more than when they're working alone”.

It is evident that different people greatly differ in their traits of character and live goals, in their professional skills and live experience. Because of this diversity, it is really stunning to live in this world, but, actually, it should become a real problem when it is necessary to solve the definite problem in definite interval of time in conditions of limited resources. The necessity and hardship of together work were realized by business and scientific communities as the “team work” problem.

To consider IT field, every project has become a very complex job, which involves not only IT professionals but experts from different non-IT fields as well who are speaking quite different language, Recent meetings with employers only confirm this industrial demand.

But in our IT educational programmes almost no attention (or, at least, a few) is paid to the issues team work competencies and skills. On the whole, these programmes are concentrated on the IT professional, technical and natural sciences competencies. Moreover, university authorities often postulates that our technical educational programmes are “overloaded” with humanitarian disciplines, and that these discipline volumes should be reduced**.**

II. PROBLEM DISCUSSION

In the course “Modern Scientific Technologies: Big Data Analysis”, team work was proposed as the main method of active learning.

This course had several specific features: firstly, it was oriented towards managerial competences acquisition because the main goal of analysis was discussed from the standpoint of strategic management, and in the focus should be data which is relevant for the issue of business competitive advantage achievement. Secondly, this course was taught by a foreign professor, and it was delivered in English. A person who ever works in one of the Russian universities knows that the problem with foreign language knowledge really exists, there are several reasons for this issue but it is not the topic of this paper and will not be discussed. These two specific course features are enough to get clear understanding that the students in experiment were in rather stressed situation, and this course learning could become a real challenge for them. In this case, to engineer good teams has become a half the battle. Moreover, it has given a real chance to implement information technologies for a real problem solution.

To realize this process of team engineering, it was decided to implement SNA-methodology’s (SNA—Social Network Analysis) concepts and techniques [2]. SNA-methodology considers every social group to be a network, in discussed case, students have become network nodes, and team relations have appeared to be network links.

This methodology is theoretically based on the concepts and structures of relational algebra, implements axioms and theorems of graph theory, and uses similar to sociometric indices and algorithms.

Following the popular sociometric techniques, primary data was collected by means of a questionnaire. Interviews or observation techniques [3], in this case, might be more time- and effort-consuming but less effective.

It was assumed that during the first stage students would express their wishes, that is why, they were asked to write the names of those who they wished to be in one team. This type of a questionnaire is called "free recall", the alternative is "roster", which assumes a choice, in most cases of a multivariate form, from a list. Free recall was implemented in order to provide a kind of freedom in number of those chosen to be co-members in a team, at the same time, it was a clear perception that it would be too difficult to write a lot of names.

43 university students from two academic groups of the 4th year of education took part in this research. It was necessary to divide them into the teams of 5 to 7 members. These teams had to deliver presentations according to the predetermined theme once a week. At the end of the course, a scientific resulting report had to be made (one report from every team).

After having processed all the answers, socio matrix (org matrix) was built. Socio matrix is a square matrix, in our case, its dimension was 43x43 where every row (column) corresponds to a single student. A row contains data about corresponding student choices, and a column contains data about those who were chose, i.e. if the *i-th* student chooses the *j-th* one for team work, then, the matrix *ij-th* element is set to 1 (it is set to 0, otherwise). Thus, a binary matrix with zero diagonal elements has been received for the further analysis.

III. Software for SNA-methodology support

There are several information instruments supporting SNA-methodology, most of them propose not only estimate different network indices but also can visualize a network using different layouts, shapes and colors. Instrument choice and its usage evaluation have become additional goals of the research.

According to Wikipedia, network analysis software generally consists of either packages based on graphical user interfaces (GUIs), or packages built for scripting/programming languages [4].

GUI packages are easier to learn, while scripting tools are more powerful and extensible. Amongst open-sourced and well-documented GUI packages Wikipedia mentions EgoWeb 2.0 (open source), NetMiner, UCINet, Pajek (freeware), GUESS, ORA, Cytoscape, Gephi, SocNetV (free software), Meerkat (SNA), and muxViz (open source).

As scripting tools Wikipedia describes NetMiner with Python scripting engine, Statnet suite of packages for the R statistical programming language, igraph, which has packages for R and Python, muxViz (based on R statistical programming language and GNU Octave), NetworkX library for Python, and SNAP package for large-scale network analysis in C++ and Python.

To summarize this information, in order to use these packages you must have skills in coding in C++, R, or Python.

Other information sources attract your attention to almost the same software products. In [4] five frameworks are compared, they are as follows: Pajek, Gephi, Netlytic, GraphViz, and Social Network Visualizer. Gephi can be used under GNU GPL (license) and the other four are freeware.

One example more, in [5] there is an attempt to classify all information instruments for social network analysis based on their functionality. In this survey, there is a special class of information systems named "systems for social network analysis in scientific research", and most of listed instruments are devoted to this class. Gephi is characterized as a system of medium size graph (up to 1 000000 nodes) analysis and visualization. NetworkX, igraph, and SNAP are discussed as libraries of computing functions, and characterized as applications for large networks with several millions of nodes.

IV. Network analysis

Although the author and IT students have experience in coding, it was considered to propose a technique for team assignment based on standalone software, i.e. GUI packages.

Two frameworks were used in this research: Pajek (Fig.1) and UCINET (Fig.2) for the aim of their comparison. The reasons of this choice are the following: these products are widely used for scientific research, and their functionality will be enough for successful information support of this study; these products have different commercial nature as Pajek is a freeware while UCINET is a proprietary product (with a free trial version).

One of the main differences is the result of their different commercial nature. UCINET is a commercial version, that is why, it tries to meet customer's demands, and one of the most common customer demand is to start using the product as soon as possible with minimal efforts to learn it, thus, its interface is quite similar to most common office products used for years on PC.

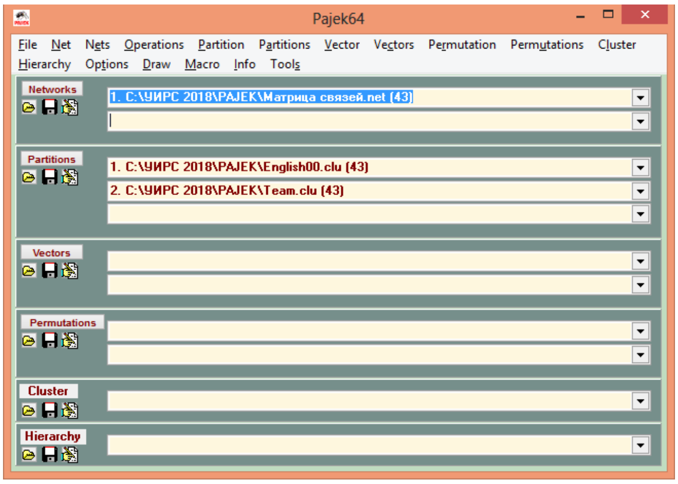

Fig. 1. Main screen of Pajek

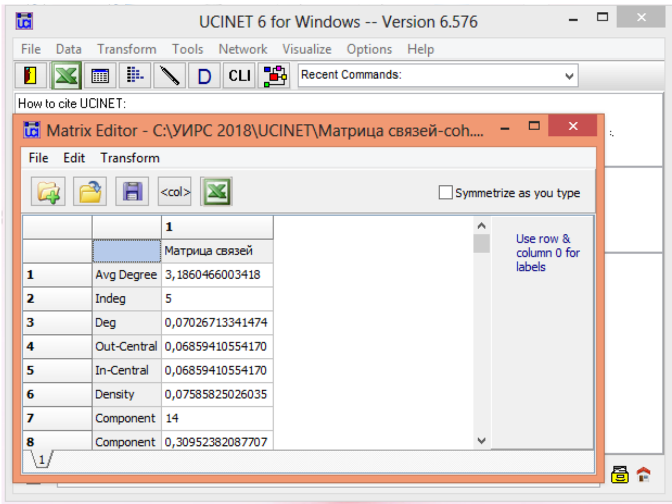

Fig. 2. Main screen of UCINET

As a rule of thumb, freeware authors try to introduce a kind of ideology into their products, their programs are aimed not only to carry out separate functions but often are focused on a whole problem solving.

One more difference revealed on the first stage of software usage was the form of data presentation: UCINET is oriented towards matrices, as it is named in its User Guide – "matrix-centered" view of the data, while most of the data in Pajek are presented in a vector form. This result in UCINET's full compatibility with Excel, while Pajek can only convert output data into the Excel format.

In Pajek network data is stored in a text file of the special format with .net extension, one can edit this file with any text editor, e.g. Notepad. In UCINET network data is stored in the special dataset which can be edited only with special software (UCINET, or some other frameworks, e.g. Pajek). This dataset files have .##h extension.

Nevertheless, having serious differences, these products are very "friendly" to each other, in another words, they are compatible: anyone may use UCINET datasets as input data for Pajek, and, vice versa, Pajek datasets can be imported into UCINET and analyzed with its help. After choosing the analytical instruments and building the socio matrix, network datasets were constructed in both frameworks and the network was visualized (in Fig.3 Pajek network presentation is shown). UCINET uses built-in NetDraw, which is a free instrument, for network visualization, and also it is possible to call Pajek from UCINET to draw a network as well.

The network under research is an oriented and non-weighted one. Main indices, which characterize this network, were calculated in both frameworks. Almost for all of them, the same values were received. The only difference was for an verage degree: Pajek has calculated it as 6.372, and UCINET – as 3.186 (twice less, may be, UCINET do not consider link orientation). The main indices are comprised in Table 1.

According to the network indices, this network is sparse (with low density value) but enough coherent (high level of clustering coefficient), and with high reciprocity (more than 62% of arcs are reciprocal).

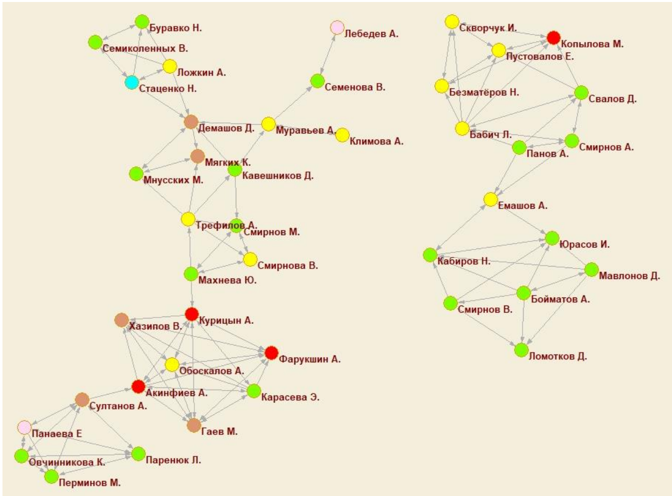

Fig. 3. Network under research with English knowledge differentiation in Pajek64 (v.5.01)

TABLE I. MAIN NETWORK INDICES

Network Index	Value
Number of vertices	43
Number of links	137
Average Degree	3.186
Density	0.076
Diameter	6.000
Arc Reciprocity	0.628
Clustering Coefficient	0.530

One interesting phenomenon was revealed: there is no boundary between academic groups – students from one group often choose students from another group (there are no preferences for the same group students).

The main goal of this research was team engineering which means that we had to divide from 43 students into 5 to 7 teams, and the most evident way to do this was to find components in the engineered network. As a component, a highly connected, in comparison to the whole network, subnetwork was comprehended.

In the first stage of the component detection network graphic representation was used. A network can be drawn in many ways, and each drawing stresses different structural features. The most useful in this case is the energy concept which states that the distance between vertices should express the strength or number of their ties as closely as possible. The energy commands "pull" vertices to better positions until they are in a state of equilibrium. Therefore, these procedures are known as "spring embedders" [8].

Pajek supports two energy representations: Kamada-Kawai and Fruchterman Reingold. One can find corresponding options in the main menu of its graphical editor window (Layout/Energy/Kamada-Kawai|Fruchterman Reingold). Draw screen of Pajek is accessed from Draw command in the Main menu of Pajek. In Fig.3 the network has Kamada-Kawai layout (with one more applied option selected from the submenu

“Separate Components”) and, due to this particular algorithm, components have become visible.

As one can notice, vertices in Fig.3 have different colors. Every color determines the level of student’s knowledge in English, in case that this fact was very important for the project. To enable this Pajek functionality, a partition was created.

“A partition of a network is a classification or clustering of the vertices in the network such that each vertex is assigned to exactly one class, or cluster. In network analysis, partitions store discrete characteristics of vertices” [8]. That is why, a partition was created where every level of English knowledge was coded with a natural number from 0 to 5 (A0 level was coded as 0, A1-1, A2-2, B1-3, B2-4, C1-5). It became evident that most of our students have the level of A2, in Fig.3 this level is depicted with the green color.

Pajek makes a report where statistic data are delivered (if you choose Partition/Info command in Pajek Main menu), statistics for English knowledge level is shown in Table 2.

TABLE II. FREQUENCY DISTRIBUTION OF ENGLISH KNOWLEDGE LEVEL IN THE NETWORK

Cluster (1)	**Freq. (2)**	**Freq (%) (3)**	**CumFreq (4)**	**CumFreq (%) (5)**	**Representative (6)**
0	1	2.3256	1	2.3256	23
1	11	25.5814	12	27.9070	1
2	20	46.5116	32	74.4186	3
3	4	9.3023	36	83.7209	8
4	5	11.6279	41	95.3488	4
5	2	4.6512	43	100.00	16

In the first column of Table 2 all possible cluster values (codes of English knowledge levels) are enumerated, in the second column there are numbers of students with corresponding English knowledge level, in the third column, values from the second column are presented in percentage form, the fourth and fifth columns delivers cumulative frequencies of the second and third columns, and the last column contains one (with minimal number) vertex which is a representative of the corresponding cluster. In our example, all vertices are marked with internal network numbers (in order to keep privacy of information).

According to components revealed in Fig.3, one more partition was developed, the partition of organized teams. This partition clusters were determined because of the following reasons:

- necessary number of teams;
- reasonable number of members in each team (different groups might have quite similar numbers of members);
- a student with high level of English knowledge has to be included in every team.

A new representation of the network which reflects their belonging to different teams by means of different colors are shown in Fig.4. The final team assignment slightly differs from that depicted in Fig.4 because the team of 3 members (37, 30, and 29) was too small, that is why 3 other members from “overcrowded” groups were added to it (under the condition of their good will).

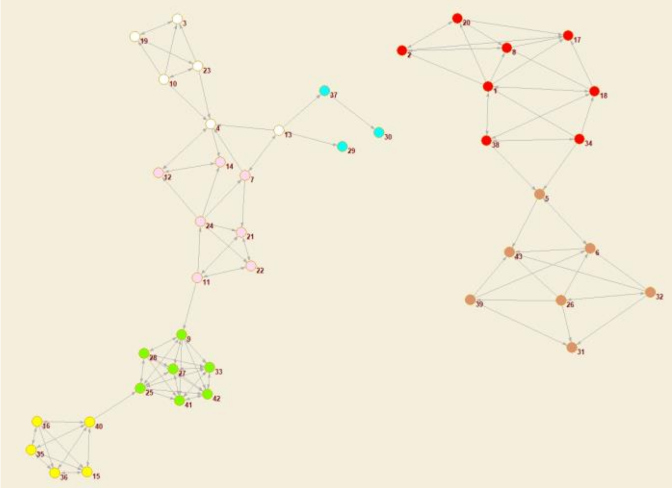

Fig. 4. Network under research with team differentiation Pajek64 (v.5.01)

Pajek proposes a special technology which can support the process of team forming: there is functionality providing shrinking of network in accordance to the partition (command Operations/Network+Partition/Shrink Network in Pajek Main menu), and the opposite action of extracting subnetwork (command Operations/Network+Partition/Extract Subnetwork). In Fig.5 the third team members are extracted while all other components are compressed.

In UCINET there are no functions realizing energy algorithms, but built-in graphical editor NetDraw with the layout “Graph Theoretic Layout” (from menu option Layout in NetDraw main menu) produces similar results (as depicted in Fig.6).

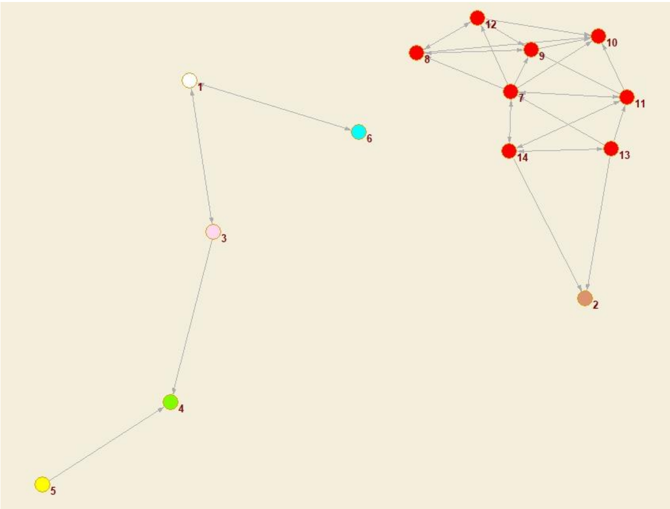

Fig. 5. The third team allocated in the network by use of shrinking command (Pajek64 v.5.01)

Moreover, UCINET proposes one more graphical representation which can help to solve the discussed problem. One can find dendrograms which are not realized in Pajek. A dendrogram can be developed by command

Network/Subgroup/Cliques). A clique is the maximum number of actors who have all possible ties present among themselves.

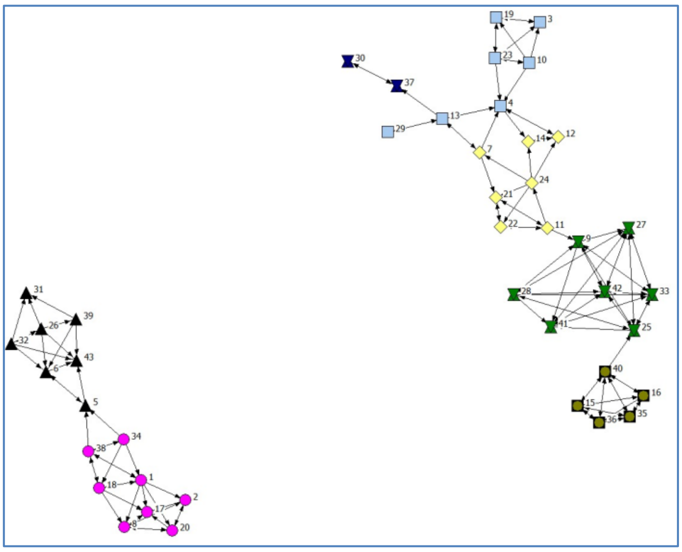

Fig. 6. Network under research with team differentiation (UCINET 6)

Ucinet has found 18 cliques in this network, they are shown in Table 2. Some of the revealed cliques might be considered as teams because they include necessary number of members, and other have to be analyzed more thoroughly to receive the final teams with the necessary number of memebers. To support this analysis, the dendrogram from Fig. 7 may be very useful.

TABLE III. NETWORK CLIQUES

Clique number	Clique members
1	1 2 8 17 20
2	1 8 17 18
3	1 18 34 38
4	3 10 19 23
5	4 7 13
6	4 10 23
7	4 12 14
8	5 6 43
9	5 34 38
10	6 26 39 43
11	6 26 32 43
12	7 21 24
13	9 25 27 28 33 41 42
14	11 21 22 24
15	12 14 24
16	15 16 35 36 40
17	26 31 32
18	26 31 39

Similar to the partition concept adopted in Pajek which was used in this research, in UCINET there is a concept of attributes and a special instrument for their creation – Attribute Editor. A list of all vertices values of the determined attribute is a structure similar to partition in Pajek, and this list of attributes can be converted by UCINET into the Pajek partition format (and, vice versa, from a single partition one can receive an attribute list).

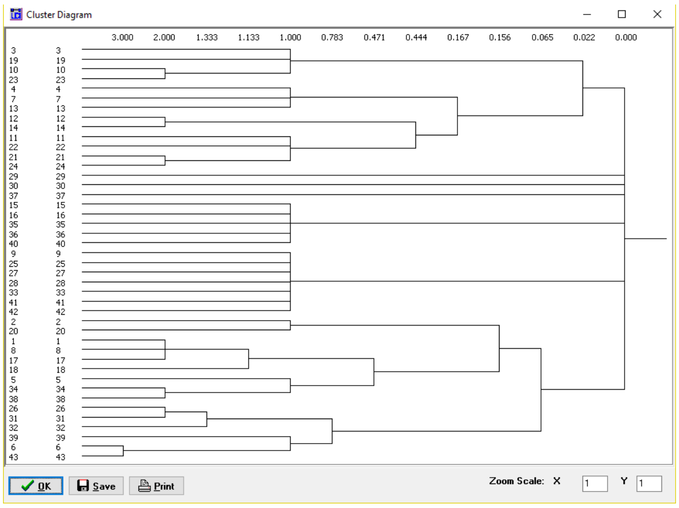

Fig. 7. Dendrogram of the network in UCINET6

V. CONCLUSION

On the basis of the data of student preferences and with the help of two information instruments, seven teams of 5 to 7 members were specified in this research. All educational tasks were done successfully, and this example was proposed for students as a practical example of SNA-methodology implementation. Moreover, comparative analysis of two software products was carried out which was useful as an example how to make a choice of a research instrument.

In the future research, it is planned to study the concept of roles played by team members, and their influence on the team work efficiency.

ACKNOWLEDGMENTS

The author is grateful for the students of the 4th year of education from IRIT-RTF department of Information Technologies and Automation, who with enthusiasm took part in this study.

VI. REFERENCES

[1] A. Doyle, “List of Teamwork Skills”. Available at: https://www.thebalance.com/list-of-teamwork-skills-2063773

[2] S. Wasserman and K. Faust, “Social network Analysis: Methods and Applications”, Cambridge University Press, 1994.

[3] C. N. Phellas, Alice Bloch, and Clive Seale, “Structured methods: interviews, questionnaires and observation”. Chapter 11. Available at: https://us.sagepub.com/sites/default/files/

[4] Wikipedia. The free encyclopedia. Social Network Analysis Software. Available at: https://en.wikipedia.org/wiki/ Social_network_analysis_software.

[5] H. Agrawai, A. Thakur, R. Slathia, Sumangali K., “A comparative Analysis of Social Networking Analysis Tools”, Journal of Information Technolgy & Software Engineering, Vol.5, Issue 3, 2015.

[6] N. Bazenkov, D. Gubanov. Information systems for social networks analysis: a survey**”,** Upravlenie Bol’shimi Sistemami, Vol. 41 (2013), 357–394

[7] S.P. Borgatti, , M.G. Everett, and L.C. Freeman, “Ucinet 6 for Windows: Software for Social Network Analysis”, Harvard, MA: Analytic Technologies”, 2002.

[8] W. de Noy, A. Mrvar, and V. Batagelj, “Explaratory Network Analysis with Pajek”, Cambridge University Press, 2005.