Baltic J. Modern Computing, Vol. 4 (2016), No. 3, pp. 454–465

The Practical Implementation of Graph Analysis to Monitoring Process of Information System Implementation in School

Alexey RASKIN[1], Natalia LEONOVA[1], Oleg GUSTUN[1], Ekaterina DODONOVA[2]

[1] National Research Nuclear University "MEPhI", Moscow
[2] School 777, Moscow

a.a.raskin@gmail.com, nmleonova@mephi.ru, gustun@mail.ru, e.a.dodonova@gmail.com

Abstract. This paper considers the problem of implementation of information system in school. We present some techniques which were used in practice to analyze and manage the process of implementation using graph model of users activities. The first problem which was solved is the problem of monitoring the process of information system implementation. To solve this problem we use clustering coefficients (global and local) and evaluation of their dynamics. The second problem was to find the co-operators who actively use information system and can help administrators to improve the system and involve other co-operators in the process of active use of the system. The results were obtained basing on real data of one of educational complexes in Moscow.

Keywords: graph analysis, centralities, clustering coefficient

1 Introduction

For the last few years the process of schools informatization has been actively conducted. Some time ago the main indicator of progress was considered to be the number of computers installed in the organization, but now managers come to understanding that it is much more important to analyze the changes that occur in the educational organization work. The number of computers means less than the answers to the questions if the productivity of employees work has increased or how many employees are really involved in informatization process.

One more reason for our research is the process of merging schools, kindergartens, centers of additional education into large educational complexes, which is in parallel with the process of schools informatization. That all means people who used to work in absolutely different organizations now become members of the same complex and must work together for the common goals. However, these people are not quite often ready to cooperate indeed.

Therefore we offer the evaluation of how actively interact employees of the organization as one of the indicators of the efficiency of the organization.

There is one more problem in case of informatization of an educational complex. A complex can consist of a few different buildings which can be located quite far from one another. So some colleagues may be interconnected only virtually with no overlap in one room or even building. It is an additional reason for computerization and implementation of information system. Employees can not regularly communicate in person, but they can exchange e-mails, work on shared documents, create collaborative projects in the network, etc.

The managers of school want to receive answers to the question: what is the result of the complex informatization? Have teachers begun to cooperate more actively in solving various problems after the system implementation?

2 Information system description

In this paper we analyze information system of Moscow educational complex number 777 in 2014. The Google Apps system has been deployed in this complex. This system includes all standard web services: mail, online storage, online office, etc.

We have information about all users' actions: how many e-mails were sent and received and the addressees of these e-mails, how many documents were created and edited, who was the author of the edited documents and other information about users activities. There was a full access to all anonymized logs of users actions and this information was used to evaluate interaction between employees.

3 Data description

The information we have consists of logs of 157 employees. The dataset covers the time period from March to May of 2014. It is quite a stable time period not containing long holidays.

We examined more than 25,000 actions: each user made 62 actions on average per month (fig.2) and each user had 7 active contacts among colleagues (fig.1). The distribution is very uneven: some of the employees used the system very actively (basically this is administration and staff responsible for information system implementation), and some others used the system rarely.

The biggest part of users actions is working with documents (sharing permissions, editing, creating). We tried to clear data from personal (not connected with working process) actions.

4 Problems to solve

4.1 Implementation of the information system control

One of the main tasks is to monitor the implementation of information system and evaluate the results of implementation. We want to create artificial criteria as the straight-

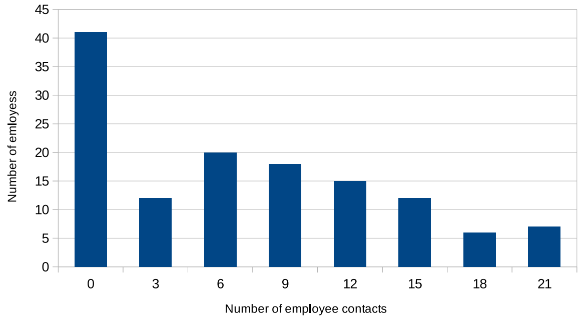

Fig. 1. Distribution of employees according to number of their active contacts

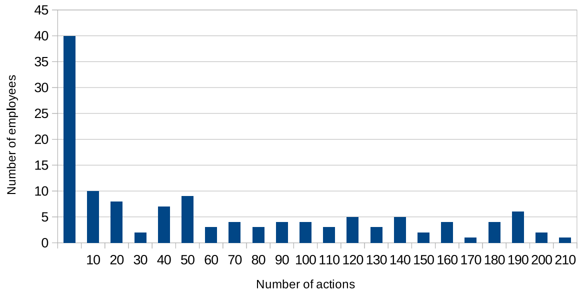

Fig. 2. Distribution of employees activities

forward effect on main school functions (education quality) is significantly delayed, meanwhile efforts need to be evaluated today.

Thus we need quantitative assessment of the effect of the system implementation. The obvious way (the calculation of the number of emails sent per unit of time, for example) cannot give an answer to the main question whether the system has taken root or the whole activity is nothing more than a reaction to the actions of the administration, implementing the system. One idea is to assess how the system is widely used in the individual groups (departments, divisions) without the participation of administrative staff. Thus, for example, the dynamics of the methodical associations and the average overall trend may be assessed.

Managers can always get people to do something but we are interested in the implementation with the increase in effective and motivated work, but not working under the lash.

4.2 Providers detection

One more important goal is to identify people who work with the system, who understand advantages of the system and can be provider of the managers ideas into small groups of employees (providers). They need to be equally involved in the overall network structure of the school and be the centers in their compact groups (departments, methodical associations). This problem is similar to finding bridges (cut-edges) in graph (Tarjan, 1974), but it has some nuances so we can't use finding bridges for problem solving.

5 Graph

The main task was to form a graph. We have no explicit social graph because we have no information about straightforward links between people. We have only information about a users interactions with other ones.

So the implicit social graph (Maayan, 2010) was created based on information about interaction between employees (fig.3). Edges were weighed by the number of actions between two employees (number of e-mails, shared documents etc.)

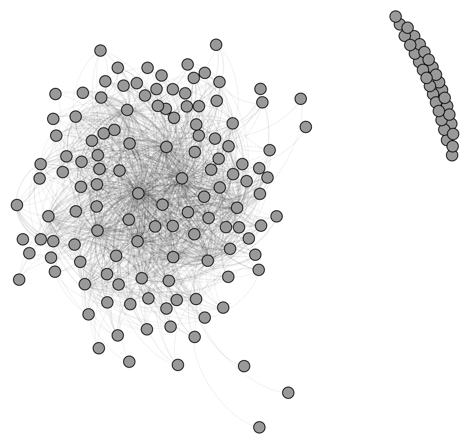

Fig. 3. Implicit social graph of school employees

The created graph consists of 157 nodes (corresponding to the number of employees) with 31 nodes standing alone (users do not use the system at all). There are about 1300 edges in this graph with average degree of the nodes which equals 8.2 (fig.4). Because of 31 nodes with 0 degree the median of nodes degree is higher than mean value and equals 9. Out degree and in degree distributions are quite similar with each other.

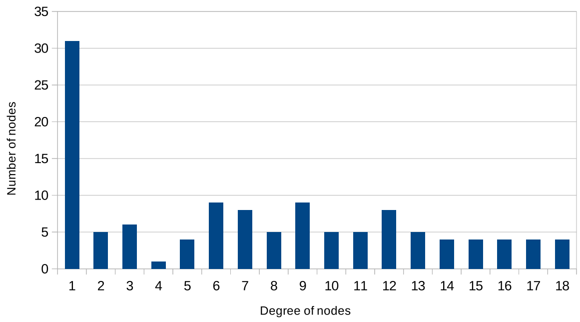

Fig. 4. Distribution of degree of nodes

6 Techniques and algorithms

6.1 Providers detection

Our approach to identify "providers" is based on the assumption that we know in advance the group in which we are looking for providers. In the case of the school complex such groups are methodical associations. We propose to look for those who have a high level of interaction within their group and with respect to colleagues outside this group.

These tools are likely not to introduce significantly new information to supervisor (if the size of the organization is not too big), but it can be useful for tracking changes and monitor the situation. The advantage of this approach can also be considered as the validity of the groups in terms of management of the organization and the intuitiveness of the proposed method to identify providers.

The main problem is to choose the index of interaction. The options are the various centralities or degrees of nodes (Freeman, 1978). It seems that the betweenness and closeness centralities have higher priority (choice of various centralities can be used as a tool to evaluate the different qualities of a member of the organization), but they cannot work well when we have disconnected fragments of graph.

The main idea of closeness centrality is the more central a node is the lower its total distance from all other nodes. This measure is more meaningful to use only with outcome edges for directed graphs, because in general a node has little control over its incoming links.

The farness of a node x is defined as the sum of its distances from all other nodes, and its closeness was defined (Bavelas, 1950) as:

$$C(x) = \frac{1}{\sum_y d(y, x)}$$

.

The main limitation of closeness (Opsahl, 2010) is the lack of applicability to networks with disconnected components: two nodes that belong to different components do not have a finite distance between them. Thus, closeness is generally restricted to nodes within the largest component of a network. That means that if methodical association or departments are divided into fragments the closeness centrality can show results totally inappropriate for the problem solving. For example, see fig.5: the closeness centrality contains a lot of maximum (1) and minimum (0) values and the reason for this is disconnected fragments of graph.

Another option of centrality, which could be discussed, is betweenness, reflecting the number of shortest paths from all vertices to all others that pass through that node (Newman, 2010). A node with high betweenness centrality has a large influence on the transfer of information through the network, under the assumption that item transfer follows the shortest paths. The concept finds wide application, including computer and social networks.

The betweenness of a vertex v in a graph $G := (V, E)$ with V vertices is computed as follows:

1. For each pair of vertices (s, t), the shortest paths between them is computed.
2. For each pair of vertices (s, t), the fraction of shortest paths that goes through the vertex v is determined.
3. This fraction over all pairs of vertices (s, t) is summed.

But the use of such a technique is also problematic for the reasons described above: the number of shortest paths in a lot of small graphs can be very similar and cannot help with the detection of more significant employees in methodical associations

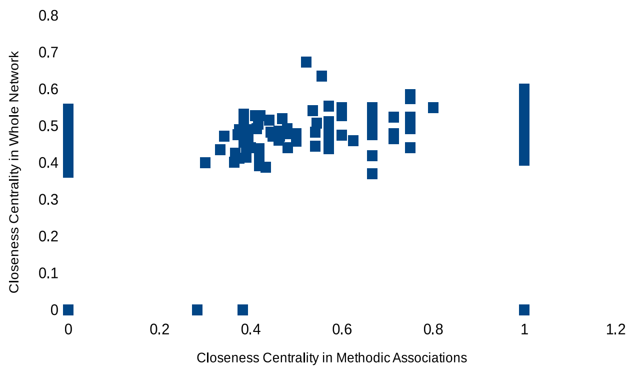

Fig. 5. Correlation between closeness centrality in small groups and in the whole graph

So we settled on a simple but effective method under these conditions. We calculate the degree in two cases: for edges within one structural unit and for all edges in the graph. We felt logical to use their product (as participants with zero degree on one of the parameters must be the last in the list) and need to go to logarithm scale because there are some employees who made many more actions than others.

$$K_{pr} = \log(Degree_{ingroup} * Degree_{wholegraph} + 1) \quad (1)$$

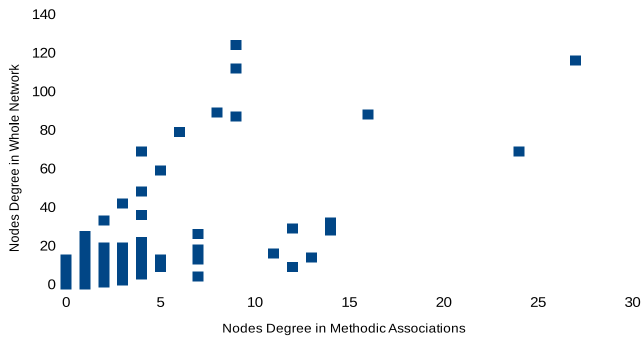

Fig. 6. Correlation between degree of vertex in small groups and in the whole graph

6.2 Implementation control

The implementation control means the evaluation of changes of graph parameters in time. As the assessment of the people involvement in the system we used the clustering coefficient, which is a measure of the degree to which nodes in a graph tend to cluster together. The clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other. In other words, it is the fraction of pairs of A's friends that are connected to each other by edges.

There are two versions of this measure: the global (gives an overall indication of the clustering in the network) and the local (gives an indication of the embeddedness of single nodes) and we use both of them for different reasons.

The local clustering coefficient of a vertex (node) in a graph quantifies how close its neighbors are to being a clique (complete graph). This measure was introduced in (Watts,1998) to determine whether a graph is a small-world network. For our problem this measure can be used to sort all employees in order of their personal involvement in information system. Obviously it is not individual indicator as it depends on behavior of his or her colleagues.

In (Kemper,2009) they present the global clustering coefficient as the average of the local clustering coefficients of all the vertices n in graph:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i. \tag{2}$$

It was important to observe three factors simultaneously:

1. The distribution of clustering indexes across employees (in order to understand the uniformity of employee engagement)
2. A clustering index, taking into account only the interactions within the methodical associations. This index must show the activity of employees without the direct administrative influence.

3. A clustering index, taking into account only collaboration across the organization with the exception of interactions within the methodical associations. This index must show the activity of interaction with the administration.

Distribution (for the first item) and dynamics (the second and third ones) allows to evaluate the development of the system over time.

7 Results of providers detection

The figure below (fig.7) shows the results of the providers detection. The size of the vertex corresponds to how active the user is inside their department and within all staff except their department (according to eq.1).

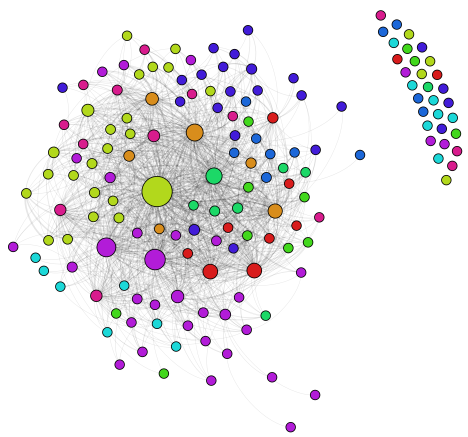

Fig. 7. Implicit social graph of school employees with providers specified by size of vertices

As an example, that the algorithm works, there is an example of employees X (left one on fig.7) and Y (right one on fig.7), which have the same number of connections to the outside, but the employee X is somewhat less active within their department and therefore has a smaller final coefficient.

The result table with final coefficients, degree with edges inside methodical associations (inside degree) and with edges out from it (outside degree) is much more useful

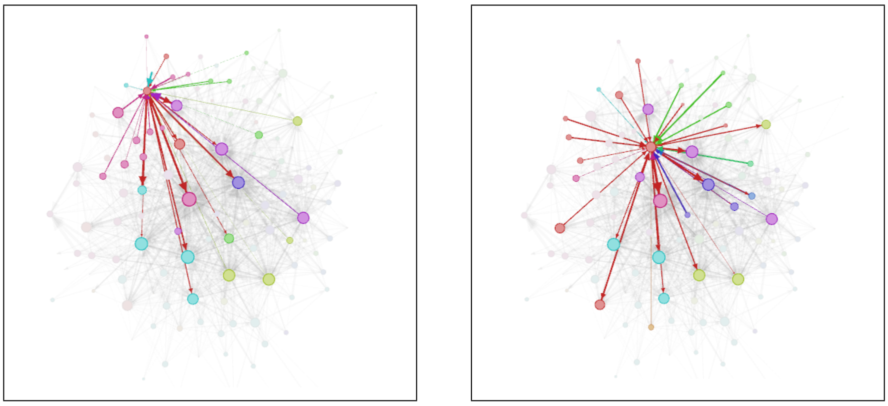

Fig. 8. Example of two employees with different same role in outer part in graph and different role inside methodical associations

from practical point of view. Manager can find (tabl. 1) people with high activity in local groups and in whole net. The user in the third row (SELDEIT14) is less active than in the fourth and fifth rows inside small group, but they are much more open for outer communication and for this reason is more suitable (than the fourth and fifth ones) to be a connector between department and administration of school complex.

Table 1. Degree coefficients for one of departments with similar inside degree for most part of employees

N	User ID	Inside degree	Outside degree	Final coefficient
1	QGASTIT12	9	124	3.048
2	FOFZAIT13	8	89	2.853
3	SELDEIT14	6	79	2.677
4	BIGDEIT23	7	26	2.262
5	DNNKUIT85	7	17	2.079
6	HKSLYIT43	3	12	1.568

Another table demonstrates the situation when the user from the first row looks like the best "provider" despite the fact that they are not the most active in communication with employees from the other part of school. They are the real center of department and the more active they are the more actively their department uses the system. All new document formats, orders, ideas should be discussed with this user and the department should be informed about them via this user to be sure that all people from the department are aware.

Table 2. Degree coefficients for one of departments (the leader has relatively low activity outside the methodical association, but active inside)

N	User ID	Inside degree	Outside degree	Final coefficient
1	DNGDOEN02	12	29	2.543
2	QGUUSEN13	5	59	2.471
3	PAALAEN14	4	69	2.442
4	PEAMIEN15	2	33	1.826
5	KALMAEN16	3	8	1.398
6	MLISAEN17	2	10	1.322
7	NGUMAEN23	2	6	1.114
8	LOIMAEN19	2	6	1.114
9	YOVDEEN96	1	6	0.845
10	UENKOEN32	1	3	0.602
11	REAAKEN81	0	7	0
12	HSAGOEN03	0	0	0
13	JOSSVEN24	0	0	0
14	KEAGOEN28	0	0	0

8 Results of implementation control technique

We analyzed the period from March to May 2014. Three coefficients described below were evaluated for each of the months (tab.8). Also we present the ratio of the coefficients of clustering for interactions within methodical associations (internal coefficient) and for interactions across the whole organization (outer coefficient).

Attention is drawn to the fact that in April the inner coefficient of clustering increased compared to outer coefficient. We believe this is an indicator of the growth of amount of independent actions without the participation of members of the school administration.

In May there was sharp burst of activity associated with the filling of accounting documents (the end of the school year), what resulted in a decrease of the ratio interaction within the teachers and interaction across the whole organization.

Table 3. Clustering coefficients for different types of interactions

Coefficient	March	April	May
Interactions within the methodical associations	0.087	0.097	0.36
Collaboration across the whole organization	0.082	0.095	0.34
Ratio	0.944	0.977	0.944

On the figure below (fig.9) we represent distribution of number of employees according to clustering coefficient. The little progress between March and April and dramatic changes of the situation in May can be noticed here.

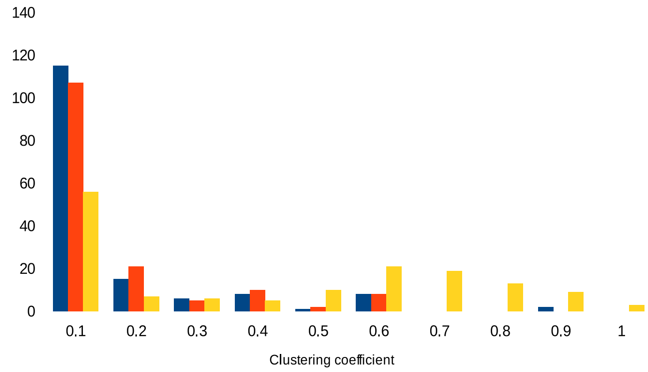

Fig. 9. Different ways to design information about authors

9 Conclusion

We present some techniques which were used in practice to understand the process of implementation of information system in school.

The first problem to solve was monitoring process of information system implementation. To solve this problem clustering coefficients were calculated for the whole system (all employees) and for little groups (methodical association), which can show how system is developing and how many people are involved in it. We have detected administration influence on this process and quantity of employees using the system for their professional purpose without administration pressure.

Another important result is defining people who can provide ideas of managers into methodical associations. We call them providers and find them based on ratio of node degree: taking into account edges in methodical association only and taking into account all edges except them. It helps managers to improve communication between administration and teachers using providers as main points of information distribution.

Acknowledgements

We thank School 777 management and IT department for their assistance in data acquisition and cleaning. All their advices, remarks and comments about the results of research greatly improved the paper.

References

Bavelas, A. (1950). Communication patterns in task-oriented groups. *J. Acoust. Soc. Am*, 22(6):725 – 730, 1950

Freeman C. L. (1978). Centrality in social networks conceptual clarification. *Social Networks*, p. 215, 1978

Kemper, A. (2009). Valuation of Network Effects in Software Markets: A Complex Networks Approach. *Contributions to Management Science*. Springer. p. 142., 2009

Maayan, R., Assaf, B.-D., Deutscher D., Flysher G., Horn I., Leichtberg A., Leiser N., Matias Y., Merom R. (2010). Suggesting Friends Using the Implicit Social Graph. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, USA

Newman, M.E.J. (2010). Networks: An Introduction. Oxford, UK: Oxford University Press, 2010

Opsahl, T., Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, 31 (2), 155-163, 2009

Opsahl, T., Agneessens, F., Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32 (3), 245-251, 2010

Tarjan R. (1974). A note on finding the bridges of a graph journal, *Information Processing Letters*, 2 (6), p. 160 – 161, 1974

Watts D. J., Strogatz S. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393 (6684): 440–442, 1998.

Received April 29, 2016 , accepted April 30, 2016