Project or Process? How to Measure the Real Type of Employees' Activity?

Dmitry Romanov
Higher School of Economics
Moscow, Russian Federation
DRomanov@hse.ru

Pavel Sidorov
Moscow Institute of Physics and Technology
Moscow, Russian Federation
pioneermipt@gmail.com

***Abstract*—By analyzing the logs of corporate e-mail networks we found a number of patterns, showing how the size of ego-networks of individual employees changes on a day by day basis. We proposed a simple model that adequately describes the observed time dependence of an employee's "social circle". Comparison of experimental data with the theoretical model showed that employees are divided into two groups - with fast and slow changes in their social circles, respectively. We believe that the presence of these groups reflects both project-type and process-type of employees' activities. Comparison of data obtained before and during the global economic crisis has shown that the crisis led to an actual reduction in project-type activities.**

Keywords: ego network; social circle; business process; project management.

I. INTRODUCTION

Both the project and business process are two basic types of activities in modern organizations. Business process is characterized by repeatable and reproducible activities, a stable set of members, and stable relationships between them. In contrast, unique tasks, new project teams, innovative characteristics and permanent changes are well known features of project activity. But neither totally project-like nor totally process-like organizations does not exist in reality. Actual activity is always a combination of these two types and sometimes it is difficult to say which management approach (process-centric or project-centric) will be more effective.

The problem of choosing the optimal business model would be easier to solve if we could measure the actual nature of the activities in the organization. In particular, the actual activity of the every employee should have impact on the speed and character of changes in his/her personal network of contacts. We can expect that process-oriented employees have relatively stable networks of contacts, while project-oriented employees' networks of contacts should be quickly updated.

Information exchange networks within social structures with different scales and backgrounds (social networks; informal communities; companies, etc.) is a new object of research being investigated by scientists in various areas of knowledge. However, the dynamic properties of social objects at the micro level are still poorly understood. Perhaps one of the probable causes of the status quo is the lack of adequate experimental data. Traditional sociological surveys and interviews are not the best choice for studying the dynamic properties of the ego-networks. The characteristic time of the study (weeks or months) can substantially exceed the characteristic time of changes in the ego-network. As a result we see a static picture of the network contacts rather than the dynamic process of change. In addition, only information on strong ties is obtained by survey studies of ego networks [1].

Building a network of contacts through the analysis of electronic communication between staff (e-mail, phone calls, instant messages, blogs, etc.) also has its drawbacks. In this case, only part of the communication channels between people is taken into account [2]. An interesting approach based on the use of special sensors that can capture all the informal ("face-to-face") contacts between the employee and his colleagues was proposed in [3]. This works well in small groups, but this approach may be too expensive for large organizations.

A lot of experimental data of communications between people are received from open social networks on the Internet. Another typical object of investigating is the institute itself, which employs the researchers. A no less interesting object of study is an information exchange network between employees of commercial companies because in such social structures typical patterns of communication and behavior may differ from institute or social networks. However this data is usually closed or partially opened. The most famous example of such data is the so-called "Enron Corpus" [4] but this dataset contains information only about the correspondence of 150 employees.

In this paper the authors have set a goal to create a mathematical model to describe dynamical properties of ego-networks and use it to analyze experimental data based on the e-mail correspondence of employees of commercial companies. Results of analysis of experimental data were related with different types of activity – process and project.

The remainder of the paper is organized as follows: Section 2 describes the input array, which was formed from the log files of internal e-mail correspondence. Also in this section, we describe the typical patterns of growth in the observed size of ego networks. Section 3 presents a simple model of the time dependence of the unique input degree of the node in a corporate e-mail network. Section 4 describes the procedure for comparing experimental data with the mathematical model. With this procedure were obtained parameters characterizing the dynamic properties of the ego-network for each employee? Final results of the analysis are presented in Section 5. Finally, Section 6 comprises of our conclusions and future work.

II. EXPERIMENT

We investigated the log of e-mail correspondence between of more than 800 employees of the one Russian IT-company for the two time intervals:

- Interval "IT1" – July 24, 2006 to June 15, 2007
- Interval "IT2" – August 18, 2008 to April 17, 2009

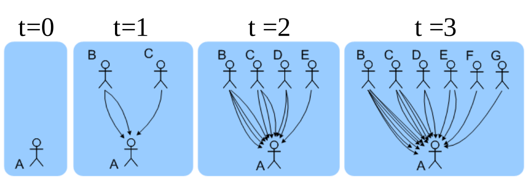
The first time period was prior to the global economic crisis, while the latter was at the beginning of the crisis. The choice was not accidental. Data sets for the two intervals of time were obtained in different economic conditions and gave us a unique chance to compare organizational behavior during very different economic environments.

The initial data contains information about who sent the message, to whom the message was addressed and the date the message was sent. From the test data set weekends and holidays were excluded, because messaging between employees during those days was of a random character. To investigate the behavior of social circles we converted the data set into the following form:

- ID – Code number of person who received message.
- t – Day message was received.
- A – Total number of messages per day.
- S – Number of unique contacts per day.

Fig. 1 shows an example of growth of an ego network. At time t=0, the example employee has no links. At time t=1 (day 1) two contacts (other employees) sent him/her a total of 3 messages. Therefore on this day S=2 and A=3. At the time of t=2 (day 2) this employee has a total of 4 contacts, but 2 of whom have already written. So at the time t=2, S=2, since new contacts were 2 of 4. Let's define $K_a(t)$ and $K_u(t)$ as the time-dependent node degree of all incoming links and by incoming links from unique contacts, respectively. We plotted the dependence $K_u(t)$ for all employees. As was found, some nodes (~25%) demonstrated a good approximation with the power law (Fig. 2). However, for many nodes (~40%) the power law is not the best choice (Fig. 3). Also there are examples of more complicated patterns:

- Acceleration of growth K_u (Fig. 4).
- Reduction of growth K_u (Fig. 5).
- Leaps of growth (Fig. 6).
- And more complicated dependences (Fig. 7).



t=0	t=1	t=2	t=3
$K_a=0$	$K_a=3$	$K_a=11$	$K_a=16$
$K_u=0$	$K_u=2$	$K_u=4$	$K_u=6$
$A=0$	$A=3$	$A=8$	$A=5$
$S=0$	$S=2$	$S=2$	$S=2$

Figure 1. Example of social circle growth

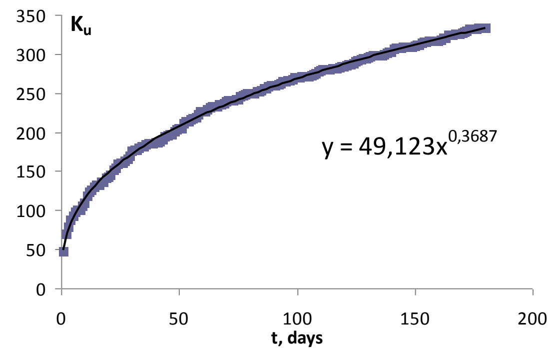


Figure 2. Graph which is well approximated by the power law

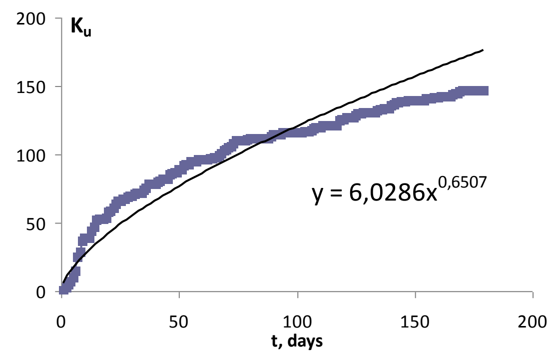


Figure 3. Graph not well approximated by the power law

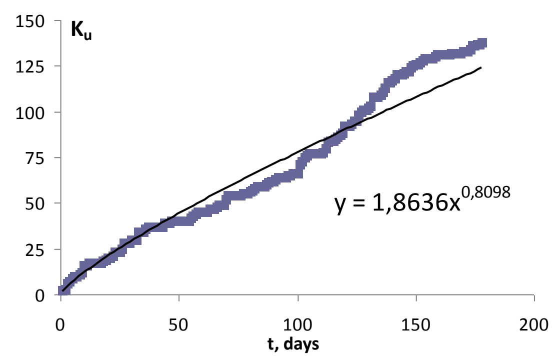


Figure 4. Graph with the a gradual increase in range

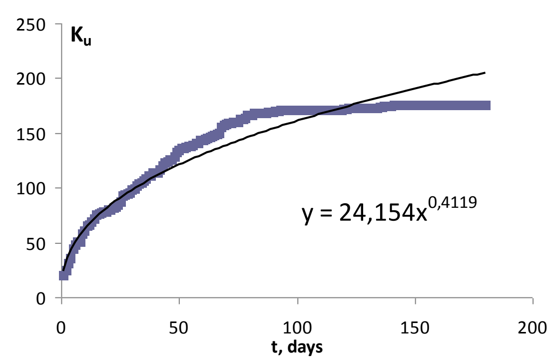


Figure 5. Graph with reduced range

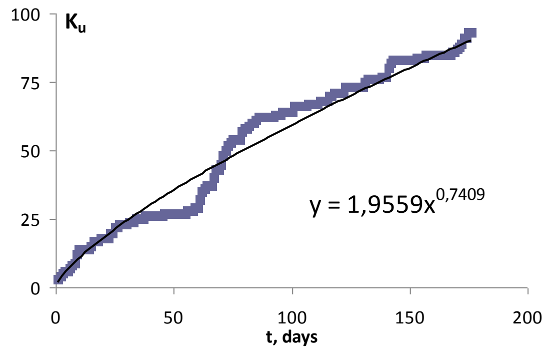


Figure 6. Graph with a leap in growth

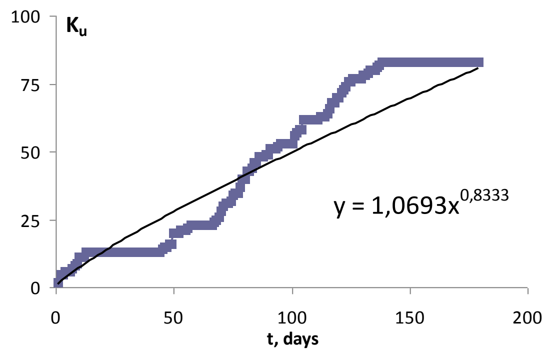


Figure 7. Graph that is not classifiable

Obviously, the power law cannot describe all the features of the real behavior of the quantity K_u. Since K_u behavior reflects the real activity of each person, the dependence of K_u can (and should) be used to understand the behavior of employees, and even the whole organization. Therefore it is important to find a pattern of growth of social circles, which would be able to reflect itself in the observed phenomena.

III. MODEL

A. Simple Model

Consider a simple model of the time dependence of the unique input degree of an employee in a corporate e-mail network. During normal work, employees of the organization share information with their colleagues. Messaging can be implemented within the framework of formalized business processes or in co-operation of staff in project teams or informal collaboration. As a result, each employee has a corporate "social circle" (more precisely, e-mail communication circle) which can be defined as a subset of other employees with which information is exchanged.

As a first approximation of our model, we assume that both the personal social circle and the state of the company remain constant over time. We also take into account preferential attachments and set as basis for the assumption that the probability of receiving a message from a contact was proportional to the amount of messages already received from this contact [5]. So, let the initial time t=0 assume that the employee has not received a single message. Obviously, the probability of receiving a message from each contact from his/her social circle is one and the same amount p_0. Let N be the number of counterparties to the social circle of the employee. Then:

$$p_0 \cdot N = 1 \tag{1}$$

Now consider the time $t \neq 0$. To simplify the equation, replace K_a=a, K_u=u. Let's say at this moment the employee received messages from u contacts. Then:

$$\sum_{i=1}^{u} p_t \cdot a_i + p_t \cdot (N - u) = 1 \tag{2}$$

where a_i is a total number of messages from contact i and

$$\sum_{i=1}^{u} a_i = a \tag{3}$$

Let $\dot{u} = \frac{dK_u}{dt}$ - number of new unique links per time dt. Then:

$$\dot{u} \propto \dot{a}$$

$$\dot{u} \propto \frac{p_t \cdot (N-u)}{\sum_{i=1}^{u} p_t \cdot a_i + p_t \cdot (N-u)} \tag{4}$$

$$\dot{u} = \alpha \cdot a \cdot \frac{p_t \cdot (N-u)}{\sum_{i=1}^{u} p_t \cdot a_i + p_t \cdot (N-u)} \tag{5}$$

Thus we obtain the following equation:

$$u = \frac{\alpha \cdot \dot{a}}{1 + \frac{a}{N-u}} \tag{6}$$

This is the equation describing growth of observed ego networks. Here N – the size of social circle, u – the number of unique contacts, a – the total number of messages and 0<α <1 – the model factor.

B. Model with shift of social circle

Now take into account the effect of a social circle shift. The idea is that contacts within the social circle can replace each other (Fig. 8). This happens for two main reasons:

- Employee turnover in the organization.
- Change in the type of activity of a particular employee.

Since we assume in Model B that social circle size does not depend on time, the total number of contacts remains constant, but their composition can change. For example, if a company employs 1,000 people and every month an average of 30 employees are dismissed or leave and, therefore, 30 new employees are recruited. Let's now suppose the selected employee has a circle size of 100. Then, every month, on average, it has replaced three of its contacts in terms of communication. It is thus obvious that the number of substitutions is proportional to the size of the social circle. To the same effect as staff turnover, also should be added is the associated involvement of the employee in new project team(s) and the termination of old projects. This means that in equation (6), we must add a component proportional to the number of existing unique links. That is

$$\dot{u} = \frac{\alpha \cdot \dot{a}}{1 + \dfrac{a}{N - u}} + \mu \cdot u \qquad (7)$$

where μ – a parameter characterizing the speed shift of the social circle.

To our knowledge, the first attempt to describe the dynamic properties of ego-networks was made in [1], where a mathematical model was built on the assumption, that the rate of loss of contacts will be an inverse function of the strength of the tie. Maybe this statement is correct in some social networks, but in corporate e-mail networks it is not necessarily true. For example, an employee that is involved in particular project has some weak (in terms of time) ties that have almost zero rate of loss, because he needs to communicate with particular people to do his job in the current project and time of life of these ties can be from a few days to a few years. Similarly, the fact that employee was dismissed is not dependent upon how many prior e-mail messages he/she received.

C. *Model with shift and variable size social circle*

Now we should take into account that the size of a social circle may not be constant. For example, rapid growth in the size of social circles is typical for new employees, as they establish new relationships with other employees. Also the change in size of a social circle may be associated with a change in the total number of employees of the organization. We assume that in the first approximation the circle varies linearly with time, i.e.

$$\dot{u} = \frac{\alpha \cdot \dot{a}}{1 + \dfrac{a}{N_0(1+\gamma t) - u}} + \mu \cdot u \qquad (8)$$

where γ – a parameter characterizing the change in size of the social circle and N_0 – initial size of the social circle.

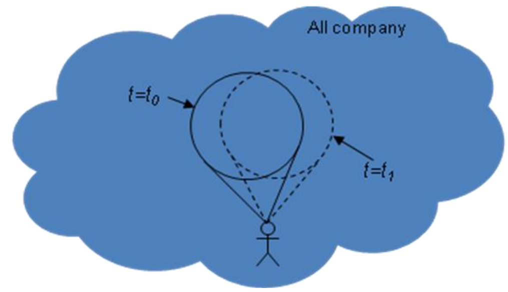


Figure 8. Illustration of social circle shift of an employee

IV. Data Analysis

Comparison of experimental data with the model was carried out as follows:

First, we found parameter γ for each employee. For this purpose, we used the following argument: the average number of unique contacts from which the employee receives messages during the day should be directly proportional to the size of the social circle of the employee. Constructing the integral dependence of this quantity on time and using its approximation as a quadratic function, we may calculate gamma. Secondly, substituting the known values of γ , a , $\dot{a}$ in to equation (8) we received theoretical dependence of the number of unique links over time and compared it with experimental data for K_u. Then we selected the optimal set of parameters α, μ, N_0 for which the discrepancy between the theoretical model and experimental results was minimal. Thus, for each contact we found the best model parameters under the given conditions.

V. Results

As we can be seen in Fig. 9 and Fig. 10, our model is well suited to describe fairly smooth and monotonic curves - those for which at the time of observation changes in nature of a social circle remained the same. However, to describe more complex dependences (abrupt changes in the nature of social circle growth) we split a common time period into smaller time intervals, each of which changes in the social circle were of a monotonous character.

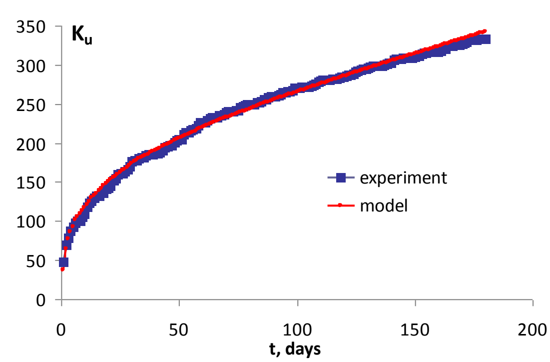


Figure 9. The power law type of K_u is in good agreement with the model

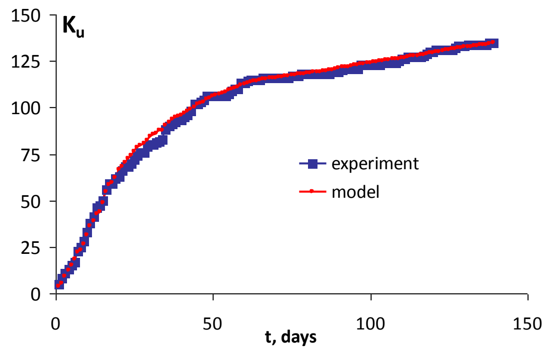


Figure 10. More complicated type of K_u also is in good agretment with the model

Once we got the parameters μ for each employee, we investigated the distribution of employees within this parameter. The results are shown in Fig. 11. We see that this graph has a maximum where the value of μ is near to $6{\cdot}10^{-3}$.

This value of μ corresponds to the natural turnover of employees. This means that the activities of these employees are of process nature - changes in social circles are determined only by the arrival of new staff replacing previous staff.

For the time interval IT1 we can see another peak near to $\mu=9{\cdot}10^{-3}$. The same peak is noticeable on the chart for the time interval IT2, although in this case it is considerably less. This peak represents employees whose circle of contacts is updated much faster than normal staff turnover. Detailed analysis showed that such employees are mostly project managers, sales staff and R&D staff. The work of such employees is the nature of the project - there are a large number of new tasks, contacts and activities.

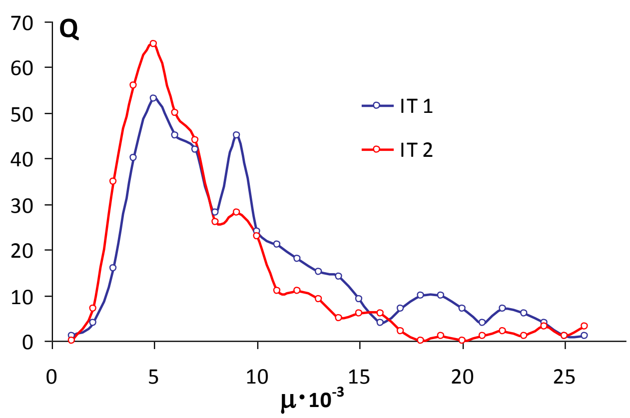


Figure 11. The distribution of the model coefficient μ for the time intervals IT1 and IT2. Here Q is the number of employees with a given value of the parameter μ

Importantly, this approach lets you define the true and fair type activities and not those declared by employees themselves or by position in the organizational hierarchy.

The Fig. 11 shows also that the first peak is slightly lower for the time interval IT1 than IT2. In addition, there are a lot of contacts with parameters μ more than 0.01 in the first interval of time. This means that the actual ratio of employees with project-like types of activities compared to the total number of employees decreased during the crisis.

VI. Conclusion and Future Work

As a result, we can conclude that the proposed model describes well the dynamic properties of ego-networks. In fact, we propose a simple method to measure the actual type of activity for each employee in the organization. It is not a problem to modify the model for more complicated phenomena like leaps of growth or other sharp shifts of social circles. Further improvement of the model should include more accurate classification of the types of dependencies that can be seen in real life and the search for a method that would allow the original curve to split it into pieces that can to be described as typical of those too.

Also of great interest is to test the model on data from other organizations, the study of the dynamics of the social circle not only for internal but also for external communication, as well as study the dynamics of typical patterns of social circles within an open social networks.

Acknowledgment

This work was conducted with financial support from the Government of the Russian Federation (Russian Ministry of Science and Education) under contract № 13.G25.31.0096 on "Creating high-tech production of cross-platform systems for processing unstructured information based on open source software to improve management innovation in companies in modern Russia". We are grateful to Bill Mangino for important comments and invaluable help in preparing this paper for publication.

References

[1] M. McPherson, “A baseline dynamic model for ego networks,” American Behavioral Scientist, May 8, 2009 doi: 10.1177/0002764209331530

[2] F. Grippa, A. Zilli, R. Laubacher, and P. Gloor, “E-mail may not reflect the social network,” in Proc. North Amer. Assoc. Comput. Social Organizational Sci. Conf., Jun. 2006, pp. 1–6.

[3] D. Olguín Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, “Sensible organizations: Technology and methodology for automatically measuring organizational behavior,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 39, no. 1, pp. 43–55, Feb. 2009.

[4] B. Klimt and Y. Yang. “The enron corpus: A new dataset for email classification research.” In ECML, 2004.

[5] Barab´asi, A.-L. and Albert, R., “Emergence of scaling in random networks”, Science 286, pp. 509–512, 1999.