

A Framework for Studying Output data Flow in a Corporate Environment

Versavia Ancusa, Razvan Bogdan and Mihai Udrescu

Computer and Software Engineering Department, Faculty of Automation and Computers,
“Politehnica” University of Timisoara, Romania
versavia.ancusa@cs.upt.ro, razvan.bogdan@cs.upt.ro, mihai.udrescu@cs.upt.ro

Abstract. This paper aims at modeling and simulating a large corporate environment. The proposed model is based on complex networks. The appliance of this model is of great importance in different problems from corporation level. Such a problem is that of data flow (e.g. printing) in such an environment. Different results are obtained and further predictions are offered.

Keywords- complex networks, social interactions, simulator, corporate printing

1. Introduction

Complex networks can be used to describe social networks and more precisely, the interaction in those networks. While the nature of those interactions may vary, the precise mathematical description of such a network makes it pertinent to study them with the help of a computer simulator / emulator.

One of the most common types of social interaction is the interaction in the research environment, the precise nature of such an environment making it a very good studying and testing environment. Based on the study of [9], the social interaction in a research environment has a very definite structure (Figure 1). The scientists are mainly grouped in a distinct area, the administrative personnel in another definite area while the management division also is clustered in a specific region. A couple of links between these areas make the whole organization work together seamlessly. It should be noted that this structure presents a clear correspondence with a pyramidal structure. The links between the departments are the top of the structure and in each clear functional area (research, administration, management) a star topology can be approximated [3]. From a mathematical point of view, complex networks have fractal complexity, but as seen in the study, the research environment structure has a limited depth of maximum three levels.

Extending this model to a corporate environment we can safely assume that the idea of a pyramidal structure still remains viable. The main difference is that there are more than three main departments. For example, in a certain corporation, there is a development department that may work independently on a number of projects. There is also a marketing department, a sales department, a logistics department, and so on. Regarding this corporate structure through the complex network, we are looking at something that resembles graph in Figure 2. Our work hereby presented,

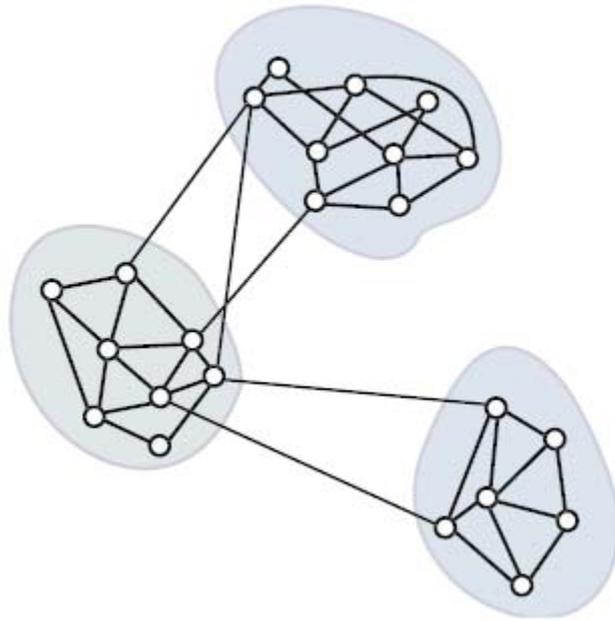


Figure 1. Social interaction in a research environment [1]

is based on this modeling of a corporate environment. Another up-side of the corporate environment and modeling it through complex networks is that the hierarchical structure of a corporate environment has immediate correspondence in the complex network. The difference from the complex network of social interactions and the hierarchical structure issues when between members of different “teams” intermingle. Still, as shown in [1], such interrelations are not very frequent and can be expressed as a probability, when modeling a general environment. All the known formal apparatus can be applied to such a model.

The rest of this paper is structured as follows: section II deals with the particular case of printing in a corporate environment. Following next, section III aims at simulating the proposed network, while section IV presents the simulation results and different discussions of the solution. Further on, different ideas for future work are established.

2. Printing In A Corporate Environment

Let us consider the simple problem of printing in a corporate environment. While there are certain commercial printing solutions (like Google printing), none of these

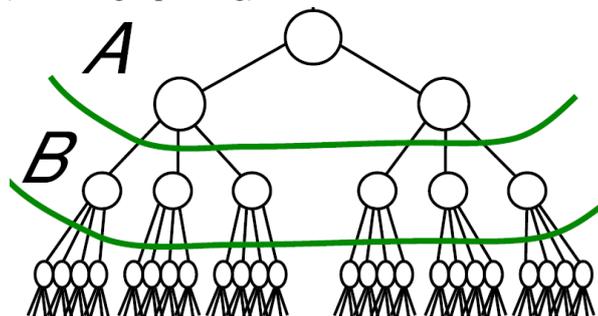


Figure 2. Corporate structure through the complex network [1]

solutions are concerned with the optimum placement of a printer. Modeling the printing network as a complex network, such a problem can be easily addressed. All kinds of cost reduction, power consumption concerns can be analyzed and various scenarios can be run on a specific network. The downside of all this modeling and simulation is the fact that it takes away the essence: the nodes in this network are human beings and human beings have a society-imposed behavior that cannot be thoroughly described by this modeling.

For example, the network in Figure 2 would be cost-efficient if all the printing is done close to the top of the pyramids. In real-life, the top of the pyramid represents some sort of manager, whether it is a senior-

researcher or the CEO of the company. Since the CEO is human, he would be less than sympathetic with a low-rank employee that wants to print some data on the CEO's printer, because the very smart printing algorithm decided that the CEO's printer is the most cost-efficient printing solution. Therefore, in modeling this network we must consider not only the presence of social interaction but also some very subjective factors, that regard the quality and quantity of that interaction, as for example the personal sympathy between humans, the willingness to help etc. Not only these factors are subjective, and consequently very hard to describe in a mathematical model, but also they are subjected to change – they are dynamic factors – on each side of the interaction relationship [6]. Besides the subjective factors influencing the activity we want to study in our network there are some very objective factors, like distance to the printing area or printing capacity of a certain node. To generalize such aspects is extremely difficult, and to model them dynamically is even more so.

3. Network Simulation

In order to simulate a network like the one presented in Figure 2, we used AweSIM [2] and modeled each node as a composition between an entities generator (representing print jobs originated from that node) that communicates with a queue of print jobs. In this arrangement in order to determine where would be the most beneficial place to position a printer, we considered the case of a printer near each node and through the print-jobs flux in that printer, we'll determine the load for each node. The higher the load, the most beneficial a printer would be in that position. The queue is a mock-up for a printer and accepts data not only from the generator but also from all the subordinate nodes that communicate directly with this node. If the queue is full than the node will forward all the unprocessed messages to the node's superior.

This composition corresponds to the hierarchical structure in a corporation. In order to simulate the inter-departmental interaction we added the possibility that the generator of print jobs can send some entities to other nodes (most of them being same-level, same-team nodes), though the probability to do so should be smaller than the probability to communicate with its own queue.

That time the entity (print-job) spends from the generator to queue will reflect the objective factors (like distance, cost, etc) and the length of the queue will reflect the subjective factors (like willingness to help, friendliness, etc).

As stated before, the selected pyramidal structure can be approximated with a composition of imbricate star topologies. This means that we have a simple star topology for all the low level employees involved in the same sub-project and their mid-level superior, but also that all the mid-level personnel involved in the same project and their upper-level management can be represented with a star topology. Only the highest level of management is not star connected. In other words, the configuration of a corporation (Figure 2) can be reduced to an interconnection of the centers of star topologies. This reduction helps when simulating large organizations, as the data can be aggregated in department-specific data, the next step being the interconnection of departmental data. In simulation, this helps to maintain the generality of the model.

For the first simulation, all the employees are identically modeled. They generate printing jobs with a frequency determined by an exponential distribution and they may initiate more than one printing job at a time. The number of emanating activities per node is modeled also with an exponential distribution.

We chose to model two interacting teams. Each team is connected in a star, and in each team, we have a "popular" member to which the other team addresses printing jobs. The members of each team will process as many entities they can, and all the overflow will be sent to their leaders. The boss of team 1 (not necessarily the boss, but the central communication node) will address any overflowing print jobs to the boss of team 2. Team 2's leader will dispatch any overflowing communication to a remote node called "lost", that will collect all the unable to perform in real (useful) time jobs.

In order to model as realistically as possible the objective factors, such as distance to the printer, or the printer's quality and capacity, we used general architectural design rules.

We established a random reference time (for example 5 ms) for the lowest communication time from an initiator node to a printer. All resulting data are expressed in relative time (that is the simulation time divided with the reference time) in order to generalize the results.

In Figure 3, we plotted the number of entities vs. the relative initiation time for print-jobs. The dashed line represents the overflowed entities. For print jobs that appear

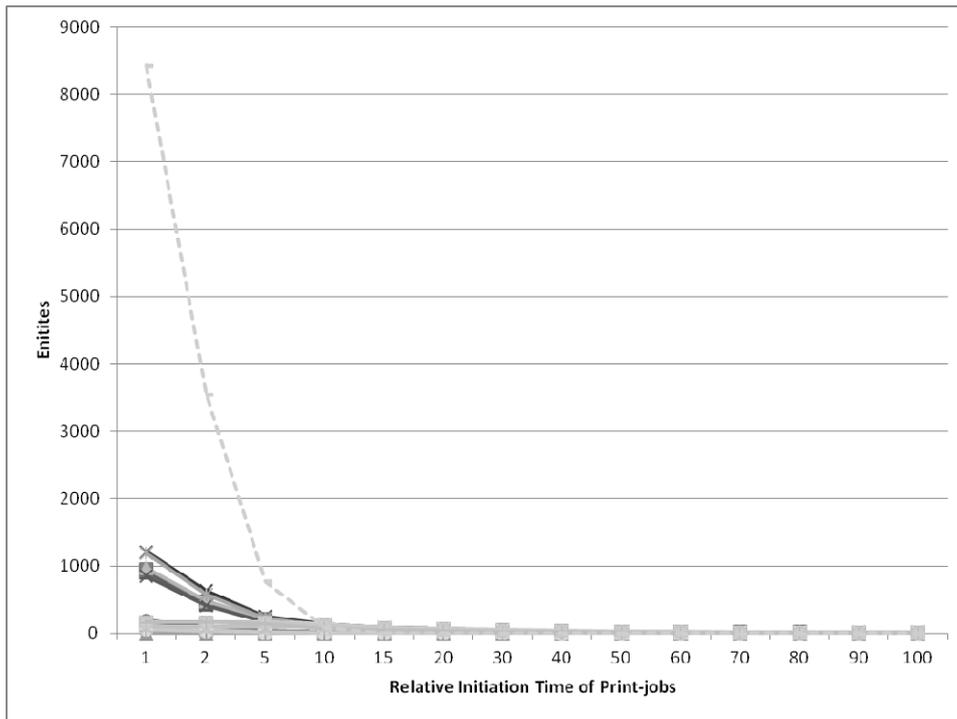


Figure 3. Number of entities vs. the relative initiation time for print-jobs

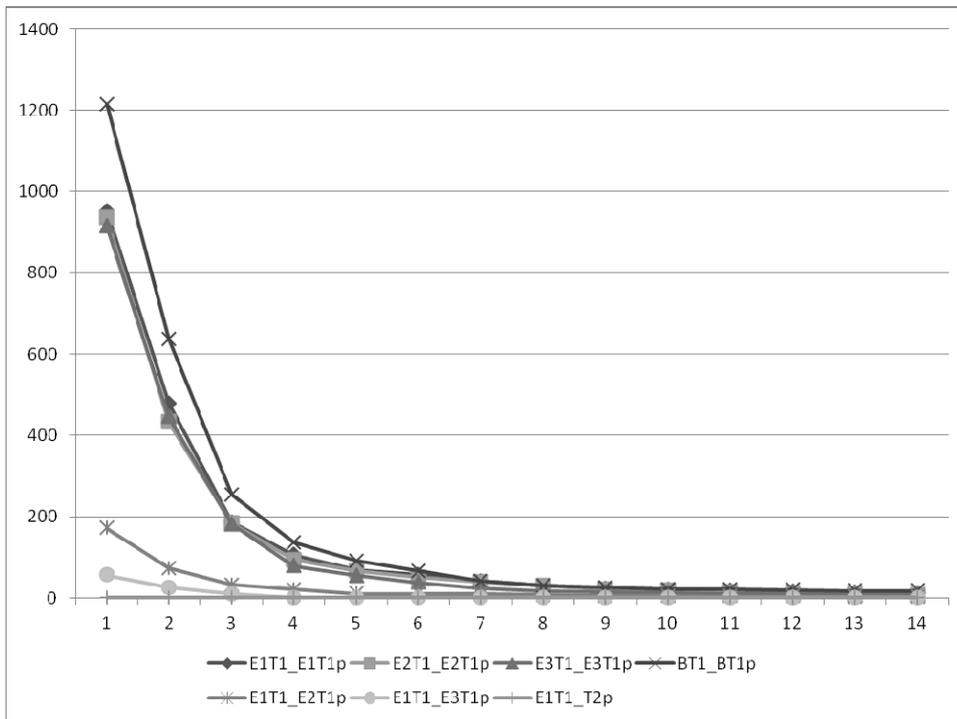


Figure 4. Entities that pass through certain lines

with a frequency ten times higher than the minimum required time to rich a printer, there are unserviceable entities. The number of unserviceable entities is smaller than the number of serviceable entities for values of the initiation time larger than four times the minimum required time to rich a printer.

In Figure 4, we plotted the entities that pass through certain lines:

- E1T1_E1T1p/ E2T1_E2T1p/ E3T1_E3T1p – the data sent by employee 1/2/3 from team 1 to his private printer (in other words – the direct data flow of that employee);

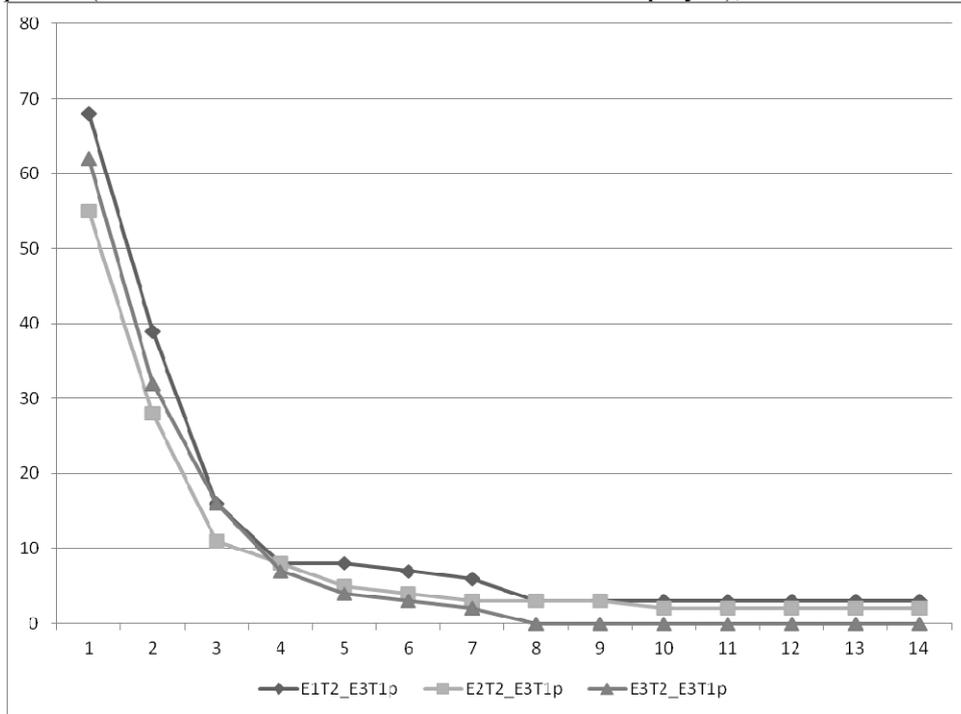


Figure 5. Team's 2 outside indirect data flow towards team 1

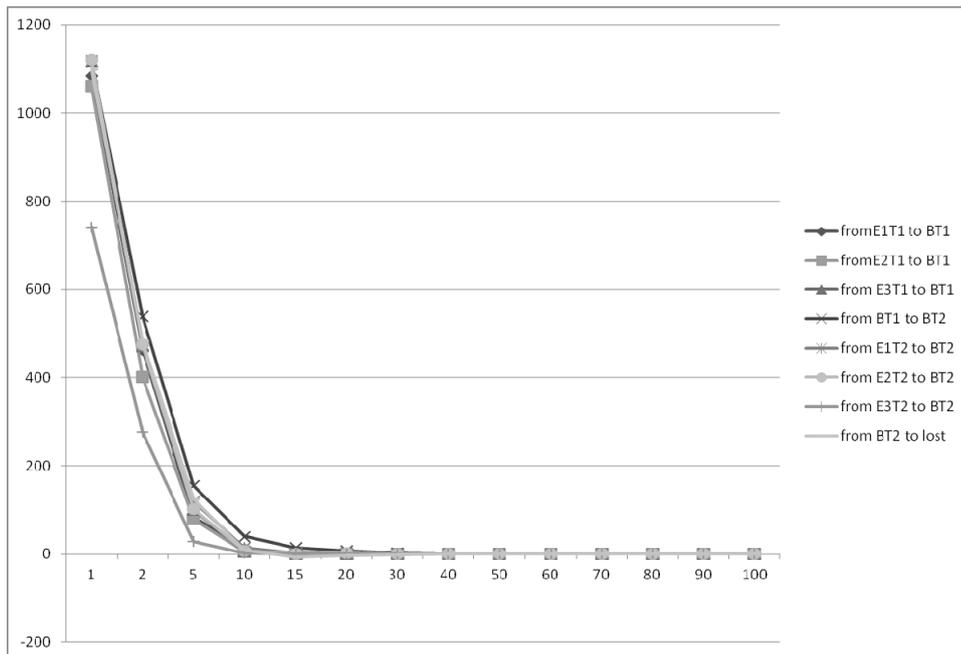


Figure 6. The number of overflowing entities

- BT1_BT1p – the direct data flow generated by the team 1's main communication node (boss);
- E1T1_E2T1p / E1T1_E3T1p - the data sent by employee 1 from team 1 to employee's 2/3 printer (in other words – the indirect data flow within team 1 of employee 1);
- E1T1_E2T1p / E1T1_E3T1p – the data sent by employee 1 from team 1 to employee's 2/3 printer (in other words – the indirect data flow within team 1 of employee 1);
- E1T1_T2p – the data sent by employee 1 from team 1 to the closest team2 printer (in other words – the indirect data flow outside team 1 of employee 1).

As expected, the direct data flow has the same profile for all the data flow of equal-level employees. The upper-level direct data flow has the same profile, though at a slightly higher rate, while the indirect data flow within the team decreases proportionally with the connection degree between nodes.

An interesting result regards the indirect data flow outside team 1 of all the low-level members of team 1. In our simulation we did not record an value for that flow. However, team's 2 outside indirect data flow towards team 1 is plotted in Figure 5. As seen, although the members of team 1 and team 2 have the same personal connection matrixes only team 2 send data to team 1. The amount of sent data is not very high, but rises with the severity of congestion.

The last figure presents the number of overflowing entities, and as we can see, the entire overflow seems to follow the same pattern and amplitude. This reflects the homogenous nature of our network. In a heterogeneous network, we may safely expect a different result.

4. Conclusions And Future Work

This paper presents a framework, which makes available an opportunity to do corporate data flow studies that can be applied to a specific type of data flow and/or organization

The simulation is based on homogenous nodes, thus an immediate development is the transformation in a heterogeneous network. For that, the initiator nodes will have different generation times, maybe even different types of probabilistically distributions to generate the numbers. Not only that, but in the future we expect to implement queues that will have different lengths, and the transmission of data between humans to have variable probabilities. The problem in modeling all this is choosing realistically values, thus some experimentation within our research facility social network is to be approached.

5. References

- [1] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, Michael W. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters," *Internet Mathematics*, vol. 6, no. 1, pp. 29-123, 2009.
- [2] Alan B. Pritsker, Jean J. O'Reilly, "AWESIM: the integrated simulation system." *Proc. of the 30th conference on Winter simulation (WSC '98)*, D. J. Medeiros, Edward F. Watson, John S. Carson, and Mani S. Manivannan (Eds.), IEEE Computer Society Press, Los Alamitos, CA, USA, 249-256, 1998.
- [3] Juyong Park, M. E. J. Newman, "Solution of the 2-star model of a network," *Physical Review E*, vol. 70, Issue 6, id. 066146, 2004.
- [4] Brian Ball, Brian Karrer, M. E. J. Newman, "An efficient and principled method for detecting communities in networks," *Physical Review E*, vol. 84, id. 036103, 2011.
- [5] S.H. Strogatz, "Exploring complex networks," *Nature*, no. 410, pp. 268-276, 2011.
- [6] S. Boccaletti, V. Latorab, Y. Morenod, M. Chavezf, D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, Issues 4-5, pp. 175-308, February 2006.
- [7] M. E. J. Newman, "Networks: An Introduction," Oxford University Press, ISBN13: 9780199206650, March 2010.
- [8] Ernst Neufert, Peter Neufert, Bousmaha Baiche, Nicholas Walliman, "Architects' Data," Wiley-Blackwell, ISBN: 978-0-632-05771-9, July 2002.
- [9] M. Girvan, M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci.*, 2002.