Modification Propagation in Complex Networks

Mary Luz Mouronte, María Luisa Vargas, Luis Gregorio Moyano,
Francisco Javier García Algarra, and Luis Salvador Del Pozo

Telefónica I+D, Madrid, Spain
mlml@tid.es
http://www.tid.es

Abstract. To keep up with rapidly changing conditions, business systems and their associated networks are growing increasingly intricate as never before. By doing this, network management and operation costs not only rise, but are difficult even to measure. This fact must be regarded as a major constraint to system optimization initiatives, as well as a setback to derived economic benefits. In this work we introduce a simple model in order to estimate the relative cost associated to modification propagation in complex architectures. Our model can be used to anticipate costs caused by network evolution, as well as for planning and evaluating future architecture development while providing benefit optimization.

Keywords: Complex system, complex networks, structures and organization in complex systems.

1 Introduction

In recent years, it has become very clear to large business organizations that their systems and networks are evolving increasingly complex. The amount of different components in a system and the various ways these elements may interconnect each other result in quite intricate network topologies. Moreover, these complex structures are seldom static, but quite the opposite: they grow and change, sometimes in a permanent basis, as different requirements and functionalities develop to satisfy other system or client needs. Meanwhile, the study of complex networks—a growing field of scientific research—is becoming more and more important as a tool to tackle new needs that derive from emergence phenomena in complex structures. There are many such systems which are relevant to illustrate this growing focus, for example, communication networks, transportation networks, computer networks, social networks and several others [1,2,3]. All these are examples of a large number of basic similar units connected in a non-trivial manner and that may be well characterized by the concepts provided by complexity science. This characterization is used to explain certain properties that in turn will lead to new, improved ways to interact with such networks.

An interesting question arises when dealing with complex systems, and refers to how a modification in one point of the network propagates through the rest

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 435–440, 2009.

of the interdependent units of the network. In the present (ongoing) work [4,5], we focus in studying the cost associated to this type of modification propagation and the relation with the network's intrinsic complexity. More specifically, we construct a simple model to describe costs associated to required software or hardware modifications (for example, as a consequence of network update or service improvement) in a complex network, as coulbe an operation support system architecture (i.e., a structure composed by software and hardware elements designed to fulfil predetermined tasks or processes, as service provision, requests management, alarm handling, inventory updates, among others). Finally, we analyze the influence of two network parameters (the average shortest distance L and the degree distribution exponent γ) in the cost of such modifications.

The costs we have in mind are those unwanted but inherent costs that arise when introducing changes in interconnected systems. The actual economic cost of implementing the modification is not considered in our model, which only tries to capture the "hidden" cost resulting from the complexity of the architecture substrate, and which is not established *a priori*. These costs can be anything which might be considered a liability, e.g. cost in economical expenditure, cost in time, cost in resources, cost in business management, and others.

2 Model

Our aim is to develop a simple model that successfully captures the observed features and costs of modification propagation due to the topological aspects of the network [6]. There are many reasons for software or hardware modifications in an architecture element, e.g., correction of defects, improvement of performance (or other related properties), adaptation to environment changes (technical updates, strategy or business priority modifications, regulatory changes), novelties in market industry, etc. Any modification of this nature has a certain impact over the rest of the components of the architecture. We aim to quantify this impact globally, particulary indirect effects that in principle are not obvious to detect. Our model will improve initial cost estimates of any modification, allowing for further analysis by implementing other relevant (but possibly secondary) features (i.e., network topology quantities, as clustering coefficient, diameter, etc.) which will improve the accuracy of those initial cost estimates.

In the present work, the impact or influence of modifications is modeled through *impact coefficients*, which to our purposes represent node relationship, i.e., the interaction of interfaces in a given system. In this way, we define a matrix K where an element K_{ij} is the coefficient describing the cost impact over node j of any modification done in node i. In the resulting matrix, links that do not exist have zero impact as well as the principal diagonal, meaning that self-impact is discarded. We consider the values of K_{ij} to be stochastic because they present fluctuations that depend on a very large number of variables. In our model, we consider these coefficients to have a well-defined statistics, in the sense that the stochastic variables involved have a precise origin (e.g. a Poissonian "arrival" of failures) and they can thus be correctly described by a given statistical distribution (more on this below). Moreover, these quantities usually vary in time, as nodes and links dynamically change inside the architecture, but for our initial purposes we consider them as stationary. To choose a suitable distribution, we take into account empirical evidence on how these coefficients should be. We consider three main characteristics: firstly, coefficients should be small (measured in arbitrary units, and compared to unity). Secondly, they must be positive (we assume the probability of a modification propagation to lower costs to be negligible). Finally, we must account also the possibility to have few, unusually high cost events (for example, when a difficult but important request from a client must be introduced) which will be described by an exponentially-tailed distribution.

Regarding the actual causes that generate this impact, we can follow basically two: firstly, costs derived from changes in functionalities or improvements in a particular node, and secondly, costs associated to defects involuntarily introduced in any node which propagate through the architecture. If we suppose these two factors to be uncorrelated, and also consider that they are well described by Gaussian probability distributions, then to account for both simultaneously we can make use of the probability distribution of the modulus, which follows a Rayleigh distribution. The Rayleigh distribution, for sufficiently low values of its variance σ^2, has the three needed characteristics we previously mentioned. Note that this is not the only choice one could make. In particular, there could be more causes of cost propagation, not taken into account here. In this case, we would have to use other (related) distributions: the Maxwell distribution (for an additional third variable), or a Chi distribution (if N variables would be needed). Indeed, we are currently working in determining the influence that a particular distribution has in the results, which in turn may lead to conclusions about the task of providing a useful characterization of the stochastic nature of modification propagation costs. To estimate the magnitude of the variance in such a way to get realistic values of impact coefficients, we choose the standard deviation in such a way that 90% of the coupling coefficients are equal or less than 0.05, i.e., one estimates the appropriate σ for which any coupling constant is less or equal than 0.05 with a 0.9 certainty probability. This is achieved with a standard deviation $\sigma = 0.04$.

To calculate the cost of a propagation of a modification done in node i, we perform the sum of all impact coefficients, taking into account all possible further ramifications of "child" nodes. We model this ramification process through the following (iterative) equation:

$$C_i = \sum_{j=1}^{J_i} K_{ij}(1 + C_j) \tag{1}$$

where J_i stands for the number of child nodes in node i. When computing subsequent C_j terms, and in order to avoid loops, the i^{th} row and column from matrix K are removed, i.e., if one calculates what is the cost on node B due to a modification on node A, afterwards there is no contribution of the impact over node A product of a change in node B. Then, the calculation proceeds on the resulting matrices, until the last nodes are reached.

Our model also takes into account the fact that some nodes suffer modifications, while others do not. This means that at any moment, there is some probability p_i for node i to be modified. To model this situation we consider two different aspects. On one hand, all nodes are taken to have a (small) probability to be modified, so there is a small but positive contribution common to all nodes. In our case, we have assigned this probability to 1% (normalized by the total number of nodes). On the other hand, important nodes (i.e., "hubs") are considered to have a larger probability of requiring some modification, which reflects the need to attend an increasing demand, or to better adapt to the child nodes it already has. Therefore, we take into account another contribution to the probability to be modified that will be proportional to a power of the number of connections k that the node has, which in our case is set to $\frac{1}{2}$. Therefore, the probability to change is

$$p_i = c\sqrt{k_i} + 0.01\,\frac{1}{N} \tag{2}$$

where p_i is the probability that must be applied to node i , k_i the number of links of node i, and N the total number of nodes in the architecture. This probability must satisfy that the sum of all nodes equals unity, so

$$c = 0.09\,(\sum_{i=1}^{N} \sqrt{k_i})^{-1}. \tag{3}$$

The average total cost due to modification propagation will be evaluated as

$$C = \sum_{i=1}^{N} p_i\, C_i\,. \tag{4}$$

3 Results

To estimate how costs depend on the complexity that a substrate architecture or network exhibits, we generated a set artificial scale-free networks [7] (close to 200 randomly generated networks). These networks have a fixed number of nodes and links ($N = 110, N_l = 270$), but variable topological parameters, such as the (scale-free) degree distribution exponent γ, and average shortest length L (i.e., the average minimum number of links needed to travel between two nodes). Our results were calculated with K_{ij} coefficients taken from a Rayleigh distribution of $\sigma = 0.04$ to describe each node pair. Then, for each artificially generated test network, we calculated the associated cost (applying Eq. 4) and directly observed the relationship between cost and structure. Our preliminary results show that there is a strong influence from the topology of the network on our cost model. In Fig. 1 we show the modification propagation cost as a function of these two complexity parameters, namely, the degree distribution exponent γ (right) and the average shortest length L (left). As we can see from the figure, our results

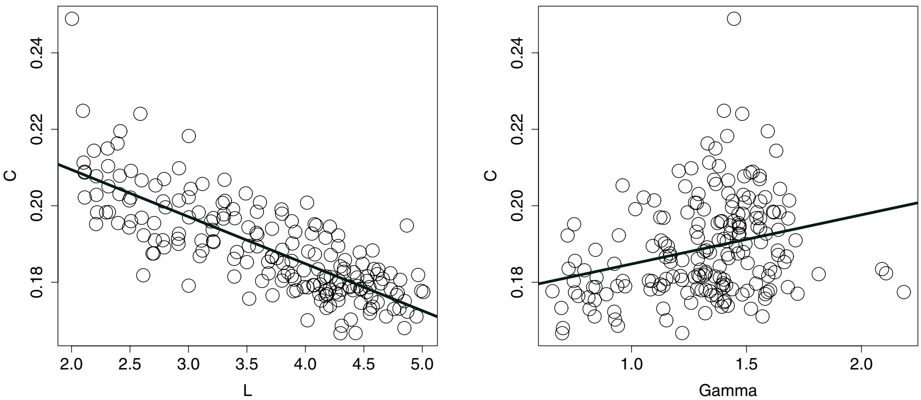

Fig. 1. Left: Propagation cost C as a function of the average shortest length of the network L. Right: Propagation cost C as a function of the degree distribution exponent γ. In solid line, a linear fitting. Both quantities present an overall relative cost change of approximately 20%, with negative slope in the case of the average shortest length L, and with positive slope regarding the γ dependence, in the ranges shown in the figures.

show that, in average, costs decrease as the average shortest length increases. and increases with a higher heterogeneity of the networks. In order to have a general estimative of the relative change, we performed linear fits in our results. In both cases, the absolute change in costs (associated with the chosen range of L and γ) is of about 20%, being the slopes $m_\gamma = 0.012(3)$ and $m_L = 0.0123(7)$. Clearly, a linear estimate is more suit for the lenght dependence than for the γ dependence, where more analysis is needed to determine its influence. Such analysis is currently being addressed.

4 Conclusions

Adapting an organization's business processes to changes in the environment is crucial for maintenance tasks, optimal network evolution, among many other fundamental business activities. The considerable cost of these tasks relates directly to the topology of the operating architecture, i.e., the system designed to carry the organization's processes. In this work, we introduce a simple model to estimate the cost of such modifications in a complex system architecture. Our study is aimed at estimating the unwanted cost of a software modification on an architecture before its actual implementation, by measuring the existing system topology. Furthermore, we plan to explore the possibility to identify the point at which architectures are no longer economically profitable, as a consequence of too high modification costs. Our preliminary results suggest that network complexity (in the sense of scale-free characteristics) implies higher economic expenses when changes need to be made, but larger systems actually decrease this cost. The correct estimation of propagation of costs due to inherent complexity in complex network systems is of both theoretical and practical interest and should be considered as an interesting and important problem.

References

1. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167 (2003)
2. Watts, D.J., Strogatz, S.H.: Collective dynamics of "small-world" networks. Nature 393, 440 (1998)
3. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Review of Modern Physics 74, 47 (2002)
4. Benito Zafrilla, R.M., Cárdenas Villalobos, J.P., Mouronte López, M.L.: Redes Complejas: El nuevo paradigma. Sociedad de la Información, Tecnología e Información Bulletin (Fundación Telefónica), Madrid (2007)
5. Mouronte, M.L., Armas, A.: Análisis de la Complejidad del Mapa de Sistemas de Telefónica de España, Sociedad de la Información, Tecnología e Información Bulletin (Fundación Telefónica), Madrid (2006)
6. Pastor-Satorras, R., Vespignani, A.: Epidemic dynamics in finite size scale-free networks. Phys. Rev. E. 65, 035108 (2002)
7. Catanzaro, M., Boguña, M., Pastor-Satorras, R.: Generation of uncorrelated random scale-free networks. Phys. Rev. E. 71, 027103 (2005)