Project Number 001907

DELIS

Dynamically Evolving, Large-scale Information Systems

Integrated Project

Member of the FET Proactive Initiative **Complex Systems**

Deliverable D5.6.2

Development of an integrated package for network evolutionary dynamics

Start date of the project:	January 2004
Duration:	48 months
Project Coordinator:	Prof. Dr. math. Friedhelm Meyer auf der Heide Heinz Nixdorf Institute, University of Paderborn, Germany
Due date of deliverable:	December 2006
Actual submission date:	January 2007
Dissemination level:	PU – public
Work Package 5.6:	The Structure of Tinkered Landscapes
Participants:	Universita di Bologna (UniBO), Italy Universitá Pompeu Fabra, Barcelona (UPF), Spain
Authors of deliverable:	Sergi Valverde (svalverde@imim.es) Ricard V. Solé (ricard.sole@upf.edu) David Hales (dave@davidhales.com)

1 Introduction

An important aspect of our research within DELIS involves the exploration of how computational or information processing networks can evolve through time. More fundamentally, there is a basic question concerning the nature of the process itself of finding better solutions and how innovations emerge. Innovations need to be properly differentiated from improvements. An improvement implies slight changes from previous designs, whereas a true innovation is associated to a major novelty in design. Previous work within this SP5 involved the exploration of simple feedforward networks or equivalent digital circuits obtained from evolutionary dynamics. Such work revealed something surprising: the underlying landscapes are surprisingly similar to those known within evolutionary biology. In particular, the landscape of feedforward networks is closely replated to the statistical behavior of RNA folding networks [2]. Such similarities indicate that some fundamental features of the evolutionary dynamics of both natural and artificial networks might be at work.

Theoretical and computational methods can be used to study the fundamental properties of evolutionary processes. For example, computer simulations have provided evidence of universal features in the evolution of both artificial and natural systems [1]. In these experiments, populations of self-replicating computer programs (i.e., digital "life") mutate and evolve in a fitness landscape towards maximizing the likelihood of survival. If the principle of universality holds, then studying these artificial organisms will be somehow equivalent to study the natural evolutionary processes with several additional advantages. Indeed, studies in digital organisms have successfully reproduced the growth of complexity in evolving systems, patterns of epistatic interactions and quasi-species dynamics [1].

However, the above systems do not scale up to complex evolutionary phenomena. For example, these systems can readily simulate biochemical viruses and bacteria but they are less adequate for studying the evolution of multi-cellular organisms spanning several layers of complexity. Our intention is to overcome these limitations and develop a powerful integrated package for multi-scale evolutionary dynamics. There are, however, a number of questions that we should address first prior to the building of the package. These questions include the definition of proper measures that characterize complex evolutionary processes and, ultimately, a full evolutionary theory or framework providing the right context for the package.

Beyond the specific choices made in previous work within SP5 concerning the structure of the evolutionary paths followed by computational networks, there is mounting evidence that the landscapes of innovation might display universal properties. In this direction, we have performed a large-scale analysis of a network describing and linking technological innovations, trying to gain further understanding on the nature of universal patterns of technological change. This network offers a unique opportunity to prove our intuitions concerning the presence of common organizational laws shared by biological and man-made systems.

2 Patent citation network and patterns of technological Innovation

In addition to computational methods, the evolution of technological systems offer a privileged window into general principles of evolutionary theory [4]. For example, innovation is one important and poorly understood aspect of evolution that also takes place in technology. New entities often result from the combination of predefined designs or building blocks and, sometimes, a completely new solution emerges. The evolution of patents is a surrogate of the ways in which innovations takes place. Patents are well-defined objects introducing a novel design, method or solution for a given problem. Additionally, the list of references cited by a patent indicate what previous novelties have been required to build the new one. This information enables us to reconstruct the full evolutionary history of any given patent.

Previous studies [6] have measured the value of any innovation by means of the analysis of patent

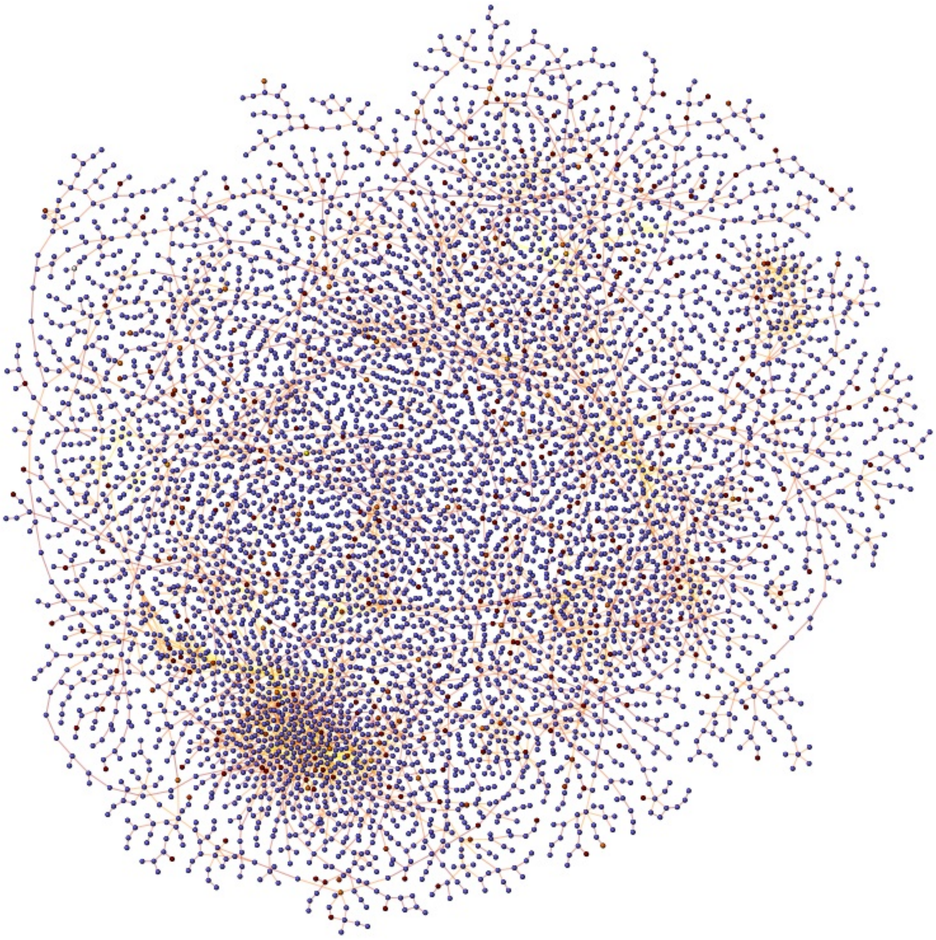

Figure 1: Patent citation network for the 7000 first patents in the USPTO database.

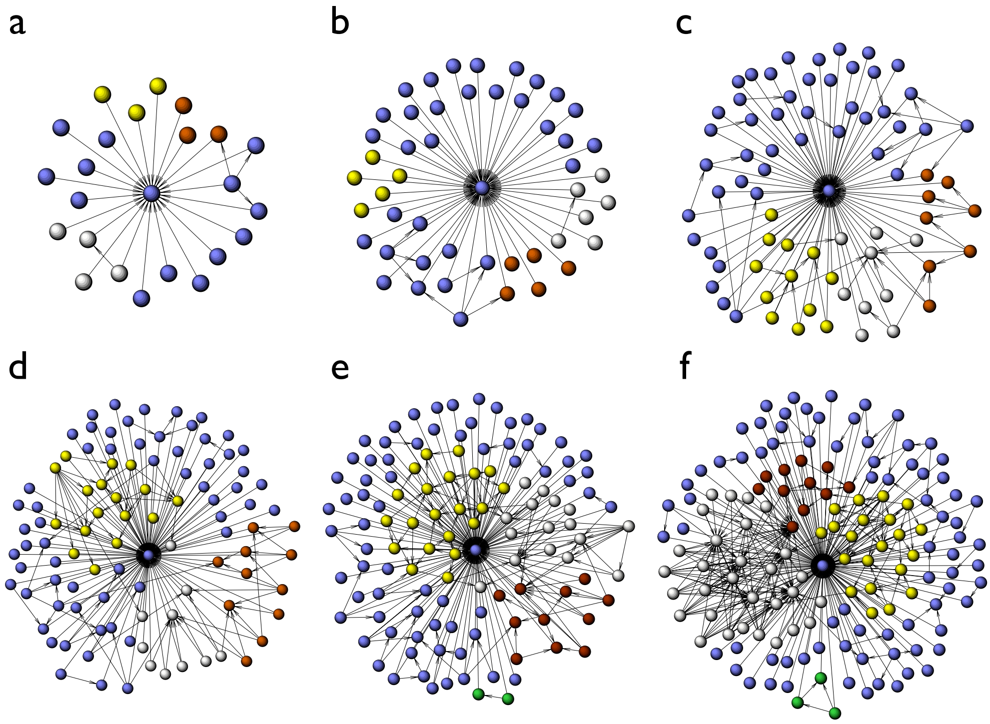

Figure 2: From (a) to (f), evolution of a patent subset related to computed tomography. The hub in the center corresponds to the precursor invention by G. Hounsfield (US patent 3778614).

citations, i.e., how many cites receive any patent. There are, however, a number of limitations associated to this simple type of analysis. Here, we propose [3] that patent citation networks are the appropriate approach to the global analysis of the process of technological innovation. We define the patent citation network G as the set of patents and their citations. This network belongs to the more general class of citation networks, which includes the scientific citation network. Interestingly, scientific citation networks share other traits with patent citation networks (see below). Figure 1 shows the first 7000 patents as registered by the US patent and trademark office (http://www.uspto.gov). Network analysis reveals interesting patterns that cannot be easily recovered by other means. For instance, patents are grouped into modules, that is, sets of nodes that exchange more links between them than with the rest of nodes. A careful inspection of patents associated to nodes within a module reveal common functional traits [3]. This is also illustrated in figure 2 for a specific example of patent growth dynamics.

We have found that the in-degree distribution for the patent citation network follows an extended power-law form [3]:

$$P_i(k) \sim (k + k_0)^{-\gamma} \tag{1}$$

where $k_0 = 19.46 \pm 0.22$ and $\gamma = 4.55 \pm 0.04$. The extended power-law reduces to a pure power-law when $k >> k_0$ and it degenerates to an exponential distribution $P_i(k) \sim exp(-\gamma k_i/k_0)$ for $k << k_0$. These extended power-laws have been associated to a mixed attachment mechanism. New links attach to target nodes according to their degree and also at random. It has been found that degree distribution displays this extended power-law for many real networks including the citation network of scientific papers an the movie actor network. However, here we propose an alternative attachment rule for the emergence of the extended power-law, namely, the combination of preferential attachment and aging.

The direct observation of the patent citation process provides a reliable estimation of the attachment rule. By tracking how new patents introduced at time $t > t_0$ cite old patents at t_0 we can evaluate the attachment rule depending on the in-degree k_i of cited patent and the age $\tau = t_0 - t$:

$$\Pi(k_i, \tau) \sim k_i^{\beta} \tau^{\alpha-1} e^{-\left(\frac{\tau}{\tau_0}\right)^{\alpha}} \tag{2}$$

Here, τ_0 controls the rightward extension of a Weibull aging term. As τ_0 increases, so does the probability of citing older papers.On the other hand, small values of τ_0 indicate strong aging that favors recently published patents [7]. Here we choose the simplest assumption (preferential attachment $\beta = 1$). Interestingly, such attachment rule is also shared by the network of scientific citations [7]. This evidence suggests there is strong convergence in the global organization and evolution of scientific ideas and technological artifacts.

3 Discussion

The results obtained in our analysis of the patent citation network, considered in abstract terms as a mapping into the landscape of technological change, support our previous ideas developped within DELIS SP5 concerning universal principles of evolution shared between biological and technological structures. It reveals the presence of a well-defined pattern of change that shares many important aspects with evolutionary search in nature. The existence and features of such landscape will be explored through the next year.

The patterns of innovation emerging in our society are the outcome of an extensive exchange of shared information linked with the capacity of inventors to combine and improve previous designs. Even very original inventions are not isolated from previous achievements. A patent can be identified as an object which needs a minimum amount of originality to be considered as truly different from previous patents. Moreover, to be obtained, it must properly refer to related patents in a fair way. Such constraints make this system specially interesting since we can wisely conjecture that it represents the expansion of real designs through some underlying technology landscape. These designs can be just small improvements or large advances. Our analysis provides a quantitative approach to this evolving structure using the approach of statistical physics.

As a final point in our discussion, it is worth noting that we have strong correlations among patents indicating a complex organization in modules. As shown by the example in figure 2, together with the nonlinearities associated with the attachment rules, there is some underlying community structure in the patent network that deserves further exploration. The emergence of modules is a natural consequence of the specialized features shared by related patents. But it might also reveal the structure of the innovation landscape itself: new patents related to previous ones can also be understood as improved solutions that explore the neighborhood of previous solutions. This view would provide a quantitative picture of the topology of technology landscapes [8, 9].

References

[1] C. O. Wilke, and C. Adami, "The Biology of Digital Organisms", Trends in Ecology and Evolution, (2002).

[2] P. fernandez and R. V. Solé, "Neutral fitness landscapes in signalling networks", J. Royal Soc. Interface (2006) in press.

[3] Valverde, S., Solé, R. V., Bedau, M., and Packard, N. H., "Topology and Evolution of Technology Innovation Networks", submitted to Phys. Rev. E. (2006).

[4] Solé, R.V., Ferrer-Cancho, R., Montoya, J. M. and Valverde, S., Complexity, **8**, 20 (2002).

[5] US Patent and Trademark Office (http://www.uspto.gov/)

[6] Jaffe, A. B., and Trajtenberg, M., *Patents, Citations and Innovations: A Window on the Knowledge Economy*, MIT Press (2002).

[7] Borner, K., Maru, J. T., and Goldstone, R. L., PNAS, **101**, 5266 (2004).

[8] Kauffman, S. A. *At home in the universe*. Oxford U. Press (1995).

[9] Kauffman, S. A. *Investigations*. Oxford U. Press (2000).