

Title: What can the study of technological evolution offer to biology?

Koon-Kiu Yan^{1,2}, Daifeng Wang^{1,2}, Mark Gerstein^{1,2,3}

1 Program in Computational Biology and Bioinformatics,

2 Department of Molecular Biophysics and Biochemistry,

3 Department of Computer Science,

Yale University, New Haven, CT 06520

The remarkable resemblance between the development of technology and the evolution of biological systems has fascinated generations of engineers, biologists, historians and philosophers [1][2][3]. Indeed, there are many apparent analogies. For instance, both biological and technological systems are adaptive, meaning their evolution is driven by some form of selection. While natural selection shapes the diversity of species, selection takes place in technological evolution in the form of the market, which combines various elements like physical constraints and customer requirements. Apart from such analogies, biological and technological evolutions share deeper commonality. For example, the bursty nature of biological evolution, punctuated equilibrium [4], has been reported in the evolution of various technological systems such as software systems and programming languages [5][6]. [As biologists, what can we learn from studying the evolution of human-designed systems?](#)

Perhaps, the most intriguing question is as follow: [What is the reason behind the resemblance, given that random tinkering drives biological evolution whereas technological evolution is human-driven in accordance to the plan of rational designers?](#) [Can we develop an unified framework or even discover laws to two evolutionary processes? To compare and contrast the two systems, it is worthwhile to look at the underlying networks.](#) The network is the common framework used to describe both biological and technological systems. By capturing the interactions between heterogeneous components in the corresponding systems, the underlying structure of the complex networks determines the function of the systems. A decade ago, Uri Alon [and collaborators](#) observed several common features in the organization of biological and engineering circuitry, such as modular organization and the existence of recurring elements called network motifs [7]. He argued that the common architectures result from common design principles adopted by nature and human design. For instance, certain network motifs make a system tolerant to noise [8], such motifs are therefore widely found in biological and engineering networks for the sake of robustness. Despite a biological system and a man-made system presenting two similar solutions to an engineering problem, the mechanisms or routes to which the solutions are derived remains in question. To answer this question, instead of merely focusing on the convergent trends, it may be useful to re-examine certain differences.

Koon-Kiu Yan 1/28/2016 12:05 PM

Deleted: Evolution, Grand unification?

Koon-Kiu Yan 1/28/2016 12:06 PM

Deleted: [1]

Koon-Kiu Yan 2/2/2016 3:14 PM

Deleted: lying at the heart of the resemblance

Koon-Kiu Yan 2/2/2016 3:15 PM

Deleted: G

Koon-Kiu Yan 2/2/2016 3:15 PM

Deleted: ,

Koon-Kiu Yan 2/2/2016 3:16 PM

Deleted: are there laws that unite

Koon-Kiu Yan 2/2/2016 3:33 PM

Deleted: t

Koon-Kiu Yan 2/2/2016 3:17 PM

Deleted: he

Koon-Kiu Yan 2/2/2016 3:21 PM

Deleted: of the two distinct classes of systems? [2]

Mapping the evolutionary patterns of components onto the underlying networks may shed light on the arrival of similar solutions. As a case study, we focus on a piece of software, the statistical computing language R, a technological system built through the collaborative efforts of many statisticians and programmers. The evolution of R is captured by the so-called package dependency network, which specifies how the proper installation of a package (node) depends on (edge) the installation of another package. Such dependency exists because most programmers tend to reuse existing code rather than developing everything from scratch. While the component dependency in software systems by itself has interesting biological analogies [9][10], here we concentrate on tracking the updates of different packages. As shown in Figure 1, central nodes (hubs or bottlenecks) in the network evolve at a faster rate, meaning they are updated more often by programmers. Of particular interest is a comparison to the protein-protein interaction network in human, a map that captures all physical interactions within the human proteome (Figure 1). In the biological network, central nodes are under strong selective constraints, exhibiting a slower rate of evolution as quantified by the ratio between synonymous versus non-synonymous mutations [11][12]. In other words, network centrality and rate of evolution is negatively correlated in biological networks, but positively correlated in technological networks.

What are the reasons and implications for this observation? Andreas Wagner formalized the idea of a genotype network in both biological and technological systems [13][14]. The term is used to describe the connections among all genotypes of a system in which two genotypes are connected if one is able to become another via a certain evolutionary step. As a genotype could be a network of constituents by itself, the genotype network could be pictured as a network of networks. In such an abstract setting, biological and technological evolution results in different trajectories on genotype networks. Their outcomes are respectively natural and the human-driven ways for innovation. Mediated by random mutations, innovation is possible in biological systems only if systems exhibit a certain level of mutational robustness, i.e. the effects of certain mutations are neutral rather than lethal. From a network perspective, mutational robustness is usually achieved by organizing the systems into decentralized modules. While the hubs are negatively selected because of many underlying constraints, the peripheral nodes provide room for innovation and they are in general under positive selection [15].

Technological evolution, however, exhibits a different innovative process. As in the package dependency network, a piece of code is frequently used by many disparate processes. Recycling code is cost-effective and indeed a common practice for a software engineer. Intuitively, increasing the sharing of packages lowers the innovability of the system. Nevertheless, being driven by rational designers, this loss is compensated by constantly modifying the hubs.

Indeed, the central points in a system are often those in the greatest use and hence are in the most need of the designer's attention. The situation is analogous to highway networks: The road planner thinks construction is possible in Manhattan without too much disruption, and one sees comparatively more construction on highly used bottlenecks (e.g. the George Washington Bridge) compared to out-of-the-way thoroughfares. In other words, with the efforts of intelligent designers, technological evolution presents an alternate route in the genotype network, in contrast to the tinkerer in biological evolution. In a sense, we could picture both the designer and tinkerer working on an optimization problem with similar underlying design objectives, but employing different criteria when balancing constraints. The difference between tinkerer and designer suggests that, as an optimization process, no approach optimizes all objectives (cost-effectiveness and mutational robustness in this case) and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there's no free lunch [16].

In short, we have presented a side-by-side comparison on the evolution of a technological network and a biological network. By stepping out of the biological domains, we see the contrasts with technological evolution on ideas like genotype networks and trade-offs provide new perspective to biological evolution. The formal theory of evolution was originated from biology. Since then the worldview of Darwin has penetrated and transformed many disciplines. So far, what we examined is merely a case study. It will be instructive to look for further data. Today, the sequencing technology enables biologists to potentially investigate the genomes of every extant species, whereas thorough documentation in the technological era presents engineers an unprecedented opportunity to study the evolution of design systems. Given the resources to explore the space of genotypes in both domains, analysis could potentially be scaled up. Perhaps, studying the evolution of systems in different domains will in return benefit biology.

- [1] G. Basalla, *The Evolution of Technology*. Cambridge University Press, 1989.
- [2] B. Arthur, *The Nature of Technology*. Free Press, 2011.
- [3] A. Wagner and W. Rosen, "Spaces of the possible: universal Darwinism and the wall between technological and biological innovation," *J. R. Soc. Interface*, vol. 11, no. 97, p. 20131190, Aug. 2014.
- [4] S. J. Gould, *The Structure of Evolutionary Theory*, 1 edition. Cambridge, Mass: Belknap Press, 2002.
- [5] A. A. Gorshenev and Y. M. Pis'mak, "Punctuated equilibrium in software evolution," *Phys Rev E*, vol. 70, no. 6, p. 4, 2004.
- [6] S. Valverde and R. V. Solé, "Punctuated equilibrium in the large-scale evolution of programming languages," *J. R. Soc. Interface*, vol. 12, no. 107, p. 20150249, Jun. 2015.
- [7] U. Alon, "Biological Networks: The Tinkerer as an Engineer," *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.
- [8] U. Alon, "Network motifs: theory and experimental approaches," *Nat Rev Genet*, vol. 8, no. 6, pp. 450–461, Jun. 2007.

Koon-Kiu Yan 1/28/2016 10:44 PM

Moved down [1]: The worldview of Darwin has penetrated to many disciplines.

Koon-Kiu Yan 1/28/2016 11:46 PM

Deleted: it is good to put both of these trends together and look for a united theoretical framework. The worldview of Darwin has penetrated to many disciplines.

Koon-Kiu Yan 1/28/2016 10:44 PM

Moved (insertion) [1]

Koon-Kiu Yan 1/28/2016 11:48 PM

Deleted:

- [9] K.-K. Yan, G. Fang, N. Bhardwaj, R. P. Alexander, and M. Gerstein, "Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks," *Proc. Natl. Acad. Sci.*, vol. 107, no. 20, pp. 9186–9191, May 2010.
- [10] T. Y. Pang and S. Maslov, "Universal distribution of component frequencies in biological and technological systems," *Proc. Natl. Acad. Sci.*, vol. 110, no. 15, pp. 6235–6239, Mar. 2013.
- [11] H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman, "Evolutionary Rate in the Protein Interaction Network," *Science*, vol. 296, no. 5568, pp. 750–752, Apr. 2002.
- [12] M. W. Hahn and A. D. Kern, "Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks," *Mol. Biol. Evol.*, vol. 22, no. 4, pp. 803–806, Apr. 2005.
- [13] A. Samal, J. Matias Rodrigues, J. Jost, O. Martin, and A. Wagner, "Genotype networks in metabolic reaction spaces," *BMC Syst. Biol.*, vol. 4, no. 1, p. 30, 2010.
- [14] K. Raman and A. Wagner, "The evolvability of programmable hardware," *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269–281, Feb. 2011.
- [15] P. M. Kim, J. O. Korbil, and M. B. Gerstein, "Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context," *Proc. Natl. Acad. Sci.*, vol. 104, no. 51, pp. 20274–20279, Dec. 2007.
- [16] A. D. Lander, "Pattern, growth, and control," *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.
- [17] The R package dependency network consists of all the available packages (5711) via R studio in OS X 10.9 at October 2014.

Figure 1.

Different evolutionary patterns in biological networks versus technological networks. The left shows the protein-protein interactions network in human, whereas the right is the R package dependency network [17]. Central nodes in a PPI network are under strong selective constraints (slow rate of evolution), whereas central nodes in the R package dependency network evolve faster. In other words, network centrality and rate of evolution is negatively correlated in biological networks (left), but positive correlated in technological networks (right).