Title: Evolution, Grand unification?

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 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], have been reported in the evolution of various technological systems such as software systems and programming languages [5][6]. Perhaps, the most intriguing question lying at the heart of the resemblance is: Random tinkering drives biological evolution whereas technological evolution is human-driven in accordance to the plan of rational designers, are there laws that unite the evolution of the two distinct classes of systems?

A common framework to describe both biological and technological systems is the network. 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 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 are results of common design principles adopted by nature and human design. For instance, certain network motifs make a system tolerant to stochastic fluctuations, from either intrinsic or external sources [8]. Such motifs are therefore widely found in biological and engineering networks for the sake of robustness. Despite the identification of the common solutions presented by biological and technological evolution bring valuable insights in terms of understanding nature and application standpoint, the process of how the solutions are arrived via completely different routes remains in question. To illuminate on this question, instead of merely focusing on the convergent trends, it may be useful to re-examine certain differences.

Mapping the evolutionary patterns of components onto the underlying networks may shed light on the arrival of similar solutions. Biologists have established standard metrics to quantify the degree of evolution, or to certain extent the selection pressure, of elements like genes and proteins. This is because all living organisms have their genomes composed of the same units: A, C, T, and G. For technological systems, there is no unified quantification simply because two technological system could be vastly different. As a case study, we focus on a piece of software,

STD? VN ARCH.

the statistical computing language ‘R’, arguably a technological system built by 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. 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 the Figure, central nodes in the network (hubs or bottlenecks) 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. In the biological network, central nodes are under strong selective constraints, exhibiting a slower rate of evolution 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 positive ly correlated in technological networks.

What are the reasons and implications? Andrea Wagner formalized the idea of genotype network it in both biological and technological systems [13][14]. The term is used to describe the connections between all genotypes of a system, whereas two genotypes are connected if one is able to become another via a certain evolutionary step. As a system 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 evolutions are simply different trajectories on their corresponding genotype networks. Their outcomes are respectively the natural and the human-driven ways for innovations. Mediated by random mutations, innovation is possible in biological systems only if the systems exhibit certain level of mutational robustness, i.e. the effects of certain mutations are neutral rather than lethal. In a network perspective, mutational robustness is usually arrived 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].

The situation is opposite in technological evolution. As in the package dependency network, a piece of code is highly called by many disparate processes. Recycling code is cost-effective and indeed a common practice for software engineer. But intuitively, such an organization lowers the innovability of the system. Nevertheless, being driven by rational designers, the innovability 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 road 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 that both the designer and tinkerer are working on an optimization problem with similar underlying design objectives, but taking different views when balancing constraints. The discrepancy between tinkerer and engineer 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].

TOO STRONG

The next-generation sequencing technology enables biologists to investigate the genomes of as many species as we could find in the tree of life, whereas the technological era presents engineers a unique opportunity to study the evolution of design systems in an unprecedented resolution. Therefore we are now given the resources to explore the space of genotypes in both domains; the next step is to unite the worldview of Darwin across natural and man-made systems into quantitative laws.

[1] "The Evolution of Technology," *Cambridge University Press*. [Online]. Available: http://www.cambridge.org/us/academic/subjects/history/history-science-and-technology/evolution-technology. [Accessed: 03-Oct-2015].

[2] *The Nature of Technology*. .

[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.

[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.

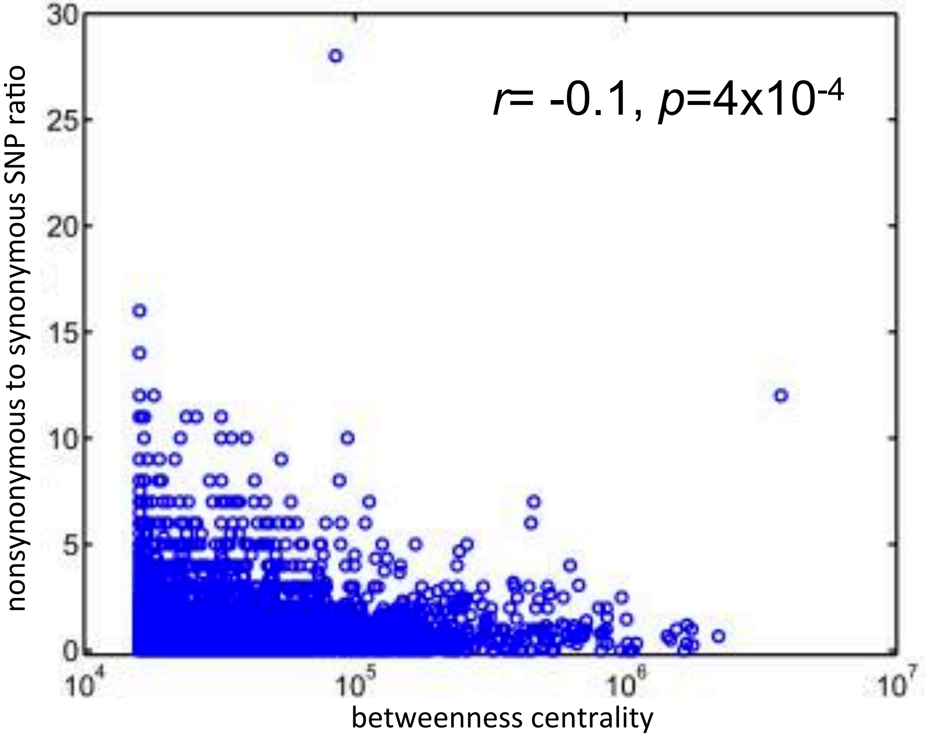
[12] M. W. Hahn and A. D. Kern, "Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks," *Mol. Biol. Evol.*, vol. 22, no. 4, pp. 803–806, Apr. 2005.

[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. Korbel, 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.

Figure 1.

Different evolutionary patterns in biological networks versus technological networks. The left shows the protein-protein interactions network in human [84], whereas the right is the R package dependency network. The R package dependency network consists of all the available packages (5711) via R studio at October 2014.

Protein-protein interaction network



R package dependency network