(?) Comparative Netomics - lessons from cross-disciplinary network comparison

A signature of biology in the "omic" era is the shift of attention from few individual components to the comprehensive collections of constituents [1]. For instance, structural biologists studied the binding of a few proteins in the past but nowadays they are able to probe the interactions between thousands of proteins. Similarly, geneticists who used to knockout a single gene for functional characterization can now employ high-throughput techniques in functional genomics to study the genetic relationships between all genes. In many cases, genome-wide information describing how components interact could be captured by a network representation [2]. While we have been astonished by the complexity of such networks found in genomics or systems biology, many are not able to gain any intuition from the hairballs [3].

Is there any clue for deciphering the hairballs? Throughout the history of science, many advances in biology were catalyzed by discoveries in other disciplines. For instance, the maturation of X-ray diffraction facilitated the discovery of the double helix, and later on the characterization of structures of thousands of different proteins. One may wonder if ideas in other areas of science could help us to decipher the hairballs. In this essay, we argue that, while the influx of ideas in the age of reductionism mostly originated from specific areas in physics or chemistry, to understand biology via a systems perspective, the concept of network serves as a mediator that brings a new wave of catalysts to biology, from disciplines as diverse as engineering, behavioral science and sociology. Toward this end, biologists should think about performing cross-disciplinary network comparison.

Drawing analogy is by no mean new to biologists. For instance, decades ago Dawkins came up with the idea of meme, which is a unit carrying cultural ideas analogous to gene in biology [4], to illustrate principles of selection. The comparison has been further elaborated in the protofield of phylomemetics, which concerns itself with phylogenetic analysis of non genetic data [5]. Nevertheless, comparing a bio-molecular network with a complex network from a disparate field, say a social network, sounds like comparing apples to oranges. So what kinds of comparison could truly deepen our understanding? We believe that it is useful to think of different descriptions of a cellular system as a spectrum (Figure 1).

**A spectrum of cellular descriptions**

Given the complexity of a cell, a certain level of simplification is necessary for useful discussion. We could picture the description of cellular systems as a spectrum (Figure 1). On one hand, there's a simple parts list that just enumerates each component without specifying any relationships. On the other hand, there is a complete three or even four-dimensional picture of how cellular molecules interact in space and time. It is well regarded that the characteristics of a cellular system cannot be explained by the characteristics of individual components – the whole is greater than the sum of its parts. Therefore, the parts list description is not fully informative. However, the full picture is often too ambitious for the current state-of-the-art in data acquisition.

Network description sits conveniently between these extremes by capturing the some of the relationships between components of the parts list in a flexible fashion, particularly those where topology rather than exact location captures the relationship. There are two ways to think about networks. The first one, referred as association network, is essentially a process of abstraction; meaning entries are connected via abstract mathematical association. While any mechanistic interaction could be abstracted as a mathematical association, the idea of association could be generalized to statistical relationships between two components. An example is the disease networks [6] in which a gene (genotype) and a disease (phenotype) are connected via the statistical association between the existence of genomic variants and the occurrence of the disease. Networks derived from co-expression relationships provide another example. The second kind of network, referred as mechanistic network, on the contrary, is a process of concretization. Unlike abstraction that is moving away from the complete 4D-picture, concretization is pointing towards this picture. It aims to understand more of the physical processes happening inside a living system, for instance the processing of information, the

chemistry of metabolites and the assembly of molecular machine, and therefore focuses on incorporating various details of interactions. Adding further mechanistic detail onto a simple nodes-and-edges skeleton can often be visualized by decorating edges with directionality, color, thickness etc. Nevertheless, the incorporation of too much detail makes the system intractable, and network formalism generally breaks down if we try to load spatial or temporal details as well as higher-order interactions onto the diagram. At certain point, the actual four-dimensional picture is required.

The advantage of focusing on rather abstract association is, mathematical formalisms are more readily transferrable. Toward this end, by comparing similar network-based mathematical formalisms across disciplines, biologists will benefit in terms of algorithms or method development. On the other hand, mechanistic networks can serve as the skeletons for describing different complex systems in detail. In this case, because of systems-specific details, it is less likely that everything could be transferred from one discipline to another. Here, it is important to focus on the conceptual resemblance instead of merely topological resemblance. And comparison of appropriately matched networks allow biologists to gain intuitions by examining analogous interactions in cross-disciplinary complex systems in the way as the interactions between molecular components in cells.

**Comparison leverages mathematical formalism**

Lying at the heart of the power of network formalism is its simplicity. In the era of Big Data, network is a very useful data structure with a wide variety of applications in both biology and other data intensive disciplines like computational social science.

*Formalism focusing on network topology*

One of the first applications of abstract network formalism is to compare the organization principles of various complex systems. The earliest and probably the most important observation is that networks organize themselves into scale free architectures in which a majority of the nodes contain very few connections (edges) while a few nodes (also called hubs) in the network are highly connected [7]. The behavior of scale-free networks is dominated by a relatively small number of nodes and this ensures that these networks are resistant to random accidental failures but are vulnerable to coordinated attacks at hub nodes [8]. In other words, just like the Internet functions without any major disruptions even though hundreds of routers malfunction at any given moment, different individuals belonging to the same biological species remain healthy in spite of considerable random variation in their genomic information. Nevertheless, a cell is not likely to survive if a hub protein is knocked out. For example, highly connected proteins in the yeasts' protein-protein interaction network are three-times more likely to be essential than proteins with only a small number of links to other proteins [9]. Another important property of scale-free networks is its small world property [10][11]: the presence of hubs ensures that the distance between any two nodes in the network is small. An example is the combination of airport and highway networks that ensure that we can travel across any two points in USA in a finite amount of time. This has led to a second measure of a nodes' centrality in the network that is based on the effect of its removal on the communication pathways between all the other nodes in the network. Similar in spirit to heavily used bridges, highways, or intersections in transportation networks, a few centrally connected nodes termed bottlenecks funnel most of the paths between different parts of the network and removal of these nodes could reduce the efficiency (increase of distance) of communication between nodes within these networks [12]. Indeed, it has been reported that changes to the sequences of bottlenecks in biological networks can be deleterious [13]. A more sophisticated way to define centrality is to take into account the importance of neighbors. Toward this end, the PageRank algorithm plays a prominent role. Faced with a search query, Google has to decide which set of results are ranked higher and appear on the first page of the results page. Originally developed in social network analysis [14], the PageRank utilizes an algorithm developed to rank relevant documents based on the rank of the websites that link to this document in a self-consistent manner - ie being linked by higher ranking nodes counts for more. The algorithm was then adopted in food webs to prioritize nodes that are in danger of extinction [15] and also to rank prognostic relevance for patients with cancers [16].

Koon-Kiu Yan 9/25/2014 11:29 AM
**Deleted:** that

Koon-Kiu Yan 9/25/2014 11:29 AM
**Deleted:** most of the nodes in these networks are connected to each other and

Besides scale-free nature, one can easily observe that social networks tend to have communities within them due to the relatively larger number of interactions between people in the same neighborhood, school, or work place. People within the same social group tend to form strong ties in the form of cliques and form a single cohesive group. Analogous to closely-knit social groups, a large number of biological components form a single functional macromolecular complex like the ribosome. More generally, a common feature of a large number of technological and biological networks are that they are organized in the form of modules such that nodes within the same module have a larger number of connections with each other as compared to nodes belonging to different modules [17]. The quantity dubbed modularity tries to quantify this, comparing the number of intra and inter module links in the network.

*Formalisms focusing on the interplay between topologies and the properties of nodes*

Network is extremely useful in data mining because it can be used as a reference for mapping additional properties or features of different nodes. Similar questions and solutions have been come up in dealing with biological data as well as data from disciplines like computational social science. An important example is the inference the missing data by the idea of "guilt by association", or the idea that nodes that have similar associations in the network tend to be more similar in nature. For example, in a social context, if your friends in Facebook use Product Y, you are more likely to use product Y and the advertisements you view online are personalized based on these recommendation systems [18]. In biological context, the assumption is based on observations like the cellular components within the same module are more closely associated with the same set of cellular phenotypes than components belonging to different modules [19], and the modules within gene coexpression networks also tend to contain genes with similar functions and genes within the same module are often involved in the same biological pathway [20]. As a result, one could infer the functions of a protein or a non-coding element based on the function of its neighbors in the underlying network. Networks play an important role in gene prioritization, an essential process for applications like disease gene discovery because of limited validation and characterization resources [21]. For examples, network properties of individual genes have been used to distinguish functionally essential and loss-of-function tolerant genes [22]. More formally, one could prioritize the candidate genes based on how they are connected to the known genes. For example, if a gene is one-step away from a group of disease genes, it is very likely that the gene is associated with disease X. Of course, the influence of a node may not be restricted to its nearest neighbors; network flow algorithms are widely used to examine the long-range influence [23][24]. In social science context, for example, ones use cascade structured models to capture the information propagation on web blog networks, and predict the blog's popularity [25].

We want to emphasize that networks are in general noisy. High-throughput experiments may result at spurious links, and missing data is very common social science. Methods for link prediction and denoising are therefore very useful. Link prediction can be done by using the network information alone, for instance, in a protein-protein interaction network, defective cliques were used to find missing interactions and determine the parts required to form a functional macromolecular complex [26]. More often, because whether two nodes are connected depends on their intrinsic properties, one could employ machine-learning techniques to explore the relationships between connections and various features [27]. Recently, generative models of networks, say stochastic block models [28], are very popular in computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standards for validation.

*Formalisms focusing on causal relationships and dynamics*

The construction of various association networks is an active area of research for both biology and computational social science. While correlational relationships could potentially be easily calculated with the appropriate data, a fundamental question is the distinction between direct and indirect interactions. For example, a transcription factor X regulates gene Y and Z, one could expect pairs like X-Y, X-Z and Y-Z are all correlated, but the key is to identify the direct regulatory

Unknown
**Field Code Changed**

Koon-Kiu Yan 9/25/2014 2:00 PM
**Deleted:** The same is true for predicting disease-associated genes: if the neighbors of a gene are all associated with Disease X,

Koon-Kiu Yan 9/25/2014 2:05 PM
**Deleted:** [21]

Koon-Kiu Yan 9/25/2014 2:00 PM
**Deleted:**

Unknown
**Field Code Changed**

Unknown
**Field Code Changed**

Koon-Kiu Yan 9/26/2014 2:36 PM
**Deleted:** [24]

Koon-Kiu Yan 9/25/2014 2:25 PM
**Deleted:** Very often whether

Koon-Kiu Yan 9/25/2014 2:27 PM
**Deleted:** depend

Unknown
**Field Code Changed**

Koon-Kiu Yan 9/26/2014 2:39 PM
**Formatted:** Font:(Default) Arial, Font color: Black

interactions X-Y and X-Z. Established mathematical machineries like Bayesian networks, Markov random fields and other information theoretical frameworks [29] have been used for this purpose.

The inference of causal relationships could be greatly benefited by time-series data. In social science, online retailers are interested to use purchase records to study how customers influence each other [30]. On the other hand, the same question is extremely common in biology, under the term “reverse engineering”. For example, how can we infer the developmental gene regulatory network from temporal gene expression dynamics? Ideally, one could write differential equations to fit the temporal data; nevertheless, temporal data in most genomics experiments do not have enough time-points. To overcome the drawback, data mining techniques such as matrix factorization are employed. For instance, given the genome-wide expression profile of at different time-points, one could project the high-dimensional gene expression data to low dimensional space and write differential equations to model the dynamics of the projections [31]. The inference of casual and direct relationships from statistical data points to the study of mechanistic networks.

Apart from the actual dynamical processes happen in a network, one could explore the evolutionary dynamics of networks. In biological context, pairs of orthologous genes can be used to align networks from different species. Based on the notion, a mathematical formalism was developed to measure the evolutionary rewiring rate between networks across species in analogous to quantifying sequence evolution [32]. It was shown that metabolic networks rewire at a slower rate compared to various regulatory networks. The same notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [33].

**Comparison gains physical intuition**

Now we shift discussion to mechanistic" networks. Here, the network framework serves as a skeleton of different complex systems. From a biologist standpoint, network comparison thus brings intuition from other disciplines to bear on molecular biology.

*Looking for universal mechanisms*

The study of mechanistic networks enables one to further explore the origin of some of the striking similarity observed in the structure of biological and non-biological networks. Probably the most important example is to model the universally observed scale-free degree distribution described above. A number of different stochastic mechanisms lead to the formation of scale-free networks. For example, when a new airport is created, it is more efficient for people to travel to other parts of the world if the airport connects to pre-existent hubs in the networks. This model is called the preferential attachment model or the “rich get richer” model as newly created hubs prefer to connect to pre-existent hubs in the network [7]. Another mechanism that can lead to the formation of hubs in the network is the so-called duplication divergence model. In this model, a new copy of a pre-existing node and its associated edges (for example, a webpage with all its pre-existing links) are created randomly and the edges associated with these two nodes are allowed to change independent of one another. This model leads to scale-free networks because the number of edges for a hub node will increase further as it is more likely that one of its many neighbors will get duplicated. One of the main mechanisms for the creation of new proteins is gene duplication. Hence, the formation of scale-free behavior in the protein-protein interaction network could be explained by the duplication-divergence model [31]. Remarkably, the same duplication-divergence mechanism has been applied to describe the patterns of “memes” in online media [32].

More recently, it has been shown that components in both bacterial genomes as well as large-scale computer software projects form multilayered dependency networks (enzyme A is used to decompose the output metabolites of enzyme B; the installation of package A depends on the installation of package B). The common underlying dependency networks leads to the same power-law components-usage frequency distribution (how often a enzyme is present in a bacterial genome; how often a certain package is installed in a computer) [36]. While it is elegant

to explain the topology of disparate networks by simple stochastic models, such universal mechanism cannot capture the full picture. For example, it has also been shown that when the protein interaction network is analyzed in terms of the structural interfaces involved (giving rise to the structural network) the duplication-divergence model only applies to hubs having a single interface as opposed to those with many (with the duplicated protein reusing the same interface as its parent) [37]. Nevertheless, as biologists, we love to think about functions and selection; it is interesting to see that, by network comparison, network organization could be a manifestation of stochasticity.

*Looking for common design principles*

Apart from universal mechanisms, comparison of networks shed light on the design principles of networks. An example is the so-called network hierarchy (see Box 1). Many biological networks, for instance transcription regulatory networks, have an intrinsic direction of information flow, forming a hierarchical organization, similar to structures like corporate management hierarchy [38]. Nodes in the middle level therefore form the information bottlenecks. To avoid break down of flow, middle regulatory factors tend to co-regulate downstream targets; the same is true for management hierarchy in where middle managers tend to communicate often [39][40].

Koon-Kiu Yan 9/25/2014 10:18 PM

**Deleted:** Most observed similarities in terms of network organization are not easy to explain by simple mechanisms or principles, for instance,

Lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. The mapping points to biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [41]. As it is in general very hard to define a "function", toward this direction, comparison with various technological or engineered networks with well-defined functions is particularly insightful. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a source of negative feedback and a source of time delay. Nevertheless, different oscillators (e.g. for circadian rhythms, for cell cycle, or from various organisms) have a certain level of variation because of additional design objectives or strategies. This is just like the case that not all electronic devices use the same oscillator design because of other design objectives. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [42]. Robustness is obviously a preferred design objective because it makes a system tolerate stochastic fluctuations, either intrinsically or from external sources. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by an interface is widely practiced [43]. The same is for biological networks because modules can be readily reused to adapt new functions.

*Looking for the commonalities and differences between tinkerer and engineer*

The comparison of biological networks and technological networks should best be performed under the light of evolution. As Alon highlighted by the phase "the tinkerer as an engineer" [42], it is remarkable that "good-engineering solutions" are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer neither designs things nor builds systems— it settles on systems that, historically, conveyed a survival benefit (and if a better way comes along, it will adopt that). On the other hand, technological networks are essentially blueprints drawn by engineers who have a grand plan that makes sure everything works harmoniously. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that constructs living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological ones like roadways and electronic circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [44]. To a certain extent, an engineer is a tinkerer (see Box 2).

Under such a united framework, we could picture that both engineer and tinkerer are working on an optimization problem with similar underlying design objectives. Like all optimization problems, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there's no free lunch [45][46]. Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited in the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many disparate processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fraction of functions are updated often [47]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. The patterns seems to be hold for other software systems like the organization of packages in the statistical computing language R (Figure 2). In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, hubs in the software system evolve rapidly. This seems to be counter to ones intuition that an engineer should not meddle too much with highly connected components. However, there is another intuition in play: rational designers may believe that they can modify a hub without disrupting it -- in contrast to the situation with random changes. Moreover, the central points in a system are often those that are in the greatest use and hence are in the most need of the designer's attention. The situation is analogous to road networks: one sees comparatively much construction on highly used bottlenecks (e.g. the George Washington Bridge) as opposed to out of the way thoroughfares (see Box 2).

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of 'omes' across species. In the "omics" era, may be it is a time to extend our tradition even further to compare networks in biology as well as other disciplines. Over the past few years, efforts have been spent on concatenating networks together forming a multiplex structure [48][49]. This direction is of particular interest to biology. First, from an abstract formalism standpoint, due to rapid advancements in data acquisition, the structure of biological data goes beyond a single layer of network to multiplex structure: the multiple layers could either be formed by different categories of relationships (co-expression, genetic interactions, etc.), analogous to social science in which an individual may participate in multiple social circles: family, friends, colleagues, or in online setting: Facebook, LinkedIn and Twitter. Second, mechanistically, biological regulation happens in multiple levels: transcriptional regulation, post-transcriptional regulation, and even post-translational in analogous to a city with electrical networks, water pipes, and cell phone lines. We are looking forward to some of the methods developed in other contexts to be applied in biology.

So far, we have already seen examples in which comparison brings new connections. For examples, there are emerging theories that unite evolved and designed systems; there is an increase of attention among biologists and sociologists on the connection between genomics information and sociological information such as whether phenotypes or genotypes are correlated in friendship networks [50]. Indeed, various scientific disciplines form a network in the intellectual universe where knowledge emerges when things connect.

**Potential exhibits:**

**Figure 1 Caption**

**Figure 2 Caption**

?A table showing examples of the two types networks.
(Give more examples of association networks, like genetic interaction networks.)

?A table highlighting problems studied in the framework of association networks, and the corresponding problems arise in computational social science.

Koon-Kiu Yan 9/25/2014 10:37 PM
**Deleted:** phenomenological

Table/Figure summarizing all comparisons/references.

Box 1 Network science 101
Betweenness centrality:
As dangerous epidemics spread in a network, it is important not only to develop vaccines but it is also important to efficiently utilize these vaccines so that the spread of the epidemic can be reduced. It has been proposed that strategies based on human contact networks are most effective at reducing the spread of a disease in a population [51].

Box 1 Hierarchical organization of networks
Many biological networks possess an intrinsic direction of information flow, forming a hierarchical network organization. The hierarchical organization in biological networks resemble the chain of command in human society, like in military context and corporate hierarchy [38]. For instance, in a transcriptional regulatory network more influential transcription factors (regulators whose expression are more highly correlated with the expression of target genes) tend to be better connected (have more interacting partners) and higher in the hierarchy [52]. Moreover, the transcription factors in the middle layer tend to be more cooperative [39]. Such a situation has been well studied in management science, where in certain corporate settings middle managers interact the most with peers to manage subordinates below them [40]. These observations reflect a democratic hierarchy as opposite to a conventional autocratic organization [53].

Of particular interest for hierarchical organization is the so-called bow-tie structure, meaning the intermediate layers have fewer components than the input and output layers. For example, in a signaling network, a large number of receptors corresponding to diverse stimuli and many transcription factors form the input and output layers, whereas the intermediate layer refers to a few key molecules like calcium and cAMP that mediate the inputs and outputs [54]. Similarly, in the networking architecture of the Internet, various protocols in the input/link layer (ARP, RARP, NDP etc) and various application protocols in the application/output layer (HTTP, FTP,DHCP etc) are essentially connected by only IPv4, the primary protocols in the internet layer. The reason for the emergence of such a common pattern is still widely open, a recent paper suggested bow-tie is a result of information compression [55].

Box 2 Tinkerer versus engineer
Despite the apparent differences, the similarity between biological systems and technological systems draws a parallel between tinkerer and engineer, and the parallel points to a common framework to unite them. Wagner further proposed an analogy between the genotype space for a biological system and the design space for a technological system. These spaces contain all the possible networks in the corresponding systems. In biology, many attempts have been made to search for solutions of common functional problems such as adaptation, oscillation and cell polarization [41]. Similar studies were performed in the context of circuit design, where a set of logic gates was evolved via rewiring in order to perform a predefined computational task [56][57]. These studies suggested that in both kinds of systems, the solution networks are close together in the genotype/design space. As each solution in genotype/design has multiple neighbors, robustness of a solution to mutation facilitates the evolvability of these systems [58][59]. Indeed, it has been demonstrated that electronic circuits can be evolved to fulfill a fluctuating evolutionary goal [56]. Similarly, metabolic networks of bacteria living in multiple habitats are evolved to decompose multiple food sources [60][61]. Both of these networks show a level of modular organization.

Very often we picture engineers design things from scratch. In reality, as a technological system evolves, engineers are subjected to various constraints like tinkerer. In the example of internet architecture, while there are frequent innovations at the input layer that interact with a variety of networking hardware and output layers that connect with many different software applications, the internet layer with very few protocols is the bottleneck under heavy constraints and such protocols can hardly be replaced [62]. The observed rapid innovation at the top and bottom layers but constraint at the middle is very common in biological system. Consider the metabolic networks of different bacteria, the anabolic and catabolic components are much more diverse whereas there are less variations between central pathways [63].


[1] M. Baker, "Big biology: The 'omes puzzle," *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013.

[2] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat. Rev. Genet.*, vol. 5, no. 2, pp. 101–113, Feb. 2004.

[3] A. D. Lander, "The edges of understanding," *BMC Biol.*, vol. 8, no. 1, p. 40, Apr. 2010.

[4] R. Dawkins, *The selfish gene*, New ed. Oxford ; New York: Oxford University Press, 1989.

[5] C. J. Howe and H. F. Windram, "Phylomemetics—Evolutionary Analysis beyond the Gene," *PLoS Biol*, vol. 9, no. 5, p. e1001069, May 2011.

[6] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, "The human disease network," *Proc. Natl. Acad. Sci.*, vol. 104, no. 21, pp. 8685–8690, May 2007.

[7] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[8] null Albert, null Jeong, and null Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000.

[9] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, May 2001.

[10] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[11] L. a. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proc. Natl. Acad. Sci.*, vol. 97, no. 21, pp. 11149–11152, Oct. 2000.

[12] M. E. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 64, no. 1 Pt 2, p. 016132, Jul. 2001.

[13] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein, "The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics," *PLoS Comput. Biol.*, vol. 3, no. 4, p. e59, Apr. 2007.

[14] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.

[15] S. Allesina and M. Pascual, "Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions?," *PLoS Comput Biol*, vol. 5, no. 9, p. e1000494, Sep. 2009.

[16] C. Winter, G. Kristiansen, S. Kersting, J. Roy, D. Aust, T. Knösel, P. Rümmele, B. Jahnke, V. Hentrich, F. Rückert, M. Niedergethmann, W. Weichert, M. Bahra, H. J.

Schlitt, U. Settmacher, H. Friess, M. Büchler, H.-D. Saeger, M. Schroeder, C. Pilarsky, and R. Grützmann, "Google Goes Cancer: Improving Outcome Prediction for Cancer Patients by Network-Based Ranking of Marker Genes," *PLoS Comput Biol*, vol. 8, no. 5, p. e1002511, May 2012.

[17] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[18] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithm for Collaborative Filtering," in *Proceedings of the 14 th Conference on Uncertainty in Artificial Intelligence*, 1998, pp. 43–52.

[19] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nat. Rev. Genet.*, vol. 12, no. 1, pp. 56–68, Jan. 2011.

[20] J. M. Stuart, "A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules," *Science*, vol. 302, no. 5643, pp. 249–255, Oct. 2003.

[21] Y. Moreau and L.-C. Tranchevent, "Computational tools for prioritizing candidate genes: boosting disease gene discovery," *Nat. Rev. Genet.*, vol. 13, no. 8, pp. 523–536, Jul. 2012.

[22] E. Khurana, Y. Fu, J. Chen, and M. Gerstein, "Interpretation of genomic variants using a unified biological network approach," *PLoS Comput. Biol.*, vol. 9, no. 3, p. e1002886, 2013.

[23] S. Navlakha and C. Kingsford, "The power of protein interaction networks for associating genes with diseases," *Bioinformatics*, vol. 26, no. 8, pp. 1057–1063, Apr. 2010.

[24] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, and R. Sharan, "Associating Genes and Protein Complexes with Disease via Network Propagation," *PLoS Comput Biol*, vol. 6, no. 1, p. e1000641, Jan. 2010.

[25] E. Adar and L. A. Adamic, "Tracking Information Epidemics in Blogspace," 2005, pp. 207–214.

[26] H. Yu, A. Paccanaro, V. Trifonov, and M. Gerstein, "Predicting interactions in protein networks by completing defective cliques," *Bioinforma. Oxf. Engl.*, vol. 22, no. 7, pp. 823–829, Apr. 2006.

[27] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, no. 7191, pp. 98–101, May 2008.

[28] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed Membership Stochastic Blockmodels," *J Mach Learn Res*, vol. 9, pp. 1981–2014, Jun. 2008.

[29] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, and A. Califano, "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC Bioinformatics*, vol. 7 Suppl 1, p. S7, 2006.

[30] P. Domingos and M. Richardson, "Mining the Network Value of Customers," in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2001, pp. 57–66.

[31] D. Wang, A. Arapostathis, C. O. Wilke, and M. K. Markey, “Principal-Oscillation-Pattern Analysis of Gene Expression,” *PLoS ONE*, vol. 7, no. 1, p. e28805, Jan. 2012.
[32] C. Shou, N. Bhardwaj, H. Y. K. Lam, K.-K. Yan, P. M. Kim, M. Snyder, and M. B. Gerstein, “Measuring the Evolutionary Rewiring of Biological Networks,” *PLoS Comput Biol*, vol. 7, no. 1, p. e1001050, Jan. 2011.
[33] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, “OrthoClust: an orthology-based network framework for clustering data across multiple species,” *Genome Biol.*, vol. 15, no. 8, p. R100, Aug. 2014.
[34] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, “Modeling of Protein Interaction Networks,” *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.
[35] M. P. Simmons, L. A. Adamic, and E. Adar, “Memes online: Extracted, subtracted, injected, and recollected,” in *In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
[36] 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.
[37] P. M. Kim, L. J. Lu, Y. Xia, and M. B. Gerstein, “Relating Three-Dimensional Structures to Protein Networks Provides Evolutionary Insights,” *Science*, vol. 314, no. 5807, pp. 1938–1941, Dec. 2006.
[38] H. Yu and M. Gerstein, “Genomic analysis of the hierarchical structure of regulatory networks,” *Proc. Natl. Acad. Sci.*, vol. 103, no. 40, pp. 14724–14731, Oct. 2006.
[39] N. Bhardwaj, K.-K. Yan, and M. B. Gerstein, “Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 15, pp. 6841–6846, Mar. 2010.
[40] S. W. Floyd and B. Wooldridge, “Middle management involvement in strategy and its association with strategic type: A research note,” *Strateg. Manag. J.*, vol. 13, no. S1, pp. 153–167, Jun. 1992.
[41] W. A. Lim, C. M. Lee, and C. Tang, “Design Principles of Regulatory Networks: Searching for the Molecular Algorithms of the Cell,” *Mol. Cell*, vol. 49, no. 2, pp. 202–212, Jan. 2013.
[42] U. Alon, “Biological Networks: The Tinkerer as an Engineer,” *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.
[43] M. A. Fortuna, J. A. Bonachela, and S. A. Levin, “Evolution of a modular software network,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 50, pp. 19985–19989, Dec. 2011.
[44] 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.
[45] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.
[46] O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, and U. Alon, “Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space,” *Science*, vol. 336, no. 6085, pp. 1157–1160, Jun. 2012.

[47] 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.
[48] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community Structure in Time-Dependent, Multiscale, and Multiplex Networks," *Science*, vol. 328, no. 5980, pp. 876–878, May 2010.
[49] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.
[50] J. H. Fowler, J. E. Settle, and N. A. Christakis, "Correlated genotypes in friendship networks," *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.
[51] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 51, pp. 22020–22025, Dec. 2010.
[52] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, "Architecture of the human regulatory network derived from ENCODE data," *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.
[53] Y. Bar-Yam, D. Harmon, and B. de Bivort, "Attractors and Democratic Dynamics," *Science*, vol. 323, no. 5917, pp. 1016–1017, Feb. 2009.
[54] N. Polouliakh, R. Nock, F. Nielsen, and H. Kitano, "G-Protein Coupled Receptor Signaling Architecture of Mammalian Immune Cells," *PLoS ONE*, vol. 4, no. 1, p. e4189, Jan. 2009.
[55] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, "Evolution of bow-tie architectures in biology," *ArXiv14047715 Q-Bio*, Apr. 2014.
[56] N. Kashtan and U. Alon, "Spontaneous evolution of modularity and network motifs," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 39, pp. 13773–13778, Sep. 2005.
[57] K. Raman and A. Wagner, "The evolvability of programmable hardware," *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269 –281, Feb. 2011.
[58] A. Wagner, "Neutralism and selectionism: a network-based reconciliation," *Nat. Rev. Genet.*, vol. 9, no. 12, pp. 965–974, Dec. 2008.
[59] J. Masel and M. V. Trotter, "Robustness and Evolvability," *Trends Genet.*, vol. 26, no. 9, pp. 406–414, Sep. 2010.
[60] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, "The evolution of modularity in bacterial metabolic networks," *Proc. Natl. Acad. Sci.*, vol. 105, no. 19, pp. 6976–6981, May 2008.
[61] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, "Toolbox model of evolution of prokaryotic metabolic networks and their regulation," *Proc. Natl. Acad. Sci.*, vol. 106, no. 24, pp. 9743–9748, Jun. 2009.

[62] S. Akhshabi and C. Dovrolis, “The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, New York, NY, USA, 2011, pp. 206–217.
[63] M. Csete and J. Doyle, “Bow ties, metabolism and disease,” *Trends Biotechnol.*, vol. 22, no. 9, pp. 446–450, Sep. 2004.