Matchmaking between hairballs – insights from cross-disciplinary network comparison

Koon-Kiu Yan[1,2], Daifeng Wang[1,2], Anurag Sethi[1,2], Paul Muir[4,5], Robert Kitchen[1,2], Chao Cheng[6], Mark Gerstein[1,2,3]

1 Program in Computational Biology and Bioinformatics,

2 Department of Molecular Biophysics and Biochemistry,

3 Department of Computer Science,

4 Department of Molecular, Cellular and Developmental Biology,

5 Integrated Graduate Program in Physical and Engineering Biology,

Yale University, New Haven, CT 06520

6 Department of Genetics, Dartmouth School of Medicine, Hanover NH 03755

Abstract

Biological systems are complex. In particular, the interactions between molecular components often form inscrutable hairballs. While important progress in understanding biological networks has been made, criticism and concerns have also been raised. We argue that one way of untangling these hairballs is through cross-disciplinary network comparison, comparing networks from biology to those from other disciplines. Such comparisons enable the transfer of mathematical formalism between disciplines, precisely describing the abstract associations between entities and allowing us to directly apply a variety of sophisticated formalisms to biology. In addition, by examining in detail the mechanistic interactions in systems for which we have much day-to-day experience and then drawing analogies to more abstruse biological networks, network comparison allows us to leverage intuition from these systems to biology. Here, we illustrate how these comparisons benefit the field with few specific examples.

Introduction

A signature of biology in the "omic" era is the shift of attention away from the isolated interrogation of a few individual molecular components toward more holistic profiling of entire cellular systems [1]. Before molecular biologists studied protein complexes consisting of a few dozen proteins, but now proteomic methods are able to probe the interactions between thousands of proteins. Similarly, geneticists who would previously manipulate a single gene for functional characterization can now employ high-throughput techniques to study the relationships between all genes in an organism. In many cases, genome-scale information describing how components interact is captured best by a network representation [2]. However, cellular molecular networks probed by genomics and systems biology have such large size and complexity that gaining intuition or novel insights about biology is not always easy [3].

What approaches might help in deciphering these 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, subsequently, the characterization of structures of thousands of proteins. Thus, one may wonder whether ideas in other areas of science could help us with the "hairball challenge". While the influx of ideas related to reductionism mostly originated from subfields of physics and chemistry, to understand biology from a systems perspective we may benefit from new catalysts originating in disciplines as diverse as engineering, behavioral science, and sociology. These new ideas are centered on the concept of the network. In fact, comparisons and analogies are not new to biology. For instance, to illustrate the principles of selection Dawkins coined the *meme*, a unit carrying cultural information analogous to the gene in biology, which undergoes a similar form of selection [4]. This comparison has been further elaborated in the proto-field 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 sociology, may appear to be comparing apples to oranges. What kinds of comparison can truly deepen our understanding? The key is to find an optimal level of abstraction and simplification.

A spectrum of cellular descriptions

Given the complexity of the cell, a certain level of simplification is necessary for useful discussion. The description of cellular systems can be seen as a spectrum (Figure 1). On one extreme, there is a complete three or four-dimensional picture of how cellular components and molecules interact in space and time. On the other extreme, there is a simple parts list that enumerates each component without specifying any relationships. However, neither extreme currently gives the best understanding for the data we have to hand. The complete 4D picture of all the molecules in a cell is far too ambitious for the current state-of-the-art in data acquisition. Conversely, it is widely appreciated that the characteristics of a cellular system cannot be explained by the properties of individual components – the whole is greater than the sum of its parts -- and the data we have to hand is considerably richer than the parts list representation. The network representation conveniently spans these extremes, capturing some of the relationships between individual components in a flexible fashion, especially where connectivity rather than exact spatial location determines function.

There are two approaches to think about networks. In the purest form, a network is an abstract representation of the connections (edges) between constituents (nodes). As physical associations between components in all sort of complex systems can be viewed as networks, such an abstract approach to networks offer a common mathematical framework for different systems. In addition to physical associations, connections can be defined more loosely, by statistical association. This is exemplified by disease networks [6][7] 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 [8]. In general, 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 in the era of 'Big Data'.

The second way of thinking about networks aims to decipher the organization principles behind a complex system. The underlying network is assumed to be a backbone that captures the essence of the system. This is particularly the case for networks that capture the mechanistic interactions within systems -- for instance, the cellular networks resulting from protein-protein interactions and transcriptional regulation. Thinking of networks in a mechanistic way is a process of concretization – as opposed to the approach in abstract, associative networks. Concrete mechanistic networks aim to get closer to the complete 4D-picture. They are intended to describe and integrate many of the physical processes happening inside a living system -- for instance, the processing of information, the chemistry of metabolites, and the assembly of molecular machines -- and therefore focus on incorporating various details of interactions. Adding further mechanistic detail onto a simple nodes-and-edges skeleton can be visualized as decorating edges with direction, color, thickness, etc. However, incorporating too much detail makes the description intractable. In particular, the network formalism breaks down if we try to load spatial or temporal information as well as higher-order interactions onto the diagram. At a certain point, the actual 4D picture is required.

The two network approaches essentially complement each other. On one hand, thinking in an abstract fashion allows one to transfer mathematical formalism readily between disciplines. This can be beneficial for the biological sciences, in that it allows the application of formalism developed elsewhere to find fruitful application in biology. On the other hand, thinking mechanistically focuses more on the conceptual resemblances between networks. Comparison of appropriately matched networks may provide additional intuition into the interactions between molecular components of cells by examining analogous interactions in complex systems for which we have more day-to-day experience.

Abstract approach: comparison leverages mathematical formalism

Let us first focus on abstract-association approach to biological networks, whose power lies in the simplicity of its formalism..A key comparison point between various complex systems focuses on topology. The earliest and probably most important observation is that many networks organize themselves into scale-free architectures in which a majority of the nodes contain very few connections while a few (also called hubs) are highly connected (see Box) [9]. A surprisingly variety of networks exhibit scale-free architecture; for example the Internet, air transport routes, and many social networks [10].

Another important notion is that of a small-world network, in which any two nodes are on average separated by only a few steps (see Box). Scale-free networks are also small-world networks because hubs ensure that the distance between any two nodes is small [11][12]. For example, the presence of hubs in the airport network are designed to make it possible to travel between any two cities in the world within a short interval of time. However, not every small-world network is scale-free. An example of this is the mammalian brain, specifically the cerebral cortex. The cortical neuronal network is subdivided into more than 100 distinct, highly modular, areas [13] that are dominated by connections internal to each area, with only ~20% of all connections being between neurons in different areas [14]. Each area is considered to have a primary feature, for example in processing sensory or cognitive signals. The cortical architecture has a high degree of clustering and a small average path-length yet exhibits an exponential degree-distribution [15].

Koon-Kiu Yan 3/26/2015 3:13 PM
Deleted: .

The behavior of scale-free networks is dominated by a relatively small number of hubs and this ensures that such networks are resistant to random, accidental failures but are vulnerable to coordinated attacks against the hubs [16]. One see analogous examples of this robustness in many contexts: just as the Internet functions without any major disruptions even though hundreds of routers malfunction at any given moment, individuals belonging to the same species in general can tolerate considerable numbers of random mutations . However, a cell is not likely to survive if a hub protein is knocked out. For example, highly connected proteins in the yeast protein-protein interaction network are three times more likely to be essential than proteins with only a small number of links [17].

The number of connections of a node reflects its centrality in the network. There are more elaborate approaches to determining centrality than just counting neighbors – e.g. most famously, the original PageRank algorithm, underlying the Google search approach (see BOX). Besides, one can try to define centrality via network paths using such quantities as “betweenness” (See BOX). It has been reported that bottlenecks (nodes with high betweenness) in biological networks are more sensitive to mutations than the rest of the network, even more so than hubs for regulatory networks [18][19]. Centrality is a property of an individual node; apart from such individual properties, it is important to define medium-scale structures called network module (See BOX). A quantity dubbed modularity attempts to measure this, comparing the number of intra and inter module links in a network [20].

Koon-Kiu Yan 3/26/2015 2:55 PM
Deleted: , influence & bottleneck

Koon-Kiu Yan 3/26/2015 2:55 PM
Deleted: , modularity

A broad class of algorithms applied in biological and other data science maps properties or features to the nodes in a network (Figure 2) [21][22]. Apart from visualization, this mapping provides ways to organize the features. For instance, it has been reported that mapping somatic mutations to gene networks allow for stratification of cancer into subtypes [23]. Another important example is the inference of missing data using “guilt by association” -- the idea that nodes having similar associations in the network tend to have similar properties. In a social context, if your friends in an online social network recommmend a particular product, you are more likely to use this product and the advertisements you view online are personalized based on these recommendation systems [24]. In a biological context, it has been observed that cellular components within the same network module are more closely associated with the same set of phenotypes than components belonging to different modules [25]. As a result, one can infer the function of a gene or a non-coding element based on its neighbors. The diseases comorbidity network [26][27][28] makes use of a similar idea. In such networks, a node represents a disease and two diseases are connected if they are carried by a same patient as shown in medical claims data. Diseases (phenotypes) next to each other in the network may operate similarly on a molecular level.

A particular way to utilize genes with special features is based on the concept of ‘seed’ genes, a form of biological prior, to drive network creation. Instead of identifying hub genes based on connectivity, these hub genes are defined from the literature or experience as being causally implicated in a particular disease or phenotype. In one such example, genes implicated through copy-number variation in autism were used to cluster an expression network in healthy brain development in order to identify larger sets of putative autism-related genes as candidates for future investigation and diagnosis [29]. Such approaches are attractive as they maintain the power and flexibility of a network-based organizational scheme, but are grounded from the start in a particular biological context.

We can further exploit the structure of a network with data on underlying dynamical processes. As mentioned earlier in terms of recommendation systems, online retailers are interested in using purchase records to study how customers influence each other [30]. 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 relative temporal data. However, cellular processes happen too fast and thus most functional genomics experiments do not contain enough time-points. To overcome this drawback, data mining techniques such as matrix factorization are employed. For instance, given the genome-wide expression profile 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].

In addition to the actual dynamic processes occurring on a network, one can explore evolutionary change by comparing networks. In a biological context, pairs of orthologous genes (nodes) can be used to define conserved edges, called interologs and regulogs for the protein-protein interaction and regulatory networks, respectively. These can then be used to align networks from different species [32] and to detect conserved or specific functional modules [33] across species. Based on a large collection of aligned networks between species, a mathematical formalism has

been developed to measure the rewiring rate between networks using methods analogous to those quantifying sequence evolution. In this context, metabolic networks rewire at a slower rate compared to regulatory ones [34].

Criticisms on the abstract approach to networks

Despite an increasing number of studies applying networks in an abstract mathematical context, concerns have been raised. A major concern of network analysis comes from the criticism that statistical patterns (e.g. the scale-free degree distribution mentioned above) offer limited insights. Other examples of these patterns include the enrichment of network motifs (small recurrent subgraphs in a network). Statistical patterns suggest that network structures are potentially interesting; nevertheless, understanding their actual functioning requires studying the detailed dynamics of each constitutive part [35]. While this is a reasonable criticism, such patterns, whether interpretable or not, can often be used as features in machine-learning frameworks for biological and clinical predictions.

More fundamentally, depending on their background, different researchers may have different interpretation of the meaning of "understanding" [3]. In the ideal scenario, everyone wants to have a complete mechanistic picture. For perfectionist, in particular, networks have often prove frustrating because their abstract patterns don't always yield easily to precise molecular description. As an illustration, many stress that systems biology is the study of the behavior of complex biological organization and processes in terms of their underlying molecular constituents [36]. Therefore, it is instructive to employ a mechanistic approach try to get at some of desired mechanistic interpretation.

Mechanistic approach: comparison gives intuition into biological complexity

The previous section discussed insights gained by applying formalisms from various social and technological networks to biological networks. Such wide-ranging insights were possible only because in the abstract approach the detailed characterization of the nodes in the network was neglected. On the other hand, if details are added to the picture, insights about a system become more specific, and in a sense, more meaningful. However, it is typically harder to apply the same formalism equivalently to two different networks, characterized in this more detailed fashion. This situation is manifest when one tried to explain the scale-free degree distribution of various networks described above.

Different mechanistic intuitation for scale free structure

It is well known that the scale-free network topology can be arrived by two mathematically similar but conceptually different models. The first is the ~~celebrated~~ *preferential attachment* model [9]. The scenario can be illustrated by the hub and spoke system of the airline network. Every time a new airport is created, the airlines have to balance available resources and customer satisfaction[1], i.e., the cost of adding a new flight and customer comfort due to connectivity between the new airport and a larger number of other airports. The most efficient use of these limited resources occurs if the new airport connects to pre-existing hubs in the network as it reduces the average travel time to any airport in the entire system – due to the small-world nature of scale-free networks. The model is called preferential attachment because the newly created nodes prefer to connect to pre-existing hubs in the network.

The second model is *duplication-divergence; it* explains the evolution and growth of the World Wide Web. Here, a random pre-existing node and its associated edges are duplicated -- for example, to make a webpage for a new product listed on Amazon.com, one could use a template shared by an existing product [37]. After duplication, the content of two nodes and their

[1] This is arguably an idealistic view. Some may argue that airline companies do not care about customer satisfaction at all; only their revenues, and they consider customer satisfaction only within limits imposed by revenue maximization. Nevertheless, the motivation behind the airlines does not affect the conclusion of the model.

connections diverge but a proportion of their edges are likely to be shared [38]. In fact, as its name suggests, the same duplication-divergence mechanism can describe the patterns and occurrence of “memes” in online media [39]. Such a duplication-divergence model is in a sense equivalent to the preferential attachment model since it is more likely for a hub to increase its connectivity, simply because it is more likely for any node to be attached to a neighbor getting duplicated. However, it provides more intuition for biological networks via comparison. As gene duplication is one of the major mechanisms driving the evolution of protein families, scale-free behavior in the protein-protein interaction network was proposed to arise via duplication-divergence [40]. Of course, no model is perfect. Upon analyzing the structural interfaces involved in protein-protein interactions, one observes that there are great differences in hubs that interact with many proteins by reusing the same interface versus those that simultaneously use many different interaction interfaces. The duplication-divergence model only applies to the former situation (with the duplicated protein reusing the same interface as its parent) [41].

Thus, many networks that exhibit similar topologies are the result of significantly different underlying growth mechanisms. Specifically, in the case of scale free networks, there exists a common topological property but somewhat different mechanistic explanation in many different domains (e.g. airline networks vs gene networks). Some of the domains share the same mechanistic explanation -- i.e. the scale-free structure in both protein-protein interaction and web-link networks can be explained by duplication and divergence. Moreover, this latter commonality provides additional intuition about the biological network through comparison to the more commonplace web network, which is conceptually much easier to understand.

More intuition from social networks

The ability to gain intuition about the often-arcane world of molecular biology by comparison to commonplace systems is even more clear-cut when considering social networks, where people have very strong intuition for how a "system" can work. Transferring understanding of organizational hierarchy to biology is a good example of this (Figure 3). Many biological networks, such as those involved in transcriptional regulation, have an intrinsic direction of information flow, forming a natural but loose hierarchy. Likewise, society has many hierarchical structures -- e.g. a militarily command chain or a corporate "org-chart" [42]. In the purest form of the military hierarchy, multiple individuals of lower rank each report to a single individual of a higher rank and there are fewer and fewer individuals on the upper levels, eventually culminating in a single individual commanding an entire army. This structure naturally leads to information flow bottlenecks as all the orders and information related to many low-rank privates must flow through a limited number of mid-level majors. In a biological hierarchy of transcription factors (TFs), one sees a similar pattern, with bottlenecks in the middle, and in many cases, the bottlenecks create vulnerabilities. Indeed, it has been shown in knockout experiments that many of the bottleneck nodes in biological networks are essential [18]. Structurally, hierarchies can insulate themselves from mid-level bottleneck vulnerability by allowing middle managers to co-regulate those under them. This eases information flow bottlenecks in an obvious way -- if one major gets knocked out, the privates under them can receive orders from a second major. Moreover, many commentators have pointed out that, in order to function smoothly, it is imperative for corporate hierarchies that middle managers work together [43]. Strikingly, biological regulatory networks employ a similar strategy by having two mid-level TFs co-regulate targets below them, and this degree of co-regulation increases with overall organism complexity [44]. Thus, one can get an intuition for the reason behind a particular biological structure through analogies to a commonplace social situation.

Moreover, further comparison provides easy intuition into the biological characteristics of regulators at different levels in the hierarchy. Conventionally, one expects the CEOs of companies to gather information from all their sources and make the widest ranging and influential decisions in the company. One also stereotypically expects people at the top of conventional social hierarchies to be the most “conservative” and resistant to change. Likewise, TFs at the top of the hierarchy tend to be more evolutionarily conserved. They are more connected in the protein-protein interaction network as they modulate gene expression based

upon internal and external stimuli through these interactions [42][45], and to be more influential in driving gene expression [46]. Rewiring the TF network at its upper levels also tends to have a larger effect on cell proliferation and survival [47].

More intuition from technological systems: connectivity and constraints

Lying at the heart of deciphering biological networks is the mapping between architecture and function. As it is often hard to define "function" in complex biological settings, comparison with simple technological or engineered components with basic and well-defined functions is particularly insightful [48]. Sarpeshkar and colleagues have explored the similarities between the biochemical reactions within cells and electron flow in analog circuits. These similarities have enabled the application of intuitive electronic circuit diagrams to describe the processes underlying transcription factor networks. In this analogy chemical concentrations are represented as electronic currents. For example, mRNA molecules can be thought of as accumulating on a capacitor while a resistor represents mRNA degradation. The analogy extends beyond simply intuitive representations since the mathematical formalisms describing electron flow in subthreshold transistors can be adapted to capture the dynamics of chemical reactions. Thus, this comparison allows us to potentially connect diagrams and mathematical models developed for electronics to transcription [49]. Similar ideas have been employed to map a transcriptional regulatory pathway to a combination of logic gates [50].

A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [51]. Robustness is a preferred design objective because it makes a system tolerant to stochastic fluctuations, from either intrinsic or external sources. Modularity, on the other hand, makes a system more evolvable as upgrades are feasible by simply replacing a particular module. For instance in software design, modular programming that separates the functionality of a program into independent parts connected by interfaces is widely practiced [52]. The same is true for biological networks because modules can be readily reused to adapt new functions.

To illuminate how biological and technological networks and systems share the common design principles, it is important to think about how both change, as both are adaptive. Manmade networks like roadways and electronic circuits are thought to change according to the plan of rational designers. In contrast, biological networks are thought to change in response to random mutations and then for the successful changes to be selected. This is analogous to the work of a tinkerer, rather than an intelligent designer. Nevertheless, the distinction is not clear-cut. There are plenty of examples showing that many of man's great innovations are the result of trial and error, and all technological systems are subjected to selection such as user requirements [53].

In a sense, we could picture that both the engineer and tinkerer are working on an optimization problem with similar underlying design objectives, but taking different views when balancing constraints [54][55]. For example, in biological networks, more connected components (as measured by their degree or betweenness) tend to be under stronger constraint than less connected ones. This is evident in numerous studies that have analyzed the evolutionary rate of genes in many networks (e.g. protein interaction and transcription regulatory networks) in many organisms (e.g humans, worms, yeast, *E. coli*) using many different metrics of selection (e.g. variation within a population or dN/dS for fixed differences) [56][57][58][59]. One's intuition here is obvious: biological systems seek to decentralize functionality, minimizing average connectivity on nodes and making the system robust to a random mutation. However, this architecture requires a few hubs to connect everything up and these more connected components are particularly vulnerable. Is this finding true in general? And if not, why?

Software systems provide insight into this question: software engineers tend to reuse certain bits of code, leading to the sharing of components between modules, arriving at highly connected components. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of its functions (routines) is distributed in a bimodal fashion; the more central components in the underlying network (call graph) are updated often. These patterns

seem to hold for other software systems. For instance, in package-dependency network of the statistical computing language 'R', packages that are called by many others are updated more often (Figure 4). In other words, unlike biological networks whose hubs tend to evolve slowly, hubs in the software system evolve rapidly. What's the implication? As a piece of code is highly called by many disparate processes – i.e. modules tend to overlap -- intuitively one would expect that the robustness of software would decrease. Our first intuition is that an engineer should not meddle too much with highly connected components, However, there is another factor to consider: rational designers may believe that they can modify a hub without disrupting it (i.e. the road planner thinks construction is possible in Manhattan without too much disruption) -- in contrast to a situation where random changes dominate. Moreover, 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 (and maintenance). This situation is again analogous to road networks: one sees comparatively more construction on highly used bottlenecks (e.g. the George Washington Bridge) compared to out-of-the-way thoroughfares. The discrepancy between tinkerer and engineer suggests that, as an optimization process, no approach optimizes all objectives (robustness and modularity 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 [60][61].

The concept of connectivity associating with constraint is evident; there are many highly conserved genes that are very well connected in physical protein-protein and regulatory networks. Mutations in these genes are more likely to be deleterious, resulting in a loss-of-function, and it is therefore useful to prioritize these as potential disease drivers [62]. The results are therefore extremely useful for therapeutics, in which a drug targeting a highly connected target can have a very efficient effect on an entire cell, albeit often with the sacrifice of low specificity. However, the measurement of connectivity/constraint depends on the cellular process. In regulatory networks and similar systems involving information transfer, this is often better conceptualized in terms of bottlenecks, while in protein-protein interactions and similar systems involved with signaling cascades it is often better to consider hubs. An example of a chemically exploitable bottleneck in the regulatory network is the bacterial ribosome, which is the target of most antibiotics that broadly inhibit protein translation leading to the rapid death of the organism [63]. A subtler, but no less useful, route to the inhibition of protein translation is through hub proteins such as mTOR and other key gates in cellular signaling cascades that are actively exploited in therapies for ailments as diverse as breast cancer [64] to depression [65]. Nevertheless, there is an exception for the connectivity versus constraint observation: in metabolic networks, highly connected hub genes have more duplicated copies. Decentralization is achieved by a pair of duplicates compensating each other, they are in general more tolerant to loss-of-function mutations [66].

Conclusion

Biology is a subject with a strong tradition of utilizing comparative methods. One hundred years ago, biologists compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then various 'omes' across species. Perhaps, we should extend further by comparing networks in biology to those in other disciplines. In fact, efforts have already been made along this direction (Figure 5). We have described how abstract approaches that focus on simple connections between entities could allow the application of mathematical formalisms across disciplines. We then showed how mechanistic details can be placed onto these simple networks and enable them to better explain a real process such as transcriptional regulation or software code development. In this case, the networks are often too detailed to allow for direct transfer of formalisms. Nevertheless, one can gain meaningful intuition about a biological system through comparing it to a more commonplace network such as a social system using a similar mechanistic description.

Seeking comparison between biological networks, social networks, and technological networks may echo the long-time ambition of finding universality in all complex systems. Indeed, the discovery of the scale-free degree distribution in many different networks initially hinted at such direction. Very soon it was argued that a universal model does not exist: there are biological networks whose degree distributions do not follow a simple power-law [67][68]; there are simply

Koon-Kiu Yan 3/26/2015 2:38 PM
Deleted: fantasy

too many ways to generate a network with a broad degree distribution [69][70][71]. Though scale-free distribution is not universal, the concept of universality has a long tradition in statistical physics literature, and the efforts of characterizing the underlying mechanisms of complex systems by a few scaling or critical exponents should be appreciated. In fact, there are still many relatively open questions. For examples, as building blocks of networks, different network motifs exhibit different occurrence frequencies [72]. It is quite remarkable that the transcriptional regulatory networks constructed in different cell lines as well as different species exhibit similar patterns [73][74]. In general, despite a lack of fundamental laws of networks, we believe that one should not be disappointed or simply turn away from network biology. As suggested by some of the examples in this essay, understanding the differences between biological networks and networks from other disciplines may be as rewarding as finding the commonality.

What's next? We envision that these cross-disciplinary network comparisons will become increasingly common as a result of data growth. One area that is especially ripe for comparison is multiplex networks, which concatenate networks to form a multiplex structure [75][76]. This framework is commonly used in social science in which an individual may participate in multiple social circles (e.g. family, friends, and colleagues), or in an online setting: Facebook, LinkedIn, and Twitter; but it has not been very well explored in biology. Nevertheless, the fundamental structure of biological data now extends beyond a single network to multiplex structures: the multiple layers could be formed by different categories of relationships (co-expression, genetic interactions, etc.), Furthermore, biological regulation occurs at multiple levels: transcriptional, post-transcriptional, and post-translational regulation in a manner in analogous to a city with electrical networks, water pipes, and phone lines. We are looking forward to multi-layers network formalisms developed in other contexts being applied to biology. Apart from leveraging the ideas and methods developed in multiple disciplines through comparison, we can even imagine that comparisons will ultimately lead to real connections (i.e. not analogies) between biological networks and those in other disciplines. For instance, biologists and sociologists have started to investigate if there is any connection between genomics information and sociological information, say whether phenotypes or genotypes are correlated in friendship networks [77].

Acknowledgements
We acknowledge Zeynep Gümüş for the courtesy of providing Figure 2, and Leonidas Salichos for critical reading of the manuscript and insightful discussion. We acknowledge the funding from NIH.

Figure Captions

Figure 1.
A spectrum of cellular descriptions. From left to right. Networks help reveal and convey the relationships between components of a biological system. Different levels of information can be represented using a network. At an abstract level, a network can denote associations between various nodes. More details, such as excitatory and inhibitory regulatory relationships, can then be layered on top of this basic network. As additional information about the nodes and the relationships between them is added, the network begins to resemble the real world entity it models. For example, the addition of 3D structural information and temporal dynamics onto a network of molecular machine components leads it to more closely resemble the molecular machine itself.

Figure 2.
Intuitions guide visualizations of a complex hairball. A mechanistic network with multiple kinds of edges (protein-protein interactions, metabolic reactions, transcription regulations, etc.) forms an ultimate hairball (left). The hairball is then visualized by scaling the size of nodes by the degree of genes (right). The red nodes are essential, and the blue nodes are loss-of-function-tolerant. The network layout was generated by Vaja Liluashvili and Zeynep H Gümüş, using iCAVE [22].

Figure 3.

Koon-Kiu Yan 3/26/2015 2:43 PM
Deleted: perspective

Koon-Kiu Yan 3/26/2015 2:40 PM
Deleted: very much

Koon-Kiu Yan 3/18/2015 4:24 PM
Deleted: Nevertheless, discouraging the search of fundamental laws is not healthy for science. The concept of universality has a long tradition in statistical physics literature, and the perspective of characterizing the underlying mechanisms of complex systems by a few scaling or critical exponents should very much be appreciated. In fact, apart from the degree distribution, there are still many relatively open questions. For examples, as building blocks of networks, different network motifs exhibit different occurrence frequencies [73]. It is quite remarkable that under proper normalization, the transcriptional regulatory networks constructed by experiments in different cell lines as well as different species exhibit similar patterns [74][75]. Whether it is an interesting technical artifact or an insightful clue on cellular information processing is still unknown.

Comparison between the hierarchical organizations in social networks versus biological networks illustrates design principles of biological networks. The hierarchical organization in biological networks resembles the chain of command in human society, like in military context. The top panel shows a conventional autocratic military hierarchy. The structure is intrinsically vulnerable in the sense that if a bottleneck agent (star) is disrupted, information propagation breaks down. The introduction of cross-links (blue) avoids the potential problem (middle panel) because the private at the bottom can then take commands from two different superiors above. The bottom panel shows the hierarchical organization of a biological network, with the existence of cross-links between pathways. These observations reflect a democratic hierarchy as opposite to an autocratic organization.

Figure 4.
Different evolutionary patterns in biological networks versus technological networks. The left shows the protein-protein interactions network in human [78], whereas the right is the R package dependency network specifying the proper function of a package (node) depends on (edge) the installation of another. 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). The R package dependency network consists of all the available packages (5711) via R studio at October 2014.

Figure 5.
Interdisciplinary network comparison. A lot of papers have addressed the similarity and difference between biological networks (circle) and networks in social/technological systems (squares). Here we represent all these comparison in the form of a network, where an edge associated with references represents a network comparison in a specific context (color). Moreover, these comparisons can take place in terms of abstract approaches where formalism is used equivalently in two domains (dotted lines) or mechanistic approaches where one only seeks analogy between disciplines (solid lines).

Box. Network terminology

Degree: the number of neighbors of a node.

Scale free networks: The degree distribution of the network is a statistical property that can be used to understand some of the organizing principles of the network. The degree distribution of a random network is a Poisson distribution. Most real world networks including biological networks are organized in the form of scale-free networks that contain a small number of hubs that are highly connected in the network. The degree distribution in a scale free network is better modeled as a power law distribution. Hubs in scale free network also lead to the formation of small-world networks.

Betweenness: the number of paths passing a node. Similar in spirit to heavily used bridges, highways, or intersections in transportation networks, a few centrally connected nodes funnel most of the paths between different parts of the network. High betweenness nodes are referred to as bottlenecks and removal of these nodes could reduce the efficiency of communication between nodes [79].

Small World: A small world network is a kind of network in which the distance between nodes in the network is much smaller than the size of the network even though most nodes are not connected to one another. Typically, the average distance between any two nodes in a small world network scales as the logarithm of the number of nodes in the network.

Cliques: Cliques are defined as sub-networks in the graph that are completely connected, i.e., every pair of nodes in a clique contain an edge connecting them. Cliques form a single cohesive

group in social networks and such groups tend to act together. Similarly, a clique can be formed from a large biomolecular complex such as a ribosome that functions as a single unit. This property of cliques has been used to find missing edges to predict function of biomolecules.

Modules (Community Structure of networks): Most real world networks can be divided into smaller modules that have a large density of internal edges but relatively fewer edges connect nodes from different modules. For instance 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. Similarly in a biological context, a large number of biological components can form a single functional macromolecular complex such as the ribosome. A wide variety of methods have been developed to uncover the modular structure of networks. Most of these methods are based on optimizing the modularity of the network that compares the number of intra and inter module links within the network.

PageRank algorithm: PageRank is a prominent example of measuring the importance of a node by taking into account the importance of its neighbors. Originally developed in social network analysis [80], 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 - i.e. being linked to by higher ranking nodes has a larger impact on the document's ranking. This algorithm has been applied to food webs to prioritize species that are in danger of extinction [81] and has also been used to rank marker genes and predict clinical outcome for cancers [82].

References

[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] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, and A.-L. Barabási, "Uncovering disease-disease relationships through the incomplete interactome," *Science*, vol. 347, no. 6224, p. 1257601, Feb. 2015.

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

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

[10] A.-L. Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. New York: Plume, 2003.

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

Koon-Kiu Yan 3/26/2015 3:06 PM
Deleted: In biology, networks play an important role important role in gene prioritization, an essential process for disease-gene discovery because of limited validation and characterization resources [82]. For example, network properties (e.g. hubbiness) have been used to distinguish functionally essential and loss-of-function tolerant genes [65]. One could also prioritize uncharacterized genes based on how they are connected to characterized ones. If a gene, say, is one step away from a group of genes associated with a particular disease, it is very likely that it too is associated with this disease. The influence of a node may not be restricted to its nearest neighbors; network flow algorithms are widely used to examine long-range influence [83][84]. For instance, in a social science context, researchers use cascade-structured models to capture the information propagation on blog networks, predicting a blog's popularity [85].

Unknown
Field Code Changed

Koon-Kiu Yan 3/26/2015 2:30 PM
Deleted: [79]

Unknown
Field Code Changed

Koon-Kiu Yan 3/26/2015 2:30 PM
Deleted: [80]

Unknown
Field Code Changed

Koon-Kiu Yan 3/26/2015 2:30 PM
Deleted: [81]

Koon-Kiu Yan 3/26/2015 3:11 PM
Deleted: In biology, networks play an important role in gene prioritization, an essential process for disease-gene discovery because of limited validati ... [1]

Koon-Kiu Yan 3/26/2015 2:50 PM
Deleted: Box 2. Network topology ... [2]

Koon-Kiu Yan 3/26/2015 3:15 PM
Deleted: ... [3]

Koon-Kiu Yan 3/26/2015 3:15 PM
Formatted ... [4]

Unknown
Field Code Changed

Koon-Kiu Yan 3/26/2015 2:30 PM
Deleted: [1] M. Baker, "Big biology: The 'omes puzzle," *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013. ... [5]

[12] 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.
[13] D. C. V. Essen, M. F. Glasser, D. L. Dierker, and J. Harwell, "Cortical Parcellations of the Macaque Monkey Analyzed on Surface-Based Atlases," *Cereb. Cortex*, vol. 22, no. 10, pp. 2227–2240, Oct. 2012.
[14] N. T. Markov, M. Ercsey-Ravasz, D. C. V. Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy, "Cortical High-Density Counterstream Architectures," *Science*, vol. 342, no. 6158, p. 1238406, Nov. 2013.
[15] D. S. Modha and R. Singh, "Network architecture of the long-distance pathways in the macaque brain," *Proc. Natl. Acad. Sci.*, vol. 107, no. 30, pp. 13485–13490, Jul. 2010.
[16] null Albert, null Jeong, and null Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000.
[17] 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.
[18] 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.
[19] P. V. Missiuro, K. Liu, L. Zou, B. C. Ross, G. Zhao, J. S. Liu, and H. Ge, "Information Flow Analysis of Interactome Networks," *PLoS Comput Biol*, vol. 5, no. 4, p. e1000350, Apr. 2009.
[20] 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.
[21] R. Saito, M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, S. Lotia, A. R. Pico, G. D. Bader, and T. Ideker, "A travel guide to Cytoscape plugins," *Nat. Methods*, vol. 9, no. 11, pp. 1069–1076, Nov. 2012.
[22] V. Liluashvili, A. Gabow, M. Wilson, J. Sun, and Z. Gümüş, "iCAVE: immersive 3D visualization of complex biomolecular interaction networks."
[23] M. Hofree, J. P. Shen, H. Carter, A. Gross, and T. Ideker, "Network-based stratification of tumor mutations," *Nat. Methods*, vol. 10, no. 11, pp. 1108–1115, Nov. 2013.
[24] 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.
[25] 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.
[26] C. A. Hidalgo, N. Blumm, A.-L. Barabási, and N. A. Christakis, "A Dynamic Network Approach for the Study of Human Phenotypes," *PLoS Comput Biol*, vol. 5, no. 4, p. e1000353, Apr. 2009.
[27] A. Chmiel, P. Klimek, and S. Thurner, "Spreading of diseases through comorbidity networks across life and gender," *New J. Phys.*, vol. 16, no. 11, p. 115013, Nov. 2014.

[28] C.-C. Liu, Y.-T. Tseng, W. Li, C.-Y. Wu, I. Mayzus, A. Rzhetsky, F. Sun, M. Waterman, J. J. W. Chen, P. M. Chaudhary, J. Loscalzo, E. Crandall, and X. J. Zhou, "DiseaseConnect: a comprehensive web server for mechanism-based disease-disease connections," *Nucleic Acids Res.*, vol. 42, no. Web Server issue, pp. W137–146, Jul. 2014.
[29] A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, and M. W. State, "Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism," *Cell*, vol. 155, no. 5, pp. 997–1007, Nov. 2013.
[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] R. Singh, J. Xu, and B. Berger, "Global alignment of multiple protein interaction networks with application to functional orthology detection," *Proc. Natl. Acad. Sci.*, vol. 105, no. 35, pp. 12763 –12768, 2008.
[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] 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.
[35] P. J. Ingram, M. P. Stumpf, and J. Stark, "Network motifs: structure does not determine function," *BMC Genomics*, vol. 7, no. 1, p. 108, May 2006.
[36] M. W. Kirschner, "The Meaning of Systems Biology," *Cell*, vol. 121, no. 4, pp. 503–504, May 2005.
[37] K. Evlampiev and H. Isambert, "Conservation and topology of protein interaction networks under duplication-divergence evolution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 29, pp. 9863–9868, Jul. 2008.
[38] R. Pastor-Satorras, E. Smith, and R. V. Solé, "Evolving protein interaction networks through gene duplication," *J. Theor. Biol.*, vol. 222, no. 2, pp. 199–210, May 2003.
[39] 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.
[40] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, "Modeling of Protein Interaction Networks," *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.
[41] 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.

[42] 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.
[43] 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.
[44] 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.
[45] C. Cheng, K.-K. Yan, W. Hwang, J. Qian, N. Bhardwaj, J. Rozowsky, Z. J. Lu, W. Niu, P. Alves, M. Kato, M. Snyder, and M. Gerstein, "Construction and Analysis of an Integrated Regulatory Network Derived from High-Throughput Sequencing Data," *PLoS Comput Biol*, vol. 7, no. 11, p. e1002190, Nov. 2011.
[46] D. H. Erwin and E. H. Davidson, "The evolution of hierarchical gene regulatory networks," *Nat. Rev. Genet.*, vol. 10, no. 2, pp. 141–148, Feb. 2009.
[47] N. Bhardwaj, P. M. Kim, and M. B. Gerstein, "Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators," *Sci. Signal.*, vol. 3, no. 146, p. ra79, 2010.
[48] 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.
[49] R. Sarpeshkar, "Analog synthetic biology," *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 372, no. 2012, p. 20130110, Mar. 2014.
[50] D. Wang, K.-K. Yan, C. Cheng, J. Rozowsky, and M. Gerstein, "Loregic – A method to characterize the cooperative logic of regulatory factors," *PLoS Comput. Biol.*, in press.
[51] U. Alon, "Biological Networks: The Tinkerer as an Engineer," *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.
[52] 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.
[53] 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.
[54] J. C. Doyle and M. Csete, "Architecture, constraints, and behavior," *Proc. Natl. Acad. Sci.*, p. 201103557, Jul. 2011.
[55] S. Akhshabi, S. Sarda, C. Dovrolis, and S. Yi, "An explanatory evo-devo model for the developmental hourglass," *F1000Research*, Dec. 2014.
[56] 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.
[57] H. B. Fraser, D. P. Wall, and A. E. Hirsh, "A simple dependence between protein evolution rate and the number of protein-protein interactions," *BMC Evol. Biol.*, vol. 3, p. 11, May 2003.
[58] G. Butland, J. M. Peregrín-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, M. Davey, J. Parkinson, J.

Greenblatt, and A. Emili, "Interaction network containing conserved and essential protein complexes in Escherichia coli," *Nature*, vol. 433, no. 7025, pp. 531–537, Feb. 2005.
[59] 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.
[60] A. D. Lander, "Pattern, growth, and control," *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.
[61] 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.
[62] E. Khurana, Y. Fu, V. Colonna, X. J. Mu, H. M. Kang, T. Lappalainen, A. Sboner, L. Lochovsky, J. Chen, A. Harmanci, J. Das, A. Abyzov, S. Balasubramanian, K. Beal, D. Chakravarty, D. Challis, Y. Chen, D. Clarke, L. Clarke, F. Cunningham, U. S. Evani, P. Flicek, R. Fragoza, E. Garrison, R. Gibbs, Z. H. Gümüş, J. Herrero, N. Kitabayashi, Y. Kong, K. Lage, V. Liluashvili, S. M. Lipkin, D. G. MacArthur, G. Marth, D. Muzny, T. H. Pers, G. R. S. Ritchie, J. A. Rosenfeld, C. Sisu, X. Wei, M. Wilson, Y. Xue, F. Yu, E. T. Dermitzakis, H. Yu, M. A. Rubin, C. Tyler-Smith, and M. Gerstein, "Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics," *Science*, vol. 342, no. 6154, p. 1235587, Oct. 2013.
[63] D. N. Wilson, "Ribosome-targeting antibiotics and mechanisms of bacterial resistance," *Nat. Rev. Microbiol.*, vol. 12, no. 1, pp. 35–48, Jan. 2014.
[64] S. Vinayak and R. W. Carlson, "mTOR inhibitors in the treatment of breast cancer," *Oncol. Williston Park N*, vol. 27, no. 1, pp. 38–44, 46, 48 passim, Jan. 2013.
[65] H. M. Abelaira, G. Z. Réus, M. V. Neotti, and J. Quevedo, "The role of mTOR in depression and antidepressant responses," *Life Sci.*, vol. 101, no. 1–2, pp. 10–14, Apr. 2014.
[66] 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.
[67] A. Clauset, C. Shalizi, and M. Newman, "Power-Law Distributions in Empirical Data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009.
[68] R. Tanaka, T.-M. Yi, and J. Doyle, "Some protein interaction data do not exhibit power law statistics," *FEBS Lett.*, vol. 579, no. 23, pp. 5140–5144, Sep. 2005.
[69] M. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemp. Phys.*, vol. 46, no. 5, pp. 323–351, Sep. 2005.
[70] E. Fox Keller, "Revisiting 'scale-free' networks," *BioEssays*, vol. 27, no. 10, pp. 1060–1068, 2005.
[71] G. Lima-Mendez and J. van Helden, "The powerful law of the power law and other myths in network biology," *Mol. Biosyst.*, vol. 5, no. 12, pp. 1482–1493, Nov. 2009.
[72] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, vol. 298, no. 5594, pp. 824–827, Oct. 2002.

[73] S. Neph, A. B. Stergachis, A. Reynolds, R. Sandstrom, E. Borenstein, and J. A. Stamatoyannopoulos, "Circuitry and Dynamics of Human Transcription Factor Regulatory Networks," *Cell*, vol. 150, no. 6, pp. 1274–1286, Sep. 2012.
[74] A. P. Boyle, C. L. Araya, C. Brdlik, P. Cayting, C. Cheng, Y. Cheng, K. Gardner, L. W. Hillier, J. Janette, L. Jiang, D. Kasper, T. Kawli, P. Kheradpour, A. Kundaje, J. J. Li, L. Ma, W. Niu, E. J. Rehm, J. Rozowsky, M. Slattery, R. Spokony, R. Terrell, D. Vafeados, D. Wang, P. Weisdepp, Y.-C. Wu, D. Xie, K.-K. Yan, E. A. Feingold, P. J. Good, M. J. Pazin, H. Huang, P. J. Bickel, S. E. Brenner, V. Reinke, R. H. Waterston, M. Gerstein, K. P. White, M. Kellis, and M. Snyder, "Comparative analysis of regulatory information and circuits across distant species," *Nature*, vol. 512, no. 7515, pp. 453–456, Aug. 2014.
[75] 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.
[76] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.
[77] J. H. Fowler, J. E. Settle, and N. A. Christakis, "Correlated genotypes in friendship networks," *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.
[78] 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.
[79] 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.
[80] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
[81] 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.
[82] 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.