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. Here we argue that one way of untangling these hairballs is through cross-disciplinary network comparison, matching biological networks with those from other disciplines. On the one hand, such comparison allows the transfer of mathematical formalism between disciplines, precisely describing the abstract associations between entities. This allows us to directly apply sophisticated formalisms developed elsewhere to biology (e.g. related to network growth and scaling). On the other hand, by examining in detail the mechanistic interactions in systems for which we have much day-to-day experience and then drawing analogies to the more abstruse biological networks, network comparison allows us to leverage intuition from these systems to biology (e.g. leveraging intuitions about bottlenecks in management hierarchies to understand the structure of transcriptional regulatory networks).

Introduction

A signature of biology in the "omic" era is the shift of attention from a few individual components to a collection of constituents [1]. In the past structural biologists studied the binding of a few proteins, but now they 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 functional genomic techniques to study the relationships between all genes. In many cases, genome-scale information describing how components interact are captured by a network representation [2]. However, given the astonishing size and complexity of the cellular molecular networks probed by genomics or systems biology, gaining easy intuition about biology from these hairballs is not guaranteed [3].

What approaches might help in deciphering 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 with the "hairball challenge". In this essay, we argue that, while the influx of ideas in the age of reductionism mostly originated from subfields of physics and chemistry, to understand biology via a systems perspective, we can further benefit from new catalysts coming from disciplines as diverse as engineering, behavioral science and sociology. These new ideas are centered on the concept of network.

Comparison and analogies are not new to biology. For instance, to illustrate the principles of selection Dawkins came up with the idea of a meme, which is a unit carrying cultural ideas analogous to the gene in biology [4]. This 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 sociology, may sound like comparing apples to oranges. What kinds of comparison can truly deepen our understanding?

To address this, it is useful to put various descriptions of a cellular system on a spectrum, in terms of abstraction and simplification.

A spectrum of cellular descriptions

Given the complexity of a cell, a certain level of simplification is necessary for useful discussion. The depth of 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 leads to a full understanding and intuition for the system as a whole. 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. To describe the full picture, one would need the 3D structures of everything in the genome as well as representation of their dynamical movements. This level of detail is too ambitious for the current state-of-the-art in data acquisition.

The network representation sits conveniently between these extremes. It captures some of the relationships between the components on the parts list in a flexible fashion, especially those where connectivity rather than exact location determines the consequence. There are two particularly useful ways to think about networks: association networks and mechanistic networks. Association networks essentially represent a process of abstraction in which entries are connected via a defined mathematical relationship. This could, for instance, be a statistical, rather than physical, relationship between nodes. This is exemplified by 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.

On the other hand, mechanistic networks represent a process of concretization. Unlike abstract association networks that move away from the complete 4D-picture, concrete mechanistic networks aim to more completely describe it. 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. Note, any mechanistic interaction can be simplified and abstracted as a mathematical association. However, the converse is not always true.

Adding further mechanistic detail onto a simple nodes-and-edges skeleton can be visualized as decorating edges with directionality, color, thickness etc. However, incorporating too much detail makes the description intractable. That is, the network formalism breaks down if we try to load spatial or temporal information as well as higher-order interactions onto the diagram. At certain point, the actual four-dimensional picture is required.

Because of their simplicity, abstract association networks allow one to transfer mathematical formalism readily between disciplines. This can beneficial for the biological sciences, in that it allows the application of formalism developed elsewhere to easily find fruitful application in biology. On the other hand, mechanistic networks can serve as the skeletons for describing complex systems in detail. In this case, because of system-specific details, it is not possible to transfer entire formalisms; instead, one focuses more on the conceptual, rather than topological, resemblances. Thus, 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.

Association Networks: Comparison leverages mathematical formalism

The power here of the network formalism lies in its simplicity. In the era of Big Data, the 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. This is, of course, particularly true for abstract association networks.

Formalism focusing on network topology

A key application focuses on the organization principles of various complex systems. 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 (edges) while a few (also called hubs) are highly connected [7]. A surprisingly large number of networks that one comes into contact with have a scale-free architecture – e.g. the Internet, air transport routes and many social networks [8].

The behavior of scale-free networks is dominated by a relatively small number of nodes and this ensures that such networks are resistant to random accidental failures but are vulnerable to coordinated attacks at hub nodes [9]. In other words, just as 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. 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 3-fold more likely to be essential than proteins with only a small number of links [10].

A scale-free network is a kind of small-world network because hubs ensure that the distance between any two nodes is small [11][12]. For example, the presence of hubs in the airport network makes it possible to travel between any two cities in the world within a short interval of time. However, not every small world network has to be scale-free. An example of a prominent small-world network that is not scale-free is the mammalian cerebral cortex. The cortical neuronal network is subdivided into more than 100 distinct, highly modular, areas [13] that are dominated by connections that are 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 small path-length and exhibits an exponential degree-distribution [15].

While counting the number of neighbors is very useful in determining the centrality of a node, a more sophisticated way to define centrality is to take into account the importance of neighbors. The PageRank algorithm is a prominent example of this approach. Faced with a search query, Google must decide which set of results to rank higher and place on the first results page. Originally developed in social network analysis [16], 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 [17] and has also been used to rank marker genes and predict clinical outcome for cancers [18].

A second method of measuring a node's centrality is based on the number of paths passing through it -- its "betweenness". 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. These are referred to as bottlenecks and removal of these nodes could reduce the efficiency of communication between nodes [19] (increasing their effective distance). Indeed, it has been reported that bottlenecks in biological networks are more sensitive to mutations than the rest of the network, even more so than hubs for regulatory networks [20][21].

Apart from measuring degrees and paths, 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 naturally form strong ties and, in the extreme, constitute a single cohesive group (or a fully connected graph, or clique). Analogous to these closely-knit social groups, a large number of biological components can form a single functional macromolecular complex such as the ribosome. More generally, a common feature of a large number of social, technological and biological networks is that they are composed of modules such that nodes within the same module have a larger number of connections to each other compared to nodes belonging to different modules. A quantity dubbed modularity attempts to measure this, comparing the number of intra and inter module links in a network [22].

Formalisms focusing on the interplay between topologies and the properties of nodes
Networks are useful in data science because they can be used as a reference for mapping additional properties or features of different nodes. An important example is the inference of missing data using "guilt by association" -- the idea that nodes having similar associations in the network tend to be similar in properties. For example, in a social context, if your friends in an online social network use a particular product, you are more likely to use this product and the advertisements you view online are personalized based on these recommendation systems [23]. 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 [24]. Furthermore, modules within gene co-expression networks tend to contain genes in the same biological pathway or have similar functions [25]. As a result, one can infer the function of a gene or a non-coding element based on its neighbors in the underlying network.

In this context, networks play an important role in gene prioritization, an essential process for disease-gene discovery because of limited validation and characterization resources [26]. For example, network properties (e.g. hubbiness) have been used to distinguish functionally essential and loss-of-function tolerant genes [27]. 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 [28][29]. 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 [30].

Another type of formalism making use of properties of nodes is link prediction. High-throughput experiments can be noisy, and the resultant networks may contain spurious links; missing data is also very common. Methods for link prediction and denoising are therefore useful. This can be done solely using network structure. For instance, in a protein-protein interaction network, defective cliques can be used to find missing interactions and determine the parts required to form a functional macromolecular complex [31]. Moving beyond network structure, whether two nodes are connected often depends on their intrinsic properties (e.g. their gene-expression level, conservation, and subcellular localization, etc.). A number of machine learning methods (e.g. collaborative filtering [32], maximum likelihood [33] and probabilistic relational models [34]) have been proposed to combine various node and edge features for link prediction [35]. One method that has not been used much in biological sciences is stochastic block models [36]. These have been popular in computational social science for link prediction [37]. They require comprehensive gold-standards for validation and may catch-on more in the biological sciences as these develop.

Formalisms focusing on causal relationships and dynamics

As mentioned above, one of the common ways to construct association networks is by correlating high-dimensional data. While correlative relationships can be readily calculated, a fundamental question is the distinction between direct (i.e. causal) and indirect interactions. For example, if transcription factor X regulates gene Y and Z, one could expect the expression of pairs like X-Y, X-Z, and Y-Z to be correlated, but the key is to identify the direct regulatory interactions X-Y and X-Z. Established mathematical machinery such as Bayesian networks and Markov random fields [38] have been used for this purpose.

The inference of causal relationships is greatly improved by time-series data. In social science, online retailers are interested in using purchase records to study how customers influence each other [39]. 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. However, 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 [40].

In addition to the actual dynamic processes occurring on a network, one can explore evolutionary dynamics 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. Furthermore, these have been used to align networks from different species [41] and to detect conserved and specific functional modules [42] across species. Based on a large collection of aligned networks between species, a mathematical formalism has been developed to measure the evolutionary rewiring rate between networks using methods analogous to those quantifying sequence evolution. In this context, it was shown that metabolic networks rewire at a slower rate compared to regulatory networks [43]. The inference of causal and evolutionary relationships from statistical data points to the study of mechanistic networks.

Mechanistic Networks: Comparison gives intuition into biological complexity

Now we shift discussion to "mechanistic" networks. Here, the network framework serves as a skeleton for different complex systems. In particular, the previous sections discussed universal frameworks and insights gained by applying the same formalism to biological networks as well as to various social and technological networks. Such wide-ranging universal insights were possible

only because the detailed characterization of the nodes in the network was neglected during the comparison. Only the abstracted "association" between the nodes was considered. On the other hand, if details are added to this 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. 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

A number of different stochastic models and explanations can lead to the formation of scale-free graphs. First let's consider one of the paradigms of scale-free structure, the hub-and-spoke system of the airline network. How does this come about? Every time a new airport is created, the airlines have to balance available resources and customer satisfaction, 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. This model is called 'preferential attachment' as newly created nodes prefer to connect to pre-existing hubs in the network [7] and, in this case, it depends on the small-world property of scale-free networks [12]. In contrast, one explains the evolution and growth of the World Wide Web, which is also scale free, in somewhat different way. Here, a random pre-existing node and its associated edges are duplicated (for example, to make a webpage for a new product in amazon, one could use a template shared by an existing product) [44]. After duplication, the content of two nodes and their connections diverge but a proportion of their edges are likely to be shared [45]. Such a duplication-divergence model leads to the formation of scale-free networks because the connectivity of a hub increases as one of its neighbors has a higher chance of getting duplicated. The same duplication-divergence mechanism can describe the patterns and occurrence of "memes" in online media [46]. As gene duplication is one of the major mechanisms for the evolution of protein families, the formation of scale-free behavior in the protein-protein interaction network was proposed to evolve via the duplication-divergence model [47]. However, for protein networks there are additional twists in this explanation because one can actually resolve each of the nodes in the network as molecules with specific 3D geometry. In particular, upon analyzing the structural interfaces involved in protein-protein interactions, there are great differences in hubs that interact with many proteins by reusing the same structural 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) [48].

A third explanation for scale free structure comes from dependency networks. In particular, the existence of common scale free topology in many networks leads to the emergence of universal patterns in complex systems, biological and otherwise. In particular, it has been reported that the frequency of appearance of individual enzymes across different bacterial genomes and the frequency of local installations of individual packages in multicomponent software platforms follow a broad distribution [49]. In the same analysis, it has been suggested that the observations can be explained by the scale free topology of the corresponding multi-levels dependency networks because incorporation of an additional component requires the presence of the depending factors in the network. (As a specific example: enzyme A is connected to enzyme B if A is used to decompose the output metabolites of enzyme B; package A is connected to package B if the installation of package A depends on the installation of package B.)

Thus, many networks that exhibit similar topologies are the result of significantly different underlying mechanisms. In the case of scale free networks, there exists a common mathematical formalism but somewhat different mechanistic explanations 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 protein interaction network through comparison to the web-link network, which is conceptually much more easy to understand.

Intuition from common design principles on large and small scales

The ability to gain intuition about the often-arcane world of molecular biology by comparison to commonplace systems is even more evident in comparisons involving social networks, where people have very strong intuition for how a "system" can work. Transferring the understanding of organizational hierarchy to biology is a good example of this type of comparison (Figure 2). Many biological networks, such as transcription regulatory networks, have an intrinsic direction of information flow, forming a loose hierarchical organization. Likewise, many social structures are naturally organized into a hierarchical structure -- e.g. a militarily command chain or a corporate "org-chart" [50]. 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 very limited number of mid-level majors. In a biological hierarchy of TFs, one sees a similar pattern with "high betweenness" bottlenecks in the middle. In many cases, these bottlenecks create vulnerabilities. Indeed, it has been shown in knockout experiments that many of the bottlenecks in biological networks are essential [20]. Hierarchies can insulate themselves somewhat 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 him can receive orders from a second major). Moreover, many commenters have mentioned that, in order to function smoothly, it is imperative for corporate hierarchies to have middle managers working together [51]. Strikingly, biological regulatory networks employ the same strategy by having two mid-level TFs co-regulate targets below them [52]. Thus, one can get an intuition for the reason behind a particular biological structure through analogies to a commonplace social situation.

The goal of this comparison is the transfer of ideas on the relationship between network structure and "function" from a social context to a less intuitive biological one. More generally, 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 that possess basic and well-defined functions is particularly insightful [53]. For example, consider the phosphorylation and dephosphorylation reactions of a protein by a pair of kinase/phosphatases. While the mathematical description of Michaslis-Menten kinetics can be a bit complicated, the reaction essentially sets up a sigmoidal signal-response curve that is analogous the thresholding behavior of transistors in analog electronic circuits [54]. Thus, the comparison allows us to potentially map some aspects of the logical gate structure of digital electronics to the phosphorylation network. It also helped inform the design of synthetic biological circuits capable of logarithmic computation [55]. Similarly, a decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [56]. 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. For instance in software design, modular programming that separates the functionality of a program into independent parts connected by interfaces is widely practiced [57]. The same is true for biological networks because modules can be readily reused to adapt new functions.

Intuition on network change: contrasting the tinkerer and engineer

By comparing biological and technological systems, we can see remarkable similarity in their design principles, in terms of their global organization (e.g. scale-free and hierarchical), as well as local structure. As both are complex adaptive systems, to shed light on the origin of such commonalities, we describe a third comparison: how biological and technological networks change. Manmade networks like roadways and electronic circuits are thought to change according to the plan of rationale designers. In contrast, biological networks are thought to change randomly 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. In a recent review, Wagner summarized nine key commonalities between biological and technological innovation, including descent with modification, extinction and replacement, and horizontal transfer [58].

In a sense, we could picture that both the engineer and tinkerer are working on an optimization problem with similar underlying design objectives, but take different views when balancing constraints. For example, in biological networks, more connected components (as measured by their hubbiness 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) [59][60][61][62]. Constraint is related to connectivity in biological systems. One's intuition here is obvious: biological systems seek to decentralize functionality, minimizing average connectivity on nodes and making the system robust. However, this architecture requires a few hubs to connect everything up and these more connected components are particularly vulnerable to random changes; Is this finding true in general? And if not, why? Comparison can provide insight.

Consider software systems: 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 [63]. 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 3). 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 [64][65].

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, it is a time to extend this tradition even further to compare networks in biology to those in other disciplines. In fact, efforts have already been made along this direction (Figure 4). Here, we have tried to describe how these comparisons are beginning to take place. First, we have described how association networks that just show simple connections between entities are abstract enough to allow the application of mathematical formalisms across disciplines. Then, we show 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.

Indeed, a proper intuition on concepts such as how essentiality and connectivity relate enables us to decipher a hairball into a more structured network. Moreover, once made evident, these intuitions often guide visualizations that allow us to literarily see the structure of a complex hairball (Figure 5).

What's next? We envision that these cross-disciplinary network comparisons will become increasingly common. Networks are a key structure used for the analysis of large datasets in the emerging field of data science. Moreover, network datasets are becoming increasingly common in many fields. We anticipate that this data growth will enable further fruitful comparisons with biology. One area that is especially ripe for comparison is multiplex networks, which concatenate networks to form a multiplex structure [66][67]. 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. However, 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 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 focused on leveraging the ideas and methods developed in multiple disciplines through comparison. We can even imagine that these comparisons will lead to real connections (i.e. not analogies) between biological networks and those in other disciplines. For instance, there is an increasing amount 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 [68].

Figures Caption

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.
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 3.
Different evolutionary patterns in biological networks versus technological networks. The left shows the protein-protein interactions network in human [69], 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 4.
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 abstract association networks where formalism is used equivalently in two domains (dotted lines) or mechanistic networks, where one only seeks analogy between disciplines (solid lines).

Figure 5.
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.

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

[8] 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.

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

[10] 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.

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

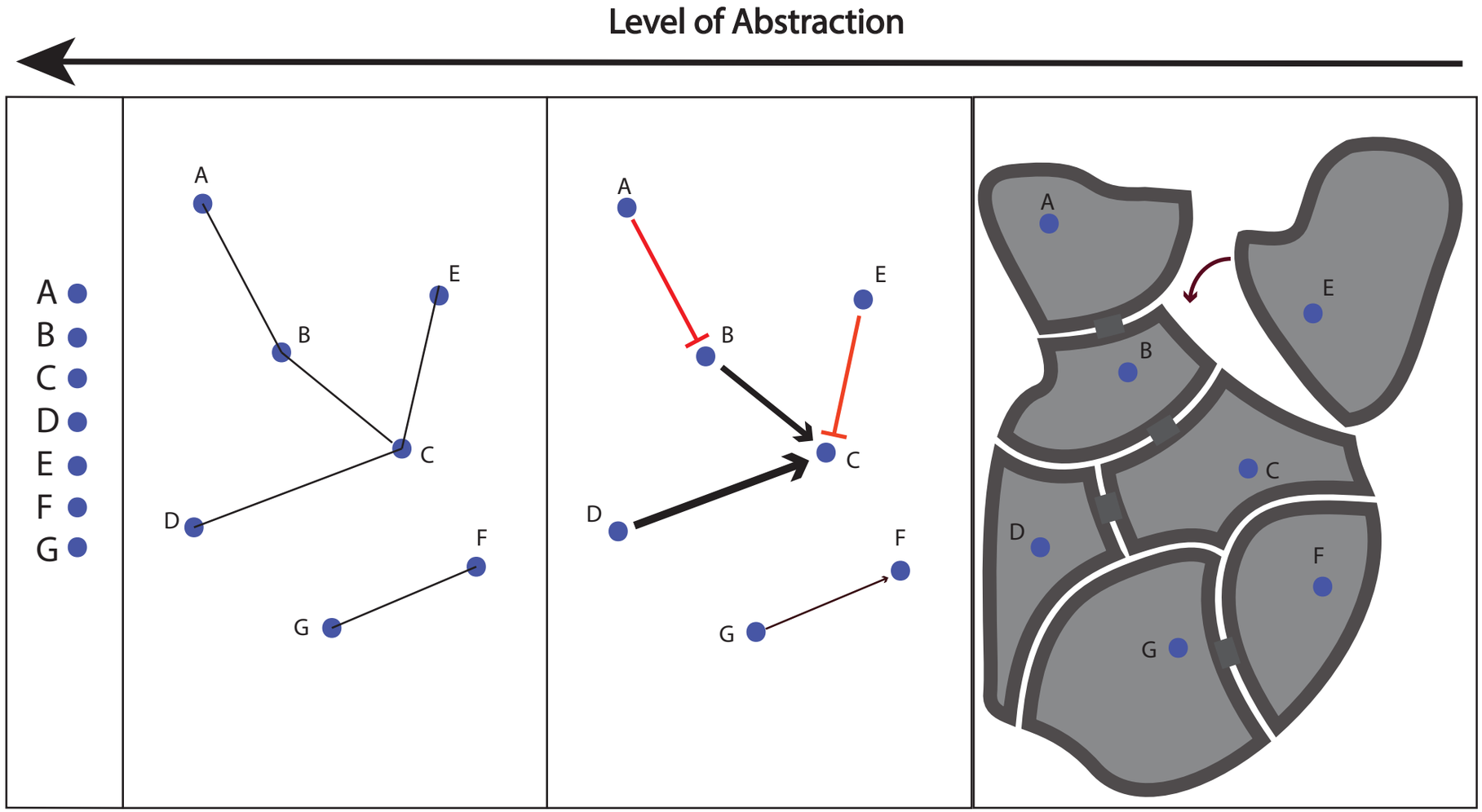
[12] L. a. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, “Classes of small-world networks,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 21, pp. 11149–11152, Oct. 2000.
[13] D. C. V. Essen, M. F. Glasser, D. L. Dierker, and J. Harwell, “Cortical Parcellations of the Macaque Monkey Analyzed on Surface-Based 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] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, “A gene-coexpression network for global discovery of conserved genetic modules,” *Science*, vol. 302, no. 5643, pp. 249–255, Oct. 2003.

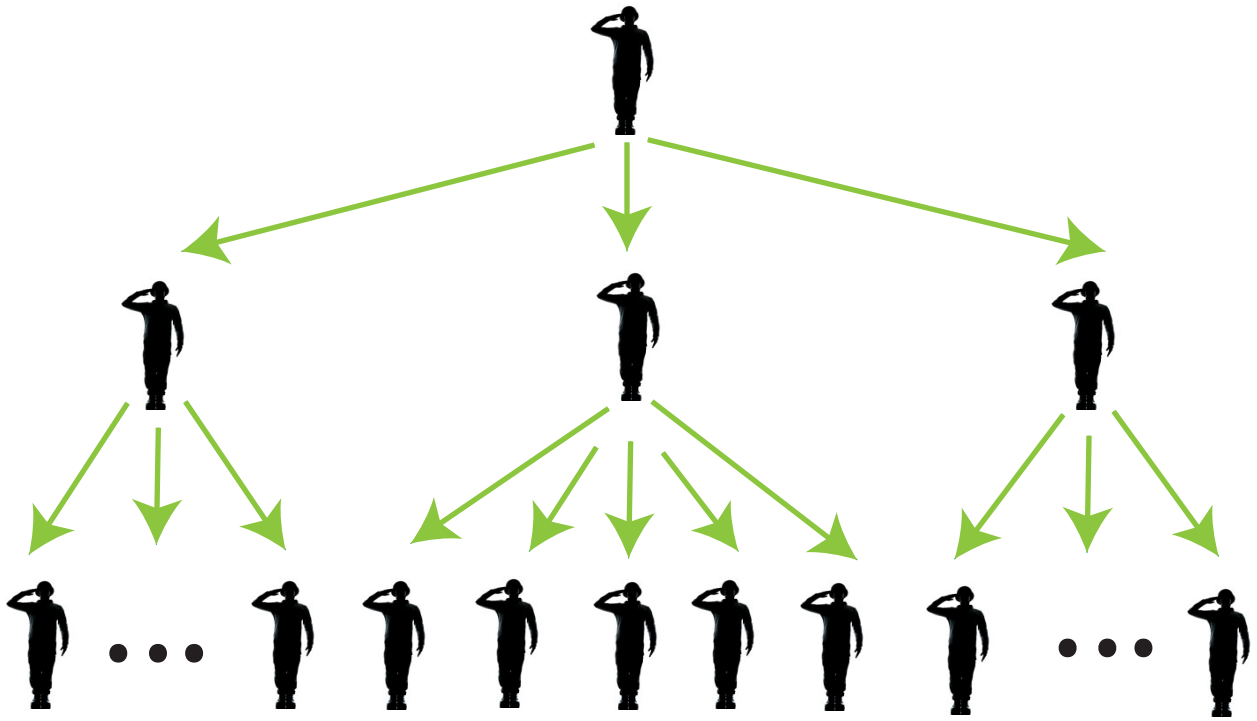
[26] Y. Moreau and L.-C. Tranchevent, “Computational tools for prioritizing candidate genes: boosting disease gene discovery,” *Nat. Rev. Genet.*, vol. 13, no. 8, pp. 523–536, Jul. 2012.
[27] E. Khurana, Y. Fu, J. Chen, and M. Gerstein, “Interpretation of genomic variants using a unified biological network approach,” *PLoS Comput. Biol.*, vol. 9, no. 3, p. e1002886, 2013.
[28] 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.
[29] 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.
[30] E. Adar and L. A. Adamic, “Tracking Information Epidemics in Blogspace,” 2005, pp. 207–214.
[31] 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.
[32] Z. Huang, X. Li, and H. Chen, “Link Prediction Approach to Collaborative Filtering,” in *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, NY, USA, 2005, pp. 141–142.
[33] 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.
[34] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, “Learning probabilistic relational models,” in *In IJCAI*, 1999, pp. 1300–1309.
[35] L. Lü and T. Zhou, “Link prediction in complex networks: A survey,” *Phys. Stat. Mech. Its Appl.*, vol. 390, no. 6, pp. 1150–1170, Mar. 2011.
[36] P. W. Holland, K. B. Laskey, and S. Leinhardt, “Stochastic blockmodels: First steps,” *Soc. Netw.*, vol. 5, no. 2, pp. 109–137, Jun. 1983.
[37] 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.
[38] 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.
[39] 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.
[40] 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.
[41] 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.

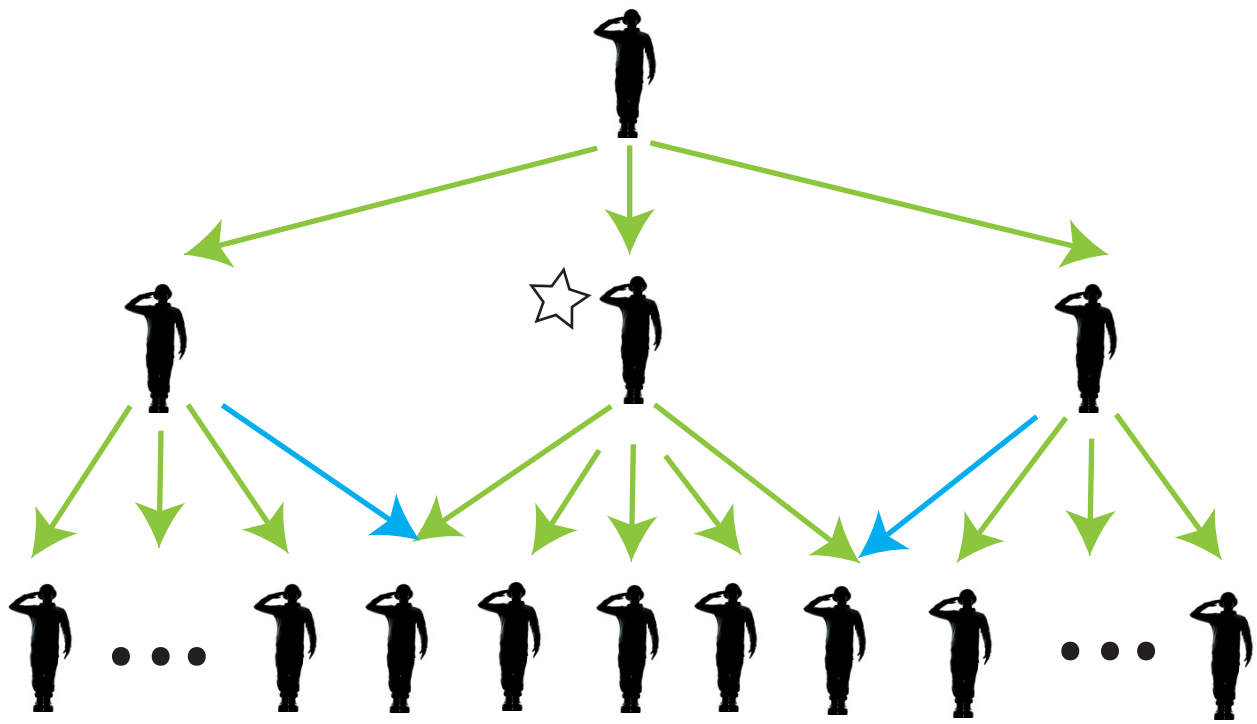
[42] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, "OrthoClust: an orthology-based network framework for clustering data across multiple species," *Genome Biol.*, vol. 15, no. 8, p. R100, Aug. 2014.
[43] C. Shou, N. Bhardwaj, H. Y. K. Lam, K.-K. Yan, P. M. Kim, M. Snyder, and M. B. Gerstein, "Measuring the Evolutionary Rewiring of Biological Networks," *PLoS Comput Biol*, vol. 7, no. 1, p. e1001050, Jan. 2011.
[44] 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.
[45] 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.
[46] 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.
[47] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, "Modeling of Protein Interaction Networks," *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.
[48] 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.
[49] 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.
[50] 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.
[51] 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.
[52] 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.
[53] 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.
[54] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell," *Curr. Opin. Cell Biol.*, vol. 15, no. 2, pp. 221–231, Apr. 2003.
[55] R. Sarpeshkar, "Analog synthetic biology," *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 372, no. 2012, p. 20130110, Mar. 2014.
[56] U. Alon, "Biological Networks: The Tinkerer as an Engineer," *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.
[57] 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.

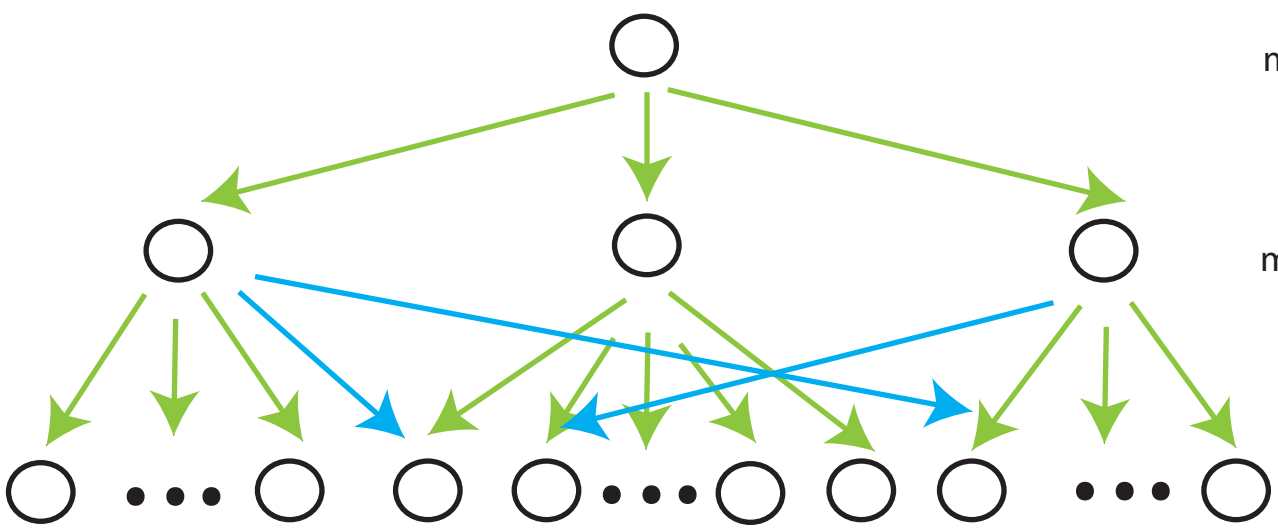
[58] A. Wagner and W. Rosen, “Spaces of the possible: universal Darwinism and the wall between technological and biological innovation,” *J. R. Soc. Interface*, vol. 11, no. 97, p. 20131190, Aug. 2014.
[59] H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman, “Evolutionary Rate in the Protein Interaction Network,” *Science*, vol. 296, no. 5568, pp. 750–752, Apr. 2002.
[60] 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.
[61] 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.
[62] 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.
[63] 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.
[64] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.
[65] 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.
[66] 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.
[67] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.
[68] J. H. Fowler, J. E. Settle, and N. A. Christakis, “Correlated genotypes in friendship networks,” *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.
[69] 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.
[70] L. Lok, “Software for signaling networks, electronic and cellular,” *Sci. STKE Signal Transduct. Knowl. Environ.*, vol. 2002, no. 122, p. pe11, Mar. 2002.
[71] T. Hase, H. Tanaka, Y. Suzuki, S. Nakagawa, and H. Kitano, “Structure of Protein Interaction Networks and Their Implications on Drug Design,” *PLoS Comput Biol*, vol. 5, no. 10, p. e1000550, Oct. 2009.
[72] 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.

[73] R. V. Solé and S. Valverde, “Information Theory of Complex Networks: On Evolution and Architectural Constraints,” in *Complex Networks*, vol. 650, E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 189–207.
[74] R. Albert, J. J. Collins, and L. Glass, “Introduction to Focus Issue: Quantitative Approaches to Genetic Networks,” *Chaos Interdiscip. J. Nonlinear Sci.*, vol. 23, no. 2, p. 025001, 2013.
[75] Yang Wang, D. Chakrabarti, Chenxi Wang, and C. Faloutsos, “Epidemic spreading in real networks: an eigenvalue viewpoint,” 2003, pp. 25–34.
[76] M. J. Keeling and K. T. D. Eames, “Networks and epidemic models,” *J. R. Soc. Interface*, vol. 2, no. 4, pp. 295–307, Sep. 2005.
[77] S. Horvath and J. Dong, “Geometric Interpretation of Gene Coexpression Network Analysis,” *PLoS Comput. Biol.*, vol. 4, no. 8, p. e1000117, Aug. 2008.
[78] A.-L. Barabási, “Network Medicine — From Obesity to the ‘Diseasome,’” *N. Engl. J. Med.*, vol. 357, no. 4, pp. 404–407, Jul. 2007.
[79] 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.

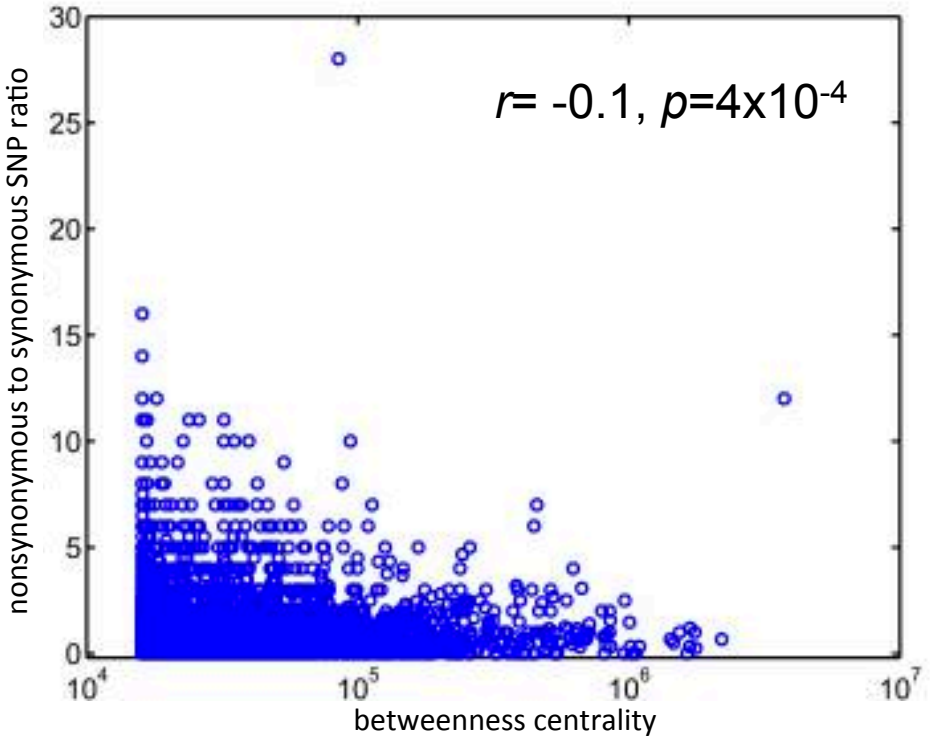

Level of Abstraction
A
B
C
D
E
F
G
A
B
C
D
E
F
G
A
B
C
D
E
F
G
A
B
C
D
E
F
G


general
major
private

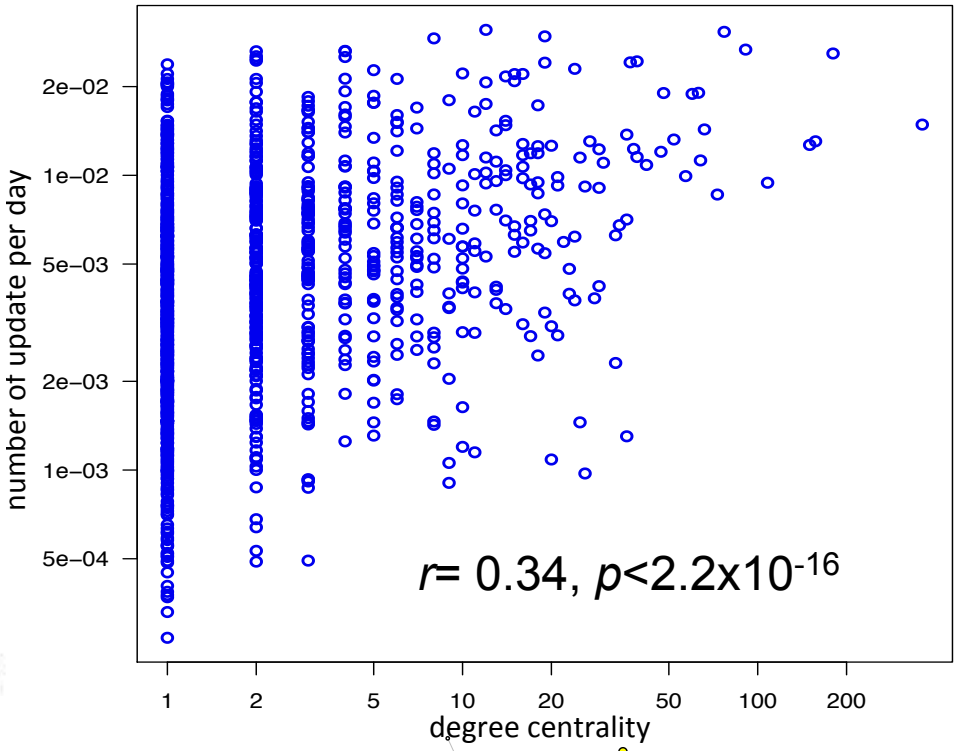

general
major
private

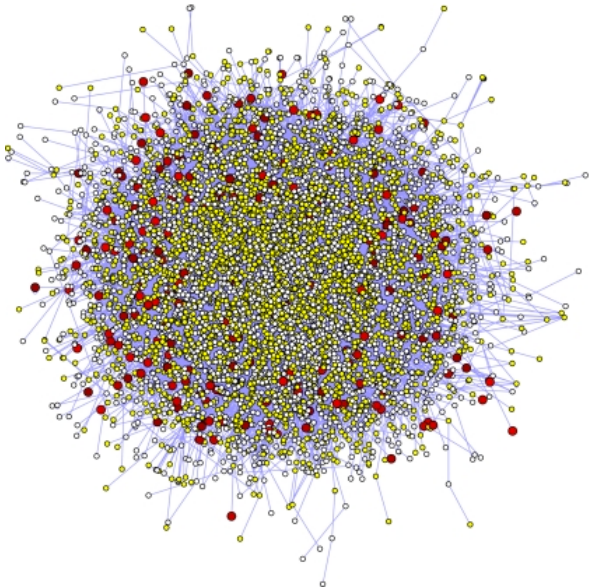

master regulator
middle manager
work horse

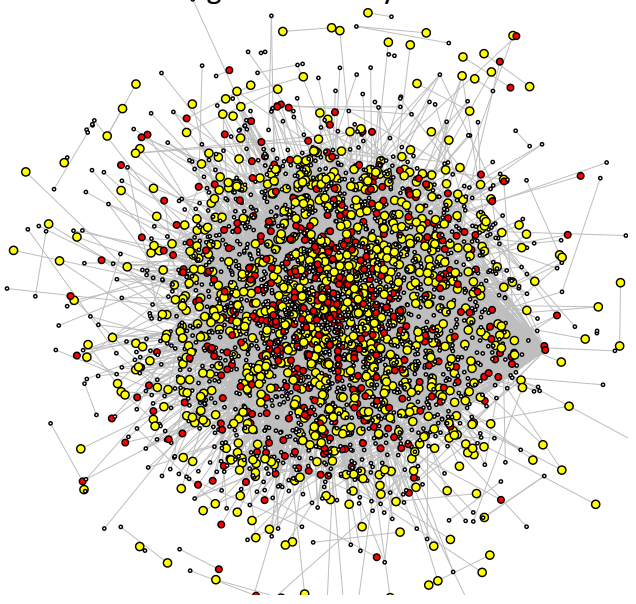
Protein-protein interaction network

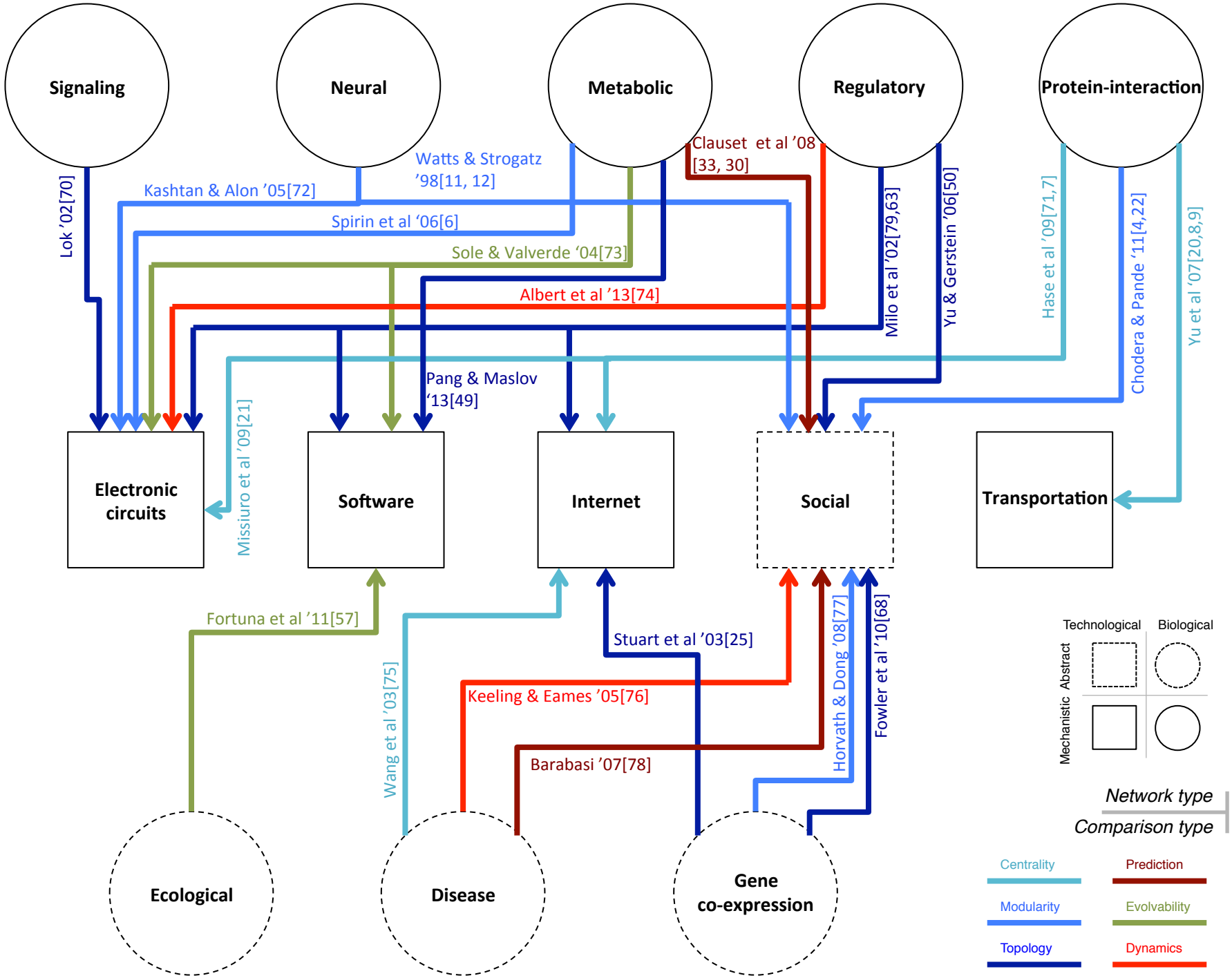


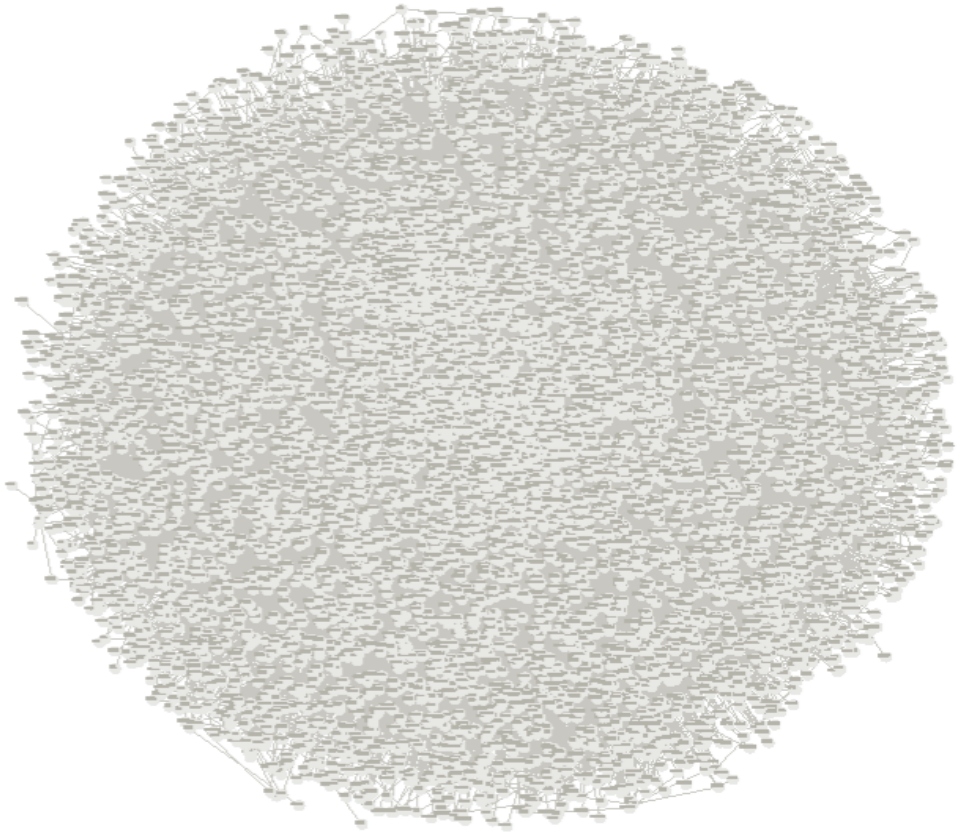
R package dependency network



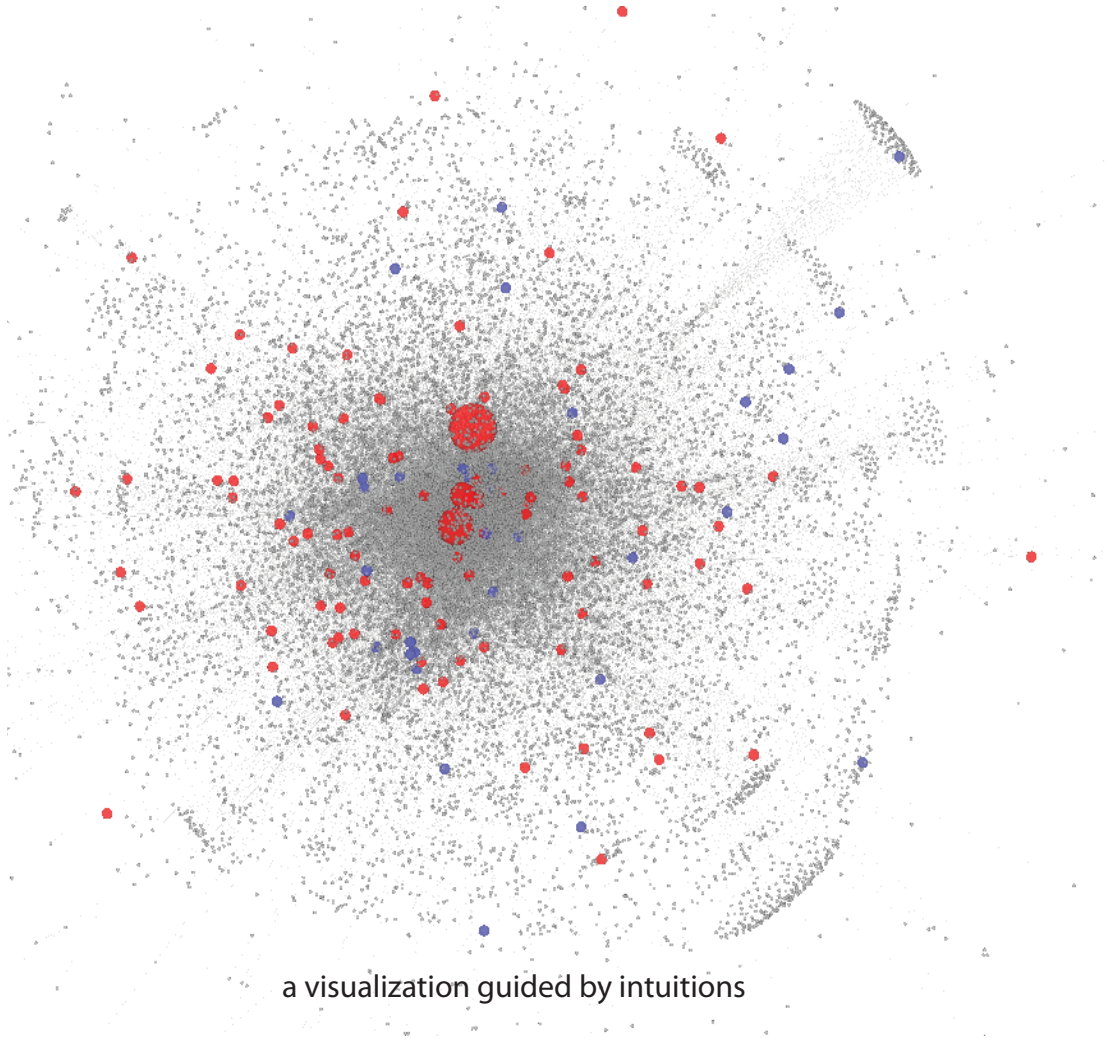





Signaling
Neural
Metabolic
Regulatory
Protein-interaction
Lok '02[70]
Kashtan & Alon '05[72]
Watts & Strogatz '98[11, 12]
Spirin et al '06[6]
Sole & Valverde '04[73]
Albert et al '13[74]
Clauset et al '08 [33, 30]
Milo et al '02[79,63]
Yu & Gerstein '06[50]
Hase et al '09[71,7]
Chodera & Pande '11[4,22]
Yu et al '07[20,8,9]
Pang & Maslov '13[49]
Missiuro et al '09[21]
Electronic circuits
Software
Internet
Social
Transportation
Fortuna et al '11[57]
Wang et al '03[75]
Stuart et al '03[25]
Keeling & Eames '05[76]
Barabasi '07[78]
Horvath & Dong '08[77]
Fowler et al '10[68]
Ecological
Disease
Gene co-expression
Technological
Biological
Abstract
Mechanistic
Network type
Comparison type
Centrality
Prediction
Modularity
Evolvability
Topology
Dynamics



a complex hairball



a visualization guided by intuitions