Submitted version:

<http://archive.gersteinlab.org/meetings/s/2015/01.05/Yan_networks_20141119b_all.pdf>

DW:

Keller EF. [Revisiting 'scale-free' networks.](http://www.ncbi.nlm.nih.gov/pubmed/16163729) Bioessays. 2005 Oct;27(10):1060-8. PubMed PMID: 16163729.

Norris V, Raine D. [On the utility of scale-free networks.](http://www.ncbi.nlm.nih.gov/pubmed/16615092) Bioessays. 2006 May;28(5):563-4. PubMed PMID: 16615092.

What Systems Biology Is (Not, Yet) Cain et al., science 2008

Although Alon’s approach provides a common language potentially scalable from biochemical

reactions, through cell-cell signaling, to organismal interactions over evolutionary time, the choice of examples “most familiar to the author” leaves out applications to nonmodel and population-level systems. Whether it will be practicable in such instances remains

to be demonstrated. <http://www.sciencemag.org/content/320/5879/1013.1.full.pdf>

The Meaning of Systems Biology, [Marc W. Kirschner](http://www.sciencedirect.com/science/article/pii/S0092867405004472#) cell 2005

<http://www.sciencedirect.com/science/article/pii/S0092867405004472>

Systems biology is not a branch of physics but differs from physics in that the primary task is to understand how biology generates variation. No such imperative to create variation exists in the physical world. It is a new principle that Darwin understood and upon which all of life hinges. That sounds different enough for me to justify a new field and a new name.

<<GOOD>>

KKY: <http://www.sciencemag.org/content/335/6069/665>

(

[**http://www.pnas.org/content/102/41/14497.short%20Doyle**](http://www.pnas.org/content/102/41/14497.short%20Doyle) **- this is not that critical**

**)**

DW: [anti systems and network biology:  
(

<http://scientificbsides.wordpress.com/2012/02/27/on-systems-biology-and-bullshit/>

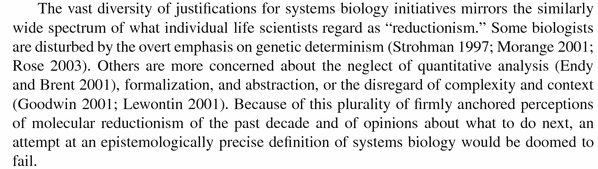
)

<http://scienceblogs.com/transcript/2008/07/27/theorizing-about-how-some-indi/>]

Systems Biology-its past, present and potential (<http://www.philsciletters.org/pdf/200914.pdf>): “A big disadvantage of the portrait of Systems Biology as “a network of disciplines” is its static character. Not

every interaction is active in a particular scenario.” (NOT SURE I agree that sys. bio. that has nothing to do w dynamics)

Dimensions of systems biology, (Not such a direct crit of networks) <http://link.springer.com/chapter/10.1007%2F112_0602#page-1>



The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

<http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>

[[RK: peer-reviewed articles]]

Network motifs: structure does not determine function

<<GOOD>>

<http://www.biomedcentral.com/1471-2164/7/108>

*“…motifs may represent "computational elements", and in the case of the feedforward motif, the possibility of the motif acting as a Boolean AND or OR gate has been investigated” .... “simply identifying the presence of particular motifs, without a detailed experimental evaluation of their respective dynamics, is unlikely to offer much insight into the functional properties of real transcriptional networks. In essence this means that knowing the structure of a network, or an inventory of the discrete modules making up that structure, doesn't provide enough information to predict how functional processes occur or how biochemical reactions proceed in a biological system.”*

Why Most Published Research Findings Are False

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/>

Why Most Published Research Findings Are False: Problems in the Analysis

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1855693/>

[[RK: articles from the popular press, not necessarily specifically anti-networks but same concepts]]

When Correlation Is Not Causation, But Something Much More Screwy

<http://www.theatlantic.com/business/archive/2012/05/when-correlation-is-not-causation-but-something-much-more-screwy/256918/>

Eight (No, Nine!) Problems With Big Data

<http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=3>

*“Molecular biologists, for example, would very much like to be able to infer the three-dimensional structure of proteins from their underlying DNA sequence, and scientists working on the problem use big data as one tool among many. But no scientist thinks you can solve this problem by crunching data alone, no matter how powerful the statistical analysis; you will always need to start with an analysis that relies on an understanding of physics and biochemistry.”*

Big data: are we making a big mistake?

<http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2ysdIXgD2>

*“... a theory-free analysis of mere correlations is inevitably fragile. If you have no idea what is behind a correlation, you have no idea what might cause that correlation to break down.”*

ANS reviewed <http://pubs.rsc.org/en/content/articlehtml/2009/mb/b908681a>

<<GOOD>>

They brought up many of the issues with the topological analysis of biological networks - statistical analysis of small-world, scale free, or power law distributions of biological networks proves that these analysis are not robust - he pushes for more dynamical analysis and analysis of smaller sub-networks. He liked the SIN analysis and pushes for more biologically meaningful analysis like SIN,

probability argument is not

how often stuffs got break down?

one components very time dependent...

LS: tree -> network

splitstree

Dear Dr XXXX,

First of all, I would like to express our appreciation to the reviewers for their thorough and in-depth reviews. We find the critical comments useful and we are eager to revise the manuscript.

Before commencing the revision in detail, we would like to ask you a preliminary question: to what degree should this be a balanced review vs a one-sided opinion piece.

Specifically, our impression is that both reviewer 2 and 3 found the manuscript reads more like a review rather than an opinion piece. Reviewer 2 asks that we cut down on this review aspect by focusing on a few examples and make it more like an opinion piece. However, reviewer 1 seems to advocate that we provide an even more balanced overview of networks by discussing both sides of well-known controversies. In a sense we see review 1 and 2 in opposition and don't think it is really possible to satisfy them both. We ask your guidance here and how to best formulate the piece.

Yours sincerely,

Mark Gerstein

==

### Essays

*PLOS Biology* publishes visionary and provocative essays that cover broad topics of general interest to life scientists. Essays might explore the implications of recent advances in a given field or global methods that promise to have broad-ranging consequences across fields, investigate an emerging trend with cross-disciplinary implications, or explain the challenges of a nascent field. (PLoS Biology author guidelines)

Dear Dr Gerstein,

Thank you for submitting your invited manuscript and for your patience during the review process. I have now heard from three expert reviewers, whose comments you can find at the end of this letter.

As you can see, all reviewers agreed that the aim of your manuscript is important and timely. However, the reviewers diverged in their recommendations, with Reviewers #1 and #2 identifying several important issues with the present version of your manuscript.

From an editorial point of view, we agree with the reviewers that your manuscript currently reads more like a review of the relevant literature, rather than an essay that provides insight and expresses a point of view. Reviewer #2 provides you with excellent advice on how to revise your manuscript in order to recast it as an Essay rather than as a literature review and ensure that it provides insight into what can be gained from network comparisons/analogies across fields. Reviewer #1 also identifies several important conceptual and terminological issues, which will need to be carefully considered if your Essay is to provide a valid and informative distinction between network elements and the types of insights that can be drawn from

them.

In sum, in light of the reviewers’ feedback, we cannot accept the present version of your manuscript for publication in PLOS Biology, but we would welcome submission of a significantly revised version

that thoroughly addresses reviewer concerns. Your revised manuscript will likely be sent for re-review, and may be edited for clarity and accessibility.

We would be grateful if you could submit your revised manuscript by Feb 13 2015 11:59PM. Please email us (plosbiology@plos.org) to discuss this if you have any questions or concerns. At this stage, your manuscript remains formally under active consideration at our journal. Therefore, please notify us by email if you do not wish to revise your manuscript for PLOS Biology and instead wish to pursue publication elsewhere, so that we may end consideration of the manuscript at PLOS Biology.

If you do still intend to submit a revised version of your manuscript, please go to<https://urldefense.proofpoint.com/v1/url?u=http://www.editorialmanager.com/pbiology/&k=dpQisR3avULHgiNaNeY%2Btg%3D%3D%0A&r=HhrYXdpxWl1L5hPq67SxwYonGQ0XmUmh5KerNvino84%3D%0A&m=s5OvBVoAS8vBRTZ7XacLInTuFTm%2FAd4zKqwuFzaeNWE%3D%0A&s=d24c7bff9464be845a73054ca34688fa8643b61c3781db65d103c088ebe487fb> and log in as an Author. Click the link labelled 'Submissions Needing Revision'. You will find your submission record there.

Thank you again for your submission to our journal. We hope that our editorial process has been constructive thus far, and we welcome your feedback at any time. Please don't hesitate to contact us if you have any questions or comments.

Sincerely,

Stavroula Kousta

Stavroula Kousta, Ph.D

------------------------------------------------------------------------

Reviewer Notes:

Reviewer #1:

Yan et al take an ambitious attempt to provide a broadly encompassing review that would present the commonalities between the concepts of "network" across a wide range of disciplines, ranging from natural to

social sciences and from science to engineering. One motivation is that biology could benefit from analogies in other disciplines outside biology where there is "intuition" of how networks function and "mechanistic" knowledge of interactions underlying the networks are better known. The latter is an idea worthwhile pursuing. There could thus be a need for such a broad overview to raise explicit awareness of using

analogies across different fields of study - a practice that many people implicitly already engage in.

My initial enthusiasm was however quickly dampened by a number of terminological and epistemological inaccuracies. First, the authors broadly divided the article into two ways to think about "networks": They distinguish between "Association" vs. "Mechanistic" networks. This is a flawed dichotomy and is not sustained by the content of the respective sections (see below). Using the two terms ’association’ and

‘mechanistic’ networks as attributes to the substantive 'network' suggests a dichotomous classification of the networks into these two types. Is that what the authors intended?

While this piece contains here and there bits of interesting, and to many readers potentially novel information (albeit mostly in a laundry list form of techniques from non-biological disciplines), it is overall flat, uses rather simplistic colloquial language unbecoming of this genre of article and exhibits not very high scholarly sophistication. I genuinely think such kind of piece as it now stands would not serve the reputation of PLoS Biology.

MAJOR ISSUES

(1) MOST GENERAL. This piece turned out to be rather disappointing: it contains epistemological errors and does not demonstrate a scholarly command of the salient technical issues which is expected for such a

commentary. By necessity then, the content remains quite superficial -even if one grants some superficiality inherent to this genre of a broad survey of a concept. Over long stretches this piece is a trivial, descriptive and rather dull narrative of what is around, without offering cogent and deeper, novel insights that one would

expect from an authoritative essays.

(2) LACK OF CRITICAL STANCE. The lack of critical analysis is disappointing because it is clear to most scholars and practitioners in the field that the concept of 'network' has been misused, often in a technically clearly flawed manner. At least the long enduring, and widely publicized controversy concerning the scale-free topology (accusations ranging from unethical scientific conduct to sloppy work to interesting technical oversight) should have been discussed. See for instance: E. Fox, Bioessays 2005 (10.1002/bies.20294) - or: <https://urldefense.proofpoint.com/v1/url?u=http://liorpachter.wordpress.com/2014/02/10/the-network-nonsense-of-albert-laszlo-barabasi/&k=dpQisR3avULHgiNaNeY%2Btg%3D%3D%0A&r=HhrYXdpxWl1L5hPq67SxwYonGQ0XmUmh5KerNvino84%3D%0A&m=s5OvBVoAS8vBRTZ7XacLInTuFTm%2FAd4zKqwuFzaeNWE%3D%0A&s=38f43205777d1bbaa41a97521ed56471756b8765d49657c3b6ea4a6561ad33d3>

### pit rev 2 v 1 - antinetwork

### reference the fox piece & barabasi blog

A second opportunity for an obviously needed critical dissection missed by the authors is that analogies across disciplines, while possibly offering a new perspective, always also come with the much feared baggage of false analogies and miss-representations.

The literature of applying network ideas to biology is riddle with such misuse, and there are already many articles that are critical. However the author miss an opportunity to summarize and present them – which would have been novel.

### not enourgh lit revie

###>>> can we find any papers critical of networks - <???>

(3) ABUNDANCE OF CATEGORY MISTAKES. While there is always a challenge in offering a cogent organization of a concept into various subclasses, the authors made rather blunt mistakes of semasiological

as well as onomasiological nature. Let me focus on just the biggest one (since they are too numerous to enumerate individually):

The division of networks into ASSOCIATION (which the authors also refer to as 'ABSTRACT') and MECHANISTIC appears at first glance meaningful

### GOOD!

and I was initially enthused that finally someone draws of line between these two natural classes: NETWORK as SCHEMATIC DIAGRAMS for representing abstract relationships (e.g. similarity between the objects that represent the nodes) and NETWORKS as maps of systems of real PHYSICAL or FUNCTIONAL interactions, such as airline networks, gene regulatory networks or Facebook friendships.

### yeah yeah

Alas, it turns out that the authors did not mean to make this elementary distinction. Instead, the authors discuss all real networks that would go under the latter class of the mechanistically concrete networks, even as physical as a neuronal connection, to the first class of abstract (ASSOCIATION) network. But what can be more concrete than synaptic network of neural cells?

This unusual classification leaves almost nothing to the MECHANISTIC networks, the second class - which remains enigmatic. What the authors perhaps mean by 'mechanistic' (based on guessing from the vague text) is a different CATEGORY than "associative" and thus could be a subset of ASSOCIATION NETWORK as well: The readers realizes only when the authors begin to explain the generating process for the scale-free network that by "MECHANISTIC NETWORK" they actually refer to a proposed mechanism that produces (hence explains) a particular TOPOLOGY. Thus, by 'ASSOCIATION' vs 'MECHANISTIC' the authors mean to imply the difference of whether there is an effort to figure out the process that generates a particular topology or not. They do not refer to the fact that there are two types of use of the network concept; but instead mean to cover two mutually non-excluding levels of APPROACHES to study networks. This is a category mistake because the focus (descriptive vs mechanistic understanding) are NOT properties of networks but of approaches – hence do not refer to ‘types’ f network. But even if by 'MECHANISTIC' they mean the networks for which one has studied the explanation for the topology, as opposed to what the informed reader would expect, namely a concrete (as opposed to diagrammatic) network, then the author also commit the additional logical error of assuming a FALSE DICHOTOMY. [ A true dichotomy of two

types, here ‘ASSOCIATION’ vs ‘MECHANISTIC’, must be such that the two are mutually exclusive, disjoint and jointly exhaustive. This is obviously not the case with ‘ASSOCIATION and 'MECHANISTIC' since many

of the examples mentioned can actually be placed under either “type”. So even if we now accept that the authors do not mean to divide the TYPE of networks in two groups but rather, intended to organize the

approaches in dealing with networks into two classes of epistemological emphasis (description vs. explanation of structure) then it is still inappropriate to call the latter (the ‘MECHANISTIC’ network) "Intuitions". This happened throughout this second part when in fact there are rigorous theories that occupy entire fields of studies of which the authors seem to be not aware of (e.g. graph evolution, bifurcation theory (structural stability), network dynamics, critical networks, non-adaptive evolutionary constraints etc)

A second level of category mistake is to subdivide the ASSOCIATION NETWORKs (which is now a meaningless class because this term refers to an approach and not a type of network) in three subclasses: (a) Formalism focusing on topology, (b) … focusing on "interplay between topology"; and (c) … focusing "on dynamics”. It is not clear whether there is a rationale between these three subclasses and what they mean: e.g., there is no definition of what is an "interplay between topologies". In this section the authors simply

list a few applications (gene prioritization and link prediction) which are neither types of networks nor types of formalism - again a category mistake. Disappointingly, here the authors come close to but fail to recognize it explicitly: the most significant dichotomy, used by all good textbooks, for approaches to network analysis: TOPOLOGY vs DYNAMICS. This is THE natural, most fundamental distinction of studies of networks. This distinction is orthogonal to the axis along which one could project the authors' 'abstract' and 'concrete' networks. Both TOPOLOGY and DYNAMICS are indeed formalisms that exist in their own right and are represented by entire disciplines in the field of math and physics (graph theory and dynamical systems theory) - but neither one is even mentioned by name.

MINOR POINTS:

There are a large number elementary technical errors and it is impossible to list them all - here are a few to give a sense of what we consider ‘technical errors’:

\* Michaelis -Menten is THE PARADIGM for non-sigmoidal kinetics - thus just the opposite of what the authors say; moreover it is an approximation (see e.g. M. Savageau, Integrative Approaches to Molecular Biology, Ed. Collado-Vides et al. MIT PRESS, 1996, 123)

\* High-betweeness and bottleneck are not equivalents. Every hub (high connectivity degree) which is not usually associated with a bottleneck, also has high betweenness. In other words, to assign a node the bottleneck property requires two conditions: high betweenness (as the authors rightly say) but also LOW connectivity degree.

\* A scale-free network is not "a kind of small network” (p. 4) but a SUBSET of the broader class of small-word networks. The preferential attachment that leads to scale-free network does not "depend" on

small-world networks (p. 7) – whatever this phrase means.

\* It is not correct that "reverse engineering" is achieved if "ideally one could write ODE to fit the data" - this is simply, plain wrong. The opposite is true: THE goal of reverse engineering is to obtain the topology, such that ODEs can be used to explore the dynamics. Thus, the authors have it the wrong way around. However, to their rescue, of course the iteration between modelling and experiment/observation may secondarily lead to the suggestion of new links to explain the observed dynamics. This error is one of many examples throughout the text for the authors' unfortunate neglect of the very field of dynamical systems theory which is THE theory for tacking complex networks.

Reviewer #2: This is a reasonably well written and, for the most part, accessible piece. The stated goal is to show how biology can benefit from the infusion of ideas and methods from the studies of networks in social, engineering and computational science. I'm sympathetic to the goal, but feel that the authors could have done a better job of achieving it. Here are some of the areas where I think attention, and revision, is needed:

1. The piece is more of a review article than an essay or opinion piece (indeed, with 82 citations, it is hard to call it anything other than a review article). There are several problems associated with this:

First, it creates real conceptual dissonance. On the one hand, the authors appear to be advocating for the introduction of something very new into biology, while on the other they are citing 20 years of literature of people already doing just what they are advocating. At the end, one comes out confused as to whether the authors are advocating or congratulating.

Second, there is a lack of judiciousness in the choices of examples and illustrations. One gets the sense that the authors are hitting us with everything they've got (as would be typical in a review article), rather than selecting those examples that best make the case for what they are advocating or proposing.

Overall, I'd recommend that the authors cut down the introductory 3 pages (which take a long time to get to the point), and select the examples they highlight from some of the more modern literature (the whole Barabasi-infused scale-free network field is getting pretty old now, and not a lot seems to have happened in it lately, at least on the biology front).

2. The piece seeks to tell us how insights can come from cross-disciplinary network comparisons, but in far too many cases it merely tells us how cross-disciplinary network comparisons can be done, rather than explaining how doing so leads to insights. In falling short of this mark, they lay the piece open to a criticism

that is all too commonly laid upon computer scientists who do biology: that they are too easily satisfied with classifying, correlating, and "inferring" (see below), which does not in and of itself produce real biological understanding. Figure 5 is a perfect illustration of this problem. The authors present it as though the re-processing of a big messy hairball into another big messy hairball with large colored dots represents some kind of triumph. It may well be true that the re-processing of data into forms that reveal intrinsic structure can lead to spectacular insights, but the conditions under which this happens are rather special, and there's an unfortunate lack of any critical discussion of that matter here. Indeed, the issue of "scale-free-ness", which occupies a fair bit of the early pages of this piece, is a case in point: after 20 years of flogging on both sides of the fence, it is still unclear whether the scale-free-ness observed in biological (and non-biological) networks has more than decorative significance, i.e. does it really convey any

deeper meaning than just being a historical artifact of the way the networks were formed? There are other examples throughout the article where the issue of what we actually learn from network analysis is simply glossed over. On one page, the authors say that, "modules within gene co-expression networks tend to contain genes in the same biological pathway or have similar functions…[so] one can infer the function of a gene or a non-coding element based on its neighbors in the underlying network". The fact that the statement is logically invalid (it is of the "a implies b therefore b implies a" type) points up the fact that the activity described here really provides a means to "assign" or "hypothesize" function; whether one actually got the function right is something that has to be addressed outside of the formalism of network analysis (to some extent it is the common but non-standard use of the work "infer" within the computer science community that is a source of confusion here). These comments are not meant to belittle the activity in question here:

crunching a huge pile of relatively unstructured data producing thousands of new hypotheses (some large number of which are likely to be correct) is no small thing. But the leap from doing that to actually obtaining insight is not insignificant. If the authors really want to reach out to people who don't do that every day, they

would do well to say something about how that leap is made, lest the average reader dismiss this article offhand. Even better, they should specifically connect these dots in some of the specific examples that

they go through. If nothing else, not doing so lends the article too much of the structure of a "compare and contrast essay", where the emphasis is on enumerating similarities and differences, rather than

explaining them.

3. A few specific points could also use some attention:

a. The term "reverse engineering" is used in a non-standard way. What they authors are referring to is properly called "system identification", i.e. inferring the architecture of a system from measurements of inputs and outputs. "Reverse engineering" is best reserved for the act of inferring the function of a system from its architecture (i.e. what is it "designed" to do?). Too many biologists have been confusing these terms in recent years, but given that the term "reverse engineering" has a long established meaning in

engineering, it would be best not to mess with it.

b. I like that the distinction is drawn between association networks and mechanistic networks. But I'm not sure I buy the assertion that mechanistic networks are necessarily incapable of being compared in the same sorts of generic way as association networks. Surely mechanistic Boolean networks (which are completely generic in structure, yet mechanistic) would be an exception?

c. Quite a bit of the ground related to constraints, fragility, and so on, which is discussed in the final pages, has been covered in detail in a variety of articles by John Doyle and colleagues from 2002 to the present. It would seem appropriate at least to cite some of this work.

Reviewer #3: I have read this review manuscript with a great deal of enthusiasm and interest. I have also learnt quite a bit from it. I totally agree and like the idea discussed in the manuscript that networks may be at the right level of abstraction for complex biological systems, near an optimal trade off between simplicity and

hence tractability, and the amount of details needed to carry out useful research leading to practically relevant and correct conclusions. I particularly liked the part of manuscript discussing how useful and reliable the conclusions and techniques that we can draw from comparing biological and other (technological, social, etc.) networks could be. I recommend the manuscript for publication. The only minor comments that I have are:

1) Overall, the manuscript reads more like a review, rather than an opinion -- some opinions appear mostly in conclusion. Adding more though-provoking opinions would not hurt the manuscript, I believe.

2) In terms of connections between biological and technological networks, an interesting parallel has been recently drawn in

<https://urldefense.proofpoint.com/v1/url?u=http://f1000research.com/articles/3-156&k=dpQisR3avULHgiNaNeY%2Btg%3D%3D%0A&r=HhrYXdpxWl1L5hPq67SxwYonGQ0XmUmh5KerNvino84%3D%0A&m=s5OvBVoAS8vBRTZ7XacLInTuFTm%2FAd4zKqwuFzaeNWE%3D%0A&s=675997bfe8b61cc5afd3a72c536aec6f10d7aadfd4142887a83a48a54045ee24> that I think fits nicely the network-comparison discussion in the manuscript.

3) A recent conceptually very similar but much higher-level opinion article<https://urldefense.proofpoint.com/v1/url?u=http://journal.frontiersin.org/Journal/10.3389/fncom.2014.00114&k=dpQisR3avULHgiNaNeY%2Btg%3D%3D%0A&r=HhrYXdpxWl1L5hPq67SxwYonGQ0XmUmh5KerNvino84%3D%0A&m=s5OvBVoAS8vBRTZ7XacLInTuFTm%2FAd4zKqwuFzaeNWE%3D%0A&s=3b4c7ec6b58de5681ac8bbf77b592b36cf914d430cc55f9e07d5ce17990c5619>

is also very relevant, especially for the discussion in the introduction.

4) The only part of the manuscript where I strongly disagree with the presented opinion is the discussion of how airport networks evolve. The presented opinion states that airline companies strive to please their customers as much as possible, given the limited amount of resources that these companies have. In my humble opinion, this is a very idealistic view. My customer experience and basic laws of economics clearly indicate that airline companies do not care about customer satisfaction at all. All they care about is maximizing their revenues. Unfortunately for them, and fortunately for us, this also involves pleasing customers to a certain degree. That is, they do care about customer satisfaction but only within limits loosely imposed by laws and by revenue maximization considerations. For example, if a particular airline company starts packing their poor customers in body-size boxes before departure to maximize the capacity and revenue even further, then those customers may consider flying by another more traditional type of carrier (if

one still exists at those futuristic times).

==

cuts

Green = DW

Magenta = ANS

Jan, 2015 (version submitted)

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, comparing 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 formalism 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[[word]], 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 an organism . In many cases, genome-scale information describing how components interact is

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. [DW: select one example?] 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]. <<keep>> [[dont implement chg]]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.

@@@@@[[KKY:

While the concept of metabolic pathways exists for decades and the study of regulatory networks was pioneered by researchers like Davidson and Kauffman, the paradigm of studying biological networks was arguably triggered by Barabasi’s efforts which began slightly more than a decade ago. Indeed, quite a number of controversies have been publicized.]]

skeptics.

@@@@@

Two schools, thought processes: to decipher the organization principles behind the network; to use the networks for integrating information or solving practical problems

Network critics

1. scale-free, bottlenecks, motifs, or any statistical patterns, so what?
   1. this is a fair comments. merely reporting a statistical pattern offer limited insight
      1. for mechanistic people who want to understand the reason behind. solely network topology is not enough. that’s why it’s important to decorate the networks with details (the whole spectrum picture). indeed, additional information helps. e.g. PPI and SIN. Furthermore, go for dynamics. While it’s useful to incorporate additional information, too many details make the system intractable. therefore we the issue of break down. only relatively small systems can be modeled with many details
      2. for methodology oriented people, without caring too much on why, those statistical patterns can be used as features for machine-learning models, making predictions for practical purpose.
2. scale-free distribution is not a “universal” feature. many of the analysis are not robust, problems in data pre-processing and fitting. There are infinite ways to arrive at a power law / broad distribution.
   1. preferential attachment, or duplication divergence etc. no single model can indeed capture the reason behind the distribution (we should downplay our original writing).
   2. Yeah, we should be very careful in fitting. more importantly, how can we exploit the broad distribution?
3. are these patterns really universal?
   1. yeah, it will be great….statistical physics perspective...

]]

**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). [DW: shorten this paragraph, or cite refs saying from list to 3D?] [[NO - no chg]]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. [ANS - replace this with next sentence?] On the other hand, a complete picture of biological systems in three or four-dimensions remains too ambitious a goal for the current state-of-the-art in data acquisition.[[make chg]]

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 coexpression

relationships provide another example. [[ok]]

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. [ANS-shorten this] <<keep or shorten>>[[no]]

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. [[no]]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.

[DW: put the following text in a box?]

BOX: 1. network topology

2. scale free

3. centrality such as betweenness, degrees,...

4. modularity

*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 scalefree

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].[[NO]]

[DW: move to “scale-free” in box][[NO]] 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].

[DW: remove this paragraph? just mention “scale-free” by a sentence like above or introduce “scale-free” in a box] 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].[[NO]]

[DW: summarize this paragraph as “network influence” in BOX] [[YES]]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 selfconsistent

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].

[DW: move to “centrality” in BOX] A second method of measuring a node’s centrality is based on the number of paths passing

through it -- its "betweenness”. [DW: just mention “bridge example”? remove all rest] 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). [[YES]]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].

[DW: shorten to “modularity” in BOX] [[YES]]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].

[DW: box stop here?]

[DW: merge the following two sections to mechanistic network, or as an transition?]

*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 [DW: mechanisms drive those properties]. Recently, 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 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 [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]. Furthermore, modules within

gene co-expression networks tend to contain genes in the same biological pathway or have

similar functions [26]. As a result, one can infer the function of a gene or a non-coding element

based on its neighbors in the underlying network.

[DW: merge to “network influence” in BOX?][[YES]]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 [27]. For

example, network properties (e.g. hubbiness) have been used to distinguish functionally essential

and loss-of-function tolerant genes [28]. 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 [29][30]. 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 [31].

[DW: shorten to one sentence about predicting missing links, or “missing link” in BOX][[YES]] 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 [32]. 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 [33], maximum likelihood [34] and probabilistic relational models [35]) have

been proposed to combine various node and edge features for link prediction [36]. One method

that has not been used much in biological sciences is stochastic block models [37]. These have

been popular in computational social science for link prediction [38]. They require comprehensive

gold-standards for validation and may catch-on more in the biological sciences as these develop.

[Dw: move to mechanistic network as comparisons at bayesian, boolean, equation levels?]

*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

[39] 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 [40]. 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 timepoints.

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 [41].

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 [42] and to detect conserved and specific functional modules [43]

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 [44]. The inference

of causal and evolutionary relationships from statistical data points to the study of mechanistic

networks

[[crit.?]][[MG: put criticisms here & use as a transition]]

**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 preexisting

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) [45]. After

duplication, the content of two nodes and their connections diverge but a proportion of their edges

are likely to be shared [46]. Such a duplication-divergence model leads to the formation of scalefree

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 [47]. 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

[48]. 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) [49].

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 [50]. 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" [51]. 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 [52].

Strikingly, biological regulatory networks employ the same strategy by having two mid-level TFs

co-regulate targets below them [53]. 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 [54]. 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 to the thresholding behavior of transistors in analog

electronic circuits [55]. 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 [56]. 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 [57]. 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 [58]. The same is

true for biological networks because modules can be readily reused to adapt new functions.

[[more biological ex.]]

*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 [59].

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) [60][61][62][63]. 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 [64]. 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 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 [65][66].

[[more biological ex.]]

**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) [67][68].[[ ADD \cite]]

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 [69][70]. 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 posttranslational

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 [71].

============= [[MG: stoppped]] ===================

Version: 10/5

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

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

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

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

**A spectrum of cellular descriptions**

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

Network description sits conveniently between these extremes by capturing the some of the relationships between components of the parts list in a flexible fashion, particularly those where topology rather than exact location captures the relationship. There are two ways to think about networks. The first one, referred as association network, is essentially a process of abstraction; meaning entries are connected via abstract mathematical association. While any mechanistic interaction could be abstracted as a mathematical association, the idea of association could be generalized to statistical relationships between two components. An example is the disease networks [6] in which a gene (genotype) and a disease (phenotype) are connected via the statistical association between the existence of genomic variants and the occurrence of the disease. Networks derived from co-expression relationships provide another example. The second kind of network, referred as mechanistic network, on the contrary, is a process of concretization. Unlike abstraction that is moving away from the complete 4D-picture, concretization is pointing towards this picture. It aims to understand more of the physical processes happening inside a living system, for instance the processing of information, the chemistry of metabolites and the assembly of molecular machine, and therefore focuses on incorporating various details of interactions. Adding further mechanistic detail onto a simple nodes-and-edges skeleton can often be visualized by decorating edges with directionality, color, thickness etc. Nevertheless, the incorporation of too much detail makes the system intractable, and network formalism generally breaks down if we try to load spatial or temporal details as well as higher-order interactions onto the diagram. At certain point, the actual four-dimensional picture is required.

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

**Comparison leverages mathematical formalism**

Lying at the heart of the power of network formalism is its simplicity. In the era of Big Data, 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.

*Formalism focusing on network topology*

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

[[RK new:

An example of a 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 [[ref](http://cercor.oxfordjournals.org/content/22/10/2227.abstract?ijkey=0791f72899a11e288235e14f592f6941bd29832d&keytype2=tf_ipsecsha)] that are dominated by connections internal to each area, with only ~20% of all connections being between neurons in different areas [[ref](http://www.sciencemag.org/content/342/6158/1238406.long)]. Each area is considered to have a primary feature, for example in processing sensory or cognitive signals, and is an excellent analogue of the modular characteristics of intra-cellular molecular networks in which proteins in tightly controlled functional groups coordinate as part of larger pathways to achieve well defined cellular functions. The cortical architecture has a high degree of clustering and small path-length and exhibits an exponential degree-distribution [[ref](http://www.pnas.org/content/107/30/13485.full)].

]]

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

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

Network is extremely useful in data mining because it can be used as a reference for mapping additional properties or features of different nodes. Similar questions and solutions have been come up in dealing with biological data as well as data from disciplines like computational social science. An important example is the inference the missing data by the idea of “guilt by association”, or the idea that nodes that have similar associations in the network tend to be more similar in nature. For example, in a social context, if your friends in Facebook use Product Y, you are more likely to use product Y and the advertisements you view online are personalized based on these recommendation systems [[KKY2DW: can you find a ref?]]. In biological context, the assumption is based on observations like the cellular components within the same module are more closely associated with the same set of cellular phenotypes than components belonging to different modules [18], and the modules within gene coexpression networks also tend to contain genes with similar functions and genes within the same module are often involved in the same biological pathway [19]. As a result, one could infer the functions of a protein or a non-coding element based on the function of its neighbors in the underlying network. Networks play an important role in gene prioritization, an essential process for applications like disease gene discovery because of limited validation and characterization resources [20]. For examples, network properties of individual genes have been used to distinguish functionally essential and loss-of-function tolerant genes [21]. More formally, one could prioritize the candidate genes based on how they are connected to the known genes. For example, if a gene is one-step away from a group of disease genes, it is very likely that the gene is associated with disease X. Of course, the influence of a node may not be restricted to its nearest neighbors; network flow algorithms are widely used to examine the long-range influence [22][23]. In social science context, [[KKY2DW/ANS, can you guys come up with an application in here, with ref?]]

We want to emphasize that networks are in general noisy. High-throughput experiments may result at spurious links, and missing data is very common in social science. Methods for link prediction and denoising are therefore very useful. Link prediction can be done by using the network information alone, for instance, in a protein-protein interaction network, defective cliques were used to find missing interactions and determine the parts required to form a functional macromolecular complex [24]. More often, because whether two nodes are connected depends on their intrinsic properties, one could employ machine-learning techniques to explore the relationships between connections and various features [25]. Recently, generative models of networks, say stochastic block models [26], are very popular in computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standards for validation.

*Formalisms focusing on causal relationships and dynamics*

The construction of various association networks is an active area of research for both biology and computational social science. While correlational relationships could potentially be easily calculated with the appropriate data, a fundamental question is the distinction between direct and indirect interactions. For example, a transcription factor X regulates gene Y and Z, one could expect pairs like X-Y, X-Z and Y-Z are all correlated, but the key is to identify the direct regulatory interactions X-Y and X-Z. Established mathematical machineries like Bayesian networks, Markov random fields and other information theoretical frameworks [27] have been used for this purpose.

The inference of causal relationships could be greatly benefited by time-series data. In social science, online retailers are interested to use purchase records to study how customers influence each other [[KKY2DW: can you figure out the reference?]] On the other hand, the same question is extremely common in biology, under the term “reverse engineering”. For example, how can we infer the developmental gene regulatory network from temporal gene expression dynamics? Ideally, one could write differential equations to fit the temporal data; nevertheless, temporal data in most genomics experiments do not have enough time-points. To overcome the drawback, data mining techniques such as matrix factorization are employed. For instance, given the genome-wide expression profile of at different time-points, one could project the high-dimensional gene expression data to low dimensional space and write differential equations to model the dynamics of the projections [28]. The inference of casual and direct relationships from statistical data points to the study of mechanistic networks.

Apart from the actual dynamical processes happen in a network, one could explore the evolutionary dynamics of networks. In biological context, pairs of orthologous genes can be used to align networks from different species. Based on the notion, a mathematical formalism was developed to measure the evolutionary rewiring rate between networks across species in analogous to quantifying sequence evolution [29]. It was shown that metabolic networks rewire at a slower rate compared to various regulatory networks. The same notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [30].

**Comparison gains physical intuition**

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

*Looking for mechanistic insights*

While the previous sections discussed universal frameworks and insights gained by comparing biological networks to various social, and technological networks. Such wide-ranging universal insights were possible only because the identities of the nodes in the networks were neglected during the comparison and only the association between the various nodes was considered. As one adds details to this picture, however, the insights become more specific. This is clearest when one tries to explain the universally observed scale-free degree distribution of various networks.

A number of different stochastic models lead to the formation of scale-free graphs. For example, every time a new airport is created, the airlines have to create a balance between the 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 airports. The most efficient use of these limited resources occur if the new airport connects to pre-existing hubs in the network as it reduces the travel time of the average customer. This model is called the preferential attachment model as newly created nodes prefer to connect to pre-existent hubs in the network [7]. In contrast, the evolution of the world wide web is explained by the duplication divergence model. In this model, a pre-existing node and its associated edges (for example, a webpage with all its pre-existing links) are duplicated randomly. After duplication, the edges associated with these two nodes diverge independent of one another. The duplication-divergence model leads to the formation of scale-free networks because the connectivity of hub nodes increases even further after one of its many neighbors get randomly duplicated. The same duplication-divergence mechanism can describe the patterns and occurrence of “memes” in online media [32]. 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 [31]. On analyzing the structural interfaces involved in protein-protein interactions, however, it was found that this model only applies to hubs that interact with all its partners through a single interface (with the duplicated protein reusing the same interface as its parent) as opposed to those that interact with their partners through multiple interfaces [34]. Nevertheless, as biologists, we love to think about functions and selection; it is interesting to see that, by network comparison, network organization could be a manifestation of stochasticity.

[[ANS2KKY: This should move up now because we are talking about models for creating scale-free networks here rather the scale-free behavior of networks]]. More recently, it has been shown that components in both bacterial genomes as well as large-scale computer software projects form multilayered dependency networks (enzyme A is used to decompose the output metabolites of enzyme B; the installation of package A depends on the installation of package B). The common underlying dependency networks leads to the same power-law components-usage frequency distribution (how often a enzyme is present in a bacterial genome; how often a certain package is installed in a computer) [33]. While it is elegant to explain the topology of disparate networks by simple stochastic models, such universal mechanism cannot capture the full picture.

*Looking for common design principles*

Apart from universal mechanisms, comparison of networks shed light on the design principles of networks. An example is the so-called network hierarchy (see Box 1). Many biological networks, for instance transcription regulatory networks, have an intrinsic direction of information flow, forming a hierarchical organization, similar to structures like corporate management hierarchy [35]. Nodes in the middle level therefore form the information bottlenecks. To avoid break down of flow, middle regulatory factors tend to co-regulate downstream targets; the same is true for management hierarchy in where middle managers tend to communicate often [36][37].

Lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. The mapping points to biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [38]. As it is in general very hard to define a “function”, toward this direction, comparison with various technological or engineered networks with well-defined functions is particularly insightful. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a source of negative feedback and a source of time delay. Nevertheless, different oscillators (e.g. for circadian rhythms, for cell cycle, or from various organisms) have a certain level of variation because of additional design objectives or strategies. This is just like the case that not all electronic devices use the same oscillator design because of other design objectives. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [39]. Robustness is obviously a preferred design objective because it makes a system tolerate stochastic fluctuations, either intrinsically or from external sources. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by an interface is widely practiced [40]. The same is for biological networks because modules can be readily reused to adapt new functions.

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

The comparison of biological networks and technological networks should best be performed under the light of evolution. As Alon highlighted by the phase “the tinkerer as an engineer” [39], it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer neither designs things nor builds systems— it settles on systems that, historically, conveyed a survival benefit (and if a better way comes along, it will adopt that). On the other hand, technological networks are essentially blueprints drawn by engineers who have a grand plan that makes sure everything works harmoniously. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that constructs living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological ones like roadways and electronic circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [41]. To a certain extent, an engineer is a tinkerer (see Box 2).

Under such a united framework, we could picture that both engineer and tinkerer are working on an optimization problem with similar underlying design objectives. Like all optimization problems, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there’s no free lunch [42][43]. Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited in the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many disparate processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fraction of functions are updated often [44]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. The patterns seems to be hold for other software systems like the organization of packages in the statistical computing language R (Figure 2). In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, hubs in the software system evolve rapidly. This seems to be counter to ones intuition that an engineer should not meddle too much with highly connected components. However, there is another intuition in play: rational designers may believe that they can modify a hub without disrupting it -- in contrast to the situation with random changes. Moreover, the central points in a system are often those that are in the greatest use and hence are in the most need of the designer's attention. The situation is analogous to road networks: one sees comparatively much construction on highly used bottlenecks (e.g. the George Washington Bridge) as opposed to out of the way thoroughfares (see Box 2).

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of ‘omes’ across species. In the “omics” era, may be it is a time to extend our tradition even further to compare networks in biology as well as other disciplines. Over the past few years, efforts have been spent on concatenating networks together forming a multiplex structure [45][46]. This direction is of particular interest to biology. First, from an abstract formalism standpoint, due to rapid advancements in data acquisition, the structure of biological data goes beyond a single layer of network to multiplex structure: the multiple layers could either be formed by different categories of relationships (co-expression, genetic interactions, etc.), analogous to social science in which an individual may participate in multiple social circles: family, friends, colleagues, or in online setting: Facebook, LinkedIn and Twitter. Second, mechanistically, biological regulation happens in multiple levels: transcriptional regulation, post-transcriptional regulation, and even post-translational in analogous to a city with electrical networks, water pipes, and cell phone lines. We are looking forward to some of the methods developed in other contexts to be applied in biology.

So far, we have already seen examples in which comparison brings new connections. For examples, there are emerging theories that unite evolved and designed systems; there is an increase of attention among biologists and sociologists on the connection between genomics information and sociological information such as whether phenotypes or genotypes are correlated in friendship networks [47]. Indeed, various scientific disciplines form a network in the intellectual universe where knowledge emerges when things connect.

Version: 9/14

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

Throughout the history of science, advancements of 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. In the era of systems biology, attention has shifted from individual molecular components to their interactions at a system level. New functional genomics assays, in particular ones based on high-throughput sequencing (\*Seq) [1], enables biologists to probe thousands of ‘omes [2] – the comprehensive collections of constituents. One may wonder which discipline will contribute the most to biology in this new scientific paradigm [3]. While the influx of ideas in the age of reductionism mostly originated from specific areas in physics or chemistry, to understand biology via a systems perspective, the new wave of catalysts come from areas of science that are far apart, as diverse as engineering, behavioral science, sociology, but are centered on the concept of network [4].

Networks are by no mean new to biologists [5]. Metabolic pathways have been studied for decades. But more recently, as a result of the advancements of high-throughput techniques, simple pathways have been expanded to intertwined wiring diagrams. While many of us have been astonished by the complexity of such networks found in genomics or systems biology, few are able to gain any intuition from the hairballs [6]. In this essay, we argue that, by cross-disciplinary network comparison, algorithms or mathematical techniques as well as intuitions developed in commonplace networks can be able to catalyze our understanding of biology. One may wonder, however, comparing a bio-molecular network with a complex network from a disparate field, say a social network, sounds like comparing apples to oranges. So what kinds of comparison could truly deepen our understanding? We believe that it is useful to think of different descriptions of a cellular system as a spectrum (Figure 1).

**A spectrum of cellular descriptions**

Given the complexity of a cell, a certain level of simplification is necessary for useful discussion. We could picture the description of cellular systems as a spectrum (Figure 1). On one hand, there’s a simple parts list that just enumerates each component without specifying any relationships. On the other hand, there is a complete three or even four-dimensional picture of how cellular molecules interact in space and time. It is well regarded that the characteristics of a cellular system cannot be explained by the characteristics of individual components – the whole is greater than the sum of its parts. Therefore, while the full picture is often too ambitious for current data acquisition, the parts list description is not fully informative. Network description sits conveniently between these extremes by capturing the relationships between components of the parts list. There are broadly two ways to define relationships based on the nature of experimental data. Firstly, networks could be defined in a phenomenological sense; meaning entries are connected via abstract mathematical relationship derived from phenomenological observables. Perhaps the most important phenomenological networks are built on the mapping between genotypes and phenotypes. An example is the disease networks [7]; 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. While phenomenological networks offer a mathematical abstraction far away from the complete picture in the spectrum, there are networks defined to capture the mechanistic interactions happening inside a living system. Such networks are built on experimental knowledge on different facets of the complex organization of an organism, for instance, a regulatory network describes part of the cellular information processing, a metabolic network traces the chemistry of metabolites, and the protein-protein interaction network captures cell signaling as well as providing a manual on how to assemble molecular machines. The integration of such mechanistic networks provides a description reasonably close to the complete picture along the spectrum. The process could be visually regarded as the decoration of edges with directionality, color, thickness etc. Nevertheless, the incorporation of too much detail makes the system intractable, and network formalism generally breaks down if spatial or temporal details as well as higher-order interactions are included. At certain point, the actual four-dimensional picture is required.[[CC: more details??]] [[CC: some network may not follow into either phenomenological or mechanistic network. e.g. genetic network from synthetic lethal: interaction are determined by experiments, but the mechanism behind of them are quite complicated. ]] [[CC: do we need to prepare a table to show examples of the two types networks??]]

Underlying mechanistic networks serve as the skeletons of different complex systems. Comparison of such networks allow biologists to gain intuitions by examining interactions in cross-disciplinary complex systems in the same ground as the interactions between molecular components in cells. Nevertheless, because of systems-specific details, not everything could be transferred from one discipline to another, and it is important to focus on the conceptual resemblance instead of merely topological resemblance. [[CC: this part is hard to understand. Better provide some concrete examples]] On the other hand, phenomenological networks are rather abstract connections between entities, and thus mathematical formalisms are easily transferrable. Toward this end, by comparing similar network-based mathematical formalisms across disciplines, biologists will benefit in terms of algorithms or method development.

**Comparing phenomenological networks to leverages mathematical machineries**

Phenomenological networks are typical products dealing with big data; they are essentially two-dimensional projection of high-dimensional data. As it is extremely common to have data with many features in the era of Big Data, especially in computational social science, networks across disciplines actually present very similar challenges. Here, we highlight a few areas where different questions arise in genomics and social science could be formulated by the very same approach. By the same token, network algorithms developed in one discipline can readily be applied in biology.

*Approaches focusing on network topology*

Even though the evolutionary process involves random changes at the molecular level, it is not surprising that natural selection organizes biological networks in an ordered fashion. Comparison of biological networks with different social and technological networks have provided valuable insights into the organizing principles of biological networks. One striking outcome of the comparison is that these natural and man made networks organize into scale free networks in which a majority of the nodes contain very few connections (edges) while a few nodes in the network are highly connected. Changes to the sequences of the centrally connected nodes (also called hubs) in biological networks are predictive of lethal and disease causing changes and have proven to be valuable in interpreting the genotype to phenotype network. [[CC: the subtitle of this section is about “phenomenological networks”, but concepts of scale-free and modularity were first proposed in PPI, which form “mechanistic” networks. ]] [[ANS: I agree with CC]]

[[More on scale-free - Barabasi versus Erdos-Renyi, centrality, bottleneck. First real result of network analysis.]]

[[Modules, cliques, - more comparison of cliques and modules in other contexts. Modularity]]

Another striking insight from these comparisons is these networks are organized in the form of modules such that nodes within the same module have a larger number of connections with each other as compared to nodes belonging to different modules. Evolutionarily this makes sense because connections within a module can be reused in a different functional context. Naturally, genes within the same module have similar biological properties. For example, the cellular components within the same module are more closely associated with the same set of cellular phenotypes than components belonging to different modules. Similarly, the modules within gene coexpression networks also tend to contain genes with similar functions and genes within the same module are often involved in the same biological pathway.

Due to rapid advancements in data acquisition, the structure of biological data goes beyond a single layer of network to multiplex structure commonly found in different technological and social networks. Multiplex networks contain multiple layers of interconnected networks - the multiple layers in these networks could either be formed by different categories of relationships (co-expression, genetic interactions, etc.) or they could be formed by relationships observed at different timepoints. The idea originated in social network analysis because an individual may participate in multiple social circles: family, friends, colleagues, or in online setting: Facebook, Linkedln and Twitter. Similarly, the different layers in a temporal network contain parts of the network that are connected at different timepoints [28]. As dynamic data in genomic information becomes available, we think that valuable insights can be gleaned by the analysis of these data using algorithms developed in the context of multiplex social networks. [[CC: do we need to give one or more examples?]].

Nevertheless, biology motivates an alternate definition of temporal network. While they exist together at the same time-point, networks from different species essentially capture the evolutionary changes to a common core. In this definition, pairs of orthologous genes can be used to connect networks from different species, forming a multi-layers structure. The notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [29]. Based on the same notion, a mathematical formalism was developed to measure the evolutionary rewiring rate between networks across species in analogous to quantifying sequence evolution [30]. It was shown that metabolic networks rewire at a slower rate compared to various regulatory networks.

*Approaches focusing on the properties of nodes and edges*

Lots of interesting questions as well as many machine-learning formalisms arise when we start to map properties of individual nodes to a network. In both biology and computational social science, very often the properties of nodes are incomplete, and we are interested to infer the missing data. The essence of these methods is the idea of “guilt by association” or the idea that nodes that have similar associations in the network tend to be more similar in nature. For example, if your friends in Facebook use Product Y, you are more likely to use product Y and the advertisements you view online are personalized based on these recommendation systems. In genomics, for example, one could infer the functions of a protein or a non-coding element based on the function of its neighbors in the underlying network. The same is true for predicting disease-associated genes: if the neighbors of a gene are all associated with Disease X, it is very likely that the gene is associated with disease X. Of course, the influence of a node may not be restricted to its nearest neighbors; network flow algorithms are widely used to examine the long-range influence. [[KKY: an example by DW on network influence can fit here.]] Mona Singh - functional flow

Nodes association is closely related to nodes prioritization, in which the PageRank algorithm plays an important role. Originated from Katz centrality in social network analysis [22], PageRank algorithm was first used by Google to rank documents based on linkages in a self-consistent way. The algorithm was then adopted in food webs to determine extinction [23] and later in an algorithm called NetRank that rank prognostic relevance for patients with cancers [24]. Generally speaking, in addition to algorithms like PageRank that prioritize nodes by network topology, expression data, sequence information, functional annotation and biomedical literature are required for further filtering [25]. In applications like disease gene discovery, nodes prioritization is an essential process because of limited resources. Similarly, social networks can be utilized to identify the people that need to be vaccinated so that the spread of a disease can be avoided. [[ Move this to centrality -prioritization]]

Very often whether two nodes are connected depend on their intrinsic properties. Therefore the inference or prioritization of nodes leads to prediction and denoising of links. Difficulties lie at the proper learning of network organization based on observable data. Recently, generative models of networks, say stochastic block models [27], are very popular in computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standard for validation. Interestingly, because of the availability of datasets, for instance the Framingham study, there is an increase of attention on the connection between genomics information and sociological information. Biologists and sociologists have started to examine the hypothesis on whether phenotypes or genotypes are correlated in friendship networks [21]. [[Discussion at the end]]

*Approaches focusing on causal relationships and dynamics*

The construction of various phenomenological and social networks an active area of research for both biology and computational social science. While correlational relationships could potentially be easily calculated with the appropriate data, a fundamental question is the distinction between direct and indirect interactions. For instance, a statistical analysis on many cancer samples can easily identify the correlation between various somatic mutations (indirect), but the key is in fact to identify the driver mutations (direct). Established mathematical machineries like Bayesian networks or Markov random fields have been used for this purpose. The inference of causal relationships could be greatly benefited by time-series data. In social science, online retailers are interested to use purchase records to study how customers influence each other. On the other hand, the same question is extremely common in biology, under the term “reverse engineering”. For example, how can we infer the embryonic developmental gene regulatory network from temporal gene expression dynamics? Ideally, one could write differential equations to fit the temporal data; nevertheless, temporal data in most genomics experiments do not have enough time-points. To overcome the drawback, for instance, given the genome-wide expression profile of at different time-points, one could perform project the high-dimensional gene expression data to low dimensional space by data mining techniques such as SVD, and write differential equations to model the dynamics of the projections [26]. The inference of causal and direct relationships from statistical data points to the study of mechanistic networks.

**Comparison of Mechanistic networks for gaining intuition**

From a biologist standpoint, comparing various mechanistic networks between biology and other disciplines can bring intuition from other disciplines into biology. In spite of the disparate fields, we believe there is several aspects biologists could find inspiration.

*Looking for universal mechanisms*

Since the burgeoning of studying networks in various disciplines, efforts have been made on explaining some of the striking similarity in terms of organization of underlying networks in biological and other complex systems. An early example is the emergence of the scale-free degree distribution in a protein-protein interactions network. The pattern of organization could be explained by the duplication divergence model [10], a simple stochastic process describing how a protein network grows by gene duplication. As a hub protein has many interactions, its number of interactions is likely to increase further simply because one of its neighbors got duplicated. The same “richer get richer” model was proposed originally to explain the same pattern in many other networks [11]. More recently, it has been shown that components in both bacterial genomes as well as large-scale computer software projects form multilayered dependency networks (enzyme A is used to decompose the output metabolites of enzyme B; the installation of package A depends on the installation of package B). The common underlying dependency networks leads to the same power-law components-usage frequency distribution (how often a enzyme is present in a bacterial genome; how often a certain package is installed in a computer) [12].

While it is elegant to explain the topology of disparate networks by simple stochastic models, such universal mechanisms are rather rare. To a certain extent, the existence of such models underlines the importance of randomness in biology. Remarkably, the same duplication-divergence mechanism has been applied to describe the patterns of “memes” in online media [13]. As biologists, we love to think about functions and selection; it is interesting to see that, by network comparison, network organization could be a manifestation of stochasticity.

*Looking for common design principles*

Of course, biological networks are not random, and so do networks from other disciplines. Most observed similarities in terms of network organization are not easy to explain by simple mechanisms or principles, for instance, the so-called network hierarchy (see Box 1). The reason is because, for most networks, it is in general very hard to define a “function”. In fact, lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. The mapping points to biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [14]. Toward this direction, comparison with various technological or engineering networks with well-defined functions is particularly insightful. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a source of negative feedback and a source of time delay. Nevertheless, different oscillators (e.g. for circadian rhythms, for cell cycle, or from various organisms) have a certain level of variation because of additional design objectives or strategies. This is just like the case that not all electronic devices use the same oscillator design because of other design objectives. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [15]. Robustness is obviously a preferred design objective because it makes a system tolerate stochastic fluctuations, either intrinsically or from external sources. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by interface is widely practiced [16]. The same is for biological networks because modules can be readily reused to adapt new functions. Because of the fundamental importance of such design objectives, an insightful network comparison should be rooted in the common design objectives rather than merely network topology.

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

The comparison of biological networks and technological networks should best be performed under the light of evolution. As Alon highlighted by the phase “the tinkerer as an engineer” [15], it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer neither designs things nor builds systems— it settles on systems that, historically, conveyed a survival benefit (and if a better way comes along, it will adopt that). On the other hand, technological networks are essentially blueprints drawn by engineers who have a grand plan that makes sure everything work harmoniously. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that construct living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological networks like roadways and circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [17]. To a certain extent, an engineer is a tinkerer (see Box 2).

Under such a united framework, we could picture that both engineer and tinkerer are working on an optimization problem with similar underlying design objectives. Like all optimization problems, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there’s no free lunch [18][19]. Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited in the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many different processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fraction of functions are updated often [20]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. The patterns seems to be hold for other software systems like the organization of packages in the statistical computing language R (Figure 2). In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, software system is very similar to a roadway system; bottlenecks under high usage like George Washington Bridge require more upgrade and more construction. While intentional tweaking on bottlenecks sounds obvious for technological systems, it is not always possible (see Box 2).

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of ‘omes across species. To nourish a system-level understanding and to leverage the tremendous amount of high-throughput data, may be it is a time to extend our tradition even further to compare with networks from other complex systems as well as other disciplines. Indeed, various scientific disciplines form a network in the intellectual universe where knowledge emerges when things connect.

**More Potential exhibits:**

A table showing examples of the two types networks.

(Give more examples of phenomenological networks, like genetic interaction networks.)

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

Box 1 Hierarchical organization of networks

Many biological networks possess an intrinsic direction of information flow, forming a hierarchical network organization. The hierarchical organization in biological networks resemble the chain of command in human society, like in military context and corporate hierarchy [8]. For instance, in a transcriptional regulatory network more influential transcription factors (regulators whose expression are more highly correlated with the expression of target genes) tend to be better connected (have more interacting partners) and higher in the hierarchy [32]. Moreover, the transcription factors in the middle layer tend to be more cooperative [33]. Such a situation has been well studied in management science, where in certain corporate settings middle managers interact the most with peers to manage subordinates below them [34]. These observations reflect a democratic hierarchy as opposite to a conventional autocratic organization [35].

Of particular interest for hierarchical organization is the so-called bow-tie structure, meaning the intermediate layers have fewer components than the input and output layers. For example, in a signaling network, a large number of receptors corresponding to diverse stimuli and many transcription factors form the input and output layers, whereas the intermediate layer refers to a few key molecules like calcium and cAMP that mediate the inputs and outputs [31]. Similarly, in the networking architecture of the Internet, various protocols in the input/link layer (ARP, RARP, NDP etc) and various application protocols in the application/output layer (HTTP, FTP,DHCP etc) are essentially connected by only IPv4, the primary protocols in the internet layer. The reason for the emergence of such a common pattern is still widely open, a recent paper suggested bow-tie is a result of information compression [47].

Box 2 Tinkerer versus engineer

Despite the apparent differences, the similarity between biological systems and technological systems draws a parallel between tinkerer and engineer, and the parallel points to a common framework to unite them. Wagner further proposed an analogy between the genotype space for a biological system and the design space for a technological system. These spaces contain all the possible networks in the corresponding systems. In biology, many attempts have been made to search for solutions of common functional problems such as adaptation, oscillation and cell polarization [14]. Similar studies were performed in the context of circuit design, where a set of logic gates was evolved via rewiring in order to perform a predefined computational task [36][37]. These studies suggested that in both kinds of systems, the solution networks are close together in the genotype/design space. As each solution in genotype/design has multiple neighbors, robustness of a solution to mutation facilitates the evolvability of these systems [38][39]. Indeed, it has been demonstrated that electronic circuits can be evolved to fulfill a fluctuating evolutionary goal [36]. Similarly, metabolic networks of bacteria living in multiple habitats are evolved to decompose multiple food sources [40][41]. Both of these networks show a level of modular organization.

Very often we picture engineers design things from scratch. In reality, as a technological system evolves, engineers are subjected to various constraints like tinkerer. In the example of internet architecture, while there are frequent innovations at the input layer that interact with a variety of networking hardware and output layers that connect with many different software applications, the internet layer with very few protocols is the bottleneck under heavy constraints and such protocols can hardly be replaced [42]. The observed rapid innovation at the top and bottom layers but constraint at the middle is very common in biological system. Consider the metabolic networks of different bacteria, the anabolic and catabolic components are much more diverse whereas there are less variations between central pathways \cite{2}.

[1] L. Pachter, “\*Seq: functional genomics assays based on high-througphput sequencing,” *Bits of DNA: Reviews and commentary on computational biology*, 2014. .

[2] M. Baker, “Big biology: The ’omes puzzle,” *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013.

[3] Tony Hey, Stewart Tansley, and Kristin Tolle, *4th Paradigm*. Microsoft Research, 2009.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*, 1 edition. Burlington, MA ; San Diego: Academic Press, 2006.

[10] A. V&aacute;zquez, A. Flammini, A. Maritan, and A. Vespignani, “Modeling of Protein Interaction Networks,” *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.

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

[12] 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.

[13] 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.

[14] 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.

[15] U. Alon, “Biological Networks: The Tinkerer as an Engineer,” *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.

[16] 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.

[17] 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.

[18] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.

[19] 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.

[20] 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.

[21] J. H. Fowler, J. E. Settle, and N. A. Christakis, “Correlated genotypes in friendship networks,” *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.

[29] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, “OrthoClust: An orthology-based network framework for expression clustering across multiple species,” *Genome Biol.*, vol. 15, p. R100.

[30] 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.

[31] N. Polouliakh, R. Nock, F. Nielsen, and H. Kitano, “G-Protein Coupled Receptor Signaling Architecture of Mammalian Immune Cells,” *PLoS ONE*, vol. 4, no. 1, p. e4189, Jan. 2009.

[32] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, “Architecture of the human regulatory network derived from ENCODE data,” *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.

[33] 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.

[34] 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.

[35] Y. Bar-Yam, D. Harmon, and B. de Bivort, “Attractors and Democratic Dynamics,” *Science*, vol. 323, no. 5917, pp. 1016–1017, Feb. 2009.

[36] 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.

[37] K. Raman and A. Wagner, “The evolvability of programmable hardware,” *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269 –281, Feb. 2011.

[38] A. Wagner, “Neutralism and selectionism: a network-based reconciliation,” *Nat. Rev. Genet.*, vol. 9, no. 12, pp. 965–974, Dec. 2008.

[39] J. Masel and M. V. Trotter, “Robustness and Evolvability,” *Trends Genet.*, vol. 26, no. 9, pp. 406–414, Sep. 2010.

[40] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, “The evolution of modularity in bacterial metabolic networks,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 19, pp. 6976–6981, May 2008.

[41] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, “Toolbox model of evolution of prokaryotic metabolic networks and their regulation,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 24, pp. 9743–9748, Jun. 2009.

[42] S. Akhshabi and C. Dovrolis, “The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, New York, NY, USA, 2011, pp. 206–217.

[43] B. Prud’homme and N. Gompel, “Evolutionary biology: Genomic hourglass,” *Nature*, vol. 468, no. 7325, pp. 768–769, Dec. 2010.

[44] 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.

[45] I. S. Peter and E. H. Davidson, “Evolution of Gene Regulatory Networks Controlling Body Plan Development,” *Cell*, vol. 144, no. 6, pp. 970–985, Mar. 2011.

[46] D. L. Stern and V. Orgogozo, “Is Genetic Evolution Predictable?,” *Science*, vol. 323, no. 5915, pp. 746–751, Feb. 2009.

[47] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, “Evolution of bow-tie architectures in biology,” *ArXiv14047715 Q-Bio*, Apr. 2014.

Version: 9/11/2014

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

Throughout the history of science, advancements of 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. In the era of systems biology, attention has shifted from individual molecular components to their interactions at a system level. New functional genomics assays, in particular ones based on high-throughput sequencing (\*Seq) [1], enables biologists to probe thousands of ‘omes [2] – the comprehensive collections of constituents. One may wonder which discipline will contribute the most to biology in this new scientific paradigm [3]. While the influx of ideas in the age of reductionism mostly originated from specific areas in physics or chemistry, to understand biology via a systems perspective, the new wave of catalysts come from areas of science that are far apart, as diverse as engineering, behavioral science, sociology, but are centered on the concept of network [4].

Networks are by no mean new to biologists [5]. Metabolic pathways have been studied for decades. But more recently, as a result of the advancements of high-throughput techniques, simple pathways have been expanded to intertwined wiring diagrams. While many of us have been astonished by the complexity of such networks found in genomics or systems biology, few are able to gain any intuition from the hairballs [6]. In this essay, we argue that, by cross-disciplinary network comparison, algorithms or mathematical techniques as well as intuitions developed in commonplace networks can be able to catalyze our understanding of biology. One may wonder, however, comparing a bio-molecular network with a complex network from a disparate field, say a social network, sounds like comparing apples to oranges. So what kinds of comparison could truly deepen our understanding? We believe that it is useful to think of different descriptions of a cellular system as a spectrum (Figure 1).

**A spectrum of cellular descriptions**

Given the complexity of a cell, a certain level of simplification is necessary for useful discussion. We could picture the description of cellular systems as a spectrum (Figure 1). On one hand, there’s a simple parts list that just enumerates each component without specifying any relationships. On the other hand, there is a complete three or even four-dimensional picture of how cellular molecules interact in space and time. It is well regarded that the characteristics of a cellular system cannot be explained by the characteristics of individual components – the whole is greater than the sum of its parts. Therefore, while the full picture is often too ambitious for current data acquisition, the parts list description is not fully informative. Network description sits conveniently between these extremes by capturing the relationships between components of the parts list. There are broadly two ways to define relationships based on the nature of experimental data. Firstly, networks could be defined in a phenomenological sense, meaning entries are connected via abstract mathematical relationship derived from phenomenological observables. Perhaps the most important phenomenological networks are built on the mapping between genotypes and phenotypes. An example is the disease networks [7], 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. While phenomenological networks offer a mathematical abstraction far away from the complete picture in the spectrum, there are networks defined to capture the mechanistic interactions happening inside a living system. Such networks are built on experimental knowledge on different facets of the complex organization of an organism, for instance, a regulatory network describes part of the cellular information processing, a metabolic network traces the chemistry of metabolites, and the protein-protein interaction network captures cell signaling as well as providing a manual on how to assemble molecular machines. The integration of such mechanistic networks provides a description reasonably close to the complete picture along the spectrum. The process could be visually regarded as the decoration of edges with directionality, color, thickness etc. Nevertheless, the incorporation of too much detail makes the system intractable, and network formalism generally breaks down if spatial or temporal details as well as higher-order interactions are included. At certain point, the actual four-dimensional picture is required. [[CC: more details??]] [[CC: some network may not follow into either phenomenological or mechanistic network. e.g. genetic network from synthetic lethal: interaction are determined by experiments, but the mechanism behind of them are quite complicated. ]] [[CC: do we need to prepare a table to show examples of the two types networks??]]

Underlying mechanistic networks serve as the skeletons of different complex systems. Comparison of such networks allow biologists to gain intuitions by examining interactions in cross-disciplinary complex systems in the same ground as the interactions between molecular components in cells. Nevertheless, because of systems-specific details, not everything could be transferred from one discipline to another, and it is important to focus on the conceptual resemblance instead of merely topological resemblance. [[CC: this part is hard to understand. Better provide some concrete examples]] On the other hand, phenomenological networks are rather abstract connections between entities, and thus mathematical formalisms are easily transferrable. Toward this end, by comparing similar network-based mathematical formalisms across disciplines, biologists will benefit in terms of algorithmic or method development.

**Comparing phenomenological networks to leverages mathematical machineries**

Phenomenological networks are typical products dealing with big data; they are essentially two-dimensional projection of high-dimensional data. As it is extremely common to have data with many features in the era of Big Data, especially in computational social science, networks across disciplines actually present very similar challenges. Here, we highlight a few areas where different questions arise in genomics and social science could be formulated by the very same approach. By the same token, network algorithms developed in one discipline can readily be applied in biology.

*The modular organization of biological networks*

Even though the evolutionary process involves random changes at the molecular level, it is not surprising that natural selection organizes biological networks in an ordered fashion. Comparison of biological networks with different social and technological networks have provided valuable insights into the organizing principles of biological networks. One striking outcome of the comparison is that these natural and man made networks organize into scale free networks in which a majority of the nodes contain very few connections (edges) while a few nodes in the network are highly connected. Changes to the sequences of the centrally connected nodes (also called hubs) in biological networks are predictive of lethal and disease causing changes and have proven to be valuable in interpreting the genotype to phenotype network. [[CC: the subtitle of this section is about “phenomenological networks”, but concepts of scale-free and modularity were first proposed in PPI, which form “mechanistic” networks. ]] [[ANS: I agree with CC]]

Another striking insight from these comparisons is these networks are organized in the form of modules such that nodes within the same module have a larger number of connections with each other as compared to nodes belonging to different modules. Evolutionarily this makes sense because connections within a module can be reused in a different functional context. Naturally, genes within the same module have similar biological properties. For example, the cellular components within the same module are more closely associated with the same set of cellular phenotypes than components belonging to different modules. Similarly, the modules within gene coexpression networks also tend to contain genes with similar functions and genes within the same module are often involved in the same biological pathway.

As biological networks are multidimensional in nature, we think a large number of insights can be gained by comparing the its organization with the multiplex structure of different technological and social networks. Multiplex networks contain multiple layers of interconnected networks - the multiple layers in these networks could either be formed by different categories of relationships (co-expression, genetic interactions, etc) or they could be formed by relationships observed at different timepoints. The idea originated in social network analysis because an individual may participate in multiple social circles: family, friends, colleagues, or in online setting: Facebook, Linkedln and Twitter. Similarly, the different layers in a temporal network contain parts of the network that are connected at different timepoints [28]. As dynamic data in genomic information becomes available, we think that valuable insights can be gleaned by the analysis of these data using algorithms developed in the context of multiplex social networks. [[do we need to give one or more examples?]]

Nevertheless, biology motivates an alternate definition of temporal network. While they exist together at the same time-point, networks from different species essentially capture the evolutionary changes to a common core. In this definition, pairs of orthologous genes can be used to connect networks from different species, forming a multi-layers structure. The notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [29]. Based on the same notion, a mathematical formalism was developed to measure the evolutionary rewiring rate between networks across species in analogous to quantifying sequence evolution [30]. It was shown that metabolic networks rewire at a slower rate compared to various regulatory networks.

*Formalisms for association and prioritization of nodes*

In both biology and computational social science, the networks are often incomplete and we may need to infer the properties of various nodes and edges based upon incomplete data. For example, if your friends in Facebook use Product Y, you are more likely to use product Y and the advertisements you view online are personalized based on these recommendation systems. The essence of these methods is the idea of “guilt by association” or the idea that nodes that have similar associations in the network tend to be more similar in nature. Such methods are also used in genomics, for example, to infer the functions of a protein or a non-coding element based on the function of its neighbors in the underlying network. The same is true for predicting disease-associated genes: if the neighbors of a gene are all associated with Disease X, it is very likely that the gene is associated with disease X. Interestingly, because of the availability of datasets, for instance the Framingham study, there is an increase of attention on the connection between genomics information and sociological information. Biologists and sociologists have started to examine the hypothesis on whether phenotypes or genotypes are correlated in friendship networks [21].

Nodes association is closely related to nodes prioritization, in which the PageRank algorithm plays an important role. Originated from Katz centrality in social network analysis [22], PageRank algorithm was first used by Google to rank documents based on linkages in a self-consistent way. The algorithm was then adopted in food webs to determine extinction [23] and later in an algorithm called NetRank that rank prognostic relevance for patients with cancers [24]. Generally speaking, in addition to algorithms like PageRank that prioritize nodes by network topology, expression data, sequence information, functional annotation and biomedical literature are required for further filtering [25]. In applications like disease gene discovery, nodes prioritization is an essential process because of limited resources. Similarly, social networks can be utilized to identify the people that need to be vaccinated so that the spread of a disease can be avoided.

[[DW]]*Formalisms for network pathways (or flows)*

Ones always are interested in how a node is related to another node if they are not neighbors in a network in both biology and other disciplines. For example, the professionals (also in Facebook) suggested by Linkedin are normally connecting to many your connections or professionals connecting to your connections (i.e., 2nd or 3rd connections). You are also interested how you relate to influential persons. Similarly, in genome, two genes may not directly interact with each other, but are associated via other intermediary genes like forming a regulatory pathway. Therefore, the network pathways connecting to two nodes have been analyzed broadly. Especially, beyond simple identification of pathway nodes and edges, ones normally quantify the pathways and prioritizethem based on their interests. For example [[DW: circuit itself may be mechanistic network, also simply adding electronic element delays is pretty abstract]], the critical pathway in VLSI design from electronic circuits is the pathway from input to output that have longest timing delay. Electrical engineers always try to minimize the critical pathways to improve circuit designs. In metabolic flux analysis, ones are also interested in finding the metabolic pathways that have fastest consumption rates.

*Formalisms for inference of edges* [[ANS: I think this should go more into dynamics now]][[DW: yes, I agree]]

*Formalisms for inference of edges from dynamics? [[DW]]*

As the construction of various phenomenological and social networks an active area of research for both biology and computational social science. [[DW: add transition: The observed network datasets are normally dynamic, rather than static in both spatial and temporal dimensions. For example, the co-expressed gene modules were found to vary during embryonic development. How can we infer the embryonic developmental gene regulatory network from temporal dynamics in gene expression; i.e., inference of the regulatory edges between TFs and target genes? This type of network edges does not describe correlational relationships because TFs and their targets are likely not co-expressed (i.e., low correlation). These edges tell us a causal relationship.]] While correlational relationships could potentially be easily calculated with the appropriate data, a fundamental question is the distinction between direct and indirect interactions. This is of particular importance for biology in terms of identifying the master regulator of a disease. The same application is true for social networks for identifying the source of influence. Established mathematical machineries like Bayesian networks or Markov random fields have been used for this purpose. The inference of causal relationships could be greatly benefited by time-series data. The question is extremely common in biology, under the term “reverse engineering”. Ideally, one could write differential equations to fit the temporal data, nevertheless, temporal data in most genomics experiments do not have enough time-points. To overcome the drawback, for instance, given the genome-wide expression profile of at different time-points, one could perform project the high-dimensional gene expression data to low dimensional space by data mining techniques such as SVD, and write differential equations to model the dynamics of the projections [26]. [[It is not clear to me how networks help in this or networks from different fields help in this section? -ANS]]

[[DW: Thus, dynamic modelling of abstract networks can help us to understand mechanistic networks that normally have causal edges. like co-expression infers co-regulation ]]

[[DW: this can be also a transition from abstract network to mechanistic network]]

Many networks in biology and social science are noisy and incomplete, leading to common challenges like link prediction and denoising. Difficulties lie at the proper learning of network organization based on observable data. Recently, generative models of networks, say stochastic block models [27], are very popular in computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standard for validation.

[[For me this whole section has been written from the point of view that algorithms are the common points for transfer of information rather than ideas from different fields - Do we want to focus on the ideas or the algorithms - For example, from the experimentalists perspective, does it matter that the algorithms are transferrable - For them, it is more important that some biological problems could be solved because of insights gained in other fields - ANS]]

**Comparison of Mechanistic networks for gaining intuition**

From a biologist standpoint, comparing various mechanistic networks between biology and other disciplines can bring intuition from other disciplines into biology. In spite of the disparate fields, we believe there are several areas biologists could find inspiration.

*Looking for universal mechanisms*

Since the burgeoning of studying networks in various disciplines, efforts have been made on explaining some of the striking similarity in terms of organization of underlying networks in biological and other complex systems. An early example is the emergence of the scale-free degree distribution in a protein-protein interactions network. The pattern of organization could be explained by the duplication divergence model [10], a simple stochastic process describing how a protein network grows by gene duplication. As a hub protein has many interactions, its number of interactions is likely to increase further simply because one of its neighbors got duplicated. The same “richer get richer” model was proposed originally to explain the same pattern in many other networks [11]. More recently, it has been shown that components in both bacterial genomes as well as large-scale computer software projects form multilayered dependency networks (enzyme A is used to decompose the output metabolites of enzyme B; the installation of package A depends on the installation of package B). The common underlying dependency networks leads to the same power-law components-usage frequency distribution (how often a enzyme is present in a bacterial genome; how often a certain package is installed in a computer) [12].

While it is elegant to explain the topology of disparate networks by simple stochastic models, such universal mechanisms are rather rare. To a certain extent, the existence of such models underlines the importance of randomness in biology. Remarkably, the same duplication-divergence mechanism has been applied to describe the patterns of “memes” in online media [13]. As biologists, we love to think about functions and selection; it is interesting to see that, by network comparison, network organization could be a manifestation of stochasticity.

*Looking for common design principles*

Of course, biological networks are not random, and so do networks from other disciplines. Most observed similarities in terms of network organization are not easy to explain by simple mechanisms or principles, for instance, the so-called network hierarchy (see Box 1). The reason is because, for most networks, it is in general very hard to define a “function”. In fact, lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. The mapping points to biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [14]. Toward this direction, comparison with various technological or engineering networks with well-defined functions is particularly insightful. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a source of negative feedback and a source of time delay. Nevertheless, different oscillators (e.g. for circadian rhythms, for cell cycle, or from various organisms) have a certain level of variation because of additional design objectives or strategies. This is just like the case that not all electronic devices use the same oscillator design because of other design objectives. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [15]. Robustness is obviously a preferred design objective because it makes a system tolerate stochastic fluctuations, either intrinsically or from external sources. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by interface is widely practiced [16]. The same is for biological networks because modules can be readily reused to adapt new functions. Because of the fundamental importance of such design objectives, an insightful network comparison should be rooted in the common design objectives rather than merely network topology.

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

The comparison of biological networks and technological networks should best be performed under the light of evolution. As Alon highlighted by the phase “the tinkerer as an engineer” [15], it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer neither designs things nor builds systems— it settles on systems that, historically, conveyed a survival benefit (and if a better way comes along, it will adopt that). On the other hand, technological networks are essentially blueprints drawn by engineers who have a grand plan that makes sure everything work harmoniously. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that construct living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological networks like roadways and circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [17]. To a certain extent, an engineer is a tinkerer (see Box 2).

Under such a united framework, we could picture that both engineer and tinkerer are working on an optimization problem with similar underlying design objectives. Like all optimization problems, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there’s no free lunch [18][19]. Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited in the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many different processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fraction of functions are updated often [20]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, software system is very similar to a roadway system; bottlenecks under high usage like George Washington Bridge require more upgrade and more construction. While intentional tweaking on bottlenecks sounds obvious for technological systems, it is not always possible (see Box 2).

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of ‘omes across species. To nourish a system-level understanding and to leverage the tremendous amount of high-throughput data, may be it is a time to extend our tradition even further to compare with networks from other complex systems as well as other disciplines. Indeed, various scientific disciplines form a network in the intellectual universe where knowledge emerges when things connect.

Box 1 Hierarchical organization of networks

Many biological networks possess an intrinsic direction of information flow, forming a hierarchical network organization. The hierarchical organization in biological networks resemble the chain of command in human society, like in military context and corporate hierarchy [8]. For instance, in a transcriptional regulatory network more influential transcription factors (regulators whose expression are more highly correlated with the expression of target genes) tend to be better connected (have more interacting partners) and higher in the hierarchy [32]. Moreover, the transcription factors in the middle layer tend to be more cooperative [33]. Such a situation has been well studied in management science, where in certain corporate settings middle managers interact the most with peers to manage subordinates below them [34]. These observations reflect a democratic hierarchy as opposite to a conventional autocratic organization [35].

Of particular interest for hierarchical organization is the so-called bow-tie structure, meaning the intermediate layers have fewer components than the input and output layers. For example, in a signaling network, a large number of receptors corresponding to diverse stimuli and many transcription factors form the input and output layers, whereas the intermediate layer refers to a few key molecules like calcium and cAMP that mediate the inputs and outputs [31]. Similarly, in the networking architecture of the Internet, various protocols in the input/link layer (ARP, RARP, NDP etc) and various application protocols in the application/output layer (HTTP, FTP,DHCP etc) are essentially connected by only IPv4, the primary protocols in the internet layer. The reason for the emergence of such a common pattern is still widely open, a recent paper suggested bow-tie is a result of information compression [47].

Box 2 Tinkerer versus engineer

Despite the apparent differences, the similarity between biological systems and technological systems draws a parallel between tinkerer and engineer, and the parallel points to a common framework to unite them. Wagner further proposed an analogy between the genotype space for a biological system and the design space for a technological system. These spaces contain all the possible networks in the corresponding systems. In biology, many attempts have been made to search for solutions of common functional problems such as adaptation, oscillation and cell polarization [14]. Similar studies were performed in the context of circuit design, where a set of logic gates was evolved via rewiring in order to perform a predefined computational task [36][37]. These studies suggested that in both kinds of systems, the solution networks are close together in the genotype/design space. As each solution in genotype/design has multiple neighbors, robustness of a solution to mutation facilitate the evolvability of these systems [38][39]. Indeed, it has been demonstrated that electronic circuits can be evolved to fulfill a fluctuating evolutionary goal [36]. Similarly, metabolic networks of bacteria living in multiple habitats are evolved to decompose multiple food sources [40][41]. Both of these networks show a level of modular organization.

Very often we picture engineers design things from scratch. In reality, as a technological system evolves, engineers are subjected to various constraints like tinkerer. In the example of internet architecture, while there are frequent innovations at the input layer that interact with a variety of networking hardware and output layers that connect with many different software applications, the internet layer with very few protocols is the bottleneck under heavy constraints and such protocols can hardly be replaced [42]. The observed rapid innovation at the top and bottom layers but constraint at the middle are very common in biological system. Consider the metabolic networks of different bacteria, the anabolic and catabolic components are much more diverse whereas there are less variations between central pathways \cite{2}.

Box 3

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

May be giving a few more examples of phenomenological networks, like genetic interaction networks.

[1] L. Pachter, “\*Seq: functional genomics assays based on high-througphput sequencing,” *Bits of DNA: Reviews and commentary on computational biology*, 2014. .

[2] M. Baker, “Big biology: The ’omes puzzle,” *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013.

[3] Tony Hey, Stewart Tansley, and Kristin Tolle, *4th Paradigm*. Microsoft Research, 2009.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*, 1 edition. Burlington, MA ; San Diego: Academic Press, 2006.

[10] A. V&aacute;zquez, A. Flammini, A. Maritan, and A. Vespignani, “Modeling of Protein Interaction Networks,” *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.

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

[12] 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.

[13] 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.

[14] 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.

[15] U. Alon, “Biological Networks: The Tinkerer as an Engineer,” *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.

[16] 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.

[17] 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.

[18] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.

[19] 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.

[20] 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.

[21] J. H. Fowler, J. E. Settle, and N. A. Christakis, “Correlated genotypes in friendship networks,” *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.

[29] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, “OrthoClust: An orthology-based network framework for expression clustering across multiple species,” *Genome Biol.*, vol. 15, p. R100.

[30] 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.

[31] N. Polouliakh, R. Nock, F. Nielsen, and H. Kitano, “G-Protein Coupled Receptor Signaling Architecture of Mammalian Immune Cells,” *PLoS ONE*, vol. 4, no. 1, p. e4189, Jan. 2009.

[32] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, “Architecture of the human regulatory network derived from ENCODE data,” *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.

[33] 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.

[34] 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.

[35] Y. Bar-Yam, D. Harmon, and B. de Bivort, “Attractors and Democratic Dynamics,” *Science*, vol. 323, no. 5917, pp. 1016–1017, Feb. 2009.

[36] 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.

[37] K. Raman and A. Wagner, “The evolvability of programmable hardware,” *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269 –281, Feb. 2011.

[38] A. Wagner, “Neutralism and selectionism: a network-based reconciliation,” *Nat. Rev. Genet.*, vol. 9, no. 12, pp. 965–974, Dec. 2008.

[39] J. Masel and M. V. Trotter, “Robustness and Evolvability,” *Trends Genet.*, vol. 26, no. 9, pp. 406–414, Sep. 2010.

[40] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, “The evolution of modularity in bacterial metabolic networks,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 19, pp. 6976–6981, May 2008.

[41] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, “Toolbox model of evolution of prokaryotic metabolic networks and their regulation,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 24, pp. 9743–9748, Jun. 2009.

[42] S. Akhshabi and C. Dovrolis, “The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, New York, NY, USA, 2011, pp. 206–217.

[43] B. Prud’homme and N. Gompel, “Evolutionary biology: Genomic hourglass,” *Nature*, vol. 468, no. 7325, pp. 768–769, Dec. 2010.

[44] 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.

[45] I. S. Peter and E. H. Davidson, “Evolution of Gene Regulatory Networks Controlling Body Plan Development,” *Cell*, vol. 144, no. 6, pp. 970–985, Mar. 2011.

[46] D. L. Stern and V. Orgogozo, “Is Genetic Evolution Predictable?,” *Science*, vol. 323, no. 5915, pp. 746–751, Feb. 2009.

[47] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, “Evolution of bow-tie architectures in biology,” *ArXiv14047715 Q-Bio*, Apr. 2014.

Version: 8/27/2014 (see the pdf attached in my mail if you are interested at the track changes)

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

Throughout the history of science, advancements of 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. In the era of systems biology, attention has shifted from individual molecular components to their interactions at a system level. New functional genomics assays, in particular ones based on high-throughput sequencing (\*Seq) [1], enables biologists to probe thousands of ‘omes [2] – the comprehensive collections of constituents. One may wonder which discipline will contribute the most to biology in this new scientific paradigm [3]. While the influx of ideas in the age of reductionism mostly originated from specific areas in physics or chemistry, to understand biology via a systems perspective, the new wave of catalysts come from areas of science that are far apart, as diverse as engineering, behavioral science, sociology, but are centered on the concept of network [4].

Networks are by no mean new to biologists [5]. Metabolic pathways have been studied for decades. But more recently, as a result of the advancements of high-throughput techniques, simple pathways have been expanded to intertwined wiring diagrams. While many of us have been astonished by the complexity of such networks, few are able to gain any intuition from the hairballs [6]. While the term “biological network” is used rather loosely in literature for all networks originating from any subfields of biology, say food web, here in this essay, we focus our attention to molecular networks coming from genomics or systems biology because it is in general harder to gain intuitions from such networks. We argue that, intuitions as well as algorithms or mathematical techniques developed in commonplace networks from emerging disparate disciplines can be able to catalyze our understanding of biology. Therefore it is instructive to initiate comparison between biological networks with networks in other disciplines.

Comparing a bio-molecular network with a complex network from a distant field, say a social network, sounds like comparing apples to oranges. So what kinds of comparison could truly deepen our understanding? We believe the focus of comparison should depend on the types of relational information captured in networks.

**Comparison of Mechanistic networks for gaining intuition**

It is well regarded that the characteristics of a cellular system cannot be explained by the characteristics of individual components – the whole is greater than the sum of its parts. The essence of network is to describe the relationships between components of the parts-list (genes, proteins, small molecules etc.). Given the complexity, a certain level of simplification is necessary in the defining the relationships. A natural way is by using various kinds of mechanistic interactions. Such networks essentially capture different facets of the complex organization of an organism, for instance, a regulatory network describes part of the cellular information processing, a metabolic network traces the chemistry of metabolites, and the protein-protein interaction network captures cell signaling as well as providing a manual on how to assemble molecular machines. Depending on the nature of interactions, mechanistic networks resemble, and should be compared with various commonplace networks. For instance, signaling networks resembles certain chains of command in human society (e.g. corporate hierarchy) in terms of information transmission [7]. Developmental transcriptional regulatory networks, on the other hand, resemble technological systems like circuits in terms of the emergence of functions (specific input-output responses) [8]. In this context, networks serve as a basic framework capturing the underlying skeletons of various complex systems. It thus allows biologists to gain intuitions by examining interactions in cross-disciplinary complex systems in the same ground as the interactions between molecular components in cells.

*Looking for universal mechanisms*

Since the burgeoning of studying networks in various disciplines, efforts have been made on explaining some of the striking similarity in terms of organization of underlying networks in biological and other complex systems. An early example is the emergence of the scale-free degree distribution in a protein-protein interactions network. The pattern of organization could be explained by the duplication divergence model [9], a simple stochastic process describing how a protein network grows by gene duplication. As a hub protein has many interactions, its number of interactions is likely to increase further simply because one of its neighbors got duplicated. The same “richer get richer” model was proposed originally to explain the same pattern in many other networks [10]. More recently, it has been shown that components in both bacterial genomes as well as large-scale computer software projects form multilayered dependency networks (enzyme A is used to decompose the output metabolites of enzyme B; the installation of package A depends on the installation of package B). The common underlying dependency networks leads to the same power-law components-usage frequency distribution (how often a enzyme is present in a bacterial genome; how often a certain package is installed in a computer) [11].

While it is elegant to explain the topology of disparate networks by simple stochastic models, such universal mechanisms are rather rare. To a certain extent, the existence of such models underlines the importance of randomness in biology. Remarkably, the same duplication-divergence mechanism has been applied to describe the patterns of “memes” in online media [12]. As biologists, we love to think about functions and selection; it is interesting to see that, by network comparison, network organization could be a manifestation of stochasticity.

*Looking for common design principles*

Of course, biological networks are not random, and so do networks from other disciplines. Most observed similarities in terms of network organization are not easy to explain by simple mechanisms or principles, for instance, the so-called network hierarchy (see Box 1). The reason is because, for most networks, it is in general very hard to define a “function”. In fact, lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. The mapping points to biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [13]. Toward this direction, comparison with various technological or engineering networks with well-defined functions is particularly insightful. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a negative feedback loop and a source of time delay. Nevertheless, oscillators of various purposes (e.g. for circadian rhythms or for cell cycle) or from various organisms are not identical but have a certain level of variation because additional design objectives or strategies are involved. Just like not all electronic devices use the same oscillator design, the importance of design objectives is not new at all in engineering systems. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [14]. Robustness is obviously a preferred design objective because it makes a system tolerate intrinsic or extrinsic stochastic fluctuations. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by interface is widely practiced [15]. The same is for biological networks because modules can be readily reused to adapt new functions. Because of the fundamental importance of such design objectives, an insightful network comparison should be rooted in the common design objectives rather than merely network topology.

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

The comparison of biological networks and technological networks should best be performed under the light of evolution. As Alon highlighted by the phase “the tinkerer as an engineer” [14], it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer starting with bits and pieces and trying to connect random nodes, whereas technological networks are essentially blueprints drawn by engineers. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that construct living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological networks like roadways and circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [16]. To a certain extent, an engineer is a tinkerer (see Box 2).

Under such a united framework, we could picture that both engineer and tinkerer are working on an optimization problem with similar underlying design objectives. Like all optimization problems, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there’s no free lunch [17][18]. Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited in the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many different processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fraction of functions are updated often [19]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, software system is very similar to a roadway system; bottlenecks under high usage like George Washington Bridge require more upgrade and more construction. While intentional tweaking on bottlenecks sounds obvious for technological systems, it is not always possible (see Box 2).

**Comparing phenomenological networks to leverages mathematical machineries**

As mechanistic networks are “skeletons” of a cell, defining edges by a particular kind of mechanistic interaction means flesh has been omitted. For example, a simple protein-protein interactions network usually does not capture the spatial or temporal properties of binding. While in some cases a certain level of details could be incorporated, the framework may either be too specific for a particular application, or simply intractable. The scenario is analogous to classical physics; writing down the equations of all the particles is intractable, and thus physicists turn to a phenomenological and macroscopic formalism, i.e. thermodynamics. Similarly, certain networks are defined in a phenomenological sense. These networks do not capture the details of biological processes happening inside a living system, but provide an abstraction for further mathematical formulation. Perhaps the most important phenomenological networks are built on the mapping between genotypes and phenotypes. An example is the disease networks [20], 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. Phenomenological networks are typical products dealing with big data; they are essentially two-dimensional projection of high-dimensional data. As it is extremely common to have data with many features in the era of Big Data, especially in computational social science, networks across disciplines actually present very similar challenges. By the same token, network algorithms developed in one discipline can readily be applied in biology. Toward this end, by comparing similar network formalisms, biologists will benefit from an algorithmic or method development standpoint (see Box 3).

*Formalisms for association and prioritization of nodes*

In both biology and computational social science, networks are often used as a map for integrating various features. A general question of interest is to infer the properties of certain nodes. Though various kernel methods have been introduced, the essence of all solutions is the idea of “guilt by association”. In genomics, a widely used approach to infer the functions of a protein or a non-coding element is based on the function of its neighbors in the underlying network. The same is true for predicting disease-associated genes: if the neighbors of a gene are all associated with Disease X, it is very likely that the gene is associated with disease X. Online advertisers use the same trick. If your friends in Facebook use Product Y, you are more likely to use product Y and thus will be targeted. Interestingly, because of the availability of datasets, for instance the Framingham study, there is an increase of attention on the connection between genomics information and sociological information. Biologists and sociologists have started to examine the hypothesis on whether phenotypes or genotypes are correlated in friendship networks [21].

Nodes association is closely related to nodes prioritization, in which the PageRank algorithm plays an important role. Originated from Katz centrality in social network analysis [22], PageRank algorithm was first used by Google to rank documents based on linkages in a self-consistent way. The algorithm was then adopted in food webs to determine extinction [23] and later in an algorithm called NetRank that rank prognostic relevance for patients with cancers [24]. Generally speaking, in addition to algorithms like PageRank that prioritize nodes by network topology, expression data, sequence information, functional annotation and biomedical literature are required for further filtering [25]. In applications like disease gene discovery, nodes prioritization is an essential process because of limited resources. In social science setting, the same is true for applications like online advertising.

*Formalisms for inference of edges*

The construction of various phenomenological networks an active area of research for both biology and computational social science. While correlational relationships could potentially be easily calculated with the appropriate data, a fundamental question is the distinction between direct and indirect interactions. This is of particular importance for biology in terms of identifying the master regulator of a disease. The same application is true for social networks for identifying the source of influence. Established mathematical machineries like Bayesian networks or Markov random fields have been used for this purpose. The inference of casual relationships could be greatly benefited by time-series data. The question is extremely common in biology, under the term “reverse engineering”. Ideally, one could write differential equations to fit the temporal data, nevertheless, temporal data in most genomics experiments do not have enough time-points. To overcome the drawback, for instance, given the genome-wide expression profile of at different time-points, one could perform project the high-dimensional gene expression data to low dimensional space by data mining techniques such as SVD, and write differential equations to model the dynamics of the projections [26].

Many networks in biology and social science are noisy and incomplete, leading to common challenges like link prediction and denoising. Difficulties lie at the proper learning of network organization based on observable data. Recently, generative models of networks, say stochastic block models [27], are very popular in computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standard for validation.

*Formalisms for multi-layers network structure*

A recent trend of network analysis is the notion of multiplex networks where multiple layers of networks form an interconnected structure. The idea is originated in social network analysis because an individual may participate in multiple social circles: family, friends, colleagues, or in online setting: Facebook, Linkedln and Twitter. The same is true in biological context because of the existence of multiple relational connections (co-expression, genetic interactions etc.) between components in networks. While different layers of networks are categorical in this formalism, a similar multi-layers generalization in network analysis is the so-called temporal networks. In short, a temporal network considers the slices of networks taking place at different time points together as a single mathematical structure [28]. Again, the current application focuses on online social networks because genome-wide data in biological systems are still not dynamics enough. However, as the number of time points increases, say in RNA-Seq experiments, algorithms developed in social contexts can be easily applied to integrate the slices of co-expression networks.

Nevertheless, biology motivates an alternate definition of temporal network. While they exist together at the same time-point, networks from different species essentially capture the evolutionary changes to a common core. In this definition, pairs of orthologous genes can be used to connect networks from different species, forming a multi-layers structure. The notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [29]. Based on the same notion, a mathematical formalism was developed to measure the evolutionary rewiring rate between networks across species in analogous to quantifying sequence evolution [30]. It was shown that metabolic networks rewire at a slower rate compared to various regulatory networks.

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of ‘omes across species. To nourish a system-level understanding and to leverage the tremendous amount of high-throughput data, may be it is a time to extend our tradition even further to compare with networks from other complex systems as well as other disciplines. Indeed, various scientific disciplines form a network in the intellectual universe where knowledge emerges when things connect.

[[KKY2MG: the texts here are just snippets extracted from pervious writing, not coherent]]

Box 1 Hierarchical organization of networks

For instance, many biological networks possess an intrinsic direction of information flow, such as signaling networks where information propagates from G-Protein coupled receptors to transcription factors [31], forming a hierarchical network organization. The hierarchical organization in biological networks resemble certain the chain of command in human society, like in military context and corporate hierarchy [7]. For instance, more influential transcription factors (regulators whose expression are more highly correlated with the expression of target genes) tend to be better connected and higher in the hierarchy [32]. Moreover, the cooperative regulatory factors in a transcriptional regulatory network tend to be in the middle layer [33]. This situation is well studied in management science, where in certain corporate settings middle managers interact the most with peers to manage subordinates below them [34]. Such observations reflect a democratic hierarchy as opposite to a conventional autocratic organization [35].

Box 2 Tinkerer versus engineer

The parallel between tinkerer and engineer points to a common framework to unite them. Wagner further proposed an analogy between the genotype space for a biological system and the design space for a technological system. These spaces contain all the possible networks in the corresponding systems. In biology, many attempts have been made to search for solutions of common functional problems such as adaptation, oscillation and cell polarization [13]. Similar studies were performed in the context of circuit design, where a set of logic gates was evolved via rewiring in order to perform a predefined computational task [36][37]. These studies suggested that in both kinds of systems, the solution networks are close together in the genotype/design space. As each solution in genotype/design has multiple neighbors, robustness of a solution to mutation facilitate the evolvability of these systems [38][39]. Indeed, it has been demonstrated that electronic circuits can be evolved to fulfill a fluctuating evolutionary goal [36]. Similarly, metabolic networks of bacteria living in multiple habitats are evolved to decompose multiple food sources [40][41]. Both of these networks show a level of modular organization.

In the above example of internet architecture, while there are frequent innovations at the input layer that interact with a variety of networking hardware and output layers that connect with many different software applications, the internet layer with very few protocols is the bottleneck under heavy constraints and such protocols can hardly be replaced [42]. The observed rapid innovation at the top and bottom layers but constraint at the middle may shed light on a remarkably pattern in developmental genetic regulatory network. Different species exhibit different patterns at the early and late stages of embryo development, but highly similar during the phylotypic stage – the so-called hourglass phenomenon [43].

Of particular interest for hierarchical organization is the so-called bow-tie structure, meaning the intermediate layers have fewer components than the input and output layers. For example, in a developmental genetic regulatory network, information propagates from genes controlling the initial stage of development (the input) to genes controlling detailed cell differentiation and morphogenesis (output) [44][45]. The intermediate layer refers to a small set of input-output genes integrating complex spatiotemporal information and trigger development of an entire program of cell differentiation [46]. In the networking architecture of the Internet, on the other hand, various protocols in the input/link layer (ARP, RARP, NDP etc) and various application protocols in the application/output layer (HTTP, FTP,DHCP etc) are essentially connected by IPv4, the primary protocols in the internet layer. A recent paper provided a first mechanism to understand its evolution by explicitly modeling information flow in feed-forward networks as a cascade of matrix multiplications (similar to neural networks in machine learning context) [47]. It showed that a bow-tie structure emerged if the goal matrix is rank deficient, i.e. the information can be compressed.

Box 3

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

May be giving a few more examples of phenomenological networks, like genetic interaction networks.

[1] L. Pachter, “\*Seq: functional genomics assays based on high-througphput sequencing,” *Bits of DNA: Reviews and commentary on computational biology*, 2014. .

[2] M. Baker, “Big biology: The ’omes puzzle,” *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013.

[3] Tony Hey, Stewart Tansley, and Kristin Tolle, *4th Paradigm*. Microsoft Research, 2009.

[4] 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.

[5] 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.

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

[7] 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.

[8] E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*, 1 edition. Burlington, MA ; San Diego: Academic Press, 2006.

[9] A. V&aacute;zquez, A. Flammini, A. Maritan, and A. Vespignani, “Modeling of Protein Interaction Networks,” *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.

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

[11] 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.

[12] 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.

[13] 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.

[14] U. Alon, “Biological Networks: The Tinkerer as an Engineer,” *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.

[15] 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.

[16] 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.

[17] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.

[18] 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.

[19] 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.

[20] 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.

[21] J. H. Fowler, J. E. Settle, and N. A. Christakis, “Correlated genotypes in friendship networks,” *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.

[29] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, “OrthoClust: An orthology-based network framework for expression clustering across multiple species,” *Genome Biol.*, vol. 15, p. R100.

[30] 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.

[31] N. Polouliakh, R. Nock, F. Nielsen, and H. Kitano, “G-Protein Coupled Receptor Signaling Architecture of Mammalian Immune Cells,” *PLoS ONE*, vol. 4, no. 1, p. e4189, Jan. 2009.

[32] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, “Architecture of the human regulatory network derived from ENCODE data,” *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.

[33] 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.

[34] 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.

[35] Y. Bar-Yam, D. Harmon, and B. de Bivort, “Attractors and Democratic Dynamics,” *Science*, vol. 323, no. 5917, pp. 1016–1017, Feb. 2009.

[36] 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.

[37] K. Raman and A. Wagner, “The evolvability of programmable hardware,” *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269 –281, Feb. 2011.

[38] A. Wagner, “Neutralism and selectionism: a network-based reconciliation,” *Nat. Rev. Genet.*, vol. 9, no. 12, pp. 965–974, Dec. 2008.

[39] J. Masel and M. V. Trotter, “Robustness and Evolvability,” *Trends Genet.*, vol. 26, no. 9, pp. 406–414, Sep. 2010.

[40] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, “The evolution of modularity in bacterial metabolic networks,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 19, pp. 6976–6981, May 2008.

[41] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, “Toolbox model of evolution of prokaryotic metabolic networks and their regulation,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 24, pp. 9743–9748, Jun. 2009.

[42] S. Akhshabi and C. Dovrolis, “The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, New York, NY, USA, 2011, pp. 206–217.

[43] B. Prud’homme and N. Gompel, “Evolutionary biology: Genomic hourglass,” *Nature*, vol. 468, no. 7325, pp. 768–769, Dec. 2010.

[44] 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.

[45] I. S. Peter and E. H. Davidson, “Evolution of Gene Regulatory Networks Controlling Body Plan Development,” *Cell*, vol. 144, no. 6, pp. 970–985, Mar. 2011.

[46] D. L. Stern and V. Orgogozo, “Is Genetic Evolution Predictable?,” *Science*, vol. 323, no. 5915, pp. 746–751, Feb. 2009.

[47] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, “Evolution of bow-tie architectures in biology,” *ArXiv14047715 Q-Bio*, Apr. 2014.

easly snippets to be fitted in/dumped:

* the control theory paragraph...
* Parameters: robustness + sloppy….
* biological circuits need to care about stochasticity than electrical circuits
* Variation as a general idea introduced…
* Methods deal with dynamics: control theory, circuit theory, information theory and game theory
* Markov decision
* personalized medicine vs networks, personalized recommendation?
* social sys lie between bio sys and tech sys
* the documentation of technological systems in compare with the lack in biological systems
* biological innovation vs technological innovation vs knowledge innovation
* punctuated evolution..
* complexity of all these systems is because of epistasis
* Synthetic biology as forward engineering, as opposed to reverse engineering.
* [[ANS - I think this is not necessarily a given - that the rules governing lots of different networks lead to similar organization in completely different disciplines - Something like once the relationships between the parts of a complex system is put in the form of a network, the power of network theory comes from the fact that the same algorithms can be used to find the commonalities in the topology of disparate natural and man-made systems - example scale free, etc]].
* For instance, transcription regulation uses cooperative binding to arrive at a sigmoidal response curve whereas the same is achieved by an amplifier in circuit design.
* Evolution keeps tinkering and adjusting to the environment - ANS. For it to evolve, there are certain criteria that ensure that the system is not in a deep evolutionary minima - i.e., small changes to the sequence and/or network are not totally harmful
* Robustness of systems and how they can be used to understand evolution of biological systems -

Version: 8/15/2014 (see the pdf attached in my mail if you are interested at the track changes)

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

Throughout the history of science, advancements of 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. In the era of systems biology, attentions have shifted from individual molecular components to their interactions at a system level[[DW: connect to network? 5 genes are also a system]]. New functional genomics assays, in particular ones based on high-throughput sequencing (\*Seq) [1], enables biologists to probe thousands of ‘omes [2] – the comprehensive collections of constituents[[DW: can we show how large bio networks can we study, esp. from recent NGS data? 20k genes, 30B bases, or >1M SNPs….]]. One may wonder which discipline will contribute the most to biology in this new scientific paradigm [3]. While the influx of ideas in the age of reductionism was most originated from specific areas in physics or chemistry, to understand biology via a systems perspective, the new wave of catalysts actually come from areas of science that are far apart, as different as engineering, behavioral science, sociology, but centered on the concept of network [4].[[DW: why engineering, bs,soc need networks? any examples like 10k parts to engineer a rocket, 100 friends per person in soc network,.... maybe a table of disciplines (Engineering, Soc, Biology) vs. network scale (100 nodes, 1k, 10k, >1M nodes) listing examples in each category in table]]

Network is by no mean new to biologists [5]. Metabolic pathways have been studied for decades. But more recently, as a result of the advancements of high-throughput techniques, simple pathways are expanded to intertwined wiring diagrams. While many of us have been astonished by the complexity of such networks, few are able to gain any intuition from the hairballs [6]. While the term “biological network” is used rather loosely in literature for all networks originated from any subfields of biology, say food web, here in this essay, we focus our attention to molecular networks coming from genomics or systems biology because it is in general harder to gain intuitions in such networks. We want to argue that, intuitions as well as mathematical formalisms [[DW: 1st time, maybe one or two sentences to introduce “formalisms”]] developed in commonplace networks from other disciplines are able to catalyze our understanding of biology, and therefore it is instructive to initiate comparison between biological networks with networks in other disciplines.

**Comparison depends on the nature of networks [[RK: by nature you mean direct-interactions vs. associations(?) I would break up this section using subheadings to make this clearer, similar to the comment by AS.]]**

Though underlying networks of various systems may resemble one another, comparing a bio-molecular network with a complex network from a distant field, say a social network, sounds like comparing apples to oranges. What kinds of comparison could truly deepen our understanding? We believe the focus of comparison should depend on the types of information captured in networks. It is well regarded that the characteristics of a cellular system cannot be explained by the characteristics of individual components – the whole is greater than the sum of its parts. The essence of network is to describe the interactions between components of the parts-list (genes, proteins, small molecules etc.). For instance, many networks are defined based on various kinds of mechanistic interactions and specific goals of performance [[Do you mean specific cellular function here? I find “specific goals of performance” confusing - ANS]]. These networks essentially capture different facets of the complex organization of an organism, for instance, a regulatory network describes part of the cellular information processing, a metabolic network traces the chemistry of metabolites, and the protein-protein interaction network captures cell signaling as well as providing a manual on how to assemble molecular machines. Such networks closely resemble, and should be compared with networks that perform specific functions [[What does specific functions here? Do you mean engineering goals? -ANS]] like networks from engineering or technological systems [[An example for this kind of engineering network would be good here that you can compare to the regulatory network or so. A few citations here (and there are many) would help too - ANS]]. In this context, biologists could gain intuitions by examining the underlying skeletons of cross-disciplinary complex systems in the same ground as the interactions between molecular components in cells. Nevertheless, in many cases, an edge represents a certain level of coarse-graining. For example, a simple protein-protein interactions network usually does not capture the structural or temporal properties of binding. While more detailed mechanistic interactions could indeed be defined in this case [7][8], the framework may no longer be useful if too many details are incorporated. [[Is this true -ANS]][[RK: I think this is the ‘curve of utility’ of networks Mark was talking about last week. For a system too simple or too complex networks don’t really add anything]] The scenario is analogous to classical mechanics; writing down the equations of all the particles is in principle possible but not really helpful [[- this is when you move from classical mechanics to statistical mechanics which is still very very useful -ANS]]. [[I think you need to break up this paragraph around here because the emphasis of change of networks is not coming out -ANS]][[RK: I like the stat mech analogy, may be worth expanding upon. Stat mech is a different tool - you can no longer predict individual events but can model properties of the whole system very reliably]] As a result, many networks are defined in a phenomenological sense. For instance, in a genetic interaction networks [9], two genes are connected based on the phenotypes of double knockout experiments; or in a disease networks [10], a gene and a disease are connected via the statistical association between analysis of genomic variants and the occurrence of the disease. It is useful for biologists to notice that such networks, which represent mathematical abstraction of complex relationships, share common graphical structures and arise in many practical problems. For instance, mathematical machinery used in the bipartite network between genes and diseases can resonate with movie recommendation scheme building on a similar bipartite network between users and movies [[DW: citations, netflix competition?]]. Toward this end, by comparing similar network formalisms, biologists will benefit from an algorithmic or method development standpoint.

[[I think a line somewhere emphasizing that “the beauty of network theory comes from the fact that algorithms or techniques developed under one network formalism have been shown to be useful to study and analysze networks emerging from disparate disciplines of science and technology” would be good because this is the whole point of comparing networks -ANS]]

[[RK: how about trying to distinguish networks of features (genes, TFs, etc) from networks of samples (i.e. people)?]]

[[KKY: I like the saying that too many details lost the general applicability of network formalism]]

[[DW: need a connection btw two sections like: mech. networks care about depth, pathways, global characteristics, but abstract networks care locality, guilty-by-assoc.]]

[[DW: no connections to abstract networks below. maybe move abstract networks to the last section, s.t. we first talk mech network construc, topology, design/formalism, and then talk abstract network consturc, topo, design/formalism.]]

**Comparison reveals common mechanisms and principles**

Since the burgeoning of studying networks in various disciplines, efforts have been made on explaining some of the striking similarity in terms of organization of underlying networks in biological and other complex systems. An early example is the emergence of the scale-free degree distribution in a protein-protein interactions network. The pattern of organization could be explained by the duplication divergence model [11], which is essentially the same as the preferential attachment model proposed originally to explain the same pattern in many other networks [12][[RK: What is your target audience? >90% of ‘real’ biologists will have no idea what any of this last sentence means!]]. More recently, it has been shown that components in both bacterial genomes as well as large-scale computer software projects form multilayered dependency networks (enzyme A is used to decompose the output metabolites of enzyme B; the installation of package A depends on the installation of package B) leading to the same power-law components-usage frequency distribution [13]. In general, the existence of such universal mechanisms is rather rare. Nevertheless, comparisons with commonplace networks do provide intuitions for biologists. For instance, many biological networks possess an intrinsic direction of information flow, such as signaling networks where information propagates from G-Protein coupled receptors to transcription factors [14], forming a hierarchical network organization. The hierarchical organization in biological networks resemble certain the chain of command in human society, like in military context and corporate hierarchy [15]. For instance, more influential transcription factors (regulators whose expression are more highly correlated with the expression of target genes) tend to be better connected and higher in the hierarchy [16]. Moreover, the cooperative regulatory factors in a transcriptional regulatory network tend to be in the middle layer [17]. This situation is well studied in management science, where in certain corporate settings middle managers interact the most with peers to manage subordinates below them [18]. Such observations reflect a democratic hierarchy as opposite to a conventional autocratic organization [19].

[[This paragraph seems to be written as a bunch of examples written together - sometimes seems more like a laundry list. - ANS]]

[[RK: I don’t mind the laundry list too much, but you might as well be explicit and break it up with bullets or sub-subsections. Having said that the only two mechanisms/principles in this para are ‘scale-free’ ‘and ‘hierarchy’. I realise you discuss the latter in more detail in the 2 paras below, but are there any more general principles? I would have thought that redundancy/modularity are equally ‘principally’ to hierarchy.]]

[[DW: display for comparison of network topo. types vs. different disciplines]]

Of particular interest for hierarchical organization is the so-called bow-tie structure, meaning the intermediate layers have fewer components than the input and output layers. For example, in developmental genetic regulatory network, information propagates from genes controlling the initial stage of development (the input) to genes controlling detailed cell differentiation and morphogenesis (output) [20][21]. The intermediate layer refers to a small set of input-output genes integrating complex spatiotemporal information and trigger development of an entire program of cell differentiation [22]. In the networking architecture of the Internet, on the other hand, various protocols in the input/link layer (ARP, RARP, NDP etc) and various application protocols in the application/output layer (HTTP, FTP,DHCP etc) are essentially connected by IPv4, the primary protocols in the internet layer.[[DW: can we say more explicitly about hourglass? (Video, pic, text….) => TCP/IP => (DSL, wireless, …) or like Fedex, send (a lot of things) => box => delivery by (truck, air, ship…)]] A recent paper provided a first mechanism to understand its evolution by explicitly modeling information flow in feed-forward networks as a cascade of matrix multiplications (similar to neural networks in machine learning context) [23]. It showed that a bow-tie structure emerged if the goal matrix is rank deficient, i.e. the information can be compressed.

Lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. Toward this direction, comparison with various technological networks is particularly insightful. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a negative feedback loop and a source of time delay. Nevertheless, oscillators of various purposes (e.g. for circadian rhythms or for cell cycle) or from various organisms are not identical but have a certain level of variation because additional design objectives or strategies are involved. Just like not all electronic devices use the same oscillator design, the importance of design objectives is not new at all in engineering systems. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [24]. Robustness is obviously a preferred design objective because it makes a system tolerate intrinsic or extrinsic stochastic fluctuations. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by interface is widely practiced [25]. The same is for biological networks because modules can be readily reused to adapt new functions. Because of the fundamental importance of such design objectives, an insightful network comparison should be rooted in the common design objectives rather than merely network topology.

[[In this section, technical language in parts that are not explained properly. Need to work more to keep attention of reader - ANS]]

[[RK: Can you work in some discussion of this paper that just appeared in Science? http://www.sciencemag.org/content/345/6198/829.full]]

[[DW: shall we need to mention U alon’s 3rd point, recurring circuits, which were mentioned below?]]

[[DW: shall we need to talk modularity/robustness for abstract networks? prediction?]]

**Comparison highlights the commonality and difference between tinkerer and engineer**

[[DW: use non-U-alon words in heading, and maybe add ‘Box’ display to introduce concepts of tinkerer and engineer, so on…

<http://www.plosbiology.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pbio.1001877&representation=PDF> ]]

The comparison of biological networks and technological networks should best be performed under the light of evolution. As Alon highlighted by the phase “the tinkerer as an engineer” [24], it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer starting with bits and pieces and trying to connect random nodes, whereas technological networks are essentially blueprints drawn by engineers. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that construct living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological networks like roadways and circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [26]. To a certain extent, an engineer is a tinkerer. [[RK: junk DNA?]]

The parallel between tinkerer and engineer points to a common framework to unite them. Wagner further proposed an analogy between the genotype space for a biological system and the design space for a technological system. These spaces contain all the possible networks in the corresponding systems. In biology, many attempts have been made to search for solutions of common functional problems such as adaptation, oscillation and cell polarization [27]. Similar studies were performed in the context of circuit design, where a set of logic gates was evolved via rewiring in order to perform a predefined computational task [28][29]. [[DW: recurring circuits?]]These studies suggested that in both kinds of systems, the solution networks are close together in the genotype/design space. As each solution in genotype/design has multiple neighbors, robustness of a solution to mutation facilitate the evolvability of these systems [30][31]. Indeed, it has been demonstrated that electronic circuits can be evolved to fulfill a fluctuating evolutionary goal [28]. Similarly, metabolic networks of bacteria living in multiple habitats are evolved to decompose multiple food sources [32][33]. [[Doesnt fit there according to me- ANS.] ]Both of these networks show a level of modular organization. While both biological and technological networks are shaped by similar underlying design objectives that impose further constraints to the solutions, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there’s no free lunch [34][35].[[DW: move to meta-table listing all possible refs?]]

Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many different processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fraction of functions are updated often [36]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, software system is very similar to a roadway system; bottlenecks under high usage like George Washington Bridge require more upgrade and more construction.[[Make it clear that random changes occur here too but the observed evolutionary rate pertaining to healthy samples are slower due to the fact that fewer mutations are allowed randomly at the hubs - unlike an artificial system where the tinkering is done with the current system in mind and the current objective in mind, evolution behaves randomly and maps the genotype to the fitness according to the environmental needs.- ANS]] While intentional tweaking on bottlenecks sounds obvious for technological systems, it is not always possible. In the above example of internet architecture, while there are frequent innovations at the input layer that interact with a variety of networking hardware and output layers that connect with many different software applications, the internet layer with very few protocols is the bottleneck under heavy constraints and such protocols can hardly be replaced [37]. The observed rapid innovation at the top and bottom layers but constraint at the middle may shed light on a remarkably pattern in developmental genetic regulatory network. Different species exhibit different patterns at the early and late stages of embryo development, but highly similar during the phylotypic stage – the so-called hourglass phenomenon [38]. [[DW: figure display, two hourglass, one is internet protocol, another is embryo development]]

[[DW: again, need a transition/connection from design section to formalism section]]

**Comparison leverages mathematical machineries [[Mention to genotype to phenotype network as a goal of genomics here]]**

Apart from networks mediated by mechanistic interactions, many networks in the literature are essentially two-dimensional projection of high-dimensional data. As big data across disciplines are often signified by the combinatorial explosion of high dimensional features, it is not surprising that network algorithms developed in one discipline can readily be applied in biology, for instance, the idea of “guilt by association” is widely used in genomics for inferring functions of a protein or a non-coding element based on the function of its neighbors in a network, the same idea is also widely used in social media like Facebook to suggest friends. Perhaps the best example is probably the PageRank algorithm. Idea originated from Katz centrality in social network analysis [39], PageRank algorithm was first used by Google[[RK: named for Larry Page, no?]] to rank documents based on linkages in a self-consistent way. The algorithm was then adopted in food webs to determine extinction [40] and later in an algorithm called NetRank that rank prognostic relevance for patients with cancers [41]. More interestingly, the idea of PageRank was able to solve the global network alignment problem, which was applied in biological context in order to detect functional orthology across species [42].[[DW: two types of formalisms:1) relationship like correlation 2) causation. maybe we need more causation examples like dynamic analysis using bayesian, differential equations in reverse engineering...]]

Networks across disciplines, despite of different origins, actually present very similar challenges. For instance, being noisy and incomplete makes procedures like link prediction and denoising necessary. Difficulties lie at the proper learning of network organization. Generative models of networks, say stochastic block models [43], are very popular in computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standard for validation. Another trend of network analysis is the notion of multiplex networks where multiple layers of networks form an interconnected structure. The idea is originated in social network analysis because an individual may participate in multiple social circles: family, friends, colleagues, or in online setting: Facebook, Linkedln and Twitter. The same is true in biological context because of the existence of multiple relational connections (co-expression, genetic interactions etc.) between components in networks. A similar multiplex generalization in network analysis is the so-called temporal networks, which consider the slices of networks taking place at different time points together as a single mathematical structure [44]. Again, the current application focuses on online social networks because genome-wide data in biological systems are still not dynamics enough. However, as the number of time points increases, say in RNA-Seq experiments, algorithms developed in social contexts can be easily applied to integrate the slices of co-expression networks. [[DW: combine dimensionality reduction + diff. equations for abstract networks: besides correlative relationships via “guilt by association”, ones have also developed network methods based on “causation” relationships to study biological network dynamics. For example, the system matrix in systematic differential equations has been used to capture the contributions of gene expression in a network at one time point to expression at next time point (causation); i.e., if we use system matrix as adjacency matrix to represent network, then this network is a “causative” network. By analytic solutions of differential equations, eigenvalues of system matrix describe system’s dynamic characteristics. Limited biological data, however, prevent us from estimating a variety of parameters of system matrix. In order to learn network dynamics, using data mining techniques such as SVD, ones have reduced high-dimensional gene expression data space to highly co-varying low-dimensional gene expression data space, so that ones can estimate effective system matrix along with dynamic characteristics in lower space using limited data. This idea of combining differential equation models and dimensionality reduction techniques has been successfully applied to find novel cell cycle genes in yeast [DOI: 10.1371/journal.pone.0028805].

[[DW: very large-scale models (vs. VLSI): markov logic network on customer networks, Innovation Award Talk by [Pedro Domingos](http://sigkdd.org/pedro-domingos-1), KDD 2014 ]]

[[DW: both paragraphs are multiplex]]

Nevertheless, biology motivates an alternate definition of temporal network. Networks from different species essentially capture a sense of temporal changes [[I would use a different term than temporal networks here because temporally both organisms can exist together - I would say something like “evolutionary changes to a common core” - ANS]]. In this definition, pairs of orthologous genes can be used to connect networks from different species, forming a multiplex structure. The notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [45]. Another mathematical formalism was developed to measure the evolutionary rewiring rate between networks across species in analogous to quantifying sequence evolution [46]. It was shown that metabolic networks rewire at a slower rate compared to various regulatory networks. The formalism can be applied to networks in social or technological contexts in general.

control theory…

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of ‘omes across species. To nourish a system-level understanding and to leverage the tremendous amount of high-throughput data, may be it is a time to extend our tradition even further to compare with networks from other complex systems as well as other disciplines. Comparison of biological networks with technological networks, and the similarity between tinkerers and engineers point toward biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [27]. Comparison of mathematical formalisms used in mining biological networks and various social networks point to further integration of two classes of data. Such integration is getting more and more important as datasets that combine genotypes, phenotypes and information like connections between individuals will become popular, for example, the Framingham study. Indeed, various scientific disciplines form a network in the intellectual universe where knowledge emerges when things connect.

[[I think it might help to give both networks names like mechanistic networks and relational networks so you can easily refer to them in text - ANS]]

[[RK: What is the goal of this essay? Is it to try to convince biologists of the utility of network analysis? If so it is probably not sufficiently focussed and is too technical. Is it to enumerate all of the applications of networks to biology? If so there could/should be some critical intuition/appraisal of how this has been done well or done poorly.]]

[[DW: I agree with RK. maybe provide existing bioinformatic tools along with ones from other areas in each ‘step’ of network intuition/construction, topology, design, and formalism. sort of give biologists a workflow about how to use networks to study their interests]]

<http://www.pnas.org/content/100/10/5944.full>

[1] L. Pachter, “\*Seq: functional genomics assays based on high-througphput sequencing,” *Bits of DNA: Reviews and commentary on computational biology*, 2014. .

[2] M. Baker, “Big biology: The ’omes puzzle,” *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013.

[3] Tony Hey, Stewart Tansley, and Kristin Tolle, *4th Paradigm*. Microsoft Research, 2009.

[4] 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.

[5] 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.

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

[7] J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, “Evidence for dynamically organized modularity in the yeast protein–protein interaction network,” *Nature*, vol. 430, no. 6995, pp. 88–93, Jul. 2004.

[8] 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.

[9] P. Beltrao, G. Cagney, and N. J. Krogan, “Quantitative Genetic Interactions Reveal Biological Modularity,” *Cell*, vol. 141, no. 5, pp. 739–745, May 2010.

[10] 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.

[11] A. V&aacute;zquez, A. Flammini, A. Maritan, and A. Vespignani, “Modeling of Protein Interaction Networks,” *Complexus*, vol. 1, no. 1, pp. 38–44, 2003.

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

[13] 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.

[14] N. Polouliakh, R. Nock, F. Nielsen, and H. Kitano, “G-Protein Coupled Receptor Signaling Architecture of Mammalian Immune Cells,” *PLoS ONE*, vol. 4, no. 1, p. e4189, Jan. 2009.

[15] 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.

[16] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, “Architecture of the human regulatory network derived from ENCODE data,” *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.

[17] 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.

[18] 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.

[19] Y. Bar-Yam, D. Harmon, and B. de Bivort, “Attractors and Democratic Dynamics,” *Science*, vol. 323, no. 5917, pp. 1016–1017, Feb. 2009.

[20] 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.

[21] I. S. Peter and E. H. Davidson, “Evolution of Gene Regulatory Networks Controlling Body Plan Development,” *Cell*, vol. 144, no. 6, pp. 970–985, Mar. 2011.

[22] D. L. Stern and V. Orgogozo, “Is Genetic Evolution Predictable?,” *Science*, vol. 323, no. 5915, pp. 746–751, Feb. 2009.

[23] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, “Evolution of bow-tie architectures in biology,” *ArXiv14047715 Q-Bio*, Apr. 2014.

[24] U. Alon, “Biological Networks: The Tinkerer as an Engineer,” *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] K. Raman and A. Wagner, “The evolvability of programmable hardware,” *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269 –281, Feb. 2011.

[30] A. Wagner, “Neutralism and selectionism: a network-based reconciliation,” *Nat. Rev. Genet.*, vol. 9, no. 12, pp. 965–974, Dec. 2008.

[31] J. Masel and M. V. Trotter, “Robustness and Evolvability,” *Trends Genet.*, vol. 26, no. 9, pp. 406–414, Sep. 2010.

[32] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, “The evolution of modularity in bacterial metabolic networks,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 19, pp. 6976–6981, May 2008.

[33] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, “Toolbox model of evolution of prokaryotic metabolic networks and their regulation,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 24, pp. 9743–9748, Jun. 2009.

[34] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.

[35] 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.

[36] 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.

[37] S. Akhshabi and C. Dovrolis, “The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, New York, NY, USA, 2011, pp. 206–217.

[38] B. Prud’homme and N. Gompel, “Evolutionary biology: Genomic hourglass,” *Nature*, vol. 468, no. 7325, pp. 768–769, Dec. 2010.

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

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.

[45] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, “OrthoClust: An orthology-based network framework for expression clustering across multiple species,” *Genome Biol.*, vol. 15, p. R100.

[46] 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.

Version: 8/8/2014.

Comparative Netomics - lessons from cross-disciplinary network comparison

Throughout the history of science, advancements of 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. In the era of systems biology, attentions have shifted from individual molecular components to their interactions at a system level. New functional genomics assays, in particular ones based on high-throughput sequencing (\*Seq) [1], enables biologists to probe thousands of ‘omes [2] – the comprehensive collections of constituents. One may wonder which discipline will contribute the most to biology in this new scientific paradigm [3]. While the influx of ideas in the age of reductionism was most originated from specific areas in physics or chemistry, to understand biology via a systems perspective, the new wave of catalysts actually come from areas of science that are very far apart, as different as engineering, behavioral science, sociology, but centered on the concept of network [4].

Networks are by no mean new to biologists [5]. Metabolic pathways have been studied for decades. But more recently, as a result of the advancements of high-throughput techniques, simple pathways are expanded to intertwined wiring diagrams published in high profile journals. While many of us have been astonished by the complexity of such networks, few are able to gain any intuition from the hairballs [6]. While the term “biological network” is used rather loosely in literature for all networks originated from any subfields of biology, say food web, here in this essay, we focus our attention to molecular networks coming from genomics or systems biology because it is in general harder to gain intuitions in such networks. We want to argue that, intuitions as well as formalism developed in commonplace networks from other disciplines are able to catalyze our understanding of biology. Indeed, not only capturing the unique flavor of systems biology, the concept of network essentially describes the interactions between individual constituents-who is interacting with whom-in any complex system. The simple description thus enables one to examine the underlying skeletons of cross-disciplinary complex systems in the same ground as the interactions between molecular components in cells. While biologists want to gain insights on the complex interactions between molecular components, the common network language makes ideas and methods developed to understand the organization of complex systems in diverse fields more accessible to biologists, and vice versa. To leverage the interdisciplinary connections, it is instructive to initiate comparison between biological networks with networks in other disciplines.

**Two classes of network comparison**

Though underlying networks of various systems may resemble one another, comparing a bio-molecular network with a complex network from a distant field, say a social network, sounds like comparing apples to oranges. What kinds of comparison could truly deepen our understanding? We believe the focus of comparison should depend on the nature of networks. Transcriptional regulatory networks, metabolic networks or protein-protein interaction networks are examples of widely studied molecular networks. They essentially capture different facets of the complex organization of an organism, for instance, a regulatory network describes part of the cellular information processing, a metabolic network traces the chemistry of metabolites, and the protein-protein interaction network captures cell signaling as well as providing a manual on how to assemble molecular machines. Despite the intrinsic difference, they all refer to specific goals of performance mediated by actual mechanistic interactions involving physical interactions between the nodes. Such networks closely resemble, and should be compared with networks that perform specific functions like networks from engineering or technological systems. There is another class of biological networks with examples like genetic interaction networks [7] or disease networks [8]. The former example is a sophisticated metrics summarizing results of double knockout experiments whereas the later describes the statistical association between genes and diseases as a result of GWAS and analysis of various genomics variants. The edges in such networks may not represent a direct mechanistic interaction between the components but could represent a mathematical abstraction of complex relationships. From an algorithmic or method development standpoint, they share common graphical structures arise in computational social science, for instance mathematical machinery used in the bipartite network between genes and diseases can resonate with movie recommendation scheme building on a similar bipartite network between users and movies.

MG: division is not expliciit enough

relational, looser...mechanisitic ones are actually more specific

RK: one net is better than others. mor

MG: power of network, is actually the 2nd class

RDF triple

on one hand, network is not enough to capture real bio, like complexes

a spectrum of edges, what kind of edges are the best?

little info., richer, break down at certain point.. (too much info. to be captured

**Comparison of design principles**

Lying at the heart of deciphering biological networks mediated by mechanistic interactions is the mapping between architecture and function. As an example, consider a biochemical oscillator. Two essential elements of an oscillator are a negative feedback loop and a source of time delay. Nevertheless, oscillators of various purposes (e.g. for circadian rhythms or for cell cycle) or from various organisms are not identical but have a certain level of variation because additional design objectives or strategies are involved. Just like not all electronic devices use the same oscillator design, the importance of design objectives is not new at all in engineering systems. The striking similarity between biological systems and technological systems has long been identified. A decade ago, Uri Alon pointed out several common design principles in biological and engineering networks such as modular organization and robustness to perturbation [9]. Robustness is obviously a preferred design objective because it makes a system tolerate intrinsic or extrinsic stochastic fluctuations. Modularity, on the other hand, makes a system more evolvable. For instance in software design, modular programming that separates functionality of a program into independent modules connected by interface is widely practiced [10]. The same is for biological networks because modules can be readily reused to adapt new functions. Because of the fundamental importance of such design objectives, an insightful network comparison should be rooted in the common design objectives rather than merely network topology, Nevertheless, in many cases, common topological patterns are the reflection of common underlying design objectives or strategies.

While networks originated from technological systems are particularly analogous to biological networks, under certain specific design objectives, comparison could be further broadened to include networks from other disciplines. An interesting example is related to how information is transferred between input and output nodes in a network. Many biological networks possess an intrinsic direction of information flow, for instance signaling networks where information propagates from G-Protein coupled receptors to transcription factors [11], and developmental gene regulatory networks where information propagates from genes controlling the initial stage of development to genes controlling detailed cell differentiation and morphogenesis [12][13]. The former example refers to the spatial organization, whereas the later is defined in a temporal fashion. The l[[[jump]]ater is similar for the regulatory networks in general; more influential transcription factors (regulators whose expression are more highly correlated with the expression of target genes) tend to be better connected and higher in the hierarchy [14]. The hierarchical organization in biological networks resemble certain the chain of command in human society, like in military context and corporate hierarchy [15]. For instance, cooperative regulatory factors in a transcriptional regulatory network tend to be in the middle layer [16]. The situation is well studied in management science, where in certain corporate settings middle managers interact the most with peers to manage subordinates below them [17]. Such observations reflect a democratic hierarchy as opposite to a conventional autocratic organization [18]. Of particular interest for hierarchical organization is the so-called bow-tie structure, meaning the intermediate layers have fewer components than the input and output layers. A recent paper provided a first mechanism to understand its evolution by explicitly modeling information flow in feed-forward networks as a cascade of matrix multiplications (similar to neural networks in machine learning context) [19]. It showed that a bow-tie structure emerged if the goal matrix is rank deficient, i.e. the information can be compressed. Of course, there are still plenty of interesting observations without explanation. For example, in developmental genetic regulatory network, the intermediate layer refers to a small set of input-output genes integrating complex spatiotemporal information (the input) and trigger development of an entire program of cell differentiation (the output) [20]. In the networking architecture of the Internet, on the other hand, various protocols in the input/link layer (ARP, RARP, NDP etc) and various application protocols in the application/output layer (HTTP, FTP, SMTP, DHCP etc) are essentially connected by IPv4, the primary protocols in the internet layer. While there are frequent innovations at the input layer that interact with a variety of networking hardware and output layers that connect with many different software applications, the internet layer is the bottleneck with under heavy constraints in which there are very few protocols and they can hardly be replaced [21]. Remarkably, the rapid innovation at the top and bottom layers but constraint at the middle happens in developmental genetic regulatory network. Different species exhibit different patterns at the early and late stages of embryo development, but highly similar during the phylotypic stage – the so-called hourglass phenomenon [22].[[MG: protocol forms network]]

**Revisiting tinkerer versus engineer**

The parallel between biological networks and technological networks should best be examined under the light of evolution. As Alon highlighted by the phase “the tinkerer as an engineer” [9], it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer starting with bits and pieces and trying to connect random nodes, whereas technological networks are essentially blueprints drawn by engineers. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that construct living organisms on purpose. Nevertheless, the distinction is not clear-cut. Both biological networks and man-made technological networks like roadways and circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. In a recent review, Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer [23]. To a certain extent, an engineer is a tinkerer.

The parallel between tinkerer and engineer points to a common framework to unite them. Wagner further proposed an analogy between the genotype space for a biological system and the design space for a technological system. These spaces contain all the possible networks in the corresponding systems. In biology, many attempts have been made to search for solutions of common functional problems such as adaptation, oscillation and cell polarization [24]. Similar studies were performed in the context of circuit design, where a set of logic gates was evolved via rewiring in order to perform a predefined computational task [25][26]. These studies suggested that in both kinds of systems, the solution networks are close together in the genotype/design space. As each solution in genotype/design has multiple neighbors, robustness of a solution to mutation facilitate the evolvability of these systems [27][28]. Indeed, it has been demonstrated that electronic circuits can be evolved to fulfill a fluctuating evolutionary goal [25]. Similarly, metabolic networks of bacteria living in multiple habitats are evolved to decompose multiple food sources [29][30]. Both of these networks show a level of modular organization. While both biological and technological networks are shaped by similar underlying design objectives that impose further constraints to the solutions, there is no way to optimize all objectives and thus tradeoffs are unavoidable in both biological and technological systems. This is essentially the conventional wisdom – there’s no free lunch [31][32].

Despite the similarity, tinkerers and engineers take different views in balancing different constraints and tradeoffs. Their optimal choices are exhibited the topology of their corresponding networks. Taking software engineering as an example, software engineers tend to reuse certain code. However, the robustness of software will be reduced if a piece of code is highly called by many different processes. Analysis of the evolution of a canonical software system, the Linux kernel, revealed that the rate of evolution of functions (routines) is distributed in a bimodal fashion and thus a significant fashion of functions are updated often [33]. Therefore, unlike biological systems in which the majority of components are rather conserved and thus prefer a more independent organization to maintain robustness, software engineers pay the price of reusability and robustness by constantly tweaking the system. Indeed, further analysis of the underlying network of Linux kernel, the so-called call graph, showed that more central components at the call graph require more fine-tuning. In other words, unlike biological networks whose hubs tend to evolve slowly because of the number of constraints, software system is very similar to a roadway system; bottlenecks like George Washington Bridge require more upgrade and more construction.[[connectiviity vs use]]

**Comparison in terms of mathematical approaches**

Apart from networks mediated by mechanistic interactions, many networks in the literature are essentially two-dimensional projection of high-dimensional data. As big data across disciplines are often signified by the combinatorial explosion of high dimensional features, it is not surprising that network algorithms developed in one discipline can readily be applied in biology, for instance, the idea of “guilt by association” is widely used in genomics for inferring functions of a protein or a non-coding element based on the function of its neighbors in a network, the same idea is also widely used in social media like Facebook to suggest friends. Perhaps the best example is probably the PageRank algorithm. Idea originated from Katz centrality in social network analysis [34], PageRank algorithm was first used by Google to rank documents based on linkages in a self-consistent way. The algorithm was then adopted in food webs to determine extinction [35] and later in an algorithm called NetRank that rank prognostic relevance for patients with cancers [36]. More interestingly, the idea of PageRank was able to solve the global network alignment problem, which was applied in biological context in order to detect functional orthology across species [37].

[[fit in Chou’s paper somewhere] ], MG fit alignment there...or Chou’s there

network is a compact representation of matices, thsquare at’s why netork is so popular, and

matrix , discretize,

Networks across disciplines, despite of different origins, actually present very similar challenges. For instance, being noisy and incomplete makes procedures like link prediction and denoising necessary. Difficulties lie at the proper learning of network organization. Generative models of networks, say stochastic block models [38], are very popular computational social science. Nevertheless, such models are not widely used in biological context yet, presumably because of the lack of gold standard for validation. Another trend of network analysis is the notion of multiplex networks where multiple layers of networks form an interconnected structure. The idea is originated in social network analysis because an individual may participate in Facebook, Linkedln and Twitter. [[a postmodern world]] The same is true in biological context because of the existence of multiple relational connections (co-expression, genetic interactions etc.) between components in networks. The multi-layers notion has recently been used to integrate co-association across different species in order to detect conserved and specific functional modules [39]. A similar generalization in network analysis is the so-called temporal networks, which consider the slices of networks taking place at different time points together as a single mathematical structure [40]. Again, the current application focuses on online social networks because genome-wide data in biological systems are still not dynamics enough. However, as the number of time points increases, say in RNA-Seq experiments, algorithms developed in social contexts can be easily applied to mine the resultant co-expression networks.

**Conclusion**

Biology is a subject with a strong tradition of doing comparison. One hundred years ago, biologist compared the phenotypes of different species. Since the discovery of DNA, biologists have been comparing the sequences of different genes, and then all sorts of ‘omes between species. To nourish a system-level understanding and to leverage the tremendous amount of high-throughput data, may be it is a time to extend our tradition even further to compare with networks from other complex systems as well as other disciplines. Comparison of biological networks with technological networks, and the similarity between tinkerers and engineers point toward biological circuits that solve common functional problems – effectively a toolbox for synthetic biology [24]. Comparison of methods in mining biological networks and networks in computational social science is bridging the two distant fields, whereas efforts have been spent on the interface [41]. Indeed, various scientific disciplines form a network in the intellectual universe in which knowledge emerge when things connect.

~2700 words

[1] L. Pachter, “\*Seq: functional genomics assays based on high-througphput sequencing,” *Bits of DNA: Reviews and commentary on computational biology*, 2014. .

[2] M. Baker, “Big biology: The ’omes puzzle,” *Nature*, vol. 494, no. 7438, pp. 416–419, Feb. 2013.

[3] Tony Hey, Stewart Tansley, and Kristin Tolle, *4th Paradigm*. Microsoft Research, 2009.

[4] 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.

[5] 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.

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

[7] P. Beltrao, G. Cagney, and N. J. Krogan, “Quantitative Genetic Interactions Reveal Biological Modularity,” *Cell*, vol. 141, no. 5, pp. 739–745, May 2010.

[8] 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.

[9] U. Alon, “Biological Networks: The Tinkerer as an Engineer,” *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.

[10] 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.

[11] N. Polouliakh, R. Nock, F. Nielsen, and H. Kitano, “G-Protein Coupled Receptor Signaling Architecture of Mammalian Immune Cells,” *PLoS ONE*, vol. 4, no. 1, p. e4189, Jan. 2009.

[12] 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.

[13] I. S. Peter and E. H. Davidson, “Evolution of Gene Regulatory Networks Controlling Body Plan Development,” *Cell*, vol. 144, no. 6, pp. 970–985, Mar. 2011.

[14] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, “Architecture of the human regulatory network derived from ENCODE data,” *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.

[15] 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.

[16] 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.

[17] 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.

[18] Y. Bar-Yam, D. Harmon, and B. de Bivort, “Attractors and Democratic Dynamics,” *Science*, vol. 323, no. 5917, pp. 1016–1017, Feb. 2009.

[19] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, “Evolution of bow-tie architectures in biology,” *ArXiv14047715 Q-Bio*, Apr. 2014.

[20] D. L. Stern and V. Orgogozo, “Is Genetic Evolution Predictable?,” *Science*, vol. 323, no. 5915, pp. 746–751, Feb. 2009.

[21] S. Akhshabi and C. Dovrolis, “The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, New York, NY, USA, 2011, pp. 206–217.

[22] B. Prud’homme and N. Gompel, “Evolutionary biology: Genomic hourglass,” *Nature*, vol. 468, no. 7325, pp. 768–769, Dec. 2010.

[23] 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.

[24] 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.

[25] 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.

[26] K. Raman and A. Wagner, “The evolvability of programmable hardware,” *J. R. Soc. Interface*, vol. 8, no. 55, pp. 269 –281, Feb. 2011.

[27] A. Wagner, “Neutralism and selectionism: a network-based reconciliation,” *Nat. Rev. Genet.*, vol. 9, no. 12, pp. 965–974, Dec. 2008.

[28] J. Masel and M. V. Trotter, “Robustness and Evolvability,” *Trends Genet.*, vol. 26, no. 9, pp. 406–414, Sep. 2010.

[29] A. Kreimer, E. Borenstein, U. Gophna, and E. Ruppin, “The evolution of modularity in bacterial metabolic networks,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 19, pp. 6976–6981, May 2008.

[30] S. Maslov, S. Krishna, T. Y. Pang, and K. Sneppen, “Toolbox model of evolution of prokaryotic metabolic networks and their regulation,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 24, pp. 9743–9748, Jun. 2009.

[31] A. D. Lander, “Pattern, growth, and control,” *Cell*, vol. 144, no. 6, pp. 955–969, Mar. 2011.

[32] 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.

[33] 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.

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

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] K.-K. Yan, D. Wang, J. Rozowsky, H. Zheng, C. Cheng, and M. Gerstein, “OrthoClust: An orthology-based network framework for expression clustering across multiple species,” *Genome Biol.*, vol. 15, p. R100.

[40] P. Holme and J. Saramäki, “Temporal networks,” *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, Oct. 2012.

[41] J. H. Fowler, J. E. Settle, and N. A. Christakis, “Correlated genotypes in friendship networks,” *Proc. Natl. Acad. Sci.*, p. 201011687, Jan. 2011.

7/29/2014

Comparative Netomics - lessons from cross-disciplinary network comparison (tentative title)

Throughout the history of science, we have seen many examples in which the advancements of biology have been catalyzed by discoveries in other disciplines. For instance, the maturation of X-ray diffraction facilitated the discovery of double helix, and later on the characterization of structures of thousands of different proteins. In the era of systems biology, attentions have been shifted from individual molecular components to their interactions in a system level. One may wonder from which discipline will biology be benefited most. In this essay, we want to argue that, while the influx of ideas in the past was most originated from specific areas in physics or chemistry, the new wave of catalysts come from areas of science vary far apart, as different as engineering, behavioral science, sociology, but centered on the concept of network.

[[DW: I think we may emphasize that biology is a complex system. if we want to understand such complex system, we must work from a network point of view. however, it was impossible to study biology at large scale due to limitations of technologies/computing resources. Fortunately,for example, esp. in recent five years, more and more advanced technologies get tons of large-scale bio datasets (Big Data) such as NGS, plus more powerful computers + efficient algorithms make systematically understanding biology feasible…this is historic time!!!]] Systems perspective important - MG

[[ANS - what I got from meeting. In the 19th and 20th centuries, reductionist approach worked. Now a systems perspective is becoming more and more important as we study more and more data from complex systems. Network theory is becoming the method on how we study and compare these complex systems.]]

Network is by no mean new to biologists. Over the last decade, plenty of intertwined wiring diagrams were published in high profile journals [[RK: analysis of molecular pathways has been going on a lot longer - these are not hairballs - people devote entire careers to small collections of molecules. To add to DW’s point above, why not say current \*omics gives us an ability to expand from pathways to whole networks, but that we need to borrow from other fields some mechanisms for understanding/reducing/simplifying these vast hairballs - Multimodal data important]]. While many of us have been astonished by the complexity of such networks, few are able to gain any intuition from the hairballs. In this essay, we want to argue that, intuitions as well as mathematical methods developed in commonplace networks from other disciplines are able to catalyze our understanding of biology. Indeed, not only capturing the unique flavor of systems biology, the concept of network essentially describes the interactions between individual constituents-who is interacting with whom-in any complex system. The simple description thus enables one to examine the underlying skeletons of cross-disciplinary complex systems in the same ground as the interactions between molecular components in cells [[ANS - I think this is not necessarily a given - that the rules governing lots of different networks lead to similar organization in completely different disciplines - Something like once the relationships between the parts of a complex system is put in the form of a network, the power of network theory comes from the fact that the same algorithms can be used to find the commonalities in the topology of disparate natural and man-made systems - example scale free, etc]]. The common network language makes ideas and methods developed to understand the organization of complex systems in diverse fields more accessible to biologists who want to gain insights on the complex interactions between molecular components, and vice versa. To leverage the interdisciplinary connections, it is instructive to initiate comparison between biological networks with networks in other disciplines.

[[DW: i think we need to break down “complex” into two or three concrete scenarios. first, we can compare network organizations, and find some similarities of network org./struc. . In some cases, network characteristics such as modules help us like in evolution/robustness[[RK: + scale-free]]. however, beyond looking at network struc./characteristics similarities, we need further predict/identify functions or mechanisms underlying these characteristics like co-expression modules. that’s why we want to borrow successful “math modeling” from other fields;

we can discuss “network characteristics” + “math modelling” at different levels, node, link, pathway, whole network as CC suggested at ISMB ]]

[[RK: Is it worth getting into how wet-lab biology / biochemistry is progressing to finer and finer resolution that allow us to create more accurate / elaborate networks? E.g:  
- CLIP/RIP-seq for miRNA/RNA-protein interaction gives better post-transcriptional regulatory maps

- ribosome profiling for rates of translation

- cell-type specific expression/chromatin profiling for cancer/neuro/general to decompose signals from multiple sources

This is a complement, and not worthy of much discussion here- but math is not the only thing that provides deeper intuition into these kinds of networks!]] [[I agree with RK that math is not enough - ANS]][[DW: I agree too that math is not enough but this essay focuses on network comparison with other fields. we may not put too much stuff on bio experimental details. of course, we can mention those adv. experiments a bit.]]

[[first few sent. go up, make classification fuzzy]]

Though underlying networks of various systems may resemble one another, comparing a bio-molecular network with a complex network from a distant field, say a social network, sounds comparing apples to oranges. What kinds of comparison could truly deepen our understanding? We believe there are a few points one should pay attention. First of all, the term “biological network” is used rather loosely in literature. It refers to all networks originated from any subfields of biology, including ecological networks or food webs. Here, however, we focus our attention to molecular networks coming from genomics or systems biology because it is in general harder to gain intuitions in such networks. Second, networks of particular interests to genomics or systems biology can be further be divided into two classes. The first class consists of examples like transcriptional regulatory networks, metabolic networks or protein-protein interaction networks. These examples are intrinsically different, for instance, a regulatory network describes part of the cellular information processing, a metabolic network traces the chemistry of metabolites, and the protein-protein interaction network is essentially a manual on how to assemble molecular machines [[ANS - not necessarily just molecular machines - signaling networks use a lot of PPI without making one giant machine[[RK: Yes! And are extremely important for therapeutics / small molecule delivery / cell-surface proteins]]]]. While these networks capture different facets of the complex organization of an organism, they all refer to specific goals of performance mediated by actual mechanistic interactions. Such networks closely resemble, and should be compared with networks with similar goals [[ANS - I dont understand similar goals here since we are talking about circuits, social networks, etc]] like networks from engineering or technological systems. T [[stress on design.. more stronlhy]]

h[[e is another class of biological networks with examples like genetic interaction networks or disease networks. The former example is a sophisticated metrics summarizing results of double knockout experiments whereas the later describes the statistical association between genes and diseases as a result of GWAS and analysis of various genomics variants. Such networks do not show mechanistic interactions but a mathematical abstraction of complex relationships. From an algorithmic or method development standpoint, they share common graphical structures arise in computational social science, for instance mathematical machinery used in the bipartite network between genes and diseases can resonate with movie recommendation scheme building on a similar bipartite network between users and movies.

[[CC: a slightly different classification: Class 1: A do sth with/on B (undirected/directed); Class 2: A and B do/are sth, some nets are very hard to be classified…

the classes are not really for networks, but more on comparison… a list of metrics makes life easier]]

[[DW: moreover, two types of bio networks are not independent. One type typically drives another such as reg. network vs. co-expression network. thus, math modeling for one network can also help understand another network. maybe use concept of “multiplex” or “mutli-mode” network…]]

[[RK: Be careful of the chicken and egg when talking about ‘driving’. The regulatory and expression networks are only separate due to the fact we need to use different experiments and it is more convenient- they are really two/three levels of the same network]]

The striking similarity between biological systems and technological systems [CC: do we want to focus on tech system. How about other network system such as social net?] has long been identified. For instance, transcription regulation uses cooperative binding to arrive at a sigmoidal response curve whereas the same is achieved by an amplifier in circuit design. Analogies of this kind have inspired the advancement of synthetic biology. In terms of global network organization, a decade ago, Uri Alon pointed out several common design principles such as modularity and the usage of recurring elements (motifs) \cite{Alon Sci 2003}. These common topological patterns are in many cases the reflection of common underlying design objectives or strategies [[I think modularity of software can be brought up here -ANS [[RK: and WWW -- PMID:10521342]]]]. For examples, modularity makes a system more evolvable, which is an advantage one would like to optimize. Therefore, an insightful network comparison should be rooted in the common design objectives. As Alon highlighted by the phase “the tinkerer as an engineer”, it is remarkable that “good-engineering solutions” are found in biological systems evolved by random tinkering. Indeed, comparison between biological and technological networks should manifest the nature of the two very different approaches: evolution as a tinkerer trying to connect random nodes until the network is good enough to work [[Evolution keeps tinkering and adjusting to the environment - ANS. For it to evolve, there are certain criteria that ensure that the system is not in a deep evolutionary minima - i.e., small changes to the sequence and/or network are not totally harmful]], whereas technological networks are essentially blueprints drawn by engineers. Biologists often tend to distinguish the two approaches cautiously so as to avoid the notion of intelligent design – the existence of an intelligent cause that construct living organisms on purpose. Nevertheless, the distinction is not clear-cut. To a certain extent, an engineer is a tinkerer. Both biological networks and man-made technological networks like roadways and circuits are complex adaptive systems, there are plenty of examples showing that many great innovations are results of trial and error, and all technological systems are subjected to selection like users requirements. Tinkerer and engineer are fundamentally similar that in a recent review, Andreas Wagner summarized nine commonalities between biological and technological innovation, such as descent with modification, extinction and replacement, and horizontal transfer \cite{Wagner 2014} [[Don’t we want to highlight some differences before the ID mafia hijacks our paper - ANS ]]. In such a common framework, biological networks and networks with specific performance goals in other disciplines are shaped by similar underlying design objectives. However, as “there’s no free lunch”, there is no way to satisfy all objectives and thus tradeoffs are unavoidable. Networks in different disciplines are perhaps the optimal configurations subjected to different constraints and tradeoffs. Here, via a few specific examples, we are going to illustrate how network comparison works in the broadened context. (~1000 words)

[[Robustness of systems and how they can be used to understand evolution of biological systems - ANS ]]

[[CC: is there any way to show the above argument using a Figure or Table?]]

[[DW: I am not sure if we need to move up above para because this para discusses why bio network and tech network have (dis)similarities, and what kinds of (dis)similarities like evolution..]]

[[DW: transition to come up with the following examples]]

[[KKY: here a few paragraphs that describe specific examples of comparisons really bring intuition to biology which was poorly understood. Possible examples:

· In terms of information propagation/hierarchy, the comparison of regulatory network, management hierarchy

· Linux as an example to illustrate the bio vs tech comparison. Emphasize on the evolution versus tinkerer,

·

~600 words]]

[[DW: please fill your references in summary table (free to add row/column), <https://docs.google.com/spreadsheets/d/1SSY6l4EyZt5vnCtcsHWZL34dWFW8jV7SkvfIjFZob-s/edit#gid=0> ]]

[[DW: time delay is smart engineering design, also appearing in regulatory FFLs]]

[[KKY: a transition sentence, then a few more paragraphs describing examples on the share of common math. structures, and how methods developed in other disciplines are applied to biology. a reasonable way to divide the contents

* Algorithms deal with nodes, e.g. common guilt by association techniques used in bio. and social network, or PageRank for prioritizing importance, network stratification..
* Algorithms deal with links, e.g. link prediction
* Algorithms deal with organization. e.g. OrthoClust and many other modules detection scheme
* Methods deal with dynamics: control theory [[DW3, PM2]], circuit theory [[DW4]], information theory [[DW5]], game theory [[DW7]].

The big data story could be mentioned in these few paragraphs. Most of the examples we found have already been used by others in biological context, are there more novel suggestions? ~800-1000 words]]

[[KKY: the essay should be ended with an outlook, it will be great if there are a few concrete lessons, ~100 words]]

(22615549)

Prioritize nodes by taking into account local topological features.

Google applies a method called PageRank to decide the most relevant web documents based on the hyperlink information between web documents. It ranks a document according to the number of highly ranked documents that point to it. A similar method NetRank has been developed to identify genes prognostic for outcome of patients with cancer. It integrates gene expression with a network of known relationships between the genes-- it assigns a score to a gene which is influenced by the scores of genes connected with it.

Global network alignment:

Giving the networks for examining the same biological questions (e.g. protein-protein interactions) in two different species, it is useful to find a global alignments that best match their nodes in the two network. Although the most straightforward method to do this is define homolog pairs between the two species based on their sequence similarities, it is sometime more desirable to determine the global alignment solely based on their topological similarities. This problem is related the Graph isomorphism problem in graph theory. Many methods have been proposed to solve this problem by first calculating the similarity scores for all pairs of nodes between the two network and then maximize an alignment score by using different alignment strategy (18725631, 19477997, 25015987,21414992,20628593). For example, the IsoRank algorithm applies a PageRank-based spectral graph theoretic principle to assign high similarity scores to node pairs with similar neighborhood. After calculating similarity scores for all node pairs, it determines global alignment by greedily matching the high-scoring node pairs (18725631). Other global network alignment methods apply different strategies for calculating similarities between nodes and optimizing global alignment score (21414992,20628593). In addition, biological scores (e.g. sequence similarity between gene nodes) can be incorporated with these topological scores to improve node similarity measurement.

[[DW: <http://www.casmodeling.com/content/1/1/8> for information theory

Control theory has been widely and successfully applied in various engineering systems for decades. In man-made systems such as engineering systems, ones always want to control systems to achieve their objectives. For example, ones can adaptively control rockets upon environments to arrive desired locations [ref]. Recently, ones have also applied control theory to man-made biomedical systems. For example, the markov decision process (MDP) models in clinical decision support systems, help doctors make optimal treatment decisions based on patient dynamic responses [ref]. Ones also use MDPs to help design optimal drug delivery strategies according to cellular responses to changing environments [ref]. In order to calculate optimal control/decisions, we normally need a huge variety of datasets, which biological systems used to lack. However, as increasing biological datasets available, it becomes feasible to study large-scale biological systems, and engineer them using control theory. For example, according to the genomic features, and interactions to dynamical environments of cancer cells, and , how can we design optimal genomic editing strategies to convert cancer cells to develop to normal cells, since CRISPRs (clustered regularly interspaced short palindromic repeats) make genomic editing much easier than before?

#------------------------------------

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

useful references:

KKY:

1. Evolution of a modular software network (PNAS 2011), <http://www.pnas.org/content/108/50/19985>
   1. an increase of modularity
   2. introduce the concepts of dependency and conflicts -> quantify functionality
2. The evolvability of programmable hardware, <http://rsif.royalsocietypublishing.org/content/8/55/269.abstract>
3. The ecology of collective behavior, <http://dx.doi.org/10.1371/journal.pbio.1001805>
4. evolution of bow-tie architectures in biology, http://arxiv.org/abs/1404.7715

DW:

1.Peter, I. S. and Davidson, E. H. Evolution of gene regulatory networks controlling embryonic development. *Cell* **144**, 970-985, 2011. (PMC3076009)<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3076009/>

Peter, I. S., Faure, E., and Davidson, E. H. Predictive computation of genomic logic processing functions in embryonic development. Proc. Natl. Acad. Sci. USA, **109**, 16434-16442, 2012. PMCID: PMC3478651. <http://www.pnas.org/content/109/41/16434.abstract>

Networking development by Boolean logic (http://www.ncbi.nlm.nih.gov/pubmed/23412653)

2. Modelling and analysis of gene regulatory networks Nature Reviews Molecular Cell Biology 9, 770-780 (October 2008) doi:10.1038/nrm2503, <http://www.nature.com/nrm/journal/v9/n10/full/nrm2503.html>

3. [Controllability of complex networks](http://www.barabasilab.com/pubs/CCNR-ALB_Publications/201105-12_Nature-TamingComplexity/201105-12_Nature-TamingComplexity.pdf) , Nature 473, 167-173 (2011) and [Observability of complex systems](http://www.barabasilab.com/pubs/CCNR-ALB_Publications/201301-28_PNAS-Observability/201301-28_PNAS-Observability.pdf) ,Proceedings of the National Academy of Sciences 110, 1-6 (2013)

4. Brophy, J.A.N. & C.A. Voigt (2014).[**Principles of Genetic Circuit Design.**](http://www.nature.com/nmeth/journal/v11/n5/full/nmeth.2926.html) *Nature Methods,* 101: 508-520.

Reg net -> Boolean -> circuits-> control &observab. ->key regulators&pathways -> cancer -> unbound riches -> world domination

<https://drive.google.com/file/d/0BxgXadzgY4rNaEN6NzNXUjl0eFk/edit?usp=sharing>

5. How Information Theory Handles Cell Signaling and Uncertainty, [Matthew D. Brennan](http://www.sciencemag.org/search?author1=Matthew+D.+Brennan&sortspec=date&submit=Submit), [Raymond Cheong](http://www.sciencemag.org/search?author1=Raymond+Cheong&sortspec=date&submit=Submit), [Andre Levchenko](http://www.sciencemag.org/search?author1=Andre+Levchenko&sortspec=date&submit=Submit), Science 19 October 2012: Vol. 338 no. 6105 pp. 334-335, <http://www.sciencemag.org/content/338/6105/334.full.pdf>

-> variablities across cells, noise -> information capacity -> single cell networks? signaling (communication) -> network/pathway capacities to transform information-> key central/hub nodes(small world networks) how to distinguish multiplexed signals -> under evolutionary pressure, bio systems have optimal information capacity. e.g., Information flow and optimization in transcriptional regulation, PNAS 2008, <http://www.pnas.org/content/105/34/12265.full.pdf>

6. Interplay between gene expression noise and regulatory network architecture.

[Chalancon G](http://www.ncbi.nlm.nih.gov/pubmed?term=Chalancon%20G%5BAuthor%5D&cauthor=true&cauthor_uid=22365642)1, [Ravarani CN](http://www.ncbi.nlm.nih.gov/pubmed?term=Ravarani%20CN%5BAuthor%5D&cauthor=true&cauthor_uid=22365642), [Balaji S](http://www.ncbi.nlm.nih.gov/pubmed?term=Balaji%20S%5BAuthor%5D&cauthor=true&cauthor_uid=22365642), [Martinez-Arias A](http://www.ncbi.nlm.nih.gov/pubmed?term=Martinez-Arias%20A%5BAuthor%5D&cauthor=true&cauthor_uid=22365642), [Aravind L](http://www.ncbi.nlm.nih.gov/pubmed?term=Aravind%20L%5BAuthor%5D&cauthor=true&cauthor_uid=22365642), [Jothi R](http://www.ncbi.nlm.nih.gov/pubmed?term=Jothi%20R%5BAuthor%5D&cauthor=true&cauthor_uid=22365642), [Babu MM](http://www.ncbi.nlm.nih.gov/pubmed?term=Babu%20MM%5BAuthor%5D&cauthor=true&cauthor_uid=22365642).

[Trends Genet.](http://www.ncbi.nlm.nih.gov/pubmed/22365642#) 2012 May;28(5):221-32, <http://www.sciencedirect.com/science/article/pii/S0168952512000157#>

<http://www.nature.com/news/brain-wave-hits-california-1.15454>

<https://www.princeton.edu/~wbialek/rome/refs/laughlin_81.pdf>

7.Networks, Crowds, and Markets:Reasoning About a Highly Connected World

By [David Easley](http://www.arts.cornell.edu/econ/deasley/) and [Jon Kleinberg](http://www.cs.cornell.edu/home/kleinber/) <http://www.cs.cornell.edu/home/kleinber/networks-book/>

* Chapter 7. [Evolutionary Game Theory](http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch07.pdf)
  + 7.1 Fitness as a Result of Interaction
  + 7.2 Evolutionarily Stable Strategies
  + 7.3 A General Description of Evolutionarily Stable Strategies
  + 7.4 Relationship Between Evolutionary and Nash Equilibria
  + 7.5 Evolutionarily Stable Mixed Strategies

“As its name suggests, this approach has been applied most widely in the area of evolutionary biology, the domain in which the idea was first articulated by John Maynard Smith and G. R. Price [375, 376]. Evolutionary biology is based on the idea that an organism’s genes largely determine its observable characteristics, and hence its fitness in a given environment. Organisms that are more fit will tend to produce more offspring, causing genes that provide greater fitness to increase their representation in the population. In this way, fitter genes tend to win over time, because they provide higher rates of reproduction.

The key insight of evolutionary game theory is that many behaviors involve the interaction of multiple organisms in a population, and the success of any one of these organisms depends on how its behavior interacts with that of others. So the fitness of an individual organism can’t be measured in isolation; rather it has to be evaluated in the context of the full population in which it lives. This opens the door to a natural game-theoretic analogy: an organism’s genetically-determined characteristics and behaviors are like its strategy in a game, its fitness is like its payoff, and this payoff depends on the strategies (characteristics) of the organisms with which it interacts. Written this way, it is hard to tell in advance whether this will turn out to be a superficial analogy or a deep one, but in fact the connections turn out to run very deeply: game-theoretic ideas like equilibrium will prove to be a useful way to make predictions about the results of evolution on a population. ”

ANS:

1. Evolution of evolvability in gene regulatory networks - PLoS CB 2008 <http://www.ploscompbiol.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pcbi.1000112&representation=PDF>

This paper models the conditions necessary for evolvability of biological networks. It basically contends that changes in the environment that affects the fitness of the GRN (sequence changes -> network changes -> environmental fitness) are required for evolution to occur in organisms. Also, it contends that the fastest way to make large changes to the network behavior and environmental fitness are through changes to the sequence and network connections of proteins that eventually become hubs of the GRN.

2. [Parameter Space Compression Underlies Emergent Theories and Predictive Models,](http://arxiv.org/abs/1303.6738) Benjamin B. Machta, Ricky Chachra, Mark K. Transtrum, James P. Sethna, [Science **342**, 604-607 (2013](http://www.sciencemag.org/cgi/content/abstract/342/6158/604?ijkey=9EZhzaxPI700I&keytype=ref&siteid=sci)

The main idea of this paper is that in many networks (from ferromagnetic to biological networks), the parameter space is highly sloppy and the behavior of the system is sensitive to a small subset of parameters (or linear combinations of these parameters) while it is robust to all the other parameters in the model. This implies that the behavior of the system may be predictable even though there are huge uncertainties in the parameters within the models. In terms of biology, these could define the nodes in the network that determine the phenotype of the cell.

3. Robustness and Evolvability - Trends in Genetics 2010

<http://www.sciencedirect.com/science/article/pii/S0168952510001101>

This paper contends that it is only possible when the fitness landscape of the organism is flat. In other words, the fitness is robust to small changes to the sequence and network. This makes the system evolvable.

4. Evolution of gene regulatory networks controlling body plan development - Isabelle Peter and Eric Davidson -Cell 2014. <http://www.sciencedirect.com/science/article/pii/S0092867411001310>

This paper reviews that GRN are hierarchical in nature not only in their topology but also in terms of the GRN changes in time and space as well. In other words, the development of a body plan is a heritable hierarchical regulatory system process. In order to understand the evolution of developmental processes, we need to study it in terms of the hierarchical GRN that leads to the evolution of organs.

RK:

‘Grid neurons’ in the entorhinal cortex and hippocampus act as a ‘directionally oriented, topographically organized neural map of the spatial environment’ [1]. There have been a number of efforts to computationally model the collective behaviour of these neurons [2]. Obviously there is a direct connection between neuronal networks and neural network modelling, but specifically modelling how the brain achieves balance despite enormous complexity, for example by regulating excitatory and inhibitory sub-circuits, is an interesting problem [3]. Finally the integration of multi-sensory information is one of the key functions of the CNS in any organism and has been postulated that a canonical set of operations is responsible for this integration [4]. On a slightly different (more sociological) track, ‘transport’ networks created by ants obtain similar efficiencies as do those created by humans, but without centralised control [5], which are both robust and low-cost [6].

[1] <http://www.nature.com/nature/journal/v436/n7052/full/nature03721.html>

[2] <http://www.cell.com/neuron/abstract/S0896-6273(11)00650-7> (review)

[3] <http://www.sciencemag.org/content/274/5293/1724.short>

[4] <http://www.sciencedirect.com/science/article/pii/S0896627314001949> (review)

[5] <http://link.springer.com/article/10.1007/s00265-008-0680-7>

[6] <http://link.springer.com/article/10.1140%2Fepjb%2Fe2004-00364-9?LI=true>

PM

Review:

Design Principles of Regulatory Networks: Searching for the Molecular Algorithms of the Cell (Lim, Lee, Tang) – Molecular Cell

<http://www.cell.com/molecular-cell/abstract/S1097-2765(13)00004-X>

1. Application of Pareto efficiency to evolutionary optimization

Pareto evolution of gene networks: an algorithm to optimize multiple ﬁtness objectives (Warmﬂash, Francois, Siggia) – Physical Biology

<http://www.physics.mcgill.ca/~paulf/DOC/PhysBiol2012.pdf>

Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space (Shoval, Alon, et al.) – Science

http://wws.weizmann.ac.il/mcb/UriAlon/sites/mcb.UriAlon/files/shoval2012.pdf

2. Application of control theory to biological networks

A Control Theoretic Framework for Modular Analysis and Design of Biomolecular Networks (Del Vecchio) – Annual Reviews in Control

<http://web.mit.edu/ddv/www/papers/IFACReviews.pdf>

3. Neural network computation with DNA strand displacement cascades (Qian, Winfree, Bruck) - Nature

<http://www.nature.com/nature/journal/v475/n7356/pdf/nature10262.pdf>

4. Maps of random walks on complex networks reveal community structure

<http://www.pnas.org/content/105/4/1118.abstract>

Mapping Change in Large networks

# <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0008694>

Common effector processing mediates cell-specific responses to stimuli

<http://www.nature.com/nature/journal/v448/n7153/full/nature06001.html>

CC:

1.HIDEN: Hierarchical decomposition of regulatory networks. Gülsoy G, Bandhyopadhyay N, Kahveci T. BMC Bioinformatics. 2012 Sep 28;13:250. doi: 10.1186/1471-2105-13-250. PMID:23016513

-- transform the problem into a mixed integer programming problem

2.HiNO: an approach for inferring hierarchical organization from regulatory networks. Hartsperger ML, Strache R, Stümpflen V. PLoS One. 2010 Nov 4;5(11):e13698. doi: 10.1371/journal.pone.0013698. PMID:21079808

-- algorithm based on breadth first search method. Their solution improves the BFS-level method, and outputs a hierarchy for every network regardless of its topological features. However all these algorithms fail to minimize the number of edges that violate the hierarchy.

3.Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. Jothi R1, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, Babu MM. Mol Syst Biol. 2009;5:294. doi: 10.1038/msb.2009.52.

-- vertex sort method. This method incorporates topological sort algorithm for addressing the network hierarchy problem

4.Genomic analysis of the hierarchical structure of regulatory networks. Yu H, Gerstein M. Proc Natl Acad Sci U S A. 2006 Oct 3;103(40):14724-31. PMID:17003135

-- This method uses breadth first search to assign hierarchies to TFs in a network. it fails to assign accurate levels for networks that contain cycles.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Earlier minutes:

MG: I envision that this essay is going to discuss how the study of biological networks can profit from comparisons with networks in other disciplines (eg social networks) and how there are some concepts used in networks in other disciplines that might be useful in biology in the future.

Draft: Mid Aug

Deadline:9/15

7/9/2014

{{

why should we do network comparisons beneficial for biology ?

\* through comparison we can study important themes - evol. , robustness, eff.

\* through comparisons we can apply mathematical formalism that’s best developed elsewhere to biology - xray - many network concepts couldn’t have been developed first in biology but they’re useful when applied

\* data science : networks provide a basic and primal metaphor & formalism for grappling with complex systems

- only recently have people realized that biology is a discipline complexity….

- only recently have we had big data

\* if want to understand complex systems, we need to develop intuition for how they work & biology is poor

Good network comparisons are beneficial but we can only appreciate in the context of potential bad

what is a bad comparison }}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

6/10/14

Theme - How networks in non-biological disciplines help understand biological networks & suggest new ways to approach them

Examples - linux call graph and regulatory networks - greater exposition

- Connectivity follows constraints typically . why Linux is different?

- Analogy.

Example 2 -Rewiring rates

Example 3 - Hierarchy

5. Davidson - Circuits to systems biology - Example

4. Community structure and how that motivates gene expression clustering - Example

Man made networks and natural networks - differences

2. Scale free structure of networks - critical phenomenon leads to power law.

Werner Heisenberg quote which interpreted as “people get intuition of how molecular networks work from examples in the real world like social networks (commonplace networks)”.

3A. Mathematical Models - Diffusion Model, Bayesian Model, Boolean Model - also has network representation.

- ISORANK

6. 6 degrees of separation.

Watts Strogatz model - small scale networks

1. Barabasi mode - preferential attachment - Matthew effect - rich get richer - gene duplication in biological context - different kinds of hubs - SIN - mutually exclusive interactions.

3B. Controllability and observability - Barabasi

Robustness - Jim Sethna