Systems approaches in understanding evolution and evolvability

Sumeet Agarwal

Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi – 110016, India

Abstract

Systems and network-based approaches are becoming increasingly popular in cellular biology. One contribution of such approaches has been to shed some light on the evolutionary origins of core organisational principles in biological systems, such as modularity, robustness, and evolvability. Models of interactions between genes (epistasis) have also provided insight into how sexual reproduction may have evolved. Additionally, recent work on viewing evolution as a form of learning from the environment has indicated certain bounds on the complexity of the genetic circuits that can evolve within feasible quantities of time and resources. Here we review the key studies and results in these areas, and discuss possible connections between them. In particular, we speculate on the link between the two notions of 'evolvability': the evolvability of a system in terms of how agile it is in responding to novel goals or environments, and the evolvability of certain kinds of gene network functionality in terms of its computational complexity. Drawing on some recent work on the complexity of graph-theoretic problems on modular

Email address: sumeet@iitd.ac.in (Sumeet Agarwal)
URL: http://web.iitd.ac.in/~sumeet (Sumeet Agarwal)

networks, we suggest that modularity as an organising principle may have its raison d'etre in its ability to enhance evolvability, in both its senses.
Keywords: Systems, Networks, Evolution, Evolvability, Modularity

1. Introduction

The last decade or so has witnessed the rise to prominence of an approach to biology that seeks to be more holistic than 'conventional' biology, by experimenting on and modelling simultaneously the interactions between different units like genes, proteins, and metabolites. This has broadly come to be known as *systems biology* (Ideker et al., 2001; Kitano, 2002; Noble, 2010). A significant aspect of the systems approach has been the use of networks (Newman, 2003, 2009) to model the connectivity structures between such units (Barabási and Oltvai, 2004). In this paper, we look at some of the key results on the evolutionary implications of different biological network models, including with regard to the evolution of reproduction.

A network consists of a set of elements (called nodes or vertices) and a set of pairwise connections between those elements (called links or edges). The type of networks we will be primarily concerned with here are *gene regulatory networks*, which have been very widely studied in recent years (Schlitt and Brazma, 2007; Hecker et al., 2009). Here the nodes are genes, and the links represent regulatory relationships between genes, which may correspond to either activation or inhibition. The detailed structure of such interactions is depicted in Figure 1. A wide range of mathematical models have been used to represent the dynamics of gene expression in such networks, ranging from full-fledged ordinary differential equations to Boolean networks, where

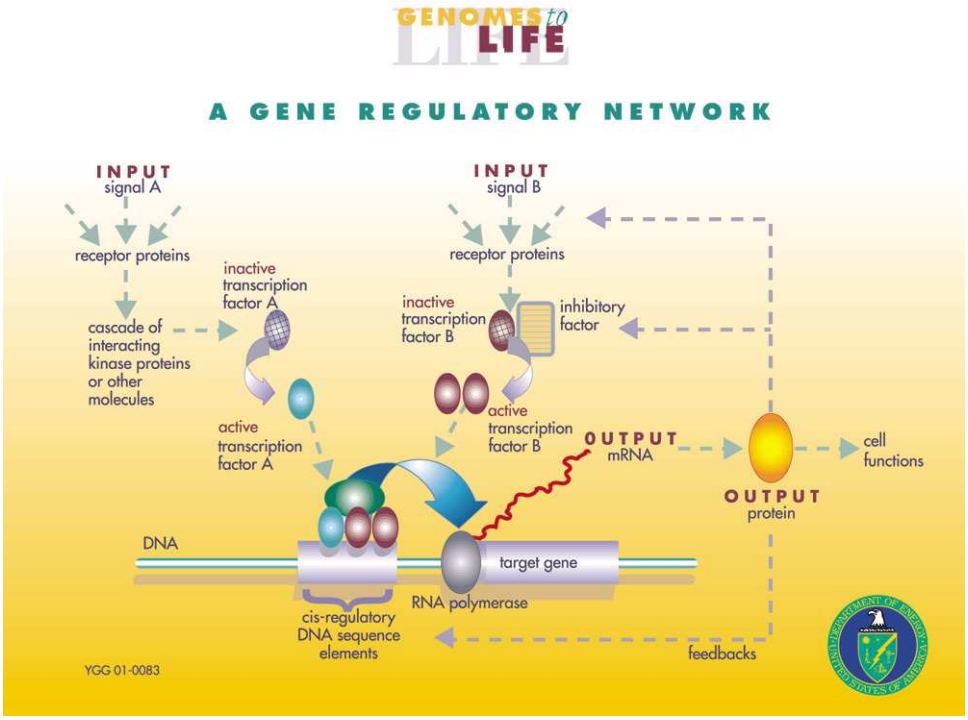

Figure 1: Detailed structure at a node of a gene regulatory network. Figure reproduced from `http://genomics.energy.gov`.

each gene is always in one of two states, on or off (Schlitt and Brazma, 2007; Hecker et al., 2009). Many studies looking at evolution and development have made use of a basic Boolean network model first proposed by Wagner (1994, 1996a), described further in Section 2.

Some of the evolutionary questions that gene network-based studies have sought to address include:

- The emergence of properties such as modularity, robustness, and evolvability in biological organisms (see Section 3).

- The conditions that might favour sexual or asexual reproduction, with particular reference to the role of epistasis, or dependencies between the expression levels of genes — such dependencies are just what network models seek to capture (see Section 4).

- What kinds of genetic circuitry and functionality are realistically evolvable, given time and resource constraints (see Section 5).

Here we review extant work in each of these directions, and then discuss some broader implications, potential connections and prospects for future research (Section 6).

2. Models of gene network dynamics and evolution

A basic Boolean model for gene network dynamics was proposed by Wagner (1994, 1996a), and has been extended or adapted for use in several other studies (Siegal and Bergman, 2002; Azevedo et al., 2006; MacCarthy and Bergman, 2007). This model makes a number of simplifying assumptions: most significantly, that the genes have just two relevant expression states, which correspond to on/off, and that the regulatory effects of genes are independent of each other. Thus, the expression state of gene i at time t is denoted by $x_i(t)$, which is either $+1$ (on or expressed) or -1 (off or not expressed). If we take a time step of τ, then the expression level of the i^{th} gene at time $t + \tau$ is specified in terms of the expression states of all genes (say we have N of them) at time t, as follows (Wagner, 1996a):

$$x_i(t + \tau) = \sigma \left[\sum_{j=1}^{N} w_{ij} x_j(t) \right]. \tag{1}$$

Here w_{ij} is a weight specifying the extent to which the j^{th} gene affects the expression of the i^{th} gene: positive weights correspond to activation, and negative weights to repression. $\sigma()$ is the sign function, i.e., $\sigma(y) = +1$ for $y > 0$ and -1 for $y < 0$. Once the initial expression states $x_i(0)$ have been specified (these may be based on certain assumptions, or thought of as determined by external factors), then the model gives the dynamics of how these states evolve with time. These dynamics will eventually lead to some equilibrium state, which may be a steady state (fixed point) where the states become constant, or a limit cycle where they oscillate in a fixed pattern.

Thus far we have a model for network dynamics. How does this network evolve with successive biological generations? To model this, one may consider that there is an optimal equilibrium state $\mathbf{x}^{\text{opt}}(\infty) = \{x_1^{\text{opt}}(\infty), ..., x_N^{\text{opt}}(\infty)\}$, where the ∞ denotes that this is a state attained after the dynamics have run for an infinitely long time. This may be motivated on the basis that if a gene pathway network is acting in a developmental process, it is thought to require a certain expression state to proceed optimally, and any deviations that perturb development will reduce the organism's fitness (Wagner, 1996a). Then, for any given network, the equilibrium expression state achieved by it may be denoted $\mathbf{x}(\infty)$, and we would like to define a distance measure between state vectors, for which a standard choice would be the Hamming distance: $d[\mathbf{x}^{\text{opt}}(\infty), \mathbf{x}(\infty)] = 1/2 - 1/(2N) \sum_{i=1}^{N} x_i(\infty) x_i^{\text{opt}}(\infty)$ (Wagner, 1996a). Using this, the fitness $f(\mathbf{x}(\infty))$ of an individual with gene network corresponding to equilibrium state $\mathbf{x}(\infty)$ was defined by Wagner (1996a) via a Gaussian function of the following form:

$$f(\mathbf{x}(\infty)) = \exp\left(-\frac{d[\mathbf{x}^{\text{opt}}(\infty), \mathbf{x}(\infty)]^2}{s}\right). \tag{2}$$

Here $s > 0$ is a parameter which specifies the variance or spread of the Gaussian; it represents the strength of selection for the optimal network, with small s implying strong selection against variation from the optimum.

Given that a selection criterion has been defined, one can now set up an evolutionary simulation structured as follows:

1. An ensemble of individuals (with given gene network characteristics) is generated via some random mechanism.
2. Their fitness is evaluated as per (2), and the fittest ones are preferentially reproduced in the next generation (via possible recombination and mutation steps).
3. Steps 1 and 2 are repeated for an appropriate number of generations or until the fitness optimum has been attained.

The use of this model to study the evolution of properties like robustness, evolvability, and sex is discussed in Sections 3 and 4.

In some cases, the evolution of generic network characteristics like modularity and frequencies of motifs (small subgraphs with particular connectivity patterns) has been studied in synthetic systems that are meant to solve certain logical computation or pattern recognition tasks, like electronic circuits and artificial neural networks (Kashtan and Alon, 2005; Kashtan et al., 2007; Clune et al., 2013). Specifically, these systems were evolved (via simulations of the fashion just described) under specific types of constraints:

- *Modularly varying goals* (Kashtan and Alon, 2005; Kashtan et al., 2007), where the fitness is defined not just by performance on a single task, but by the ability to adapt quickly to multiple tasks that contain common subtasks (a situation that often seems to apply in real life). For

example, Kashtan and Alon (2005) evolved logic circuits that had to compute the Boolean functions (X XOR Y) AND (Z XOR W) and (X XOR Y) OR (Z XOR W); they kept switching the goal from one to the other every 20 generations, and found that the system evolved two separate modules to calculate the two subgoals, (X XOR Y) and (Z XOR W). This made it possible to quickly rewire just the inter-module connections to adapt each time the goal was switched.

- *Minimal connection cost* (Clune et al., 2013), where in addition to performance on a given task, the defined fitness also included a component for how low the total cost of the connections between the network nodes was. This cost may be defined just as the number of connections in the network (for unweighted links), or the total sum of all link weights. Thus the evolved networks were attempting to simultaneously maximise performance and minimise total connection cost.

In the following section, we discuss what simulations with these models revealed about the evolution of modularity in networked systems.

3. Modularity, robustness, evolvability

Many biological and engineered systems are observed to have the property of modularity, i.e., they consist of distinct subunits which function largely independently of each other (Hartwell et al., 1999; Kashtan and Alon, 2005; Wagner et al., 2007). In the context of networks, 'modularity' has been given a more formal mathematical definition by Newman and Girvan (2004). Many real-world networks display some sort of modular organisation, as they can

be partitioned into cohesive groups of nodes such that there is a relatively high ratio of internal (within-group) to external (between-group) link density (the number of links as a fraction of the number of possible links). Such sub-networks, known as communities, are often construed to correspond to distinct functional units (Girvan and Newman, 2002; Fortunato, 2010; Porter et al., 2009; Agarwal, 2012). The Newman-Girvan technique attempts to partition a network into groups of nodes so as to maximise the number of links between nodes in the same group, whilst minimising links between nodes in different groups. For a given partition, it quantitatively defines modularity as essentially the excess of within-group links, relative to a comparable random network with no modular structure (Newman and Girvan, 2004; Newman, 2006).

Kashtan and Alon (2005) used this measure of modularity to study whether, in their model systems of logic circuits and neural networks mentioned earlier, higher modularity would emerge under certain kinds of evolutionary selective pressures. They found that evolution under rapidly switching, modularly varying goals leads to networks with high modularity and pronounced motifs, whereas evolution under a single goal led to relatively nonmodular solutions with low motif frequencies.

Their results further suggested that networks evolved under the first condition appear to discover the common subtasks between the different goals, and they evolve separate network modules for dealing with each subtask.

Thus, when the goal switches, the system is able to quickly evolve to adapt, by keeping the same modules and just rewiring a few connections between them (Kashtan et al., 2007). This is an instance of how system

modularity (itself evolved in an environment with modularly varying goals) can be an enabler of evolvability (Wagner, 1996b; Wagner and Altenberg, 1996).

There have also been other similar studies, in particular those by Hod Lipson and collaborators, showing that modularity in networks can emerge in response to environmental variation (Lipson et al., 2002), requirements for stable linear dynamics (Variano et al., 2004), or the minimal connection cost criterion (Clune et al., 2013). The last interestingly suggests that the pressure to reduce connection costs may serve as an initial driver for the emergence of modularity, a kind of bootstrapping process that creates sufficient modularity to then allow selection for its evolvability benefits to set in, in environments like the one with modularly varying subgoals. This may be necessary because it has been suggested that the evolvability selection cannot get started until modularity has reached a certain threshold, where its benefits for adaptation speed begin to become apparent (Wagner et al., 2007; Clune et al., 2013).

Robustness is another property of biological systems that has been suggested to be related to modularity and evolvability (Wagner, 1996b; Pigliucci, 2008). Robustness can have various senses: one is mutational robustness, which is naturally enhanced in a modular system, since a mutation in a gene would generally affect only the corresponding functional module (i.e., there would be no pleiotropy), with the functioning of the other modules being essentially independent (Wagner et al., 2007). A general notion of robustness is in terms of the genotype-to-phenotype mapping: random or accidental changes in the genotype should get buffered to preserve phenotypic outcomes, like expression states in developmental pathways as indicated earlier. This

property was referred to by Waddington (1942) as 'canalisation', and is also known as epigenetic stability (Wagner, 1996a). A key idea has been that both modularity and robustness are building blocks of evolvability, and that natural selection leads to systems which are largely robust to mutations and perturbations, but that can also respond quickly to changed environments and take on new functionality, due to the effectiveness of certain special types of mutations (e.g., those leading to rewiring between existing modules) (Wagner, 2005, 2008; Pigliucci, 2008). Andreas Wagner's notion of 'neutral spaces' (Wagner, 2005) proposes that the canalisation or epigenetic stability allows a range of genetic variants to build up in a population, which are in essence functionally equivalent; the mutations mapping between them are 'neutral' in phenotype space. However, the existence of these variants enhances evolvability, as in a novel environment, some of them may be able to adapt more easily and thus get selected. Modularity may even be a consequence of selection for such robustness, as it may imply lower pleiotropy (Wagner et al., 2007). It has also been demonstrated, using gene network models of development dynamics as described in Section 2, that canalisation may emerge just as a result of the need for developmental stability, and thus may require no explicit selection for suppressed phenotypic variation (Siegal and Bergman, 2002). Another recent simulation-based study by Steiner (2012) has suggested that robustness and evolvability, just like modularity, can emerge in response to fluctuating environments, in particular if the noise or variation is autocorrelated.

Interesting correlations have also been suggested between modularity, robustness and modes of reproduction; we look at some of these next.

4. The evolution of reproduction

The rationale for the evolution of sexual reproduction has long been a matter of debate (de Visser and Elena, 2007). Studies suggest that the nature of epistasis has a key role to play in determining the favourability of sex and recombination, in terms of their contribution to mutational robustness (Kondrashov, 1988; Azevedo et al., 2006; MacCarthy and Bergman, 2007). There are two kinds of epistasis possible: negative/synergistic, where the mutation of multiple genes is more harmful to fitness than the sum of the effects of the individual mutations, and positive/antagonistic, where the combined effect is less harmful than the individual ones. It has been hypothesised (the *deterministic mutation hypothesis*) that sexual reproduction confers particular robustness benefits when there is negative epistasis, because recombination can cause individual deleterious mutations to come together, and then if there is a substantial fitness loss such individuals will get selected against, leading to the purging of the mutations from the population (Kondrashov, 1988; Azevedo et al., 2006). However this effectively just pushes the question back a step, as one might now ask why negative epistasis itself evolved?

Azevedo et al. (2006), working with a gene network model of the type discussed in Section 2, show via simulations that negative epistasis can actually evolve as a consequence of sexual reproduction itself. Thus there is the possibility of a bootstrapping process; once some degree of negative epistasis has emerged initially, it favours the evolution of sex, which in turn selects for further negative epistasis, thus enabling its own maintenance. In their simulations, network evolution was implemented both sexually (via recombination) and asexually (mutation only), with selection for networks that

exhibited stable gene expression patterns; it was found that sexual reproduction leads to the emergence of greater mutational robustness than asexual reproduction. Also, starting from an initial condition of average positive epistasis, only sexual reproduction led to a switchover to negative epistasis. In similar work, Misevic et al. (2006) also found that sexual reproduction leads to greater robustness, and also greater modularity in the genome. Steiner (2012) found that both robustness and evolvability evolve more strongly in sexually reproducing populations.

MacCarthy and Bergman (2007) also carried out simulations with the same type of gene network model, but they allowed both epistasis and the reproductive mode (sexual vs. asexual, i.e., the amount of recombination) to coevolve, rather than just fixing one and looking at the effect on the other. They found that in this setting, asexual reproduction actually outdoes sexual reproduction in terms of fitness benefits, thus casting doubt on the deterministic mutation hypothesis. Their results indicate that epistasis does not determine the fate of the reproductive mode, but that the converse determination does happen: asexual reproduction leads to positive epistasis, whilst sexual reproduction encourages negative epistasis, in accordance with earlier studies (Azevedo et al., 2006; Misevic et al., 2006). Thus, MacCarthy and Bergman (2007) demonstrate the importance of looking at the coevolution of parameters like epistasis and recombination, rather than examining one in isolation, and tells us that we still do not have a clear-cut explanation for the evolution of sex.

5. Evolution as computational learning

The term evolvability has been used so far in our discussion to mean the capacity of an organism to produce new variants and functionality and thus respond to environmental changes (Wagner and Altenberg, 1996). Recent path-breaking work by Feldman, Valiant, and collaborators (Feldman, 2008; Feldman and Valiant, 2008; Valiant, 2009; Feldman, 2009; Kanade et al., 2010) has sought to formalise a somewhat different notion of evolvability: what kinds of functionality can evolve at all, given reasonably constrained time and resources? This work places evolution with the context of computational learning theory (Valiant, 1984; Anthony and Biggs, 1997), viewing it as a form of learning from experience (across generations). This is a particularly interesting perspective because it implies that evolved characteristics are also 'learnt' (just as the learning that happens over an individual's lifetime); and thus the traditional 'nature vs. nurture' debate may be irrelevant to understanding the emergence of biological behaviour.

The puzzle of evolvability is how complex mechanisms can evolve without the occurrence of unlikely events (Valiant, 2009). Darwin's key insight was that natural selection could provide a plausible answer. In order to reach a specific functional target, there needs to be an evolutionary path consisting of small steps, each of which conveys some discernible fitness benefit. One might ask, what are the conditions in biology which allow such paths to be taken routinely and efficiently? Are some kinds of mechanisms too complex to be evolvable via such paths in a feasible amount of time? These kinds of questions had not been formally addressed prior to Feldman and Valiant. Their major contribution is to have laid out a mathematical

framework within which this can be done. If we view biological mechanisms as mathematical functions, then considerations of computational complexity allow us to examine what sorts of function classes are feasibly evolvable.

The following are the fundamental notions underlying the framework for evolvability set out in Valiant (2009):

- The targets of evolution are essentially many-argument functions. In the context of gene networks, for instance, each gene's expression level can be thought of as being determined by a function of all other genes that regulate it (possibly including itself). Suppose there are N genes that could regulate gene i; we denote the expression levels of these genes by $\mathbf{x} = \{x_1, x_2, ..., x_i, ..., x_N\}$ (in this model time has not been considered explicitly, as it is assumed that these are steady state expression levels). Then the expression level of gene i is some function of all these expression levels: $x_i = f(x_1, x_2, ..., x_N)$; for instance, $f()$ could be a simple sign function of a linear function, of the form of (1). Here we assume that each x_i is restricted to being either $+1$ or -1, as also in the model described in Section 2.

- For any gene, there is assumed to be an optimal function $f(\mathbf{x})$ for its expression state; this is the function, which if achieved, would confer maximum fitness to the organism. Suppose the actual function evolved is $r(\mathbf{x})$, then a performance measure is defined which quantifies the similarity between the actual and optimal functions, given a probability distribution over input states $D_N(\mathbf{x})$:

$$\mathrm{Perf}_f(r, D_N) = \sum_{\mathbf{x}} f(\mathbf{x}) r(\mathbf{x}) D_N(\mathbf{x}). \tag{3}$$

This can be seen as a fitness landscape over the space of possible functions $r(\mathbf{x})$: it effectively captures the correlation between $r(\mathbf{x})$ and $f(\mathbf{x})$, with a value of 1 corresponding to perfect correlation and -1 to perfect anti-correlation (see Figure 2). These functions can effectively be thought of as representing the regulatory network controlling the expression of a given gene, just as described in the model presented in Section 2. If the distribution $D_N(\mathbf{x}$ is non-uniform, then the performance measure becomes a weighted correlation over input states, with each state weighted by its likelihood of occurrence. This allows it to capture more general situations where not all gene expression states are equally likely in a given environment.

- It is assumed that the size of the population that can be produced in each generation is limited (limited resources).

- It is assumed only a limited number of generations are available for the emergence of any given mechanism or function type (limited time).

Given these guidelines, Valiant (2009) uses the mathematical framework of Probably Approximately Correct (PAC) learning (Valiant, 1984) to evaluate whether certain classes of functions can be evolved (with high probability and small error) within time and resources that are polynomial functions of the number of genes N. Evolution occurs via mutation and selection based on the performance measure (3). Mutations are appropriate small changes, based on the form of the function class being dealt with; for instance, if we are optimising over the space of all disjunctive functions (those of the form x_1 OR x_2 OR x_4..., so that if any one of the inputs is on then the output is

x_1	x_2	f	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1
-1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	-1
1	-1	1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1
1	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1

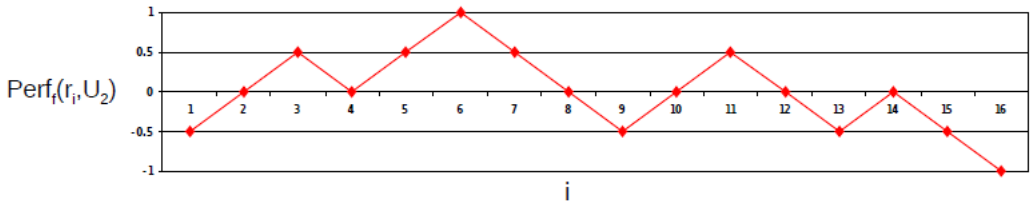

Figure 2: Example of how the performance measure defines a fitness landscape over possible functions. Here f is assumed to be the ideal or target function; r_1 to r_{16} are the 16 possible binary two-input functions; and U_2 denotes the uniform distribution over possible input states for the 2 genes, x_1 and x_2.

on), then a single mutation may involve removing or adding a single variable.

Within this setup, Valiant is able to prove that some function classes are evolvable whilst others are not. For example, monotone Boolean conjunctions (e.g., x_1 AND x_2 AND x_4...) and disjunctions (e.g., x_1 OR x_2 OR x_4...) are shown to be evolvable if the distribution D_N over input instances is uniform, meaning that if the search space is restricted to one of these classes, then the optimal function within the class can be evolved in polynomial time and resources. For arbitrary input distributions evolvability remains an open question. On the other hand, parity functions (functions that return +1 if and only if an odd number of the inputs are +1) are not evolvable in this sense for any sort of distribution over input instances.

These results are primarily interesting because they demonstrate that the Darwinian notion of evolution can be mathematically formalised and the tractability of the evolution of different sorts of complex mechanisms can actually be studied analytically. Since certain function classes are not feasibly evolvable, biology must have made a choice of which classes are actually implemented in living systems (like gene networks); it is interesting to ask how this choice got made. Does it reflect physical constraints on the sorts of mechanisms that are actually possible in biochemistry? Could the restriction on possible types of mechanisms lie behind the emergence of modularity?

6. Discussion

In this paper we have surveyed a variety of approaches involving the use of 'systems' or network-based approaches to try and understand the evolu-

tion of biological functionality. One set of studies has used simulations of evolving network models, and suggested possible explanations for the emergence of properties like modularity: it may facilitate adaptation to modularly varying goals (Kashtan and Alon, 2005), and/or it may simply reflect selection for lower connection costs (Clune et al., 2013). Similar models have also been employed to study the effects of sexual and asexual reproduction. Whilst it was shown that in a sense, sexual reproduction could 'forge its own path', by favouring the emergence of negative epistasis, which in turn favours the maintenance of sex for robustness reasons (Azevedo et al., 2006), another study indicates that if both reproduction mode and epistasis are allowed to coevolve, the picture becomes more complicated and in fact asexual reproduction appears to outcompete sexual reproduction (MacCarthy and Bergman, 2007). Thus, the results of such simulation studies need to be interpreted with caution. Finally, we discussed some recent work on a different, more formal notion of evolvability, which models evolution as a form of computational learning, and (using a number of simplifying assumptions), is able to mathematically prove that certain kinds of complex mechanisms are feasibly evolvable whilst others are not.

The possibilities for combining the insights from some of the work presented here are of great interest. In particular, can the analytic approach of Valiant be extended to examine the questions that have been sought to be addressed by simulation thus far? To take a specific example, a running theme through this survey has been the property of modularity shown by biological systems, and what evolutionary explanations it might have. We suggest exploring the connections between the following three strands of enquiry:

- Simulations which have shown that modular gene networks can achieve certain tasks or functionality either quicker (Kashtan and Alon, 2005; Kashtan et al., 2007) or more cheaply (Clune et al., 2013).

- Theoretical work on evolvability which allows us to put bounds on what is feasibly evolvable. It has also been suggested by Valiant (2009) that modularity in biology could be "a consequence of the limitations of evolvability". At the same time, this work has also suggested that learning occurring in multiple phases, with varying target functions, can lead to the evolution of arbitrarily complex genetic circuits; how does this link to the simulations showing the emergence of modularity and greater evolvability under modularly varying goals (Kashtan and Alon, 2005; Kashtan et al., 2007)? Also, can the Valiant approach be extended somehow to show that the existence of modularity reduces the complexity (in terms of either time or resources) of evolving certain kinds of functionality on gene networks, providing analytical backing to the simulation results of Kashtan and Alon (2005); Kashtan et al. (2007); Clune et al. (2013)?

- Some very recent work from a more general perspective has demonstrated that modular structure in networks can make it easier to solve certain hard graph-theoretic problems, both empirically (Agarwal, 2012) and theoretically (Bui-Xuan and Jones, 2013). In particular, these have looked at problems involving finding the shortest paths or walks traversing all the nodes in a given network or graph (the famous travelling salesman problem, or variants thereof). Could this also be relevant

to the evolution of modularity? If we conceive of the primary role of biological systems as being information processing, then it is certainly plausible that tasks like information flow on gene networks might also be aided by structural properties that make routing easier. Can this notion be formalised and analysed within the Valiant framework? Can the graph-theoretic complexity analysis of Bui-Xuan and Jones (2013) help to modulate the evolvability of certain complex mechanisms on networks that have high modularity?

In the 150 years since Darwin proposed his seminal theory, it has led to huge advances in our understanding of biology; but many questions have also remained unanswered. The powers and limits of evolution continue to be fascinating topics for enquiry. In particular, the evolution of sexual reproduction remains a significant mystery, despite the existence of several varied and competing hypotheses attempting to explain it. Systems approaches, brought to prominence by the high-throughput data revolution, have led to some promising insights into the nature of evolution in recent years. But much work still remains to be done to build on these and obtain experimental verification of their outcomes, in order to connect the simple abstractions of networked models with the real-world complexity of living organisms.

References

Agarwal, S., 2012. Networks in Nature: Dynamics, Evolution, and Modularity. Ph.D. thesis. University of Oxford.

Anthony, H., Biggs, N., 1997. Computational Learning Theory. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press.

Azevedo, R.B.R., Lohaus, R., Srinivasan, S., Dang, K.K., Burch, C.L., 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature 440, 87–90.

Barabási, A.L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. Nature Reviews Genetics 5, 101–113.

Bui-Xuan, B.M., Jones, N.S., 2013. How modular structure can simplify tasks on networks. arXiv preprint arXiv:1305.4760 .

Clune, J., Mouret, J.B., Lipson, H., 2013. The evolutionary origins of modularity. Proceedings of the Royal Society B: Biological Sciences 280.

Feldman, V., 2008. Evolvability from learning algorithms, in: Proceedings of the 40th annual ACM symposium on Theory of computing, ACM. pp. 619–628.

Feldman, V., 2009. Robustness of evolvability, in: Proceedings of COLT, pp. 277–292.

Feldman, V., Valiant, L.G., 2008. The learning power of evolution, in: Proceedings of COLT, pp. 513–514.

Fortunato, S., 2010. Community detection in graphs. Physics Reports 486, 75–174.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99, 7821–7826.

Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell biology. Nature 402, C47–C52.

Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., Guthke, R., 2009. Gene regulatory network inference: Data integration in dynamic modelsa review. Biosystems 96, 86 – 103.

Ideker, T., Galitski, T., Hood, L., 2001. A new approach to decoding life: Systems biology. Annual Review of Genomics and Human Genetics 2, 343–372. PMID: 11701654.

Kanade, V., Valiant, L.G., Vaughan, J.W., 2010. Evolution with drifting targets, in: Proceedings of COLT, pp. 155–167.

Kashtan, N., Alon, U., 2005. Spontaneous evolution of modularity and network motifs. Proceedings of the National Academy of Sciences of the United States of America 102, 13773–13778.

Kashtan, N., Noor, E., Alon, U., 2007. Varying environments can speed up evolution. Proceedings of the National Academy of Sciences 104, 13711–13716.

Kitano, H., 2002. Systems biology: A brief overview. Science 295, 1662–1664.

Kondrashov, A.S., 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336, 435–440.

Lipson, H., Pollack, J.B., Suh, N.P., 2002. On the origin of modular variation. Evolution 56, 1549–1556.

MacCarthy, T., Bergman, A., 2007. Coevolution of robustness, epistasis, and recombination favors asexual reproduction. Proceedings of the National Academy of Sciences 104, 12801–12806.

Misevic, D., Ofria, C., Lenski, R.E., 2006. Sexual reproduction reshapes the genetic architecture of digital organisms. Proceedings of the Royal Society B: Biological Sciences 273, 457–464.

Newman, M., 2009. Networks: An Introduction. OUP Oxford.

Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Review 45, 167–256.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proc. Natl Acad. Sci. USA 103, 8577–8582.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69, 026113.

Noble, D., 2010. Biophysics and systems biology. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368, 1125–1139.

Pigliucci, M., 2008. Is evolvability evolvable? Nature Reviews Genetics 9, 75–82.

Porter, M.A., Onnela, J.-P., Mucha, P.J., 2009. Communities in networks. Notices of the American Mathematical Society 56, 1082–1097, 1164–1166.

Schlitt, T., Brazma, A., 2007. Current approaches to gene regulatory network modelling. BMC Bioinformatics 8, S9.

Siegal, M.L., Bergman, A., 2002. Waddington's canalization revisited: Developmental stability and evolution. Proceedings of the National Academy of Sciences 99, 10528–10532.

Steiner, C.F., 2012. Environmental noise, genetic diversity and the evolution of evolvability and robustness in model gene networks. PLoS ONE 7, e52204.

Valiant, L.G., 1984. A theory of the learnable. Communications of the ACM 27, 1134–1142.

Valiant, L.G., 2009. Evolvability. J. ACM 56, 3:1–3:21.

Variano, E.A., McCoy, J.H., Lipson, H., 2004. Networks, dynamics, and modularity. Phys. Rev. Lett. 92, 188701.

de Visser, J.A.G.M., Elena, S.F., 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. Nature Reviews Genetics 8, 139–149.

Waddington, C.H., 1942. Canalization of development and the inheritance of acquired characters. Nature 150, 563–565.

Wagner, A., 1994. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. Proceedings of the National Academy of Sciences 91, 4387–4391.

Wagner, A., 1996a. Does evolutionary plasticity evolve? Evolution 50, pp. 1008–1023.

Wagner, A., 2005. Robustness and evolvability in living systems. Princeton University Press Princeton.

Wagner, A., 2008. Robustness and evolvability: a paradox resolved. Proceedings of the Royal Society B: Biological Sciences 275, 91–100.

Wagner, G.P., 1996b. Homologues, natural kinds and the evolution of modularity. American Zoologist 36, 36–43.

Wagner, G.P., Altenberg, L., 1996. Perspective: Complex adaptations and the evolution of evolvability. Evolution 50, pp. 967–976.

Wagner, G.P., Pavlicev, M., Cheverud, J.M., 2007. The road to modularity. Nature Reviews Genetics 8, 921–931.