(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Theoretical Biology 282 (2011) 65–70

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Functionality and metagraph disintegration in boolean networks

Jamie X. Luo [a,b], Matthew S. Turner [a,b,*]

[a] Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK
[b] Department of Physics, University of Warwick, Coventry CV4 7AL, UK

ARTICLE INFO

Article history:
Received 19 November 2010
Received in revised form
3 April 2011
Accepted 5 May 2011
Available online 13 May 2011

Keywords:
Metagraph
Evolution

ABSTRACT

We study regulatory networks of N genes giving rise to a vector expression profile $v(t)$ in which each gene is Boolean; either *on* or *off* at any time. We require a network to produce a particular time sequence $v(t)$ for $t \in 1,\ldots,T$ and parameterize the complexity of such a genetic function by its duration T. We establish a number of new results regarding how functional complexity constrains genetic regulatory networks and their evolution. We find that the number of networks which generate a function decreases approximately exponentially with its complexity T and show there is a corresponding weakening of the robustness of those networks to mutations. These results suggest a limit on the functional complexity T of typical networks that is polynomial in N. However, we are also able to prove the existence of a, presumably small, class of networks in which this scales *exponentially* with N. We demonstrate that an increase in functional complexity T drives what we describe as a metagraph disintegration effect, breaking up the space of networks previously connected by neutral mutations and contrast this with what is found with less restrictive definitions of functionality. Our findings show how functional complexity could be a factor in shaping the evolutionary landscape and how the evolutionary history of a species constrains its future functionality. Finally we extend our analysis to functions with more exotic topologies in expression space, including "stars" and "trees". We quantify how the properties of networks that give rise to these functions differ from those that produce linear functional paths with the same overall duration T.

1. Introduction

Organisms evolve and adapt to their environment. Neutral evolution allows organisms the freedom to explore their genotype space while still maintaining a specific functional response phenotype (Leigh, 2007). Here we investigate how the structure of the neutral space and the robustness of the genetic regulation are affected by the complexity of that function. We show how increasing functional complexity can lead to a process of metagraph disintegration.

Differential equation models are commonly used to understand genetic networks (Chen et al., 2000; Locke et al., 2005a,b, 2006). We utilise a Boolean approach in which both gene expression levels and time are discretized: Genes are on (1) or off (0) and time is treated as proceeding in discrete steps. We simplify the network interactions so that genes can only either up- or down- regulate other genes or have no effect on them. Despite the abstract nature of this approach it has been used to provide high level models reproducing the qualitative behaviour of the yeast cell cycle (Li et al., 2004; Davidich and Bornholdt, 2008) and the p53-Mdm2 gene circuitry (Ge and Qian, 2009), while a variant of this model has been useful in predicting the mutant phenotypes of *Drosophilia* (Albert and Othmer, 2003). It is intellectually attractive in that it simplifies the state space in a manner that many experimental scientists will find intuitive and already utilise anecdotally. While noise can be incorporated into these models they are otherwise numerically deterministic, unlike nonlinear (chaotic) differential equations where the choice of, e.g. time discretization, can affect the network behaviour at the qualitative level. Boolean Networks have also been evolved for greater robustness to noise (Mihaljev and Drossel, 2009; Szejka and Drossel, 2010; Braunewell and Bornholdt, 2008). Our study investigates the evolutionary constraints imposed by increasing functional complexity without the need to perform explicit evolutionary simulations.

Our model represents a gene regulatory network of N transcriptional regulators which are represented by their gene expression patterns $\underline{v}(t) = (v_1(t), v_2(t), \ldots, v_N(t)) \in \{0,1\}^N$ at some discrete time t during a biological cell process. An interaction matrix (which we also refer to as a network) $A = (a_{ij})$ defines the regulatory interactions between genes. The entry a_{ij} expresses the strength of interaction gene j has on gene i. We restrict ourselves to the case where $a_{ij} \in \{-1,0,1\}$. So interactions either

* Corresponding author at: Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK.
E-mail address: m.s.turner@warwick.ac.uk (M.S. Turner).

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.05.006

inhibit ($a_{ij}=-1$) or promote ($a_{ij}=+1$) gene i, or are absent. Given a state $\underline{v}(t)$ and a network A then the state of the system at the next time step is determined by $\underline{h}(t)=A\underline{v}(t)$. $v_i(t+1)=1$ if $h_i(t)>0$ (turns on), $v_i(t+1)=0$ if $h_i(t)<0$ (turns off) and $v_i(t+1)=v_i(t)$ if $h_i(t)=0$ (retains previous state). This is effectively a consensus model in the presence of multiple regulatory inputs.

We define a biological function for a cell to be a path $\{\underline{v}(t)\}_{t=0}^{T}$ through the gene expression state space. Previous authors (Ciliberti et al., 2007a,b; Martin and Wagner, 2008) have defined networks which map the initial condition $\underline{v}_0$ to the fixed point $\underline{v}_\infty$ as belonging to the same phenotype. We denote this the *unconstrained path duration* (UPD) definition of a function. In essence this definition equivocates different paths which explore different areas of the gene expression space and many with different durations T. We take the view that the UPD definition of a function is not always biologically appropriate. For example let us consider the two gene model of the p53-Mdm2 regulatory network (Ge and Qian, 2009) depicted in Fig. 1A. The p53 gene is activated in response to stress signals and consequently activates Mdm2. Mdm2 then acts to degrade p53. Fig. 1B depicts in Boolean form this key function of the p53-Mdm2 network. Any path differing from the one depicted would fail to replicate the required cell process. Under the UPD definition of functionality any network which mapped the initial state in this trajectory to the final one along any sequence of states would also be considered as functional. However, most of these other trajectories would fail to capture the observed dynamics of the p53-Mdm2 system. The type of analysis we perform here was first implemented on a biological path (Boldhaus and Klemm, 2010) for a yeast cell cycle model (Li et al., 2004) with results that differed significantly from those reported in Ciliberti et al. (2007b).

We also examine the relationship between network topology and the choice of path $\{\underline{v}(t)\}_{t=0}^{T}$ where we characterise a path by its duration T. We consider this value to be a proxy measurement of the complexity of a cellular response. By network topology we mean the structure of those matrices A which define the interaction between genes. Genetic point mutations correspond to changes in matrix elements (Fig. 1C). We investigate how path duration influences robustness and whether a path specific phenotype definition infers different topological results from the UPD definition of functional phenotype.

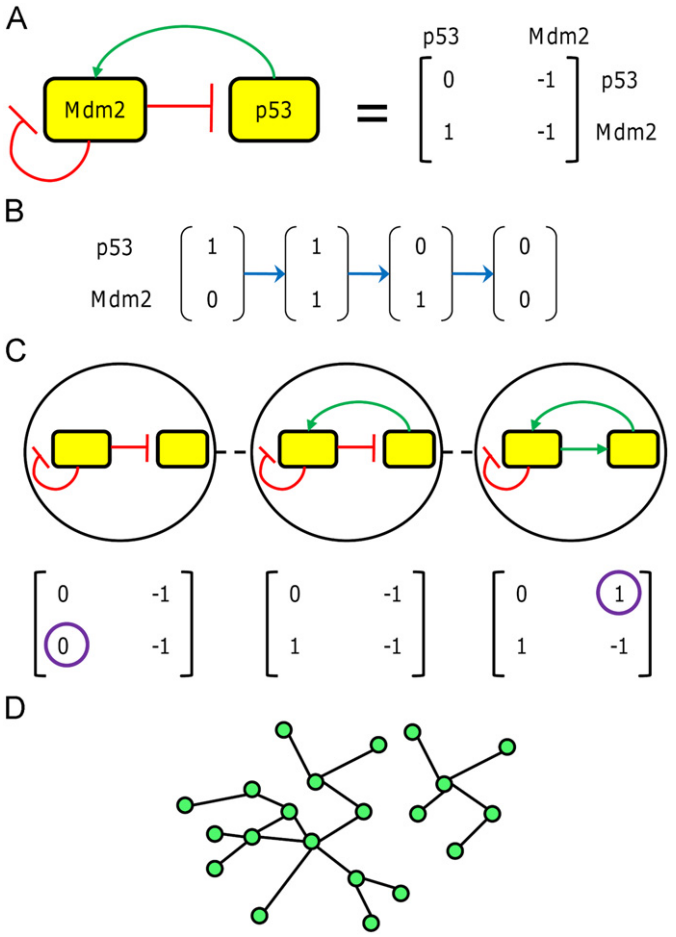

Fig. 1. p53-Mdm2 Gene Regulatory Network Model and a schematic Metagraph. (A) Boolean network model of the p53-Mdm2 gene regulatory network in both network and matrix form. Pointed arrows are excitatory interactions (+1) and the flat headed arrows are inhibitory interactions (−1). (B) An example Boolean biological function representing the p53-Mdm2 network's response when p53 is activated. p53 activates Mdm2 which inhibits p53 before degrading itself. (C) The p53-Mdm2 wild type network (middle) and two 1-mutants (one interaction difference) with their corresponding matrix forms. (D) A schematic of a metagraph where the nodes represent networks and the edges exist between networks that differ in only one interaction.

2. Results

2.1. Functional complexity constrains network topology

Here a biological function is a path $\{\underline{v}(t)\}_{t=0}^{T}$ through the state space $\{0,1\}^N$. Any network $A\in\{N\times N\ matrices|a_{ij}\in\{-1,0,1\}\}$ attains a function $\{\underline{v}(t)\}_{t=0}^{T}$ if A maps $\underline{v}(t)\rightarrow\underline{v}(t+1)$ for all $0\le t\le T-1$. We aim to classify how biologically attainable certain different classes of functions $\{\underline{v}(t)\}_{t=0}^{T}$ are in this space of Boolean networks as we vary the path duration T. Let us define $\{A\}$ to be the set of matrices which attain $\{\underline{v}(t)\}_{t=0}^{T}$ and for $|\{A\}|$ to be the cardinality of this set. For convenience we also use the notation $|A|$ to mean $|\{A\}|$. Attainability is measured by the number of networks that attain a function, $|A|$. We also refer to this value as the number of networks which are functional with respect to the function $\{\underline{v}(t)\}_{t=0}^{T}$. Attainability is dependent on the path duration T and the path generation parameters we set. For genes up to $N=10$ we generated random sample paths using the two methods outlined below and then identified all the networks which attained each sampled path.

Sample paths were generated with two different algorithms. Both start by drawing uniformly at random a $\underline{v}(0)\in\{0,1\}^N/\underline{0}$, where $\underline{0}$ is the zero vector. The first algorithm then generates $\underline{v}(1)$ by flipping each node (gene) with some probability θ which we set (we denote this the θ method). These steps are iterated until a self-avoiding path of length T is produced. Here θ controls the average number of flips in the expression state of a gene between adjacent state vectors in the path and can be thought of as a *speed* in that it controls the distance moved through the expression space per time step. The second method starts identically by drawing a $\underline{v}(0)$ but then randomly samples a matrix $A^0\in\{-1,0,1\}^{N\times N}$ which maps $\underline{v}(0)$ to some $\underline{v}(1)$. Then for each subsequent step in the path another A^t is drawn from the remaining set $\{A\}$, of matrices which attain the path $\{\underline{v}(0),\ldots,\underline{v}(t)\}$, to generate a $\underline{v}(t+1)$ which is again self avoiding. We denote this the A^t sampling method. This method tends to generate longer paths on average than the θ path generation method.

Using the θ path generation method we sampled paths, extending their length T until $|A|=0$. For the A^t sampling method, paths were extended by sampling upto a 1000 matrices that would extend the path for each t in a self-avoiding manner. If none of the 1000 sampled matrices successfully extended the path then the algorithm would terminate. For both path generation methods the mean number of matrices which attain a path $\{\underline{v}(t)\}_{t=0}^{T}$, $|A|$

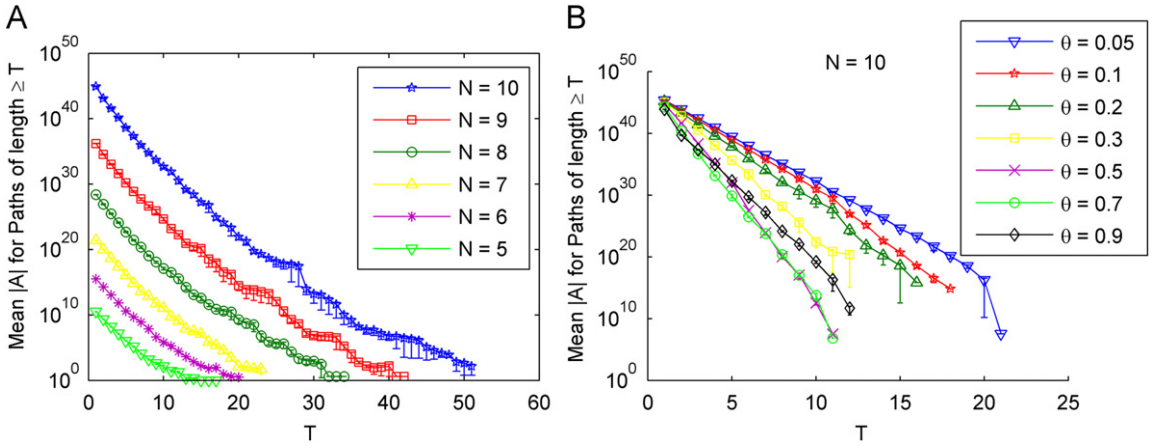

Fig. 2. An exponential decrease in the number of available networks with increasing duration. On both plots the *x*-axis shows the duration T of the path and the *y*-axis the mean number of networks $|A|$ which attain a path of duration T. In both plots $|A|$ decreases approximately exponentially with T. Note that for both path generation methods, as T increases fewer samples are available to compute $|A|$. (A) 1000 sample paths were generated using the A^t sampling method for every $5 \leq N \leq 10$. (B) 1000 sample paths were generated using the θ sampling method for $N=10$ and at various speeds θ. The error bars in both plots are a single standard deviation over $\sqrt{n}$ where n is the number of paths of that reached that length.

decreases roughly exponentially with increasing T. Fig. 2A shows the decline in mean $|A|$ under the A^t method for various N, whereas Fig. 2B is a representative example, with $N=10$, of the decrease in mean $|A|$ for different speeds θ. Hence we find that, as a function becomes more complex, the number of network topologies which can attain the function typically decreases exponentially with T. We suggest later that evolutionary considerations may mean that the number of topologies available to a single species could be even more strongly constrained. It should be noted that increasing T makes paths more difficult to attain under both methods and so there are fewer samples of paths the longer they are.

The effect of changing the speed θ is non-trivial and does not relate monotonically to $|A|$. For instance $|A|$ is higher for $\theta = 0.9$ for $T > 5$ than both $\theta = 0.5$ and 0.7 (Fig. 2B). Each θ in the first method corresponds to a different set of (random) biological functions from which we sample. There is a non-monotonic relationship between θ and the durations at which functions become unattainable, which we call T_{end} (Fig. S1). This figure also indicates that the mean attainable functional duration $\langle T_{end} \rangle$ scales approximately linearly with N. Under our second pathgeneration method $\langle T_{end} \rangle \sim N^2$ (Fig. S2). These results suggest that the genome size N corresponds to no more than $O(N)$ or $O(N^2)$ gain in attainable functional complexity for the types of random functions studied here. Furthermore, it suggests that functions with durations $O(e^N)$ should be vanishingly rare. However, we are able to construct networks which possess a path whose duration $T_{end} \sim 2^{N/6}$ (see Text S1). This is done by constructing a network that counts in base 2, using $N/6$ digits. Thus a small proportion of networks are capable of supporting anomalously complex functions.

2.2. Metagraph disintegration

For any two networks $A,B \in \{-1, 0, 1\}^{N \times N}$, we can define a distance between them as the sum of the number of the interactions which differ, $d(A,B) = \sum_{i,j} 1 - \delta_{a_{ij} b_{ij}}$ where δ_{ab} is the Kronecker delta function. Thus if $d(A,B)=1$ then the two networks differ in only one interaction and a single point mutation can change one network into the other (Fig. 1C). We define a metagraph the nodes of which represent networks and where the edges connect networks which are exactly distance one apart (Fig. 1D). Thus any connected component of the metagraph M can be traversed by point mutations (i.e. single entry changes).

In Ciliberti et al. (2007a) it was found that the vast majority of networks that take $\underline{v}_0$ to $\underline{v}_\infty$ are on one connected metagraph component. This lead to the suggestions (i) that through gradual topological changes (point mutations) any viable network could evolve towards greater robustness (Ciliberti et al., 2007a) and (ii) that gene circuits could innovate through access to many other phenotypes by traversing long distances across the large dominant connected metagraph component (Ciliberti et al., 2007b). However, the metagraph of networks corresponding to the yeast cell cycle's functional path was discovered to be very disconnected (Boldhaus and Klemm, 2010) where it was hypothesised that this may have been due to the imposition of a greater number of constraints than the UPD functions defined in Ciliberti et al. (2007b).

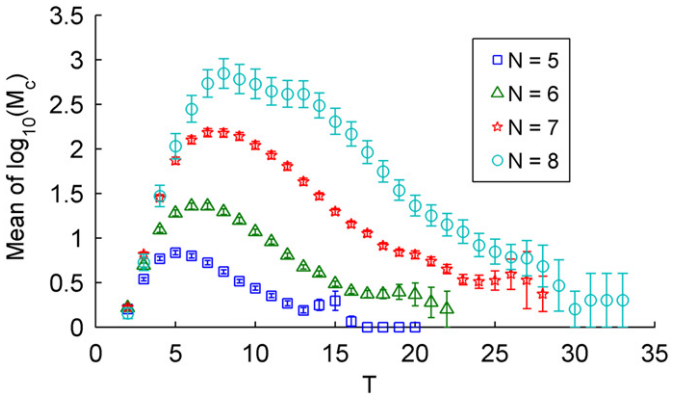

Fig. 3. The number of metagraph components depends on the path duration. The *x*-axis shows the path duration T. The *y*-axis is the mean of the log_{10} of the number of connected components $log_{10}(M_c)$ for 1000 (100 for $N=8$) sampled paths using the A^t sampling method. The initial increase in the number of connected components with increasing T creates a disconnected (neutral) evolutionary landscape. The decrease in the number of components as T increases further occurs as some components can no longer cope with the demanded functionality. Thus the evolutionary choices made previously constrain a species'future functionality. The error bars are a single standard deviation over $\sqrt{n}$ where n is the number of paths of that reached that duration T.

To investigate the effect of path complexity on the metagraph we generated paths of varying durations and found that the number of metagraph components connected by neutral mutations is dependent on the path duration T (Fig. 3). As T increases so does the number of connected components. Indeed, a distinct pattern emerges in which the number of connected components first increases before decreasing as T increases further. The metagraph literally disintegrates. The connected metagraph

components are typically of comparable size so no single dominant component exists but the number of components can vary over many orders of magnitude as the duration of a function increases. Organisms that end up on these different metagraph components have experienced what we choose to describe as *topological speciation*, in relation to their network interactions. Even though they are all producing the same functional response their networks cannot interchange via neutral mutations alone.

For all paths where $T=1$ we offer a simple proof that there can only be one connected component of the metagraph. Let the path be $\{\underline{v}(0),\underline{v}(1)\}$ and $\{A_i\}$ be the set of all row solutions for row i (node i). Consider the case when $\underline{v}_i(0)=0$ and $\underline{v}_i(1)=1$. This constraint implies that for any row, $\underline{r}\in A_i$ the value of $h_i(0)=\sum_{k=1}^{N} r_k\ v_k(0)>0$. Note that the row of all ones $\underline{1}$ always satisfies this constraint. Furthermore for any $\underline{r}\in A_i$ one can through point mutations map $\underline{r}$ to $\underline{1}$ by changing each -1 and 0 entry into 1. Each intermediate row will satisfy the constraint and so every row in A_i can reach every other row via point mutations. Similarly for the case $\underline{v}_i(0)=1$ and $\underline{v}_i(1)=0$ all rows are connected via $-\underline{1}$. It is clear that for the cases $\underline{v}_i(0)=0$, $\underline{v}_i(1)=0$ and $\underline{v}_i(0)=1$, $\underline{v}_i(1)=1$ the same arguments hold. Therefore, for any two matrices which attain a path $\{\underline{v}(0),\underline{v}(1)\}$, their rows can be mapped onto each other's through point mutations and thus they lie on the same connected component. Therefore, the metagraph is connected for any path of the type $\{\underline{v}(0),\underline{v}(1)\}$. A corollary to this result is that the metagraph is connected for any path of the type $\{\underline{v}(0),\underline{v}(1),\underline{v}(1)\}$ where we have simply specified $\underline{v}(1)$ to be a fixed point.

This simple result goes someway to explaining the differences between our metagraph disintegration result (the metagraph typically has many connected components) and the large connected component found in Ciliberti et al. (2007b) using UPD. Here UPD includes the ensemble of all paths (of all durations T) from $\underline{v}(0)$ to $\underline{v}_\infty$. As demonstrated above the path $\{\underline{v}(0),\underline{v}(1),\underline{v}(1)\}$ where $\underline{v}_1=\underline{v}_\infty$ produces a connected metagraph for our path approach. It will also be large relative to the number of networks for longer duration paths. Thus any connected components corresponding to longer paths which are within a distance one of the $\{\underline{v}(0),\underline{v}(\infty),\underline{v}(\infty)\}$ metagraph are then connected to each other. We hypothesise that the connected metagraph for the $\{\underline{v}(0),\underline{v}(\infty),\underline{v}(\infty)\}$ path may form a backbone to connect most of the components from other paths. Small, disconnected components would then necessarily correspond only to longer duration paths.

2.3. A trade-off between functional complexity and mutational robustness

The mutational robustness, R_M of a network A with respect to $\{\underline{v}(t)\}_{t=0}^{T}$ is defined as the proportion of 1-mutants of A which also attain the function $\{\underline{v}(t)\}_{t=0}^{T}$. This is equivalent to the degree of A in the relevant metagraph M of $\{\underline{v}(t)\}_{t=0}^{T}$. It was reported in Ciliberti et al. (2007a) that a broad distribution of mutational robustness exists in networks sampled from any of their metagraphs. However, as different networks are utilising different paths of different duration then a duration dependency in R_M could explain the broad distribution. As Fig. 4 demonstrates there is a strong negative correlation between T and R_M. So we can see that when duration T paths are taken to determine the metagraphs rather than the UPD approach we see an explicit negative correlation between a network's mutational robustness and path duration. We conclude that in this model a trade-off typically exists between mutational robustness and functional complexity.

2.4. The correlation between mutational and noise robustness

A small amount of noise can be represented by the random flipping of a node's state. A network can be considered robust if the perturbation is corrected—the desired path is immediately recovered, despite the perturbation from an initial condition. We measure the robustness to noise of a network A which attains a path $\{\underline{v}(t)\}_{t=0}^{T}$ by the proportion of one bit flip perturbations from $\underline{v}(0)$ which recover the remaining path under A. So the noise robustness, R_n, is the fraction of the N neighbouring states of $\underline{v}(0)$ that are mapped by A to $\underline{v}(0)$ or $\underline{v}(1)$.

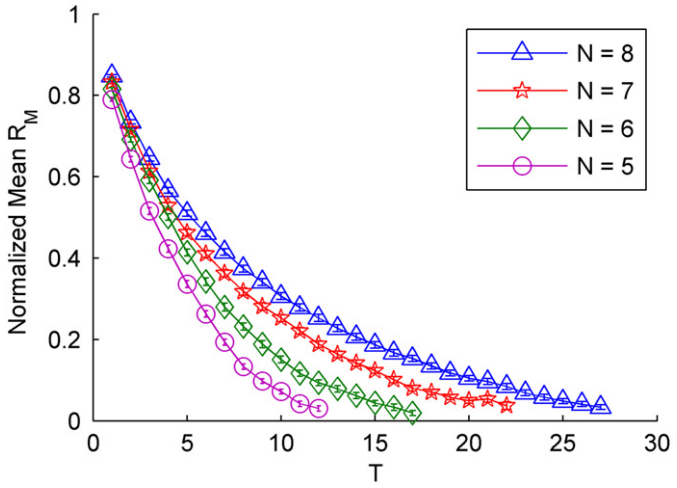

Fig. 4. Mutational Robustness and path length are negatively correlated. The x-axis shows the path duration T. The y-axis is the mean degree from 100 sampled paths normalised by the maximum degree of $2N^2$ using the A^t sampling method. For each path at each length T 100 networks were sampled to estimate the mean degree for that path. The error bars are a single standard deviation over $\sqrt{n}$ where n is the number of paths of that reached that duration T.

It has been reported in Ciliberti et al. (2007a) that a strong correlation exists between noise and mutational robustness. This correlation was also reported in evolutionary experiments on Boolean networks (Mihaljev and Drossel, 2009). However, we find that the strength of this correlation decreases with increasing T (Fig. S3). As the UPD approach allows an ensemble of paths of different duration it is possible that sampling favours shorter trajectories, which have a large number of solutions. This could dominate the sampling process and explain the stronger correlation reported in Mihaljev and Drossel (2009); Ciliberti et al. (2007a).

2.5. Multi-functionality

Up to this point we have restricted our analysis to paths of the form $\{\underline{v}(t)\}_{t=0}^{T}$. It is natural to ask what are the properties of functions, which are not single paths? We investigated the effect of requiring multiple paths/functions to be simultaneously satisfied and the effect of creating star and tree-like functional paths as depicted in Fig. 5. The star multi-functional form will naturally increase its noise robustness and the tree form is conceptually somewhere between the star and single path functions. Formally a multi-functional path of m functions has the form $\{\{\underline{v}^k(t)\}_{t=0}^{T_k}\}_{k=1}^{m}$. In this section the duration T of a multi-function is defined to be $\sum_{k=1}^{m} T_k$. It is worth reporting that for disjoint multi-functions (Fig. 5B) all the results are nearly identical to those reported in relation to single paths (Fig. S4). However, this is not the case for the tree and star-like functions.

Star paths of the form $\{\{\underline{v}^k(0),\underline{v}^*\}\}_{k=1}^{m}$ where $\underline{v}^*$ is the same point in state space for each k (Fig. 5D) were generated at random under different θ. Paths of this form are always attainable and only ever have one connected metagraph component, which follows similarly from our proof for single paths of duration $T=1$. The mean total number of matrices which attain a path

decrease with T (Fig. 6A) but at a much slower rate of decline than was found for single paths and multiple disjoint paths. Furthermore, mutational robustness was still found to decrease as a function of T, although again more slowly (Fig. 6C), whereas noise robustness increases consistently towards one (Fig. 6D). Interestingly the correlation between mutational and noise robustness remains much higher on average, at around 0.5 (Fig. 6B).

The tree-like paths are formed using the θ method by fixing a branching number b and then forming a star with b branches into that point in the state space. Then one of the tips of this star is taken (a state with no incoming states) and b new branches are generated for that tip and so on creating multi-functional paths like those depicted in Fig. 5C. Here the branching number has a curious effect on all the quantities of interest. In effect one sees a step like effect imposed on previous results. The number of matrices which attain a path $|A|$ and the mutational robustness M_d both still decrease as a function of T. However, a b duration wait is introduced where these values fall at a lower rate as each new star is formed before taking a larger decrease as a new star is started (Fig. 6A and C). The correlation between noise and mutational robustness decreases in a similar fashion (Fig. 6B).

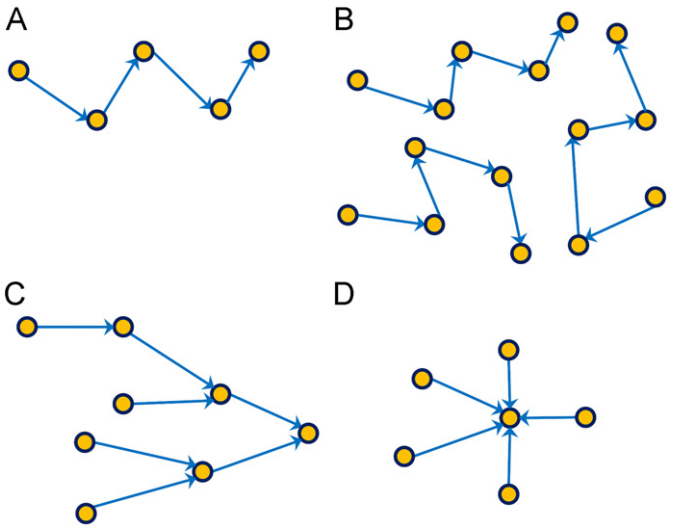

Fig. 5. Examples of multifunctions. These are state-space directed graphs of the different types of multi-functions studied. Nodes represent gene expression states and the arrows indicate which state is mapped to which. (A) Single function. (B) Multi-function with three disjoint single paths. (C) Tree multi-function with a branching number of two. (D) Star multi-function with five branches.

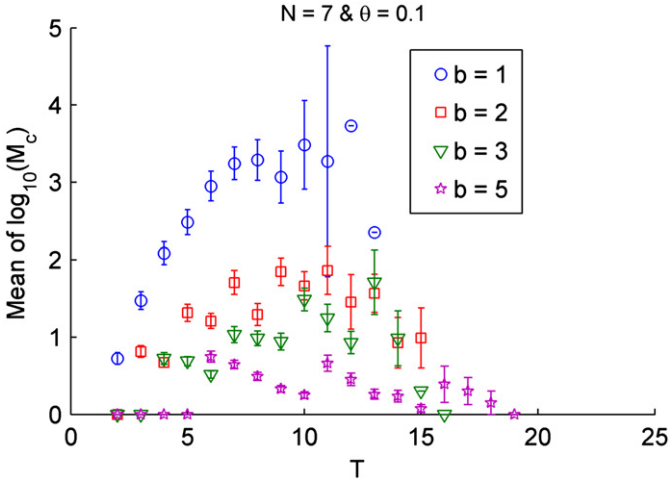

Fig. 7. The connectedness of the metagraph for tree-like multi-functions. The x-axis shows the path duration T. The y-axis is the mean of the log_{10} of the number of connected components, $log_{10}(M_c)$ for 1000 sampled tree-like paths at different branching numbers ($b=1$, 2, 3 and 4). $\theta=0.1$ was used for the path sampling method. The error bars are a single standard deviation over $\sqrt{n}$ where n is the number of paths of that reached that duration T. A step like pattern emerges with the number of components increasing when a branch is extended and then decreasing as more branches are added making the multi-function more star like.

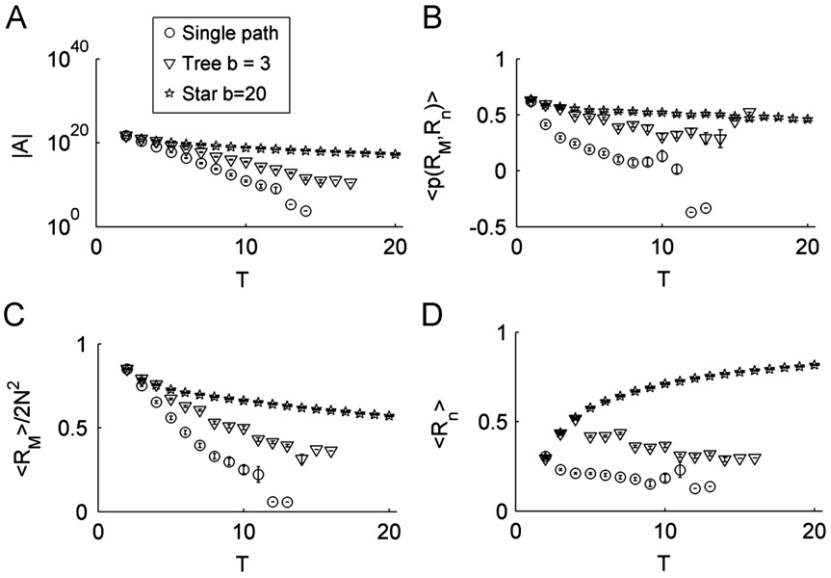

Fig. 6. Single path, star and tree-like multi-functions. In all the plots the x-axis shows the path duration T. 100 single paths, star-like paths and tree-like paths with $b=3$ were generated for $N=7$ and $\theta=0.1$. The error bars are a single standard deviation over $\sqrt{n}$ where n is the number of paths of that reached that duration T. (A) The mean $|A|$ (y-axis) declines exponentially for the single path but much more like a power law for the star multi-function. The tree falls somewhere in between with a general exponential fall which is slightly shallower than exponential as each point fills its branches. (B) The mean correlation between mutational robustness and noise was estimated from 100 sampled matrices for each function. The star retains a higher correlation than the single path and again the tree falls in between with a reduction in correlation as each unextended branch tip is extended. (C) Estimated from 100 sample matrices per function, the normalised mean mutational robustness $\langle R_M \rangle/2N^2$ falls in all cases but most slowly for the star. The tree function falls at two rates depending on whether a new tip is being extended. (D) Estimated from 100 sample matrices per function, noise robustness increases towards one for the star while falling for the single path. The tree sees a fall when a tip is extended and then a rise as more branches are added to that tip.

The noise robustness R_n decreases as each new tip is extended but recovers somewhat as more branches are added (Fig. 6D). A similar effect occurs in relation to the number of metagraph connected components as is shown in Fig. 7. These results suggest that certain types of multi-functionality, particularly those which are star like, can decrease the breakup of the available neutral space and that biological functions of this type might exhibit a much weaker tendency to speciate.

3. Discussion

We have identified the duration T of a function, or equivalently the total sum duration for a multi-function as the key factor which influences its attainability, the connectedness of its metagraph and the robustness of those networks which attain the function. These results highlight the crucial differences that arise when one adopts a path definition of functionality as opposed to UPD, which fixes only its start and end points. We believe that in many cases, particularly for genetic functions involving timing (circadian networks, cell cycle etc), our definition of functionality is most appropriate.

The connectedness of the metagraph may play an important role in the evolution of biological function. Consider, for example, populations of genetically identical organisms exposed to constant environmental stress over many generations. If UPD is adequate to define a biologically fit function the separate populations will be free to traverse more of the network space through neutral moves. Different genetic populations will arise but they will still be able to mutate (back) into each other. However, if biological fitness requires fixed duration functions the metagraph is more likely to become disconnected. Then different populations are likely to find themselves on separate islands of the metagraph after evolving the necessary function. All further mutational evolution is now constrained by the genetic structures adopted in each population. This topological speciation effect is irreversible under neutral mutations and dictates that, as more functions are attained, different population strains will be unable to access all the same future functions. This is evident from the decrease in the number of connected components after some peak (Fig. 3). Thus future innovation is topologically constrained by our evolutionary past.

It should be noted that we examine only the metagraph of neutral point mutations. One benefit of more aggressive mechanisms for genetic alteration, such as sex, would be to overcome historically imposed genetic constraints that may isolate an individual from a fitter metagraph component. We also note that we have only examined the case where interactions between genes have discrete weights (i.e. $a_{ij} \in \{-1,0,1\}$). Allowing these weights to be real or arbitrarily finely discrete would certainly affect our results in relation to metagraph connectivity. Indeed it may be that the metagraph will typically become connected in this case.

We have shown that the number of available topologies decreases exponentially as the functional complexity increases. Furthermore the mutational robustness will also typically decrease. These two features suggest that for any fixed number of genes a limit exists to the system's attainable functional complexity, and that exceeding this limit may prove unstable due to the increased mutational vulnerability of any specific networks. We did construct networks with anomalously ($O(2^{(N/6)})$) long functions and, consistent with the above trend, these are fragile to any mutation.

A form of multi-functionality, under the UPD definition, (Martin and Wagner, 2008) has been proposed elsewhere. Here pairs of disconnected start-endpoint functions were simultaneously combined to produce a single multi- (bi-) function. Because UPD is still employed this is still somewhat different to a similar fixed duration function, which might correspond to a function involving a pair of disconnected paths, each with duration $T_k=1$. As discussed above the properties of these disjoint functions were very similar to those involving a single connected path with the same total duration. Indeed, the reduction in mutational robustness in going from mono to bi-functions reported in Martin and Wagner (2008) can be understood in the context of the present work by the fact that the minimum duration T increases from 1 to 2.

We find that star and tree-like functions have markedly different properties. Star-like functions typically have higher mutational robustness. We see also that as the branching number increases we begin to lose the metagraph disintegration effect present for single functional paths. This implies that the exact structure of a biological function influences the accessible evolutionary trajectories. It is impractical to analyse every possible kind of functional path structure, those we have investigated form a very selective subset. However, microarray data has been used to define the key biological function of the yeast cell cycle (Li et al., 2004).

Our findings have implications for how functional responses might be studied in biology. We have shown that different functional topologies have different robustness and evolvability. Therefore, in experiments where these properties are crucial we suggest careful examination of the function's topology.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.jtbi.2011.05.006

References

Albert, R., Othmer, H.G., 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. Journal of Theoretical Biology 223, 1–18.

Boldhaus, G., Klemm, K., 2010. Regulatory networks and connected components of the neutral space*. European Physical Journal B — Condensed Matter and Complex Systems 77 (2), 233–237 doi:10.1140/epjb/e2010-00176-4.

Braunewell, S., Bornholdt, S., 2008. Reliability of genetic networks is evolvable. Physical Review E 77, 060902.

Chen, K.C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., Tyson, J.J., 2000. Kinetic analysis of a molecular model of the budding yeast cell cycle. Molecular Biology of the Cell 11, 369–391.

Ciliberti, S., Martin, O.C., Wagner, A., 2007a. Innovation and robustness in complex regulatory gene networks. Proceedings of the National Academy of Sciences USA 104, 13591–13596.

Ciliberti, S., Martin, O.C., Wagner, A., 2007b. Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Computational Biology 3 (2), e15.

Davidich, M.I., Bornholdt, S., 2008. Boolean network model predicts cell cycle sequence of fission yeast. PloS One 3, e1672.

Ge, H., Qian, M., 2009. Boolean network approach to negative feedback loops of the p53 pathways: synchronized dynamics and stochastic limit cycles. Journal of Computational Biology 16, 119–132.

Leigh, E.G., 2007. Neutral theory: a historical perspective. Journal of Evolutionary Biology 20, 2075–2091.

Li, F., Long, T., Lu, Y., Ouyang, Q., Tang, C., 2004. The yeast cell-cycle network is robustly designed. Proceedings of the National Academy of Sciences USA 101, 4781–4786.

Locke, J.C.W., Kozma-Bognár, L., Gould, P.D., Fehér, B., Kevei, E., Nagy, F., Turner, M.S., Hall, A., Millar, A.J., 2006. Experimental validation of a predicted feedback loop in the multi-oscillator clock of arabidopsis thaliana. Molecular Systems Biology 2, 59.

Locke, J.C.W., Millar, A.J., Turner, M.S., 2005a. Modelling genetic networks with noisy and varied experimental data: the circadian clock in arabidopsis thaliana. Journal of Theoretical Biology 234, 383–393.

Locke, J.C.W., Southern, M.M., Kozma-Bognár, L., Hibberd, V., Brown, P.E., Turner, M.S., Millar, A.J., 2005b. Extension of a genetic network model by iterative experimentation and mathematical analysis. Molecular Systems Biology 1, 2005.0013.

Martin, O.C., Wagner, A., 2008. Multifunctionality and robustness trade-offs in model genetic circuits. Biophysical Journal 94, 2927–2937.

Mihaljev, T., Drossel, B., 2009. Evolution of a population of random boolean networks. European Physical Journal B 67, 259–267.

Szejka, A., Drossel, B., 2010. Evolution of boolean networks under selection for a robust response to external inputs yields an extensive neutral space. Physical Review E 81 (2), 021908(9).