Bioinformatics, 37(4), 2021, 551–558
doi: 10.1093/bioinformatics/btaa831
Advance Access Publication Date: 11 December 2020
Original Paper

Systems biology

Neighbourhoods in the yeast regulatory network in different physiological states

Arthur M. Lesk[1,*] and Arun S. Konagurthu[2]

[1]Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, The Pennsylvania State University, University Park, PA 16802, USA and [2]Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia

*To whom correspondence should be addressed.

Associate Editor: Jinbo Xu

Received on May 21, 2020; revised on August 23, 2020; editorial decision on September 7, 2020; accepted on September 10, 2020

Abstract

Motivation: The gene expression regulatory network in yeast controls the selective implementation of the information contained in the genome sequence. We seek to understand how, in different physiological states, the network reconfigures itself to produce a different proteome.

Results: This article analyses this reconfiguration, focussing on changes in the *local* structure of the network. In particular, we define, extract and compare the 1-neighbourhoods of each transcription factor, where a 1-neighbourhood of a node in a network is the minimal subgraph of the network containing all nodes connected to the central node by an edge. We report the similarities and differences in the topologies and connectivities of these neighbourhoods in five physiological states for which data are available: cell cycle, DNA damage, stress response, diauxic shift and sporulation. Based on our analysis, it seems apt to regard the components of the regulatory network as 'software', and the responses to changes in state, 'reprogramming'.

Contact: aml25@psu.edu

1 Introduction

Regulatory networks organize activities within cells, and their interactions with their surroundings. Targets of regulation include the steps of the 'central dogma' and beyond (Fig. 1).

Here, we focus on the gene expression regulatory network, which controls transcription. The genome encodes enzymes that catalyse reactions of the metabolic pathways, and proteins that regulate expression. To a large extent, the network of metabolic pathways has a stable geometry ('hardware'), but regulatory networks are capable of reconfiguration, in response to changes in conditions internal and external ('software'). In different physiological states, cells 'reprogram' the regulators of transcription to select a different set of enzymes, or to vary the amounts of transcription of these components.

For yeast, datasets are available specifying the transcription control network in five states: cell cycle, DNA damage, stress response, diauxic shift and sporulation (Luscombe *et al.*, 2004). Each network is a directed graph, the dataset consisting of a list of pairs of nodes connected by an edge in the graph. Parent nodes in any edge in any of the networks are *transcription regulators*. Some nodes are *not* parent to any other node in any of the networks; they are regulated but do not regulate other genes. These are *target genes*.

Although these data represent a major component of the transcription regulation network, they are not a complete statement of the molecules that control protein expression in yeast (see Fig. 1): the datasets were limited to promoter-binding transcription factors. In particular, they do not include transcription control by RNA molecules (Morris, 2011), much less miRNAs and other non-coding RNAs that act post-transcriptionally. Nevertheless, we feel that these data support our contribution to an understanding of general principles of the structure and variability of regulatory networks, and that our results will retain interest even when additional data allow for extension of the conclusions.

Altogether the networks described in this work contain 1475 genes, corresponding to approximately half the known proteome of *Saccharomyces cerevisiae*. Of these genes, 113 encode transcription regulators and 1362 encode target genes.

Previous analysis of these networks (Luscombe *et al.*, 2004) describes general statistics of the topologies of the graphs, including:

- The distribution of the number of incoming connections to target genes has a mean value of 2.1, and is distributed exponentially. Most target genes receive direct input from about two transcriptional regulators. The probability that a gene is controlled by k

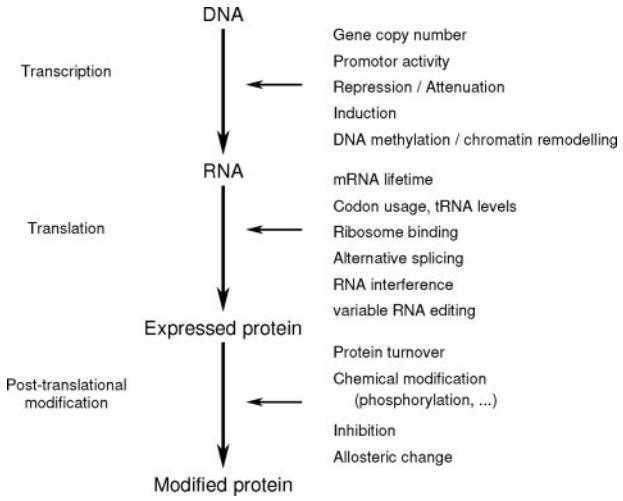


Fig. 1. Starting with DNA, cells carry out transcription, translation and modification and control of activity of proteins. Each of these steps is equipped with several types of regulatory mechanisms, that keep the overall activity organized and under control

transcription regulators, $k = 1, 2, \ldots,$ is proportional to $e^{-\alpha k}$, with $\alpha = 0.8$.

- The distribution of the number of outgoing connections has a mean value of 49.8, and obeys a power law. The probability that a given transcriptional regulator controls k genes is proportional to $k^{-\beta}$, with $\beta = 0.6$. Power-law behaviour characterizes topologies in which a few nodes—the 'hubs'—have many connections, and many nodes have few. In regulatory networks, hubs tend to be fairly far upstream, forming important foci of regulation with far-reaching control.
- The average number of intermediate nodes in a minimal path between a transcriptional regulator and a target gene is 4.7. The maximal number of intermediate nodes in the shortest path between two nodes is 12.
- The clustering coefficient of a node is a measure of the degree of local connectivity within a network. The mean clustering coefficient, averaged over all nodes, is a measure of the overall density of the network. For the yeast transcriptional regulatory network, the mean clustering coefficient is 0.11.

Previous analysis of the local structure of the networks has made use of the canonical 'motifs' described by Milo *et al.* (2002). These are the Single-Input Motif (SIM), Multiple-Input Motif (MIM) and Feed-Forward Loop (FFL) (Fig. 2). Luscombe *et al.* (2004) and Konagurthu and Lesk (2008a, b) have studied the distributions of these canonical motifs in the different states of the yeast regulatory network. Other studies of the networks have focussed on the global structure (Balaji *et al.*, 2006; Yu and Gerstein, 2006).

We are interested in studying and comparing the local structure of the networks at levels higher than the individual motifs. In the language of protein structure, if the motifs of Milo *et al.* (2002) are the secondary structure of the networks, what are the supersecondary and tertiary structures?

To address these questions, we define a *1-neighbourhood* of any node in any of the networks. A 1-neighbourhood of any node in one of the networks is a subgraph of the network consisting of the selected node as the central node, all parent nodes from the network connected directly—i.e. by a single edge—to the central node, all child nodes connected directly to the central node, plus any additional edges from the overall network that connect any pair of nodes from the set connected directly to the central node. (Technically: the vertex-induced subgraph involving the selected node, its parents and its children.) Clusters of nodes containing

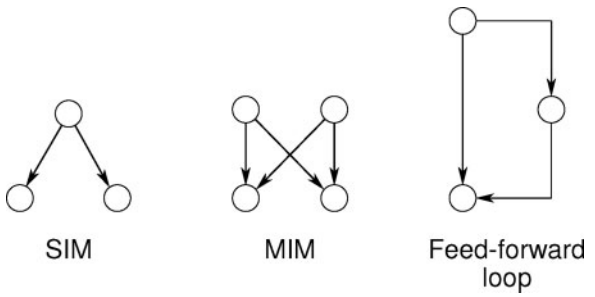


Fig. 2. The *SIM*, or fork, transmits a single incoming signal to two outputs. Successive forks, or forks with higher branching degrees, are an effective way to activate large sets of genes from a single impulse. Generalizations of the binary fork include more child genes under common control (more 'tines' to the fork). Forks can achieve general mobilization. The *MIM* can function as a logical 'or' operation: both child targets become active if *either* of the input impulses is active. (A real-life example is a fire-alarm button in each room of a building such that activation of any one produces a signal in every room to evacuate the building.) Generalizations of the square scatter pattern shown may contain different numbers of nodes on both layers. Note that MIMs are superpositions of SIMs. The *FFL* affects the output both directly through the vertical link; and indirectly and subsequently, through the intermediate link. [This motif can show interesting temporal behaviour if activation of the target requires simultaneous input from both direct and indirect paths (logical 'and'). Because buildup of the intermediate requires time, the direct signal will arrive before the indirect one. Therefore, a short pulsed input to the complex will not activate the output—by the time the indirect signal builds up, the direct signal is no longer active. The system can thereby filter out transient stimuli in noisy inputs.]

pathways of length >1 to a central node would naturally define higher-order neighbourhoods.

Here, we describe the similarities and differences in the structures and topologies of the 1-neighbourhoods of individual nodes in different states.

2 Materials and methods

We used the transcription regulatory networks of *S.cerevisiae* under various physiological conditions published by Luscombe and coworkers: http://networks.gersteinlab.org/regulation/dynamics.

Luscombe *et al.* (2004) assembled these datasets from the results of genetic, biochemical and ChIP (chromatin immunoprecipitation-chip) experiments. They demonstrated the statistical significance of differences reported between two (or more) sub-networks.

We wrote all programmes used.

3 Results

3.1 There is substantial lack of overlap between networks

The nodes are divided into transcription factors and target genes. A transcription factor is any node that appears as the parent node in some edge in any network. Transcription factors can also appear as child nodes in edges in some or all networks. A target gene is a node that appears *only* as a child node in any network.

There is a total of 1475 unique nodes in the union of all five networks. Of these 113 are transcription factors (nodes connected to another child node in one or more of the networks). There is a total of 2479 unique edges in the union of all five networks.

3.2 Differential usage of parent nodes

Table 1 shows the usage of transcription factors in different networks. It reports, for each network, the number of transcription factors that appear as a parent node in an edge, the number that appear as a child node in an edge and the number that appear as both parent and child nodes in different edges.

There are no edges in any of the network in which the two nodes connected by the edge are identical; i.e. in no network is there an edge in which the parent and child nodes are the same, technically: no self-loops. In only two cases, in all the networks, is there a

Table 1. Transcription factor usage in different networks

	Cell cycle	DNA damage	Stress response	Diauxic shift	Sporulation
Parent	70	72	63	71	74
Child	54	44	33	44	46
Both	54	35	33	36	45

reciprocal interaction between two nodes; i.e. a cycle of length 2. The stress response network contains:

YDR259C	→	YOR028C	YDR259C	→	YPR065W
YOR028C	→	YDR259C	YPR065W	→	YDR259C

Each network uses ∼60% of the 113 total transcription factors. Each network uses more transcription factors as parent nodes than as child nodes. This implies that numerous transcription factors are the initiators of paths through the network (not being preceded by any parent node; technically: source nodes for the whole network).

In the cell cycle and stress response networks, all transcription factors that appear as the child nodes of edges also appear as the parent nodes in different edges.

In the DNA damage, diauxic shift and sporulation space networks, some nodes appear as child nodes only (see Table 2). Were we to analyse only those networks, these nodes would be considered target genes and not transcription factors. But we have defined a transcription factor as a node that appears as a parent node in *any* network. The reason for this is that, as a structure, we should expect the molecule corresponding to such a node to interact with DNA to control transcription, and not to be a metabolic enzyme. Under this hypothesis, some of the control pathways have 'dead ends' in which the expression of a transcription factor is under control, but, in those networks, such transcription factors do not control either other transcription factors, or target genes.

There is also the possibility that some of these molecules have dual functions—both interacting with DNA to control transcription, and *also* carrying out some 'housekeeping' enzymatic activity. Such a molecule could then truly be a transcription factor in one network and a target gene in another. Many people would consider this to be a remote possibility, but the ligase/repressor BirA of *Escherichia coli* is one of the several known examples of such dual functionality. The N-terminal domain of BirA can bind to the biotin operator site and inhibit transcription of the operon; the C-terminal domain can function catalytically as a biotin–[acetyl-CoA-carboxylase] ligase.

Table 2. Transcription factors appearing as child nodes only in different networks

Network	Number of nodes	List of nodes
Cell cycle	0	
Diauxic shift	8	YCR065W YHR084W YIR023W YKL185W YLR403W YNL216W YOR358W YPR104C
DNA damage	9	YDR043C YDR146C YDR451C YDR501W YKL109W YKL185W YML027W YNL216W YOR032C
Sporulation	1	YDL170W
Stress response	0	

Table 3. Number of shared parent nodes

	Diauxic shift	DNA damage	Sporulation	Stress response
Cell cycle	45	49	53	37
Diauxic shift		59	54	45
DNA damage			55	42
Sporulation				45

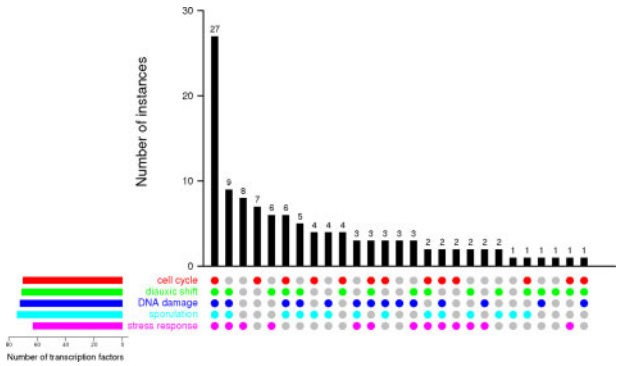


Fig. 3. Frequencies of observed combinations of transcription factor usages, including cases of transcription factors appearing in only a single network. Only 27 of the 31 possible combinations are observed. The height of the bar in the histogram shows the number of transcription factors appearing in the combination of networks shown in the circles below the bar

Table 4. Distribution of edges among networks

Total number of edges

Cell cycle	Diauxic shift	DNA damage	Sporulation	Stress response
550	1217	1082	481	566

Number of common edges in pairs of networks

	Diauxic shift	DNA damage	Sporulation	Stress response
Cell cycle	136	166	127	45
Diauxic shift		476	154	353
DNA damage			174	258
Sporulation				54

Number of shared edges among different numbers of networks

Number of networks, *n*	1	2	3	4	5
Number of edges appearing in *n* networks	1476	678	259	43	23

Are the different networks using largely the same subset of transcription factors, or are they 'mixing and matching'? Table 3 shows the number of transcription factors shared between each pair of networks. Figure 3 shows the distribution of combinations, using the representation introduced as UpSet (Lex *et al.*, 2014). A total of 27 transcription factors appear in all five networks, as parent nodes of edges. This is approximately half of the total number of transcription factors appearing as parent nodes of edges, in each of the networks; and approximately a quarter of the total of 113 transcription factors. Conversely, only 20 transcription factors are unique to one of the networks; most are shared.

3.3 Differential usage of edges

There is a total of 2479 unique edges in the union of the five networks. Table 4 shows the total number of edges in the individual networks, and the number of edges shared among them. There is considerable variation in the total number of edges: it varies from 481 for sporulation to 1217 for diauxic shift. Different networks select different sets: sporulation uses ∼20% of the total observed edges; diauxic shift uses ∼50%. With respect to choices of edges as for choices of transcription factors, the networks are 'mixing and matching'. Only 23 ($< 1\%$) appear in all five networks, and only 325 edges (∼13%) appear in three or more of the networks. Of the

481 edges appearing in the sporulation network, only 174 are shared with a second network, DNA damage.

Contrast the observations that (i) only 20 transcription factors out of 113 (18%) are unique to a single network, but (ii) 1476 edges out of 2479 (60%) are unique to a single network. This is consistent with the general idea that a relatively small number of elements is copiously recombined to integrate their function.

4 Analysis of the 1-neighbourhoods

What is the nature of the connectivity, in the networks, between an individual node and its neighbours, and how do these patterns vary among the networks?

Recall that for each node, in each network, the '1-neighbourhood' of that node is the subgraph of that network that contains, as nodes, the chosen molecule, and all molecules connected to or from the chosen central molecule by an edge.

We have examined how these 1-neighbourhood graphs of a single molecule compare, in the five networks.

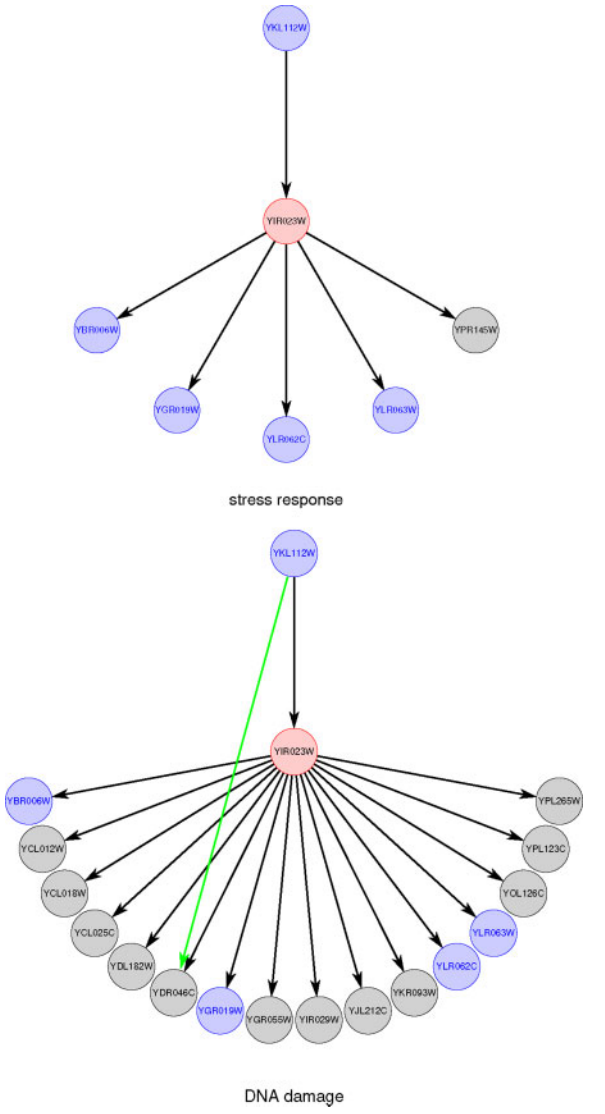


Fig. 4. The 1-neighbourhoods of the selected central node YIR023W in two networks: stress response and DNA damage. In each of these networks, the central node has only one parent neighbour. In the stress response network, YIR023W has five child connections. In the DNA damage network, YIR023W has 16 child connections, four of which are shared edges in stress response and DNA damage networks

4.1 Some interesting examples

We adopt, temporarily, a case-study mode of presentation. The nodes chosen have 1-neighbourhoods that show either unusual degrees of complexity, or interesting differences in different networks.

Figure 4 shows the comparison of the 1-neighbourhoods of node YIR023W in the stress response and DNA damage networks. The central node, YIR023W, is shown in pink in both frames. Black arrows show edges connecting the central node to parent or to child nodes. Green arrows show edges connecting neighbours of the central node, but not the central node itself. For instance, the green arrow in the DNA damage panel connects a parent of the central node to a child of the central node. If multiple parent nodes are present (not in this case) they are shown in alphabetical order clockwise. If multiple child nodes are present they are shown in alphabetical order counterclockwise.

Nodes shown in blue are common to the 1-neighbourhoods of YIR023W in both networks. Each of these 1-neighbourhoods of YIR023W contains a single, common parent node, YKL112W. This is a relatively simple example, in which the 1-neighbourhood of the central node in the stress response network is *almost* a subset of the 1-neighbourhood in the DNA damage network, except for the child node YPR145W which does not appear in the DNA damage network. Incidentally YPR145W is a target gene; i.e. in no network is there an edge for which YPR145W is the parent.

To what extent are the basic motifs (see Fig. 2) visible in the 1-neighbourhood? Any node with more than one child node exhibits one or more SIMs, or forks (or, alternatively, a single fork with multiple tines). Every SIM in the network will appear in the 1-neighbourhood of the parent node of the SIM. The green arrow, connecting a parent node of YIR023W to a child node in the DNA damage network, indicates an FFL. Every FFL in a network will appear in the set of 1-neighbourhoods, with the central node of the 1-neighbourhood corresponding to the intermediate node of the FFL, as in the case of the 1-neighbourhood of YIR023W in the DNA damage network. An MIM *might* occur within a 1-neighbourhood. But not all MIMs in the network will appear in the 1-neighbourhood of some node.

Figure 4 showed a relatively simple example. Contrast it with Figure 5, which shows an unusually complicated example: the 1-neighbourhoods of YDR207C in the cell cycle and sporulation networks.

The 1-neighbourhood of YDR207C in the cell cycle network contains edges from three parent and to four child nodes. There are in addition two edges directly between a parent node and a child node (shown in green). The 1-neighbourhood of YDR207C in the cell cycle network does contain a MIM:

It also contains two FFLs.

In contrast, the 1-neighbourhood of YDR207C in the sporulation network is among the most complicated observed in these networks. Not only is it very rich in edges to child nodes, but there are 16 edges directly connecting a parent of the central node YDR207C to a child and 7 edges connecting two child nodes. In this case also, the cell cycle 1-neighbourhood is almost a subset of the sporulation neighbourhood: they have the same parent nodes, and share three child nodes (of the four in the cell cycle network). The two edges between the parent node YOL004W and the two child nodes YGL073W and YGR044C are also common to both networks. However, overall, one can see that in the sporulation network YDR207C interactions are significantly more complex, than the more moderate demands on the cell cycle network. Of course, it contains many MIMs and FFLs.

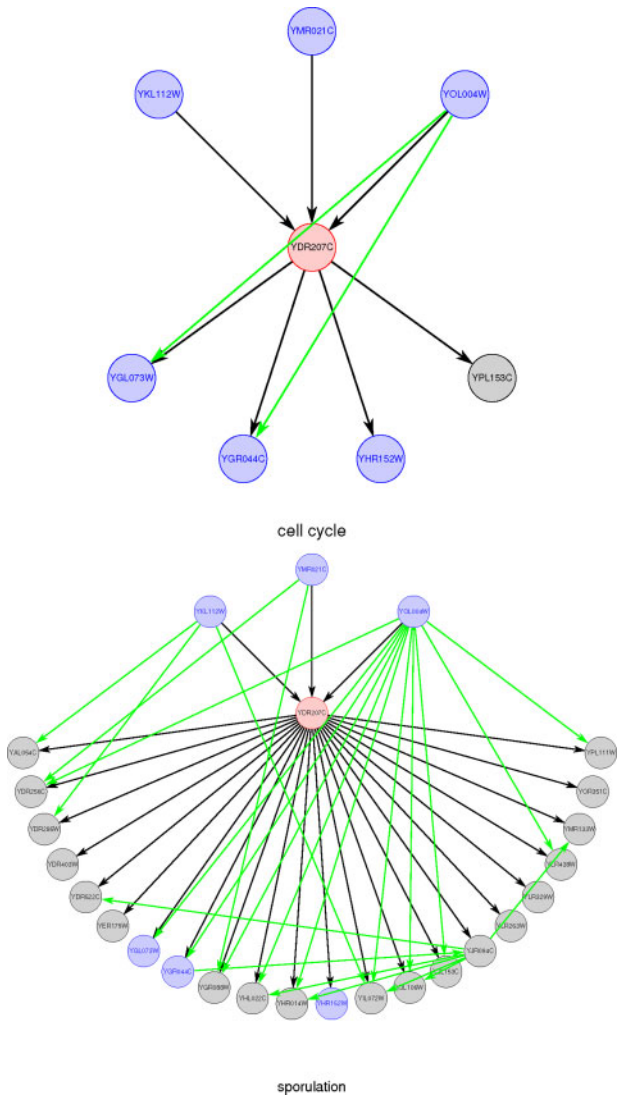


Fig. 5. The 1-neighbourhoods of node YDR207C in the cell cycle and sporulation networks. The cell cycle 1-neighbourhood, on the left, is of moderate complexity, containing edges from three parent and to four child nodes. The sporulation 1-neighbourhood is more complex, containing edges from the same three parent nodes, but edges to 22 child nodes; and many additional 'green' edges

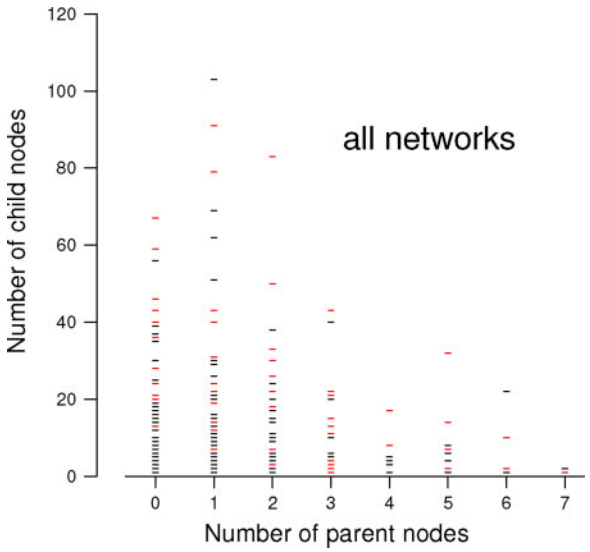


Fig. 6. Different combinations of numbers of parent and child nodes in 1-neighbourhoods of transcription factors, combining all networks. This does not indicate the *numbers* of nodes that have the same number of edges from parent and to and child nodes

4.2 Edge-count statistics

We have computed comparisons of the 1-neighbourhoods of every node in every pair of networks. Figure 6 shows the possible observed values of the combination of numbers of parent nodes and number of child nodes. Thus e.g. the hash mark at (0, 72) means that there is *some* neighbourhood that has 0 parent nodes and 72 child nodes.

Data appearing in red correspond to 1-neighbourhoods for which there are ≥ 4 additional edges, between parent and child nodes, or between parent and other parent, or between child and other child nodes (i.e. the edges that appear green in the figures). This gives a sense of the distribution of complexities of nodes—complexity in terms of numbers of edges in addition to edges connecting parent and child nodes to the central nodes.

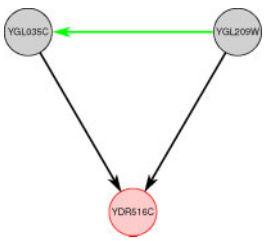
Figure 6 combines the data for all five networks. The different networks differ in their distributions of edge counts. Figure 7 shows the corresponding graphs for the two individual networks that are the most dissimilar.

Many nodes have the same numbers of input and/or output nodes in two different networks, but different sets of parent or child nodes. Many different nodes share these same 1-neighbourhood connectivity patterns. These 1-neighbourhoods have the same topology but the partners have been reprogrammed. (Many of them contain only one or two edges.)

Given that there are 113 transcription factors, they have potentially 565 1-neighbourhoods in the five networks. In 197 cases, a transcription factor makes NO edges in one or more networks, leaving 368 non-trivial 1-neighbourhoods. Of these, 198—slightly over half—have no edges that do not contain the central node—i.e. no 'green arrows' in pictures such as Figure 4. (Let us call these green arrows 'non-central edges'.) Such 1-neighbourhoods have a particularly simple topology. For the others, a rough measure of the complexity of a 1-neighbourhood is to add up the nodes that do *not* involve the central nodes. Figure 8 shows the histogram of the distribution of numbers of 1-neighbourhoods containing different numbers of non-central edges. For 119 1-neighbourhoods, in the five networks, the number of these non-central edges is nonzero but ≤ 3; these have more than the simplest possible topology, but could still be counted as relatively simple. This leaves 51 1-neighbourhoods with more complex topologies (<10%).

4.3 Are 1-neighbourhoods of individual nodes conserved between different networks?

No node has the same 1-neighbourhood in all five networks. Node YDR516C, which is a target node, not a transcription factor, has the same 1-neighbourhood in the diauxic shift, DNA damage, sporulation and stress response networks, but is not present in the cell cycle network (i.e. there are no edges joining YDR516C to any other node in the cell cycle network).



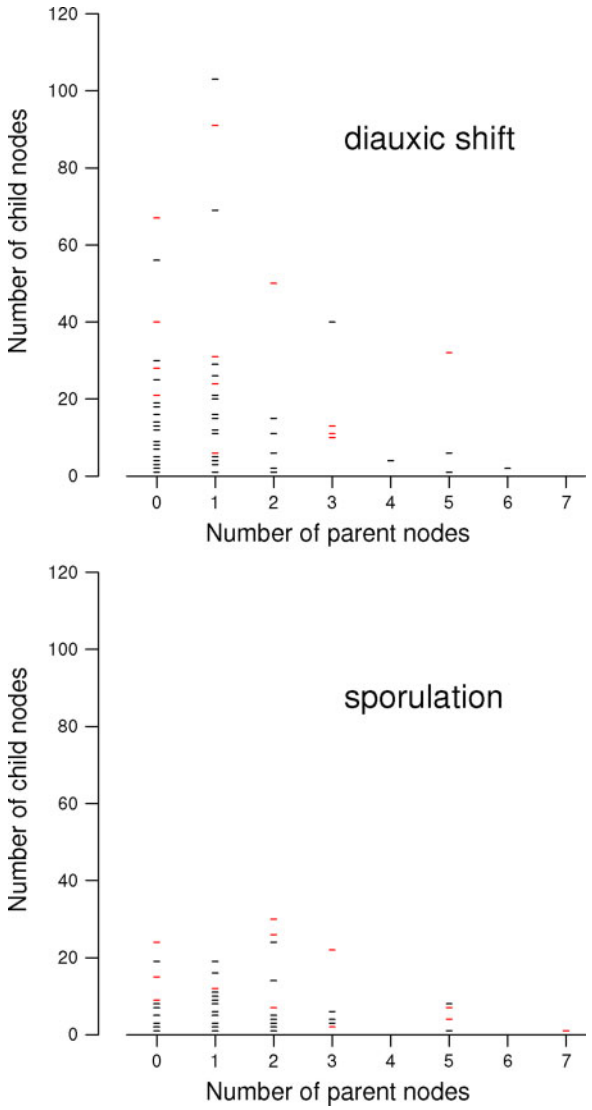
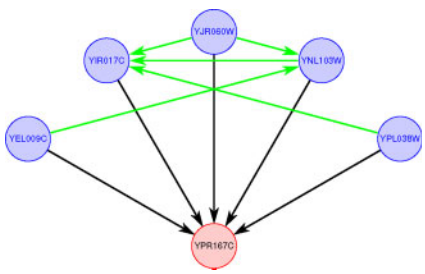


Fig. 7. Different combinations of numbers of parent and child nodes in 1-neighbourhoods of transcription factors, in (top) diauxic shift and (bottom) sporulation networks

There are many nodes with the same 1-neighbourhoods in two different networks, but most of them are relatively trivial, in that they contain only one or two edges. One interesting example is YPR167C, which has the following 1-neighbourhood in *both* the cell cycle and DNA damage networks:



These 1-neighbourhoods in two different networks conserve not only all of the parents, but also all links between pairs of parents. This is quite unusual.

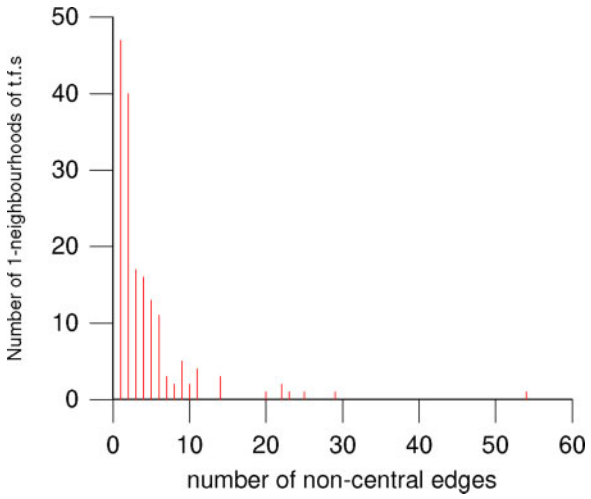


Fig. 8. Distribution of number of 1-neighbourhoods, in all five networks, that contain different numbers of non-central edges. (Counting 1-neighbourhoods of all nodes in each network separately, there are 7012 1-neighbourhoods with 0 non-central edges; these do not appear in this graph.) One node has a very unusual 1-neighbourhood: YER111C in the cell cycle network has 54 non-central edges: There are 46 edges between a parent node and a child and 8 edges between two child nodes

YGL209W illustrates somewhat more complicated sets of relationships between 1-neighbourhoods in different networks (see Fig. 9).

The 1-neighbourhoods of YGL209W appear to separate into two sets. The 1-neighbourhood in the cell cycle network is almost a subset of the 1-neighbourhood of the sporulation network (the node YIL162W in the cell cycle network does not appear in the sporulation network). Both edges between common child nodes that appear in the cell cycle network also appear in the sporulation network: YGL035C → YDR146C and YGL035C → YKL109W.

In turn, the 1-neighbourhood in the sporulation network is almost a subset of the 1-neighbourhood in the stress response network: The parent node and most of the child nodes are shared. Again, the set of edges between pairs of common child nodes is the same.

The 1-neighbourhoods of YGL209W in the diauxic shift and DNA damage networks are somewhat similar to each other, but quite different from the other three. There is no parent node in the diauxic shift network; note, however, that the parent node in the DNA damage network, YCR065W, is the same as the parent node in the other three networks (see Fig. 9a). With one exception—14 out of 15—the set of child nodes in the DNA damage network is a subset of the child nodes in the diauxic shift network. All the edges between pairs of child nodes in the DNA damage network also appear in the diauxic shift network.

In all five networks, with one exception all edges between child nodes of YGL209W have YGL035C as the parent node.

What is the relationship between the simpler 1-neighbourhoods in the cell cycle, sporulation and stress response networks, and the more complicated ones in the diauxic shift and DNA damage networks? With only two exceptions, the first three are subsets of the second two. The exceptions are node YLR044C which is a child node in the 1-neighbourhood of YGL209W in the stress response network but not in the DNA damage network (although it *is* present in the diauxic shift network), and the missing parent node YCR065W in the diauxic shift network. It would be possible to describe the pattern of similarities and differences among these 1-neighbourhoods by a tree.

Perhaps the most extreme examples are the 1-neighbourhoods of YER111C (not illustrated). In the cell cycle network, YER111C has 3 edges from parents, 43 edges to child partners, 46 edges from parent to child partners and 8 edges between child partners. However, YER111C has NO edges in any of the other four networks. This is an extreme example of reprogramming of the networks.

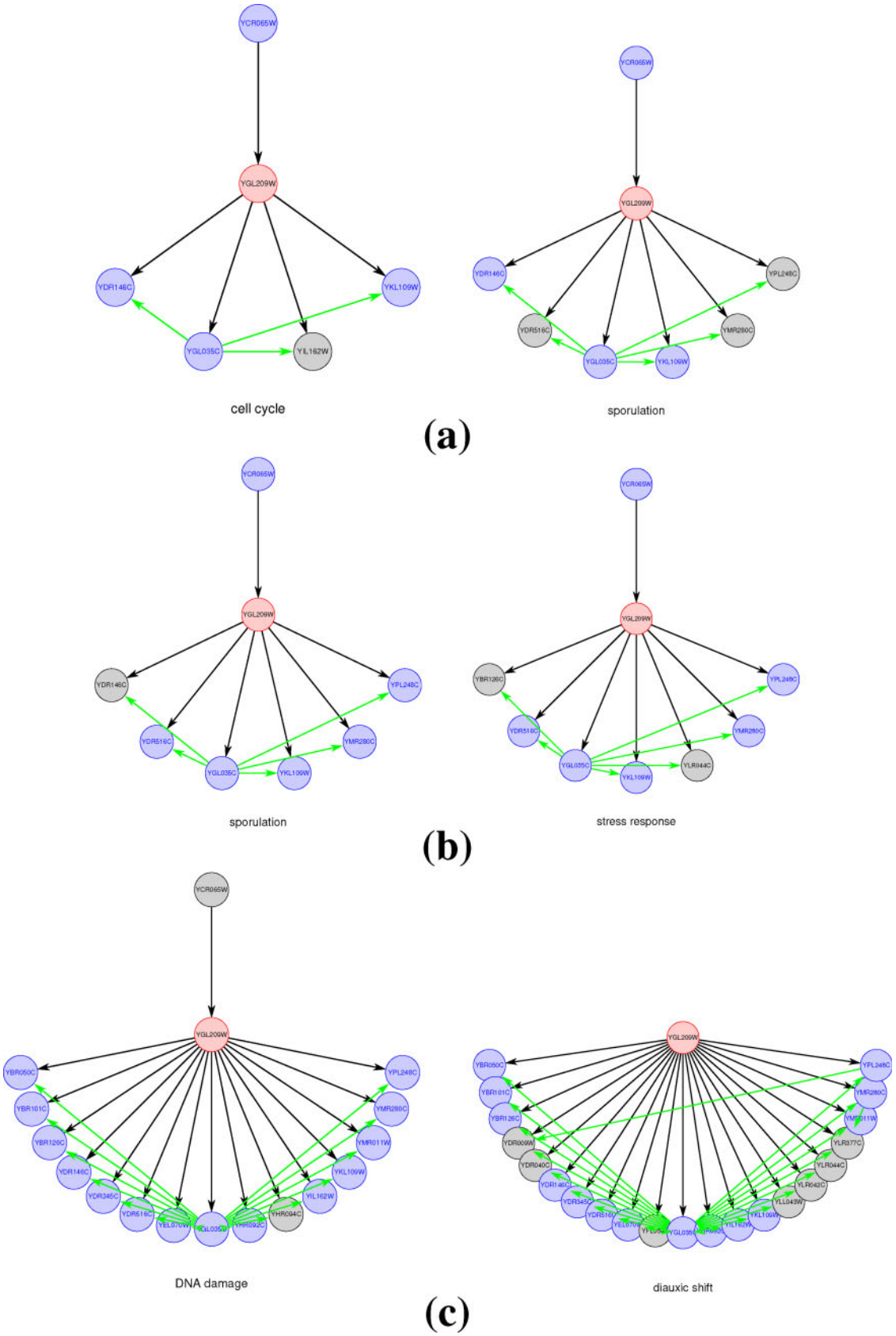


Fig. 9. 1-neighbourhoods of YGL209W. (**a**) Comparison of cell cycle and sporulation networks. (**b**) Comparison of sporulation and stress response networks. (The sporulation 1-neighbourhood is repeated because the colour patterns are different.) (**c**) Comparison of DNA damage and diauxic shift networks

5 Conclusions

We have analyzed and compared the structures of 1-neighbourhoods of all nodes in the yeast regulatory networks in five physiological states. This approach differs from earlier investigations of the structures of these networks that focussed on counting motifs, but not comparing individual components of the networks characterizing different states.

We were guided, in approaching this problem, by two analogies:

1. The enzymes that catalyse metabolic reactions are like 'hardware; the regulatory molecules are like 'software', and can be reprogrammed.
2. If the elementary motifs of networks—SIM, MIM and FFL (see Fig. 2)—are like the secondary structures of proteins, what, in networks, corresponds to the tertiary structure?

Indeed, our investigations show that:

1. The components of the regulatory networks do behave like components of 'software', in that (a) different networks make different choices of which transcription factors to use and (b) even for transcription factors common to different networks there has been extensive reprogramming of the topologies of their 1-neighbourhoods.
2. We catalogued and examined the structures of the 1-neighbourhoods. Many of them comprise larger segments of the networks than the elementary motifs, and, indeed, often contain them. The 1-neighbourhoods exhibit a great richness of variety of structure, some of which we have presented.

Our approach here was limited to structures involving immediate neighbours of each central node in each network; i.e. in subgraphs containing nodes connected directly by edges to each central node. The limitation to immediate neighbours clearly suggests the generalization, for future work, to larger neighbourhoods.

Acknowledgement

We are grateful to M. Madan Babu and A.G. Murzin for helpful discussion.

Financial Support: none declared.

Conflict of Interest: none declared.

References

Balaji,S. *et al.* (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.*, **360**, 213–227.

Konagurthu,A.S. and Lesk,A.M. (2008a) On the origin of distribution patterns of motifs in biological networks. *BMC Syst. Biol.*, **2**, 73.

Konagurthu,A.S. and Lesk,A.M. (2008b) Single and multiple input modules in regulatory networks. *Proteins Struct. Funct. Bioinform.*, **73**, 320–324.

Lex,A. *et al.* (2014) Upset: visualization of intersecting sets. *IEEE Trans. Visual Comput. Graph.*, **20**, 1983–1992.

Luscombe,N.M. *et al.* (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, **431**, 308–312.

Milo,R. *et al.* (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.

Morris,K.V. (2011) The emerging role of RNA in the regulation of gene transcription in human cells. *Semin. Cell Devel. Biol.*, **22**, 351–358.

Yu,H. and Gerstein,M. (2006) Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci. USA*, **103**, 14724–14731.