The influence of local topology and dynamics on robustness

Daniel F. Simola
Kim Lab, UPenn
15 Nov 2005
Outline

1. Why - Examples of robustness
2. What - Definitions of robustness
3. How - Causes of phenotypic robustness
   1. Network motifs as epistatic, buffering mechanisms
Observing robustness

Three examples illustrating the effect of systematic gene deletions on yeast phenotypes

• 2026 single gene deletion haploid yeast strains (~33% of genes)

• 83% of deletions nonlethal in rich medium

• 138 (8.5%) nonessential and 4 essential genes have a close homolog (redundancy)

• competition assays to estimate relative strain abundance and growth rates (to wt) using Affy arrays

• 60% of 500 deletion strains displayed no quantitative growth defects in rich or poor medium
• Update of previous study, now with 5916 (96%) deletion strains
• Now 81.3% (83% before) of deletions nonlethal in rich medium
• 15% (vs 40%) of strains show a reduced growth phenotype in YPD (growth rate < 90% of wt)
• Morphological changes
  • 15% of deletion strains show some kind of morphological alteration
  • clumped and elongated strains enriched for mutations in cell growth, division, and DNA synthesis
  • round strains enriched for mutations in protein synthesis
Morphological changes
- P Jorgensen et al, “Systematic Identification of Pathways That Couple Cell Growth and Division in Yeast”
- The G1 to S transition should be made robustly - don’t want to divide based on random changes in environment or gene activity
- 18.7% deletions lethal
- 441/4812 viable haploid deletion strains (9%) show “extreme” change in cell size (top/bottom 5% of overall size distribution)
- Remaining 90% of strains both viable and show no gross phenotypic changes in the cell division decision
Sources of phenotypic robustness

- Examples show that on average, deletion of a gene yields negligible change in viability, growth rate, and morphology, compared to wild type.
- This suggests existence of either simple, single gene redundancies or complex, epistatic relationships among genes in yeast cells.
Sources of phenotypic robustness

• More than gene duplication, because less than 10% of all yeast genes have a very similar homolog

• Possible that a nonsimilar gene encodes a protein that does function similarly to the deleted gene (functional redundancy)

• Also possible that the nature of interactions among genes buffers the cell against phenotypic changes due to removal of individual genes

• These experiments performed in a constant environment
Definitions

- **Phenotype**: any measurable trait or characteristic of an organism resulting from the expression of one or more genes

- **Phenotypic robustness**: the reduced sensitivity of a phenotype to perturbations in the conditions (parameters) that affect its expression

Types of change

- There are mechanisms that buffer against
  - heritable change
    - mutations and genetic changes resulting from reproduction
  - nonheritable change
    - environmental perturbations, such as temperature or chemical concentration (external) as well as random fluctuations in gene expression levels (internal)
Causes of robustness

- A mechanism can generate robustness in two ways
Causes of robustness

- A mechanism can generate robustness in two ways
- buffering against perturbations
Causes of robustness

• A mechanism can generate robustness in two ways

• eliminating perturbations
Evolution of robustness via redundancy or antiredundancy

- Recall population size ($N$) and cost of redundancy ($s$)
- Eliminating perturbations
- Buffering against perturbations

Evolution of robustness via redundancy or antiredundancy

- Eliminating perturbations
  - antiredundant mechanisms evolve due to large $N, s$
  - e.g. “overlapping reading frames, absence of tRNA suppressor genes, codon bias, loss of DNA error repair, reduced number of promoters, coordinated expression of genes, and checkpoint genes. All of these mechanisms remove mutant genomes from populations (either of individuals or cells).”

Evolution of robustness via redundancy or antiredundancy

- Buffering against perturbations
  - redundant mechanisms evolve due to small N, s
  - e.g. “duplicated genes, correlated gene functions, tRNA suppressors, heat shock proteins, molecular quality control, and alternative metabolic pathways. These mechanisms, which incur a cost (through increased genome size or greater need for resources), mask the effects of mutation.”

Three types of buffering mechanisms, all inherently epistatic

- **Redundancy** - multiple copies of similarly functioning gene or protein, e.g. cyclins 1, 2

- **Epistasis among unrelated genes** - the genes responsible for buffering against phenotypic change are not the genes that cause the phenotype itself

- **Epistasis among genes encoding phenotypic trait of interest** - interaction of genes that cause the trait is inherently buffering
Epistasis among unrelated genes

- E.g., heat-shock protein chaperones buffer against thermal fluctuations by assisting in the proper folding of other proteins
Epistasis among genes encoding phenotypic trait of interest

- E.g. protein-DNA binding and protein-protein interaction properties of transcription factors to activate gene transcription despite varying abundance of individual TFs
- Special case is dominance at single locus (among alleles), which buffers against effect of mutant allele
Epistasis among genes encoding phenotypic trait of interest

• Locally studied as “network motifs”, where robustness is the stable expression of a gene in the face of the varying abundances (activities) of the particular regulatory TFs

• At a more global scale, robustness can be assessed by a network's degree of sensitivity, as a function of average gene activity and network topology

  • $S = 2Kp(1-p)$ or $\bar{S} = 2\bar{K}p(1-p)$
  • $S < 1$ (fully buffered, ordered)
  • $S = 1$ (critical)
  • $S > 1$ (chaotic)
RJ Prill, PA Iglesias, A Levchenko

“Dynamic properties of network motifs contribute to biological network organization”

Representing real interactions

A

\[ X \rightarrow Y \] represents

transcription network

\[ \text{gene } x \quad \text{gene } y \]

neuron synaptic connection network

ecological food web

B

1 2 3 4 5 6 7 8 9 10 11 12 13

Milo et al (2002)
Motif connectivity

- Transcriptional regulatory networks have a fixed topological structure
- Concept of a network motif, "a directed subgraph, consisting of a few nodes, that is embedded in a larger directed graph"
- Generally not presumed that motif instances are functionally coherent units
- Motif instances are over/under-represented in various networks, e.g. *E. coli* transcriptional network, the internet
Motivation and hypothesis

• Why are certain motifs overrepresented in biological networks
  • motif instances have fixed connectivity, but dynamical properties vary based on input activities
  • a motif’s dynamical behavior may be an important criterion to assess functional significance
  • and thus may potentially influence its own abundance

• Motif robustness
  • stability to small-scale perturbations and noise in node activities
Illustrations for a general 2-node network motif

![Illustrations for a general 2-node network motif](image-url)
Properties of a general 2-node network motif

- **Gain**: $k = a_{12} a_{21}$
- **Positive feedback** (+) results from positive gain ($k > 0$)
  - obtained when inter-node edges have same signs
- **Negative feedback** (-) results from negative gain ($k < 0$)
  - obtained when internode edges have different signs
- **Instability** results from extreme positive feedback
- **Oscillation** results from extreme negative feedback

- Region of stability depends on magnitude of self-edges
- Surprising: negative feedback does not always result in stability
Scheme for modeling the local stability of a network motif

Infer network motifs

normalize matrix under constraints

I  motif

\[ \begin{align*}
\text{2} & \quad a_{22} \\
\text{3} & \quad a_{32} \\
\text{1} & \quad a_{11} \\
\end{align*} \]

\[ a_{12}, a_{13}, a_{23} \]

II  dynamical system

\[ \begin{align*}
\dot{x}_1 & = f_1(x_1, x_2, x_3) \\
\dot{x}_2 & = f_2(x_1, x_2, x_3) \\
\dot{x}_3 & = f_3(x_1, x_2, x_3) \\
\dot{x} & = 0 \iff x = x^* \\
\end{align*} \]

III  linearize

\[ \begin{align*}
\text{eig} & \left[ \begin{array}{ccc}
  a_{11} & a_{12} & a_{13} \\
  0 & a_{22} & a_{23} \\
  0 & a_{32} & a_{33} \\
\end{array} \right] \\
\end{align*} \]

\[ a_{ii} \in [-1,0) \]

\[ a_{ij} \in [-1,1] \]

IV  simulate

\[ \dot{x}_\delta = \left\{ \frac{\partial f_i}{\partial x_j} \right\}_{x=x^*} x_\delta \]

assume system has a steady state \((x^*)\)

analyze dynamics within a linear neighborhood of \(x^*\)
Stability metric

- **Structural stability score (SSS)** - probability that a motif’s dynamical system displays non-oscillatory, damped response to a small perturbation from a steady-state

- Estimate stability by sampling matrices from a uniform distribution (10,000 per motif)

- SSS equivalent to the fraction of sampled matrices containing only real numbers
Stability classes

- Analysis of all 13 3-node and 199 4-node network motifs wrt SSS
- Three distinct classes
  - class I: ultra stable ($SSS = 1$)
    - motifs are all DAGs
  - class II: moderately stable ($SSS \approx 0.4$)
    - single 2-node feedback loop
    - assuming the feedback loop is negative, motifs guaranteed either stable or damped oscillatory
  - class III: unstable ($SSS < 0.2$)
    - various more complicated motifs, including multiple 2-node loops, 3-node and 4-node loops, nested loops, etc
    - stability cannot be guaranteed

- In general, SSS correlates negatively with number and length of loops
Data sources

- **E. coli**: transcription interactions between regulatory proteins and genes
  - Source: RegulonDB

- **S. cerevisiae**: transcriptional network
  - Source: Young lab ChIP data + de novo TFBS prediction

- **C. elegans**: synaptic connections between neurons
  - Source: deterministic fate map

- **mammals**: signal transduction interactions in mammalian cells
  - Source: signal transduction knowledge environment (STKE)

- **Drosophila**: transcription networks guiding development
  - Source: GeNet
Motif abundance and SSS for 5 networks

Three nodes

Four nodes

Motif Abundance

Structural Stability Score (dashed)

Motif ID

Motif ID

10^6

10^4

10^2

10^0

Bact transcription

Yeast transcription

STKE signal

Dros transcription

Worm neuron

10^8

10^6

10^4

10^2

10^0

1

50

100

150

199

0

0

0

0

Remarkably, the perturbation can be monotonically increasing, decreasing, or otherwise variable, depending on the parameter values. Provided the integral gain is not negative, positive feedback dominates the system. The feedback can then be specified as negative, neutral, or positive, corresponding to the sign of the feedback gain. There is a necessary assumption that the feedback gain is not negative, because a negative feedback gain would result in an increase of the perturbation. Therefore, positive feedback and negative feedback cannot coexist, at least for second-order systems.
Importance of stability

- If stability is an important aspect of network organization, expect that motifs with high SSS would be overrepresented compared to less stable motifs.
- Test: Compare frequency of each motif in a real network to that in random (null) networks.
  - Erdos-Renyi networks
    - Given N nodes, draw an edge between a pair of nodes with probability p.
    - Degree distribution of a node is Poisson.
    - Low connectivity variance.
  - Scale-free networks
    - Degree distribution modeled as power law: $P(k) \sim k^{-y}$, $y \sim 2$ or 3.
    - High connectivity variance (nodes with large $k$, aka hubs).
Computing Z scores

- For a given motif, count\* number of instances of motif in real network ($N_{\text{reali}}$)
- Generate 100 random networks with same number of nodes and edges as real network
- Count* motif instances in each random network ($N_{\text{randi}}$)

\[
Z_i = \frac{N_{\text{reali}} - \overline{N}_{\text{randi}}}{\text{SD}(N_{\text{randi}})}
\]

\[
NZ_i = \frac{Z_i}{(\sum Z_i^2)^{1/2}}
\]

*Really sampling due to large network size
Distribution of Z scores for 3-node motifs

Each network contains 13 network motifs classified by stability. The figure shows the distribution of normalized Z scores for each motif. The motifs are color-coded based on their stability group: light blue for group I, green for group II, gray for group III, and red for group IV. The bars represent the density of the motifs, and the lines indicate the distribution of the Z scores.

Includes damped oscillations in SSS
3-node Z score calculation using random networks generated with a degree distribution identical to that estimated from the real networks

**ER rand**

**SF rand**

Claim ER is less biased because focus remains on local connectivity

Because motif abundance correlates with SSS, negative Z scores are not informative.
Motifs found in biological and technological networks

<table>
<thead>
<tr>
<th>Network</th>
<th>Nodes</th>
<th>Edges</th>
<th>(N_{\text{real}})</th>
<th>(N_{\text{rand}} \pm \text{SD} )</th>
<th>(Z) score</th>
<th>(N_{\text{real}})</th>
<th>(N_{\text{rand}} \pm \text{SD} )</th>
<th>(Z) score</th>
<th>(N_{\text{real}})</th>
<th>(N_{\text{rand}} \pm \text{SD} )</th>
<th>(Z) score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(transcription)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E.\ coli)</td>
<td>424</td>
<td>519</td>
<td>40</td>
<td>7 ± 3</td>
<td>10</td>
<td>203</td>
<td>47 ± 12</td>
<td>13</td>
<td>1812</td>
<td>300 ± 40</td>
<td>41</td>
</tr>
<tr>
<td>(S.\ cerevisiae)†</td>
<td>685</td>
<td>1,052</td>
<td>70</td>
<td>11 ± 4</td>
<td>14</td>
<td>1,460,000</td>
<td>18 ± 8</td>
<td>13</td>
<td>15,000</td>
<td>14 ± 6</td>
<td>25</td>
</tr>
<tr>
<td><strong>Neurons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(C.\ elegans)†</td>
<td>252</td>
<td>509</td>
<td>125</td>
<td>90 ± 10</td>
<td>3.7</td>
<td>127</td>
<td>55 ± 13</td>
<td>5.3</td>
<td>227</td>
<td>35 ± 10</td>
<td>20</td>
</tr>
<tr>
<td><strong>Food webs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little Rock</td>
<td>92</td>
<td>984</td>
<td>3219</td>
<td>3120 ± 50</td>
<td>2.1</td>
<td>7295</td>
<td>2220 ± 210</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ythan</td>
<td>83</td>
<td>391</td>
<td>1182</td>
<td>1020 ± 20</td>
<td>7.2</td>
<td>1357</td>
<td>230 ± 50</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St. Martin</td>
<td>42</td>
<td>205</td>
<td>469</td>
<td>450 ± 10</td>
<td>NS</td>
<td>382</td>
<td>130 ± 20</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chesapeake</td>
<td>31</td>
<td>67</td>
<td>80</td>
<td>82 ± 4</td>
<td>NS</td>
<td>26</td>
<td>5 ± 2</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coachella</td>
<td>29</td>
<td>243</td>
<td>279</td>
<td>235 ± 12</td>
<td>3.6</td>
<td>181</td>
<td>80 ± 20</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skipwith</td>
<td>25</td>
<td>189</td>
<td>184</td>
<td>150 ± 7</td>
<td>5.5</td>
<td>397</td>
<td>80 ± 25</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Brook</td>
<td>25</td>
<td>104</td>
<td>181</td>
<td>130 ± 7</td>
<td>7.4</td>
<td>267</td>
<td>30 ± 7</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Electronic circuits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(forward logic chips)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s15850</td>
<td>10,383</td>
<td>14,240</td>
<td>424</td>
<td>2 ± 2</td>
<td>285</td>
<td>1040</td>
<td>1 ± 1</td>
<td>1,200</td>
<td>480</td>
<td>2 ± 1</td>
<td>335</td>
</tr>
<tr>
<td>s38584</td>
<td>20,717</td>
<td>34,204</td>
<td>413</td>
<td>10 ± 3</td>
<td>120</td>
<td>1739</td>
<td>6 ± 2</td>
<td>800</td>
<td>711</td>
<td>9 ± 2</td>
<td>320</td>
</tr>
<tr>
<td>s38417</td>
<td>23,843</td>
<td>33,661</td>
<td>612</td>
<td>3 ± 2</td>
<td>400</td>
<td>2404</td>
<td>1 ± 1</td>
<td>2550</td>
<td>531</td>
<td>2 ± 2</td>
<td>340</td>
</tr>
<tr>
<td>s9234</td>
<td>5,844</td>
<td>8,197</td>
<td>211</td>
<td>2 ± 1</td>
<td>140</td>
<td>754</td>
<td>1 ± 1</td>
<td>1050</td>
<td>209</td>
<td>1 ± 1</td>
<td>200</td>
</tr>
<tr>
<td>s13207</td>
<td>8,651</td>
<td>11,831</td>
<td>403</td>
<td>2 ± 1</td>
<td>225</td>
<td>4445</td>
<td>1 ± 1</td>
<td>4950</td>
<td>264</td>
<td>2 ± 1</td>
<td>200</td>
</tr>
<tr>
<td><strong>Electronic circuits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(digital fractional multipliers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s208</td>
<td>122</td>
<td>189</td>
<td>10</td>
<td>1 ± 1</td>
<td>9</td>
<td>4</td>
<td>1 ± 1</td>
<td>3.8</td>
<td>5</td>
<td>1 ± 1</td>
<td>5</td>
</tr>
<tr>
<td>s420</td>
<td>252</td>
<td>399</td>
<td>20</td>
<td>1 ± 1</td>
<td>18</td>
<td>10</td>
<td>1 ± 1</td>
<td>10</td>
<td>11</td>
<td>1 ± 1</td>
<td>11</td>
</tr>
<tr>
<td>s838‡</td>
<td>512</td>
<td>819</td>
<td>40</td>
<td>1 ± 1</td>
<td>38</td>
<td>22</td>
<td>1 ± 1</td>
<td>20</td>
<td>23</td>
<td>1 ± 1</td>
<td>25</td>
</tr>
<tr>
<td><strong>World Wide Web</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nd.edu§</td>
<td>325,729</td>
<td>1.46e6</td>
<td>6.8e6</td>
<td>5.4e4 ± 4e2</td>
<td>800</td>
<td>6.8e6</td>
<td>5e4 ± 4e2</td>
<td>15,000</td>
<td>1.2e6</td>
<td>1e4 ± 2e2</td>
<td>5000</td>
</tr>
</tbody>
</table>

Milo et al (2002)
Z score of 3-node motifs, classified by stability

E. coli

yeast

STKE

Dros. dev. trans.

C. elegans neuron
For example, the six-edge density group is comprised of 47 network motifs: one class I, nine class II, and 37 class III. The Drosophila developmental network contains a highly overrepresented class I motif, which has the highest Z score in the density group. The remaining high Z scores correspond to class II motifs. None of the 37 low-stability motifs have high Z scores. Overall, high Z scores correspond to high SSS, but an SSS equal to unity does not always result in a high Z score, implying that structural stability may be necessary, but not sufficient, for network motif overrepresentation.

In the preceding investigation, we extended a structural analysis technique (decomposition of a network into subgraphs) to a dynamical analysis. Our characterization of motif dynamics implicitly assumed that subsystems consisting of a few nodes could behave relatively autonomously despite being embedded in a large network. Conceptually, the assumption is valid when the activity of a node does not affect the other nodes in the network.
4-node Z scores using random networks constrained by degree distribution

For example, the six-edge density group is comprised of 47 network motifs: one class I, nine class II, and 37 class III. The Drosophila developmental network contains a highly over-represented class I motif, which has the highest Z score in the density group. The remaining high Z scores correspond to class II motifs. None of the 37 low-stability motifs have high Z scores. Overall, high Z scores correspond to high SSS, but an SSS equal to unity does not always result in a high Z score, implying that structural stability may be necessary, but not sufficient, for network motif overrepresentation.

In the preceding investigation, we extended a structural analysis technique (decomposition of a network into subgraphs) to a dynamical analysis. Our characterization of motif dynamics implicitly assumed that subsystems consisting of a few nodes could behave relatively autonomously despite being embedded in a large network. Conceptually, the assumption is valid when the activity of a node does not affect other nodes.

Figure 4. Distribution of Normalized Z Scores of Four-Node Network Motifs

(Top panel) All 199 motifs are sorted on the x-axis first, according to increasing number of edges (solid red line). For a given number of edges, they are then sorted from high to low SSS (black bars).

(Bottom five panels) Normalized Z scores (green bars) for all 199 network motifs of the indicated biological networks, shown with outlines of the SSS from the top panel (dotted black outline provided as a guide to the eye). Each vertical red line indicates a change in the number of motif edges, indicating boundaries of density groups. The composition of the four-node density groups is specified in Protocol S1.
**Z score of 4-node motifs, classified by stability**

P-value is the probability that the observed differences in Z score among stability classes are expected by chance (Kruskal-Wallis)

---

**E. coli**

**yeast**

**STKE**

**Dros. dev. trans.**

**C. elegans neuron**
Table 1. High Z Scores of Four-Node Network Motifs

<table>
<thead>
<tr>
<th>Motif</th>
<th>Z Score</th>
<th>Bacteria</th>
<th>Yeast</th>
<th>STKE</th>
<th>Fly</th>
<th>Neuron</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Four-edge network motifs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Z score</td>
<td>49</td>
<td>1,461</td>
<td>35</td>
<td>17</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>153</td>
<td>4,113</td>
<td>48</td>
<td>18</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>203</td>
<td>5,574</td>
<td>83</td>
<td>35</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>High Z scores&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>652</td>
<td>17,978</td>
<td>198</td>
<td>68</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>341</td>
<td>6,877</td>
<td>145</td>
<td>55</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Stability&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
<td></td>
</tr>
<tr>
<td>Class III</td>
<td>0/3</td>
<td>0/3</td>
<td>0/3</td>
<td>0/3</td>
<td>0/3</td>
<td></td>
</tr>
<tr>
<td><strong>Five-edge network motifs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Z score</td>
<td>14</td>
<td>2,651</td>
<td>47</td>
<td>37</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>85</td>
<td>13,683</td>
<td>54</td>
<td>41</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>99</td>
<td>16,333</td>
<td>101</td>
<td>79</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>High Z scores&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>510</td>
<td>81,084</td>
<td>213</td>
<td>222</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>185</td>
<td>106</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>149</td>
<td>96</td>
<td>198</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>146</td>
<td>81</td>
<td>189</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>103</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stability&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>1/6</td>
<td>1/6</td>
<td>3/6</td>
<td>3/6</td>
<td>1/6</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>0/14</td>
<td>0/14</td>
<td>3/14</td>
<td>1/14</td>
<td>5/14</td>
<td></td>
</tr>
<tr>
<td>Class III</td>
<td>0/17</td>
<td>0/17</td>
<td>1/17</td>
<td>0/17</td>
<td>0/17</td>
<td></td>
</tr>
<tr>
<td><strong>Six-edge network motifs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Z score</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>378</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>157</td>
<td>634</td>
<td></td>
</tr>
<tr>
<td>High Z scores&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>369</td>
<td>1,862</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>310</td>
<td>1,621</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stability&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>0/1</td>
<td>0/1</td>
<td>0/1</td>
<td>1/1</td>
<td>0/1</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>0/9</td>
<td>0/9</td>
<td>0/9</td>
<td>3/9</td>
<td>3/9</td>
<td></td>
</tr>
<tr>
<td>Class III</td>
<td>0/37</td>
<td>0/37</td>
<td>0/37</td>
<td>0/37</td>
<td>0/37</td>
<td></td>
</tr>
</tbody>
</table>
Class III motifs do not have high Z scores (> 1SD above mean of edge group)
Table 1. High Z Scores of Four-Node Network Motifs

<table>
<thead>
<tr>
<th>Motif</th>
<th>Z Score</th>
<th>Bacteria</th>
<th>Yeast</th>
<th>STKE</th>
<th>Fly</th>
<th>Neuron</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-edge network motifs</td>
<td>Mean Z score</td>
<td>49</td>
<td>1,461</td>
<td>35</td>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>153</td>
<td>4,113</td>
<td>48</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>203</td>
<td>5,574</td>
<td>83</td>
<td>35</td>
<td>56</td>
</tr>
<tr>
<td>High Z scores&lt;sup&gt;ab&lt;/sup&gt;</td>
<td></td>
<td>652</td>
<td>(I)</td>
<td>17,978 (I)</td>
<td>198 (I)</td>
<td>68 (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>341</td>
<td>(I)</td>
<td>6,877 (I)</td>
<td>145 (I)</td>
<td>55 (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5,598</td>
<td>(I)</td>
<td>38 (I)</td>
<td>62 (I)</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
<td>0/10</td>
<td></td>
</tr>
<tr>
<td>Class III</td>
<td>3/6</td>
<td>3/6</td>
<td>3/6</td>
<td>1/6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five-edge network motifs</td>
<td>Mean Z score</td>
<td>14</td>
<td>2,651</td>
<td>47</td>
<td>37</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>85</td>
<td>13,683</td>
<td>54</td>
<td>41</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>99</td>
<td>16,333</td>
<td>101</td>
<td>79</td>
<td>145</td>
</tr>
<tr>
<td>High Z scores&lt;sup&gt;ab&lt;/sup&gt;</td>
<td></td>
<td>510</td>
<td>(I)</td>
<td>81,084 (I)</td>
<td>213 (II)</td>
<td>222 (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>185</td>
<td>(II)</td>
<td>106 (I)</td>
<td>210 (II)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>149</td>
<td>(I)</td>
<td>96 (I)</td>
<td>198 (II)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>146</td>
<td>(I)</td>
<td>81 (II)</td>
<td>189 (II)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>115</td>
<td>(II)</td>
<td>164 (II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>103</td>
<td>(I)</td>
<td>149 (II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>1/6</td>
<td>1/6</td>
<td>3/6</td>
<td>3/6</td>
<td>1/6</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>0/14</td>
<td>0/14</td>
<td>3/14</td>
<td>1/14</td>
<td>5/14</td>
<td></td>
</tr>
<tr>
<td>Class III</td>
<td>0/14</td>
<td>0/14</td>
<td>0/14</td>
<td>0/14</td>
<td>0/14</td>
<td></td>
</tr>
<tr>
<td>Six-edge network motifs</td>
<td>Mean Z score</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>378</td>
</tr>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>157</td>
<td>634</td>
</tr>
<tr>
<td>High Z scores&lt;sup&gt;ab&lt;/sup&gt;</td>
<td></td>
<td>369</td>
<td>(I)</td>
<td>1,862 (I)</td>
<td>310 (II)</td>
<td>1,621 (II)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>170</td>
<td>(II)</td>
<td>170 (II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class I</td>
<td>0/1</td>
<td>0/1</td>
<td>0/1</td>
<td>1/1</td>
<td>0/1</td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>0/9</td>
<td>0/9</td>
<td>0/9</td>
<td>3/9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class III</td>
<td>0/9</td>
<td>0/9</td>
<td>0/9</td>
<td>3/9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Being a stable motif (class I, II) does not guarantee high Z score
- But motifs with high Z scores are stable (previous slide)
- Thus structural stability may be necessary but not sufficient for overrepresentation of a motif
Model assumptions

- Characterization of network motif dynamics implicitly assumes motifs behave autonomously despite being embedded in a larger network
  - i.e. post-transcriptional details ignored

- Assumption is reasonable when activity of a node does not propagate far in the larger network. This occurs if
  - a small fraction of genes (nodes) are active
  - a small fraction of regulatory interactions (edges) are active
  - (consistent with $S = 2Kp(1-p)$)
Verifying model assumptions

• Context-dependent network (CDN): subset of nodes and edges active in a certain context
  • e.g. nutrient deprivation, heat-shock, etc.

• Data
  • Gasch et al. (2000): yeast environmental stress microarrays
  • Harbison et al. (2004): ChIP + de novo TFBS prediction

• Method
  • Identify genes with $\geq 2$ fold change (cf wt)
  • Infer the full network
  • CDN retains all inbound links to active nodes
Inferred networks are valid under the assumption that the environmental perturbations are “small”, and thus generate an approximately linear amount of activity change in the entire network.

Thus networks estimated from this data are consistent with the 3 and 4 node networks analyzed.

Inclusion of upstream edges attempt to compensate when expression change < 2 fold?


**Results**

Table 2. Size of Active Regulatory Motifs in the Yeast Gene Regulation Network, for Various Experimental Contexts

<table>
<thead>
<tr>
<th>Experimental Context</th>
<th>Regulated Genes (&gt; 2-fold)</th>
<th>Regulated Genes in Network&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Context Nodes&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Context Links&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Average Motif Size (Median)</th>
<th>Average Motif Size (Mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat shock</td>
<td>803</td>
<td>406</td>
<td>519</td>
<td>896</td>
<td>3</td>
<td>3.4 ± 2.4</td>
</tr>
<tr>
<td>H₂O₂</td>
<td>467</td>
<td>225</td>
<td>333</td>
<td>495</td>
<td>2</td>
<td>3.5 ± 2.8</td>
</tr>
<tr>
<td>Dithiothrietol</td>
<td>474</td>
<td>227</td>
<td>331</td>
<td>497</td>
<td>2</td>
<td>3.5 ± 2.8</td>
</tr>
<tr>
<td>Diamide</td>
<td>402</td>
<td>184</td>
<td>283</td>
<td>385</td>
<td>2</td>
<td>3.0 ± 2.1</td>
</tr>
<tr>
<td>Hyper-osmotic</td>
<td>176</td>
<td>81</td>
<td>149</td>
<td>194</td>
<td>2</td>
<td>3.4 ± 3.0</td>
</tr>
<tr>
<td>Full network</td>
<td>2,929</td>
<td>6,170</td>
<td>7</td>
<td></td>
<td></td>
<td>11.3 ± 11.1</td>
</tr>
</tbody>
</table>

<sup>a</sup> The transcriptional regulatory network includes only those genes that have at least one link with p < 0.001.

<sup>b</sup> All regulated genes, plus transcription factors that directly influence regulated genes.

<sup>c</sup> All inbound links to regulated genes.

DOI: 10.1371/journal.pbio.0030343.t002

- Analysis of 3 and 4 node motifs consistent with average indegree
- All active motifs have SSS = 1
Summary of evidence that network motifs act as a driving force in organizing biological networks

- Certain motifs are overrepresented in biological networks (given)
- Certain motifs are more “stable” than others
- Abundance of 3 and 4 node motifs correlates generally with stability (SSS) (i.e. the overrepresented motifs are more stable)
- \( Z \) score (real vs random abundance) correlates inversely with number of edges (density groups) and stability class (I, II, III)
  - Within density groups, the more stable motifs are overrepresented
  - Networks differ in which density groups are overrepresented (variation)
- Small perturbation array analysis reveals activation of 3-4 node subgraphs
Conclusion

• “...in principle, the non-random character of the yeast transcriptional network could have resulted from selection acting on small network motifs that are functionally relevant in specific environmental contexts...”

• Selection could also act to stabilize the global network, which is enforced through local changes.

• In the former case, would expect particular stable motifs to be overrepresented (stability is necessary but not sufficient).

• In the latter case, expect an unbiased choice of stable motifs (stability is sufficient).
Causes of robustness

- Authors claim that this study gives strong support for the importance of transcriptional buffering, because of the overrepresentation of stable motifs, having ignored post-transcriptional details.

- e.g. although heat-shock response (post-transcriptional buffering) involves interactions with chaperone proteins, heat-shock network motifs were detectable.