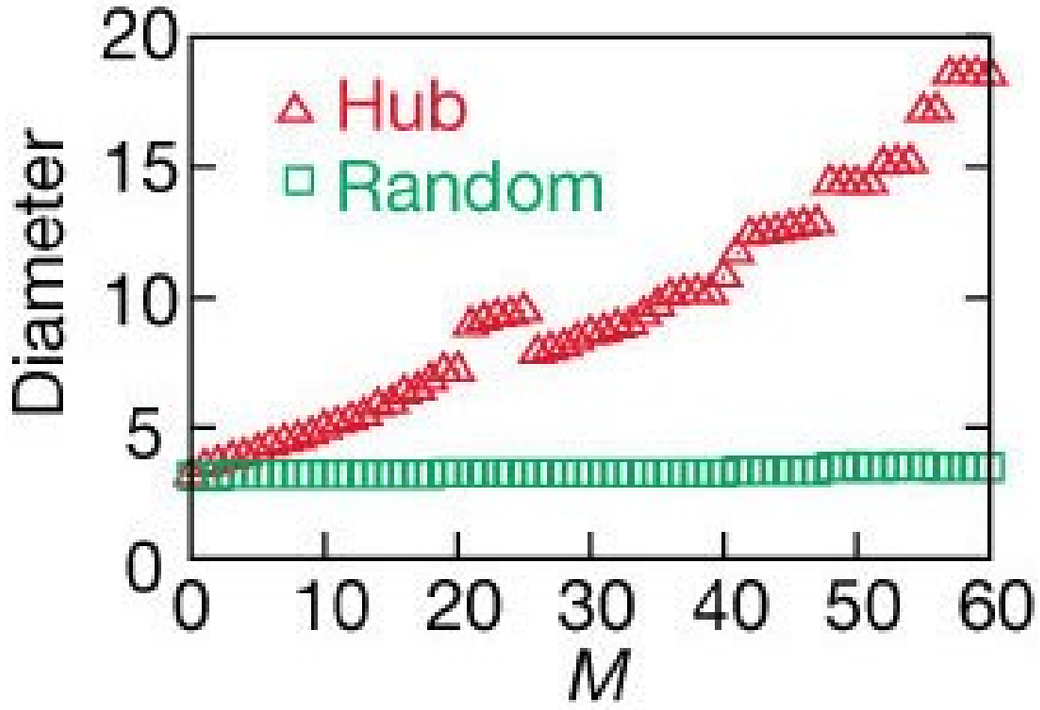
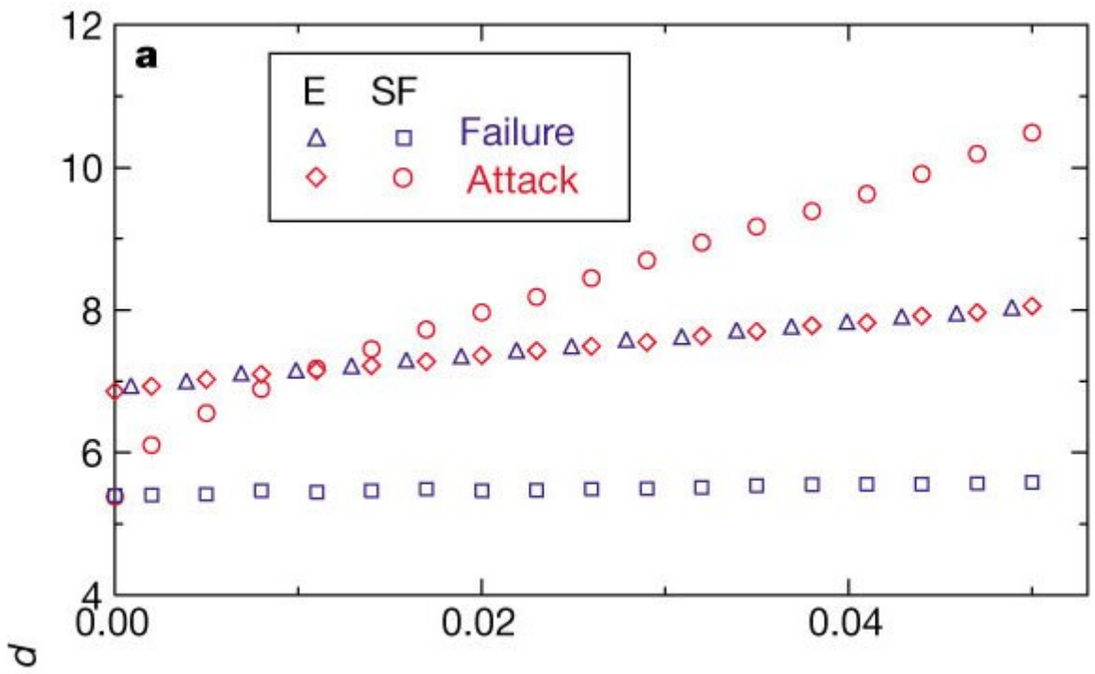
Algorithms in Nature

Robustness in biological systems

Failure and attacks on networks





- Is this okay?
  - From the perspective of an attacker?
  - From the perspective of the biological system?

Essentiality / Fragility

- Of the 5796 genes in yeast, 1122 (19.4%) are *essential* or *fragile*
  - A single KO of any essential gene kills the cell, i.e. results in failure of the network

- Where are they located in the network?
- Can we predict how fragile a node is based on its topology?
- Why are these genes “not protected”?

Predicting gene essentiality using network topology

What features should we use?

correlating a gene's topological feature with essentiality (1=essential, 0=not essential)

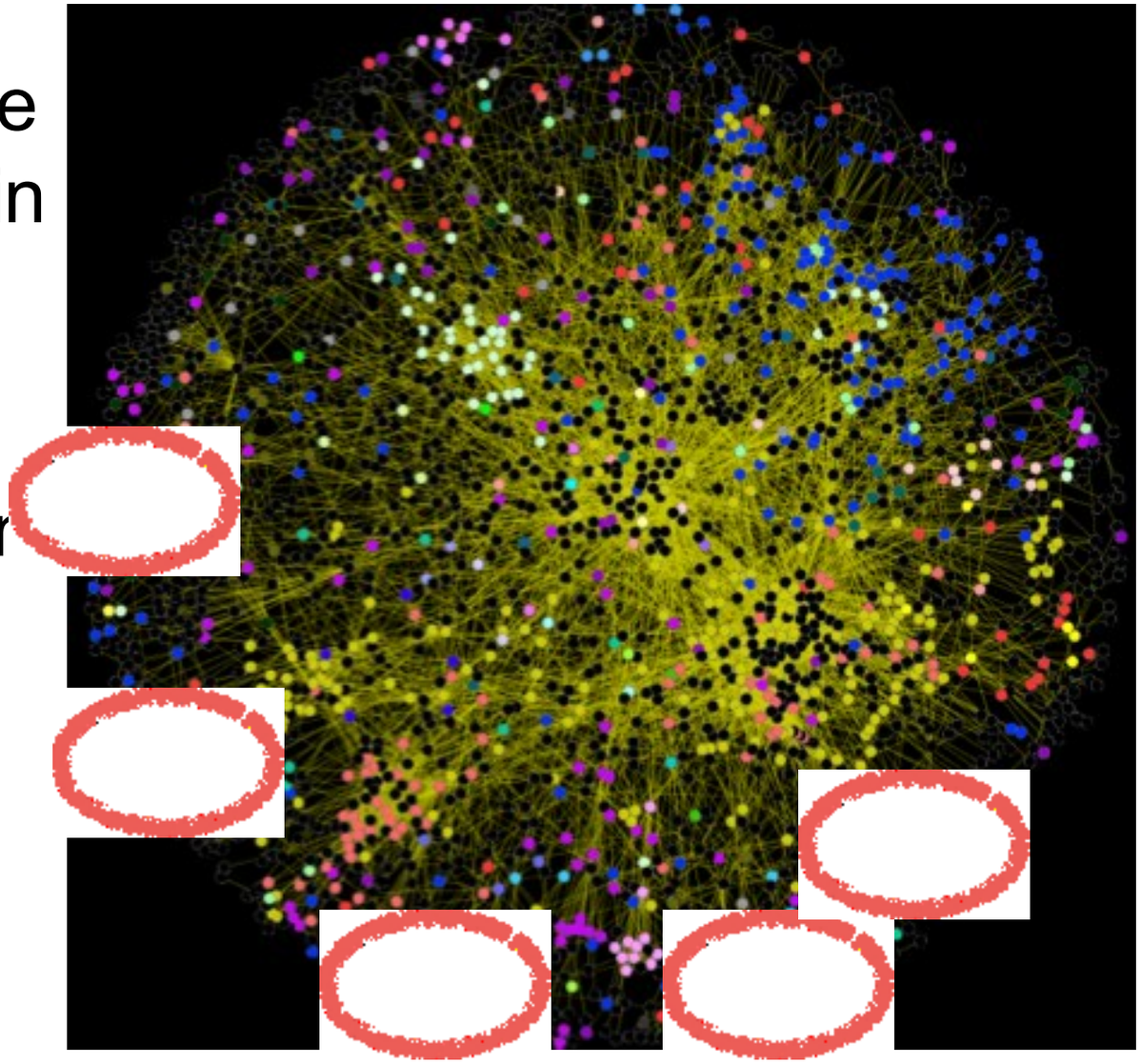
	Network	
	0.352	
PageRank	0.363	
Centrality	0.314	

The higher a gene's degree or the more "central" it is, the more likely that gene is essential

How can the biological system improve this?

Biological modules

- The previous correlations were using features computed within the *global* interaction network
- But most processing occurs within localized *modules* within the network
- A set of proteins that are all involved in a similar biological process, function, or complex



Predicting gene essentiality using network and module-level topology

correlating a gene's topological feature with essentiality (1=essential, 0=not essential)

	Network	Module
Degree	0.352	0.497
PageRank	0.363	0.404
Centrality	0.314	0.385

Consistently higher correlation with module topology than with global topology

The higher a gene's degree or the more "central" it is, the more likely that gene is essential

A gene's essentiality depends both on its module (its function) and its topological role within the module

Modeling the spread of noise

- When a node is attacked, nearby nodes are also affected
  - On the internet: computer virus attacks
  - In biology: environmental and signaling noise,which is more common than knock-outs

- Infect value of a gene u = the % of nodes in the module or network that become “infected” with a virus that begins at u and proceeds using a susceptibility-infectious model

Predicting gene essentiality using network and module-level topology

correlating a gene's topological feature with essentiality (1=essential, 0=not essential)

	Network	Module
Degree	0.352	0.497
PageRank	0.363	0.404
Centrality	0.314	0.385
Infect	0.302	0.453

Consistently higher correlation with modules than with the global topology

The higher a gene's degree or the more "central" it is, the more likely that gene is essential

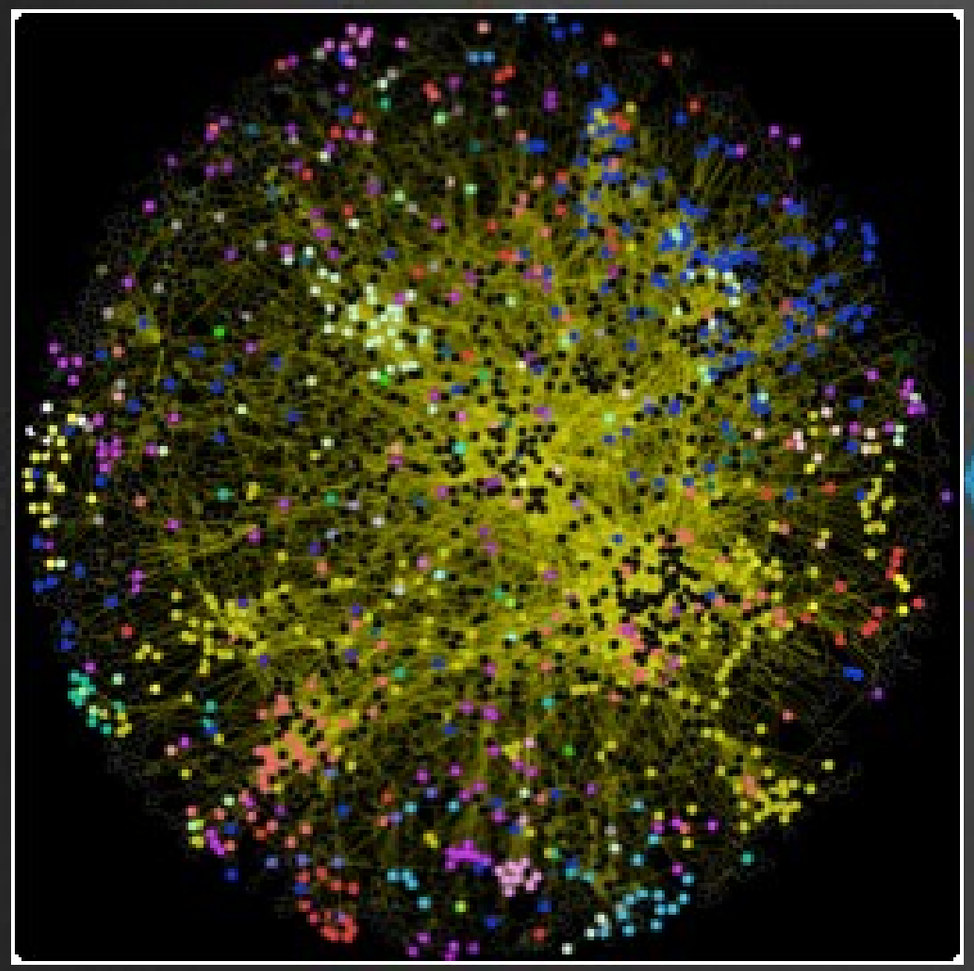
When noise spreads from an essential node, many other nodes are affected

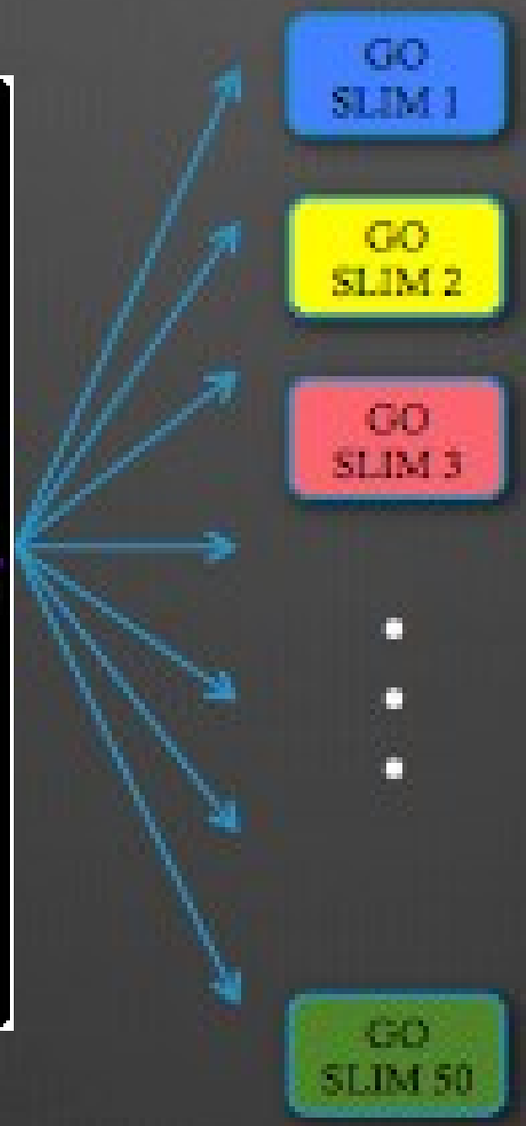
Robust and fragile modules

- We established that robustness is a module-level property
- Is essentiality distributed “equally” across all modules?
  - If not, are robust and fragile modules designed “equally”?
    - If not, what features can distinguish robust from fragile?

- Module essentiality = % of genes in the module that are essential
  - High module essentiality ⇛ many essential genes ⇛ not robust
  - Low module essentiality ⇛ few essential genes ⇛ very robust

1. Divide PPI network into module subnetworks





2. Extract subnetwork features

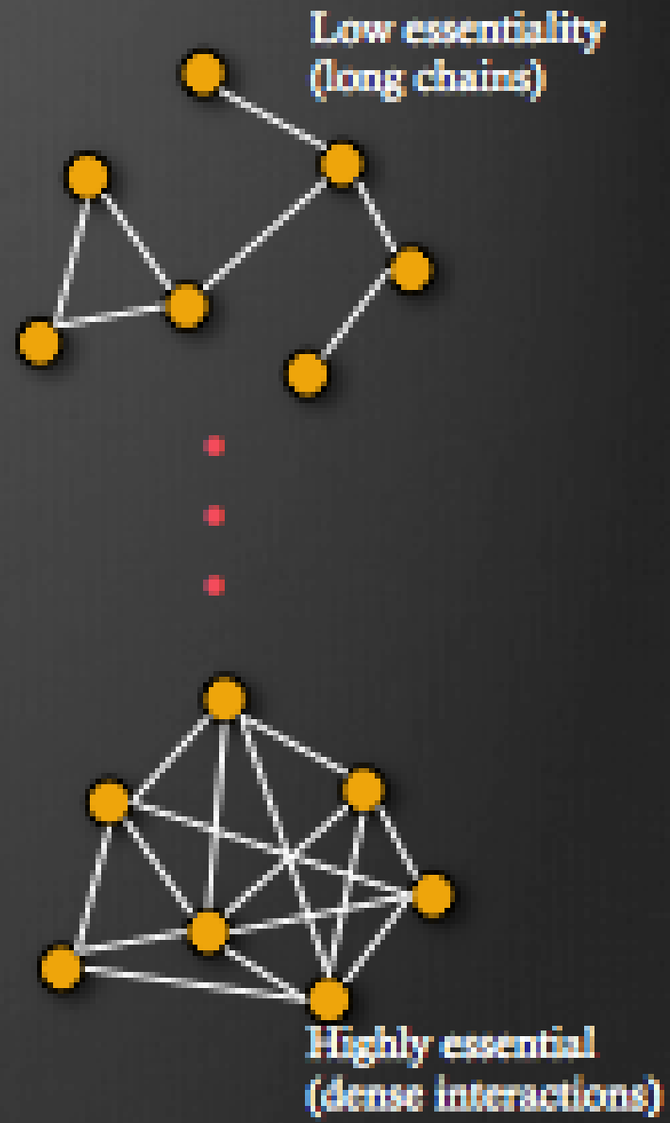
Ess %	Feat 1	Feat 2	...	Feat k
0.58				
0.10				
0.87				
...				
0.23				

3. Do any features correlate highly with essentiality?

Network features

- Eigenvalue: 1st (largest) eigenvalue of the adjacency matrix, normalized
- Degree: average degree of a node
- MIS: size of the maximal independent set
- Articulation points: number of nodes whose removal disconnects the graph
- APSPL: average shortest path length between any two nodes
- Triangle: number of triangles
- Density: num edges / num possible edges
- Clustering coefficient: number of triangles / number of possible triangles
- And others: Betweenness centrality, K-core size, Cover time, Assortativity, Modularity, etc.

GO Term	Essentiality	Density	NmEval
cellular ion homeostasis	0.089	0.0241	0.0716
endosomal transport	0.090	0.0346	0.0558
cell wall organization	0.095	0.0087	0.0181
precursor metabolites and energy	0.105	0.0363	0.0803
peroxisome organization	0.118	0.0689	0.1196
…	…	…	…
rRNA processing	0.664	0.0996	0.1252
transcription initiation	0.667	0.1613	0.2135
transcription RNA PIII	0.692	0.2510	0.3217
snoRNA processing	0.759	0.1991	0.2807
ribosomal subunit export	0.805	0.1585	0.1698
Correlation with essentiality:		0.699	0.647



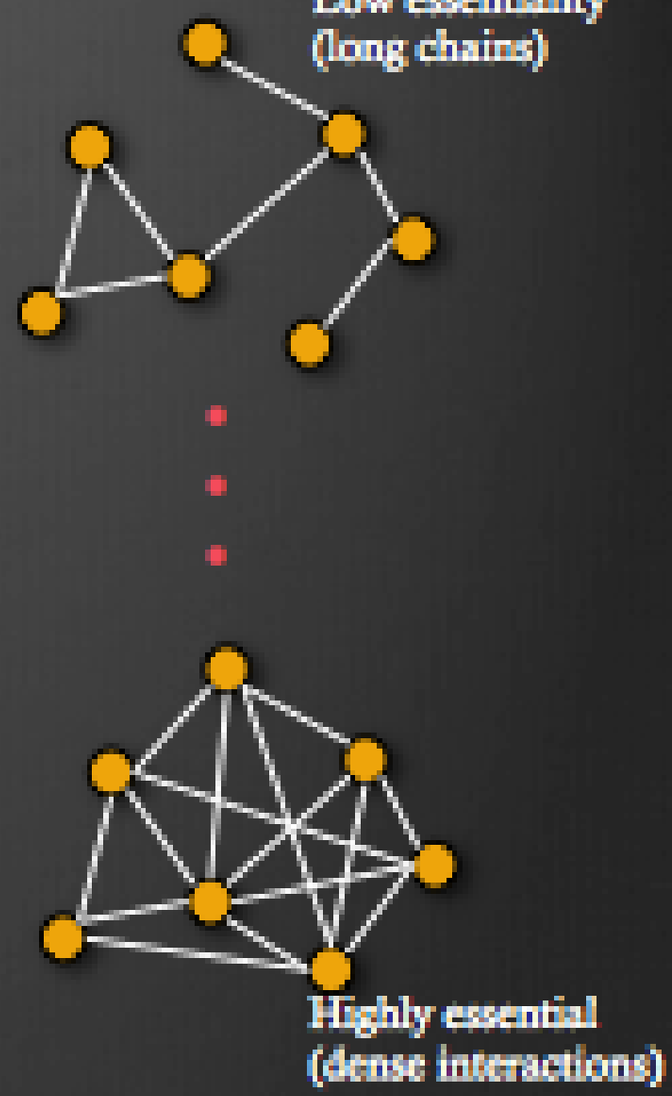
Why is the eigenvalue a good predictor of noise take-over?

Under many models of virus propagation, it was proved that the epidemic threshold $= 1/\lambda_{1,A}$ [Charkrabarti+ 2008]

Mostly *external* cellular processes

Mostly *internal*

GO Term	Essentiality	Density	NmEval
cellular ion homeostasis	0.089	0.0241	0.0716
endosomal transport	0.090	0.0346	0.0558
cell wall organization	0.095	0.0087	0.0181
precursor metabolites and energy	0.105	0.0363	0.0803
peroxisome organization	0.118	0.0689	0.1196
…	…	…	…
rRNA processing	0.664	0.0996	0.1252
transcription initiation	0.667	0.1613	0.2135
transcription RNA PIII	0.692	0.2510	0.3217
snoRNA processing	0.759	0.1991	0.2807
ribosomal subunit export	0.805	0.1585	0.1698
Correlation with essentiality:		0.699	0.647

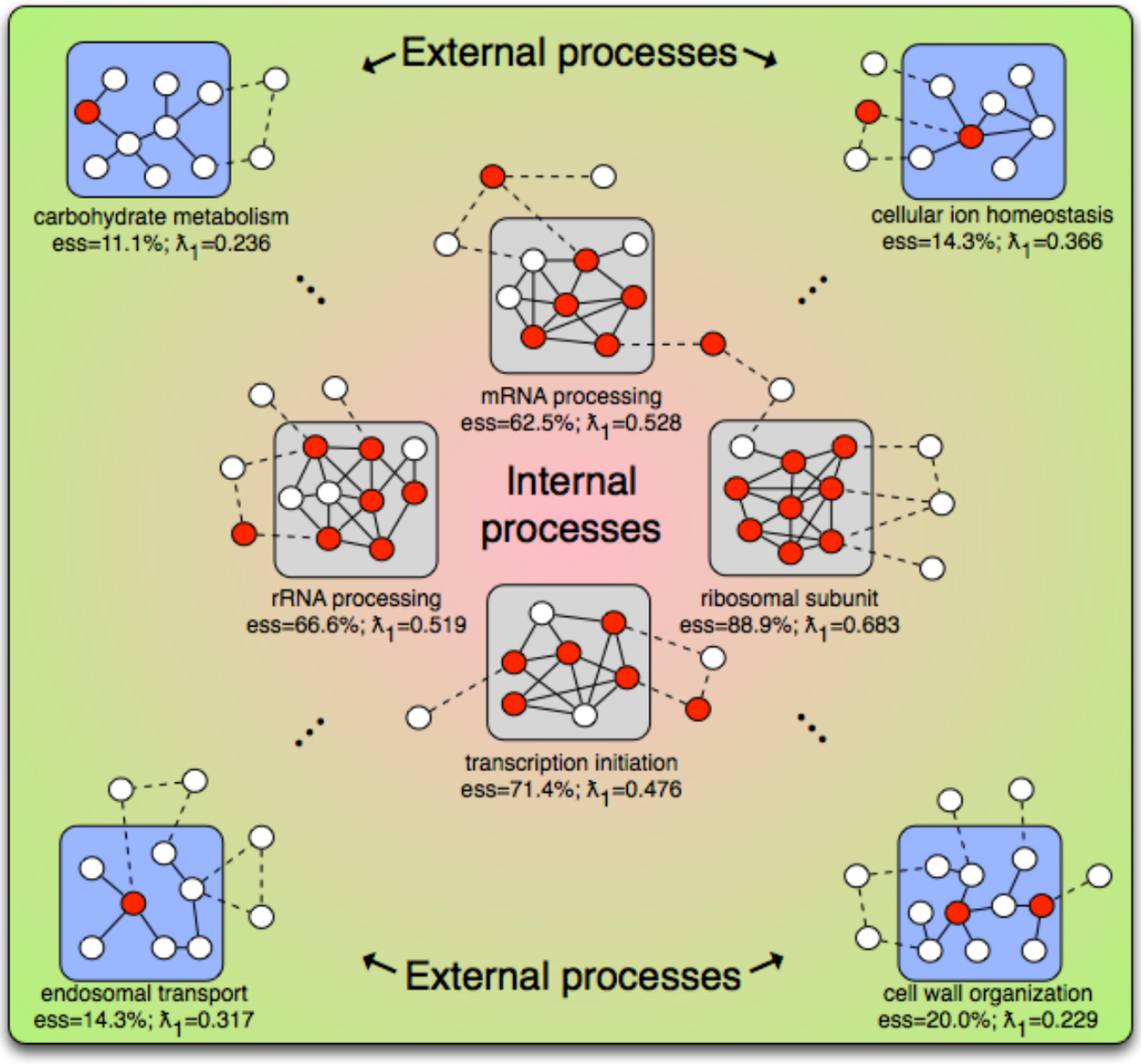


Why is the eigenvalue a good predictor of noise take-over?

Under many models of virus propagation, it was proved that the epidemic threshold $= 1/\lambda_{1,A}$ [Charkrabarti+ 2008]

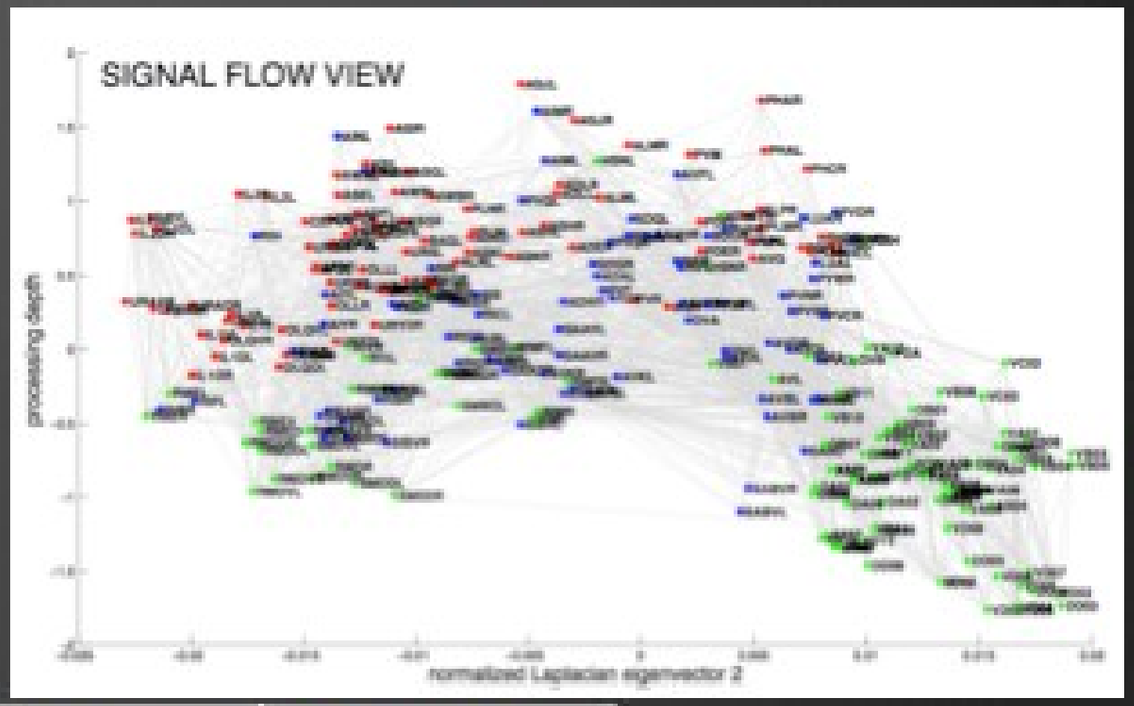
- Is essentiality distributed “equally” across all modules? NO
  - If not, are robust and fragile modules designed “equally”? NO
    - If not, what features can distinguish robust from fragile?

Internal: more connections
External: fewer connections



C. elegans neuronal network

- 279 nodes, 2990 synapses
- 3 neural types: external (sensory / motor) and internal (inter)
- Divided into 8 modules (ganglion)



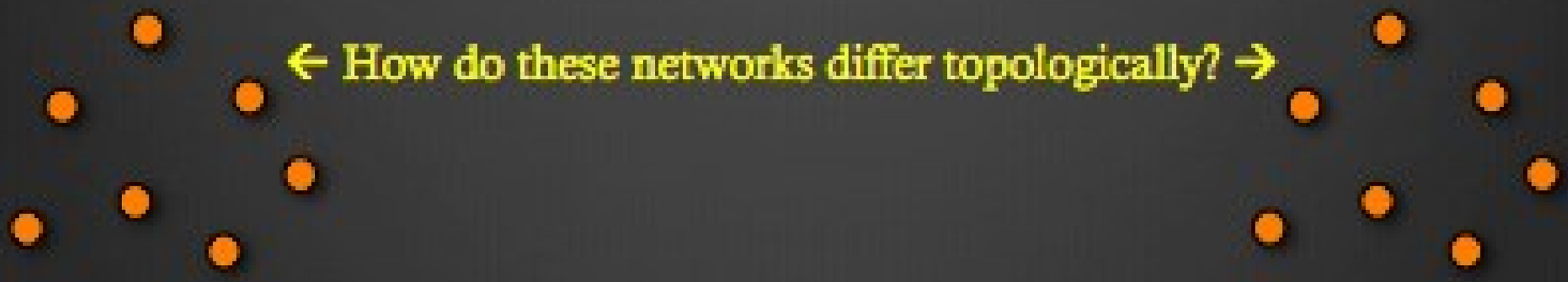
Ganglion	External %	Density	NmEval
Ventral	40.6		
Lateral	56.2		
Posterolateral	57.1		
…			
Ventral cord	100.0		

Correlation with external %: -0.537 -0.703

Metabolic networks

- Bacteria thrive in diverse environments, which are characterized by their **niche breadth**:
  - A score from 1 (narrow and stable environment) to 5 (highly complex and dynamic) indicating the diversity and fluctuation of metabolites under which the organism's metabolism must function.

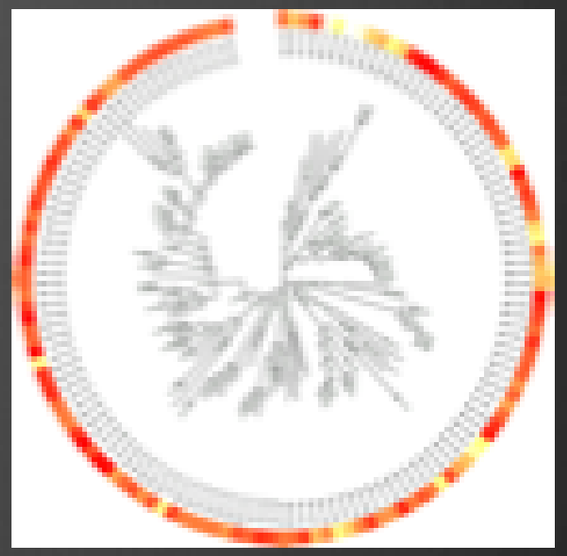
← How do these networks differ topologically? →



Breadth 1: narrow & stable … Breadth 5: complex & dynamic

We obtained 75 bacterial metabolic networks [Morine+ 2009] and looked for topological features that correlated with their niche breadth.

Species	Niche Breadth	Density	NmEval
Borrelia burgdorferi	1	0.0061	0.0112
Treponema pallidum	1	0.0055	0.0126
Buchnera aphidicola	1	0.0044	0.0101
…			
Streptomyces coelicolor	5	0.0020	0.0057
Caulobacter crescentus	5	0.0018	0.0059
Bacillus halodurans	5	0.0019	0.0058

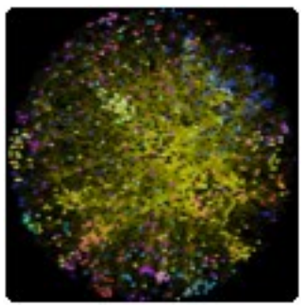


Correlation with niche breadth: -0.421 -0.431

More noise (higher niche breadth) → higher robustness

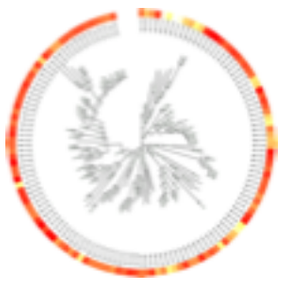
Less noise (lower niche breadth) → higher efficiency

3 case studies from biology



- Yeast protein-protein interaction network
  - Internal modules: more essential, protected ⇛ less need to buffer noise ⇛ higher connectivity
  - External modules: less essential, more exposed ⇛ need to buffer noise ⇛ lower connectivity

- C. elegans neural network:
  - Internal ganglion: integrates signals and coordinate responses efficiently ⇛ higher connectivity
  - External ganglion: deal with variable signals ⇛ buffer noise via lower connectivity



- Bacterial metabolic networks:
  - Stable environments: higher, efficient connectivity
  - Variable environments: lower, robust connectivity

Module-dependent topologies

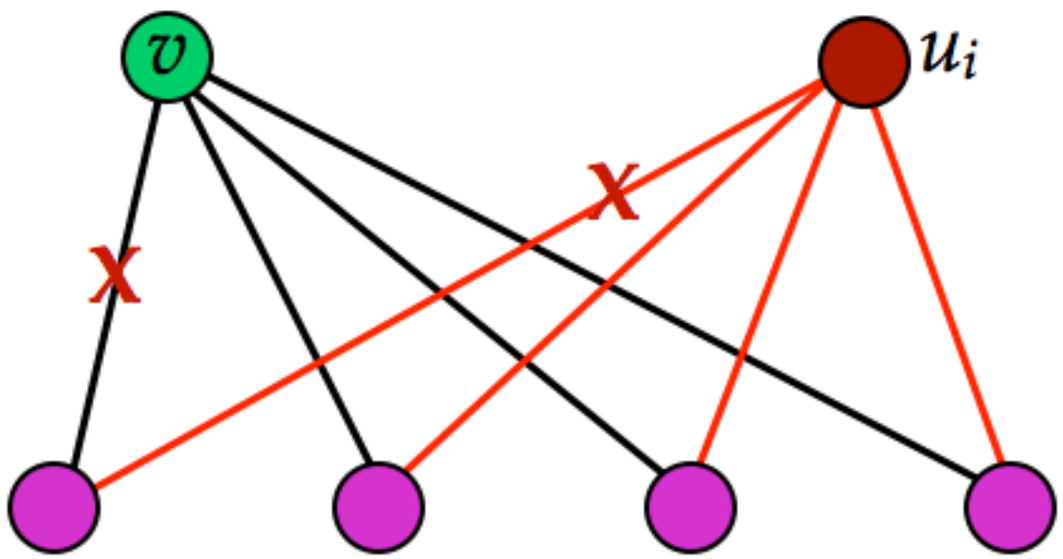
- If topology depends on the module, what does this say about the models we discussed? (preferential attachment, duplication-based, etc)
- How do we generate networks with module-topologies adjusted based on its “environmental exposure”?

DMC – Duplication, Mutation with Complementarity

Choose a node v at random, and connect u_i to all the neighbors of v. u_i is now a "clone" of v.

For every added edge, decide to delete with probability q_{del}

If you decide to delete, delete the **new or corresponding old** edge (choosing which one by flipping a coin):



How to adapt this model?

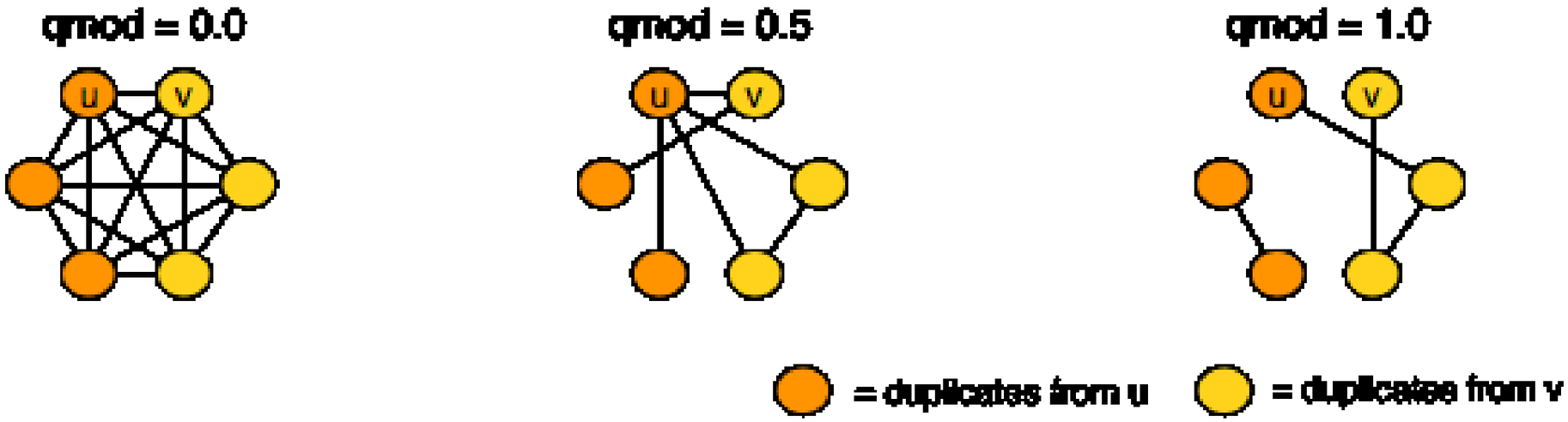
Finally, with probability q_{con}, add an edge (v, u_i)

Slide from Carl Kingsford

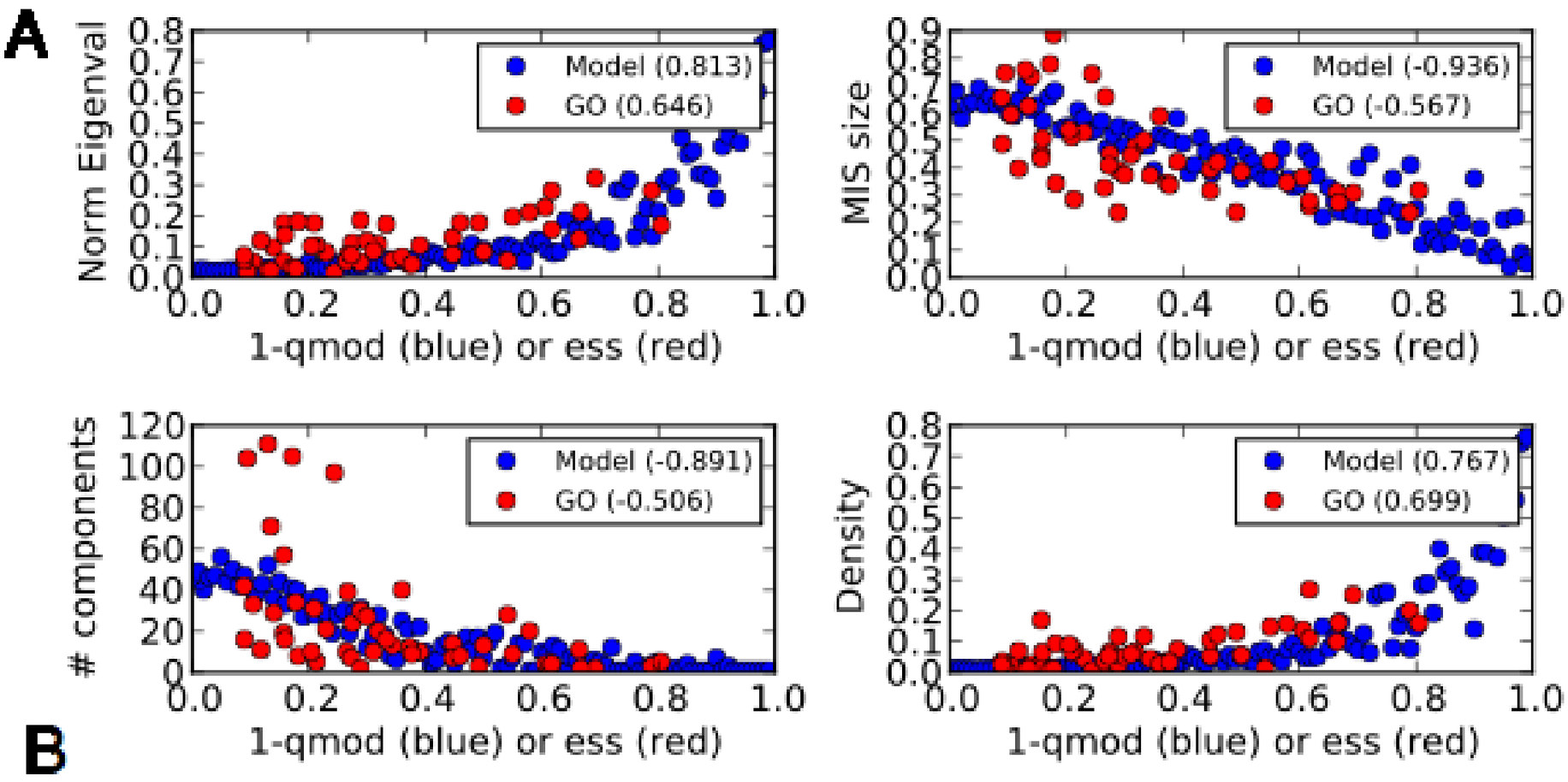
Module-dependent topologies

Stable, internal environment

Variable, external environment



Module-dependent topologies

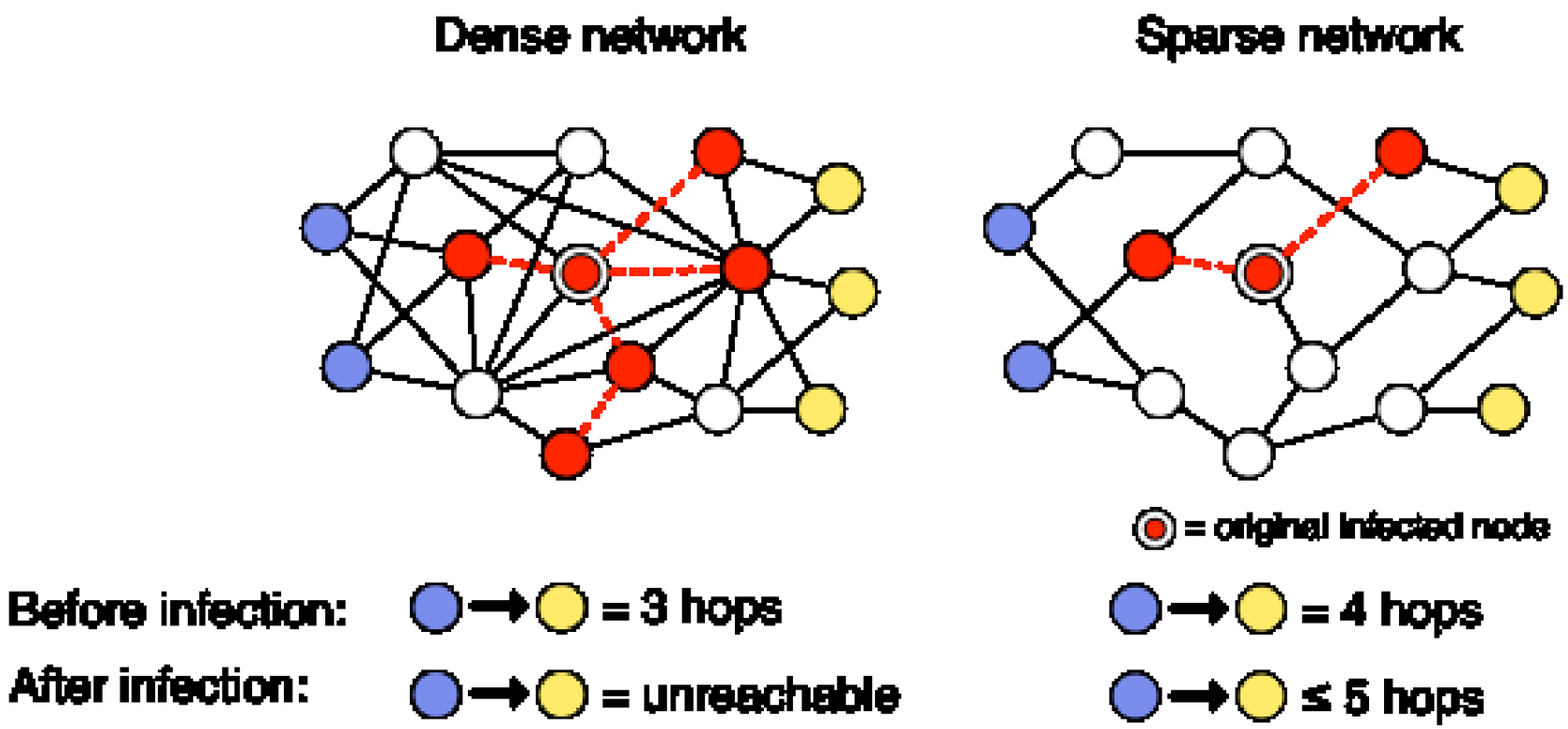


Similar diversity of features across real biological modules (red) and model-based modules using different values of qmod (blue)

Carrying these insights to CS..

- Internet is regularly targeted with worms that compromise machines
- Typically, infected machines are detected following an attack and then isolated for maintenance (e.g. wipe and reinstall OS)
- How does such removal affect the ability of the remaining nodes to communicate? This requires a delicate balance:
  - Very dense connectivity ⇛ everyone gets infected
  - Very sparse connectivity ⇛ worm will break the network apart

Measuring residual connectivity



Identifying vulnerable nodes and modules in real-world networks

Vulnerables nodes: nodes that would result in lots of damage if infected

Vulnerable modules: modules that would be quickly swamped by noise if infected

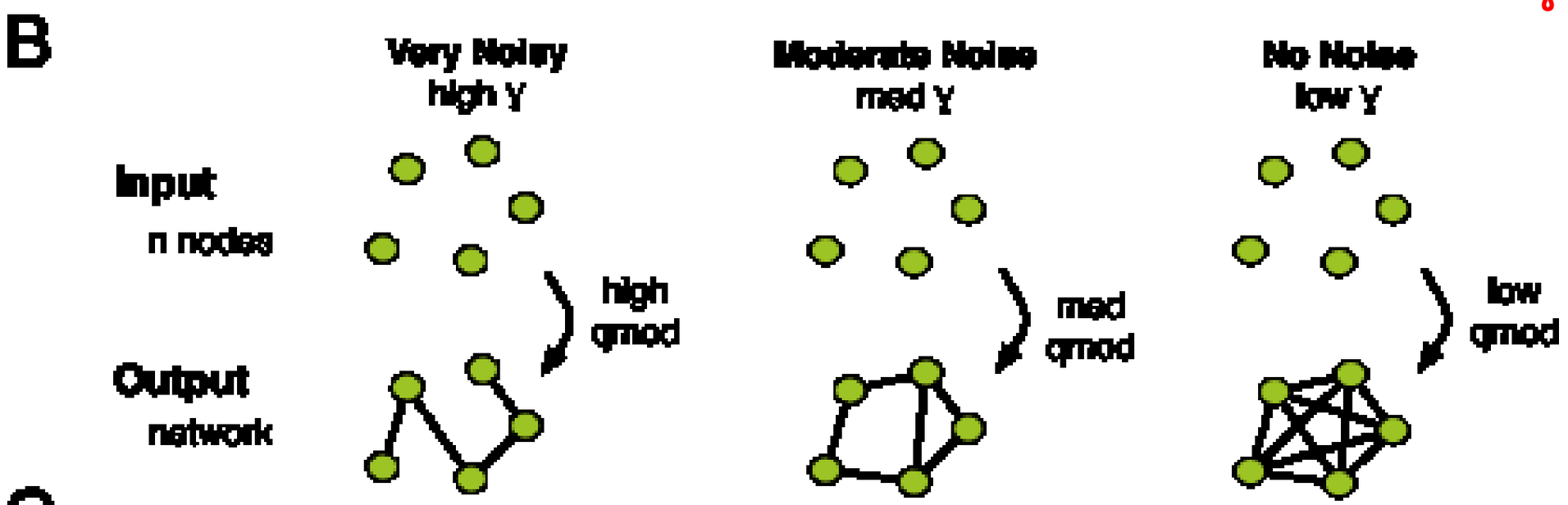
	Residual connectivity vs Infect size	Residual connectivity vs Eigenvalue
Powergrid	0.721	0.944
Internet	0.669	0.846

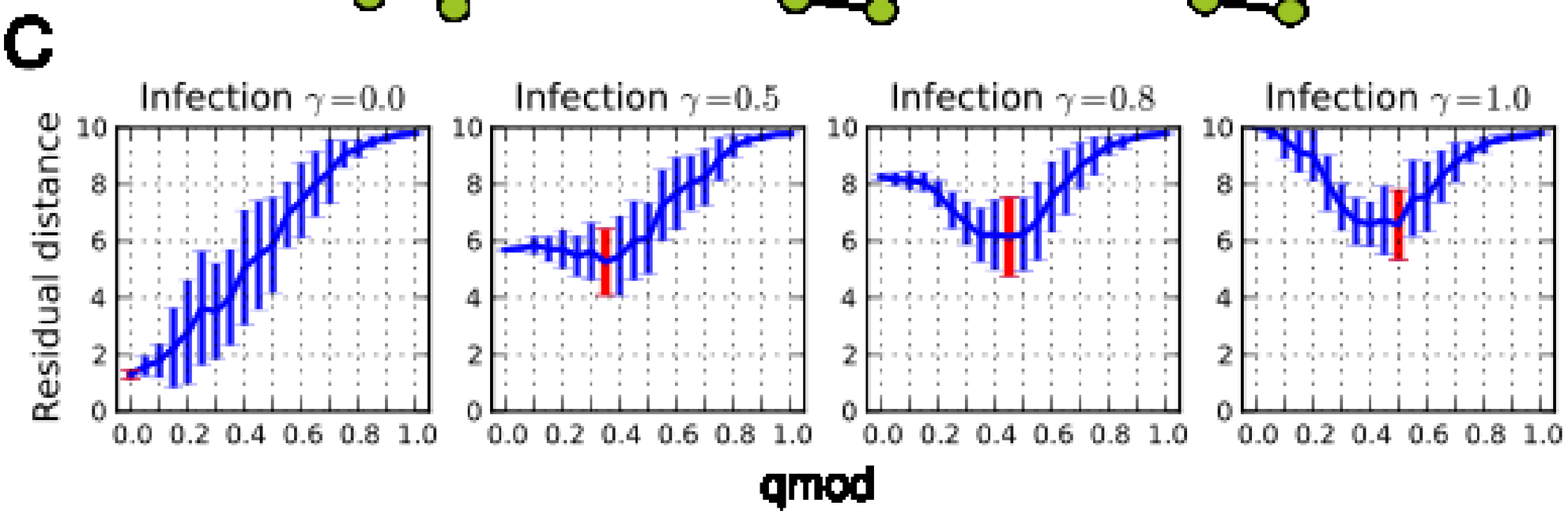
- Unclear if these are true vulnerabilities or if they represent protected/internal parts of the system (a nice project to investigate this further..)

Project Idea

Designing networks specifically tailored for different environments

γ = probability a node will be attacked

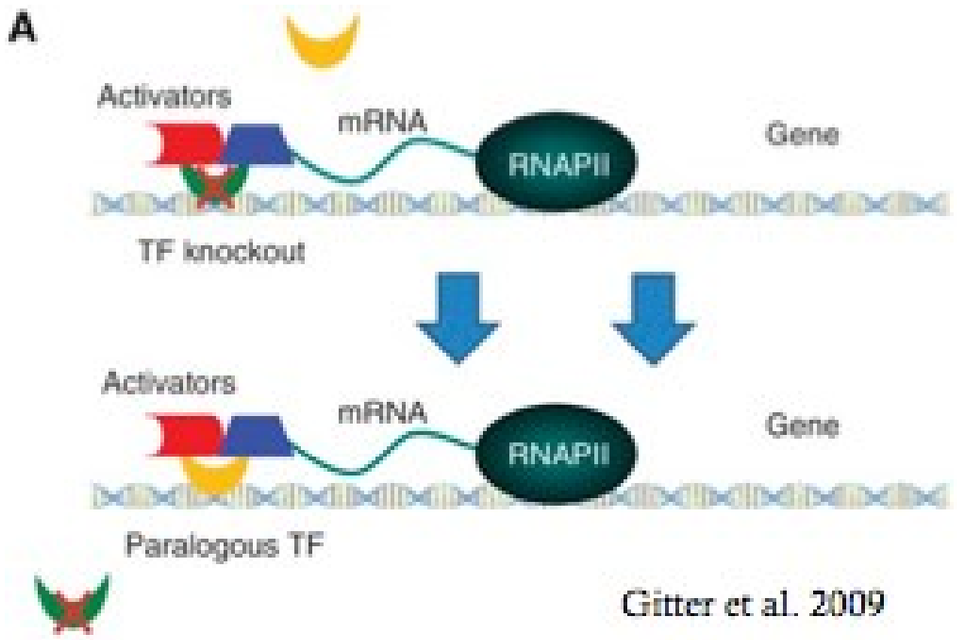




Aside: backup mechanisms

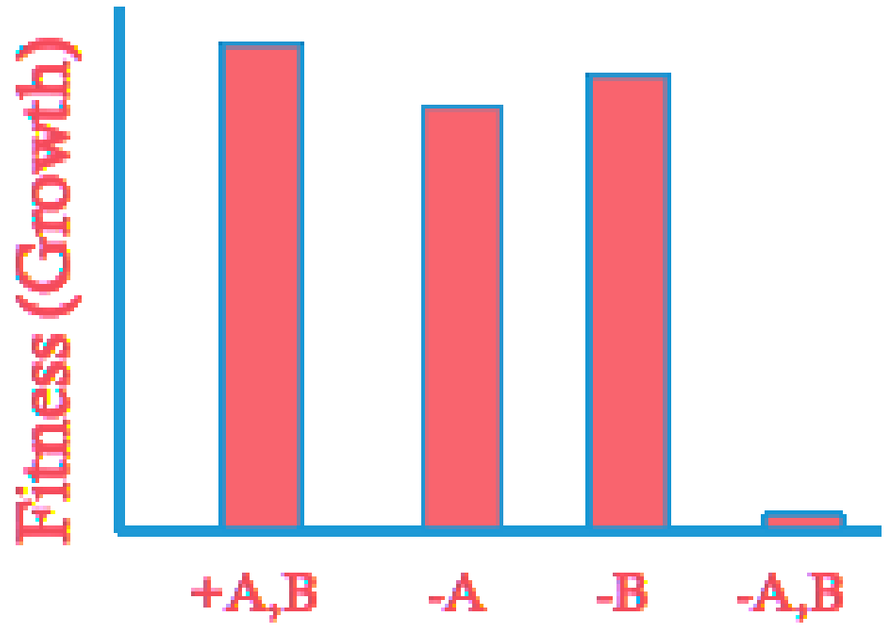
How does the cell deal with the loss of non-essential genes?

Backup in regulatory networks



Paralogous TFs compensate for one another

Backup in interaction networks



Genetic interactions: double KO confers larger phenotypic effect than expected from single KOs

Conclusions

Biology:

* the most vulnerable points are in physically hard to reach places
* the most exposed points are built to be robust to spreading noise

Computer science:

* similar trade-offs are desired and should reflect the design
* generative model to produce environment-dependent topologies
* benchmark to measure the robustness of a module or network