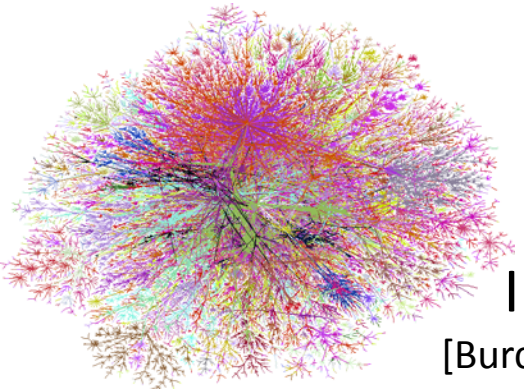
生物信息学与系统生物学

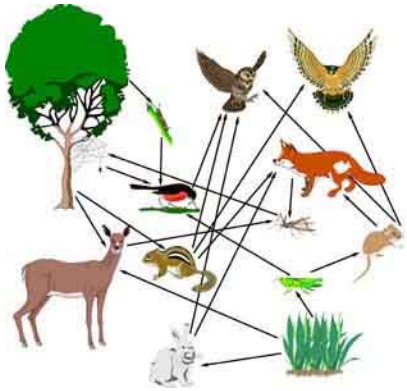
张世华

中国科学院数学与系统科学研究院

Networks as a universal language

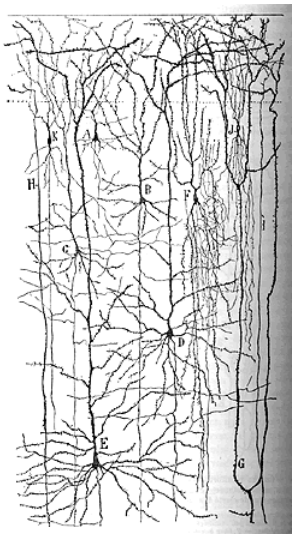


Internet
[Burch & Cheswick]

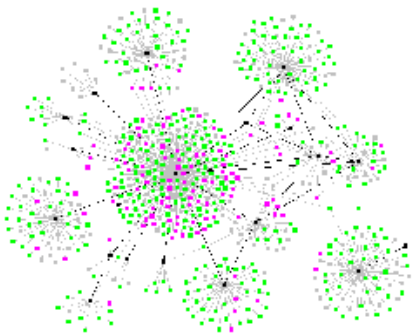


Food Web

Electronic Circuit

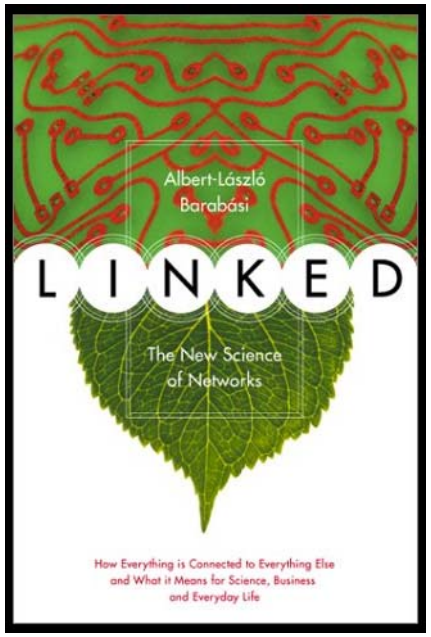


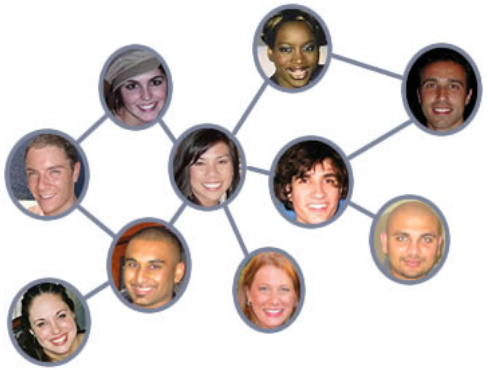
Neural Network
[Cajal]



Disease Spread
[Krebs]

Protein Interactions
[Barabasi]





Social Network

Reduction of Networks

- Networks are powerful!
- Networks are complicated!
- Can we reduce the network?

RESEARCH ARTICLE

COMPUTATIONAL BIOLOGY

Reduction of Complex Signaling Networks to a Representative Kernel

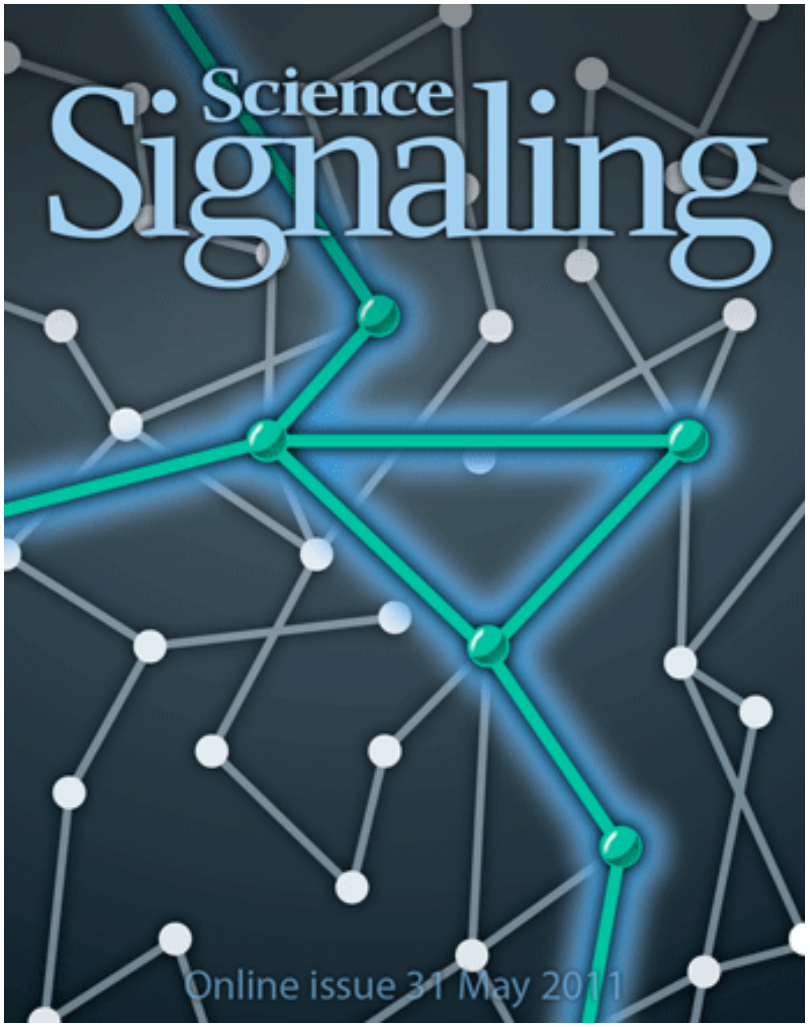
Jeong-Rae Kim,[1,2] Junil Kim,[1] Yung-Keun Kwon,[1,3] Hwang-Yeol Lee,[1] Pat Heslop-Harrison,[4] Kwang-Hyun Cho[1]*

Editor's Summary: Reducing Complexity

- The large and complex nature of the biochemical regulatory networks that govern cell behavior **provides a major challenge** to the systematic analysis of cell signaling.
- However, **most processes that reduce network complexity fail to reproduce the dynamic properties of the original network**. **Kim *et al*. describe an algorithmic approach to network reduction and simplification that preserves the dynamics of the network**.
- They applied their approach to several networks in species from bacteria to humans, producing simplified networks called "**kernels**". Examination of the genes represented by the kernel nodes provided **insight into the evolution of these core network genes**.
- Furthermore, the genes represented by the kernel nodes were enriched in disease-associated genes and drug targets, suggesting that this type of analysis may be therapeutically beneficial.

J.-R. Kim, *et al*. *Sci. Signal.* **4**, ra35 (2011).

Online cover



This week features a Research Article that describes an algorithmic approach to **simplifying complex signaling networks** and then **examines the properties of the nodes** in those simplified networks, which may have implications for *drug targeting*. The image shows an artist's rendition of a complex network overlaid with the simplified one.

Introduction: Data reduction

- Reduction

- Data reduction is the transformation of numerical or alphabetical digital information derived empirical or experimentally into a corrected, ordered, and **simplified form** (from wiki).

Introduction: Network **kernel**

- Kernel (computing), the central component of most operating systems (from wiki)
- We speculated that … networks were built around certain core structures or "**kernels**," which would be simpler to analyze without losing essential information.
- An individual kernel can be defined broadly as a simplified framework of a given complex interaction network that **preserves the dynamics and the output of the original network**.

How to study biological network?

Two general approaches:

(i) component-wise analysis of individual components in the networks, as in studies of "minimal gene sets"

One limitation of these component-wise approaches is that they cannot take into account regulatory interactions among the genes.

How to study biological network?

Two general approaches:

(ii) **computational analysis of simplified networks.**

One limitation of these simplified network approaches is that, by primarily focusing on preserving static topological properties of general complex networks, they **fail to preserve the dynamical properties** of cellular signaling networks.

How to study biological network?

More discussion:

The **spanning tree network reduction** approach reduces only the number of edges while preserving all the nodes of the original network. Because the resulting simplified network is a tree, it cannot preserve the dynamics of the original network if the original contains feedback loops (22–25) or feedforward loops (26, 27).

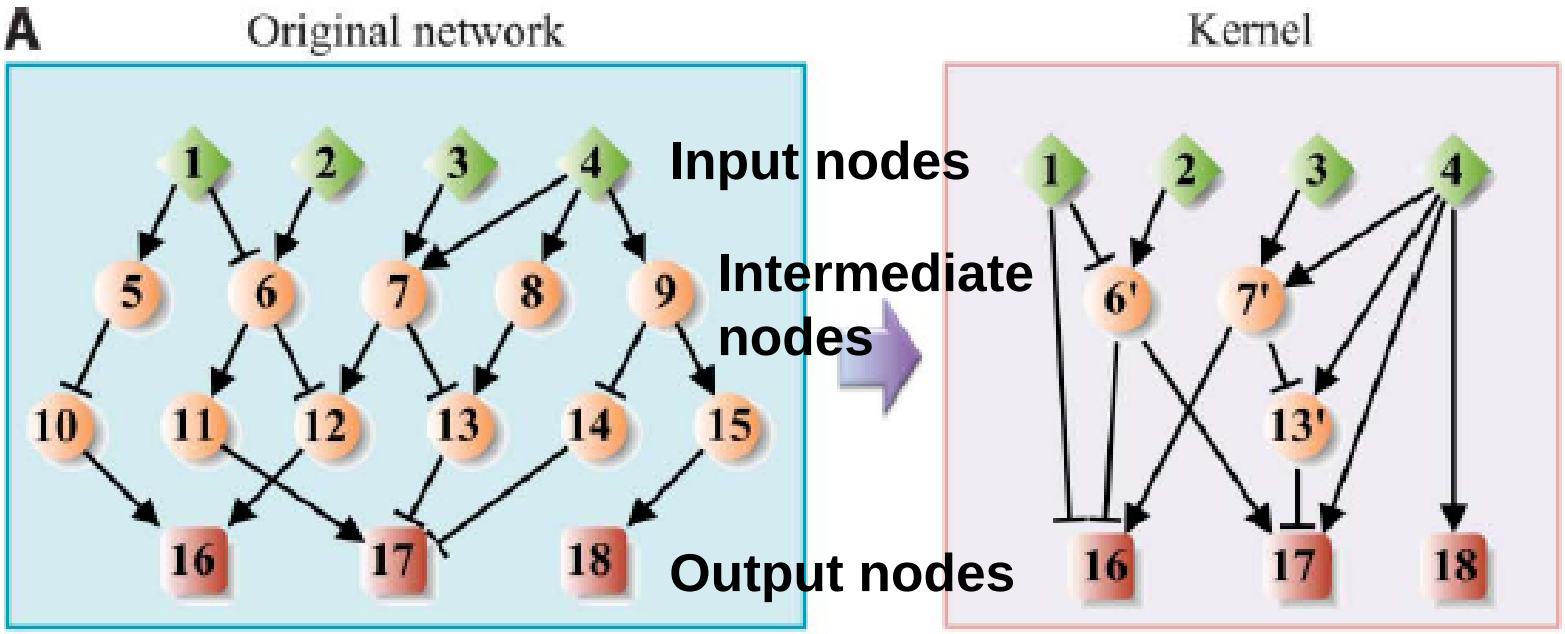
The approach taken by Itzkovitz et al. (12) **replaces network motifs** with CGUs, which in principle can preserve the dynamics of a network only if the intrinsic dynamics of each network motif are identically implemented in the CGU of the reduced network. However, it remains unclear how to implement such identical dynamics at each CGU.

Song et al. (18) proposed a reduction scheme that tiles a network with boxes such that the shortest path length of any two nodes in a box is less than a given number called a box size, where the size of the box is 1 + m, with m the maximum of the shortest paths between two nodes in the box. However, the resulting network does not contain any information on the direction or interaction type (activation or inhibition) of the edges; thus, preservation of dynamic properties is not possible.

Using network symmetry, Xiao et al. (19) proposed a network reduction scheme in which a set of nodes is grouped as one node if the rearrangement of their position within the set does not change the network topology. This approach can be effectively applied to a gene network containing many functionally redundant genes, but it is not effectively applicable to cell signaling networks that usually contain many long cascades.

Kernel identification algorithm

- an algorithm that identifies a kernel systematically by considering the relationship between a **network's structure** and its **dynamics**.



Kernel identification algorithm

- Can not simultaneously taking into account the dynamics of all possible subnetwork cases.

- To overcomes this difficulty by ***recursive sequential replacement of the neighborhood subnetwork of each node with a smaller one that preserved the same dynamics***.

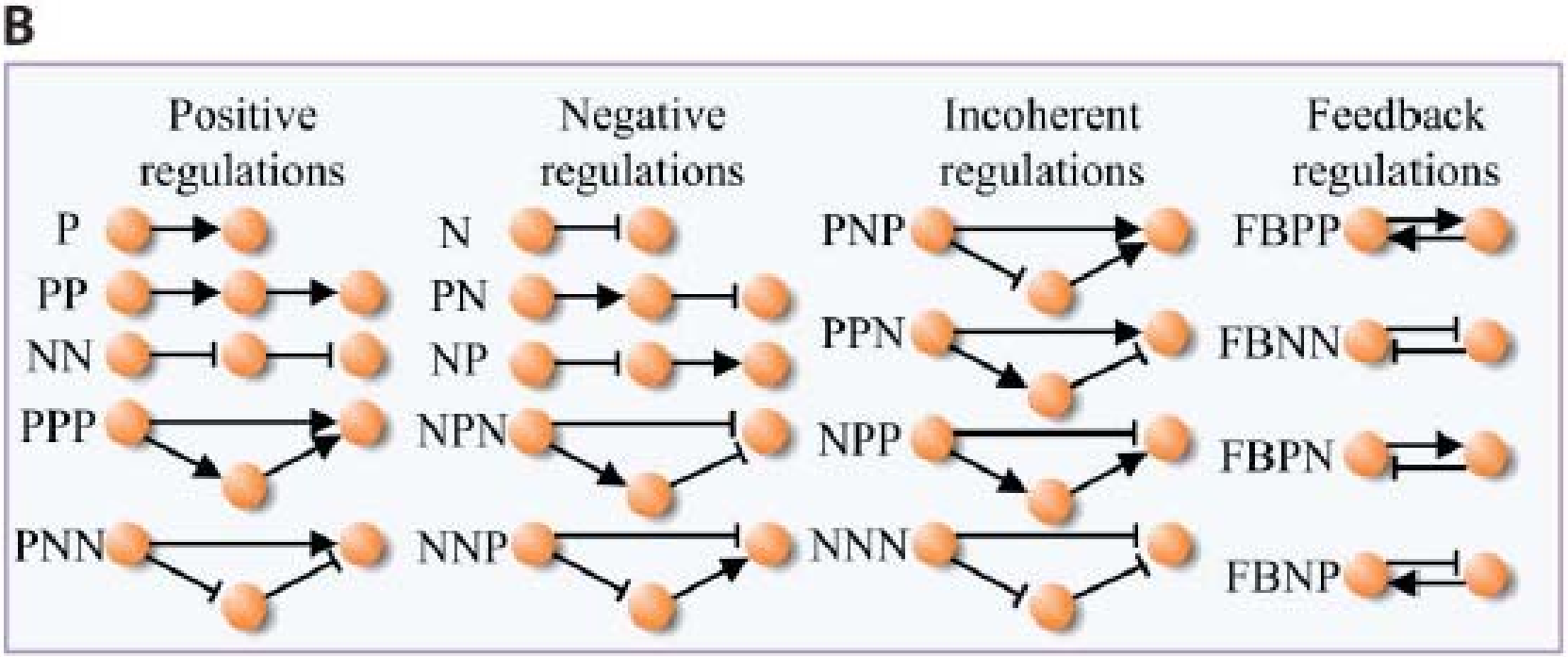
- Generally coarse-graining *fail to preserve the dynamical properties* of a network.

Kernel identification algorithm

- **Subnetwork replacement rule**:
- simulated the mathematical models of all two- and three-node networks with ordinary differential equations.
- then clustered the two- and three-node networks according to the similarity in their dynamics (Fig. 1B).
- We verified that the clustering assignments were similar between **linear and Hill-type mathematical models** and among the parameter values used.

Kernel identification algorithm

- simulated the mathematical models of all two- and three-node networks with ordinary differential equations.



(i) Linear models

We constructed linear models $\frac{dX}{dt} = AX + B$ for the two- and three-node networks (P, PP, NN, PPP, PNN, N, PN, NP, NPN, NNP, PNP, PPN, NPP, NNN, FBPP, FBNN, FBPN, FBNP) in Fig. 1B. For $X = (X_1, X_2, \cdots, X_n)$ $(n = 2 \text{ or } 3)$, the stimulus (S) was given on X_1 (see Supplementary Figure S13 for the detailed stimulus pattern) and, therefore, B is given by $B = (S, 0)^t$ or $B = (S, 0, 0)^t$. The matrix A for each network structure is given as follows:

P: $A = \begin{pmatrix} -a_{11} & 0 \\ a_{21} & -a_{22} \end{pmatrix}$, N: $A = \begin{pmatrix} -a_{11} & 0 \\ -a_{21} & -a_{22} \end{pmatrix}$, FBPP: $A = \begin{pmatrix} -a_{11} & a_{12} \\ a_{21} & -a_{22} \end{pmatrix}$, FBNN: $A = \begin{pmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{pmatrix}$

FBPN: $A = \begin{pmatrix} -a_{11} & -a_{12} \\ a_{21} & -a_{22} \end{pmatrix}$, FBNP: $A = \begin{pmatrix} -a_{11} & a_{12} \\ -a_{21} & -a_{22} \end{pmatrix}$

PP: $A = \begin{pmatrix} -a_{11} & 0 & 0 \\ a_{21} & -a_{22} & 0 \\ 0 & a_{32} & -a_{33} \end{pmatrix}$, NN: $A = \begin{pmatrix} -a_{11} & 0 & 0 \\ -a_{21} & -a_{22} & 0 \\ 0 & -a_{32} & -a_{33} \end{pmatrix}$, PN: $A = \begin{pmatrix} -a_{11} & 0 & 0 \\ a_{21} & -a_{22} & 0 \\ 0 & -a_{32} & -a_{33} \end{pmatrix}$,

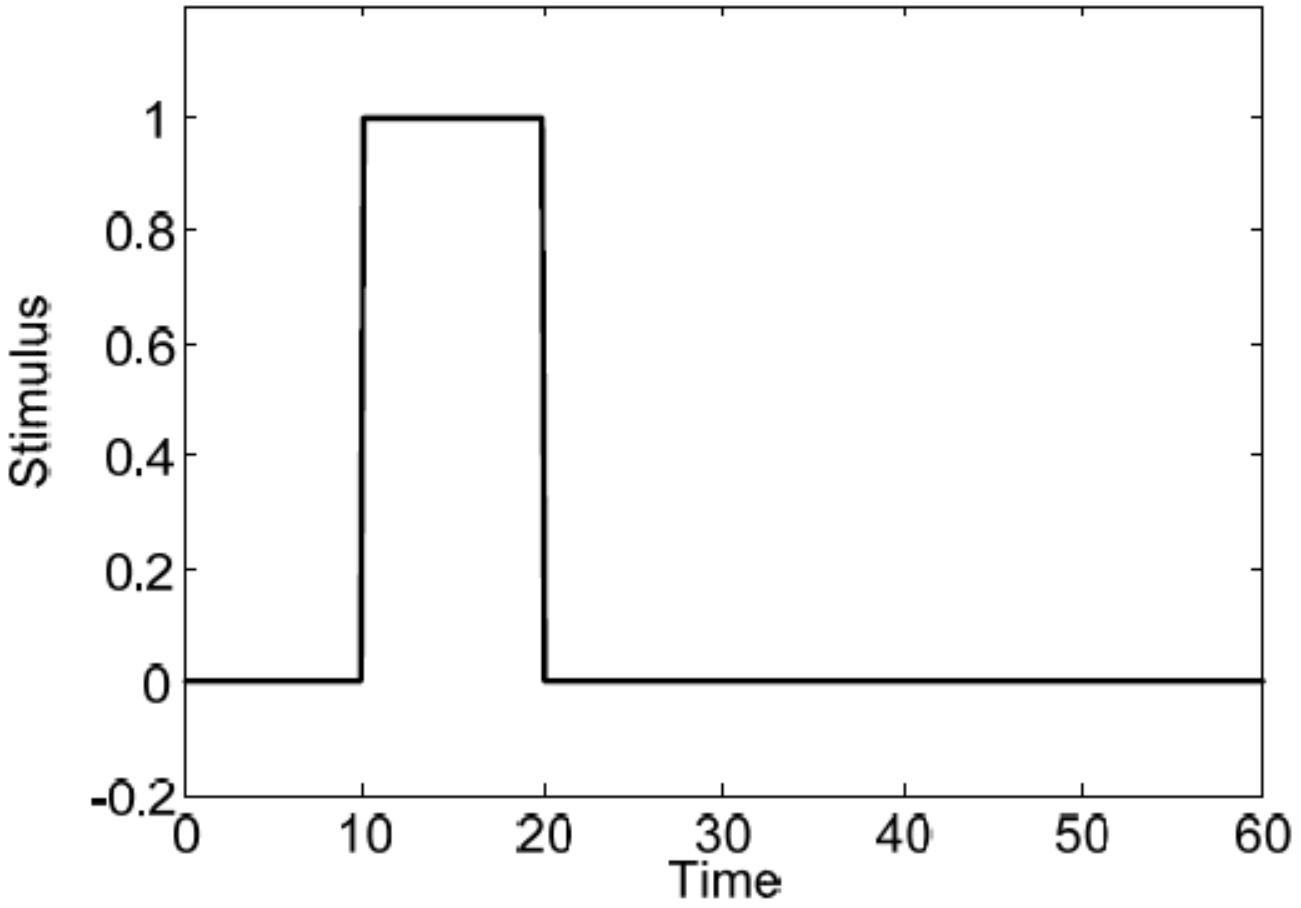
NP: $A = \begin{pmatrix} -a_{11} & 0 & 0 \\ -a_{21} & -a_{22} & 0 \end{pmatrix}$, PPP: $A = \begin{pmatrix} -a_{11} & 0 & 0 \\ a_{21} & -a_{22} & 0 \end{pmatrix}$, PNN: $A = \begin{pmatrix} -a_{11} & 0 & 0 \\ -a_{21} & -a_{22} & 0 \end{pmatrix}$

The initial point was assumed to be (0, 0) or (0, 0, 0) and the entries of *A were randomly* selected between 0 and 1 (if a matrix A is multiplied by a constant, then the time scale of the solution changes but the dynamical pattern does not change.

Stimulus pattern

B = (S,0)' or B = (S,0,0)'

The stimulus pattern used for the simulation of two- and three-node network models. This pattern was used for both linear models and nonlinear models.



Response patterns

Six representative response patterns used for classification of two- and three-node networks.

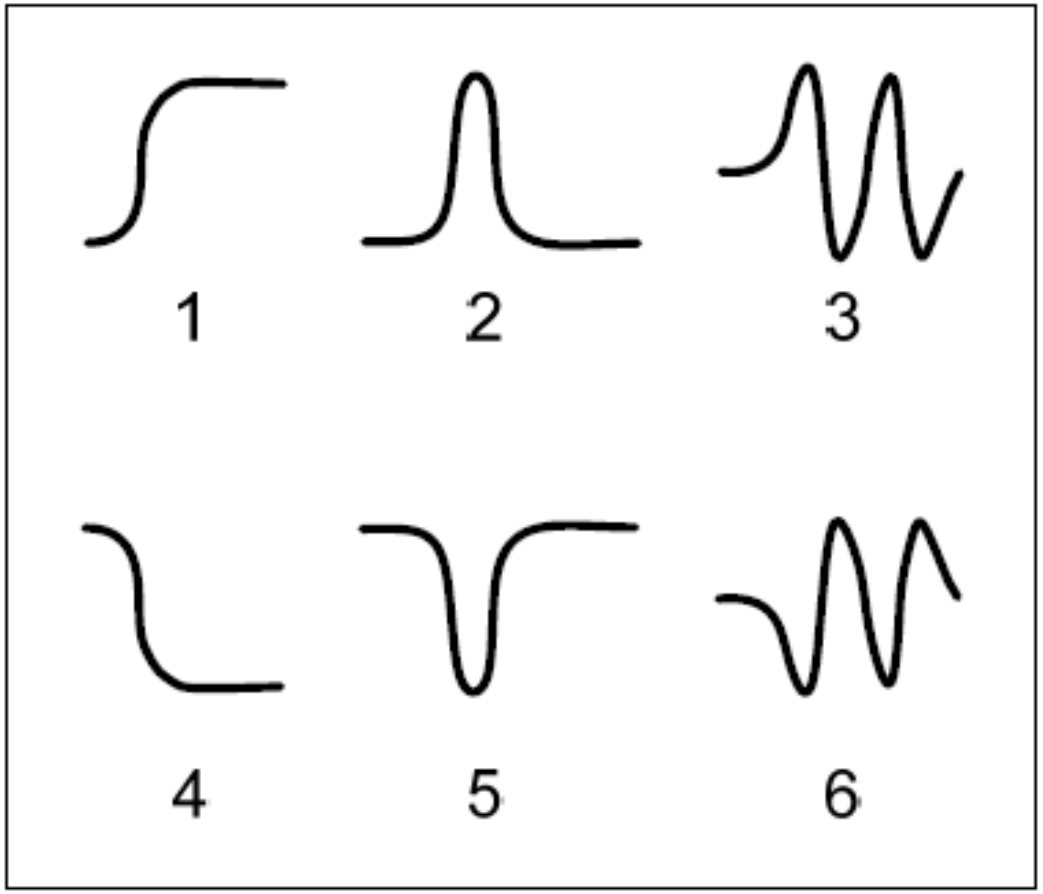
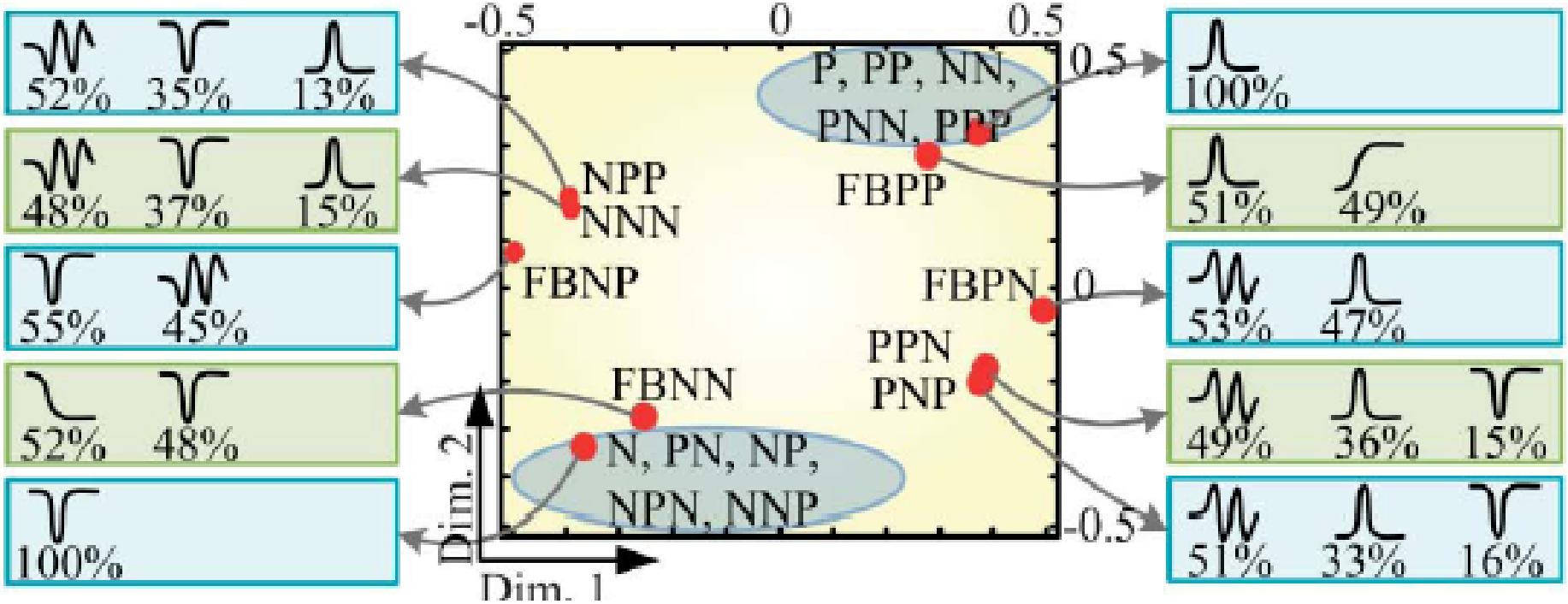


Table S10. Simulation results for linear models of 18 network structures. We repeated the simulation 1000 times for randomly selected parameter sets. The table shows the distribution of the six response types (see fig. S14 for the 6 response types) for the 1000 simulation results for each network.

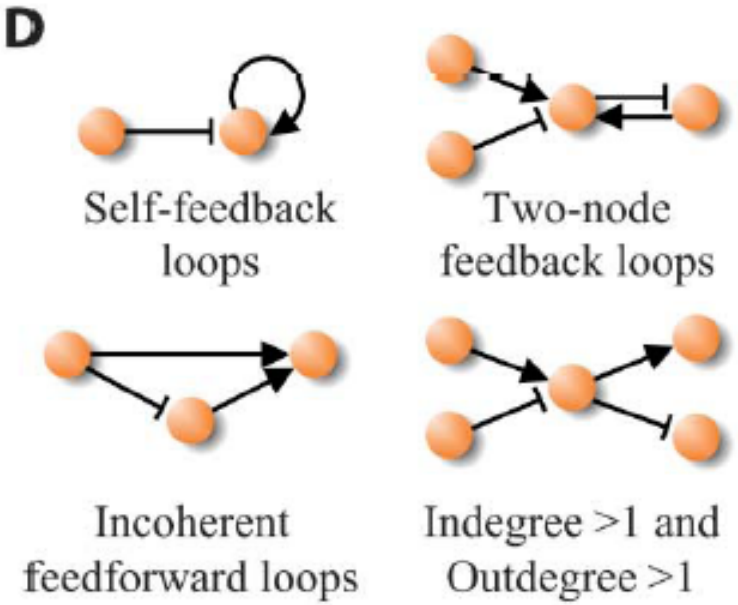
Network	Response type					
	1	2	3	4	5	6
P	0	1000	0	0	0	0
PP	0	1000	0	0	0	0
NN	0	1000	0	0	0	0
FF-PPP	0	1000	0	0	0	0
FF-PNN	0	1000	0	0	0	0
N	0	0	1000	0	0	0
PN	0	0	1000	0	0	0
NP	0	0	1000	0	0	0
FF-NPN	0	0	1000	0	0	0
FF-NNP	0	0	1000	0	0	0
FF-PNP	0	333	154	513	0	0
FF-PPN	0	362	147	491	0	0
FF-NPP	0	131	345	0	524	0
FF-NNN	0	153	372	0	475	0
FB-PP	0	513	0	0	0	0
FB-NN	523	0	477	0	0	0
FB-PN	0	473	0	527	0	0
FB-NP	0	0	451	0	549	0

Multi-dimensional scaling map for responses

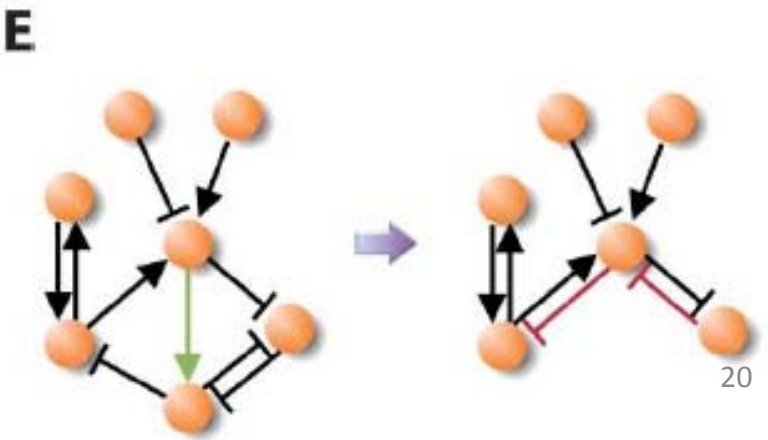


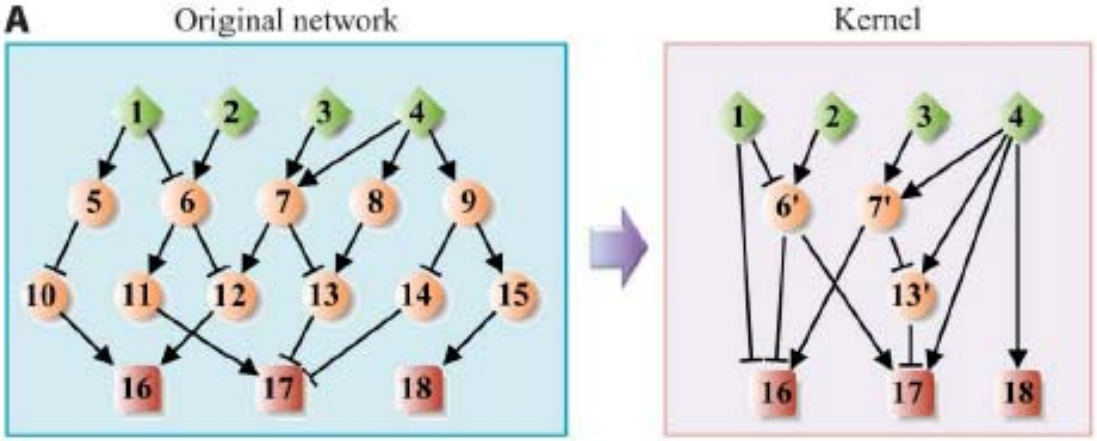
Subnetworks can be reduced or not!

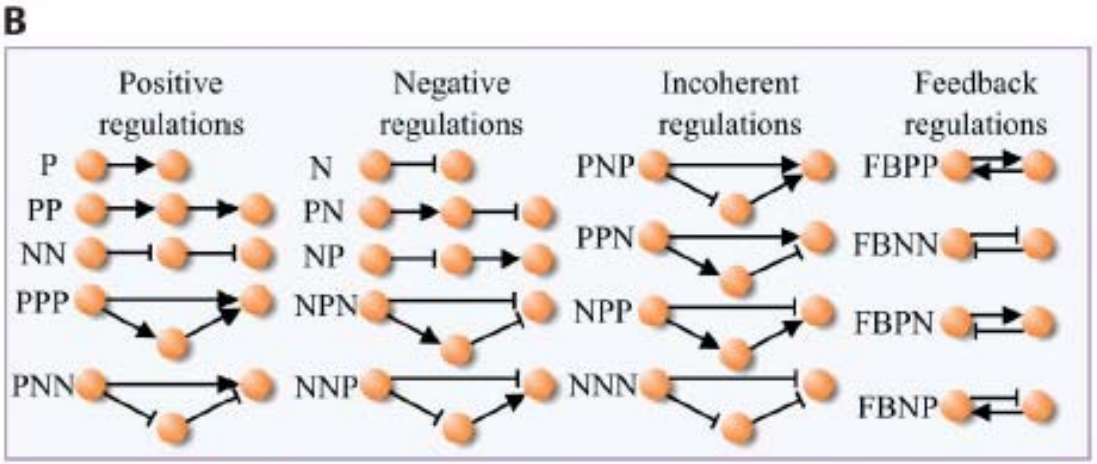
The algorithm cannot replace subnetworks: (i) when one node in a three-node subnetwork is also a component node of a self-feedback loop, a two-node feedback loop, or an intermediate node of an incoherent feedforward loop, or (ii) when both the indegree and the outdegree of the node are >1

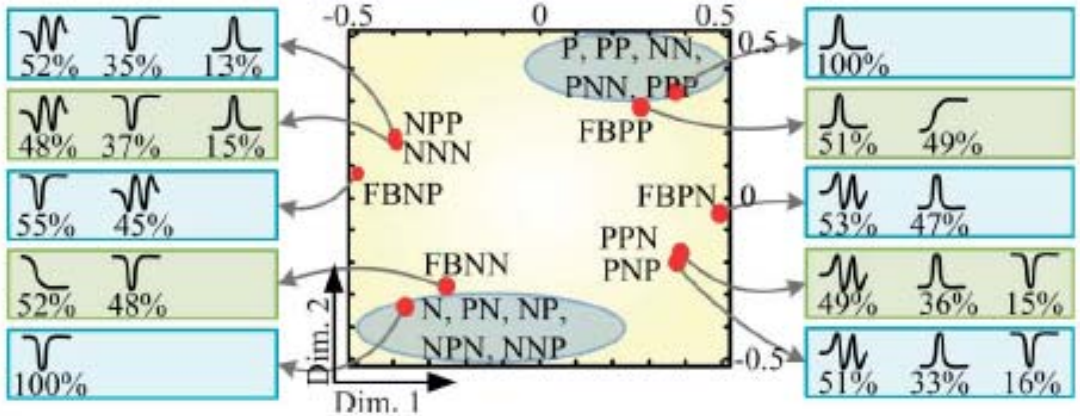


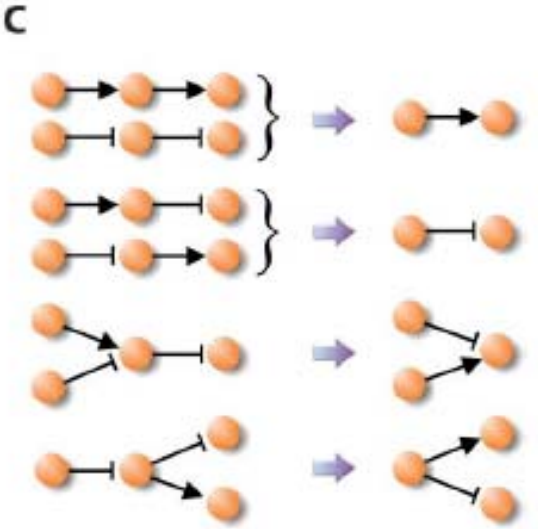
When a network cannot be reduced any further by the above reduction process, the algorithm reduces the network by replacing the neighborhood subnetwork of a set of edges, taking into account consistency of the types of regulation among the neighboring edges

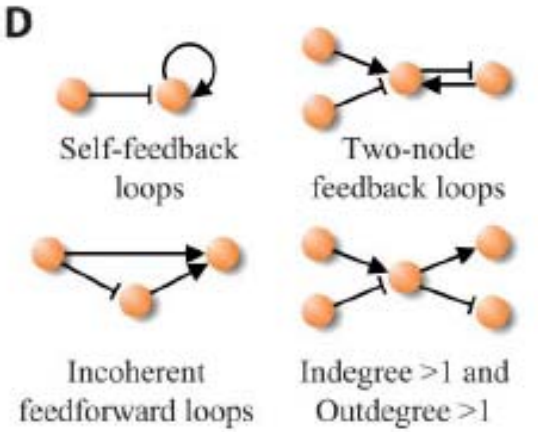


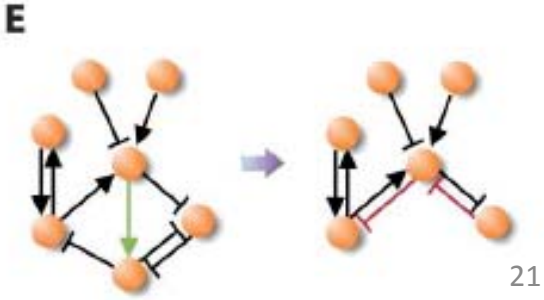

A
Original network
Kernel
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
6'
7'
13'


B
Positive regulations
Negative regulations
Incoherent regulations
Feedback regulations
P
PP
NN
PPP
PNN
N
PN
NP
NPN
NNP
PNP
PPN
NPP
NNN
FBPP
FBNN
FBPN
FBNP

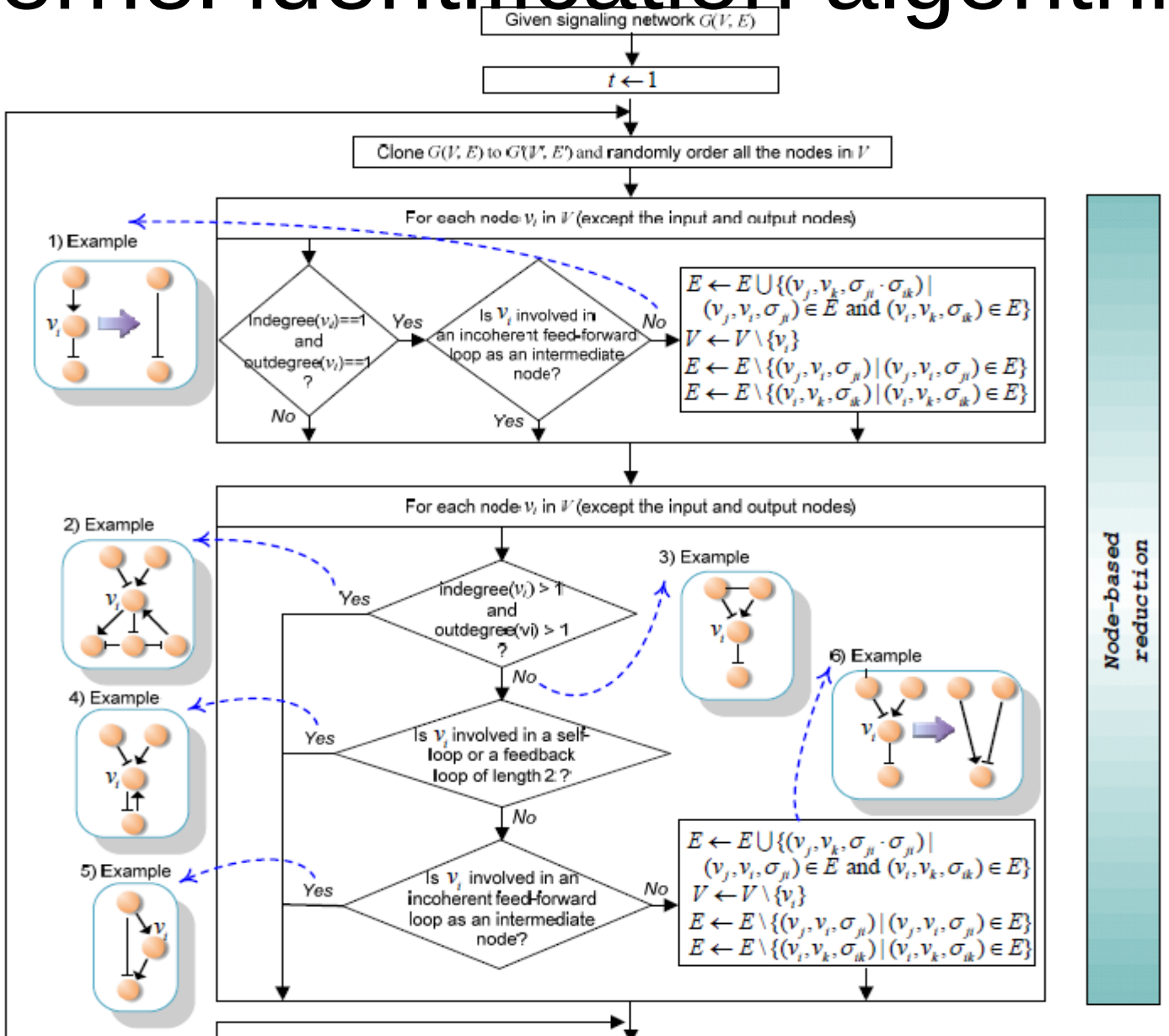

52% 35% 13%
48% 37% 15%
55% 45%
52% 48%
100%
-0.5
0
0.5
P, PP, NN,
PNN, PPP
NPP
NNN
FBPP
FBNP
FBPN
FBNN
PPN
PNP
N, PN, NP,
NPN, NNP
Dim. 2
Dim. 1
100%
51% 49%
53% 47%
49% 36% 15%
51% 33% 16%


C

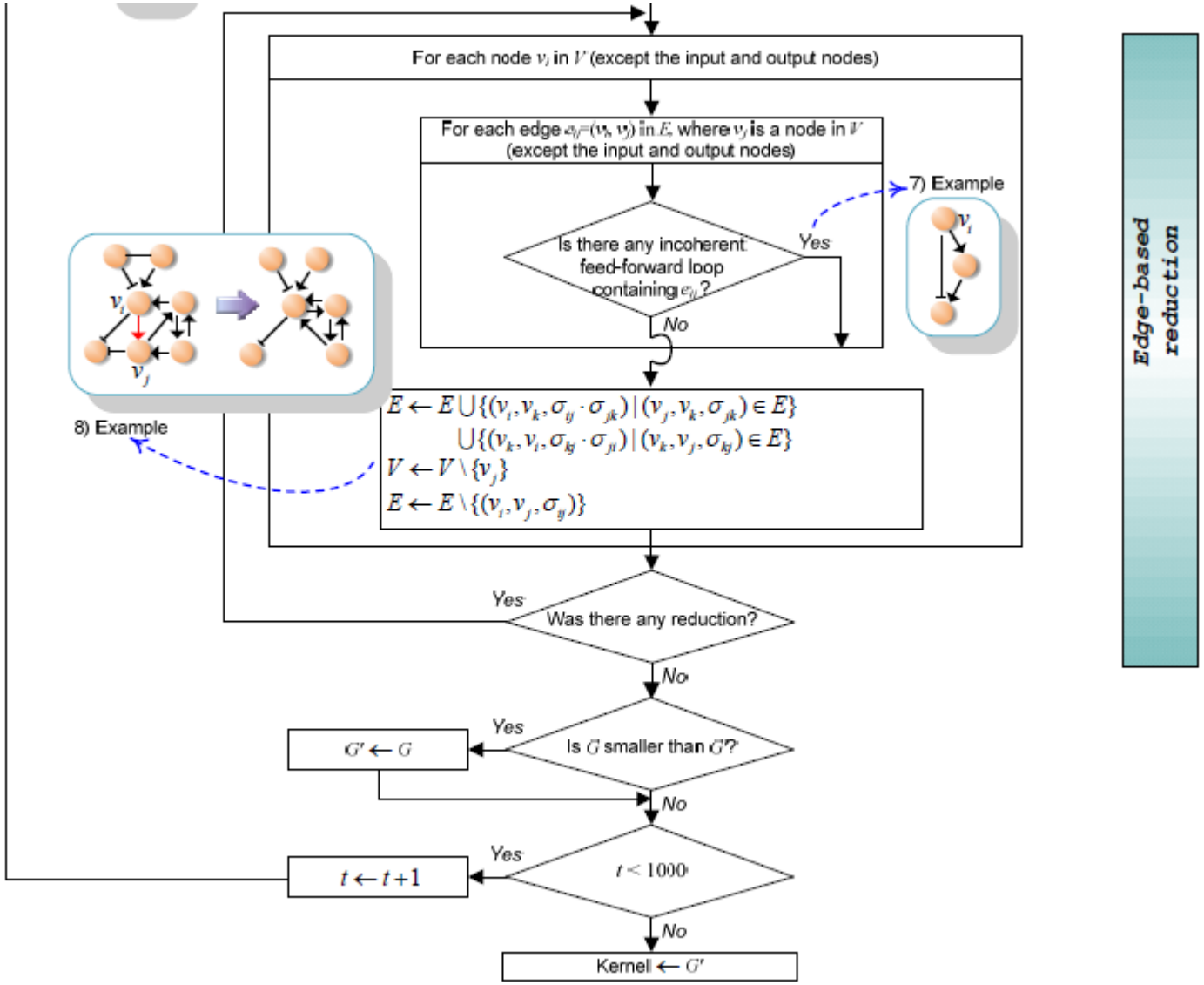

D
Self-feedback loops
Two-node feedback loops
Incoherent feedforward loops
Indegree >1 and Outdegree >1


E

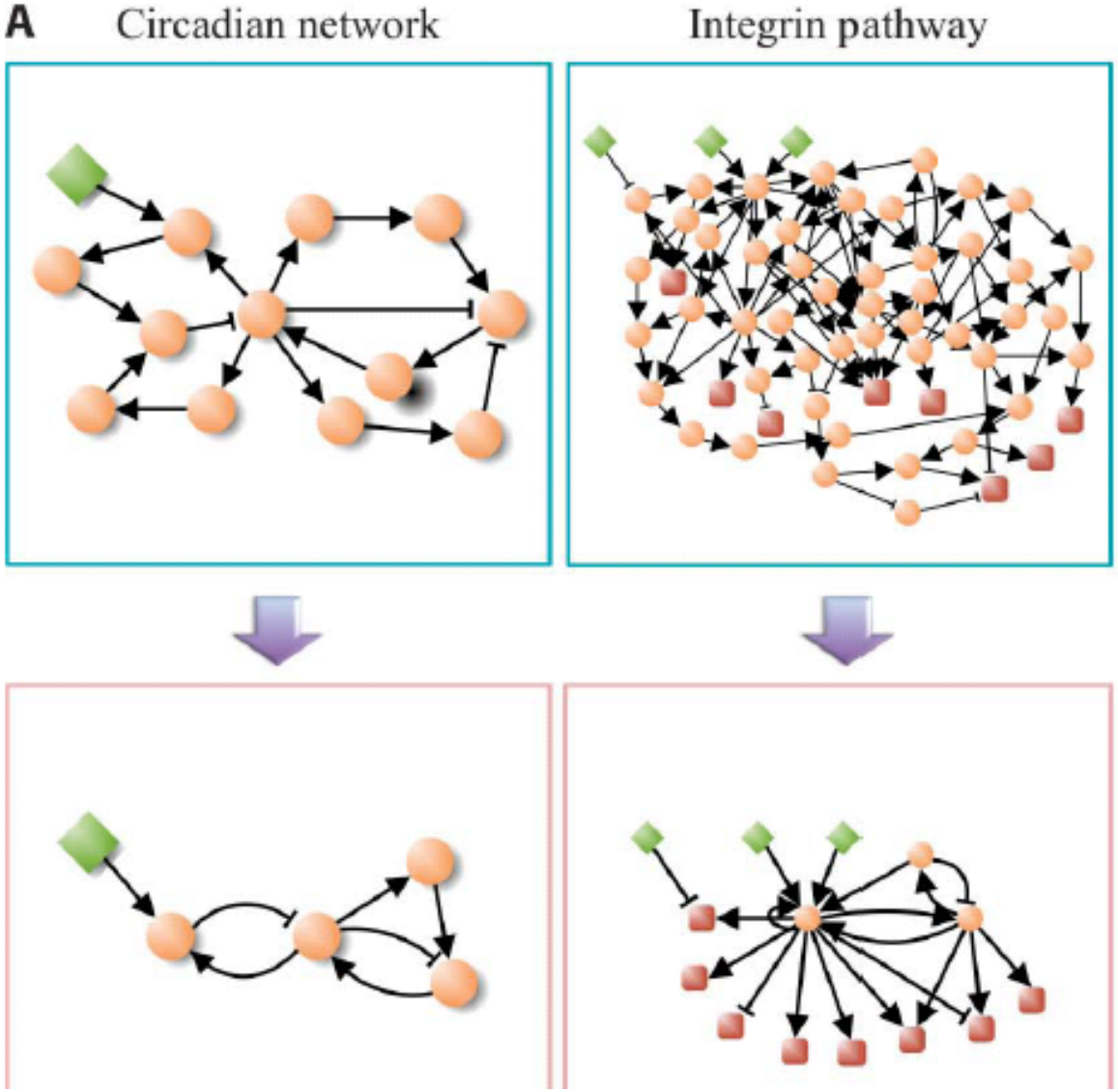
Kernel identification algorithm



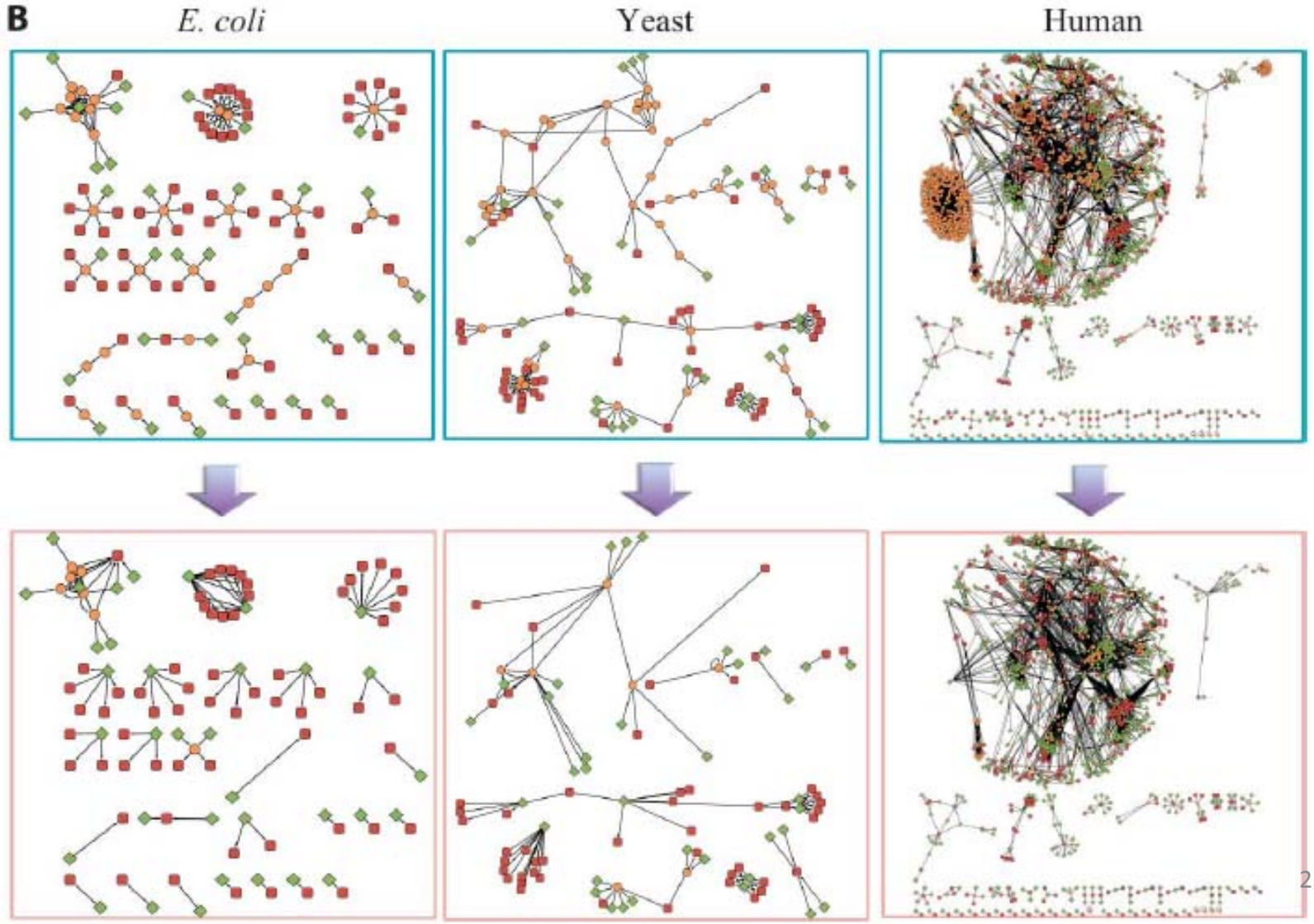
Kernel identification algorithm



Kernels of the circadian network and integrin pathway with node reduction percentages of 67% and 94%, respectively.



Kernels of the networks of E. coli, yeast, and human with node reduction rates of 77%, 81%, and 81%



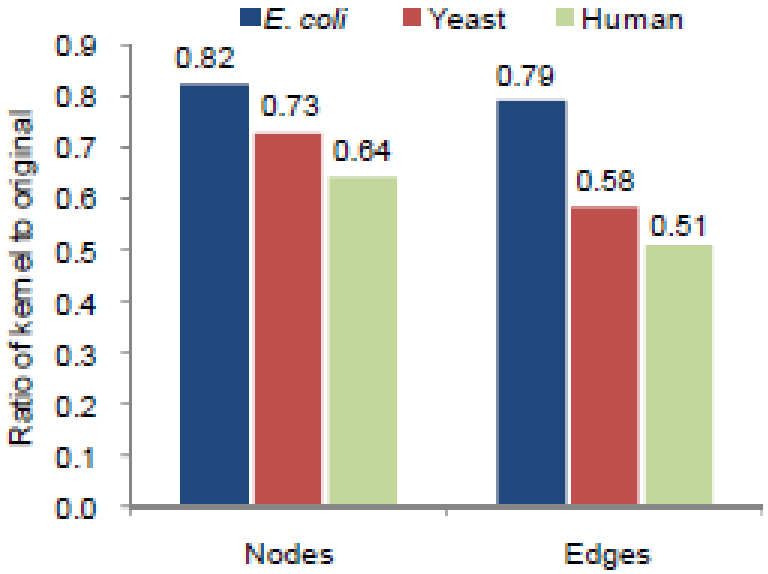
“node reduction percentage”

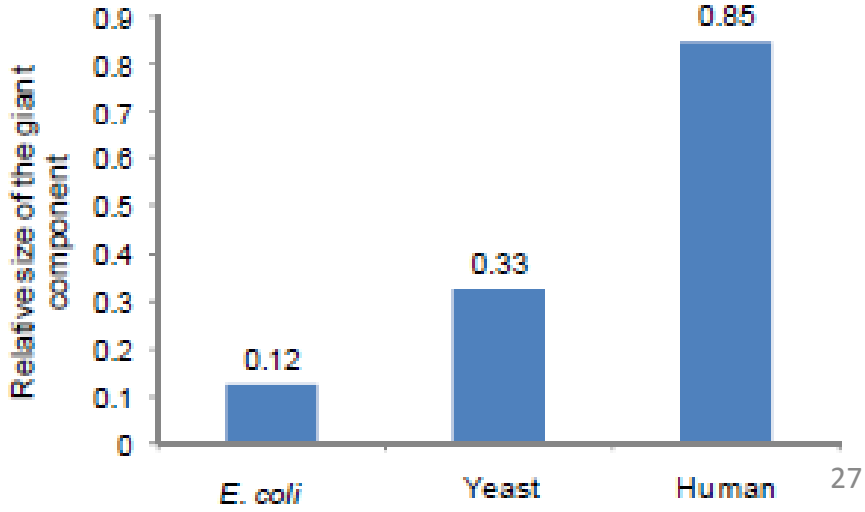
C

	E. coli	Yeast	Human
Number of the total nodes	129	129	1953
Number of the input nodes	35	36	669
Number of the intermediate nodes	30	43	867
Number of the output nodes	64	50	417
Number of the reduced nodes	23	35	699
Node reduction percentage (%)*	77	81	81

*(Number of the reduced nodes/Number of the intermediated nodes) x 100

Topological characteristics of kernels (Ratio!!)





Validation of the story

- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels
- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes encoding drug targets, or
  - genes that are part of synthetic lethal gene pairs.
  - kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
  - encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.

Validation of the story

- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels
- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes encoding drug targets, or
  - genes that are part of synthetic lethal gene pairs.
  - kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
  - encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.

Preserved the input-output dynamics or not

Boolean models

To compare the similarity of dynamics between each signaling network and the corresponding kernel, we constructed Boolean models as follows:

1) Every node has one value: 0 or 1.

2) Because we usually do not know the exact Boolean logic of large-scale networks, we consider only two cases: every Boolean logic is 'AND' or every Boolean logic is 'OR'.

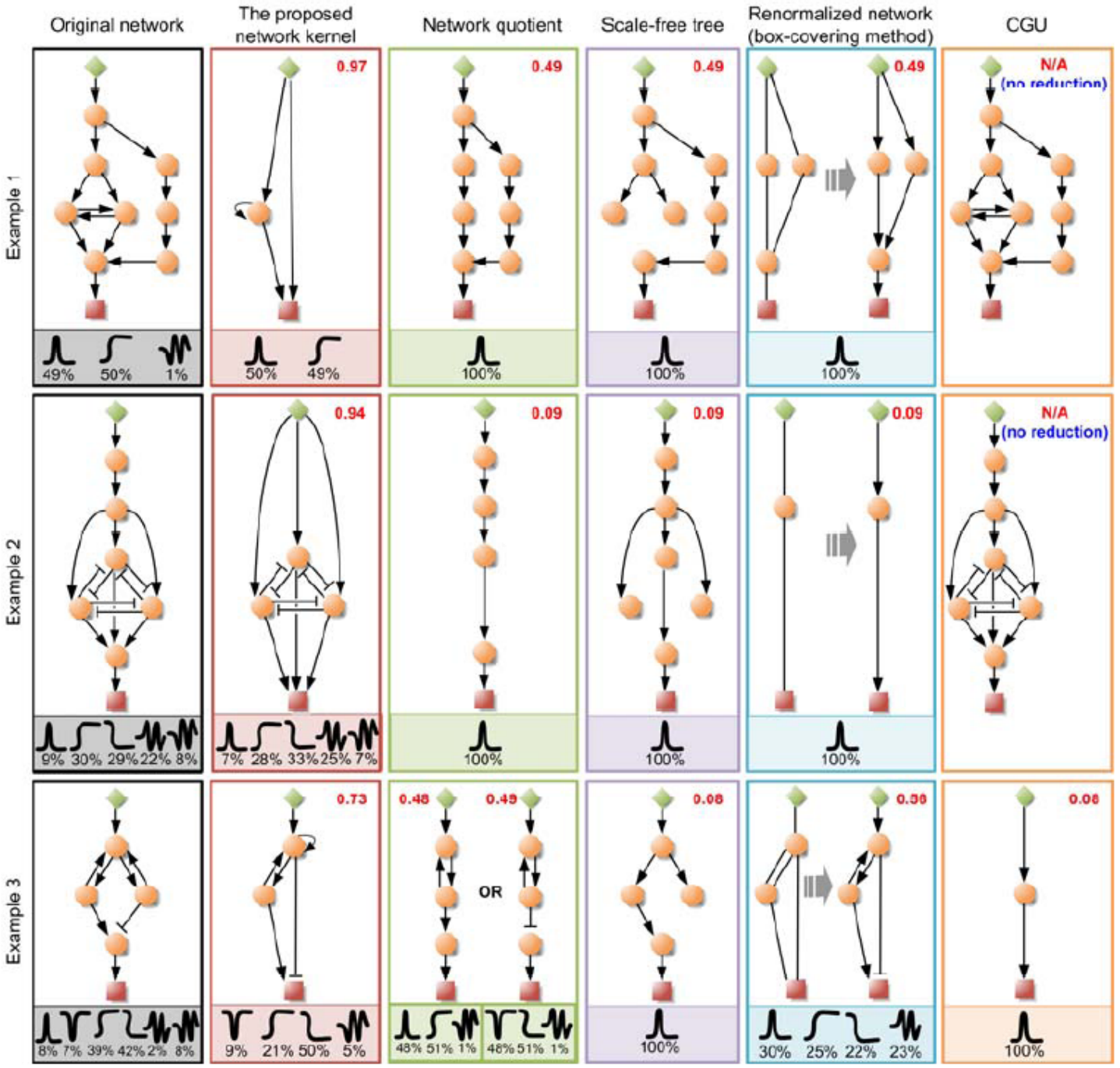
3) We used the standard Boolean model. For example, let X activate Y and Z inhibit Y. If the Boolean logic is 'AND', then the next state of Y is determined by the previous states of X and Z as follows:

X	Z	Y
0	0	0
1	0	1
0	1	0
1	1	0

If the Boolean logic is 'OR', then the next state of Y is determined by the previous states of X and Z as follows:

X	Z	Y
0	0	1
1	0	1
0	1	0
1	1	1

4) For a constant stimulation given to input nodes, we compared the steady state responses of the output nodes (we used the average of the output node states over the last 100 time steps as the steady state value of an output node). As an example, let us consider two steady state response vectors, S1 and S2, composed of the steady state values of N output nodes. Then, the response coherency is defined as the average of 1-d(S1,S2)/N for all possible input node perturbations where d(A,B) denotes the absolute distance measure between two vectors A and B. The response coherency between the original network and its kernel is a measure of how similar the response profiles of the kernel are to those of the original network.


Original network
The proposed network kernel
Network quotient
Scale-free tree
Renormalized network (box-covering method)
CGU
Example 1
Example 2
Example 3
0.97
0.49
0.49
0.49
N/A
(no reduction)
49% 50% 1%
50% 49%
100%
100%
100%
0.94
0.09
0.09
0.09
N/A
(no reduction)
9% 30% 29% 22% 8%
7% 28% 33% 25% 7%
100%
100%
100%
0.73
0.48
0.49
OR
0.08
0.36
0.08
8% 7% 39% 42% 2% 8%
9% 21% 50% 5%
48% 51% 1%
48% 51% 1%
100%
30% 25% 22% 23%
100%

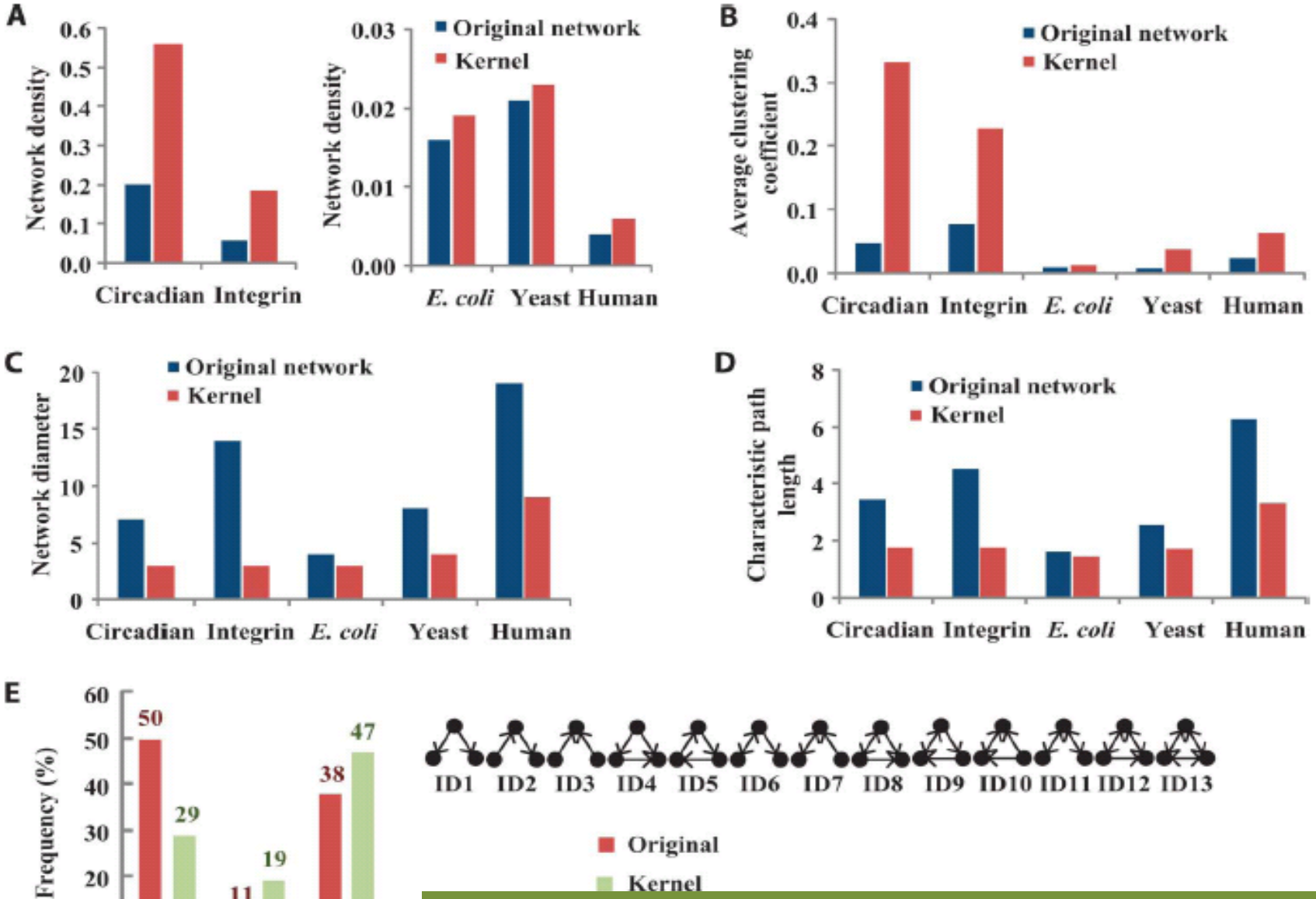
Preserved the input-output dynamics or not

Table S1. Response coherency between the original signaling network and the corresponding kernel. See Supplementary Model Descriptions for details.

	AND logic	OR logic
E. coli	1	0.998
Yeast	0.981	0.981
Human	0.938	0.966

Validation of the story

- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels
- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes encoding drug targets, or
  - genes that are part of synthetic lethal gene pairs.
  - kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
  - encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.



This implies that the human network includes a large number of signal splitters and many signaling pathways are connected by signal integrators.

Motif comparison in …

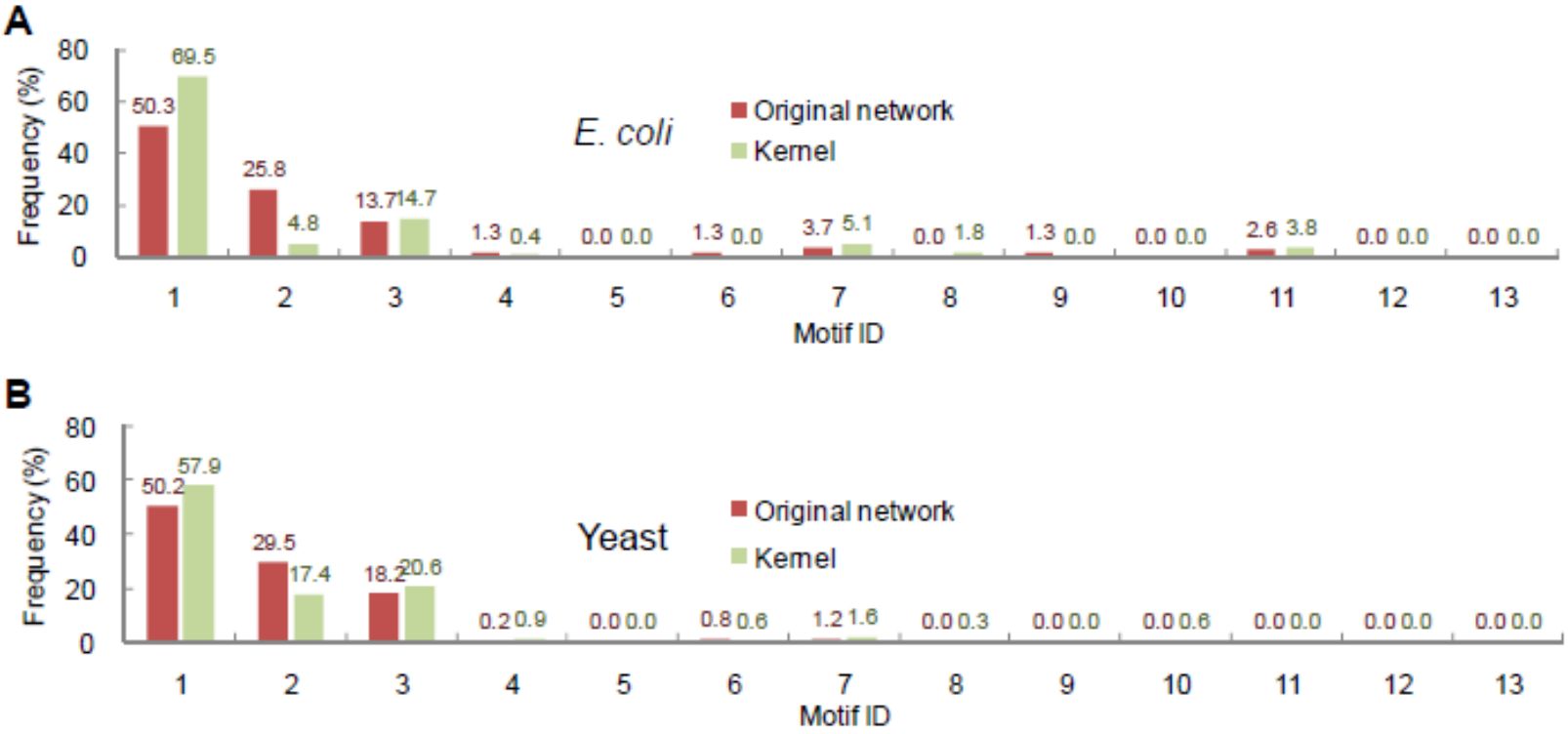
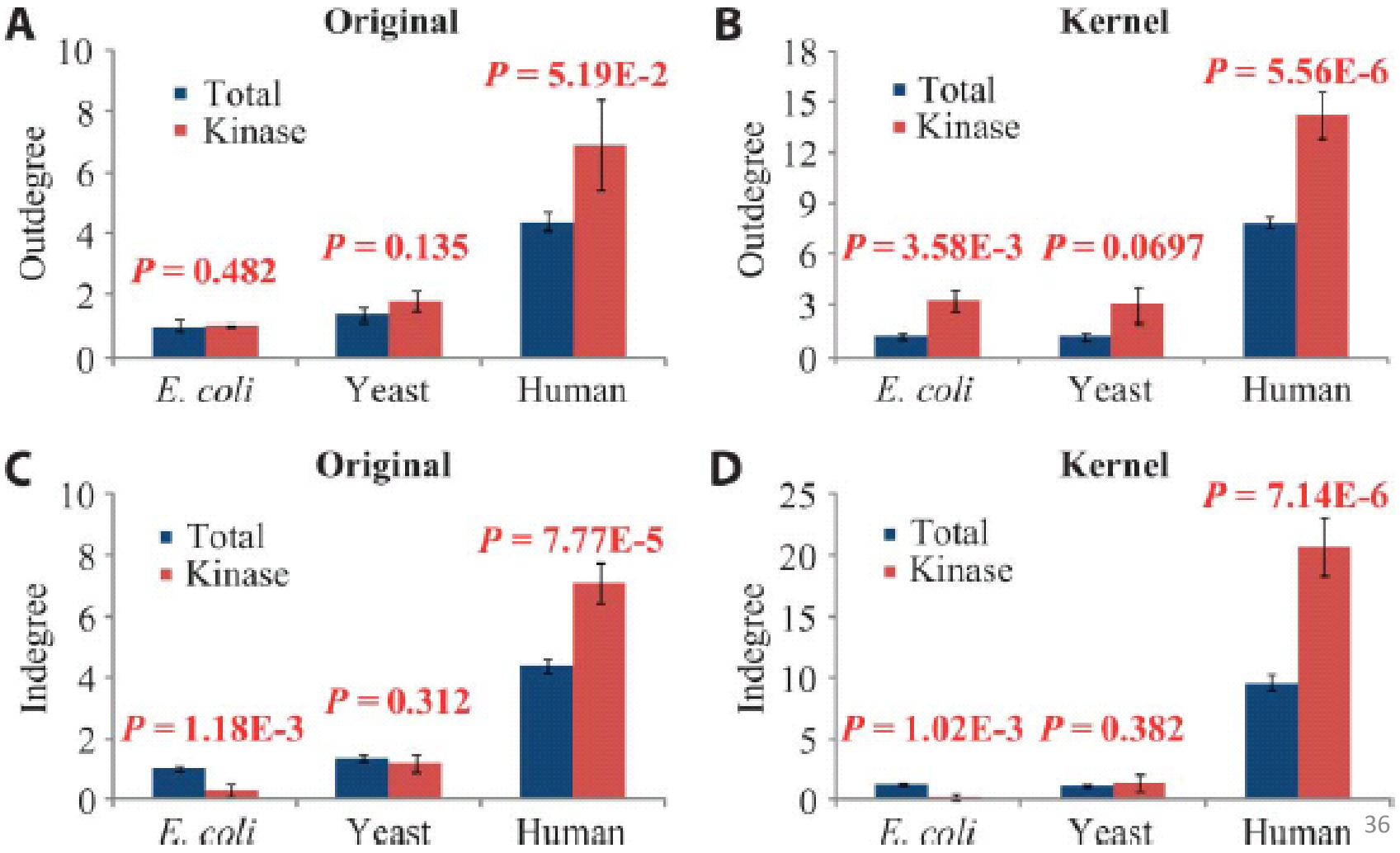


Fig. S5. The frequency distributions of three-node subnetworks in the signaling networks of *E. coli* and yeast compared with the distributions of these subnetworks in their kernels.

Human kinases are more connected than those of E. coli or yeast.

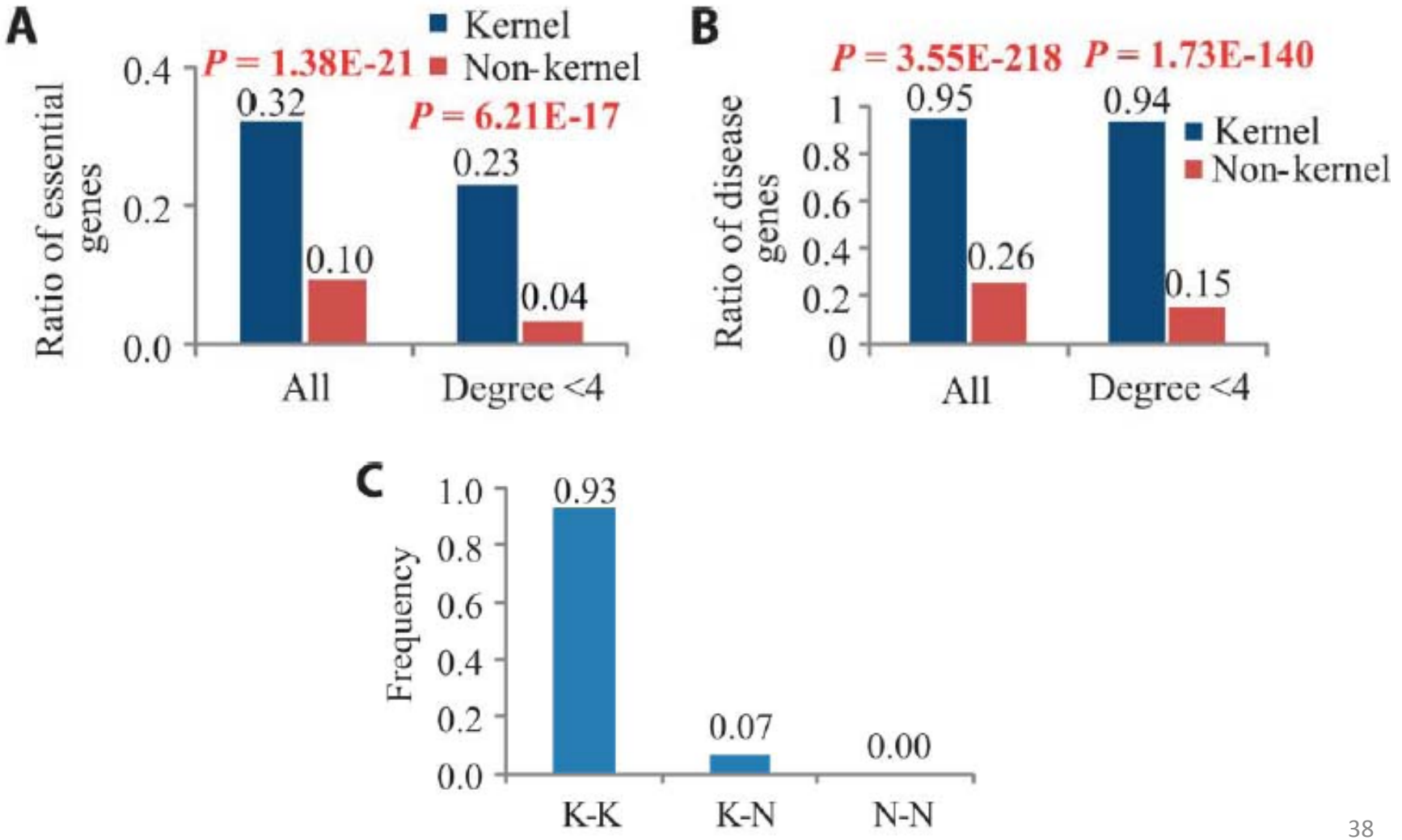


Validation of the story

- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels

- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes that are part of synthetic lethal gene pairs.
  - kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
  - encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
  - genes encoding drug targets
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.

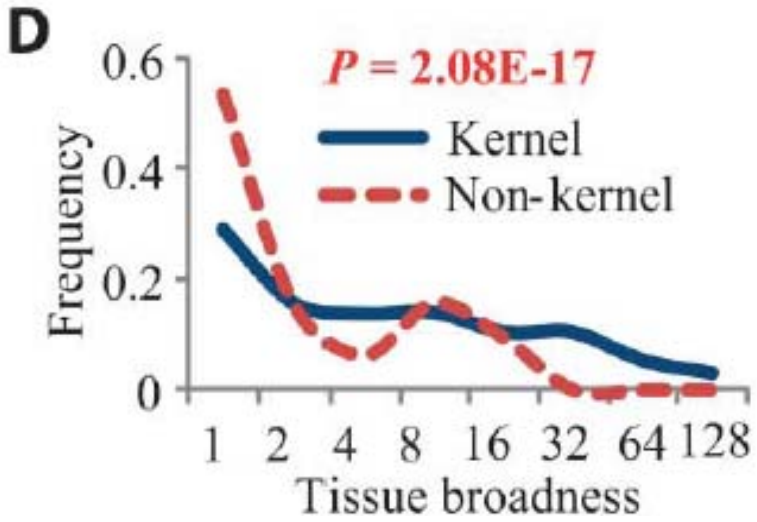
Enrichment of essential genes, disease genes, and synthetic lethal genes

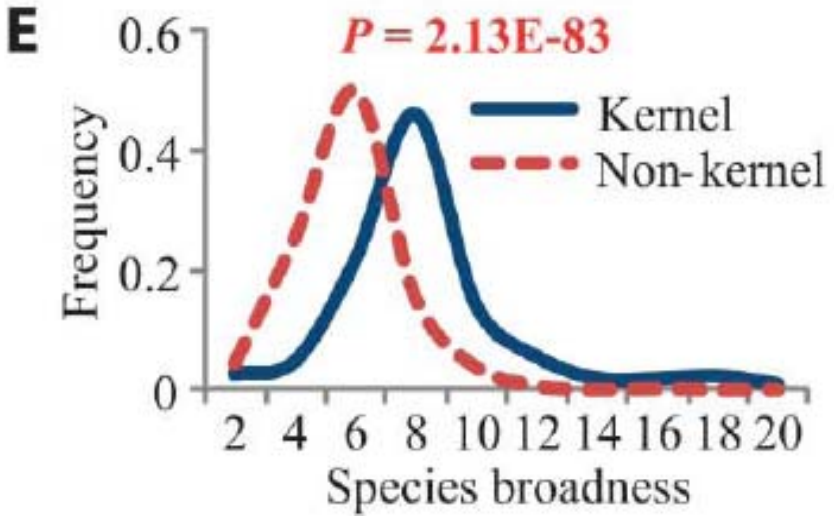


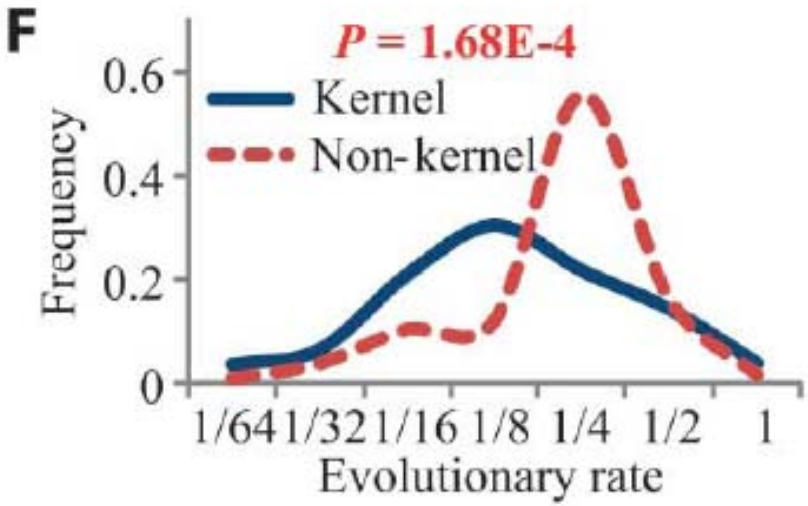
Validation of the story

- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels
- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes that are part of synthetic lethal gene pairs.
- kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
- encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
  - genes encoding drug targets
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.

Tissue broadness and species broadness of kernel nodes





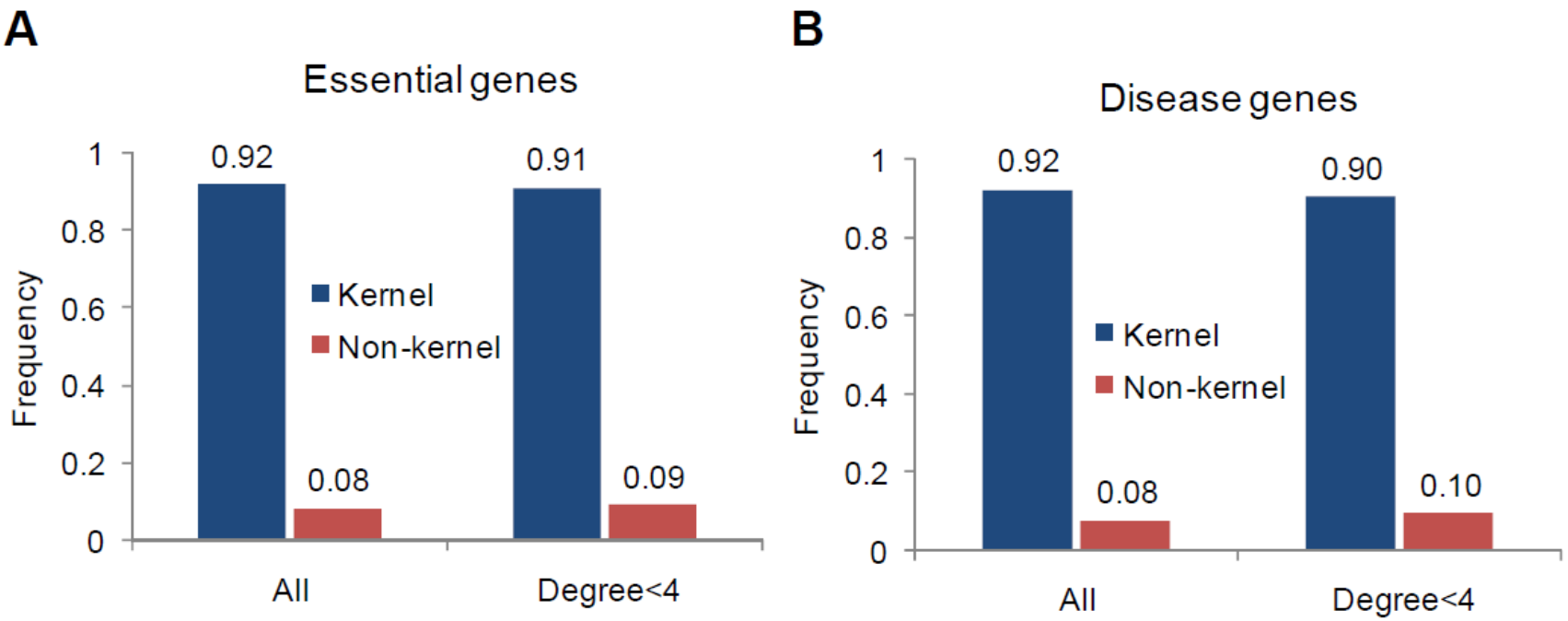


Validation of the story

- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels
- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes that are part of synthetic lethal gene pairs.
  - kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
  - encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
  - genes encoding drug targets
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.

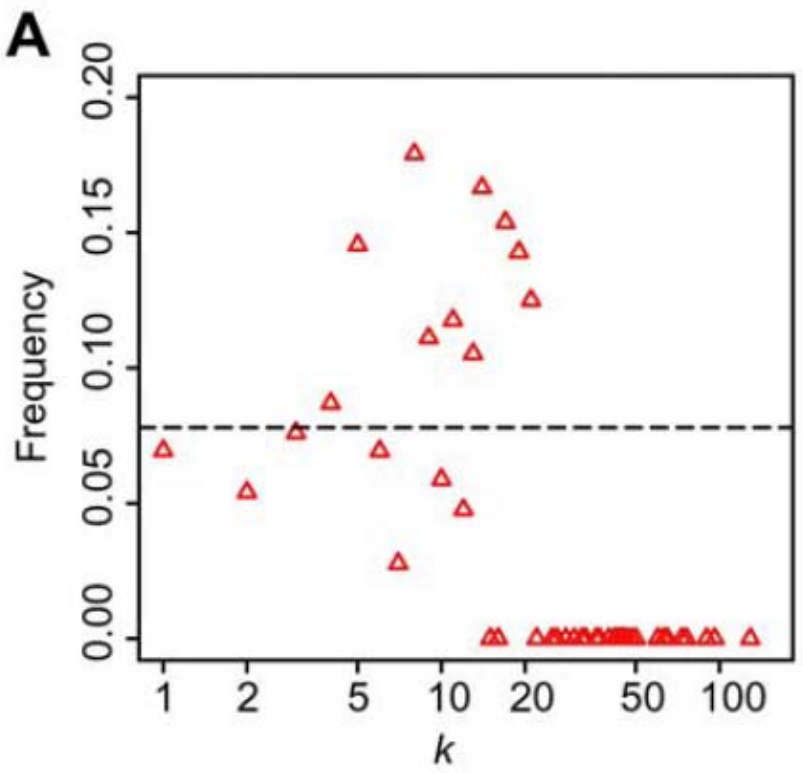
Network kernel and drug targets

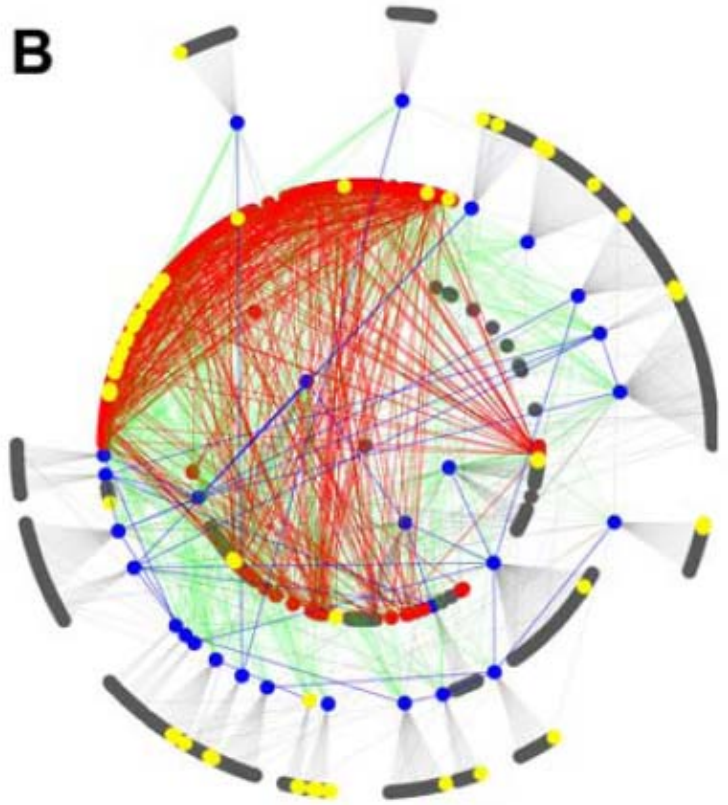
Because most kernel nodes in the human signaling network can be mapped to diseases, so how about **network kernel vs. drug targets?**



Network kernel and drug targets

Drug targets are enriched in the backbone network composed of middle-degree nodes (6 to 38 connections)

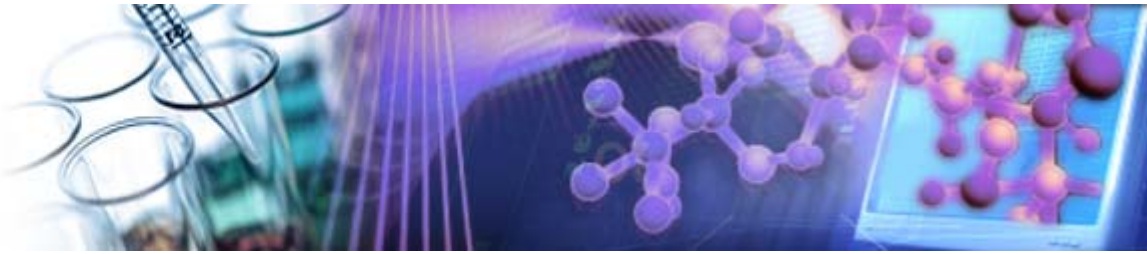




PLoS Comput. Biol. 5, e1000550 (2009)

Network kernel and drug targets

DRUGBANK
Open Data Drug & Drug Target Database



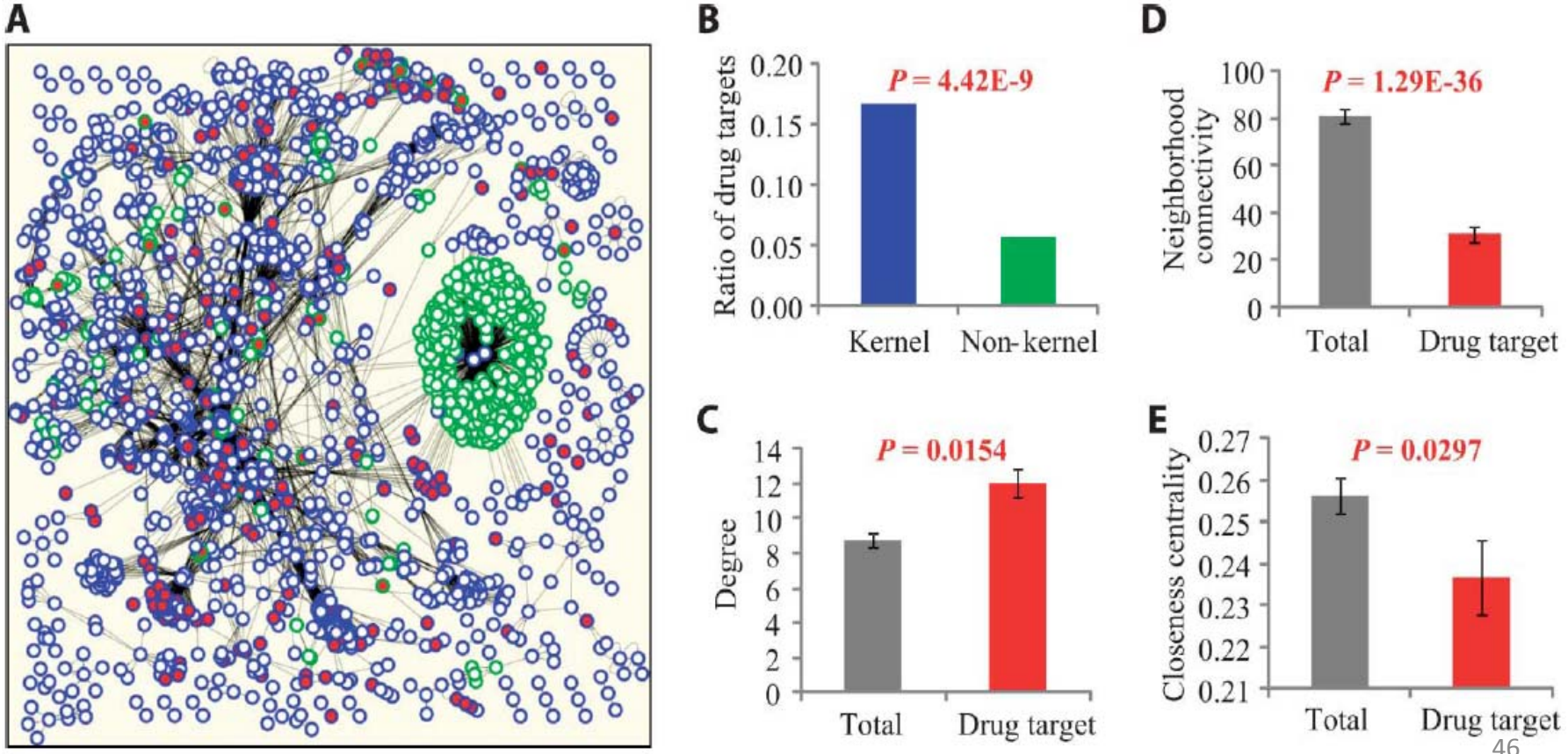
What is DrugBank?

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains **6829 drug entries** including **1435 FDA-approved small molecule drugs**, **134 FDA-approved biotech (protein/peptide) drugs**, 83 nutraceuticals and **5210 experimental drugs**. Additionally, 4438 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 150 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

Nucleic Acids Res. 2011 Jan;39:D1035-41.

Network kernel and drug targets

Drug targets were enriched in the kernel (Fig. 6, A and B), which is consistent with the previous work



drug targets and node degree

Summary

- Kernel identification algorithm
- The identified kernels preserved the input-output dynamics of the original networks.
- Structural characteristics of networks and kernels
- a large proportion of the nodes within the kernels corresponded to
  - essential genes,
  - disease-associated genes,
  - genes that are part of synthetic lethal gene pairs.
  - kernel nodes were encoded by genes conserved in multiple species, suggesting low evolutionary rates,
  - encoded proteins present in various tissues, suggesting that these kernel-associated genes may serve core cellular functions.
  - genes encoding drug targets
- provide a reduced form of a given network, and this smaller network may ***provide insight into the design principles of complex*** biomolecular interaction networks, as well as suggest effective ***ways to perturb or manipulate the network***.

谢谢！

问题？

讨论！