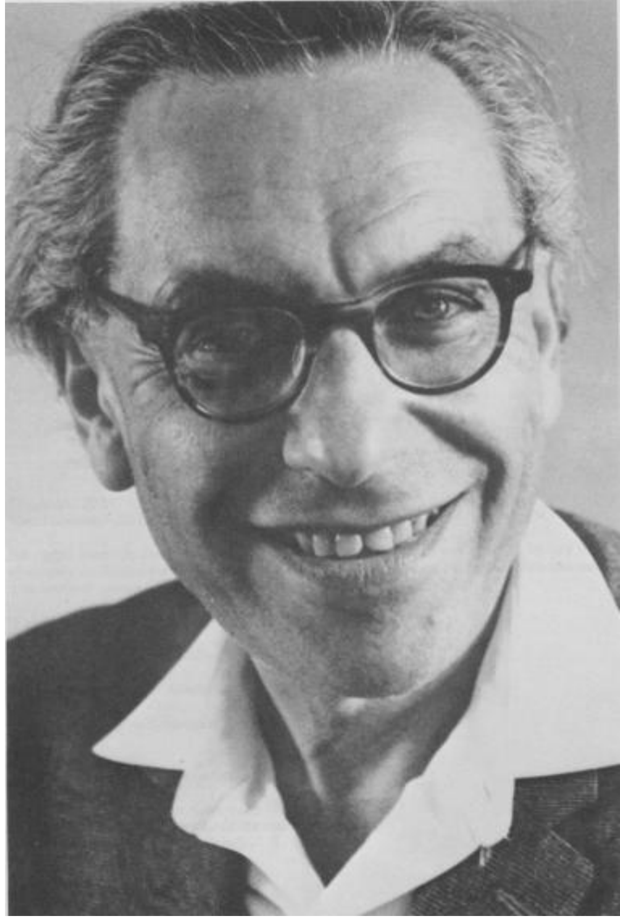
Network Science

Lecture 13

Random Graphs

"A mathematician is a device for turning coffee into theorems"

~1,500 publications (1 every 2 weeks, 35 in 1985)

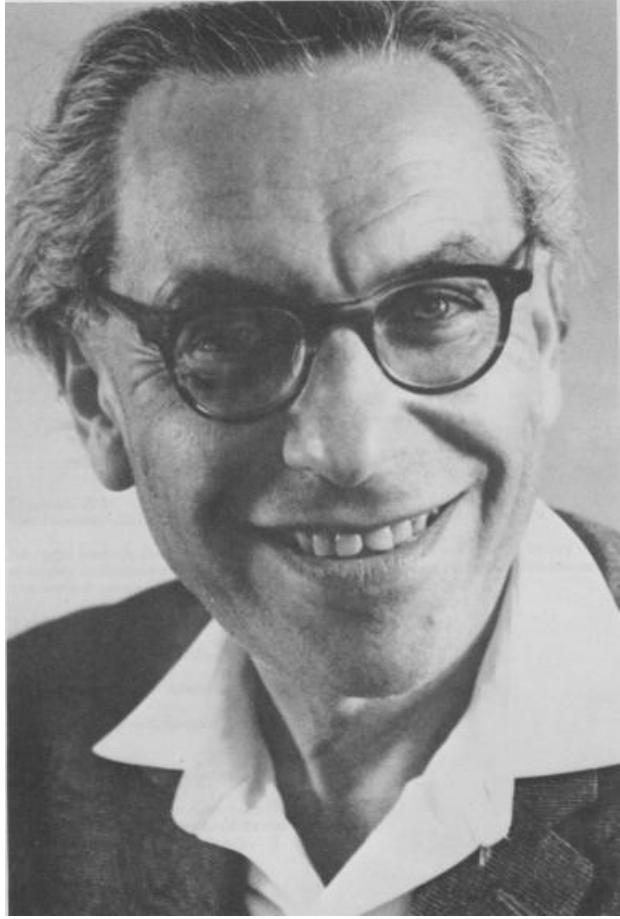


Pál (Paul) Erdõs (1931-1996)
Hungarian Mathematician

Random Graphs

"A mathematician is a device for turning coffee into theorems"

~1,500 publications (1 every 2 weeks, 35 in 1985)



Pál (Paul) Erdõs (1931-1996)
Hungarian Mathematician

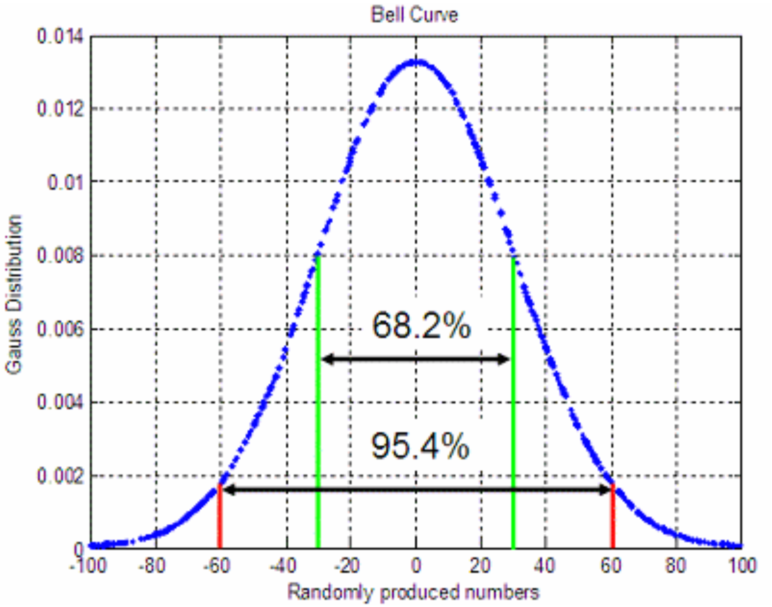
World Wide Web

- The web is a graph with pages as nodes and links to pages as edges
- In 1999 the expectation was that the distribution of links to pages would be normally distributed

Interstate Road Map

Each city is a node, each interstate is
An edge.

The degree distribution is Normal

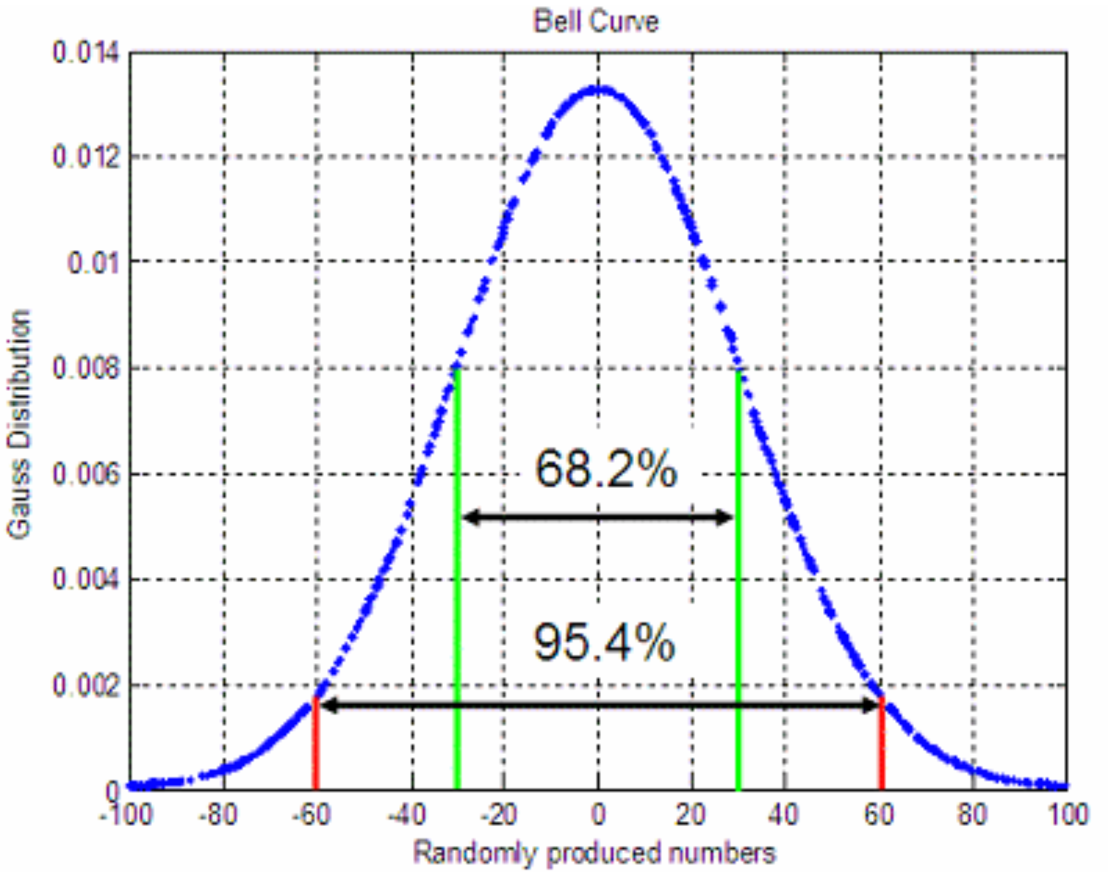


$$P(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(k-\mu)^2/2\sigma^2}$$

$$P(k) \approx e^{-k^2}$$

World Wide Web

- The web is a graph with pages as nodes and links to pages as edges
- In 1999 the expectation was that the distribution of links to pages would be normally distributed

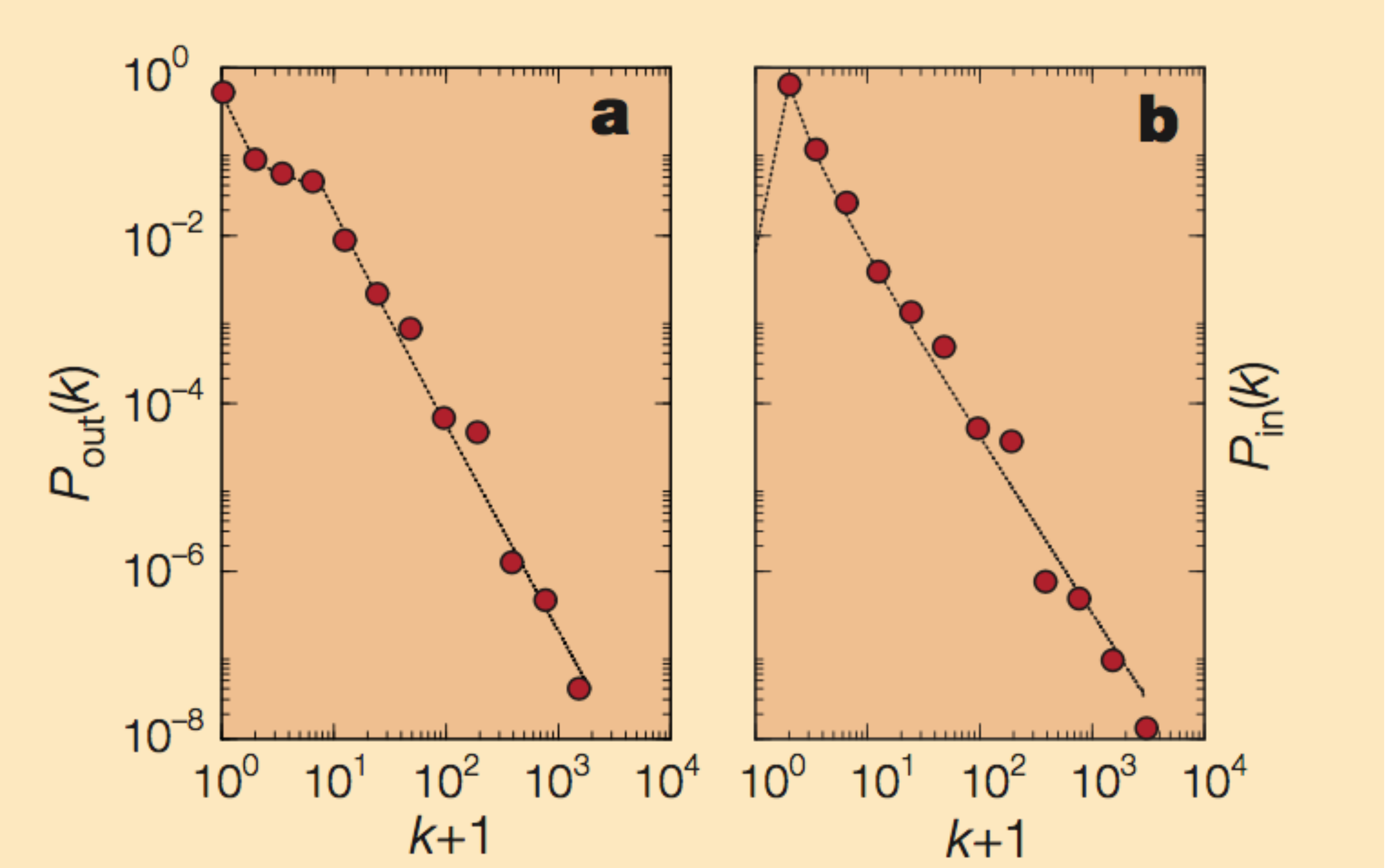


$$P(k) \approx e^{-k^2}$$

The World Wide Web

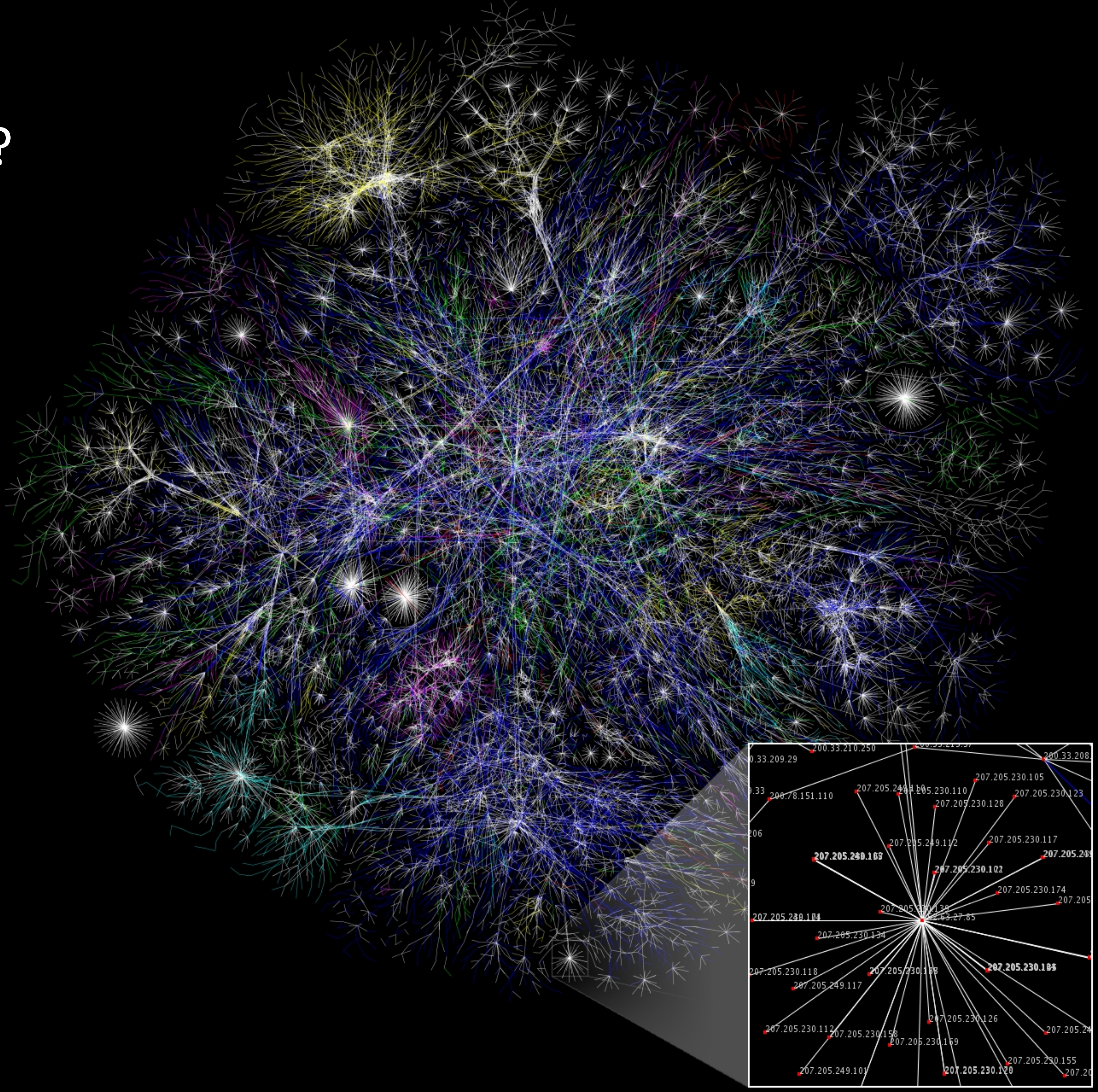
- Actually found

$$P(k) \approx k^{-\gamma}$$



Albert, Réka, Hawoong Jeong, and Albert-László Barabási. "Internet: Diameter of the world-wide web." *nature* 401.6749 (1999): 130-131.

The function is different. So what?



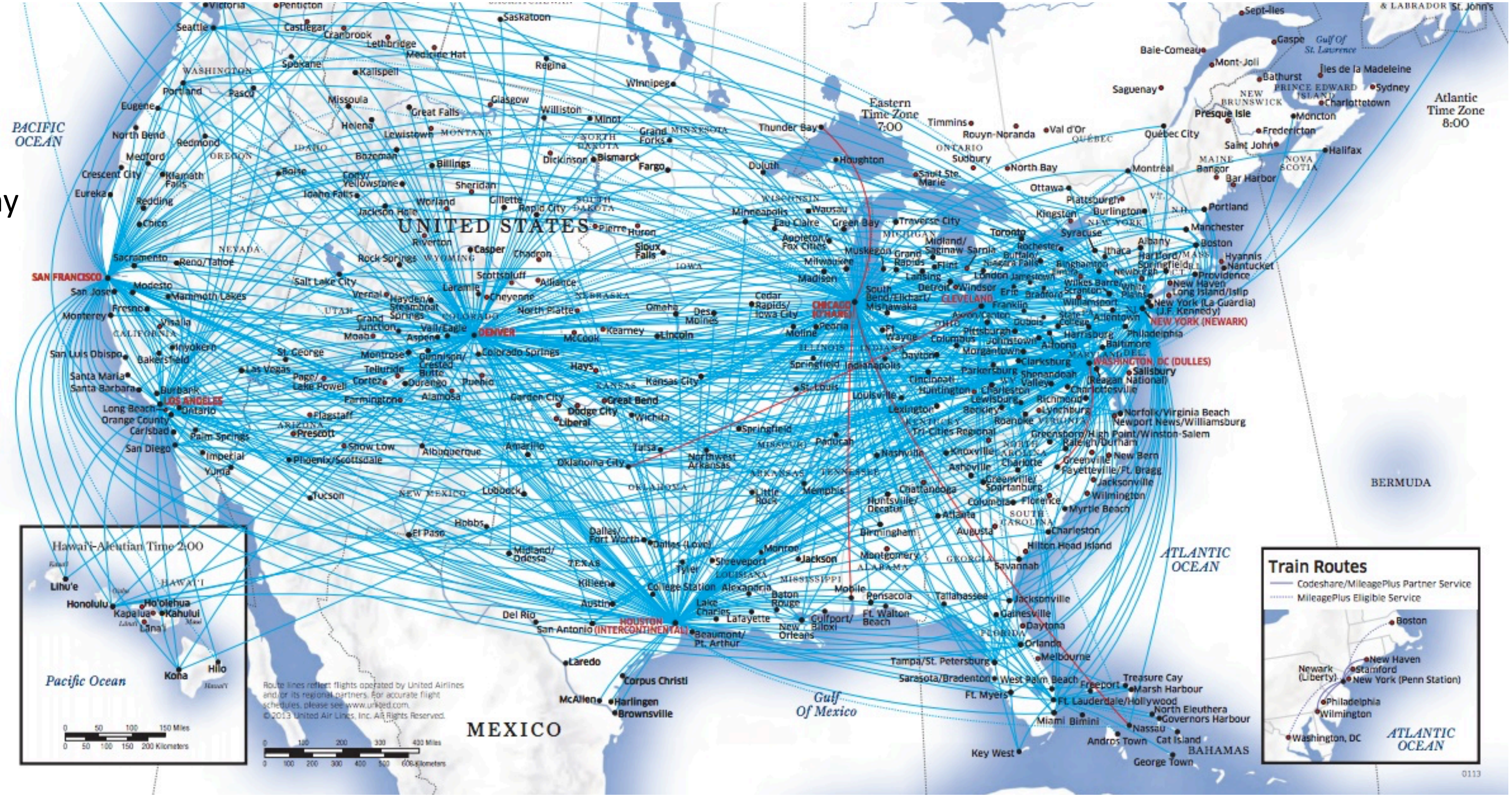
Looks much more like the structure of airline traffic

Nodes are airports

Edges are routes

Most airports are tiny

Some are huge



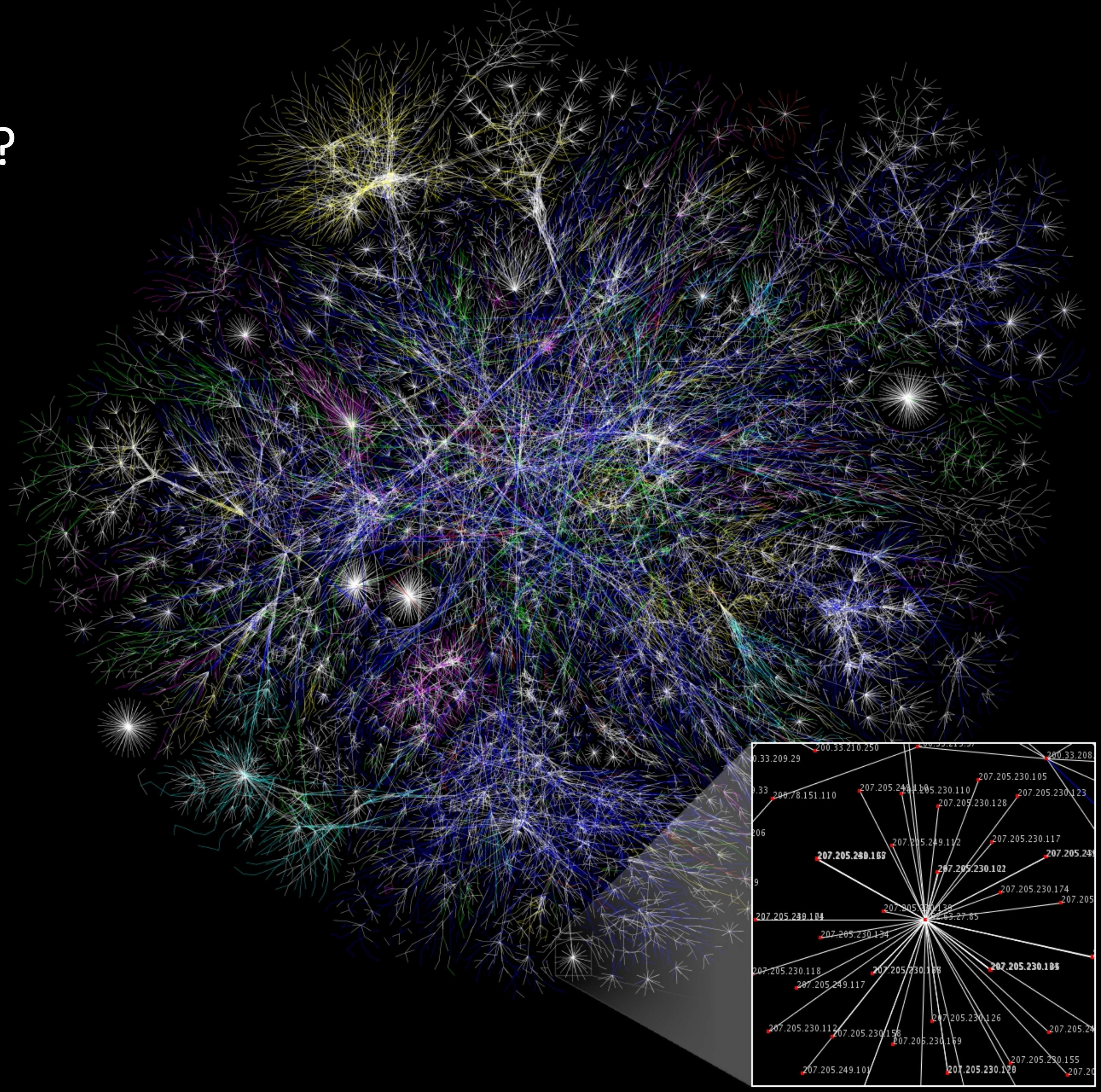
The function is different. So what?

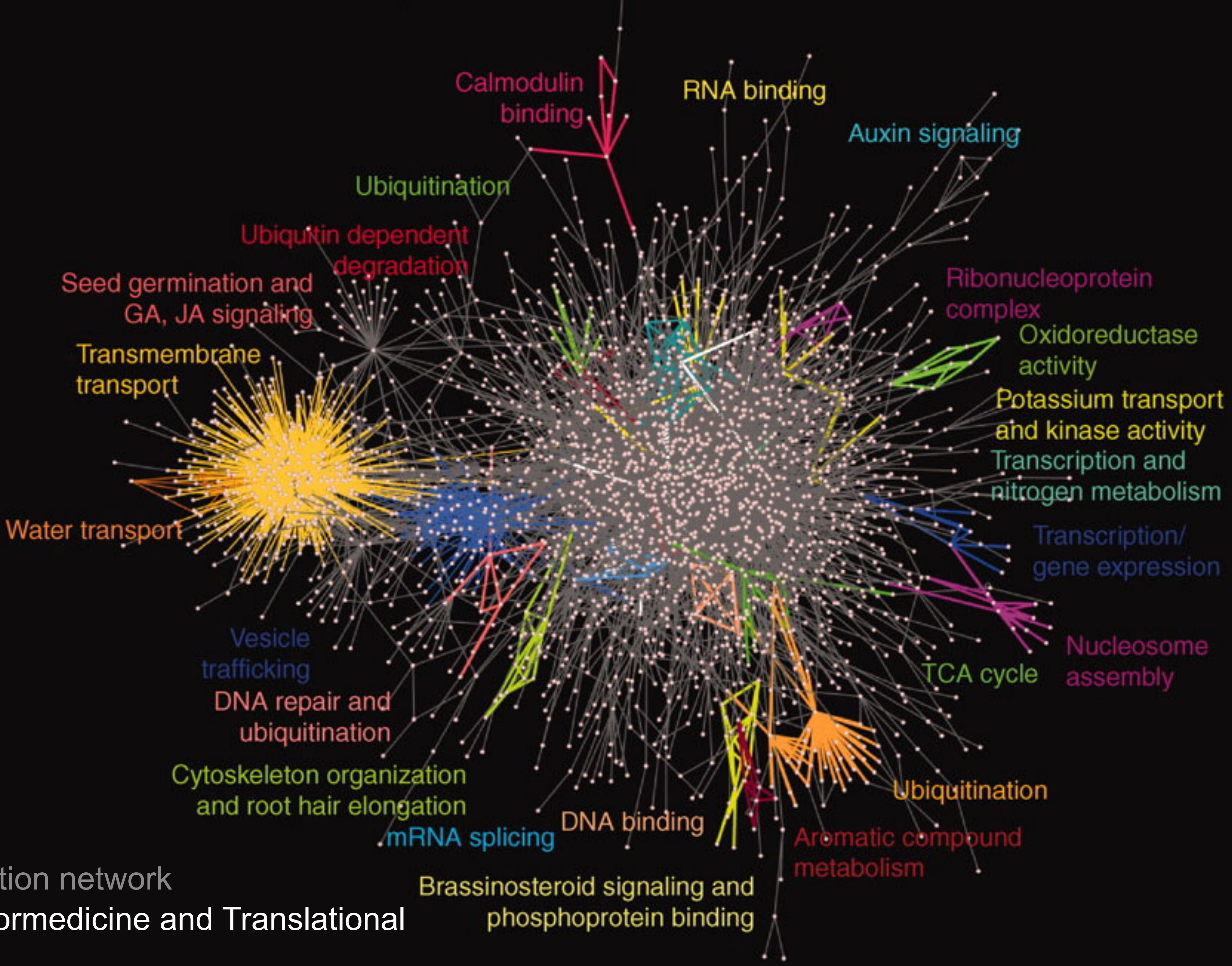
The vast majority of pages
Few people care about

A few pages are very popular

Scale free network

Why it matters…



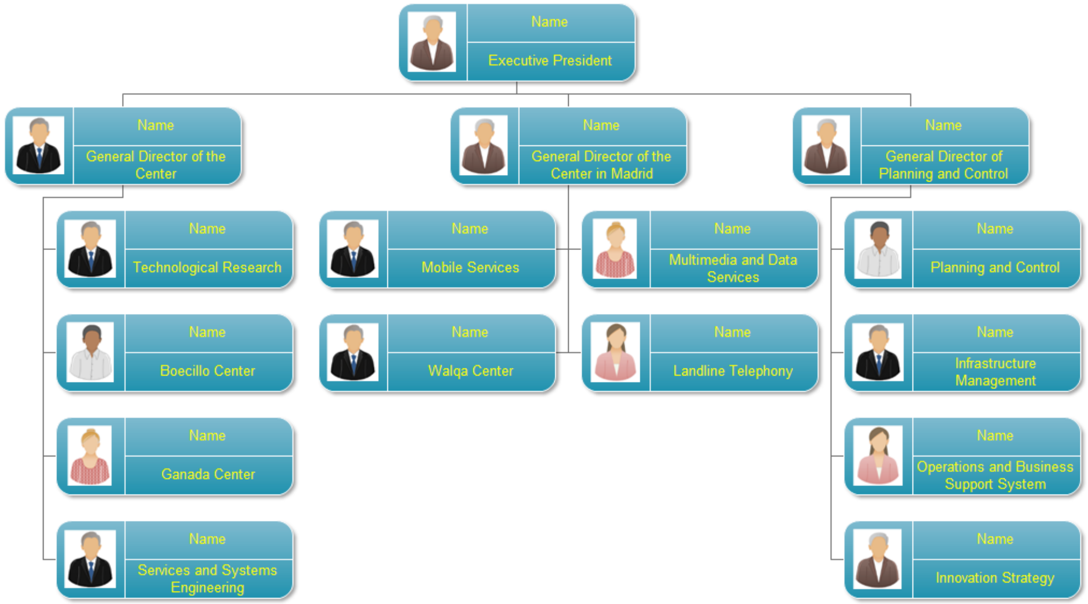


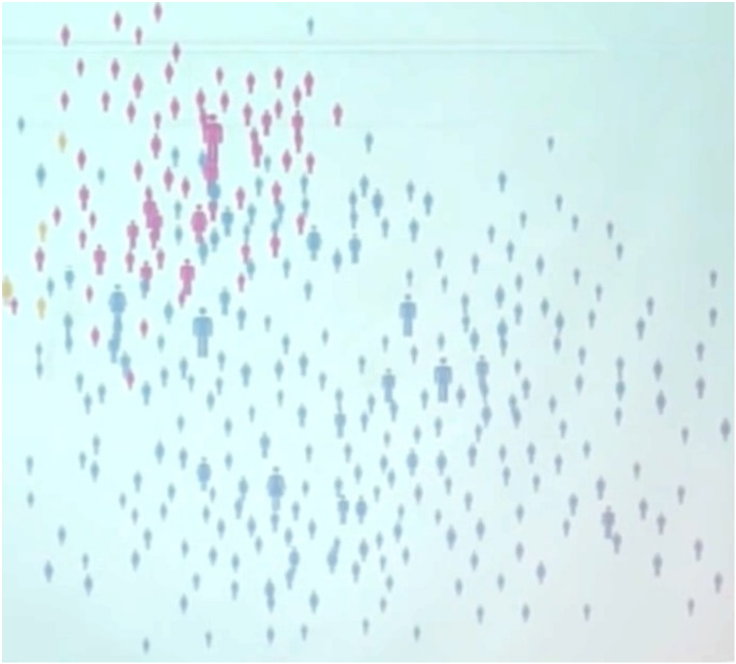
yeast protein protein interaction network

Institute of Integrated Bioinformedicine and Translational Science

A management case study (true story)

- A company is having a problem
- The management discovers that the things the employees are doing are unrelated to the directives from management.





Taken from "Networks are Everywhere," Albert Barabasi

Management Case Study

- The company hires Barabasi to help (an author of the Nature WWW scaling paper)
- All the employees list the people they get assignments from, who they delegate assignments to, and who they get information from
- This reveals the network structure of the company

Taken from “Networks are Everywhere,” Albert Barabasi

A management case study

The network is scale free.

A few employees are information hubs

Most employees have low degree

Taken from “Networks are Everywhere,” Albert Barabasi

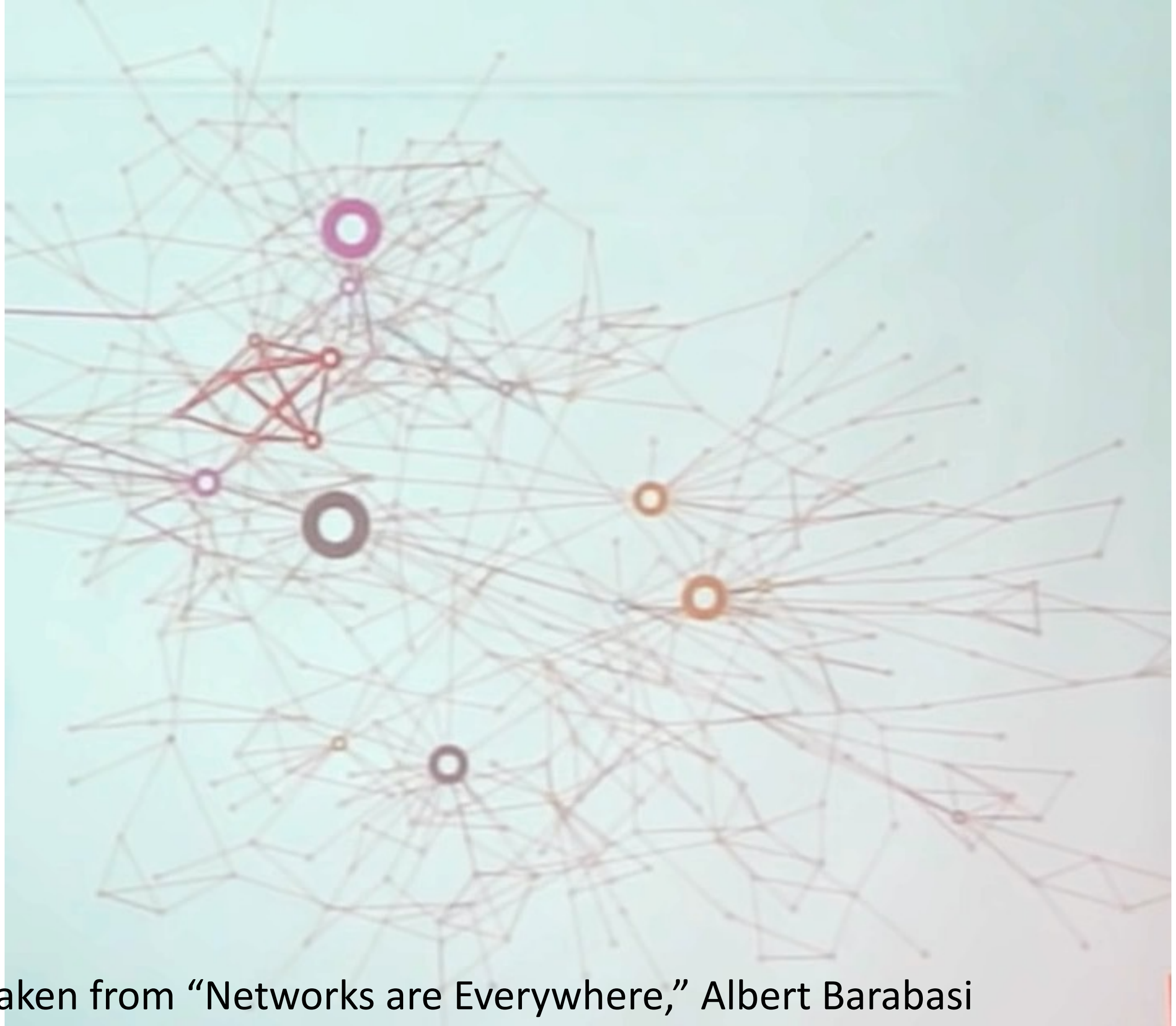
A management case study

The network is scale free.

A few employees are information hubs

Most employees have low degree

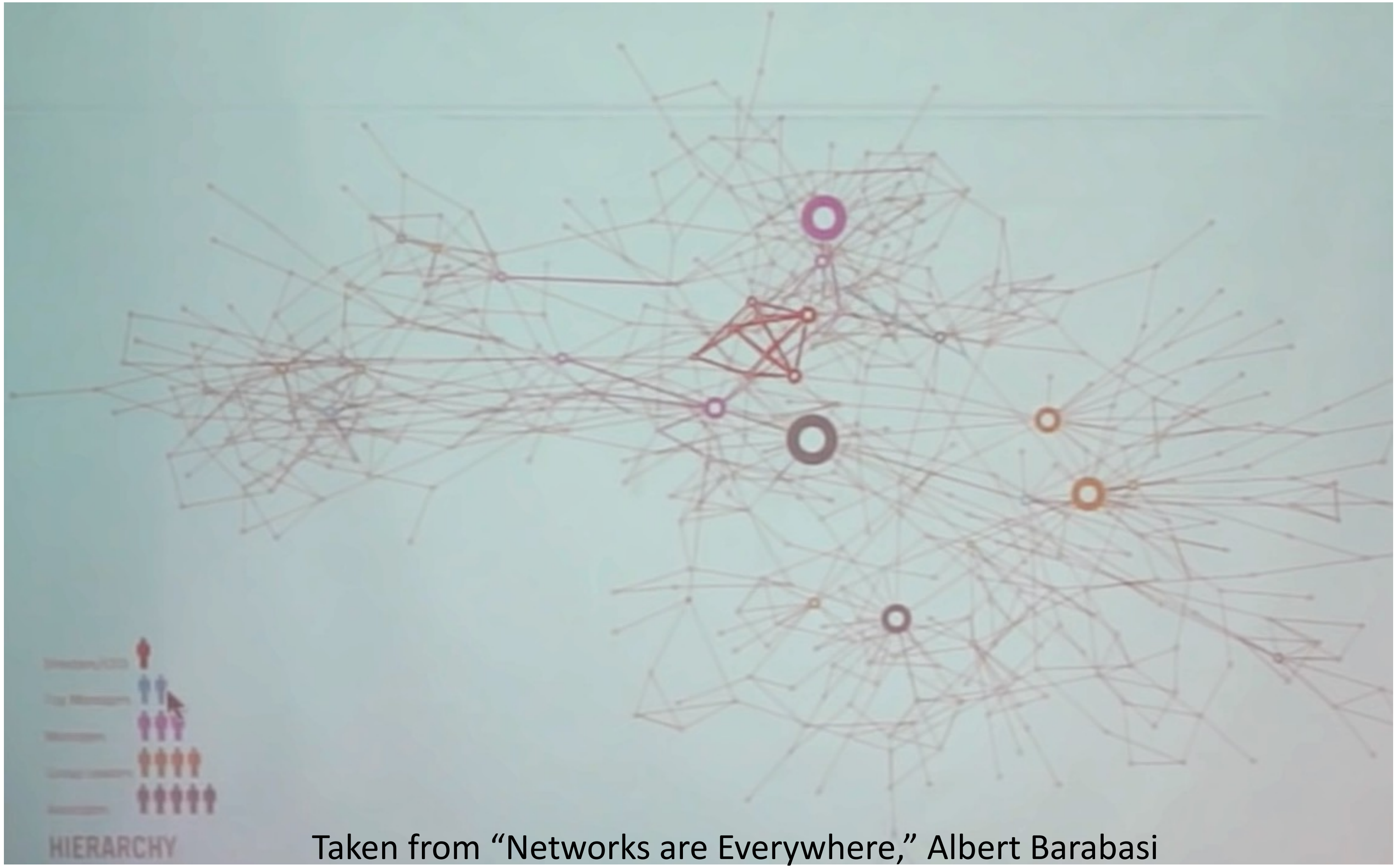
Color the nodes by the hierarchical org chart



Taken from “Networks are Everywhere,” Albert Barabasi

A management case study

Color the nodes
by their seniority in the
company org chart

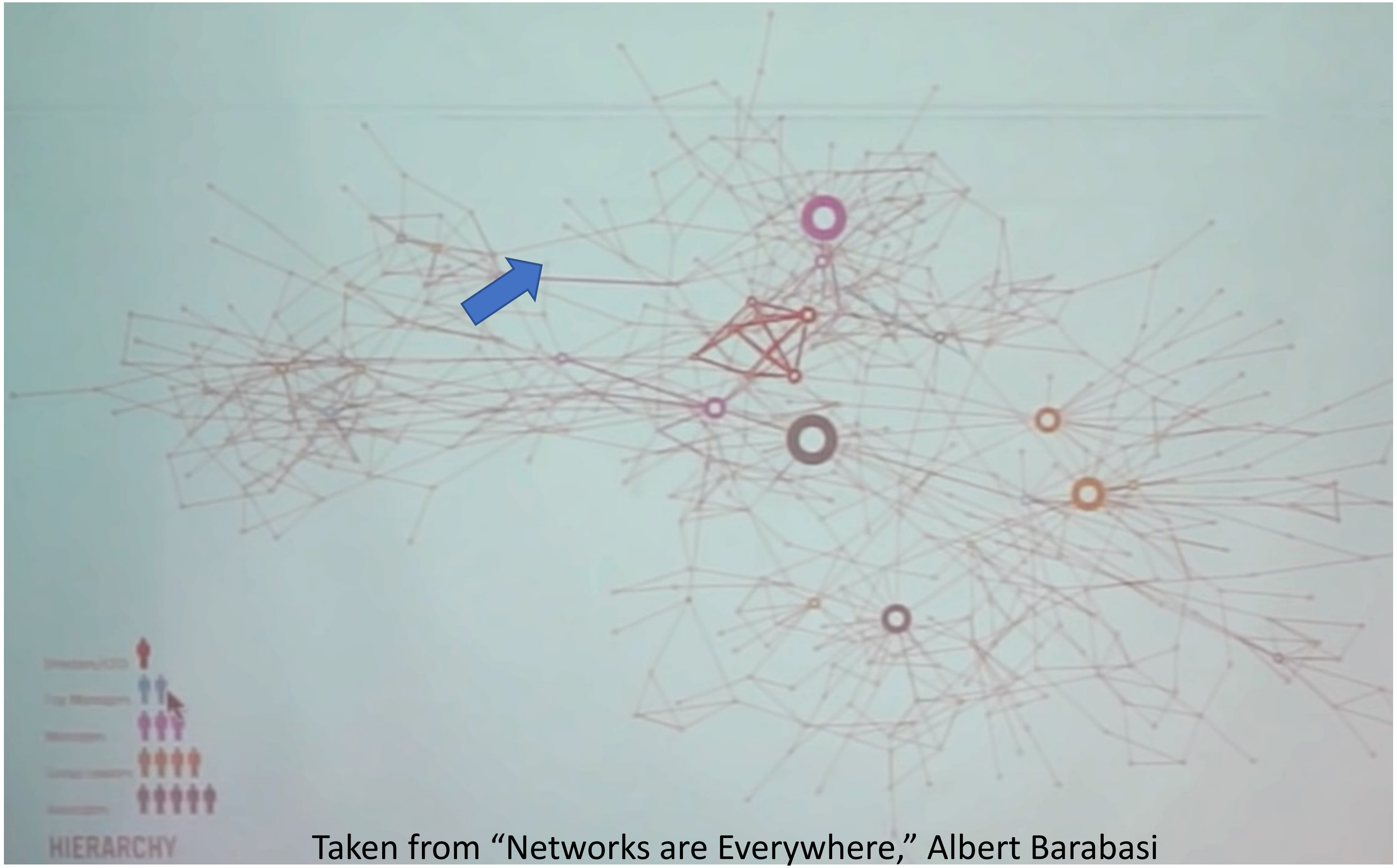


Taken from “Networks are Everywhere,” Albert Barabasi

A management case study

Color the nodes
by their seniority in the
company org chart

This is the company
director



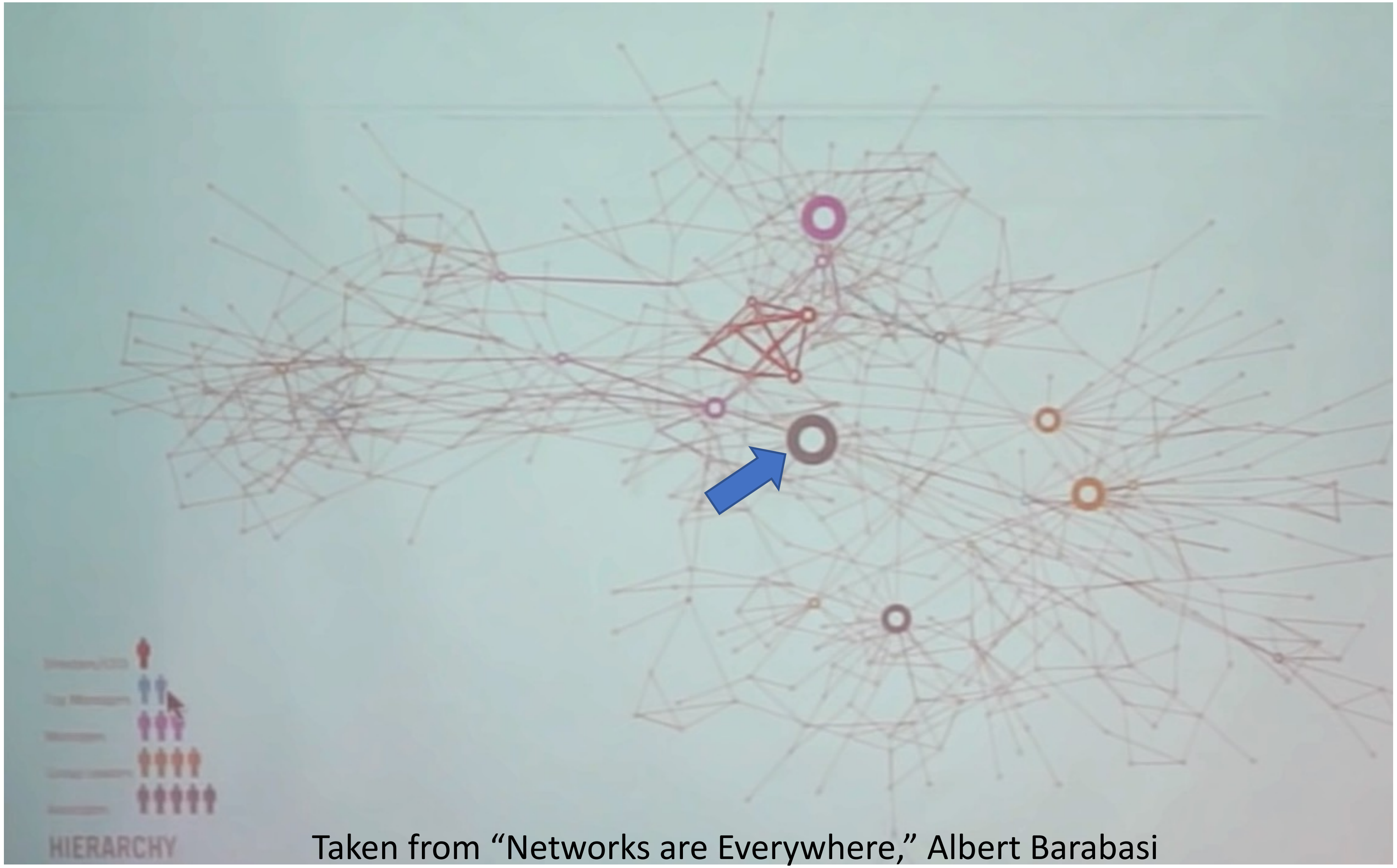
Taken from “Networks are Everywhere,” Albert Barabasi

A management case study

Color the nodes
by their seniority in the company org chart

This is the company director

Who is this??



Taken from “Networks are Everywhere,” Albert Barabasi

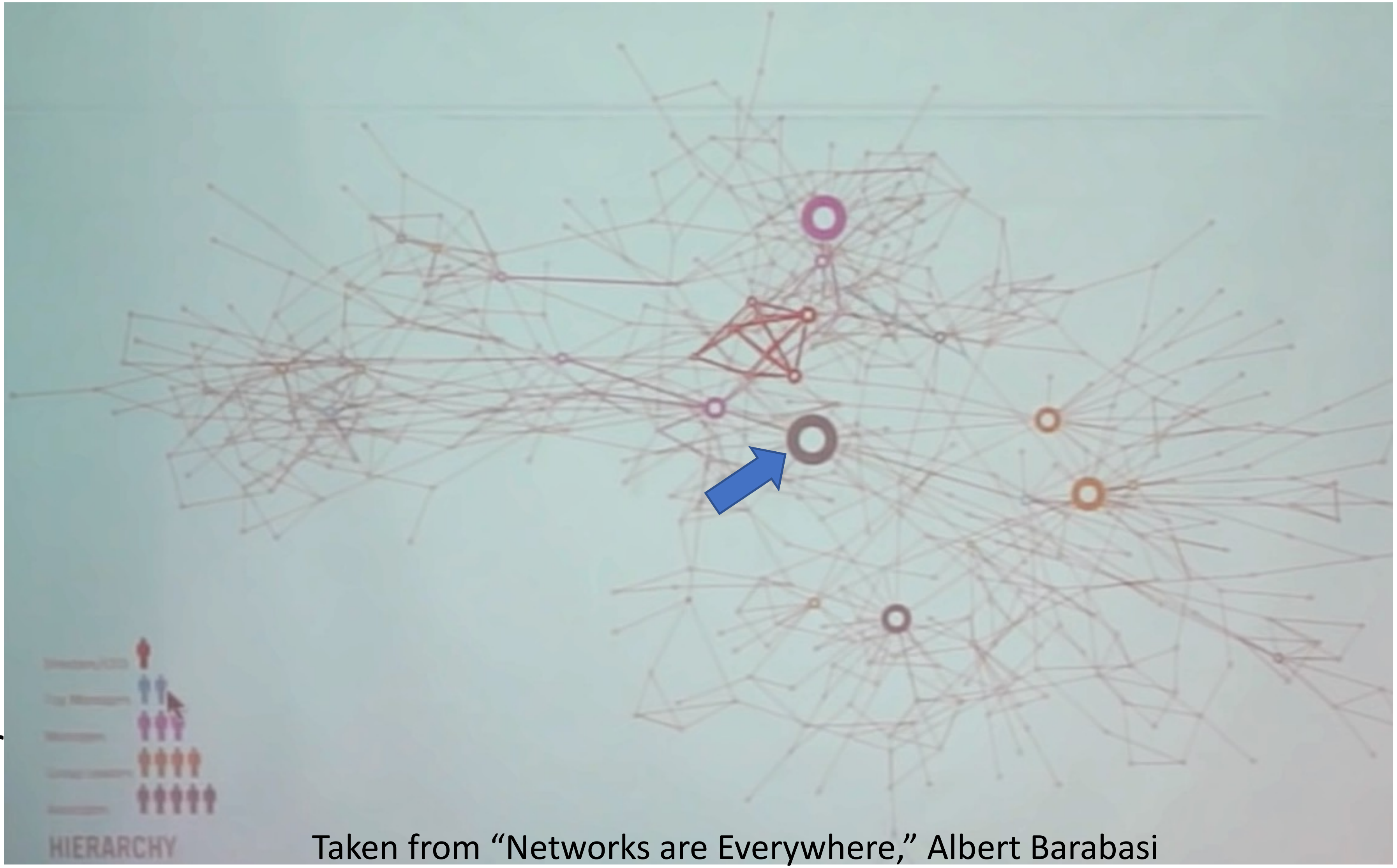
A management case study

Color the nodes
by their seniority in the
company org chart

This is the company
director

Who is this??

The safety officer who
visits all the departments
frequently but almost never
talks to management.

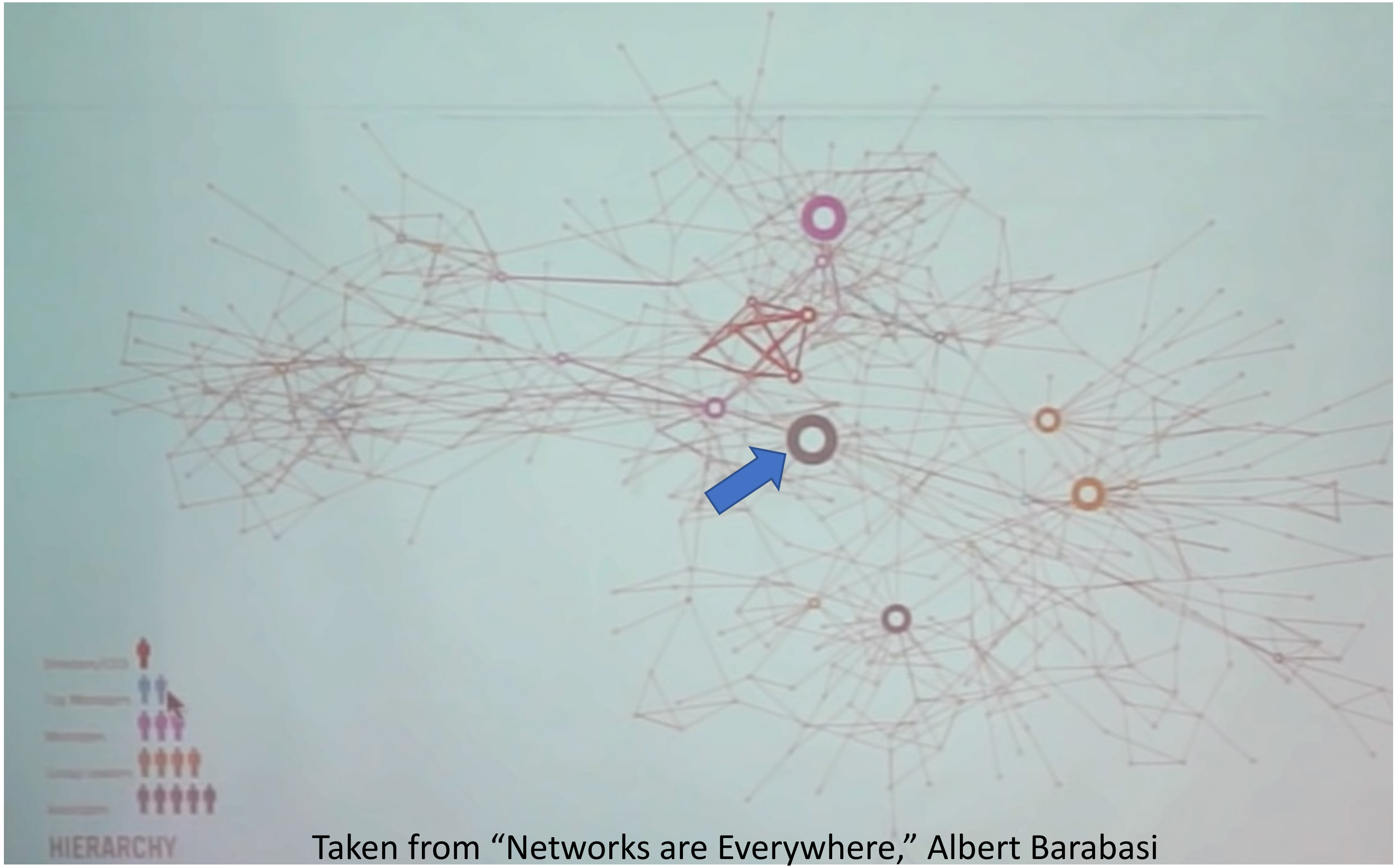


Taken from "Networks are Everywhere," Albert Barabasi

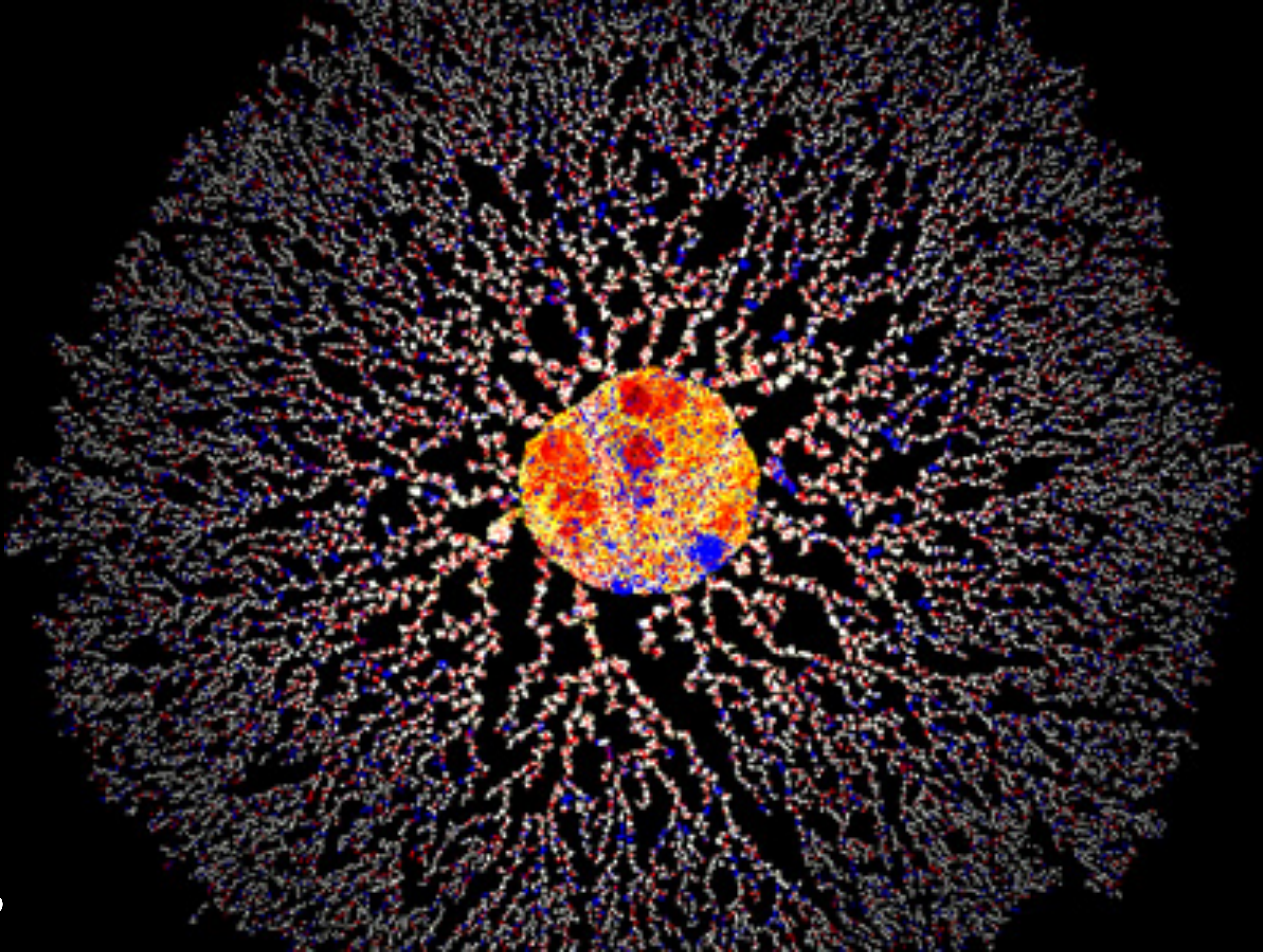
A management case study

The safety officer who visits all the departments frequenty but almost never talks to management.

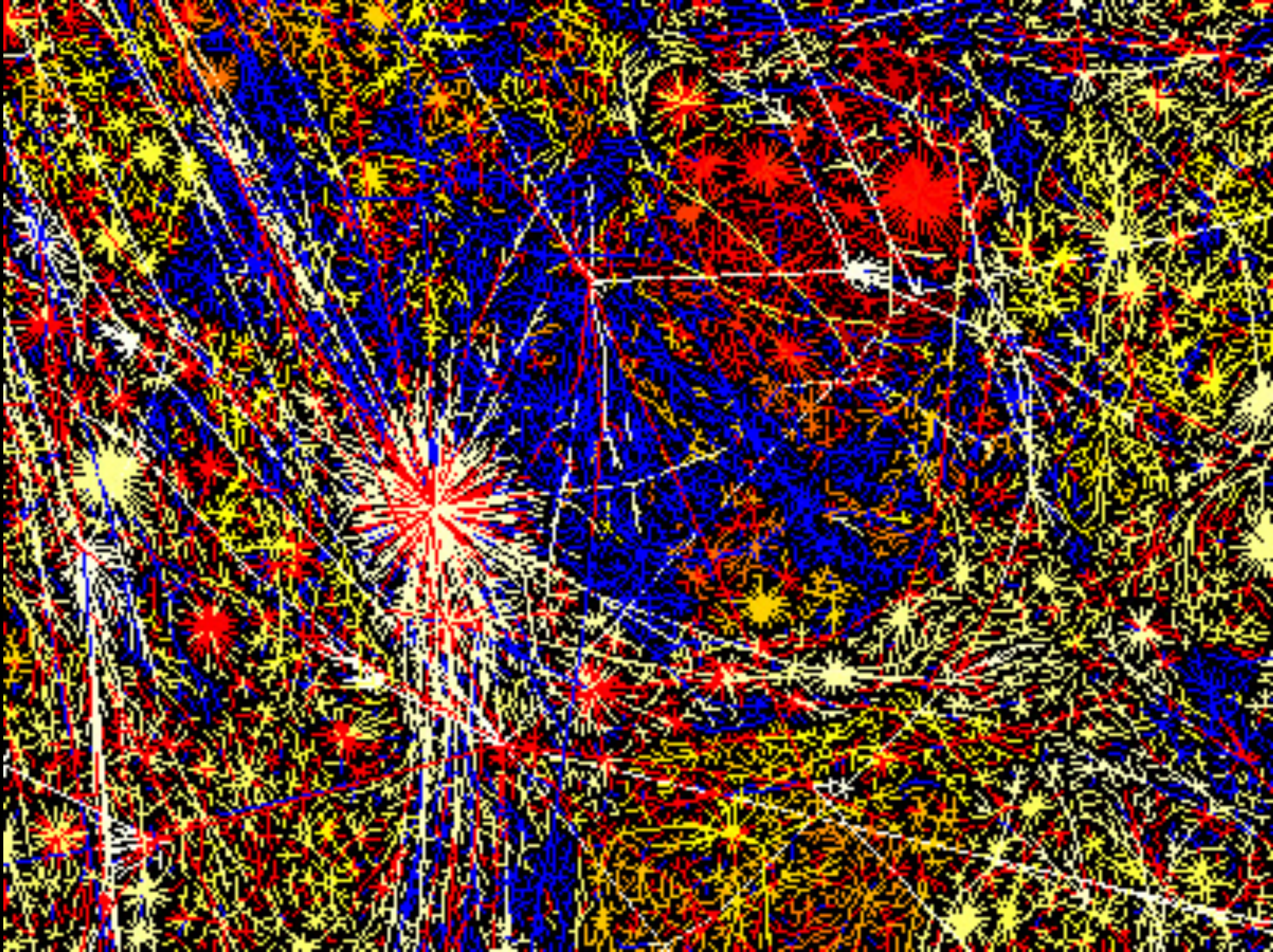
This is who is really running the company



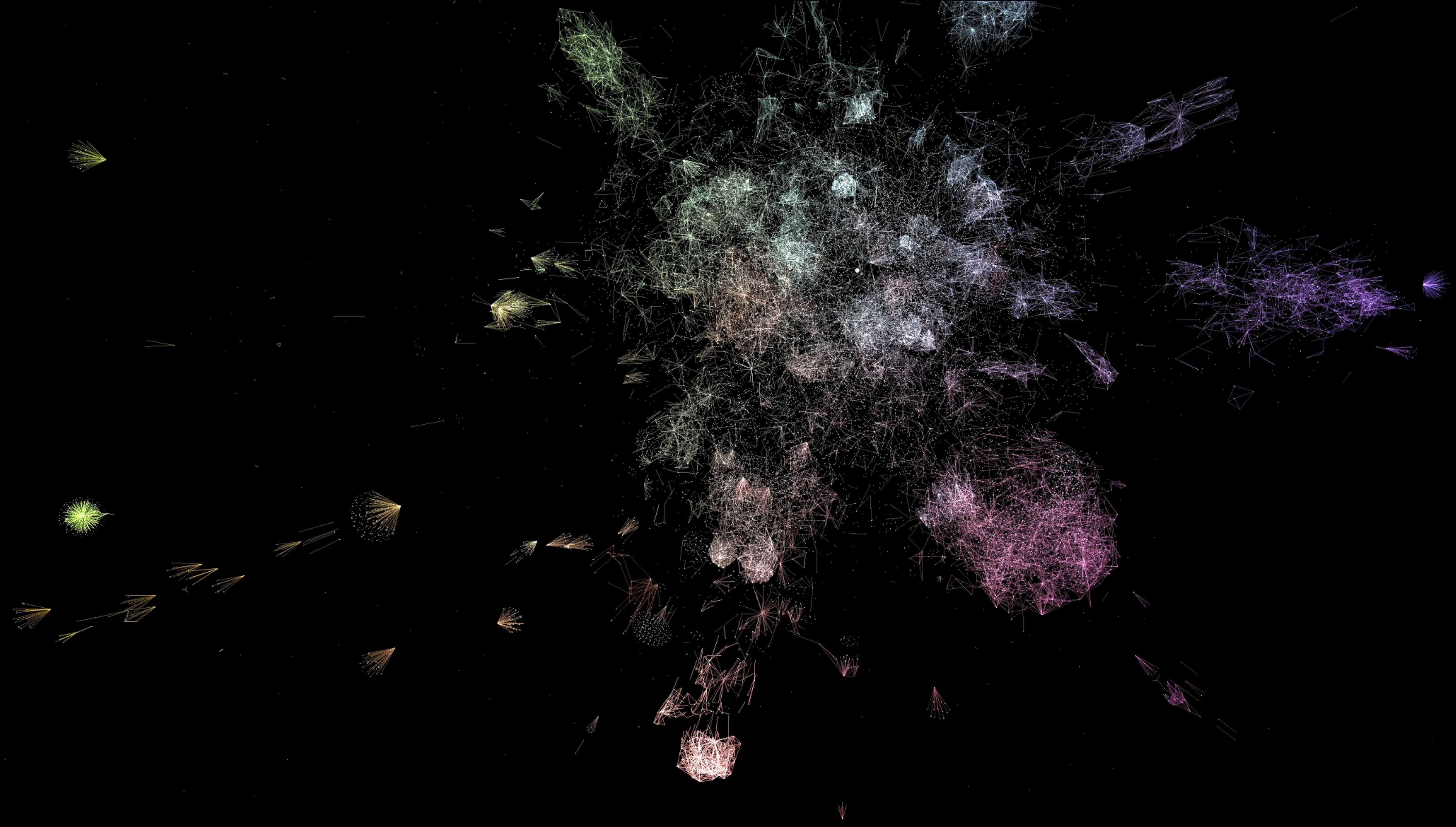
Taken from “Networks are Everywhere,” Albert Barabasi



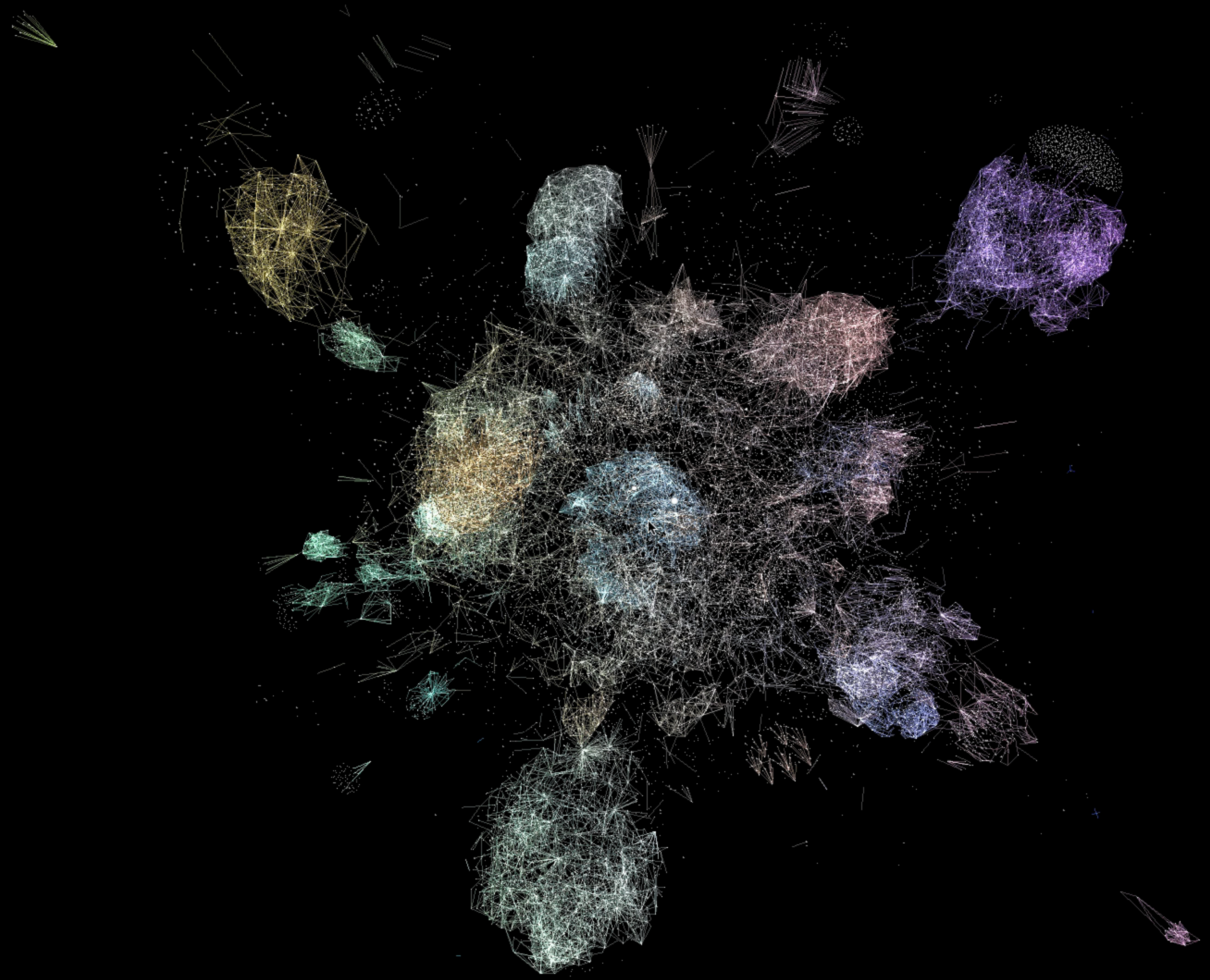
Institute of Integrated Bioinfo
Science



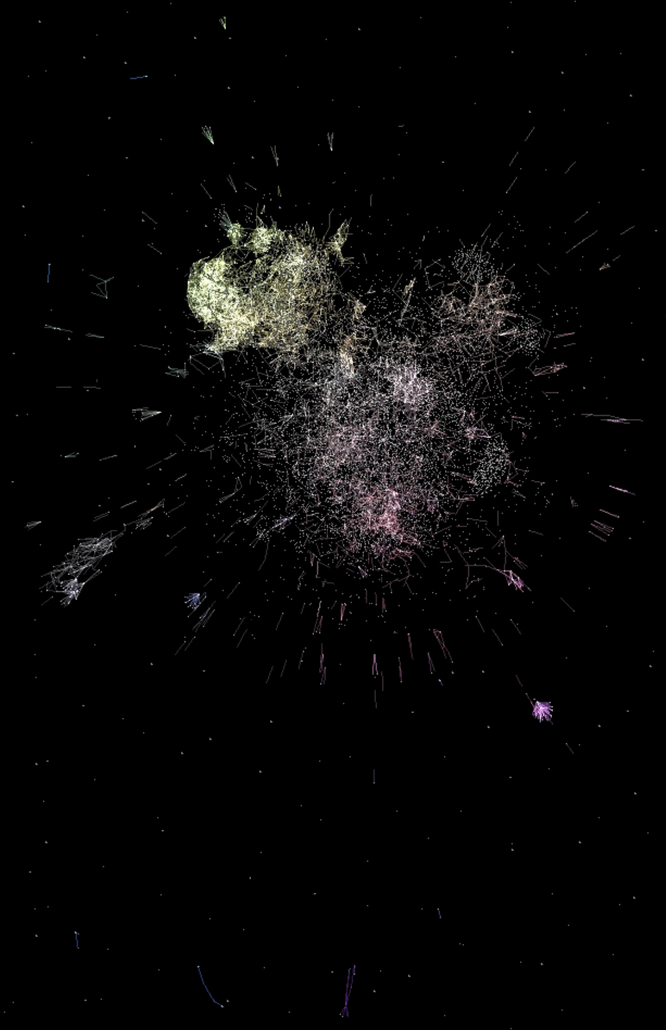
Zoom



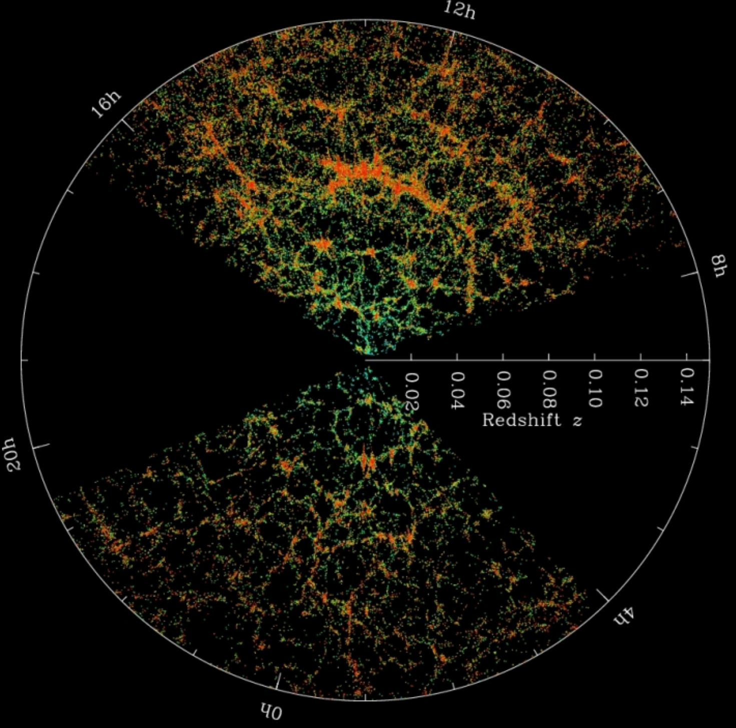
Andrei Kashcha, Code Galaxies, Debian



Fedora



Python



Sloan Sky Survey

Millenium Simulation, 10^9 agents (each agent is 10^9 solar masses, 2 billion l.y. cube)

Cosmic Web
Max Planck Society
Supercomputing Centre

Small world networks

- Closely related to scale-free networks
- The distance between any two nodes for a graph of size N is

$$L \propto \log(N)$$

- For example a million node graph would have an expected distance between any two nodes of 13.

Where do scale free networks come from?

- So the WWW has the same underlying mathematical structure as the cellular reaction networks evolution discovered… !

- Or… what is wrong with the Erdõs-Renyi graphs?

Where do scale free networks come from?

- So the WWW has the same underlying mathematical structure as the cellular reaction networks evolution discovered… !

- Or… what is wrong with the Erdõs-Renyi graphs?

- Two assumptions are not realistic for the systems we see:

1) Real graphs are dynamic. Nodes are added over time.
2) Existing nodes influence the connections of new nodes.

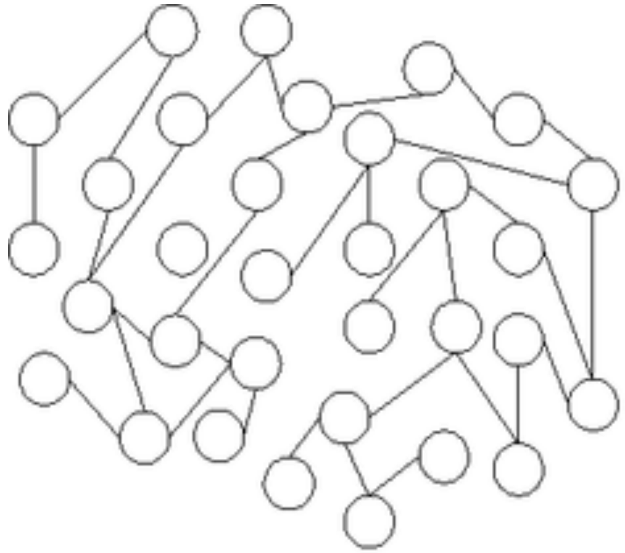
Preferential attachment is one way

- If the probability of adding a link between a new node and an existing one is biased by the number of edges the existing node has a scale free network results.

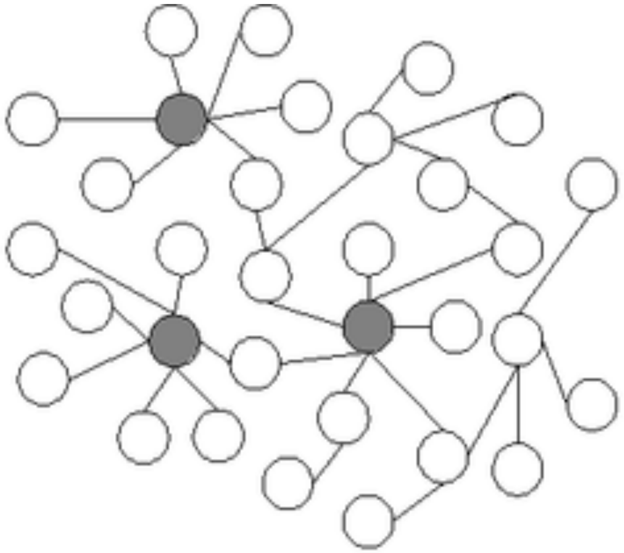
$$\Pi(k_j) = \frac{k_j}{\Sigma_j k_j}$$

The probability of connecting
to an existing node is proportional
to its degree.

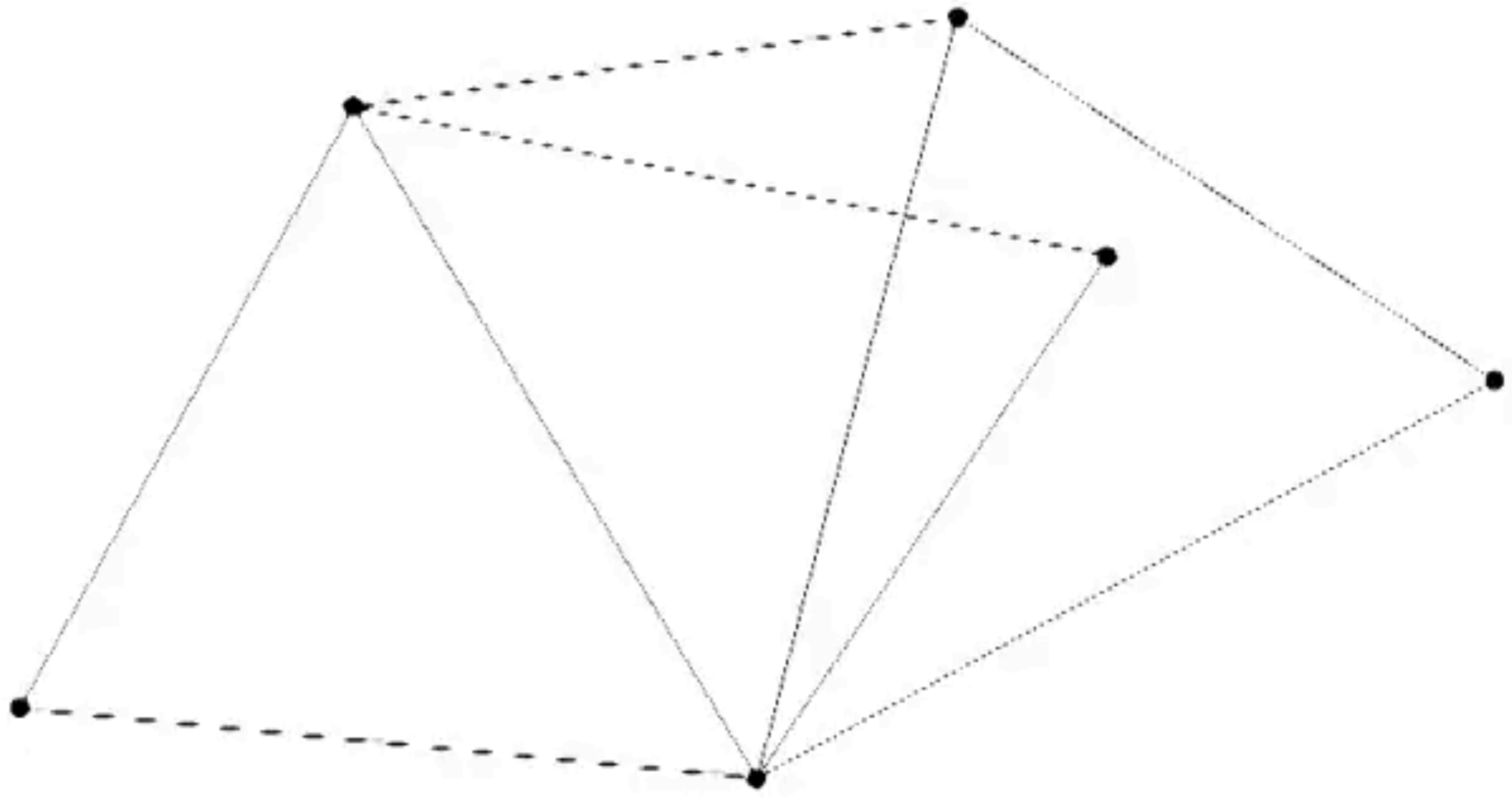
Positive feedback loop.



(a) Random network



(b) Scale-free network



Preferential Attachment

- Cell protein interactions
- The WWW
- Cosmic structures
- Information flow in a company
- Economics (income distribution)
- Disease propagation

Do we see why it matters yet?

- We have an explanation for how it happens
- But why does it happen? Why are they preserved rather than fought against?
- Scale free networks are *robust*

Do we see why it matters yet?

- We have an explanation for how it happens
- But why does it happen? Why are they preserved rather than fought against?
- Scale free networks are *robust*

Robustness

- At any given moment:
  - *Millions of failures are occurring in your cell biology*
  - *Thousands of routers on the internet are down or producing errors*

Robustness

- Consider knocking out nodes in networks
- For Erdõs-Renyi graphs and grids there are critical points where once enough nodes are removed the network collapses suddenly

Robustness

- Consider knocking out nodes in networks
- For Erdõs-Renyi graphs and grids there are critical points where once enough nodes are removed the network collapses suddenly

- Not true for scale-free networks
- This is because the chance of hitting an important hub is tiny

Robustness

- Consider knocking out nodes in networks
- For Erdõs-Renyi graphs and grids there are critical points where once enough nodes are removed the network collapses suddenly

- Not true for scale-free networks
- This is because the chance of hitting an important hub is tiny

- What is the obvious vulnerability with scale-free networks?

- Not robust to targeted attack.

Network Control

- How can we use the topologies of networks to control them?

Graph Theory and Real World Networks

- Social Networks - Stanley Milgram performed a by now famous experiment in the 1960s. He distributed a number of letters addressed to a stockbroker in Boston to a random selection of people in Nebraska. The task was to send these letters to the addressee (the stockbroker) via mail to an acquaintance of the respective sender. In other words, the letters were to be sent via a social network.

Graph Theory and Real World Networks

- Social Networks –

Stanley Milgram performed a by now famous experiment in the 1960s. He distributed a number of letters addressed to a stockbroker in Boston to a random selection of people in Nebraska. The task was to send these letters to the addressee (the stockbroker) via mail to an acquaintance of the respective sender. In other words, the letters were to be sent via a social network.

- Six Degrees of Separation

About 20% of Milgram's letters did eventually reach their destination. Milgram found that it had only taken an average of six steps for a letter to get from Nebraska to Boston. This result is by now dubbed "six degrees of separation" and it is possible to connect any two persons living on earth via the social network in a similar number of steps.

The Small-World Effect

- The “small-world effect” denotes the result that the average distance linking two nodes belonging to the same network can be orders of magnitude smaller than the number of nodes making up the network.

Foundations: Thermodynamic Limit

- Mathematical graph theory is often concerned with the thermodynamic limit.
- *The Thermodynamic Limit.* The limit where the number of elements making up a system diverges to infinity is called the "thermodynamic limit" in physics.
- A quantity is *extensive* if it is proportional to the number of constituting elements
- *intensive* if it scales to a constant in the thermodynamic limit.

Foundations

- *Coordination Number* is the average degree of the nodes
- This along with the number of nodes, N, in the graph characterise random graphs

Foundations

- *Coordination Number* is the average degree of the nodes
- This along with the number of nodes, N, in the graph characterise random graphs
- Alternatively the *Connection Probability* is the probability, *p*, that an edge occurs between a pair of nodes.

Foundations

- For Erdös-Renyi Random graphs:

$$p = \frac{zN}{2} \frac{2}{N(N-1)}$$

$$p = \frac{z}{N-1}$$

Network Diameter

- The network diameter is the maximum degree of separation between all pairs of vertices.

Network Diameter

- The network diameter is the maximum degree of separation between all pairs of vertices.
- For a random network with *N* vertices and coordination number *z* we have

$$z^D \approx N$$

- Since any node has z neighbors, z^2 next-nearest neighbors, etc.
- *D* is the network diameter

Network Diameter

- The network diameter is the maximum degree of separation between all pairs of vertices.
- For a random network with N vertices and coordination number z we have

$$z^D \approx N$$

$$D \propto \frac{\log N}{\log z}$$

Average Distance

- *Average Distance.* The average distance l is the average of the minimal path length between all pairs of nodes of a network.

Average Distance

- *Average Distance.* The average distance l is the average of the minimal path length between all pairs of nodes of a network.

- The average distance l is generally closely related to the diameter *D*; it has the same scaling as the number of nodes *N*.

Clustering

- Real networks have strong local recurrent connections. Recall the cellular chemical reaction networks and Debian package networks.

- This leads to topological features such as loops and clusters.

Clustering

- *The Clustering Coefficient.* The clustering coefficient *C* is the average fraction of pairs of neighbors of a node that are also neighbors of each other.

Clustering

- *The Clustering Coefficient.* The clustering coefficient *C* is the average fraction of pairs of neighbors of a node that are also neighbors of each other.

- The clustering coefficient is a normalized measure of loops of length 3

Clustering

- *The Clustering Coefficient.* The clustering coefficient C is the average fraction of pairs of neighbors of a node that are also neighbors of each other.

- The clustering coefficient is a normalized measure of loops of length 3.

- In a Erdös-Renyi random graph a typical site has $z(z\ 1)/2$ pairs of neighbors.
- The probability of an edge to be present between a given pair of neighbors is $p = z/(N - 1)$

Clustering

- In a Erdös-Renyi random graph a typical site has $z(z\ 1)/2$ pairs of neighbors.
- The probability of an edge to be present between a given pair of neighbors is $p = z/(N - 1)$

$$C_{\text{rand}} = \frac{z}{N-1} \approx \frac{z}{N}$$

Clustering

- In a Erdös-Renyi random graph a typical site has $z(z\ 1)/2$ pairs of neighbors.
- The probability of an edge to be present between a given pair of neighbors is $p = z/(N - 1)$

$$C_{\mathrm{rand}} = \frac{z}{N-1} \approx \frac{z}{N}$$

- This scales to zero in the thermodynamic limit. $$\lim_{N \to 0} C_{\mathrm{rand}} \to 0$$

- Since the clustering coefficient is just the probability of a pair of neighbors being interconnected.

Cliques and Communities

- *Cliques.* A clique is a set of vertices for which (a) every node is connected by an edge to every other member of the clique and (b) no node outside the clique is connected to all members of the clique.

Cliques and Communities

- The term “clique” comes from social networks. A clique is a group of friends where everybody knows everybody else.

- The number of cliques of size *K* in an Erdös–Renyi graph with *N* vertices and linking probability *p:*

$$\binom{N}{K} p^{K(K-1)/2} (1 - p^K)^{N-K}$$

Communities

- Collections of strongly connected cliques are called *communities.*

Network	N	*l*	C	C_{rand}
Movie actors collaborations	225226	3.65	0.79	0.00027
Neural Network of *C. elegans*	282	2.65	0.28	0.05
US Western Power grid	4941	18.7	0.08	0.0005

Small values of *l* indicate these are small world networks

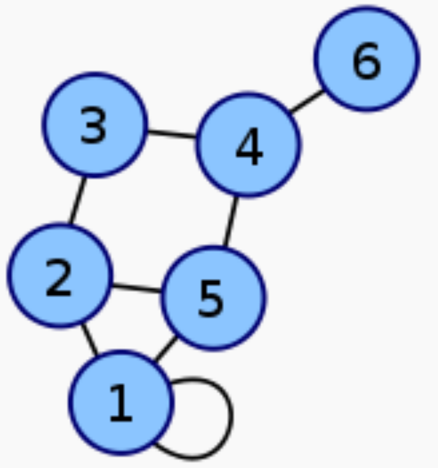
Watts and Strogatz, 1998

Graph Spectra

- Any graph *G* with *N* nodes can be represented by a matrix encoding the topology of the network, the adjacency matrix.
- *The Adjacency Matrix.* The $N \times N$ adjacency matrix *A* has elements *Aij* =1 if nodes *I* and *j* are connected and *Aij* =0 if they are not connected.

This matrix has N eigenvalues.

In matlab use the **eig** function to calculate the eigenvalues.

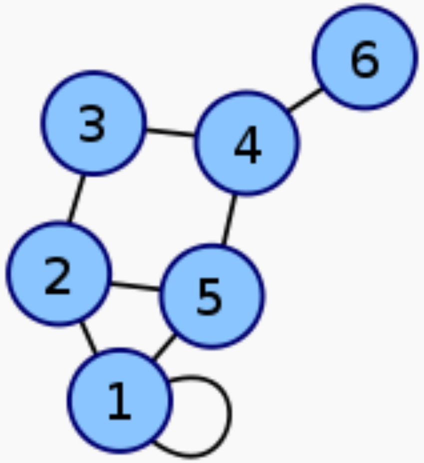
Labeled graph	Adjacency matrix
	$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

Graph Spectra

- *The Spectrum of a Graph.* The spectrum of a graph *G* is given by the set of eigenvalues λ_i of the adjacency matrix *A*.

This matrix has N eigenvalues.

In matlab use the **eig** function to calculate the eigenvalues.

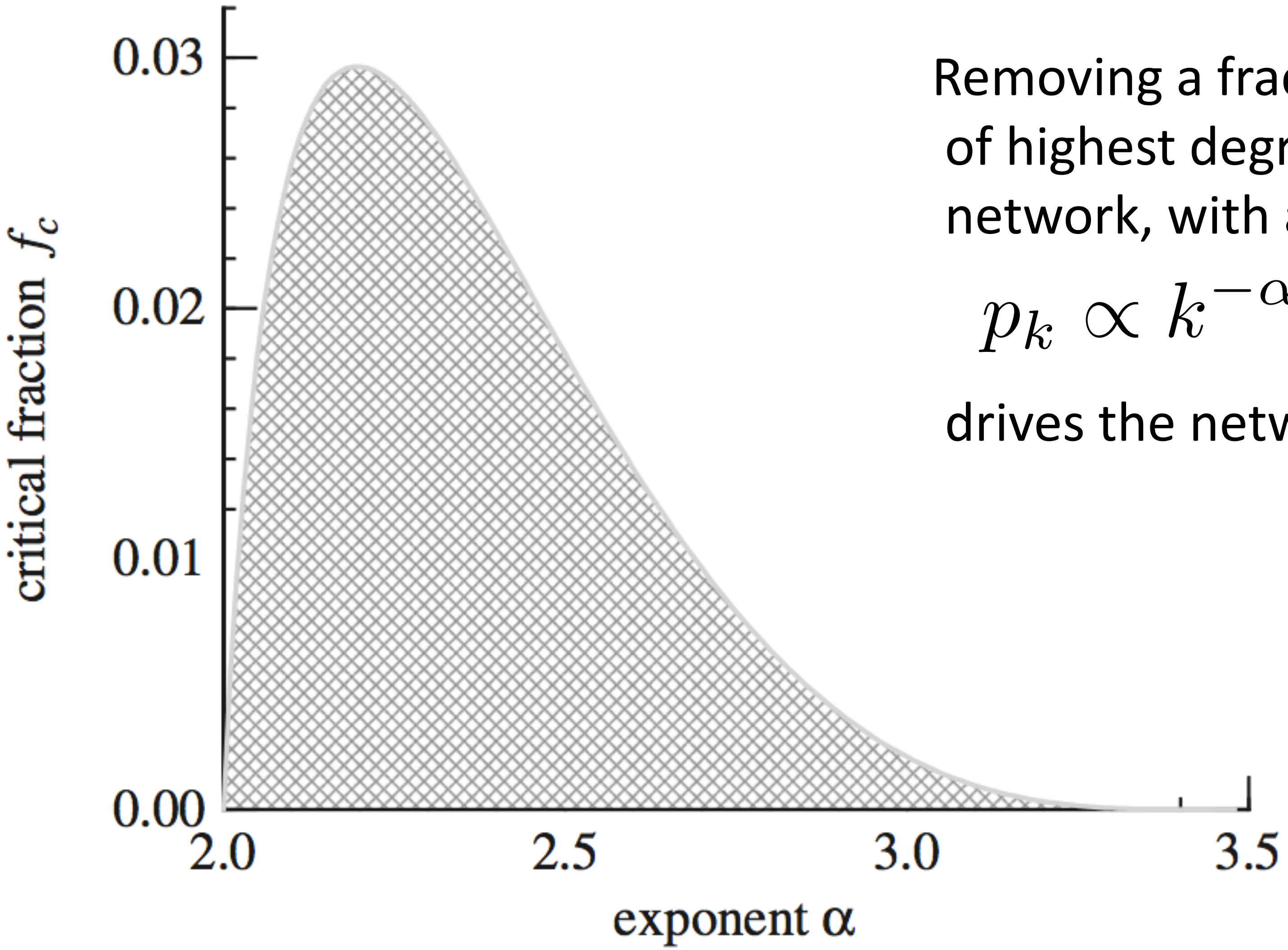
Labeled graph	Adjacency matrix
	$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

Application Example

- The largest eigenvalue λ plays an important role in modelling virus propagation in computer networks.
- The smaller the largest eigenvalue, the larger the robustness of a network against the spread of viruses.
- In fact, it was shown in that the epidemic threshold in spreading viruses is proportional to $1/\lambda$
- This allows the design of graphs with minimal λ given numbers of vertices and edges, and having a given diameter, *D*

Wang Y., Chakrabarti D., Wang C., Faloutsos C., Epidemic spreading in real networks: An eigenvalue viewpoint, 22nd Symp. Reliable Distributed Computing, Florence, Italy, Oct. 6–8, 2003.

Robustness in Scale Free Graphs



Removing a fraction greater than *fc* of highest degree vertices from a scale-free network, with a power-law degree distribution

$$p_k \propto k^{-\alpha}$$

drives the network below the percolation limit.

Newman 2002

Boolean Networks

- Problem in the 1960s. How can a single genome give rise to multiple cell types.
- Organisms have one set of genes but produce liver cells, skin cells, T cells, neurons, etc. …

Stuart Kauffman

Boolean Networks

- Kauffman built Boolean networks as sets of switches with lightbulbs.
- The lightbulbs were connected randomly with wire.

Stuart Kauffman

Boolean Networks

- Kauffman built Boolean networks as sets of switches with lightbulbs.
- The lightbulbs were connected randomly with wire.
- The network is updated *synchronously.*
- The on/off state of the lightbulb depends on the incoming signals from neighbouring nodes.

Stuart Kauffman

Boolean Networks

- Kauffman built Boolean networks as sets of switches with lightbulbs.
- The lightbulbs were connected randomly with wire.
- The network is updated *synchronously.*
- The on/off state of the lightbulb depends on the incoming signals from neighbouring nodes.
- Each node has a random Boolean function that maps its input to its state (or, and, etc).

Stuart Kauffman

Boolean Networks

- These are also called N-K networks because they are defined by, N, the number of nodes and K, the degree of the nodes (recall the coordination number).

Stuart Kauffman

Boolean Networks

- Boolean networks are a generalisation of cellular automata.
- The can exhibit many of the same dynamics (attractors, chaos, criticality, order, etc.)
- Though there are only 2^N possible states for the the network. Once any one of those states is repeated the whole history repeats.

Stuart Kauffman

Boolean Networks

- The dynamic on/off patterns in the same Boolean network initialised with slightly different values can vary greatly.
- Some networks fall into patterns of activation that are stable over time (dynamical attractors).
- This is exactly how we now understand gene regulatory networks to operate.

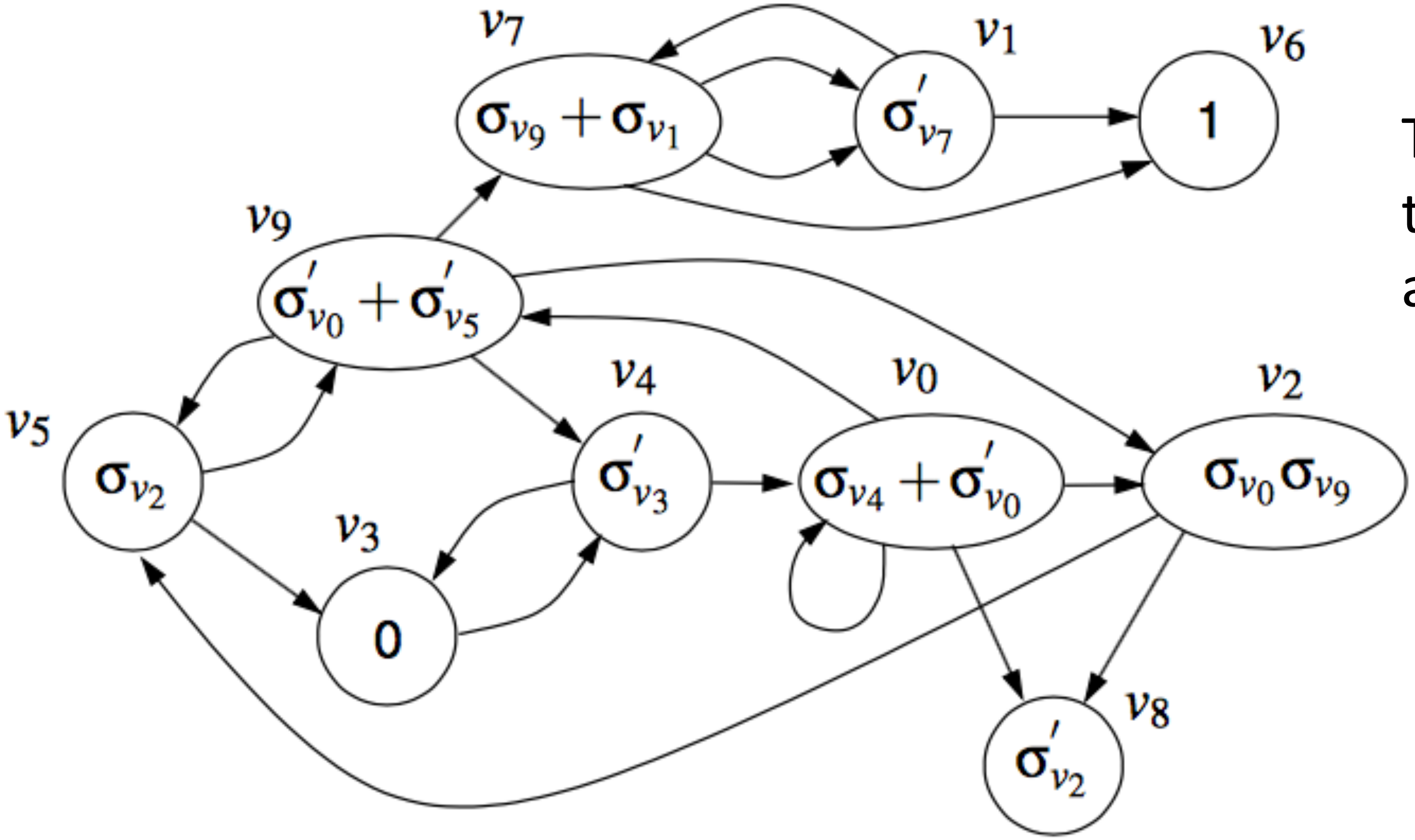
Stuart Kauffman

Boolean Variables and Graph Topologies

- *Boolean Variables.* A Boolean or binary variable has two possible values, typically 0 and 1.
- *Boolean Coupling Functions.* A Boolean function $\{0, 1\}^K$ -> $\{0, 1\}$ maps K Boolean variables onto a single one.
- The dynamics has discrete time (map not a flow).

Boolean Variables and Graph Topologies

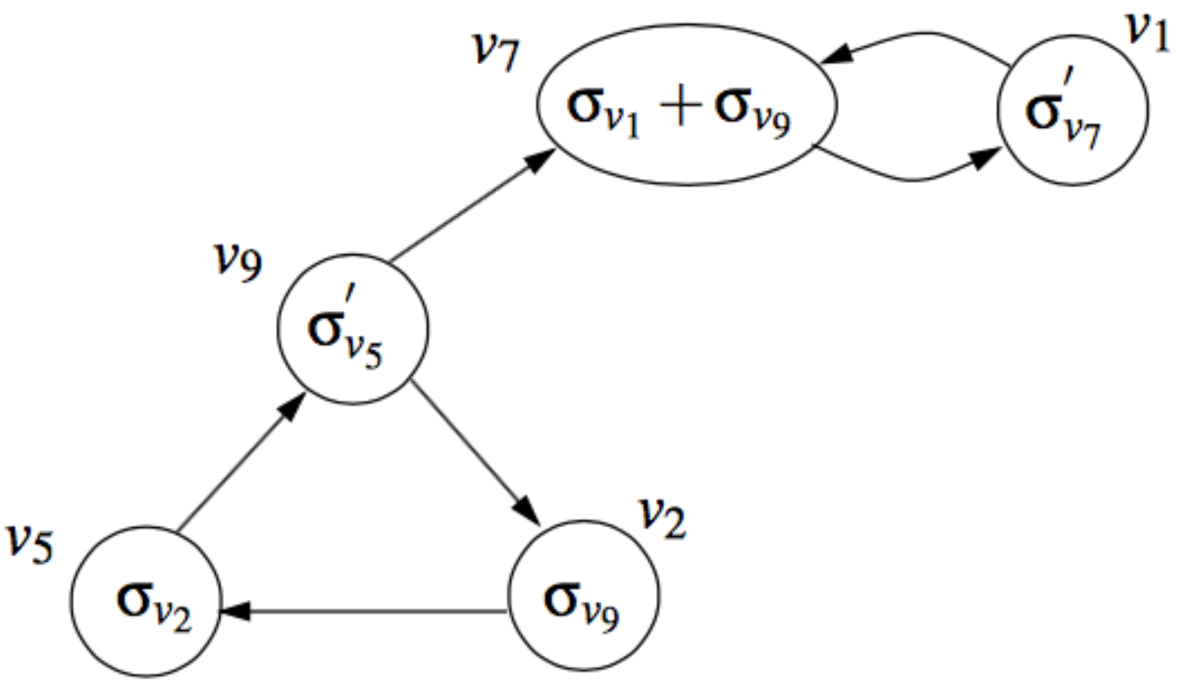
- *Boolean Variables.* A Boolean or binary variable has two possible values, typically 0 and 1.
- *Boolean Coupling Functions.* A Boolean function $\{0, 1\}^K$ -> {0, 1} maps *K* boolean variables onto a single one.
- The dynamics has discrete time (map not a flow).
- *The Boolean Network.* The set of Boolean coupling functions inter-connecting the *N* Boolean variables can be represented graphically by a directed network, the Boolean network.



They use “·”, “+” and “′” to denote the Boolean operations AND, OR and NOT, respectively.

Fig. 1. Example of a Kauffman network. The state of a vertex v_i at time $t+1$ is given by $\sigma_{v_i}(t+1) = f_{v_i}(\sigma_{v_l}(t), \sigma_{v_r}(t))$, where v_l and v_r are the predecessors of v_i, and f_{v_i} is the Boolean function associated to v_i.

Dubrova, Elena, Maxim Teslenko, and Andres Martinelli. "Kauffman networks: Analysis and applications." *Proceedings of the 2005 IEEE/ACM International conference on Computer-aided design*. IEEE Computer Society, 2005.

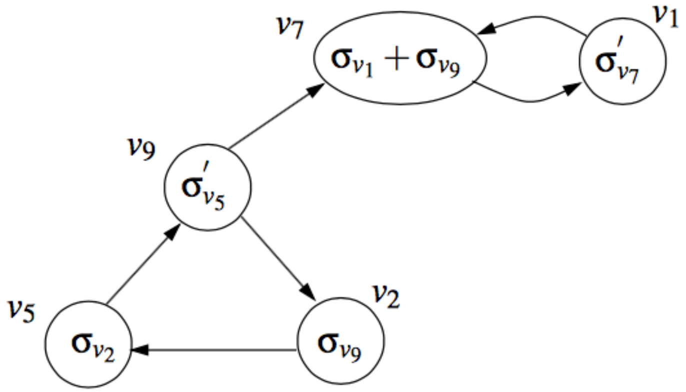


They use “·”, “+” and “′” to denote the Boolean operations AND, OR and NOT, respectively.

Reduce the network by removing Redundant states.

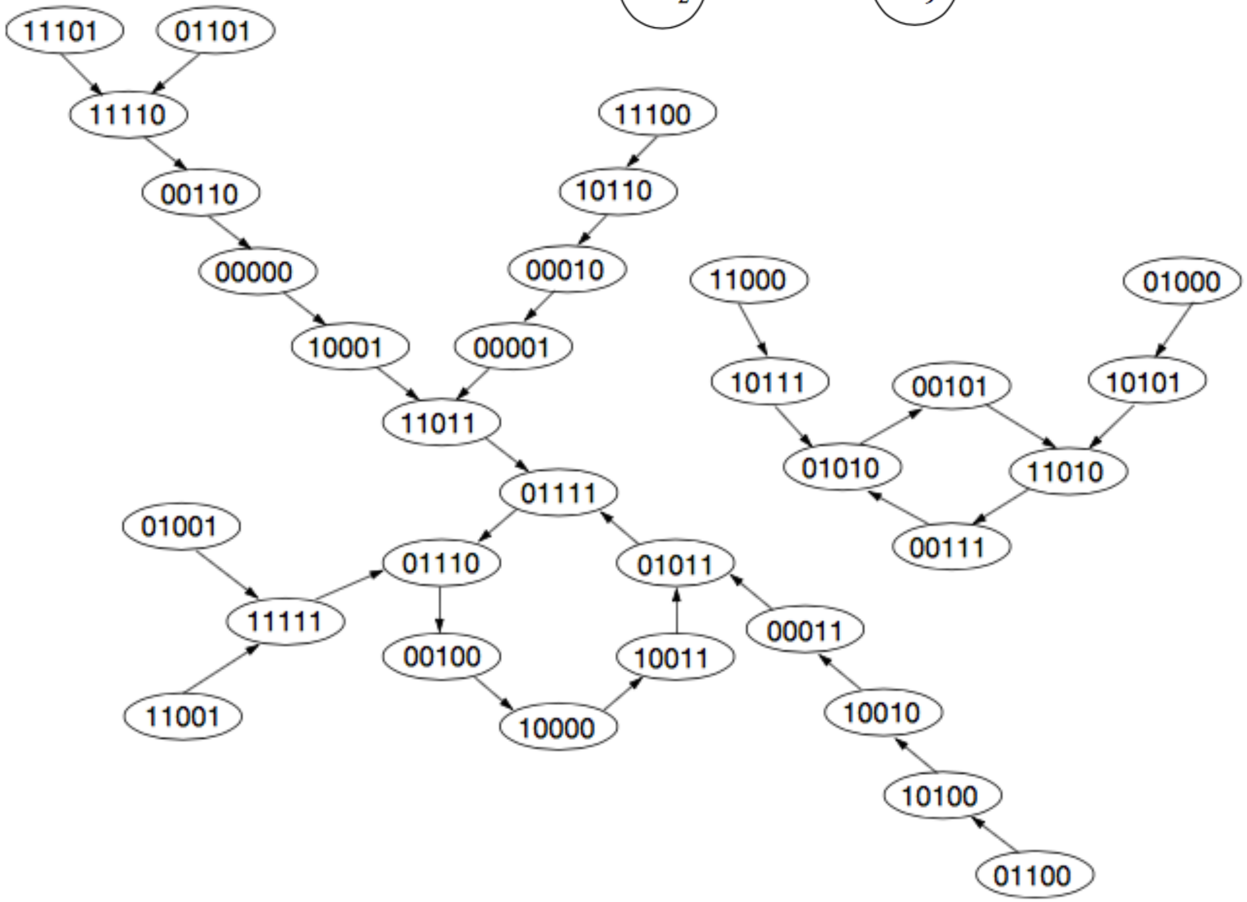
Dubrova, Elena, Maxim Teslenko, and Andres Martinelli. "Kauffman networks: Analysis and applications." *Proceedings of the 2005 IEEE/ACM International conference on Computer-aided design*. IEEE Computer Society, 2005.

State Transition Graph



We define another graph that shows the possible states and transitions of the reduced Boolean network.

Each node of this network is a 5-tuple corresponding to the state of the nodes in the Boolean graph.



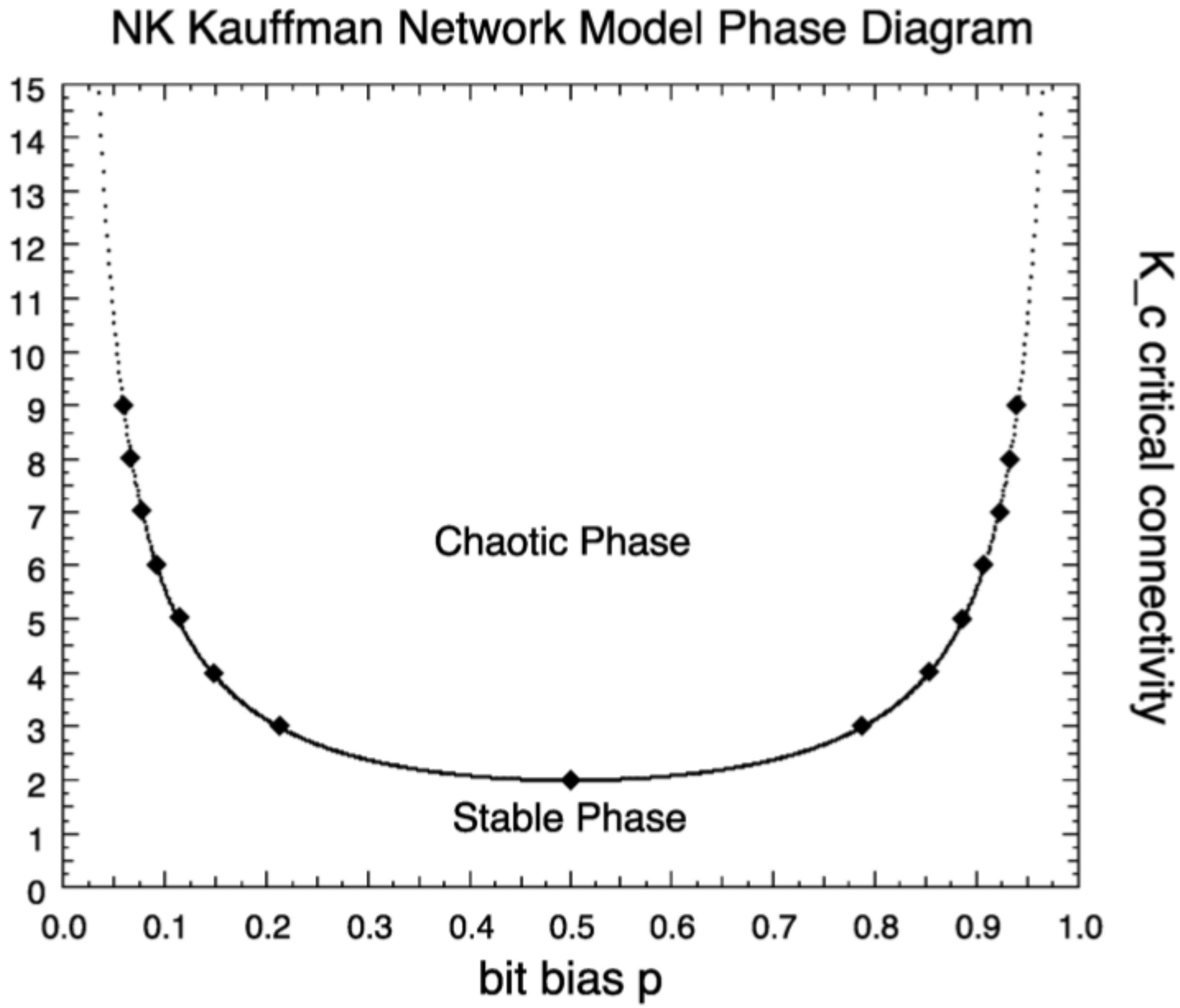
N-K Network Criticality

In Boolean networks we can define a value *p.*

p is the mean Boolean output of the Boolean functions over all the nodes.

For p near 0.5 there is a critical region of Boolean networks.

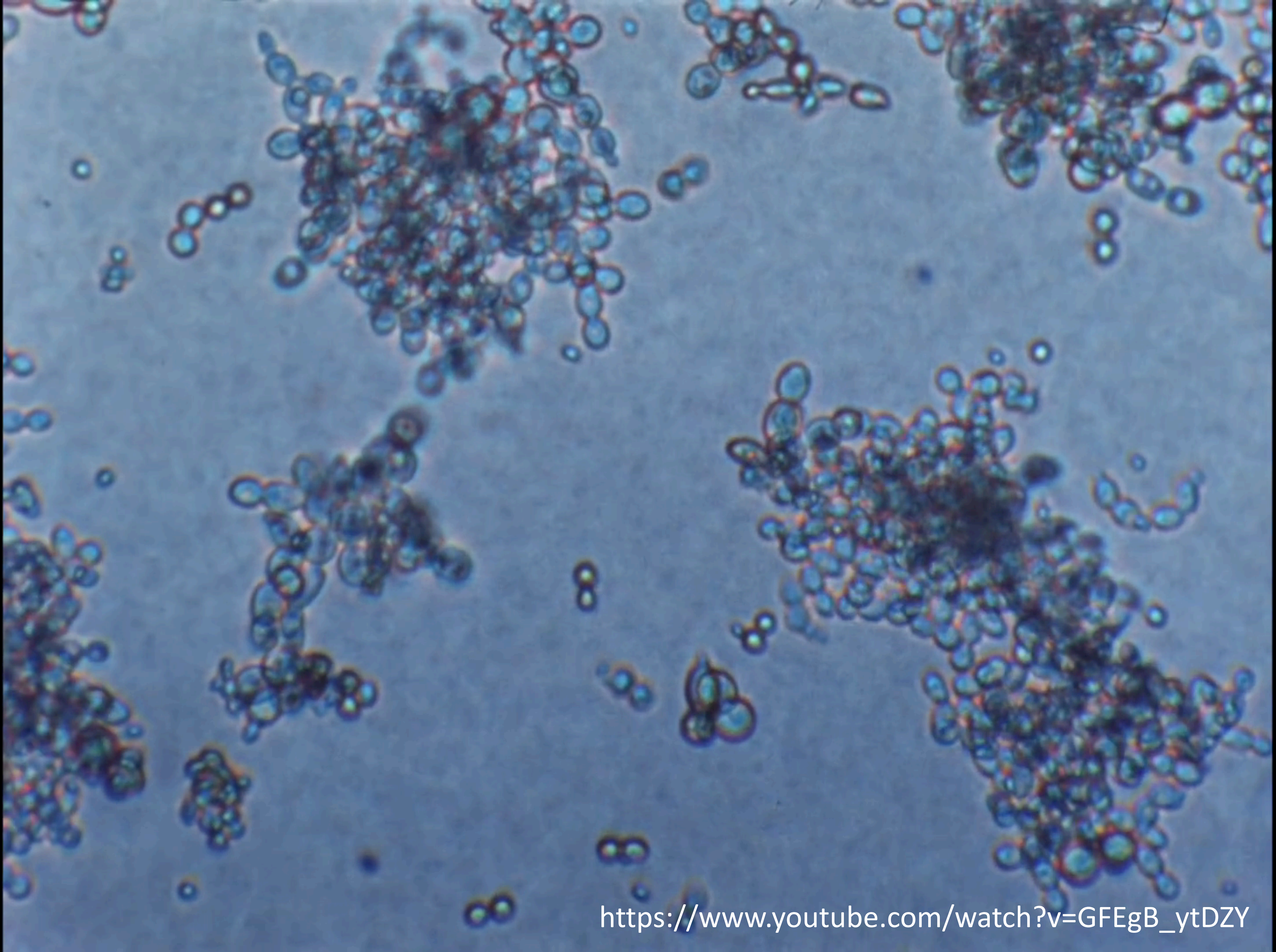
This critical region is the phase transition from stability to chaos.



Hawick, K. A., H. A. James, and C. J. Scogings. "Circuits, attractors and reachability in mixed-k kauffman networks." *arXiv preprint arXiv:0711.2426*(2007).

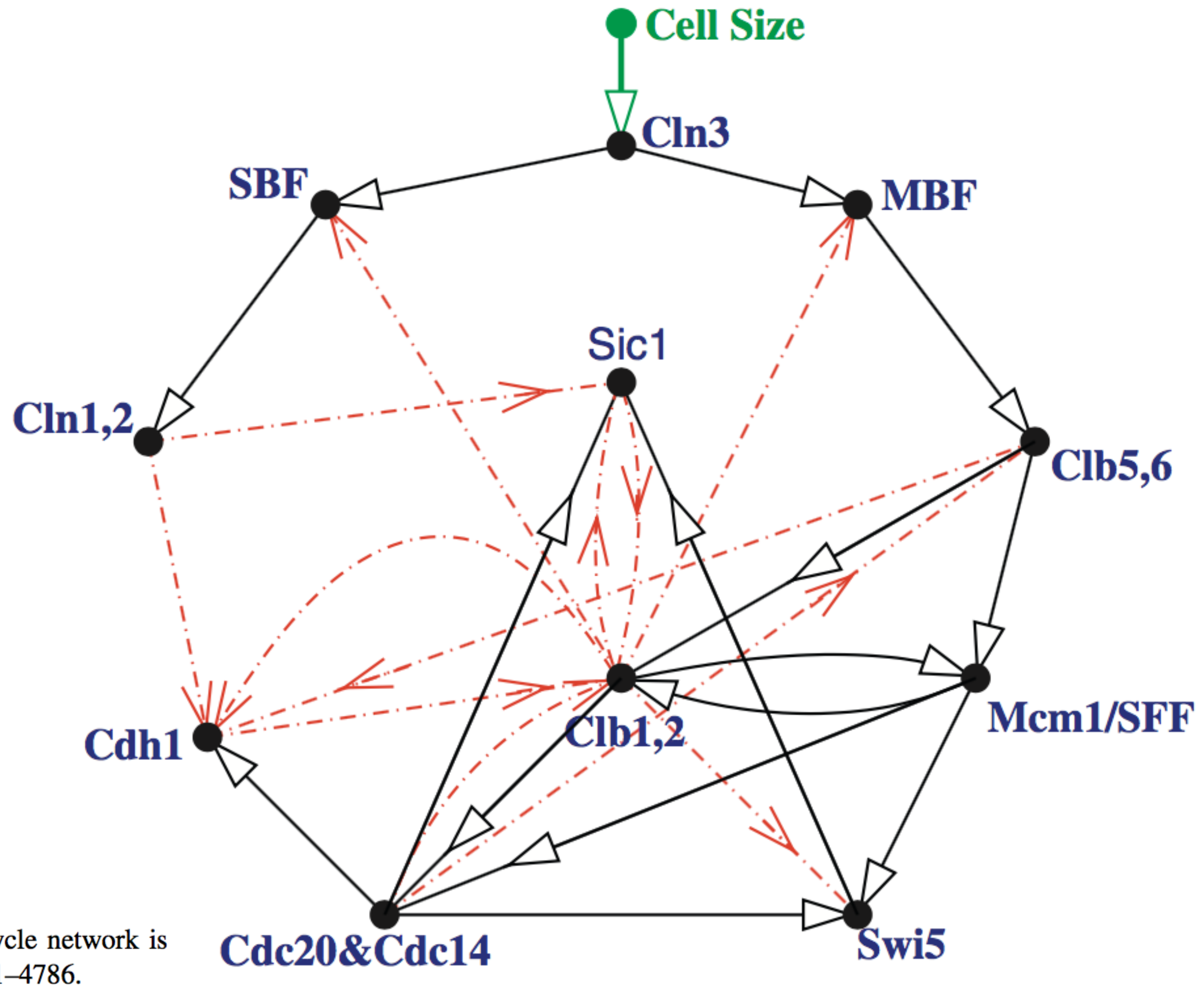
N-K Network Sim

../Code/kauffman/kauffman.html


https://www.youtube.com/watch?v=GFEgB_ytDZY

Reduced Network in Yeast Life Cycle

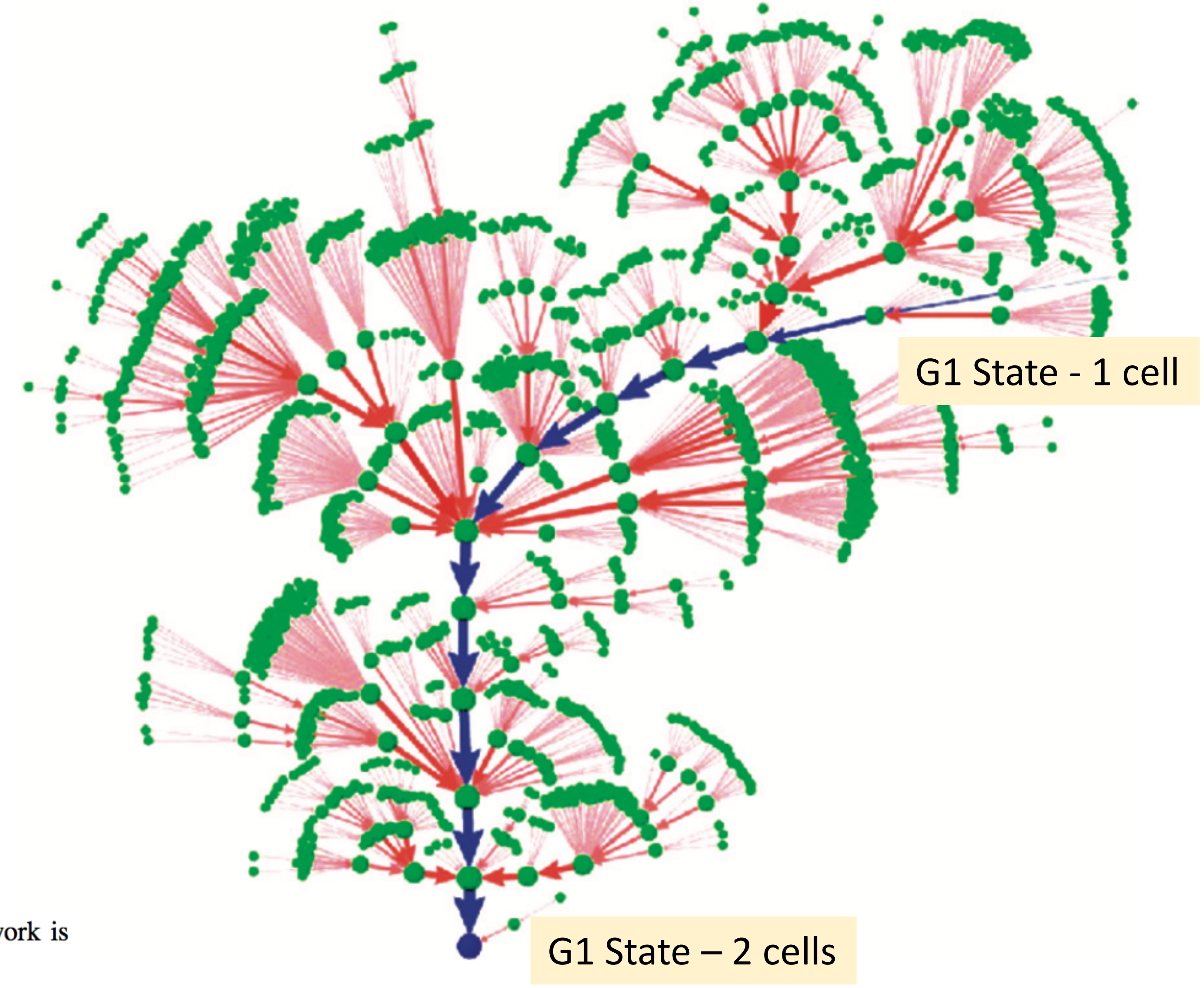
- Gene expression network.
- Dashed arrows are inhibitory (not gates)
- Solid arrows are excitatory (true)



LI, F., LONG, T., LU, Y., OUYANG, Q., TANG, C. 2004 The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy Science* **101**, 4781–4786.

State Transition Graph in Yeast Life Cycle

- Gene expression network.
- As genes are swtiched on and off the blue arrows show transitions to the G1 attractor.
- Green: 1764 gene expression states



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