Vniversitat de València Máster Universitario en Bioinformática

Computational Systems Biology

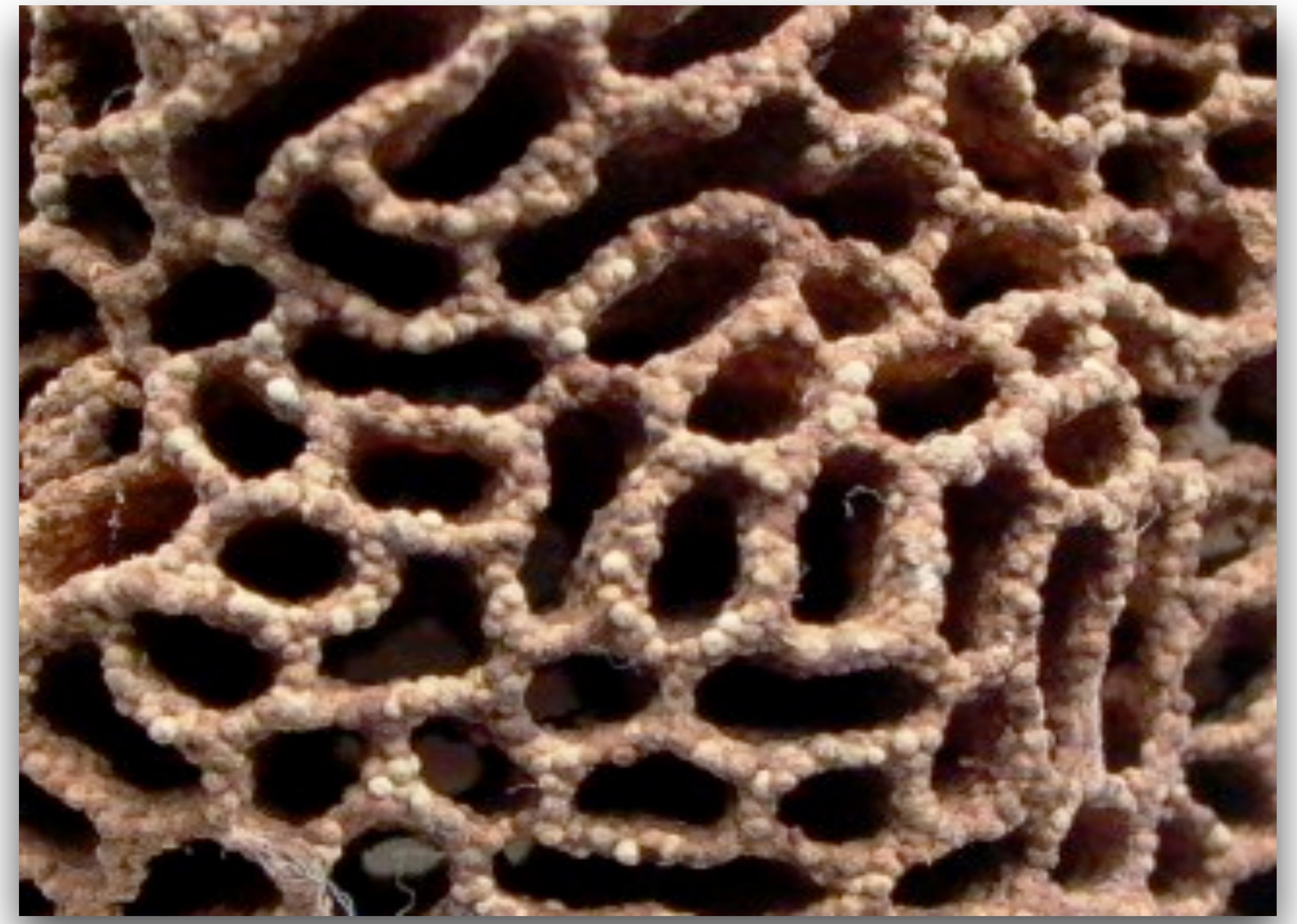
Introduction to Complex Networks

Sergi Valverde
Evolution of Networks Lab (ETL)
Institute of Evolutionary Biology (CSIC-UPF)
Consejo Superior de Investigaciones Científicas
@svalver

What is Complexity?

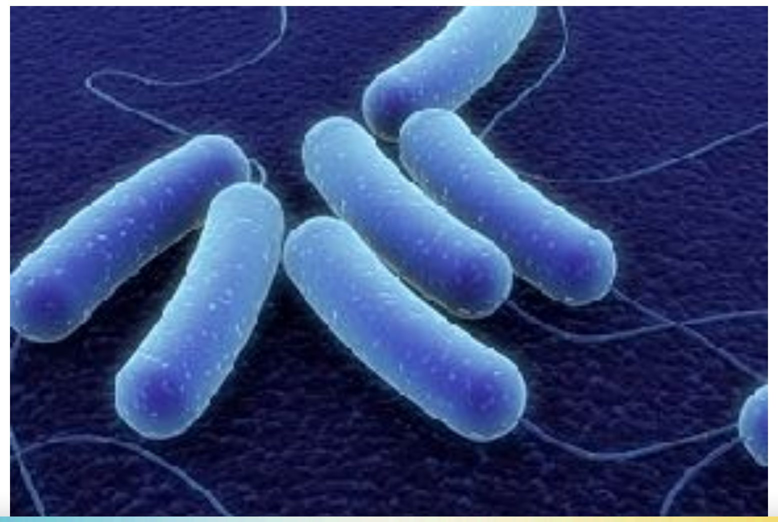
Complex systems involve emergence: the presence of higher-level phenomena that cannot be reduced to the analysis of lower-level entities.

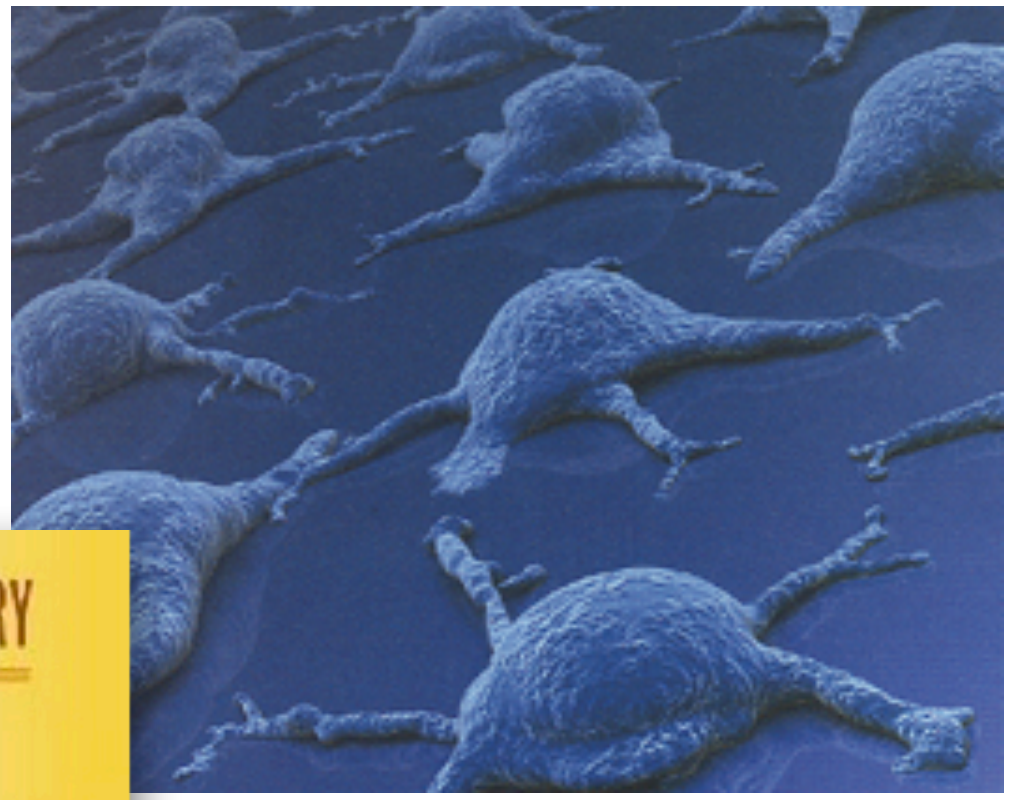
Complexity requires interactions among different units. New interactions are key to innovations.

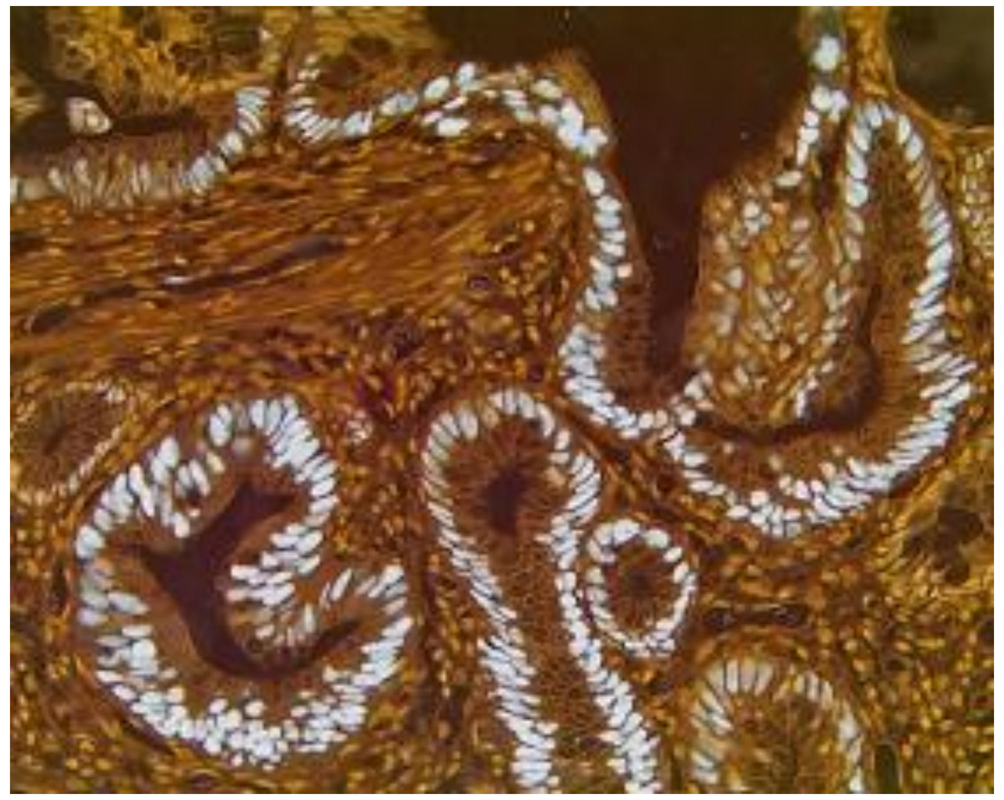


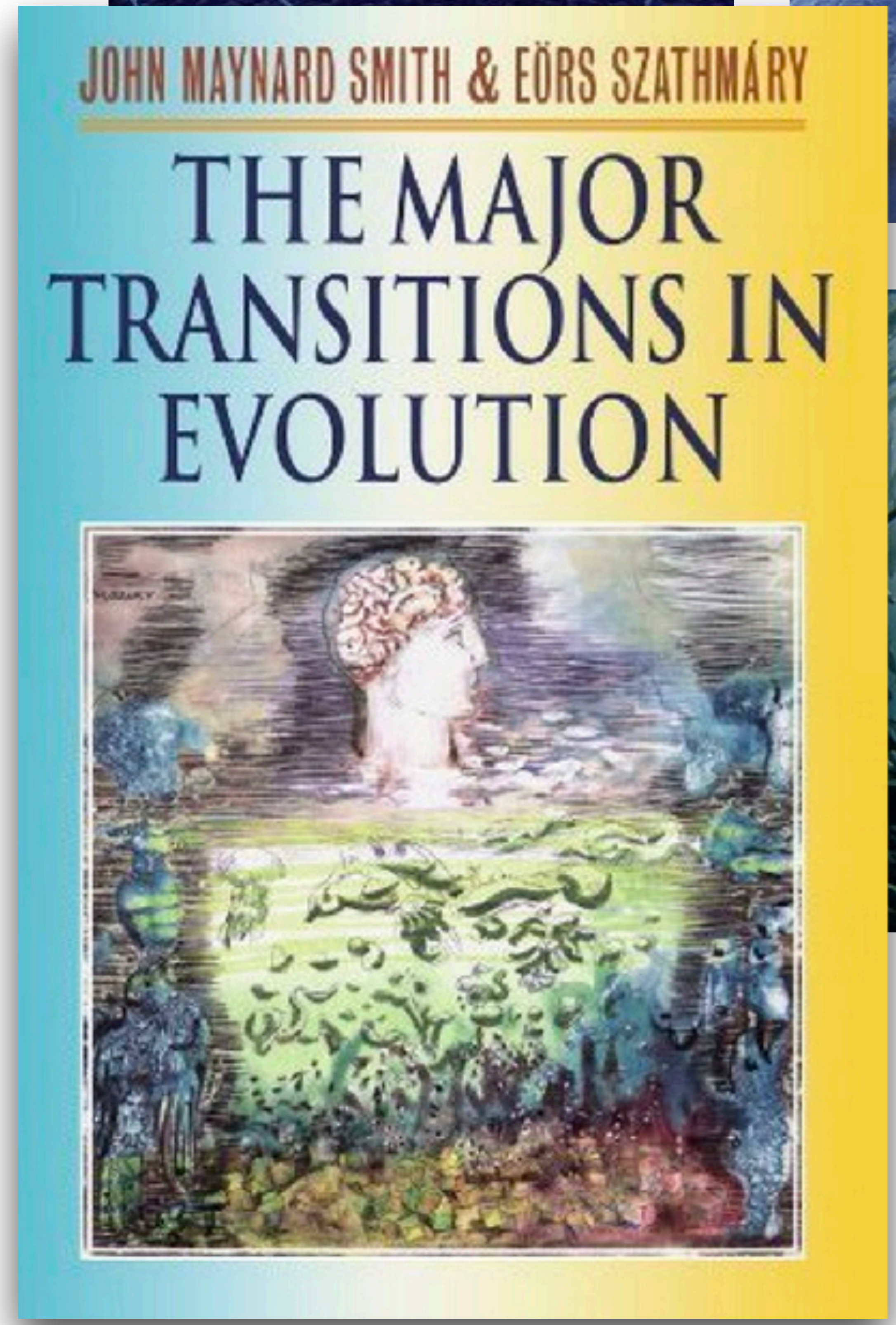
Evolution of Complexity

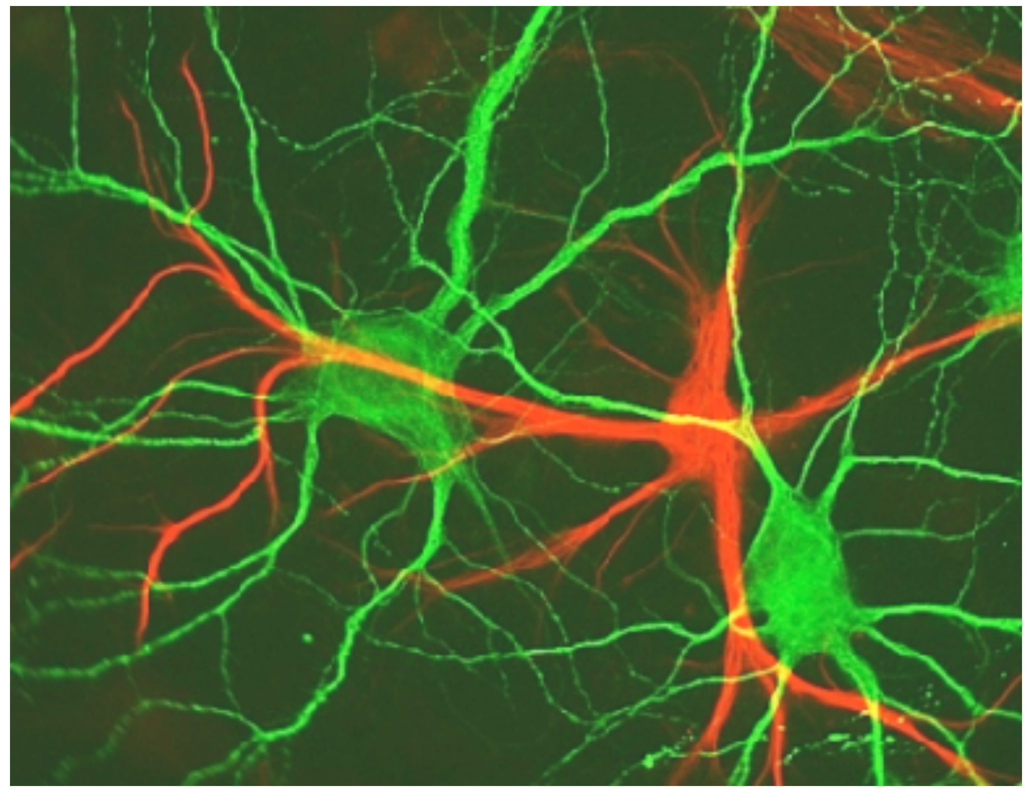
Adaptations and Innovations taking place at Multiple Scales



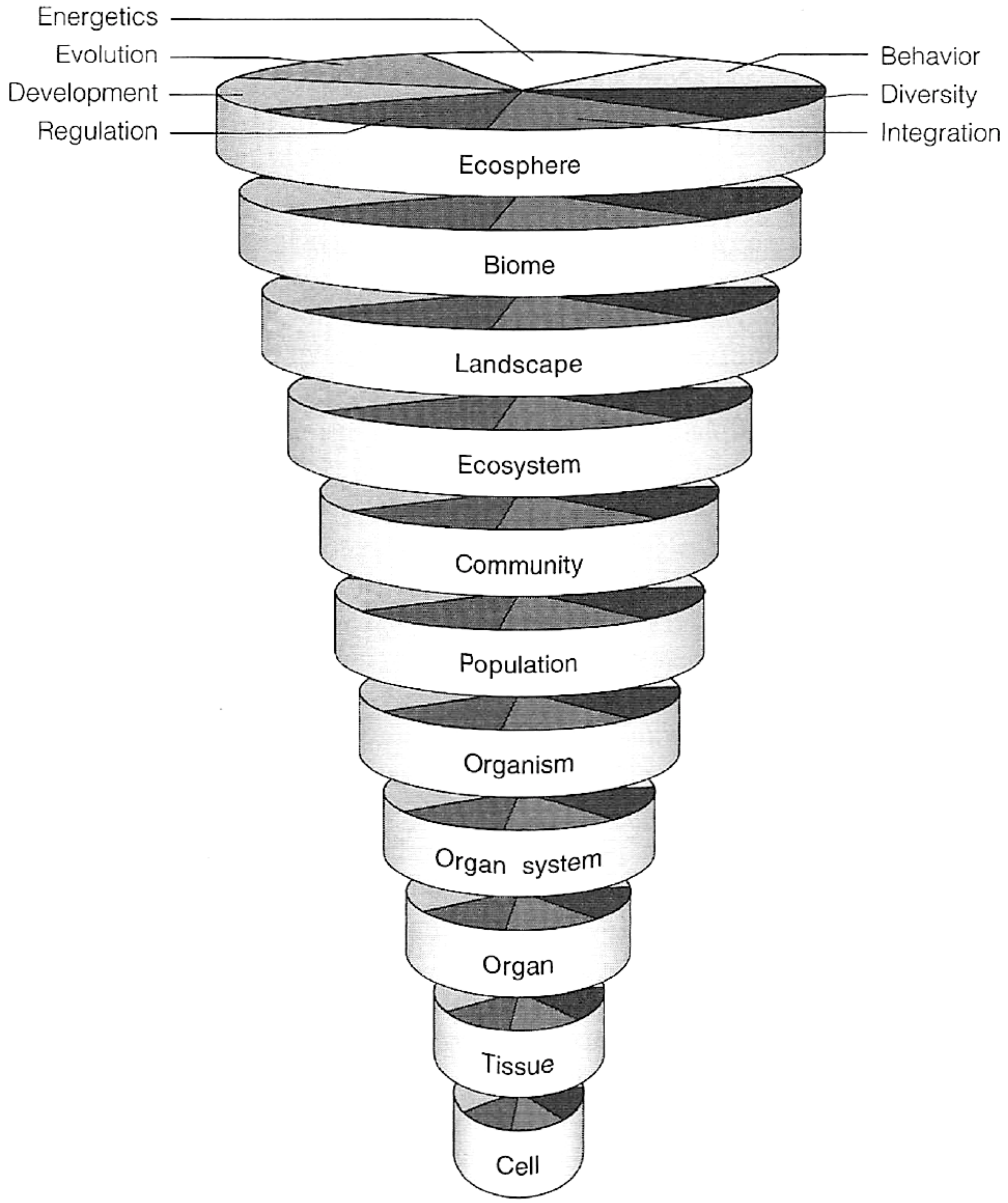






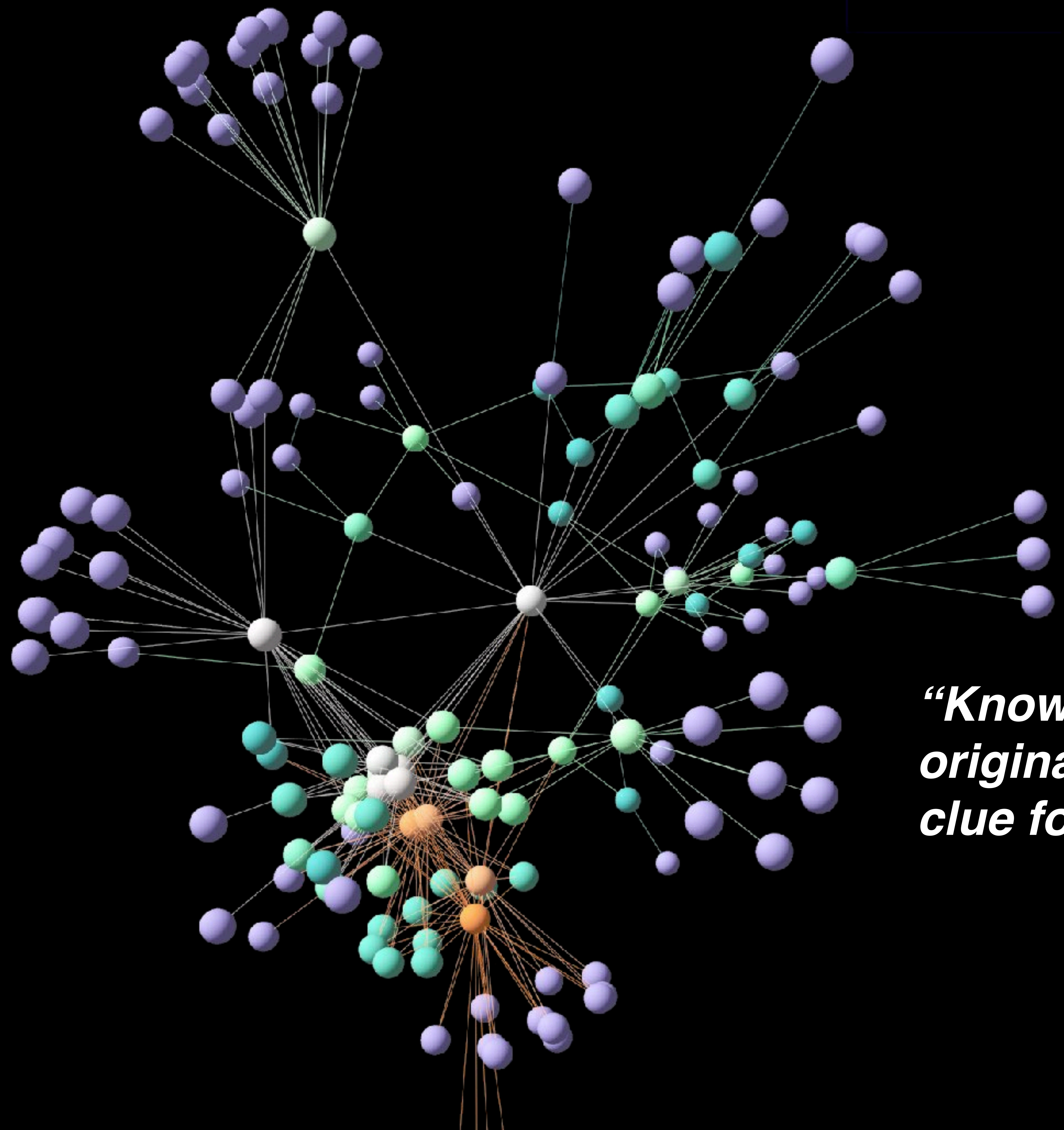


New qualitative behaviours, structures and patterns naturally emerge when crossing **phase transition points**



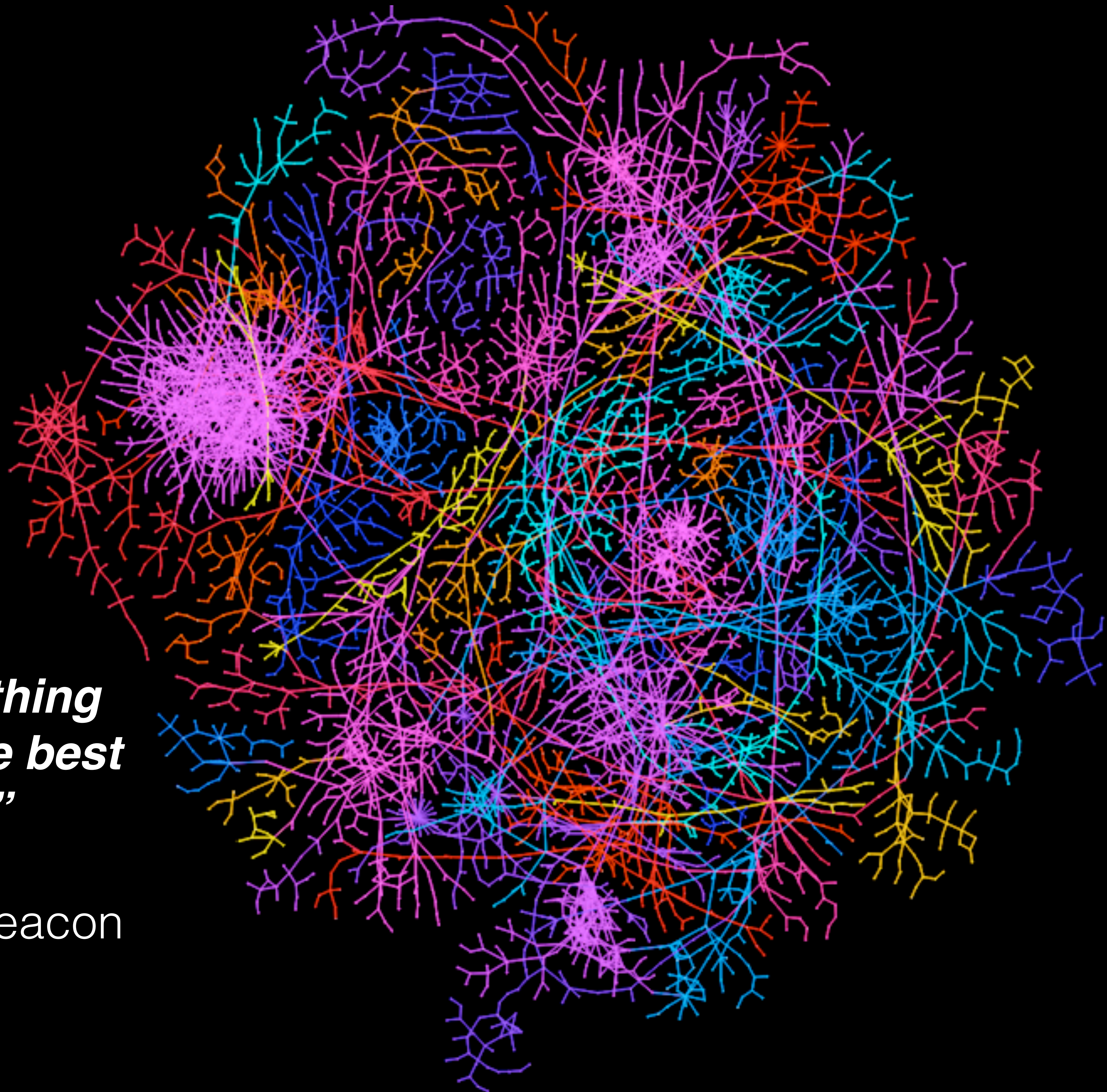
Universality

Do life and technology share the same basic architecture?



"Knowing how something originated often is the best clue for how it works"

- Terrence Deacon



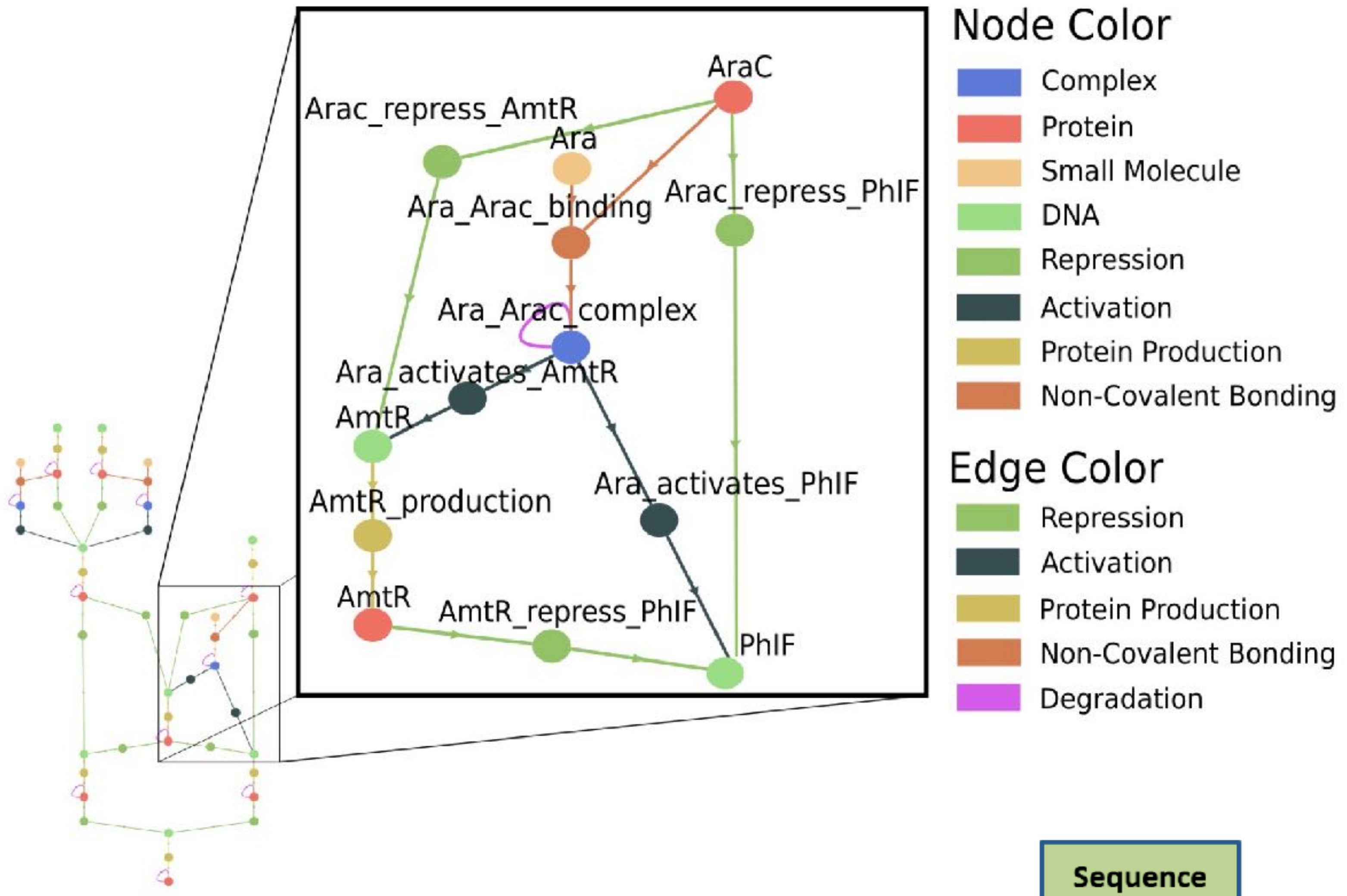
A Network Language for Biology

Can we find a good notation for biological systems?

Bertrand Russell

"A good notation has a subtletly and suggestiveness which at times make it seem almost like a live teacher ... and a perfect notation would be a substitute for thought"

quoted by Woodger (1937) The Axiomatic Method in Biology, pp. 18



Angel Goñi

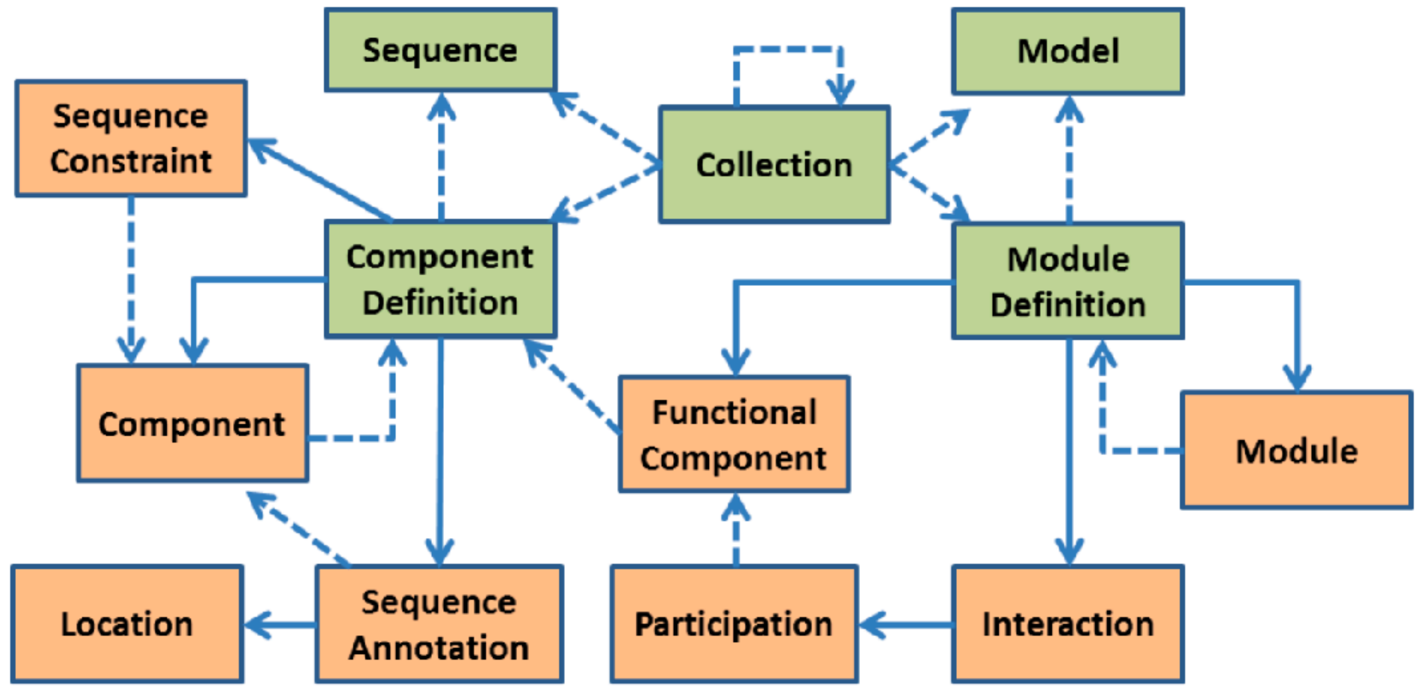
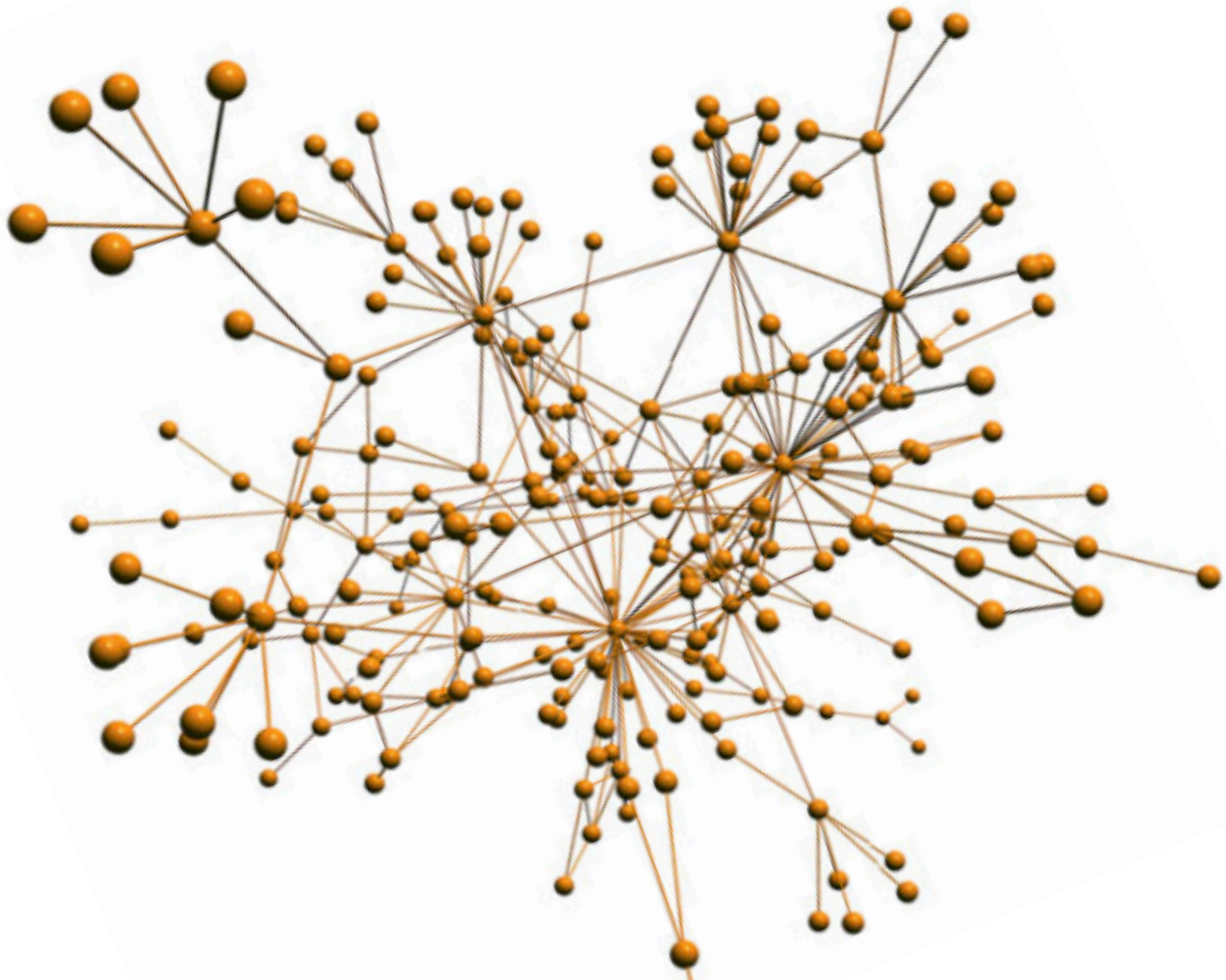


Figure 3: *Main classes of information represented by the SBOL 2.x standard, and their relationships. Green boxes are "top level" classes, while the other classes are in support of these classes. Solid arrows indicates ownership, whereas a dashed arrow indicates that one class refers to an object of another class.*

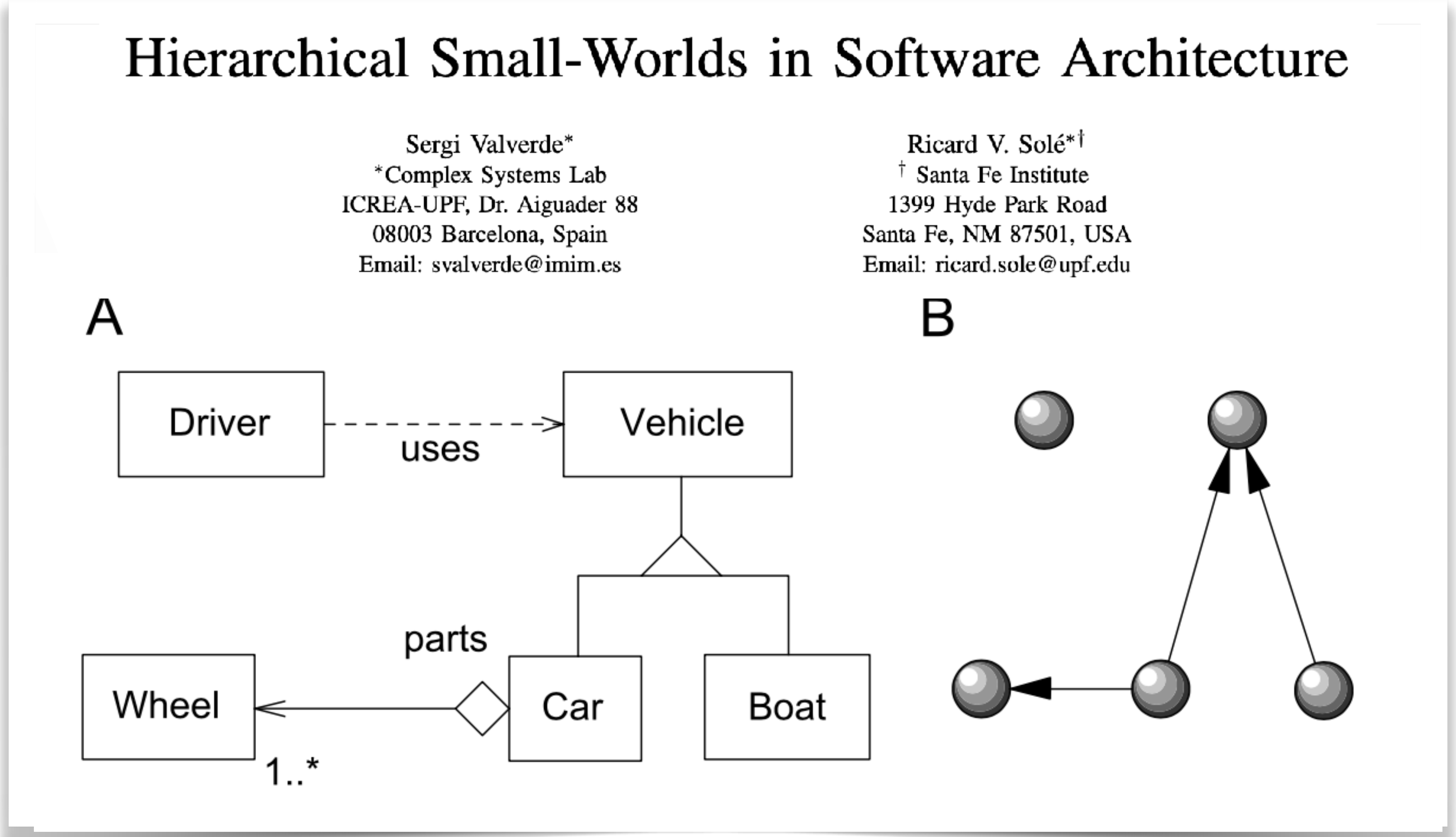
Madsen et al. (2019) *Synthetic Biology Open Language (SBOL) v 2.3*

A Network Language for Technology



Valverde et al. (2002) *Scale-Free Networks from Optimal Design*

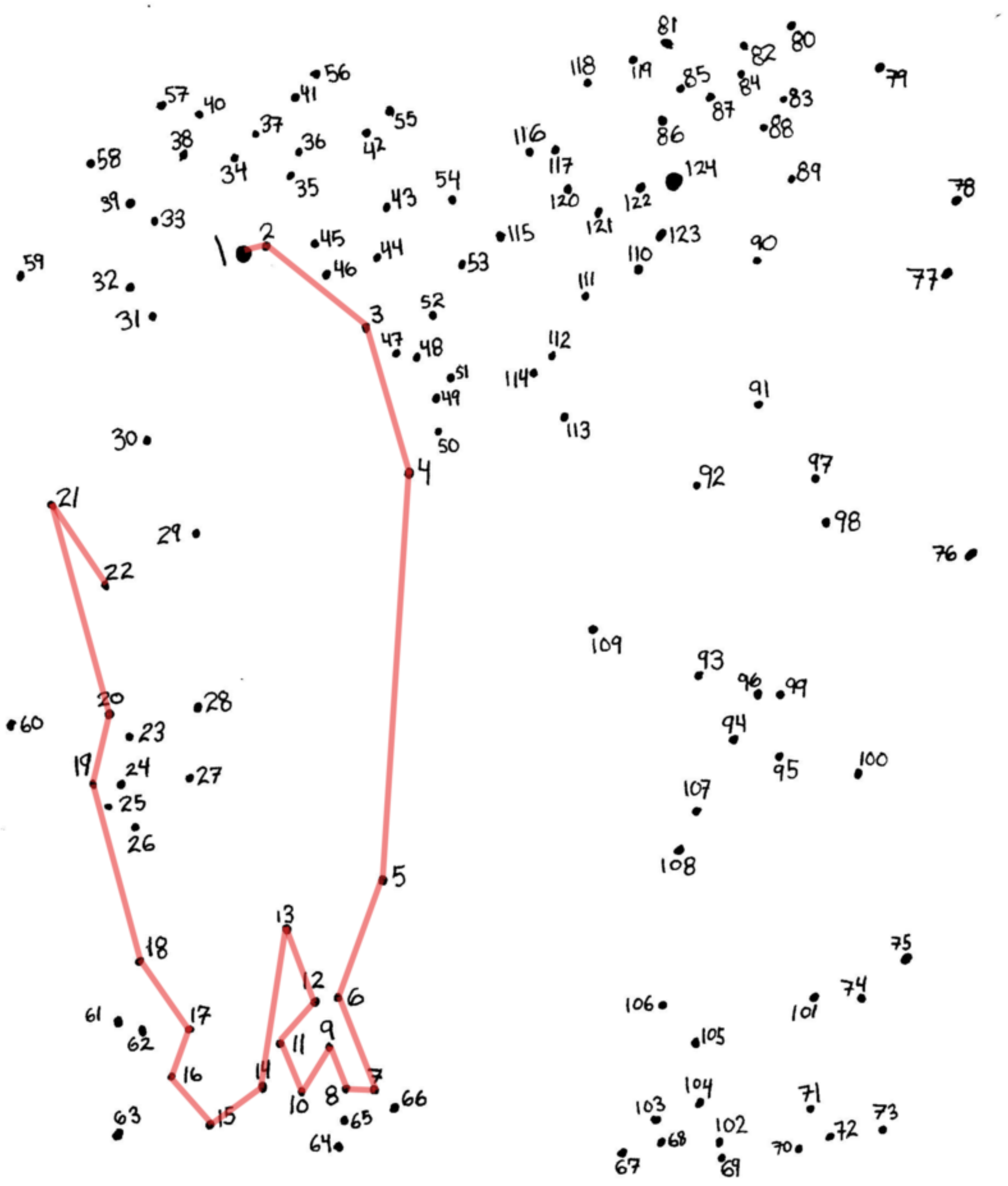
Alan Kay


Hierarchical Small-Worlds in Software Architecture

Sergi Valverde*
*Complex Systems Lab
ICREA-UPF, Dr. Aiguader 88
08003 Barcelona, Spain
Email: svalverde@imim.es

Ricard V. Solé*†
† Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501, USA
Email: ricard.sole@upf.edu

Index



Basic Properties

Robustness and Fragility

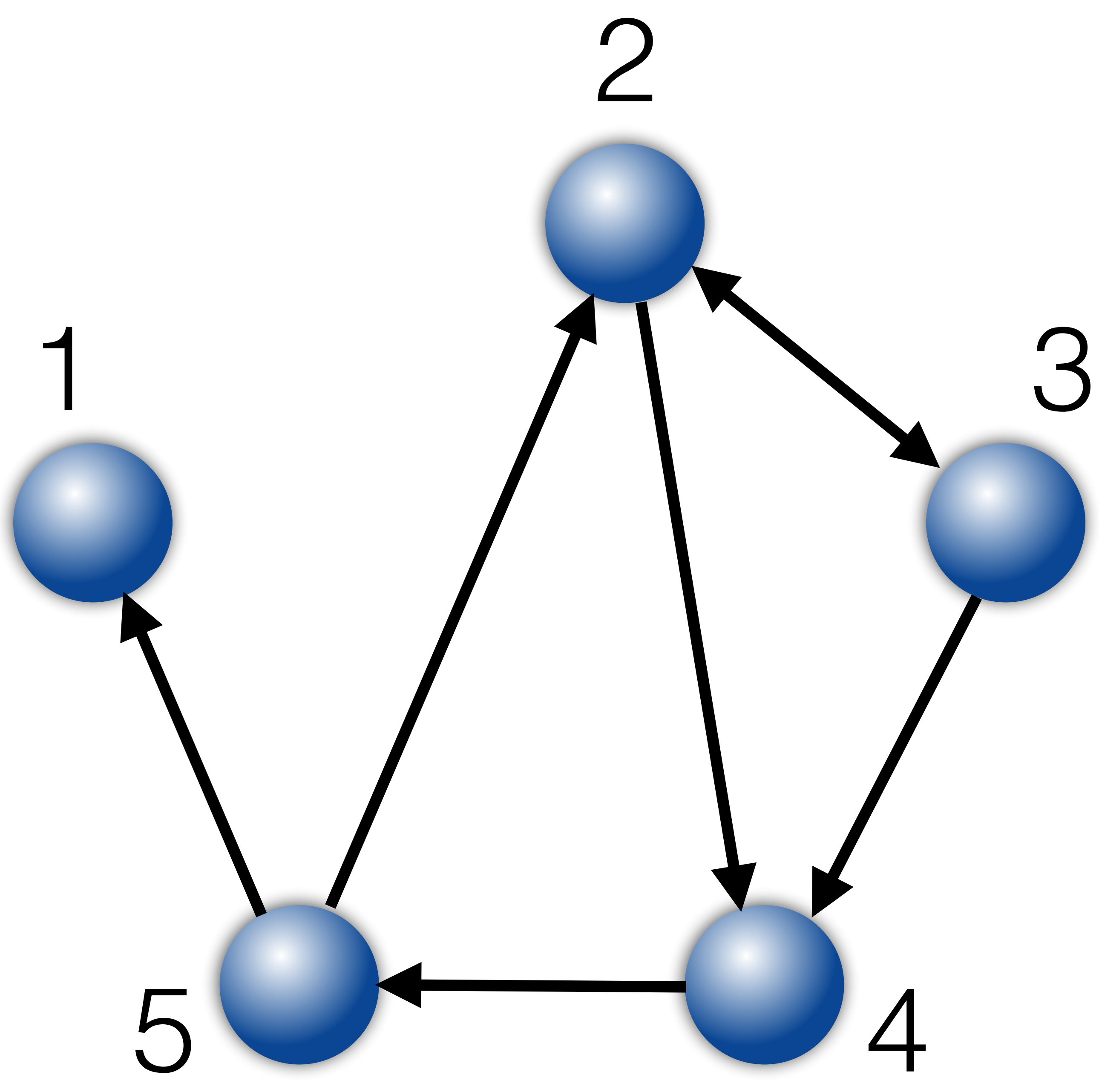
Hubs, Connectors and Paths

Evolution of Networks

Community Structure

Network Representation

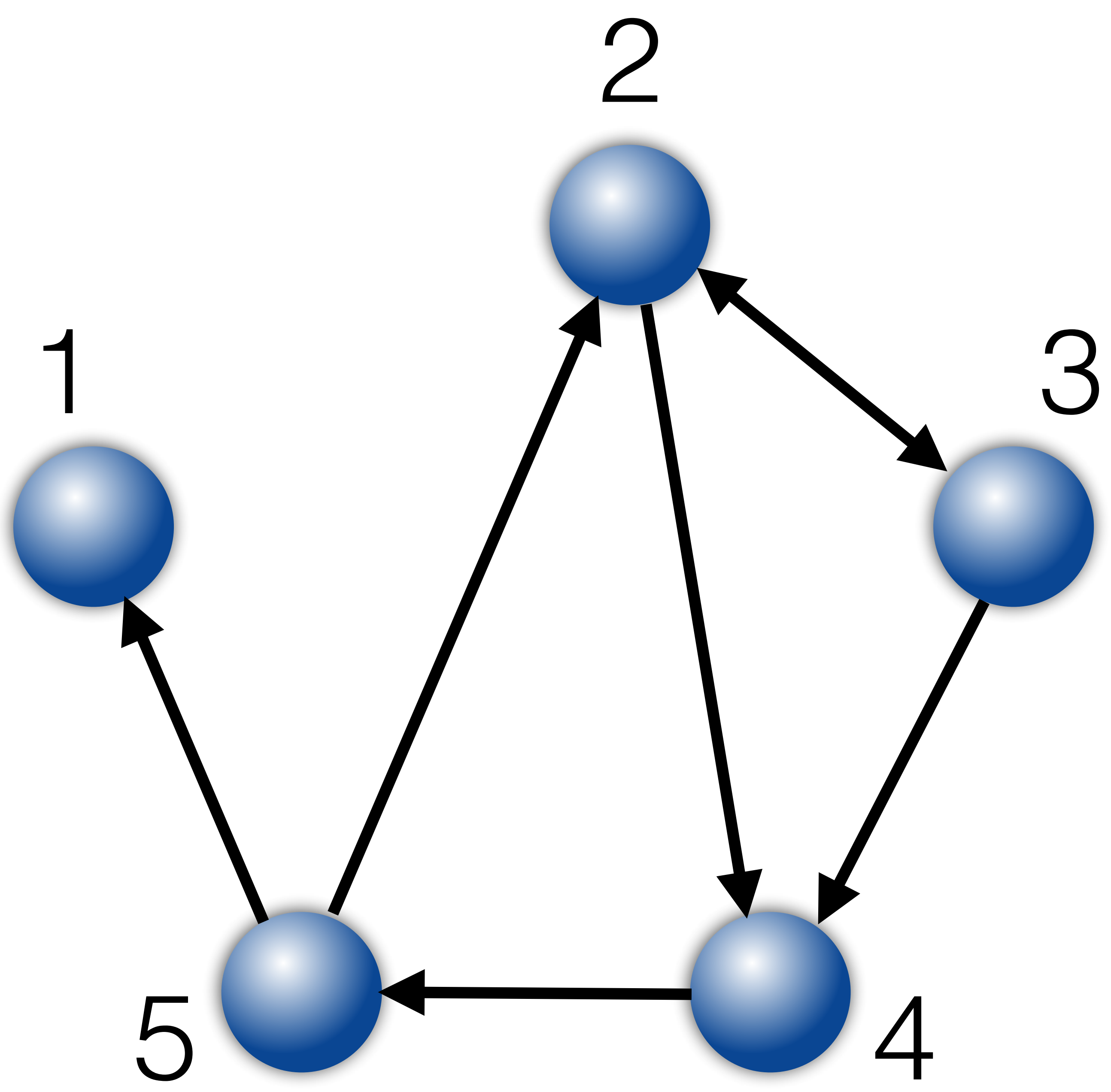
Adjacency Matrix



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Network Representation

Edge List



2	3
3	2
2	4
3	4
4	5
5	2
5	1

https://svalver.github.io/course

Introduction to Networks

42589 - Biologia de Sistemas Computacional

VNIVERSITAT DE VALÈNCIA Máster Universitario en Bioinformática

This website contains a collection of online activities that are part of the curriculum for the Universitat de Valencia course "Biologia de Sistemas Computacional". These lessons can be used in combination Netlab, an online application designed to assist students to develop evolutionary models of complex networks.

Sergi Valverde, a CSIC tenured scientist from the Institute of Evolutionary Biology (CSIC-UPF), teaches the course.

Online activities

The following online activities require a WebGL compliant web browser.

- **Defining a network (link)**:Input a simple network by hand and adjust its layout parameters.
- **A Random Graph (link)**: When determining the relevance of network patterns, random graphs are utilized as null models. The Erdös-Renyi model generates random graphs with a fixed connection probability (p) and a

NetworkX Network Analysis in Python

graph-tool

Pajek

Cytoscape

Gephi

Methods in Ecology and Evolution

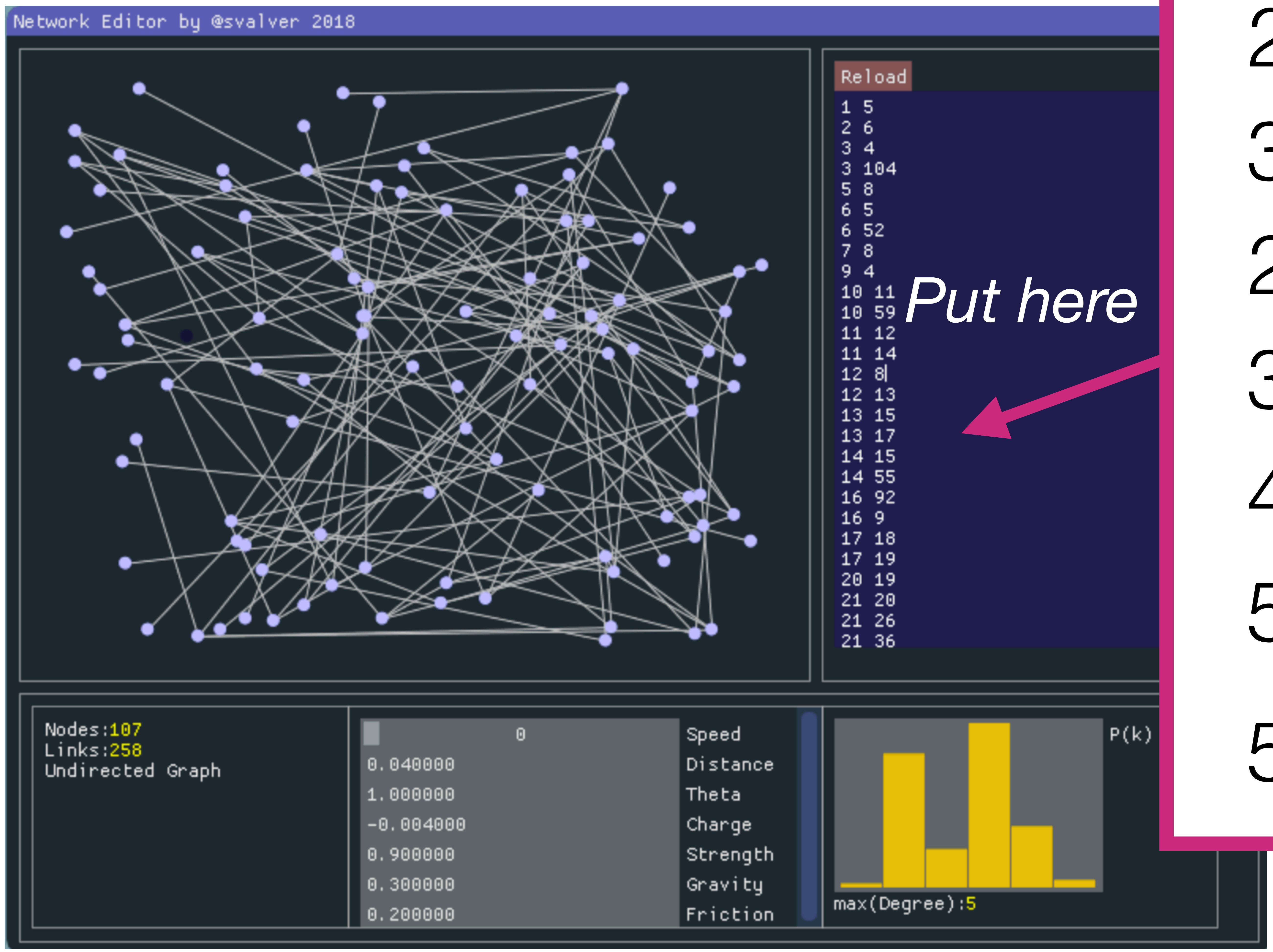
Methods in Ecology and Evolution 2016, 7, 127–132 doi: 10.1111/2041-210X.12458

APPLICATION

BiMat: a MATLAB package to facilitate the analysis of bipartite networks

Activity: Defining Networks

https://tinyurl.com/24e3n5tf

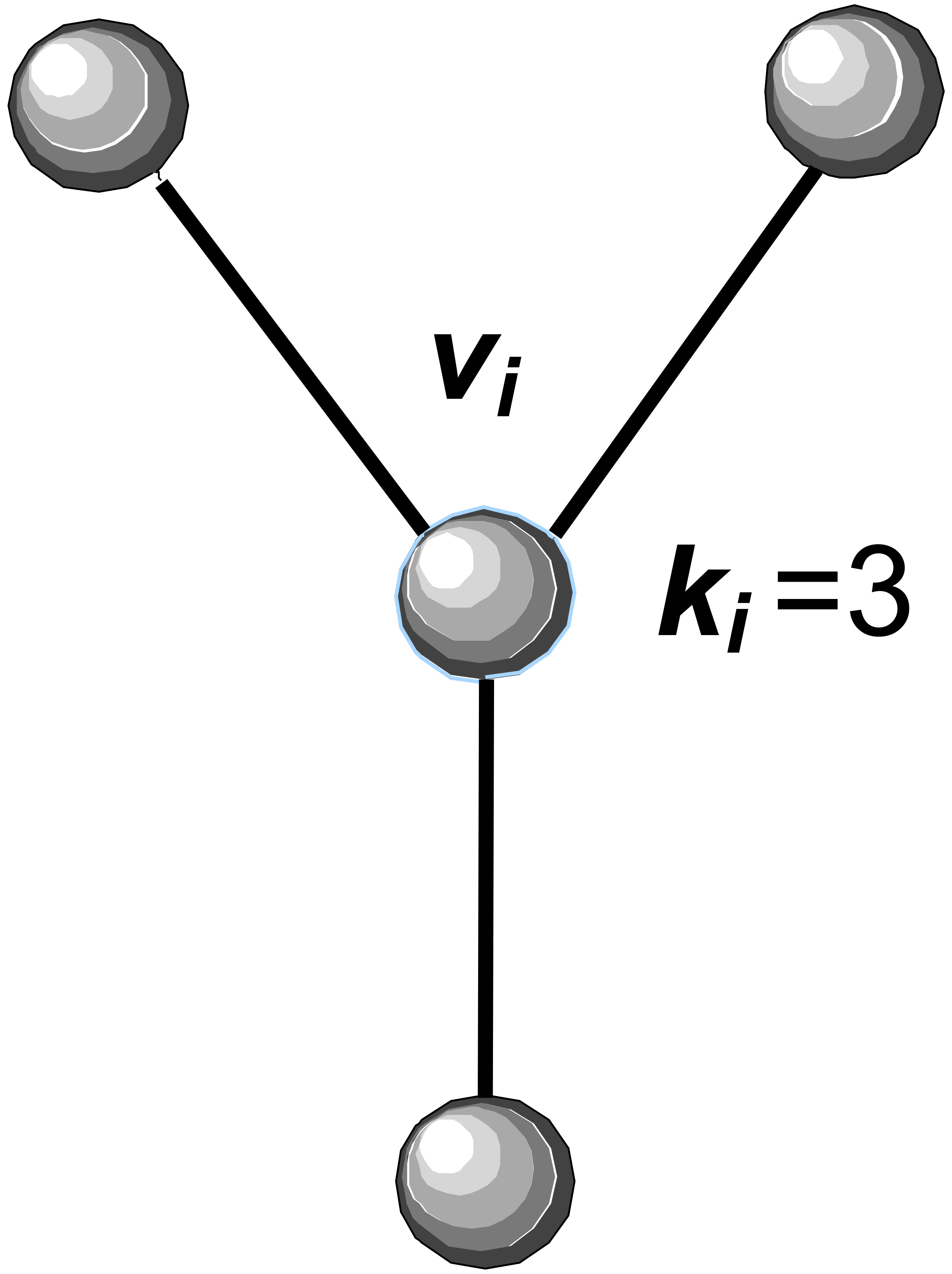


2 3
3 2
2 4
3 4
4 5
5 2
5 1

1. Explain how many bytes are needed to store this network using the adjacency list and the matrix representations.

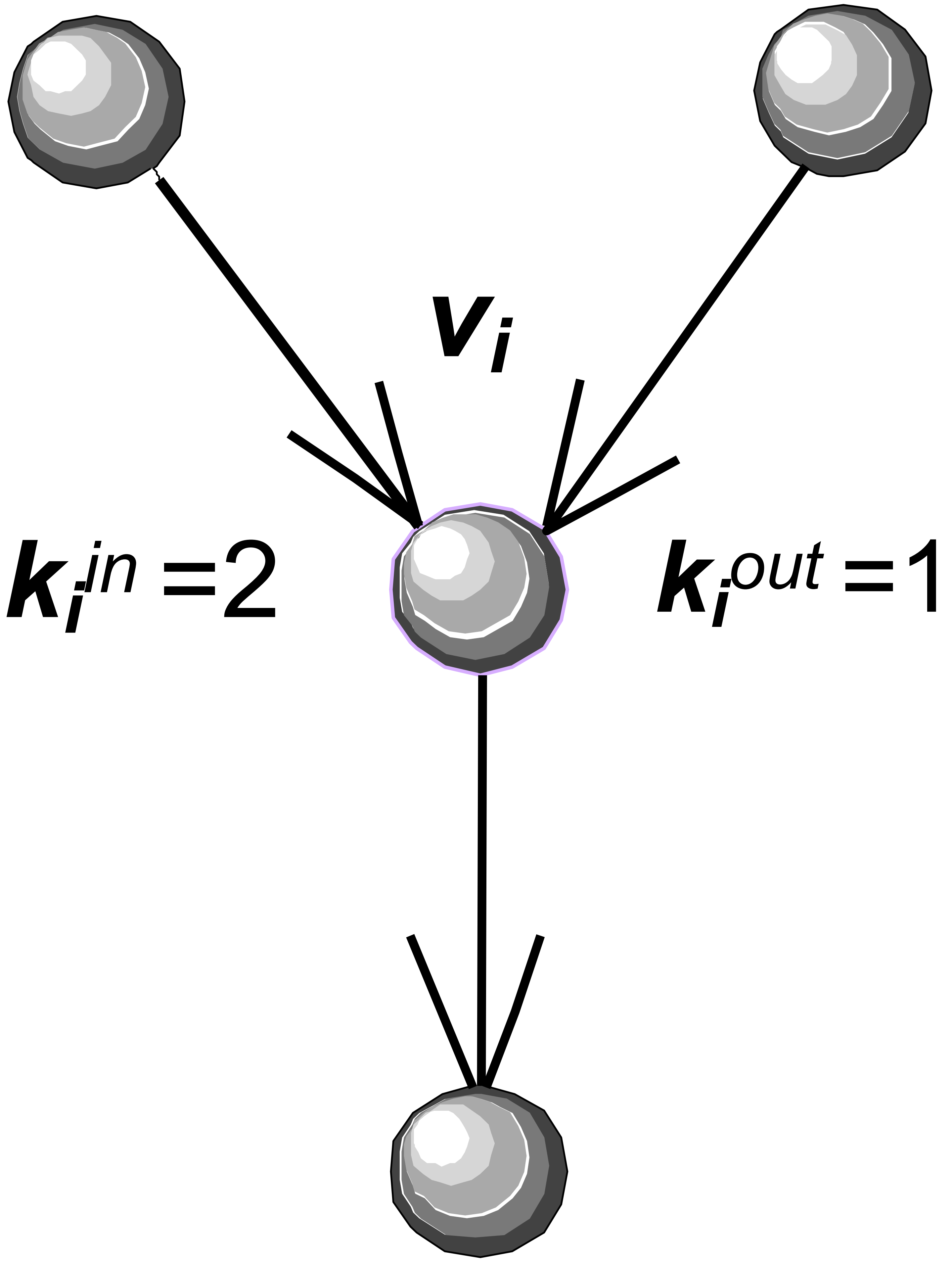
2. Consider an alternative method for representing networks. Explain.

Degree



In-degree and Out-degree

Dominance hierarchies



$$k_i^{in} = \sum_{j=1}^{N} A_{j,i}$$

$$k_i^{out} = \sum_{j=1}^{N} A_{i,j}$$

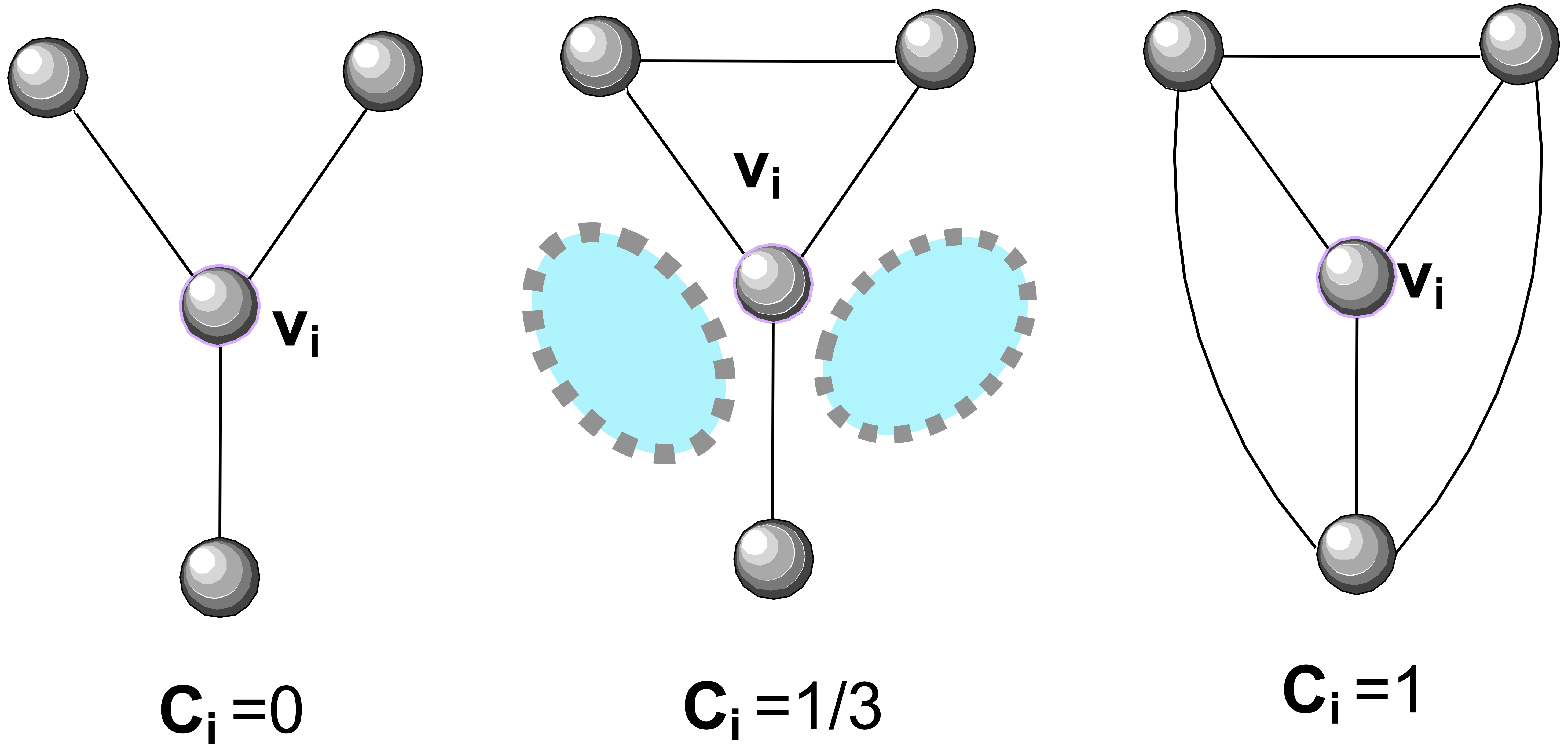
Number of Edges

$$m = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out} = \sum_{i,j} A_{i,j}$$

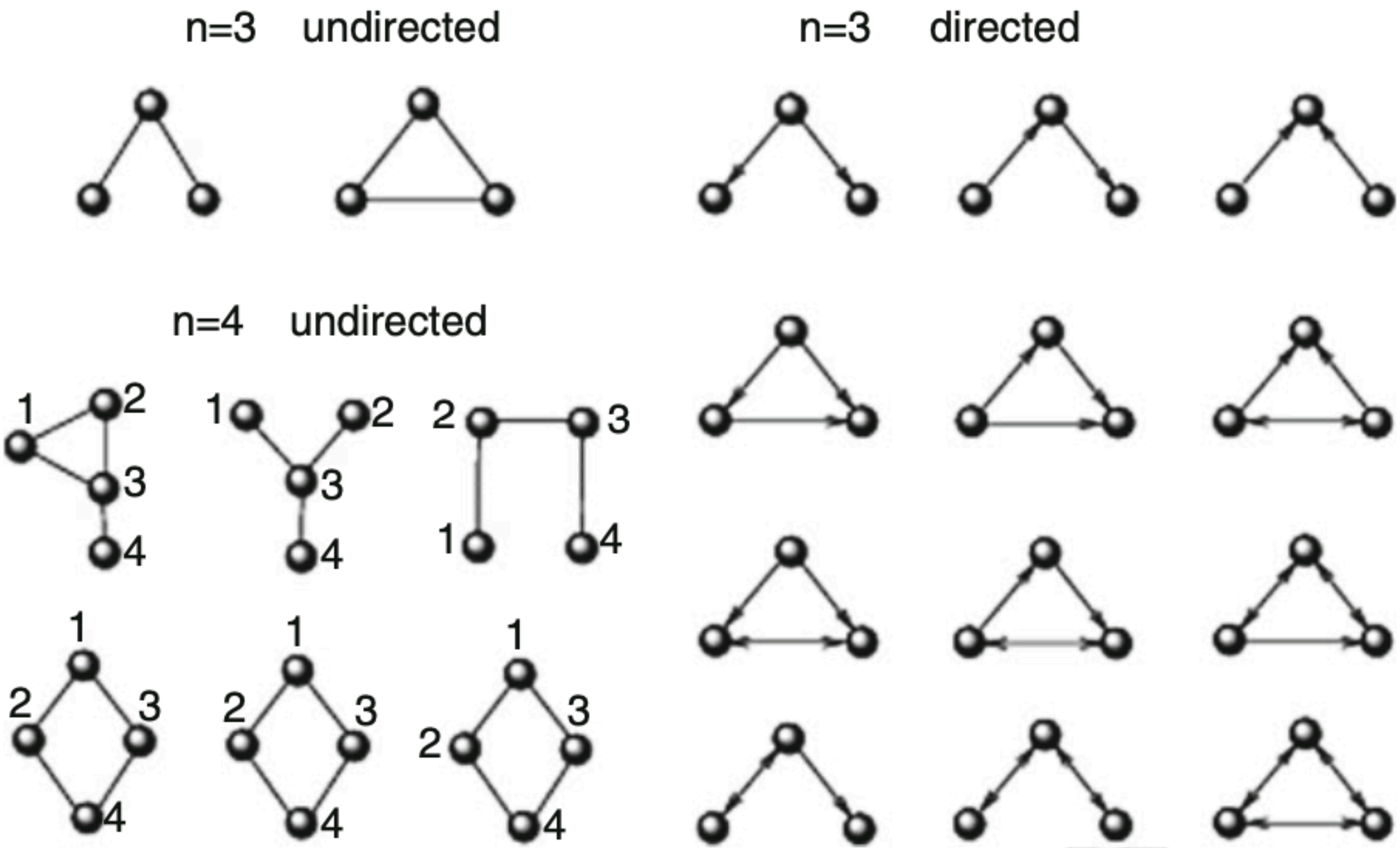
Local Clustering

$$c_i = \frac{e_i}{\binom{k_i}{2}}$$

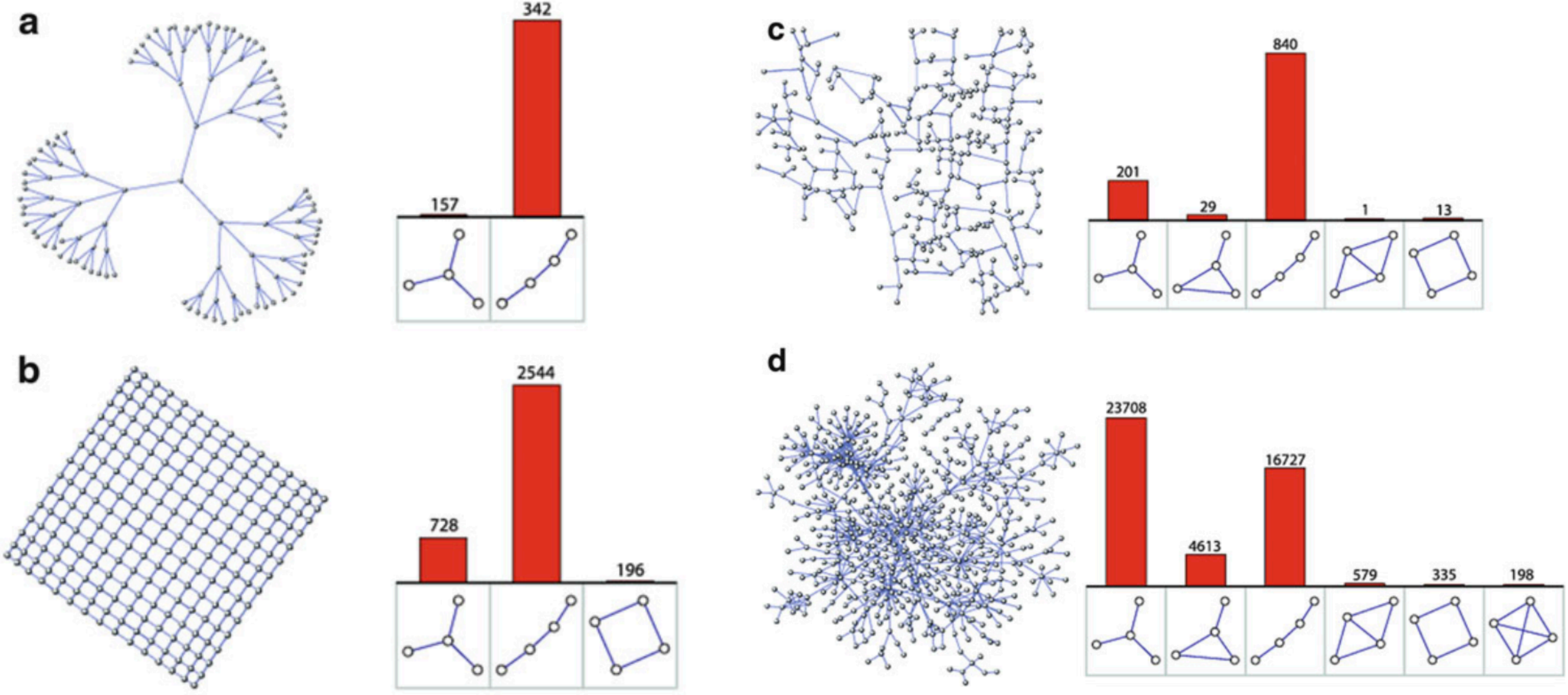
$$= \frac{2e_i}{k_i(k_i - 1)}$$



Motifs



Motifs

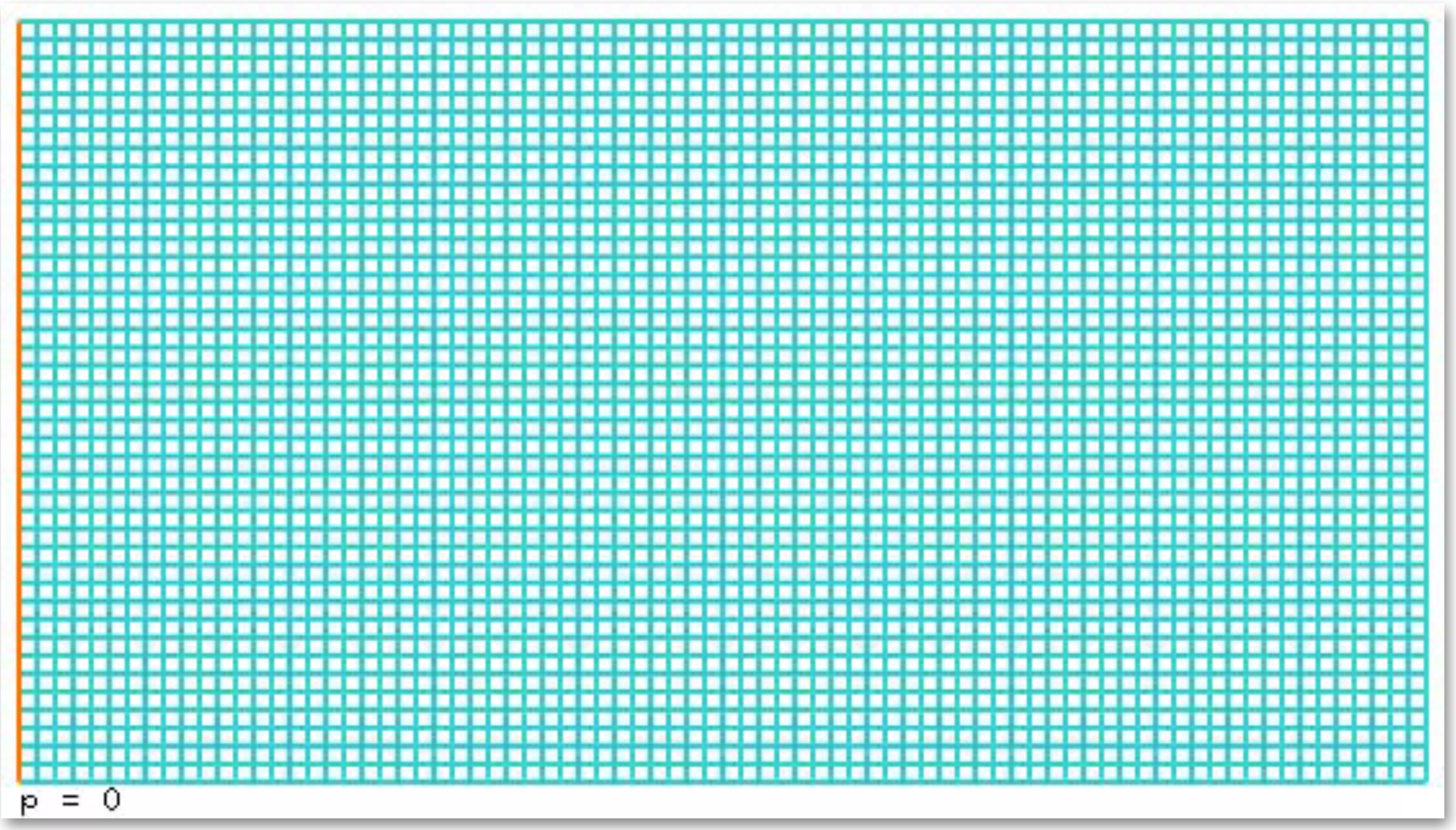


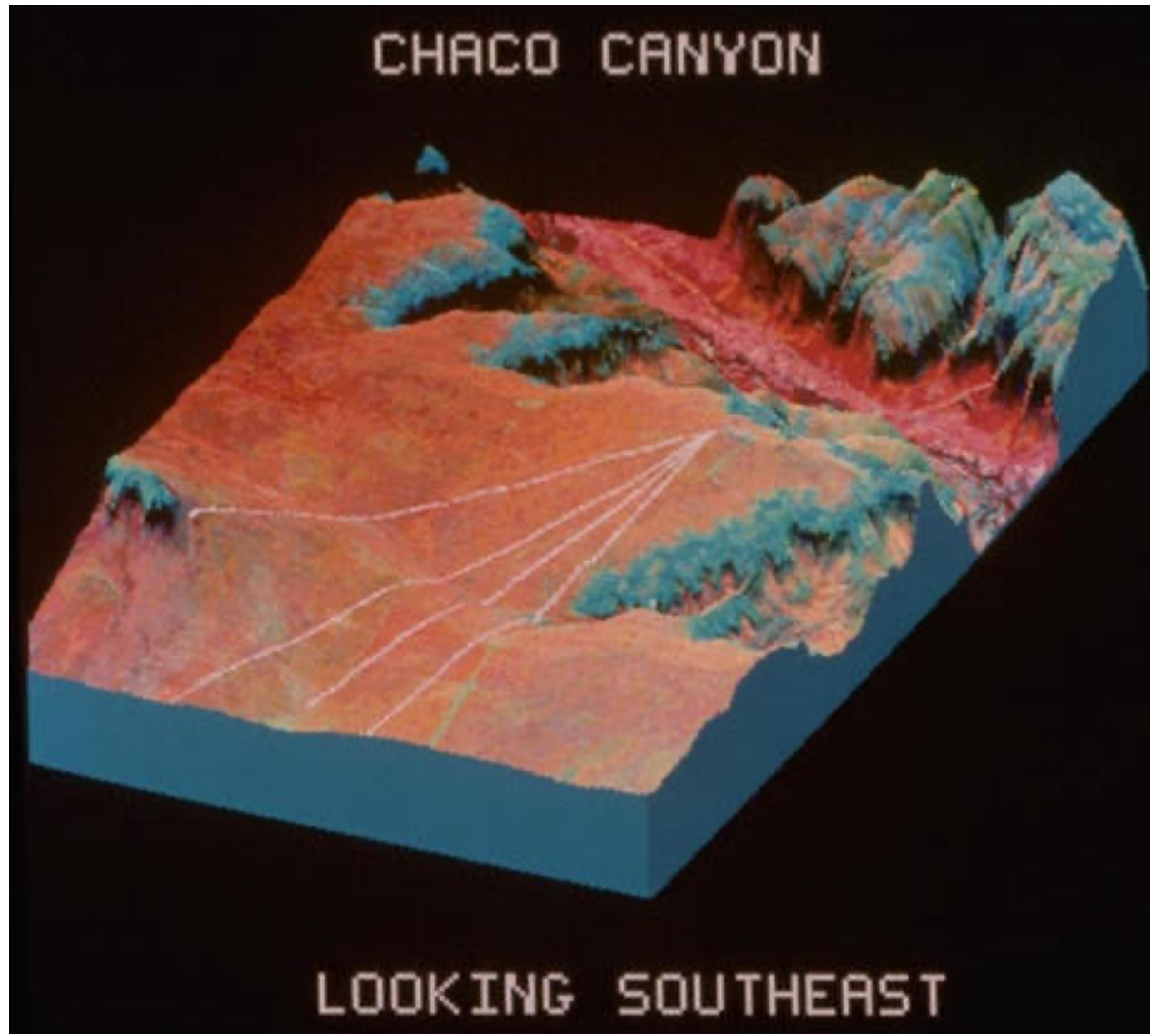
a
157
342
b
728
2544
196
c
201
29
840
1
13
d
23708
4613
16727
579
335
198

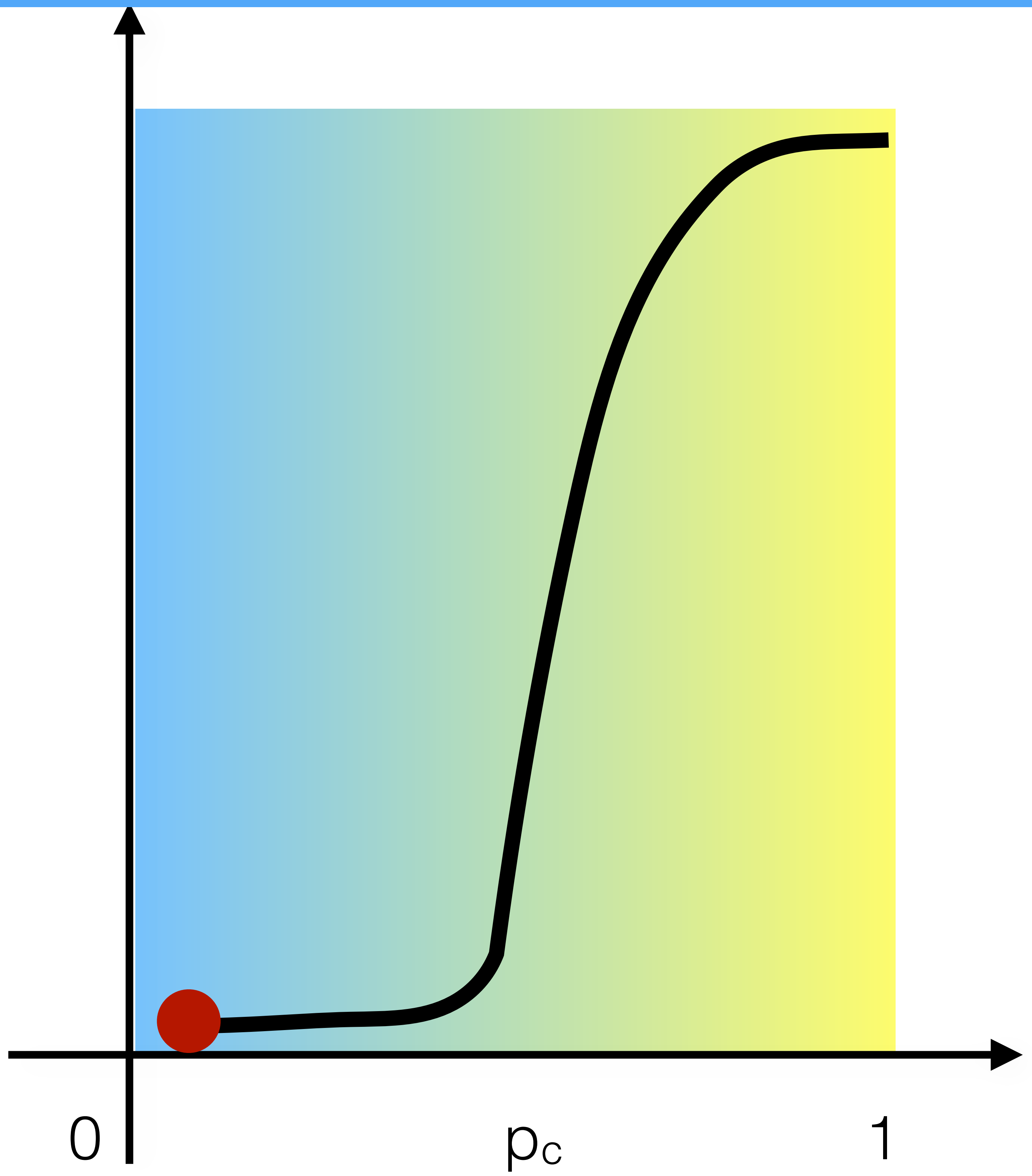
Random Networks :

Robustness & Fragility

Percolation



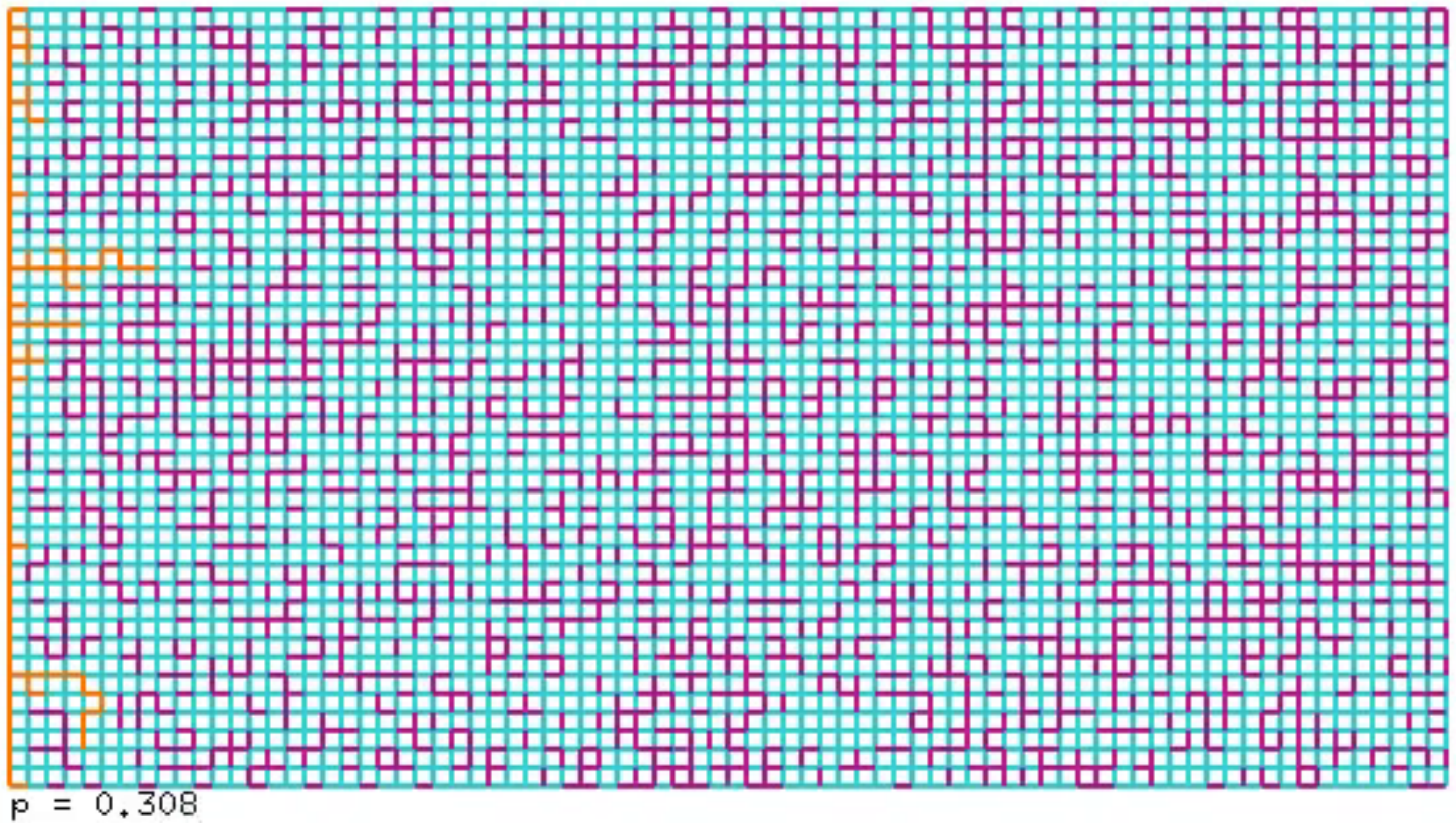


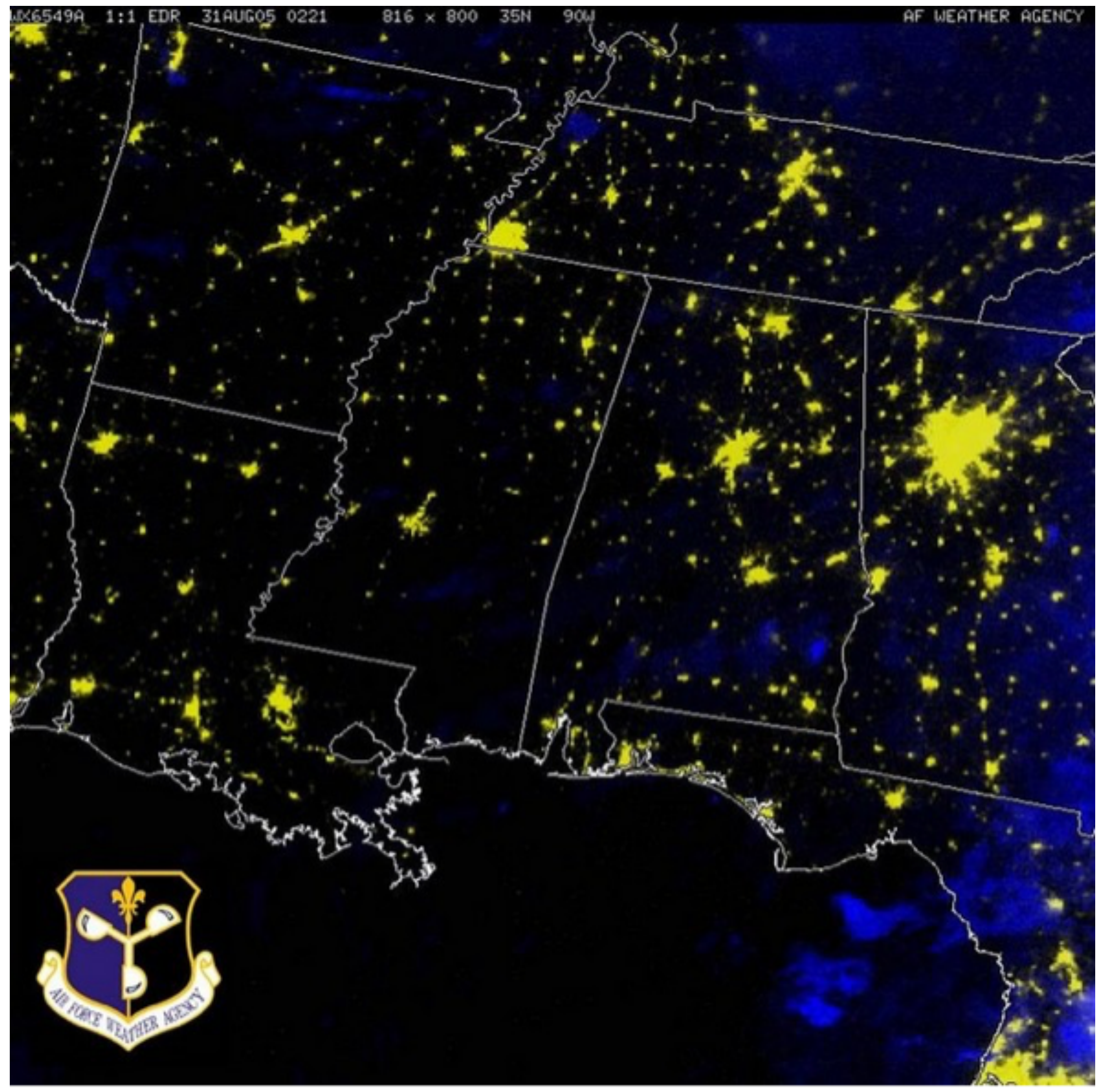


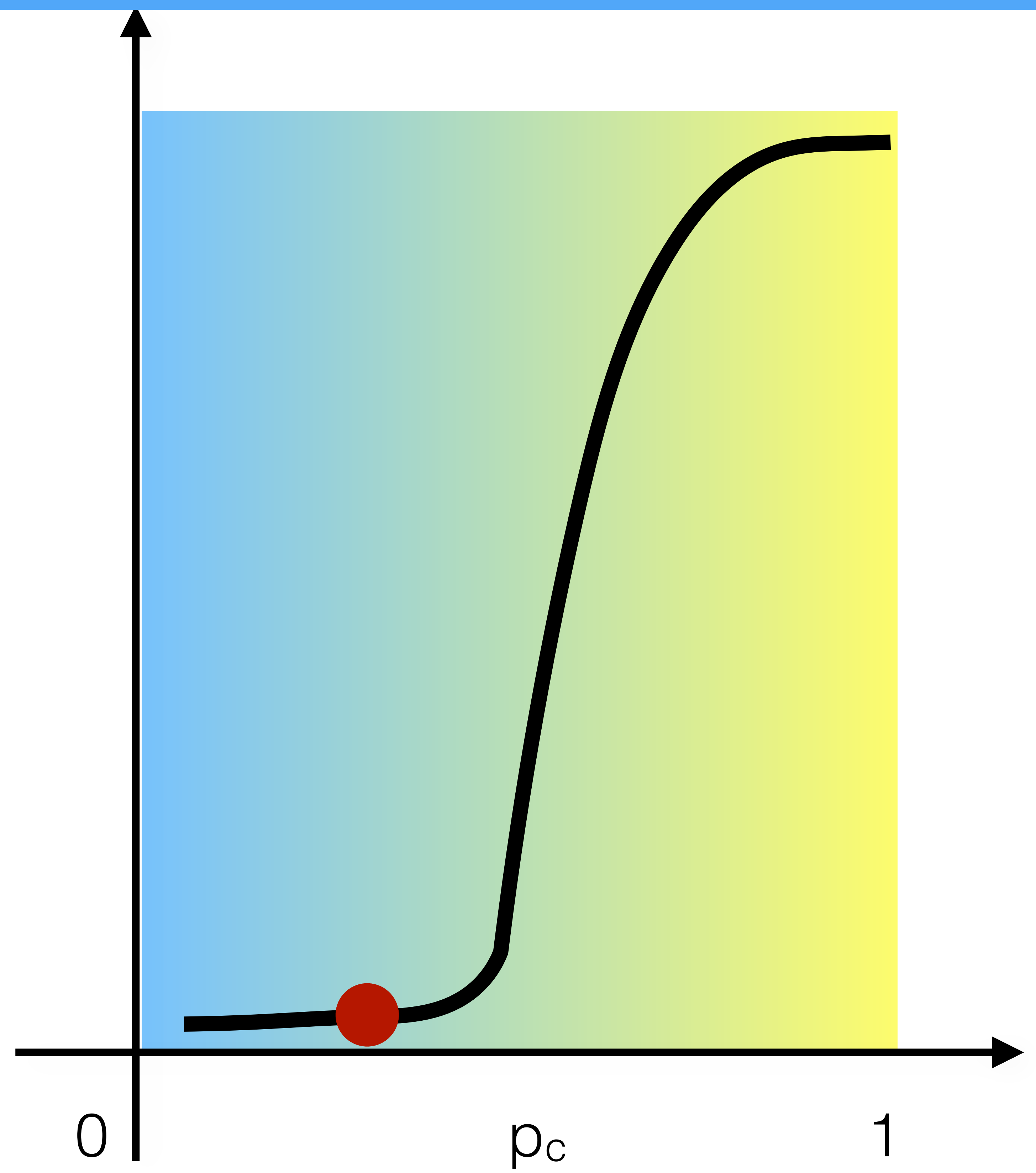
How does connectivity affects behaviour?

Kesten, Harry (1982), Percolation theory for mathematicians, Birkhauser

Disconnected Phase

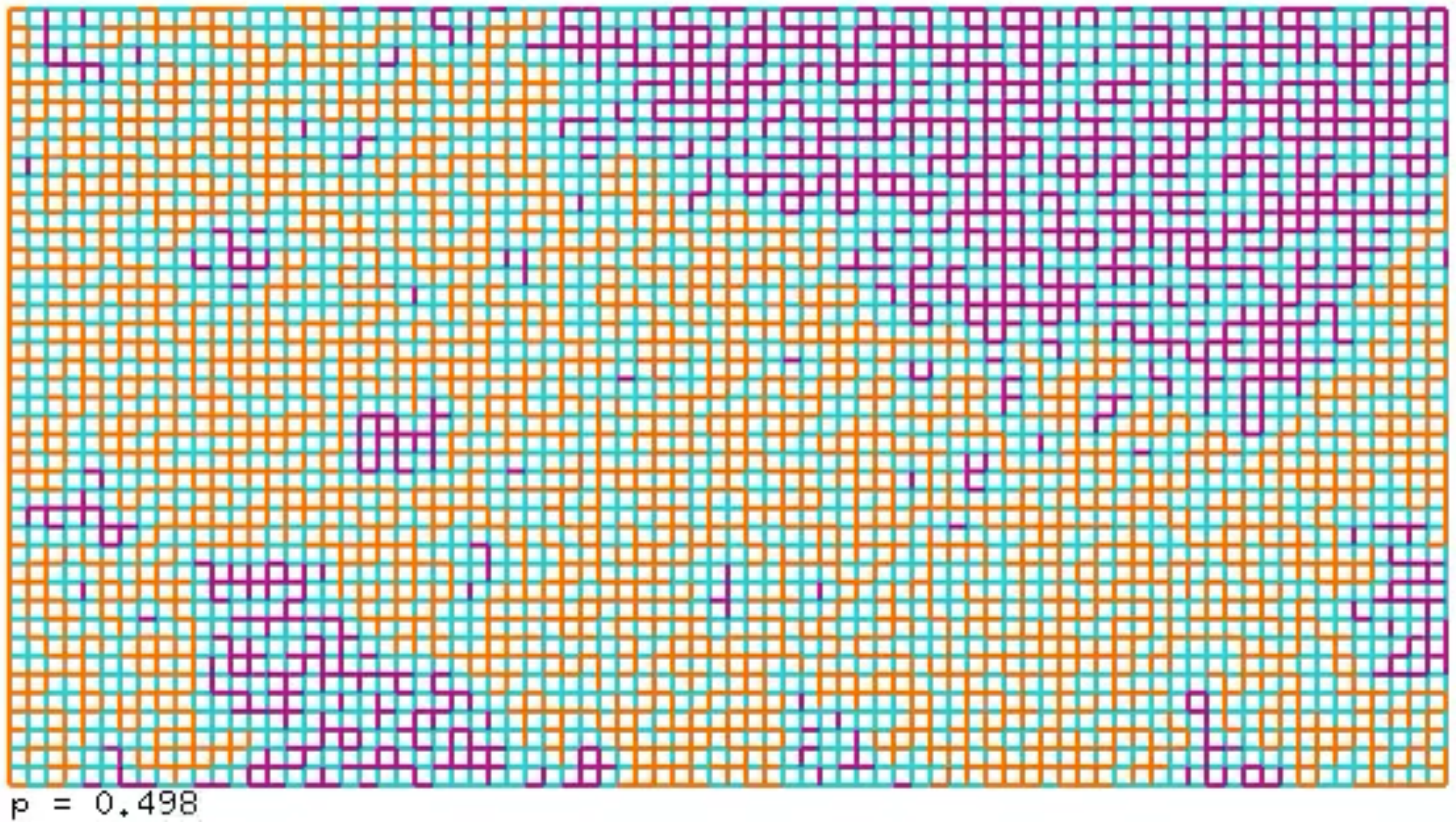


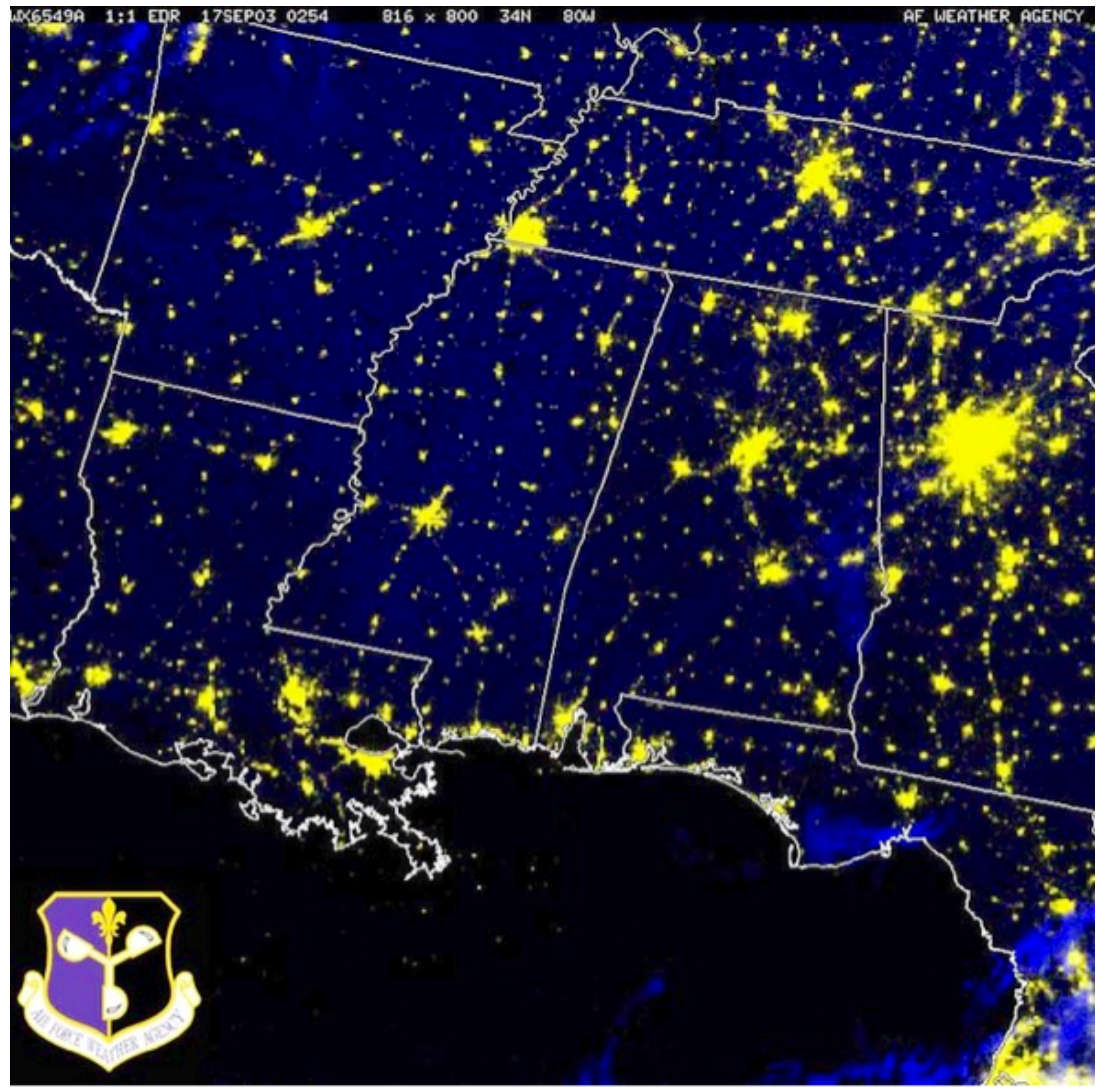


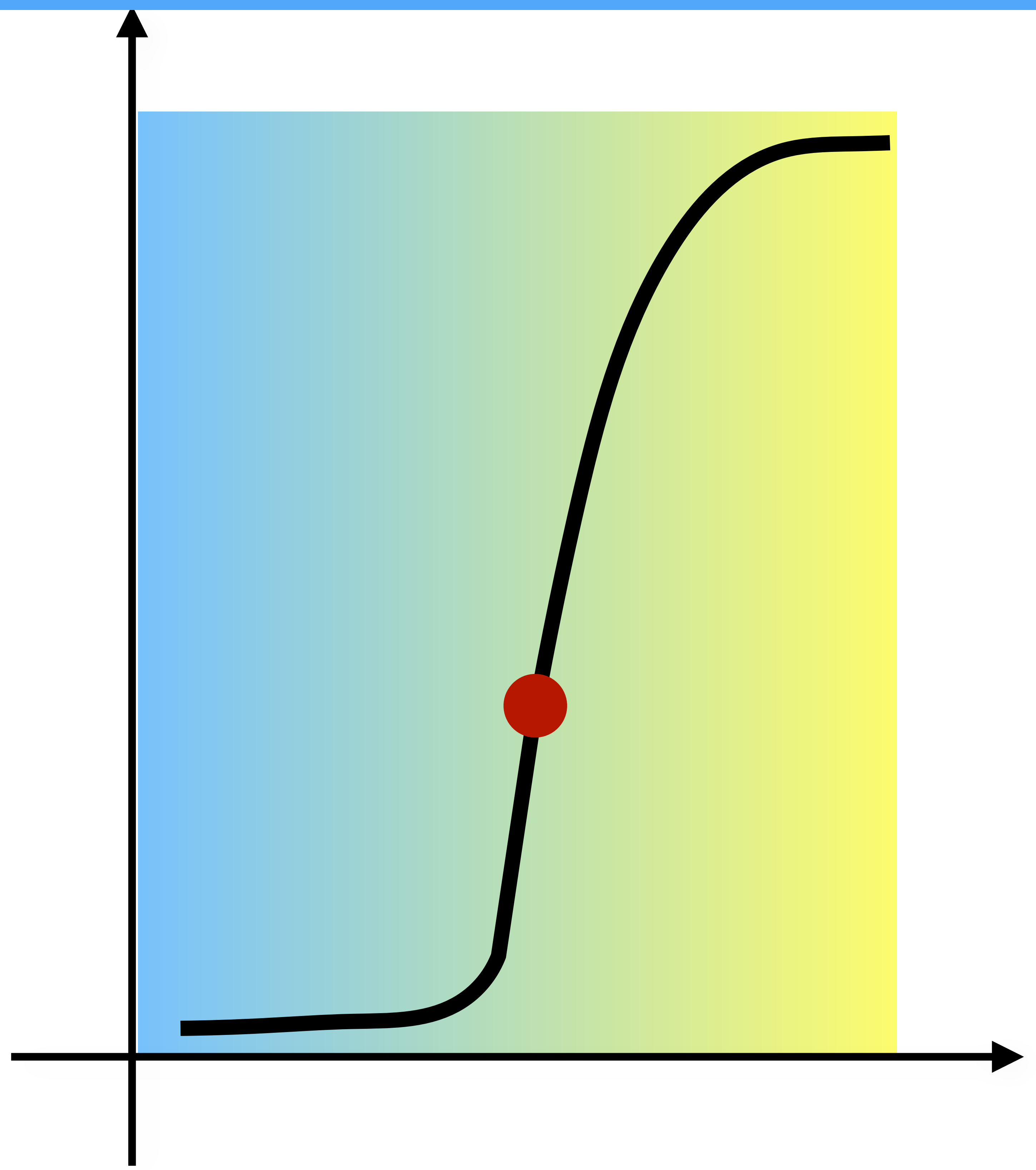


Power outage after Hurricane Katrina hit the Gulf Coast
This image was take Aug 30 and shows the widespread power outages across the Gulf Coast after Hurricane Katrina ravaged the area. U.S. Air Force Image.

Connected Phase





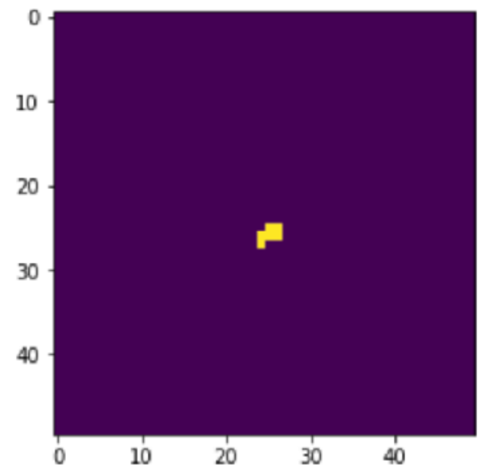


Power grid before the Hurricane Katrina hit the Gulf Coast
This image was taken Sept. 17,2003 and shows the city lights in the Gulf Coast clearly visible. U.S. Air Force Image.

Theorem (Kesten, 1980)

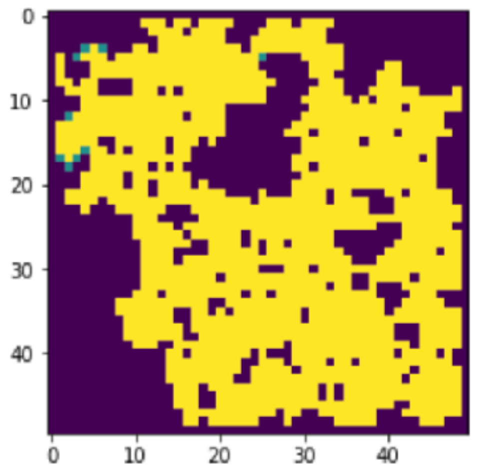
In Bernoulli percolation with parameter *p* on the infinite square grid,

if *p* <= 1/2, the
P(infinite cluster) = 0,



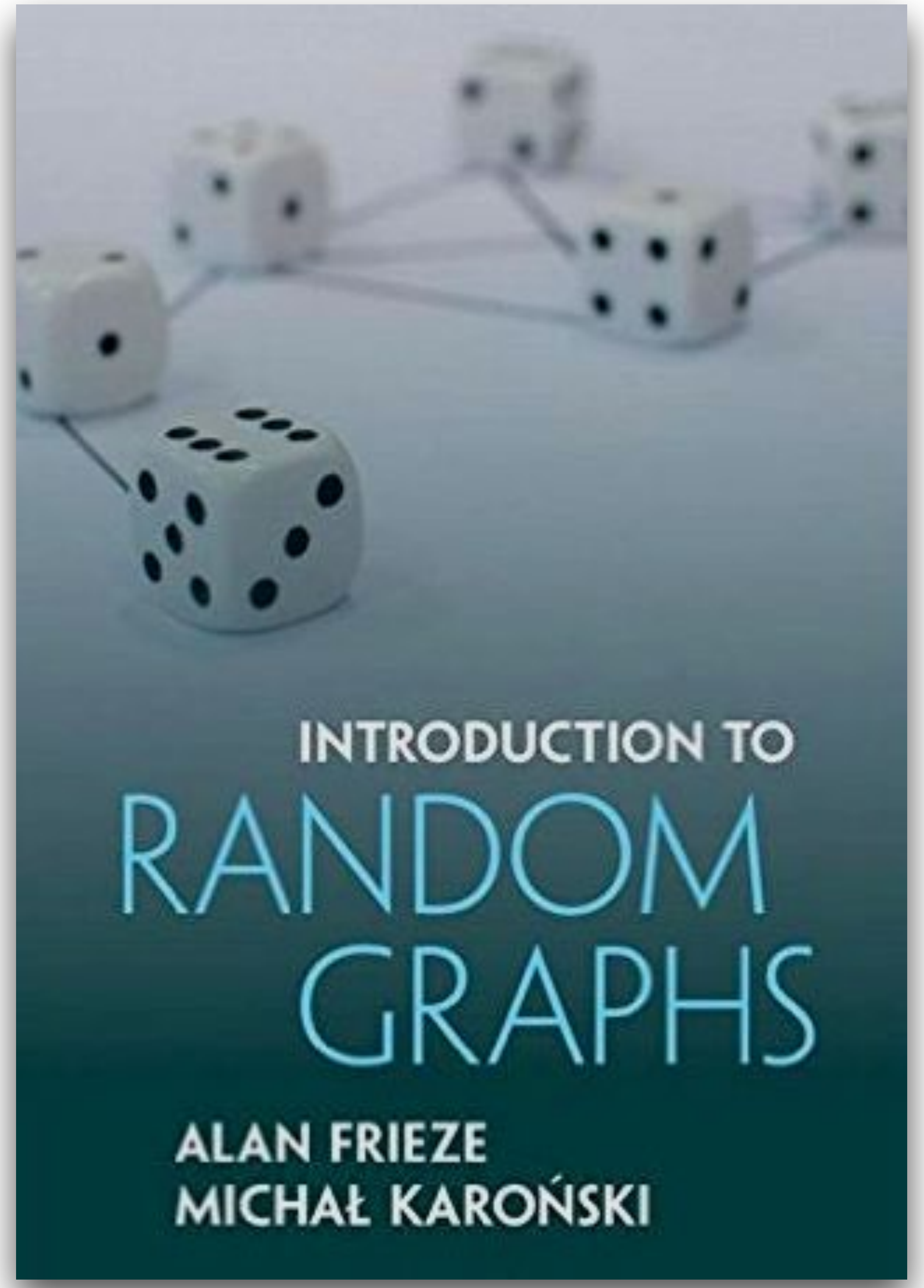
and

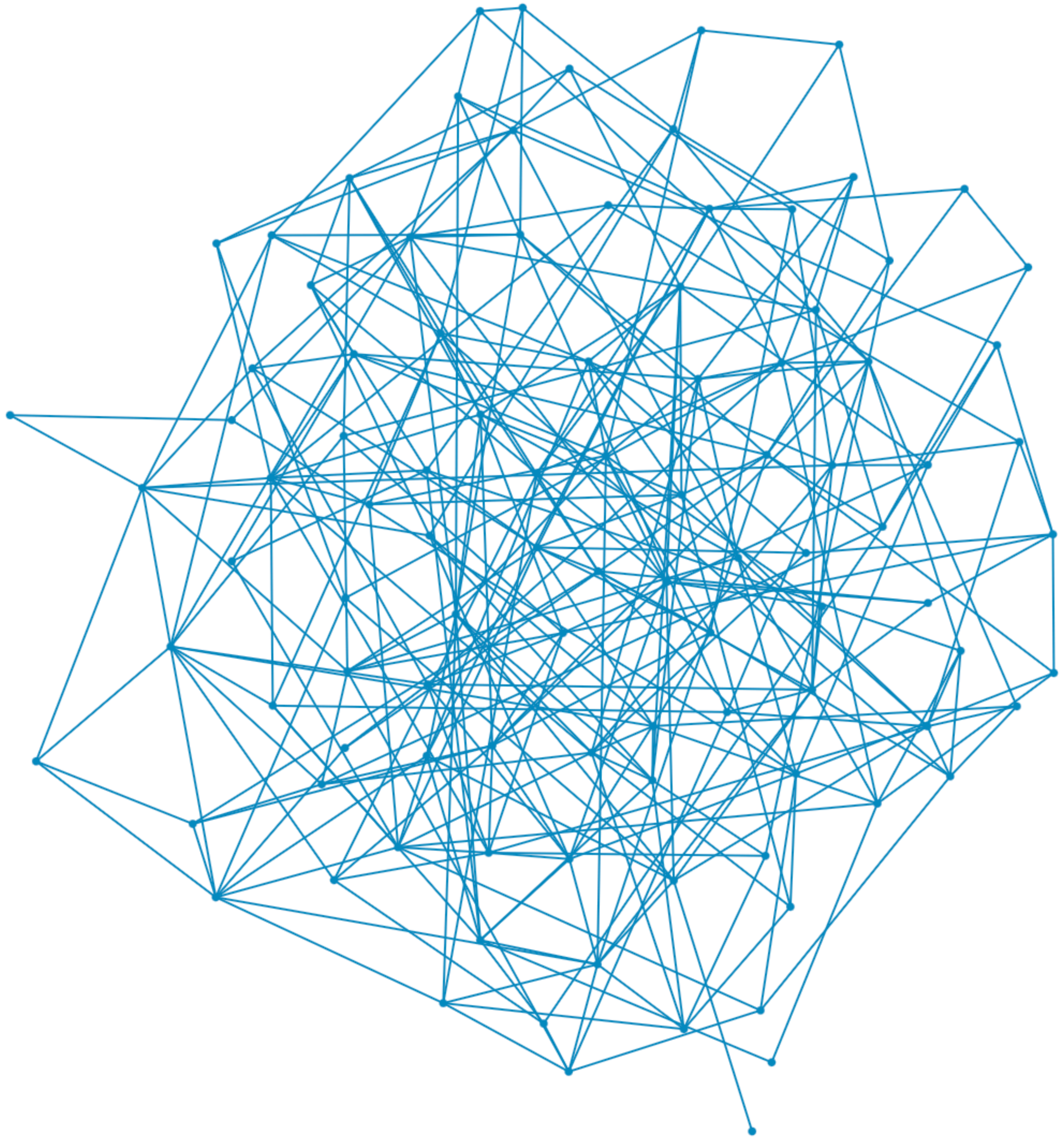
if *p* > 1/2 then
P(infinite cluster) = 1

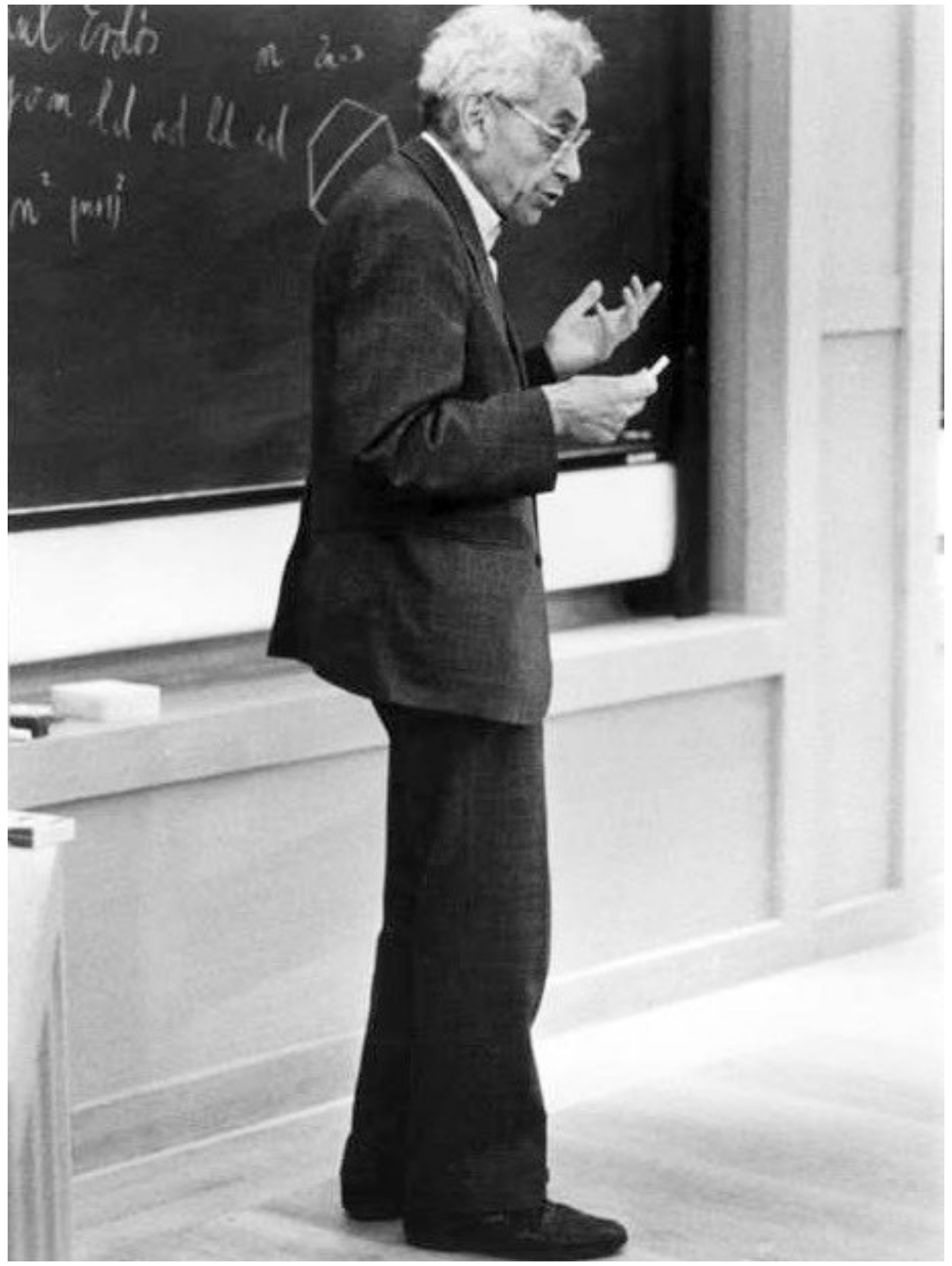


Randomness

The simplest model of a network : everything is ***boring***







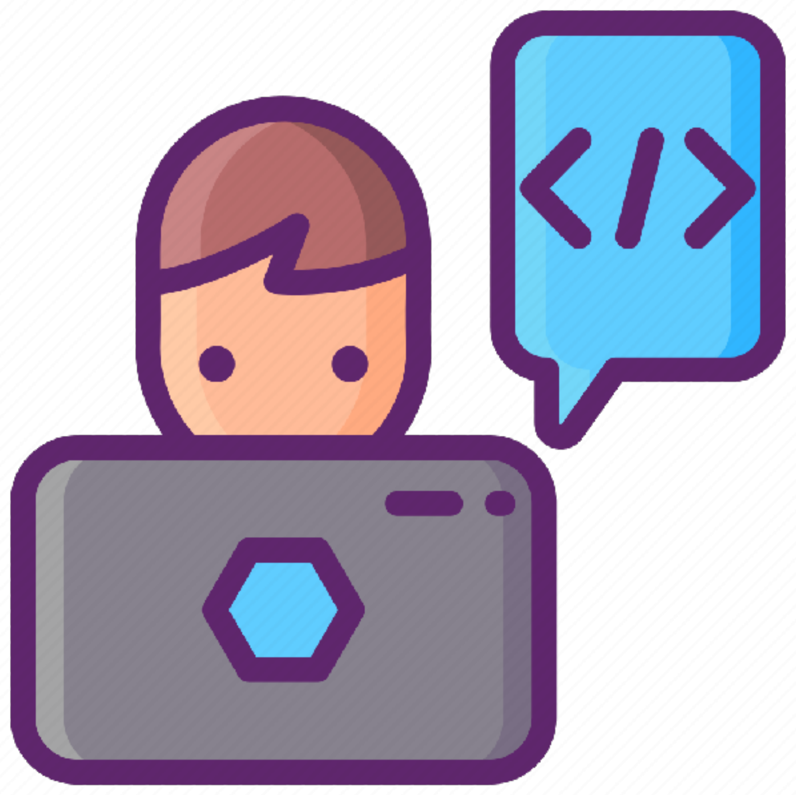
Paul Erdös (1913-1996)

Simulating Random Graphs

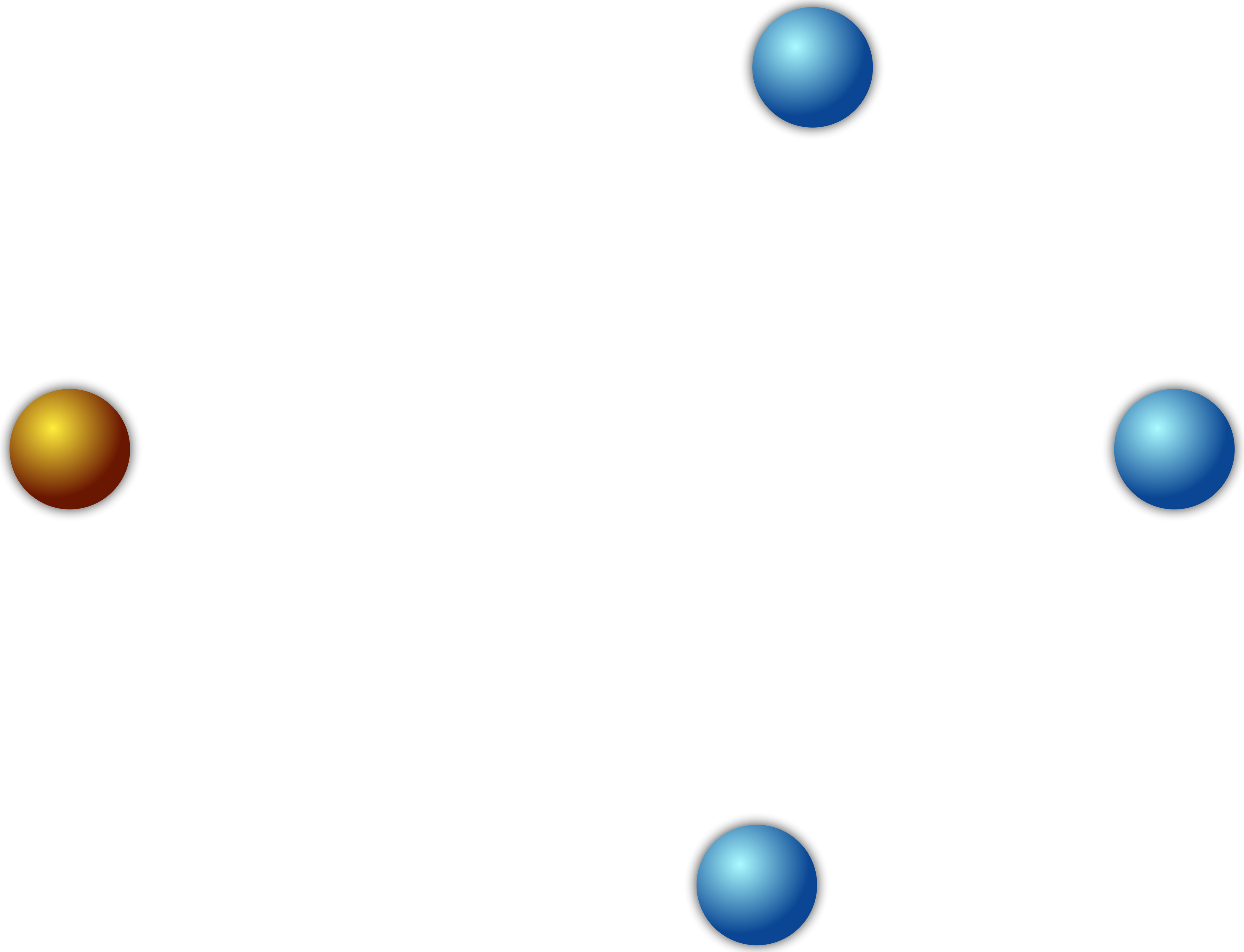
A static world without geography

N = number of nodes

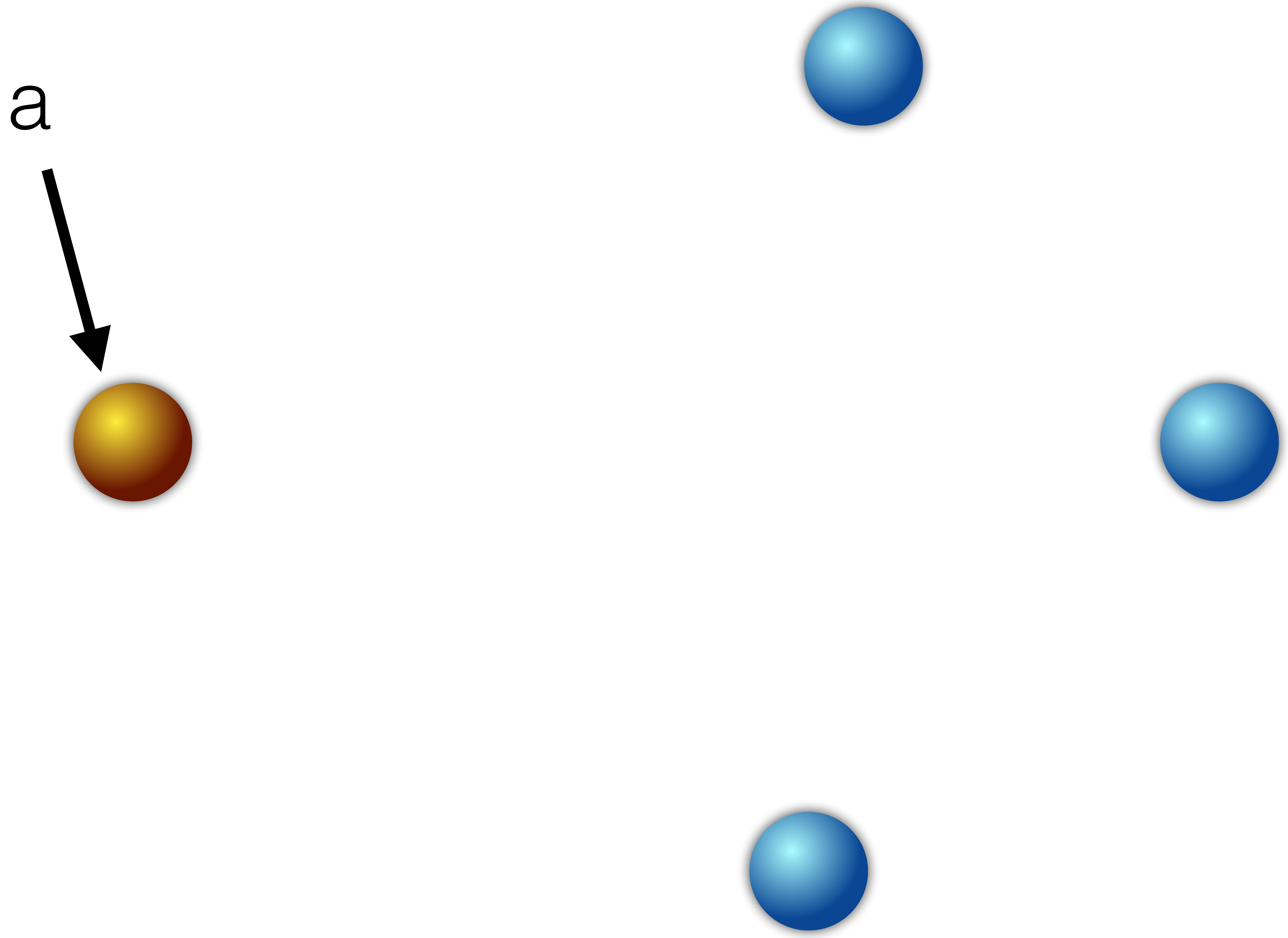
p = probability of connecting a pair of nodes



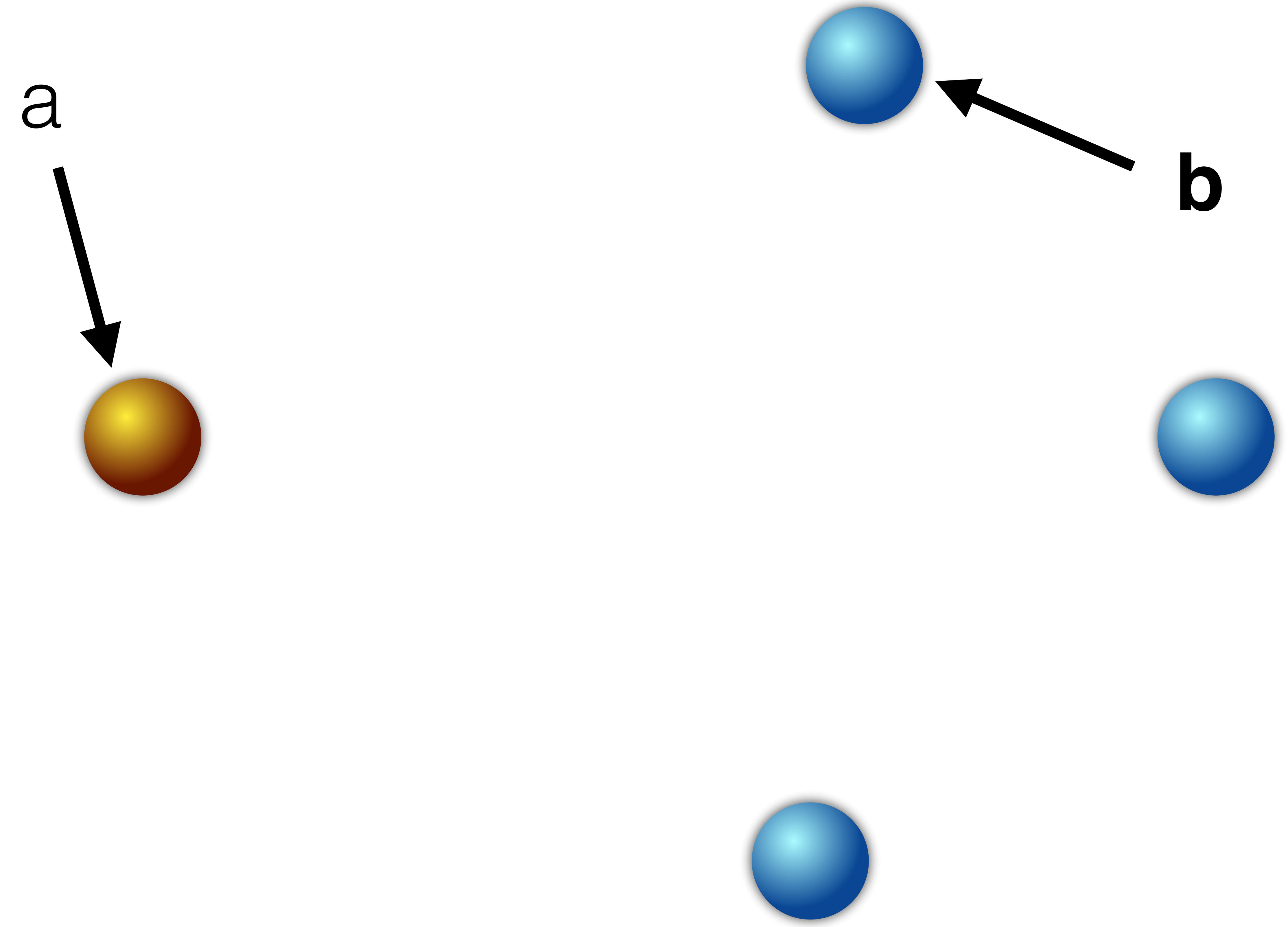
`create(4)`



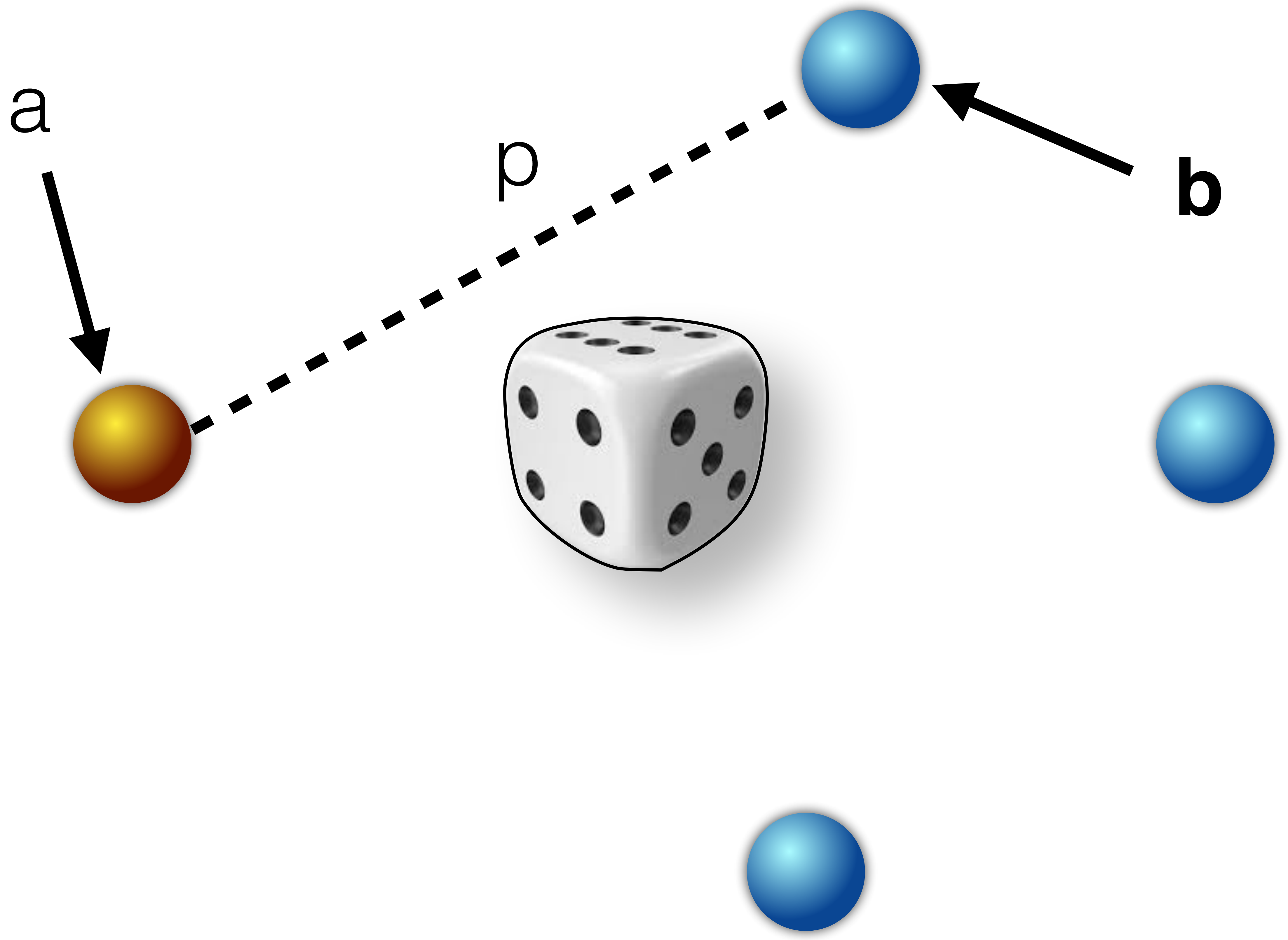
for each (a)



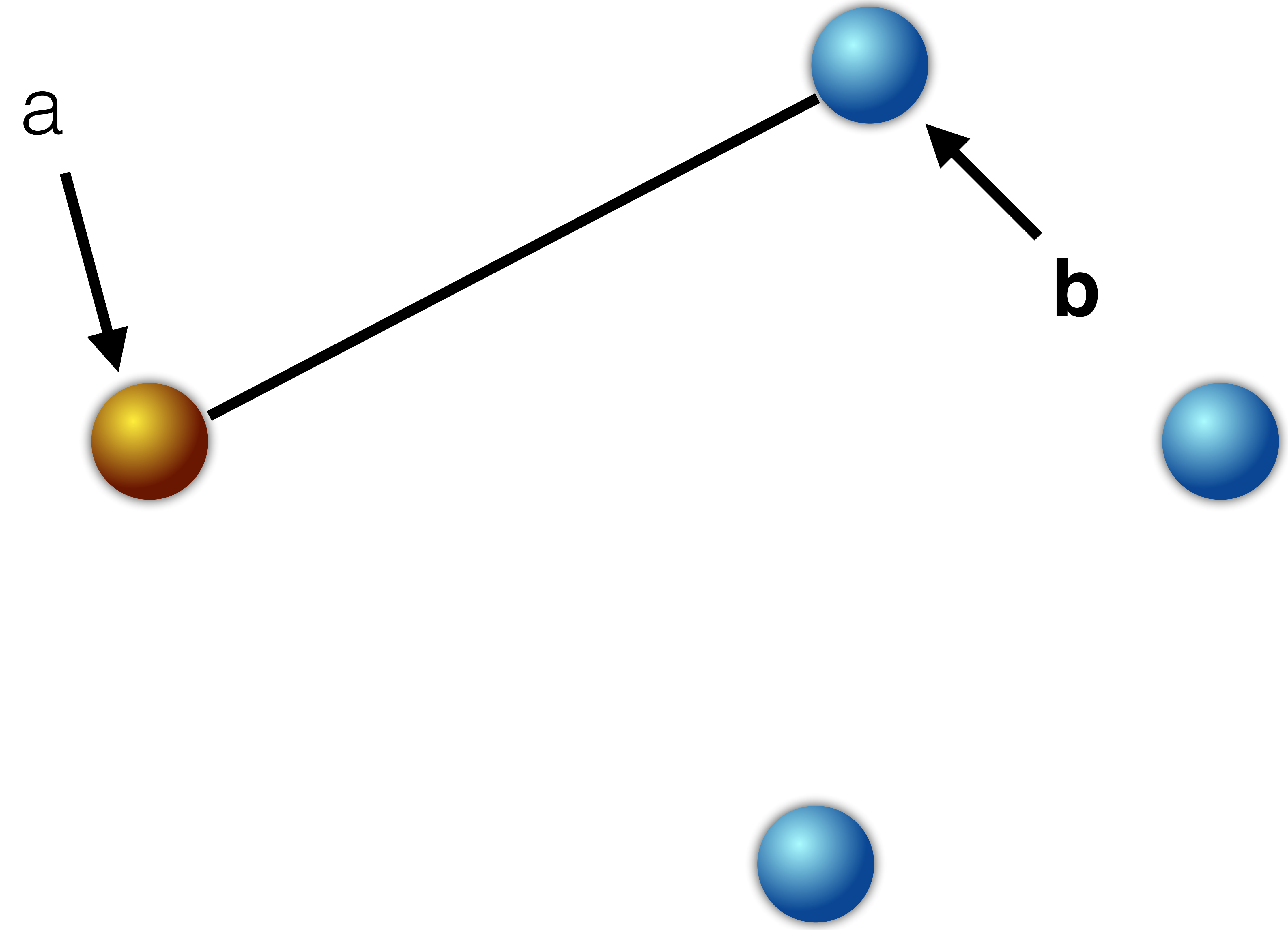
```
for each (a)
    for each (b)
```



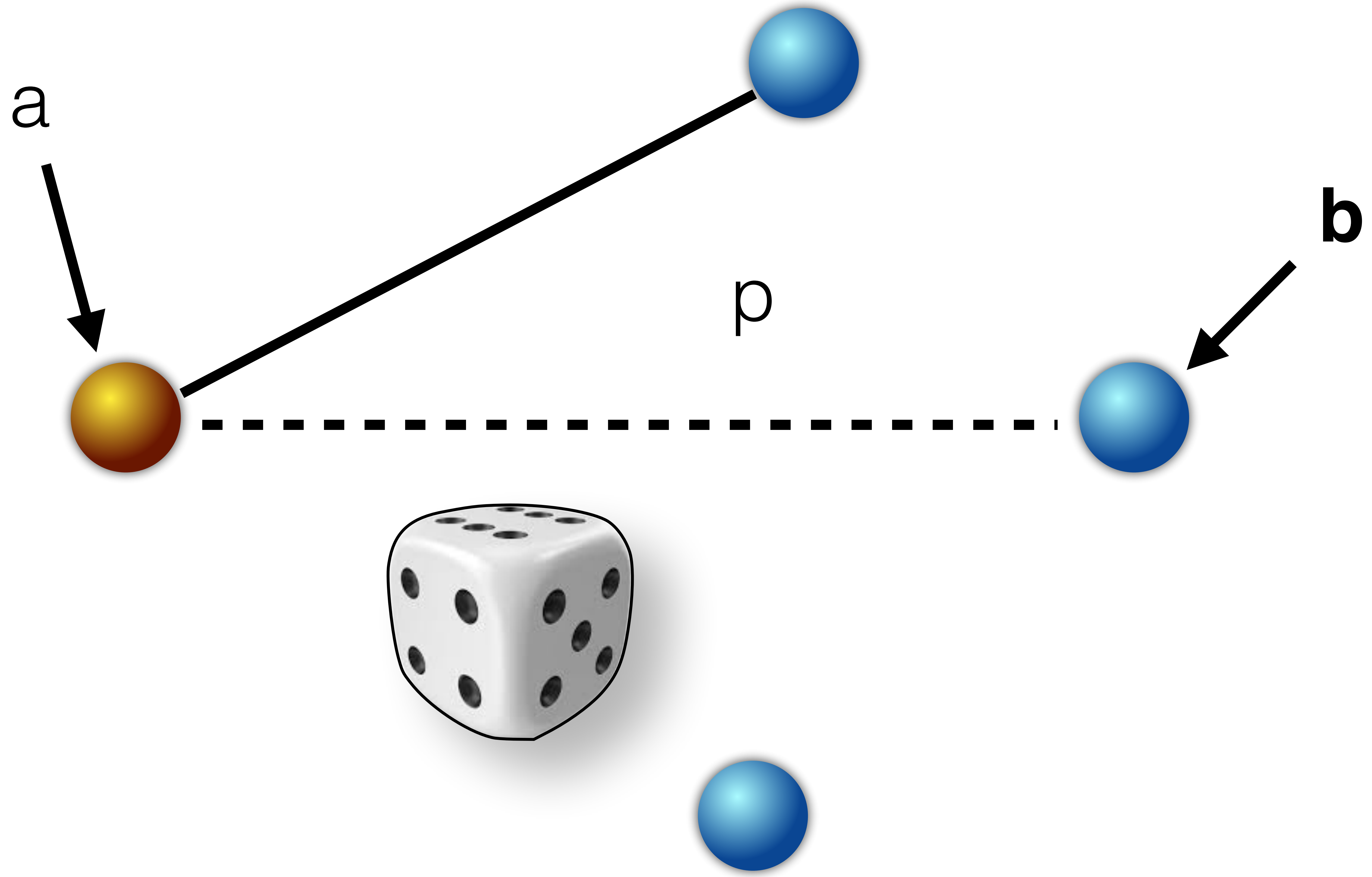
`random-float (1) < p`



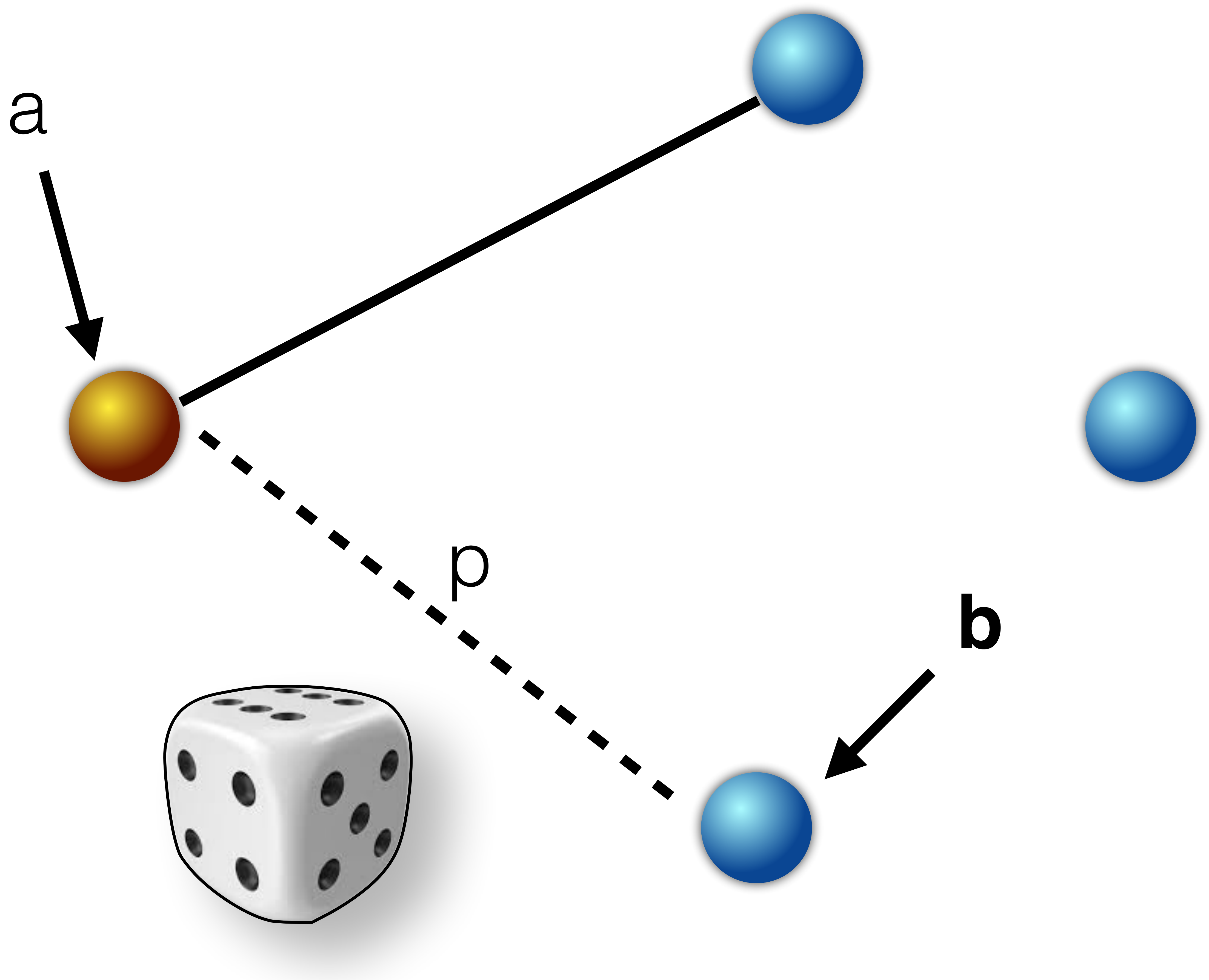
```
add_edge(a, b)
```

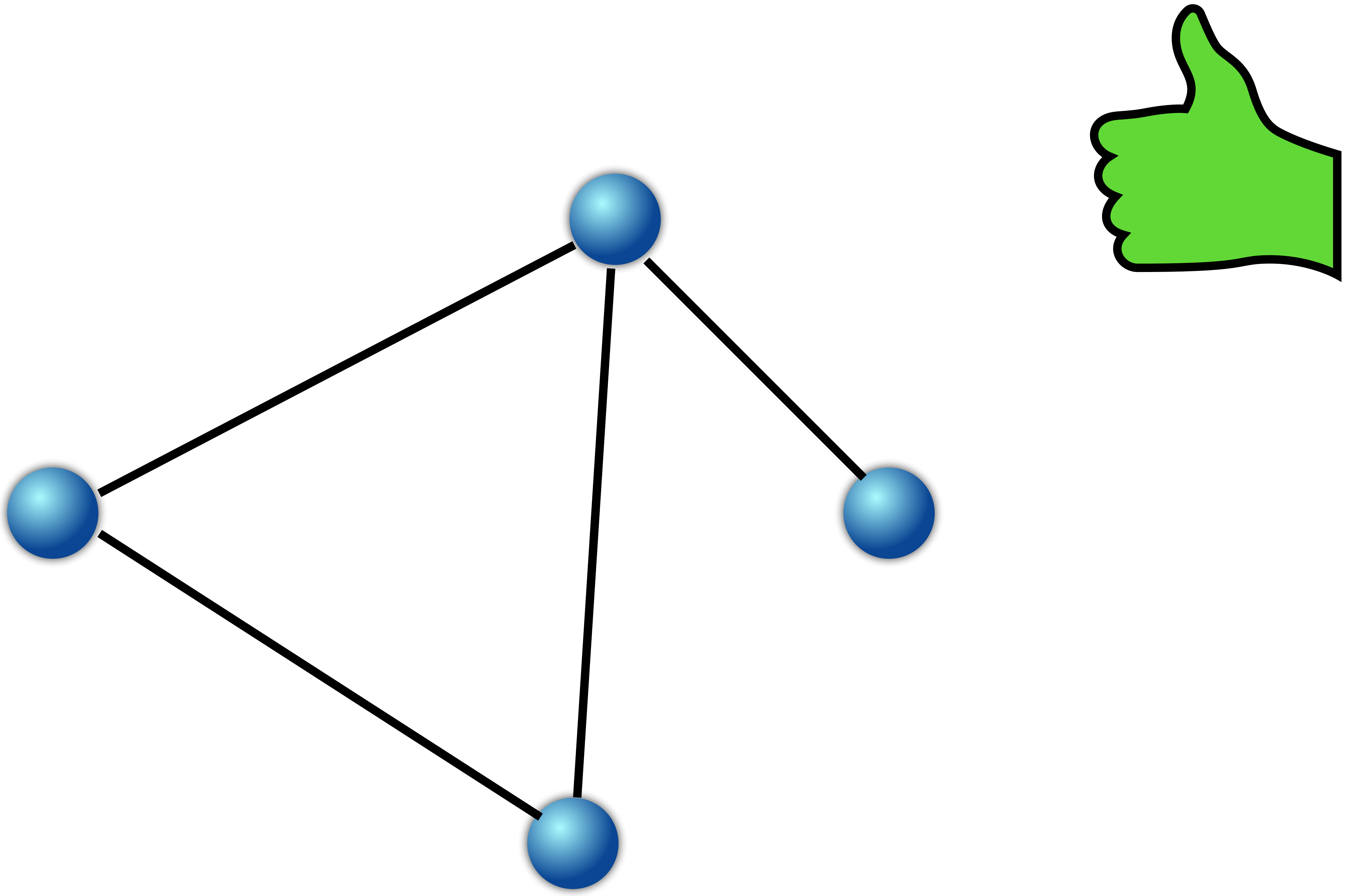


```
for each (a)
    for each (b)
```

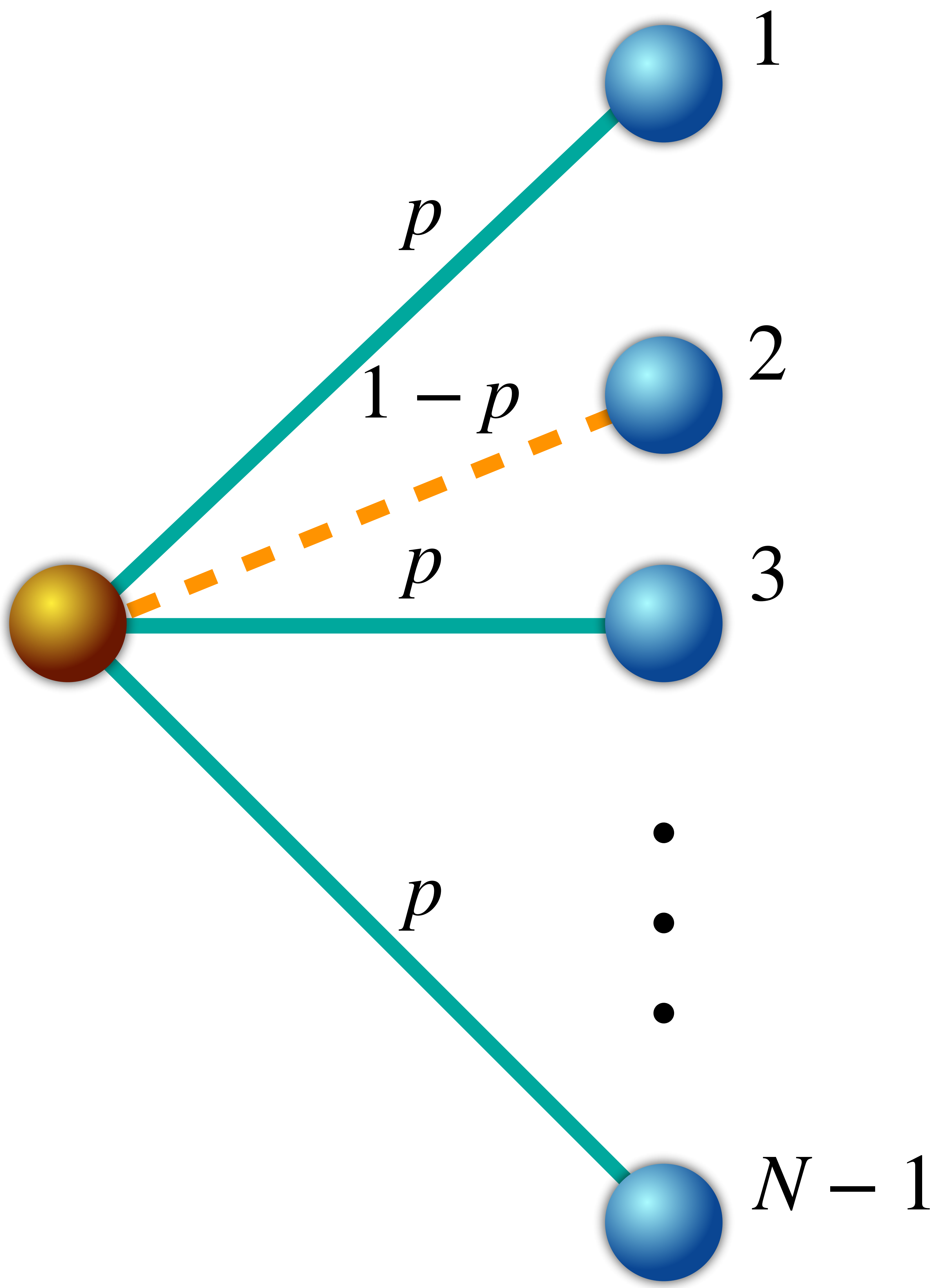


```
for each (a)
    for each (b)
```





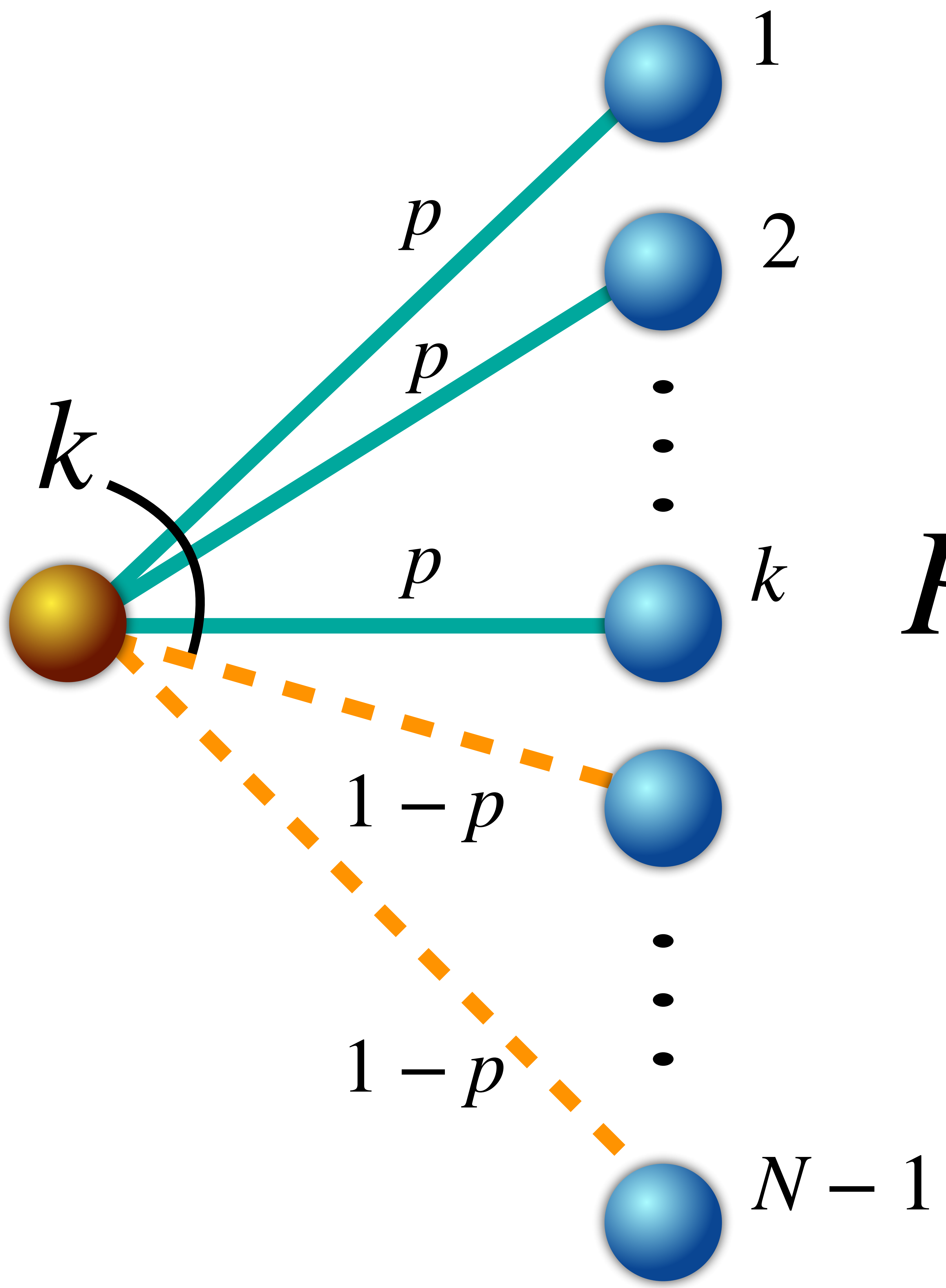
Average degree



$$L = p\binom{N}{2} = \frac{pN(N-1)}{2}$$

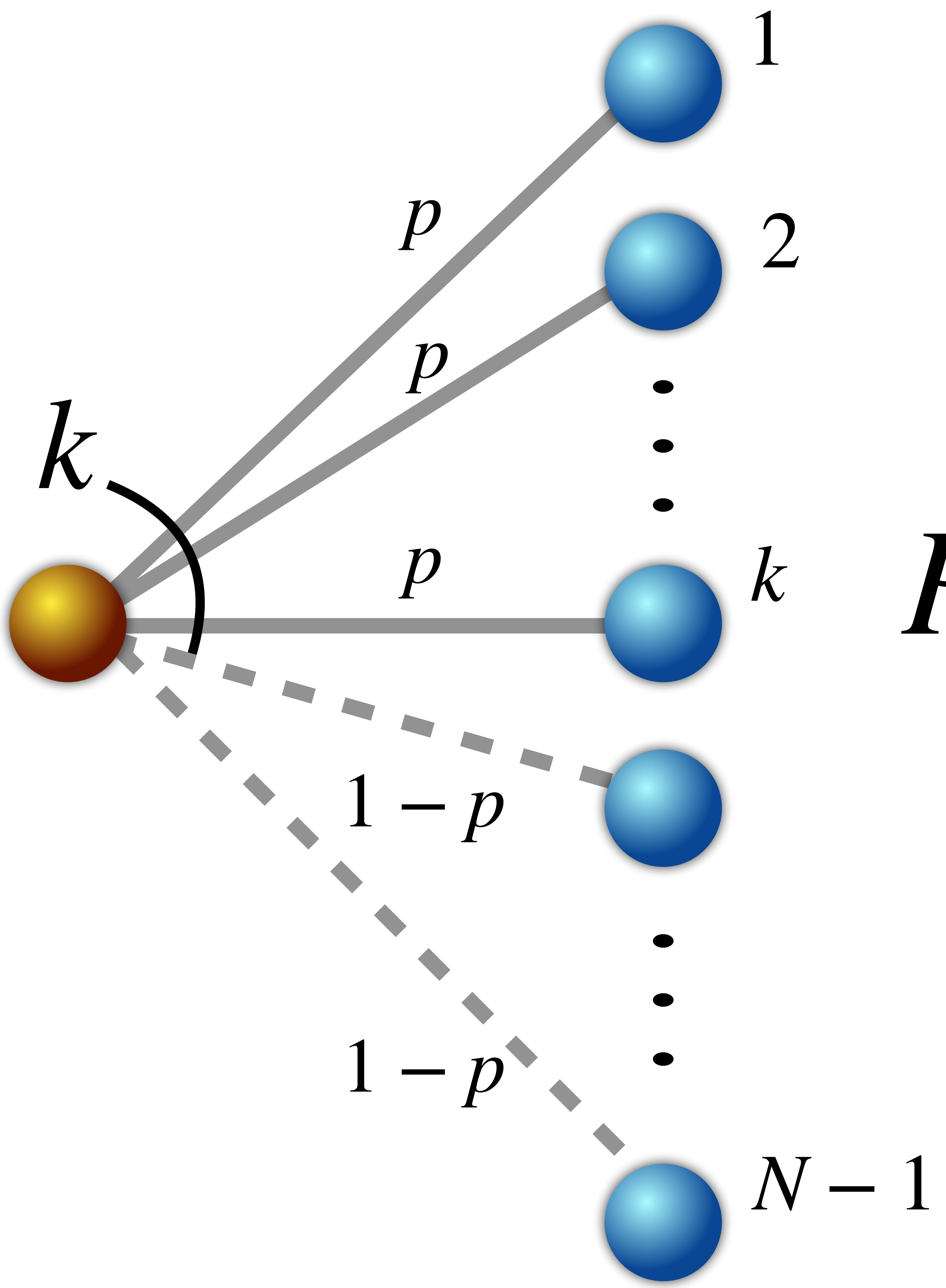
$$\langle k \rangle_{rand} = \frac{2L}{N} = (N-1)p$$

Degree Distribution



$$P(k) = p^k (1-p)^{N-1-k}$$

Degree Distribution



Discrete Binomial

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Degree Distribution

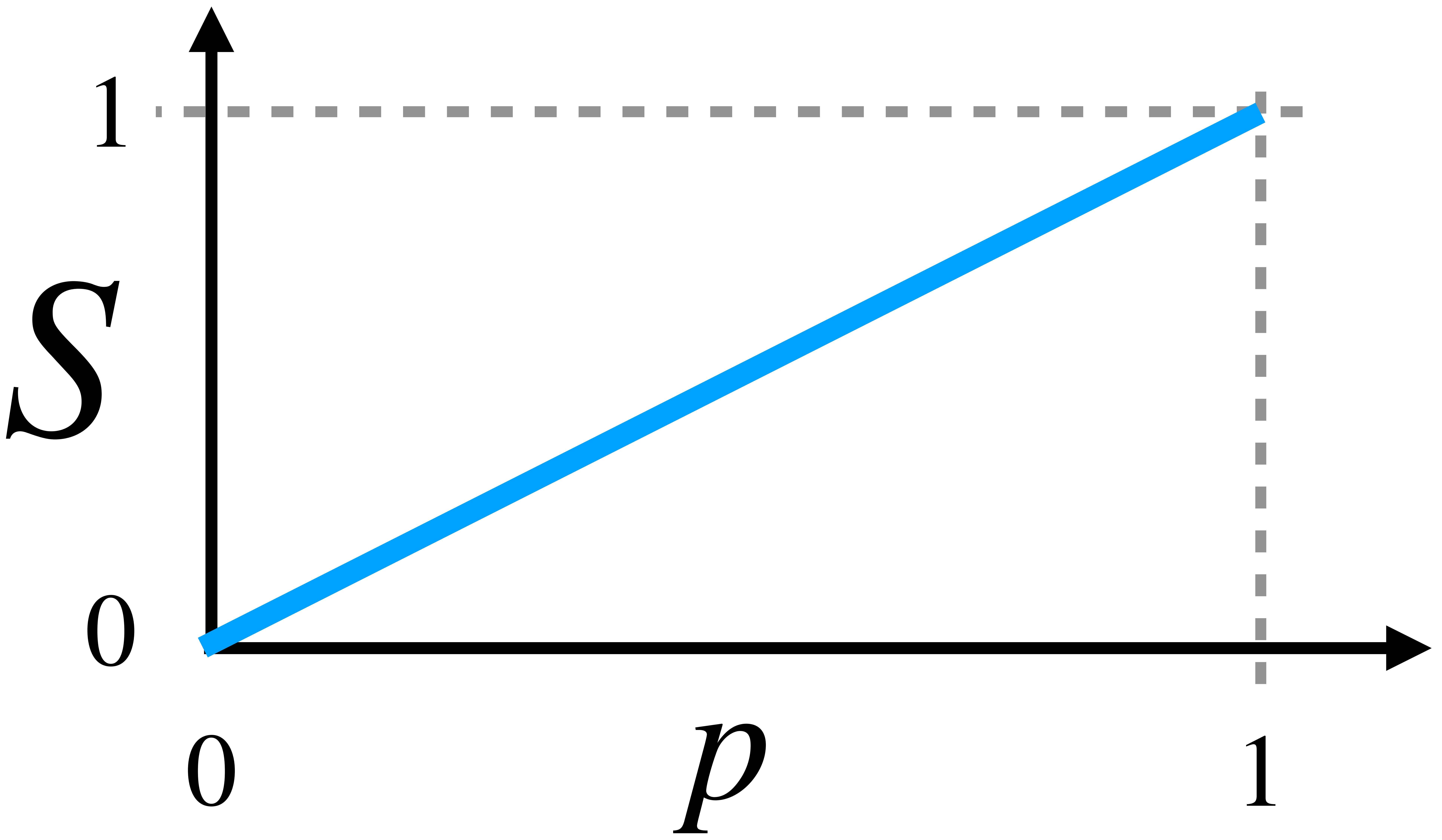
Poisson Distribution

$$P(k) = e^{-z}\left(\frac{z^k}{k!}\right)$$

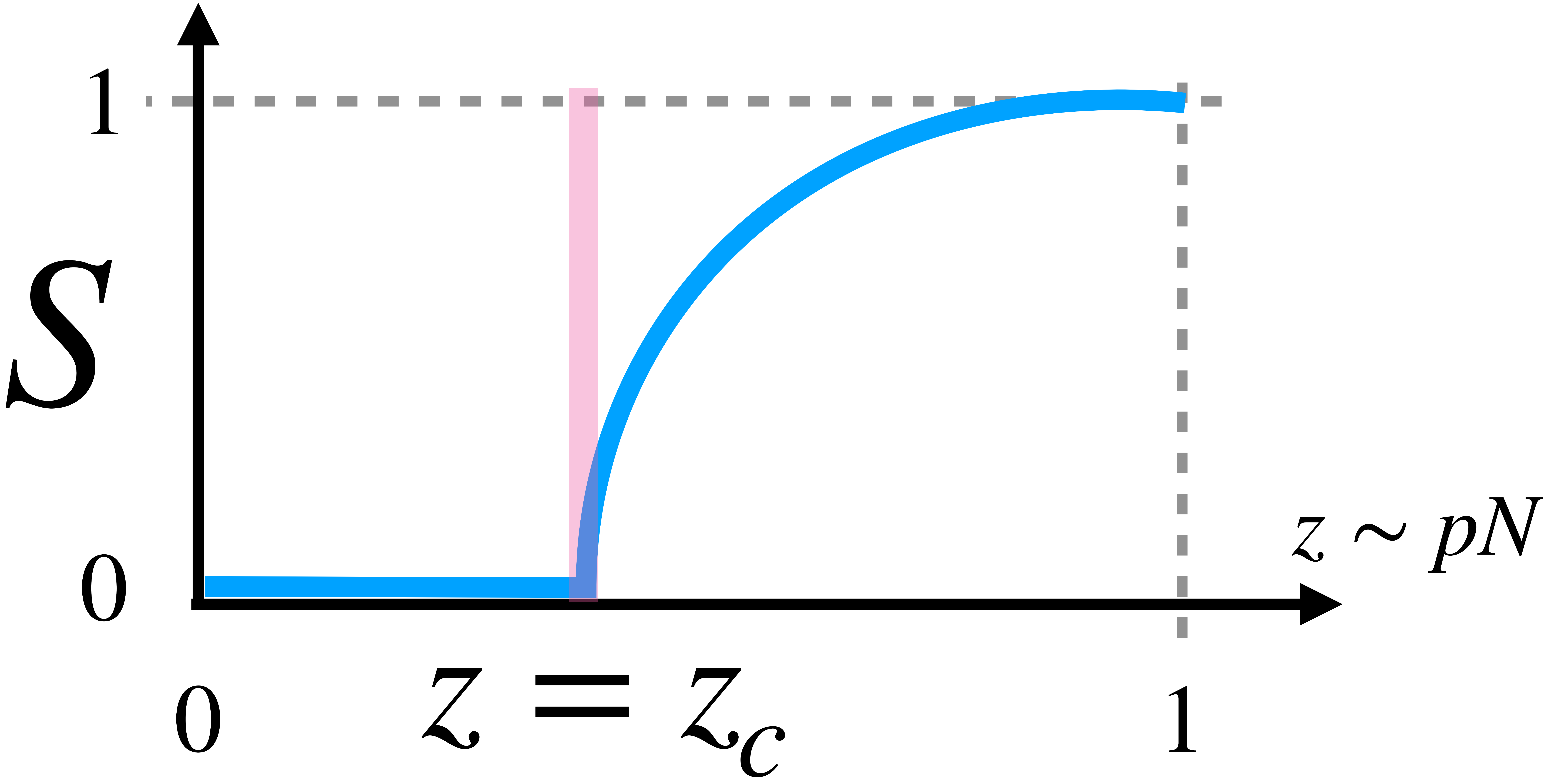
Percolation Transition

$p = 0$ $p = 1$

Percolation Transition



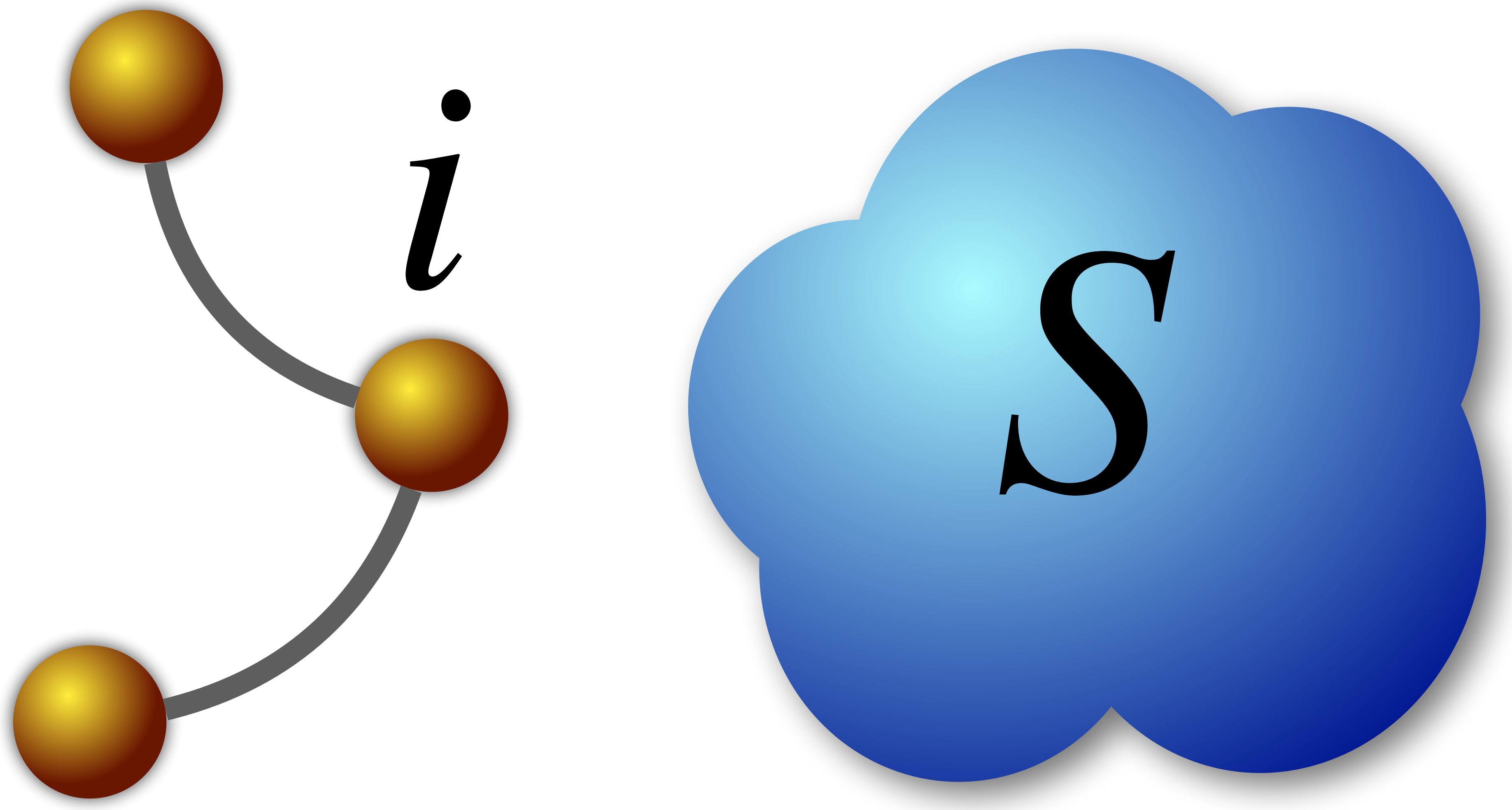
Percolation Transition



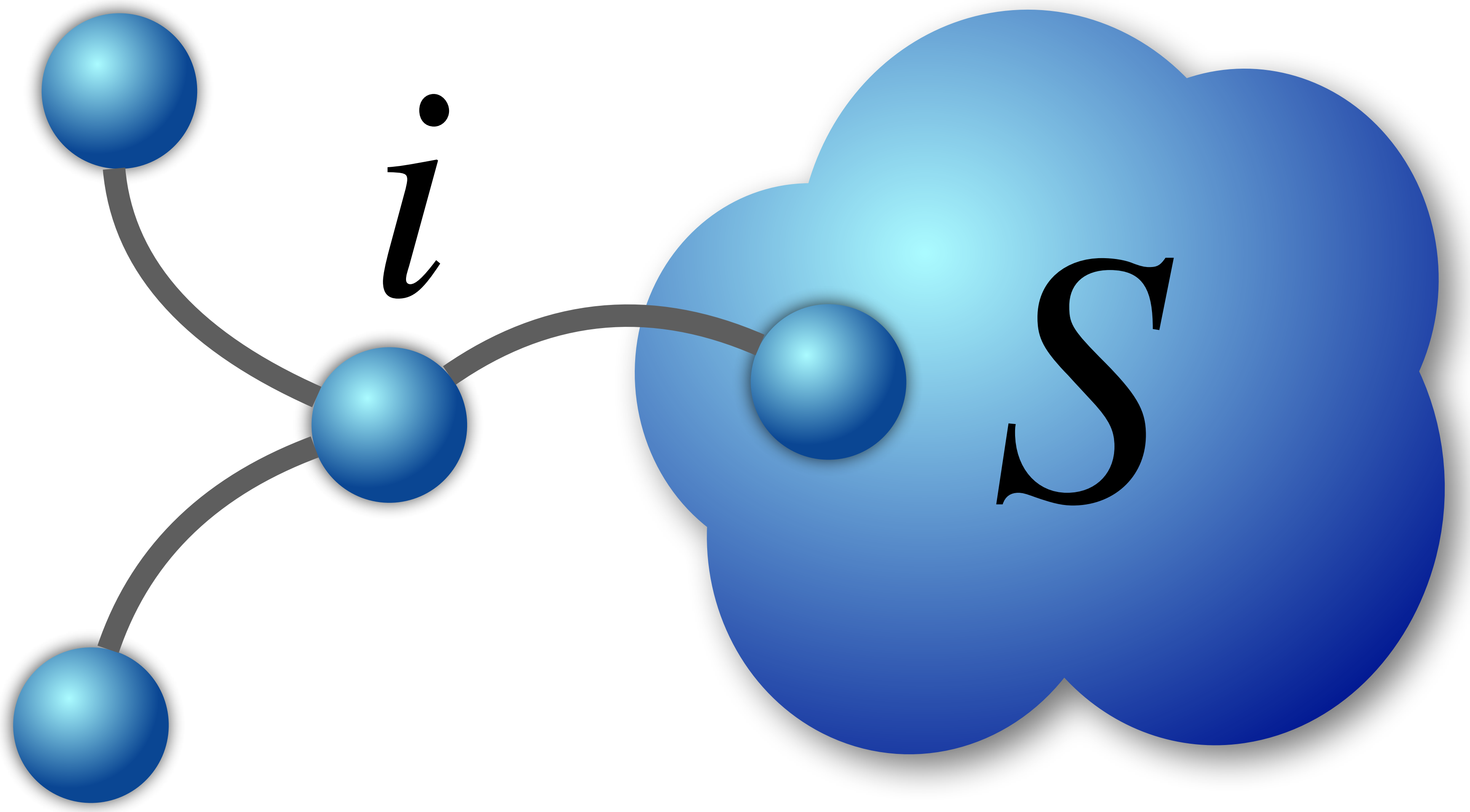
Percolation Transition

$Q = 1 - S =$ *Probability that the vertex i does* ***not*** *belong to the giant connected component*

Disconnected

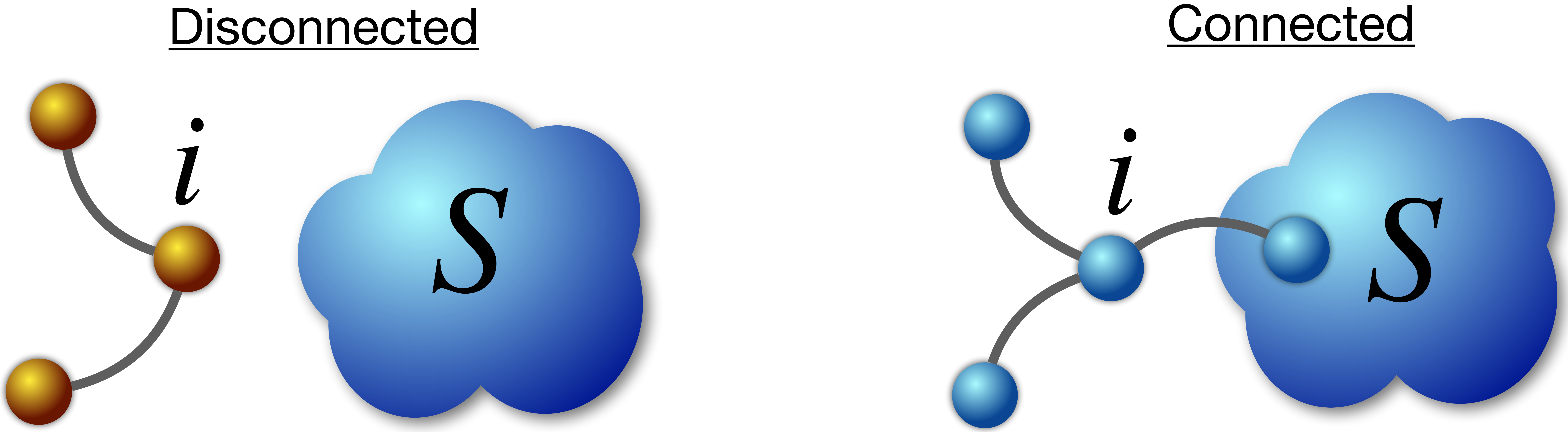


Connected



Percolation Transition

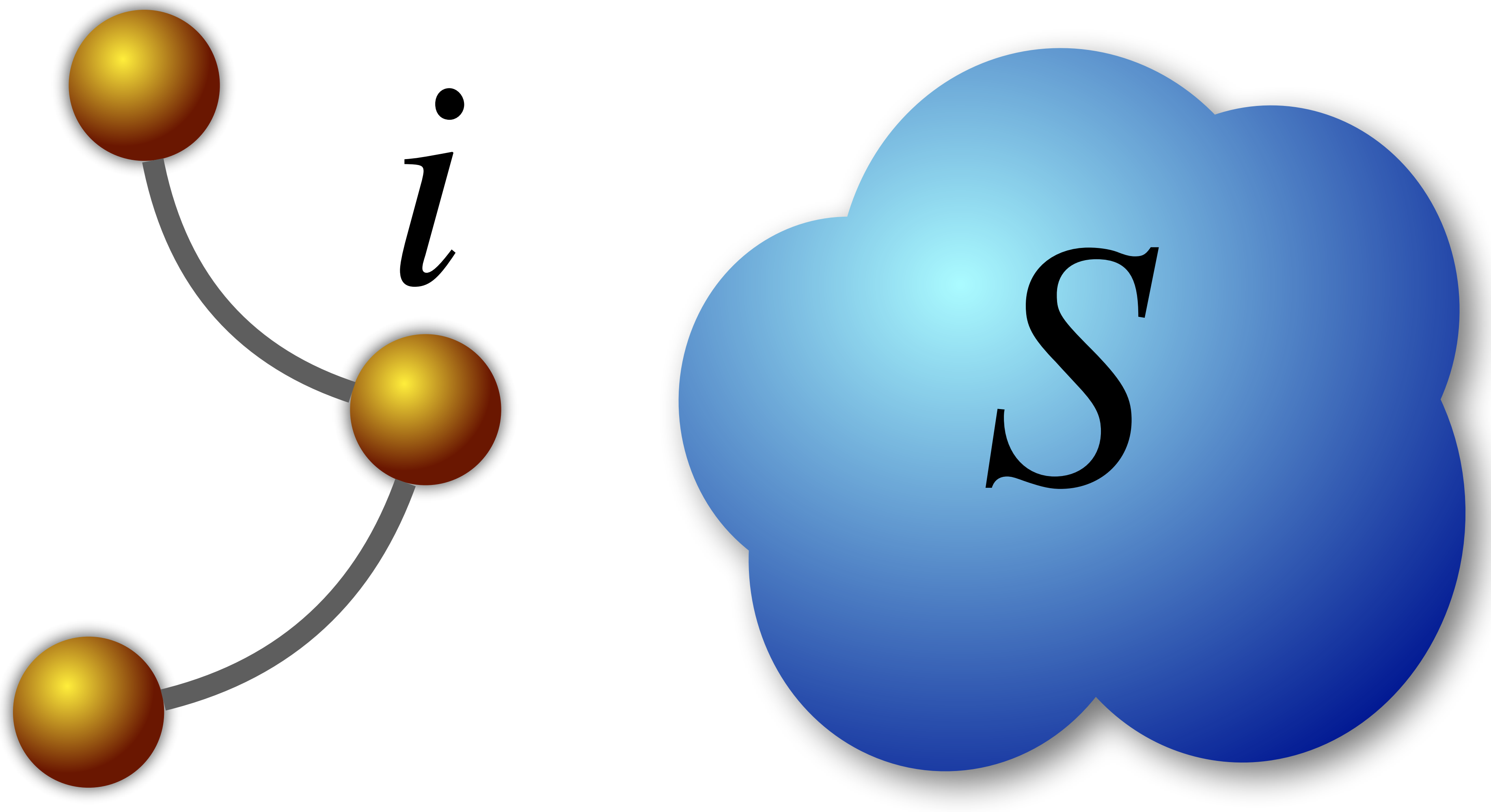
$Q^k =$ *Probability that* ***none*** *of its k neighbours belongs to the giant connected component*



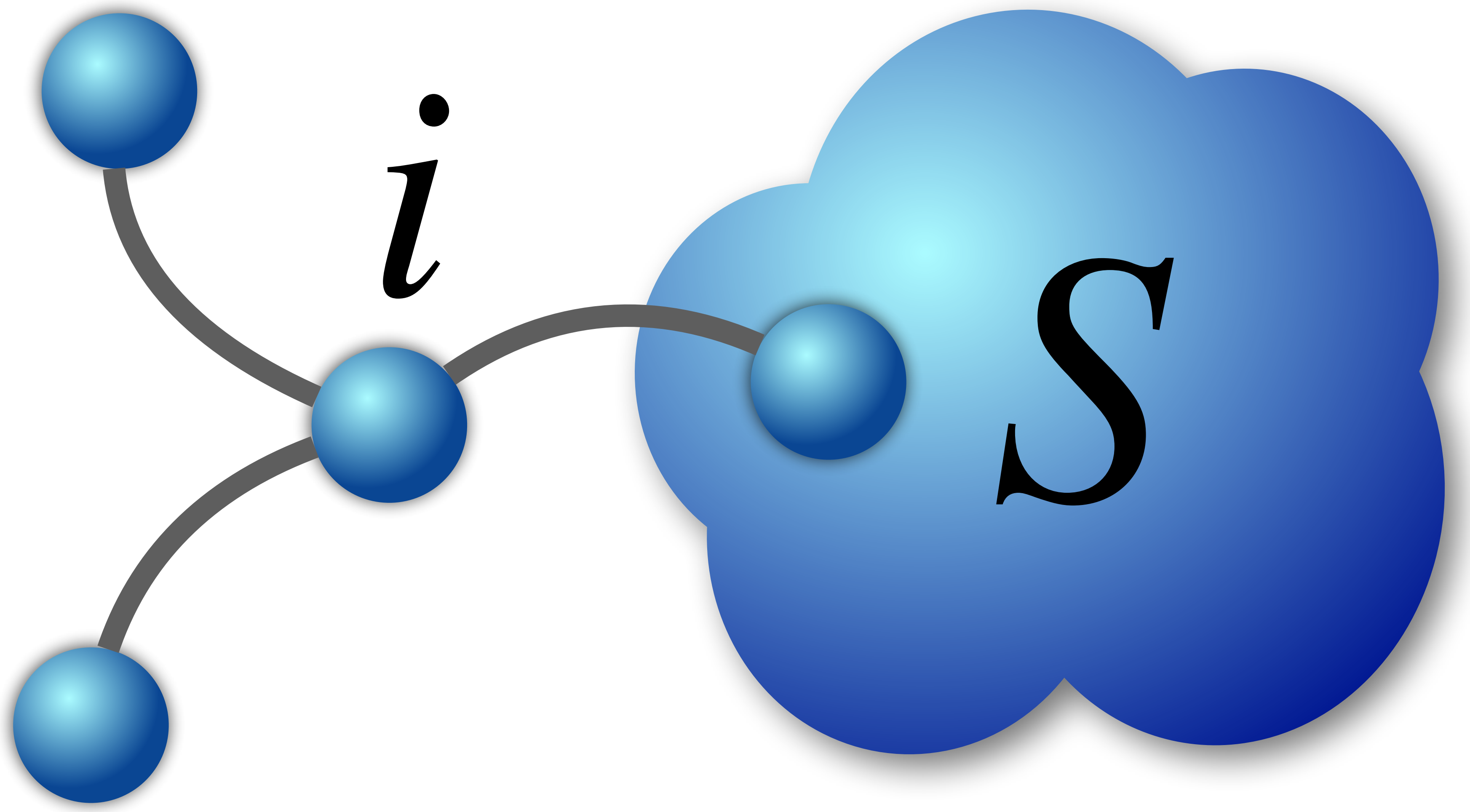
Percolation Transition

$$Q \equiv \langle Q \rangle = \sum_{k \geq 0} P(k) Q^k$$

<u>Disconnected</u>



<u>Connected</u>



Percolation Transition

$$Q = \sum_{k\geq 0} P(k) Q^k$$

$$= e^{-z} \sum_{k\geq 0} \frac{z^k}{k!} Q^k = e^{-z} \overbrace{\sum_{k\geq 0} \frac{(zQ)^k}{k!}}^{e^{zQ}} = e^{-z(1-Q)}$$

Percolation Transition

$$Q = e^{-z(1-Q)}$$

$$1 - S = e^{-zS}$$

$$\boxed{S = 1 - e^{-zS}}$$

Closed Form

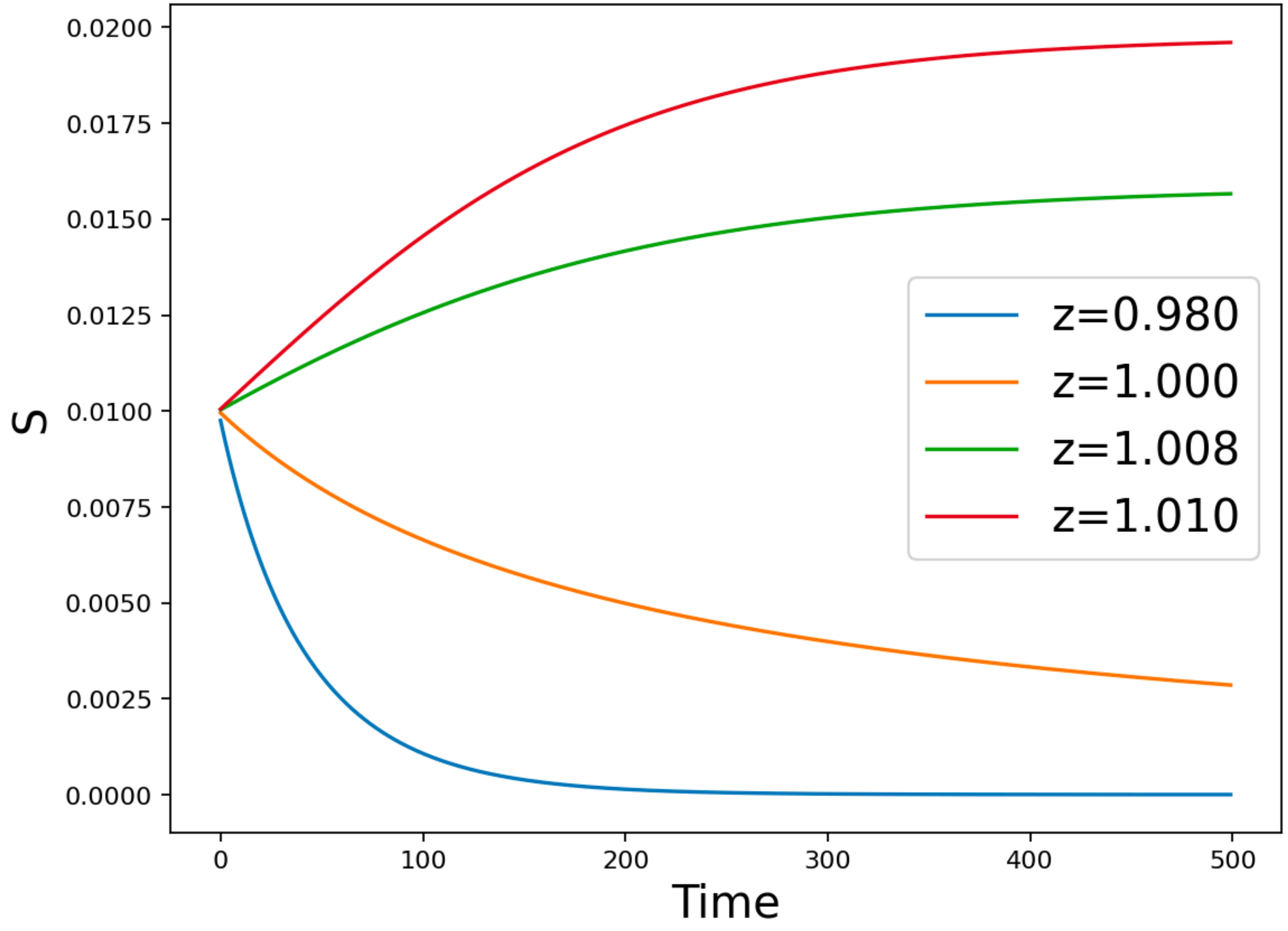
$$S = 1 - e^{-zS}$$

$$S^* = 0$$

$$S^* = 0\ ,\ \ z = 1$$

Numerical Solution

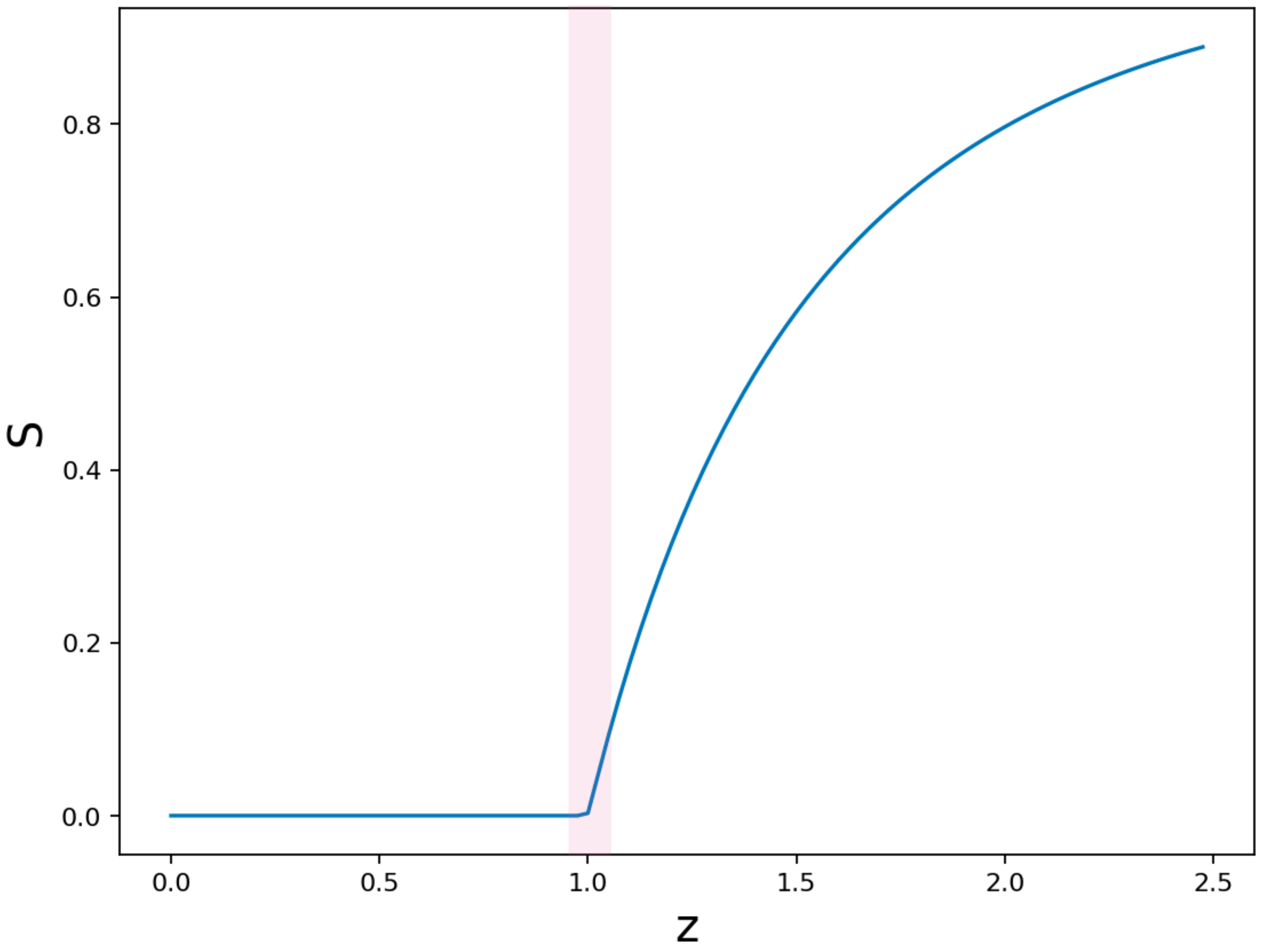
$$S = 1 - e^{-zS}$$




```
import matplotlib.pyplot as plt
import numpy as np
plt.figure(figsize=(8,6), dpi = 160)
x = range(500)
for z in [0.98, 1, 1.008, 1.01]:
    y = []
    S = 0.01
    for i in x:
        S = 1 - np.exp( -z * S)
        y.append (S)
    plt.plot ( x, y, label = "z=%0.03f"% z)
plt.xlabel ("Time", fontsize= 18)
plt.ylabel ("S", fontsize = 18)
plt.legend(fontsize = 18)
plt.show()
```

Numerical Solution

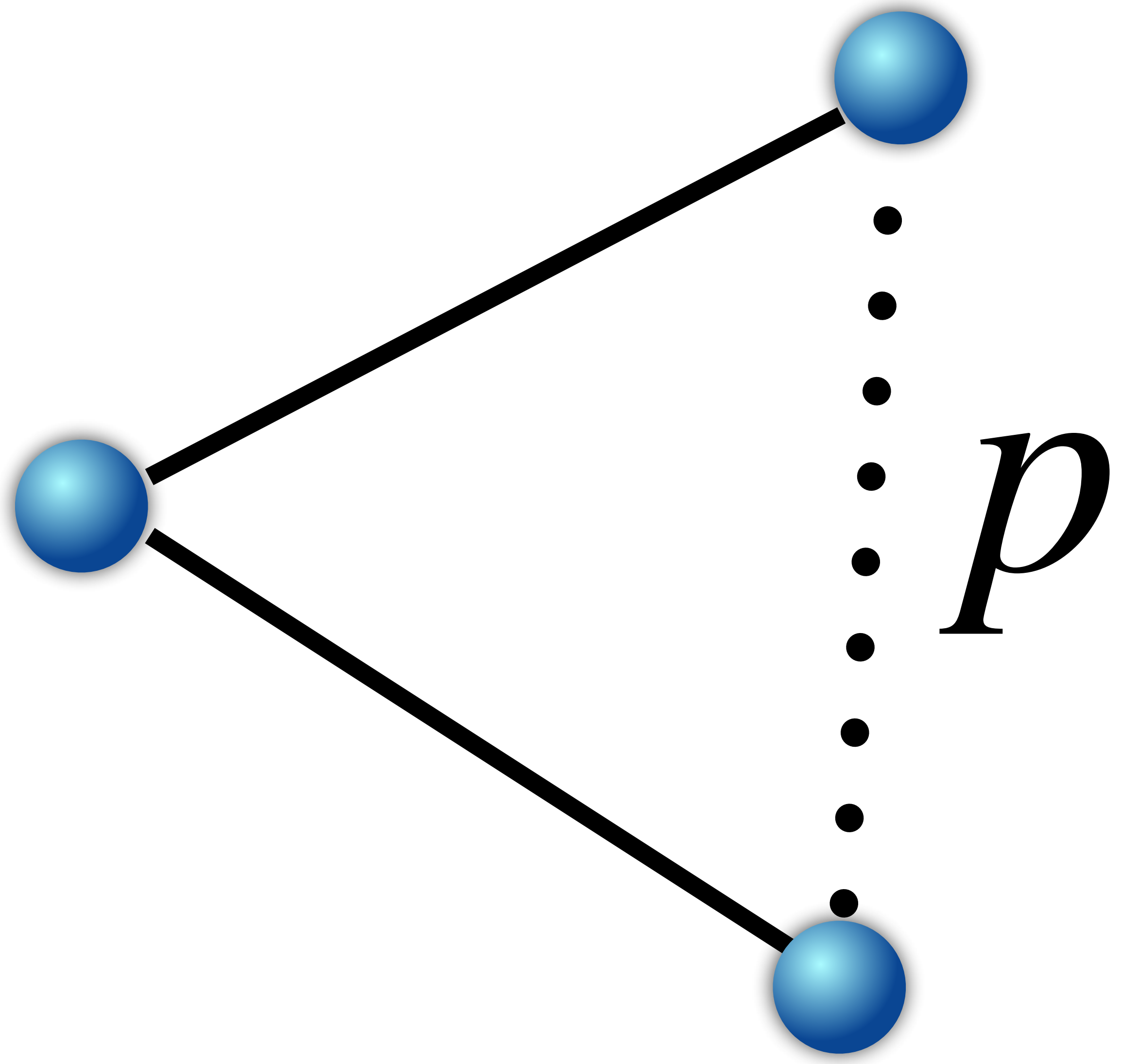
$$S = 1 - e^{-zS}$$




```python
import matplotlib.pyplot as plt
import numpy as np
plt.figure(figsize=(8,6), dpi = 160)
S_values = []
z_values = [float(i)/40.0 for i in range(100)]
for z in z_values:
    S = 0.01
    for j in range(500):
        S = 1 - np.exp( -z * S )
    S_values.append (S)
plt.xlabel ("z", fontsize= 18)
plt.ylabel ("S", fontsize = 18)
plt.plot (z_values, S_values)
plt.show()
```

Clustering

Random graphs do not display clustering



$$\langle C \rangle_{rand} = p$$

$$\langle C \rangle_{rand} = p = \frac{\langle k \rangle_{rand}}{N-1}$$

Clustering

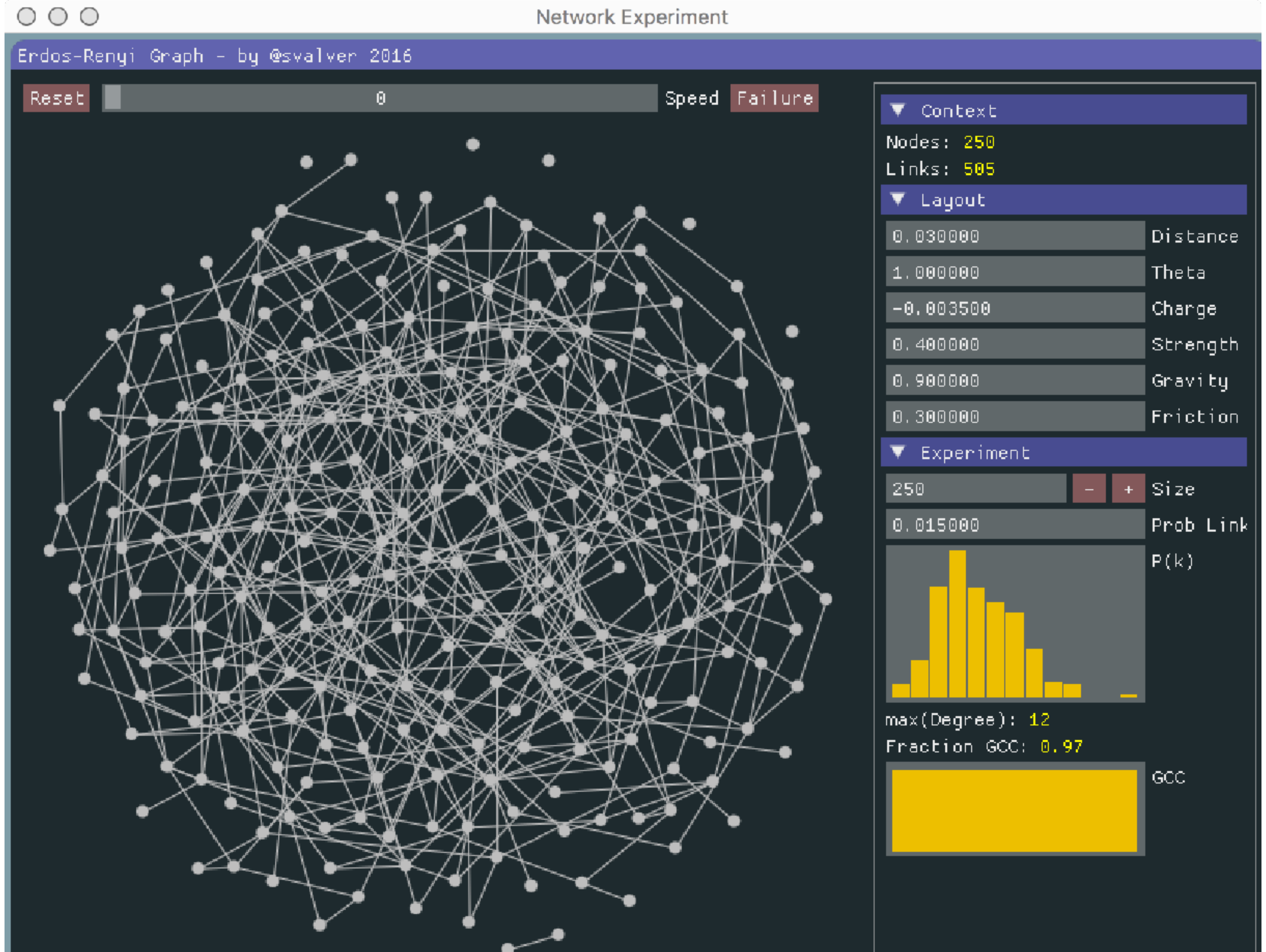
… but real-world ***graphs*** *do!*

$$0.01 \leq \langle C \rangle_{Facebook} \leq 0.5$$

$$\langle C \rangle_{rand} = \frac{\langle k \rangle}{N-1} = \frac{10^3}{10^9} \approx 0.00000001$$

Activity: Random Networks

https://tinyurl.com/3p9fxnsc

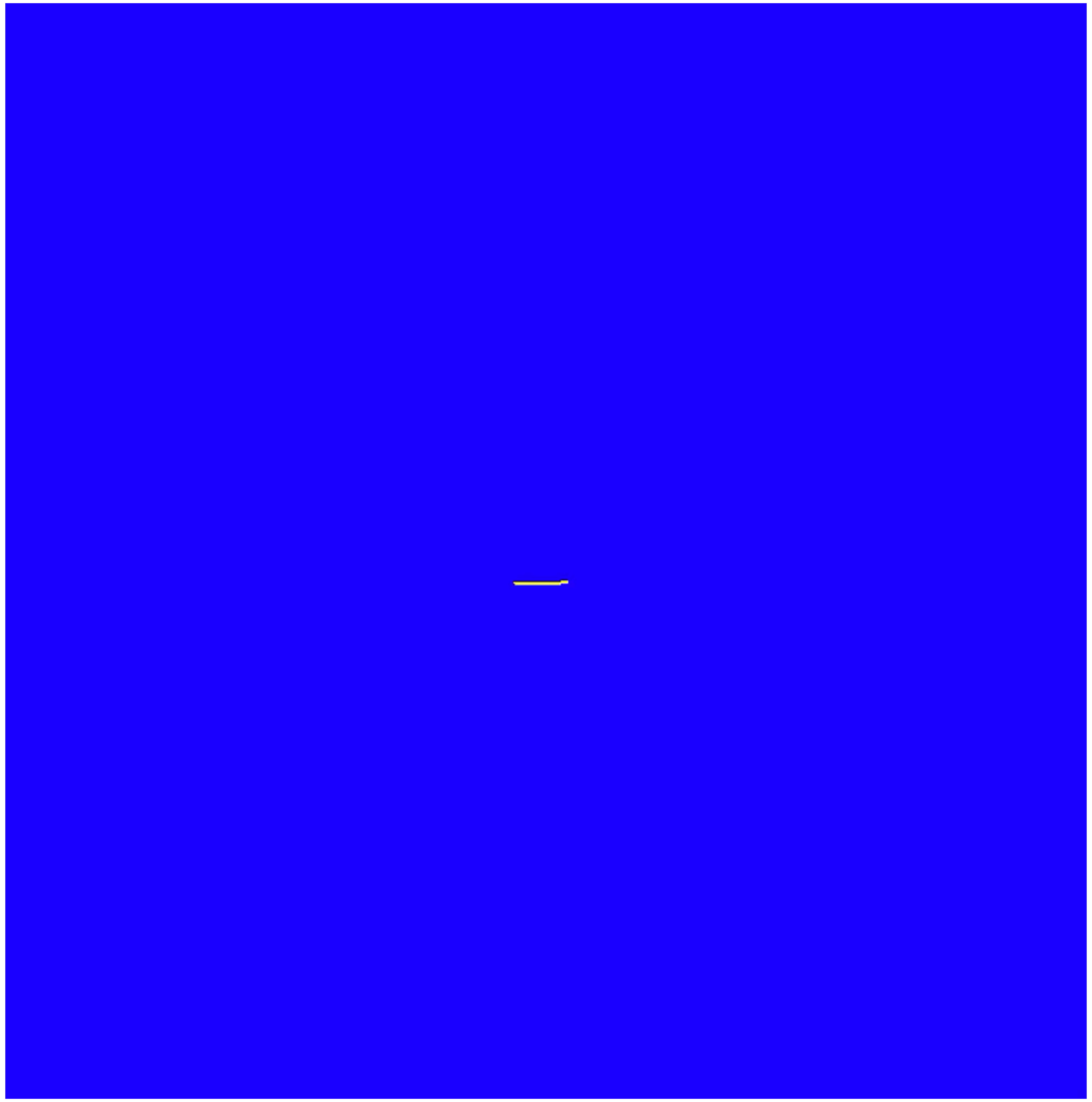


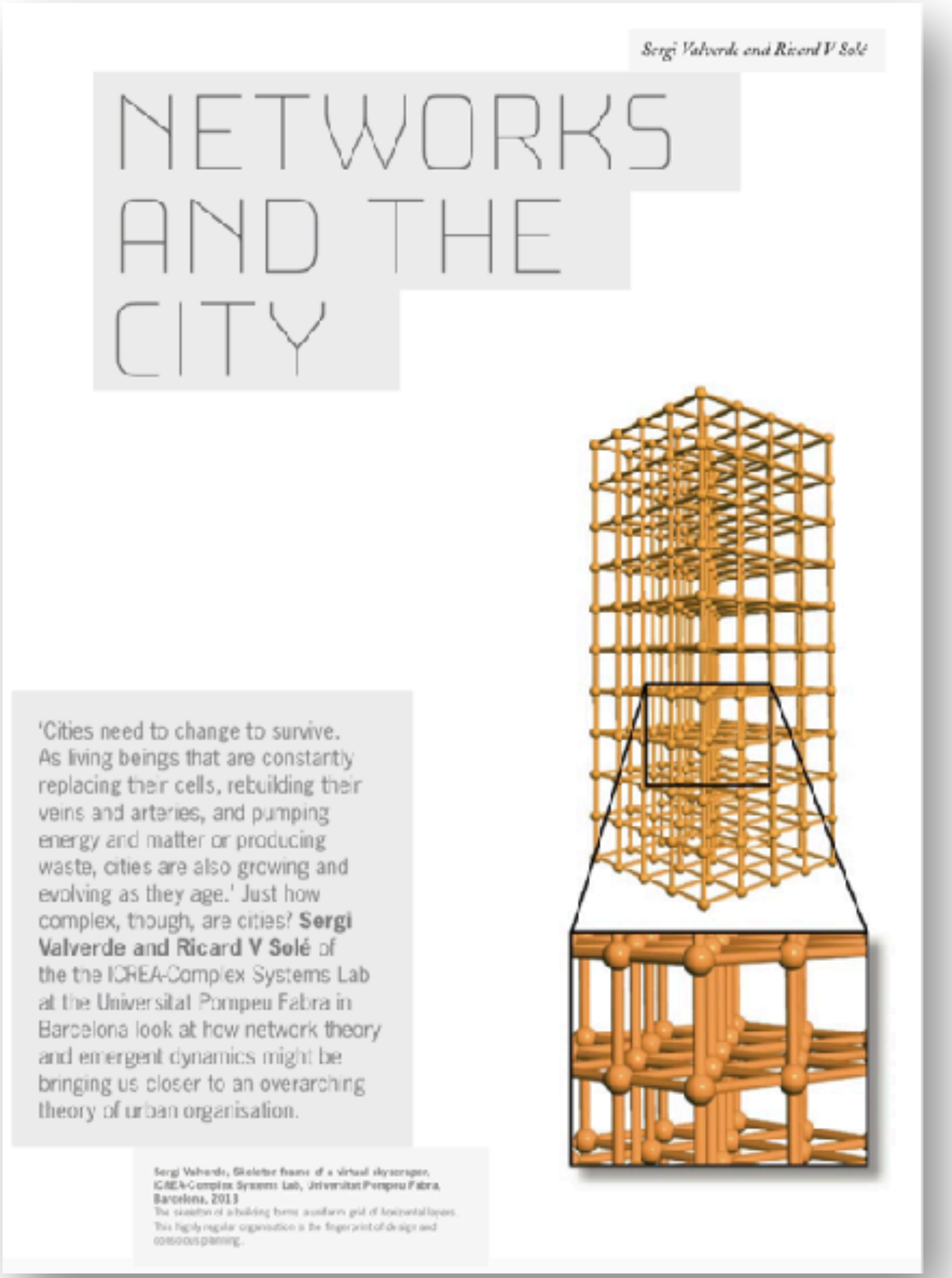
3. Can you predict the average degree before running the simulation?

4. Is it possible to obtain a node with a very large number of links?

Growth: City Networks

Man-made objects can be geometrically complex and do not resemble ideal forms such as points, lines, planes, cubes, circles of spheres.





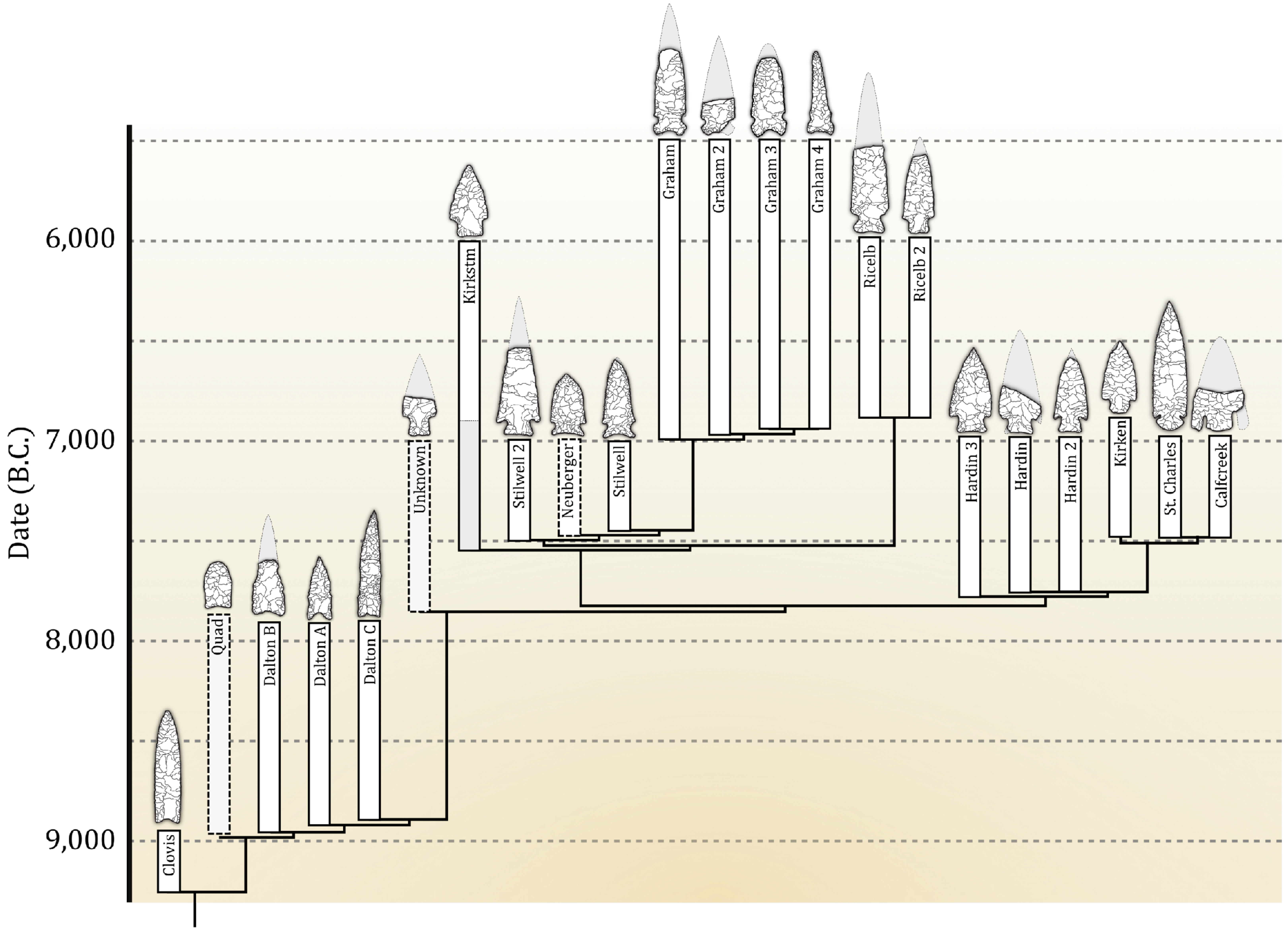
Sergi Valverde and Ricard V Solé

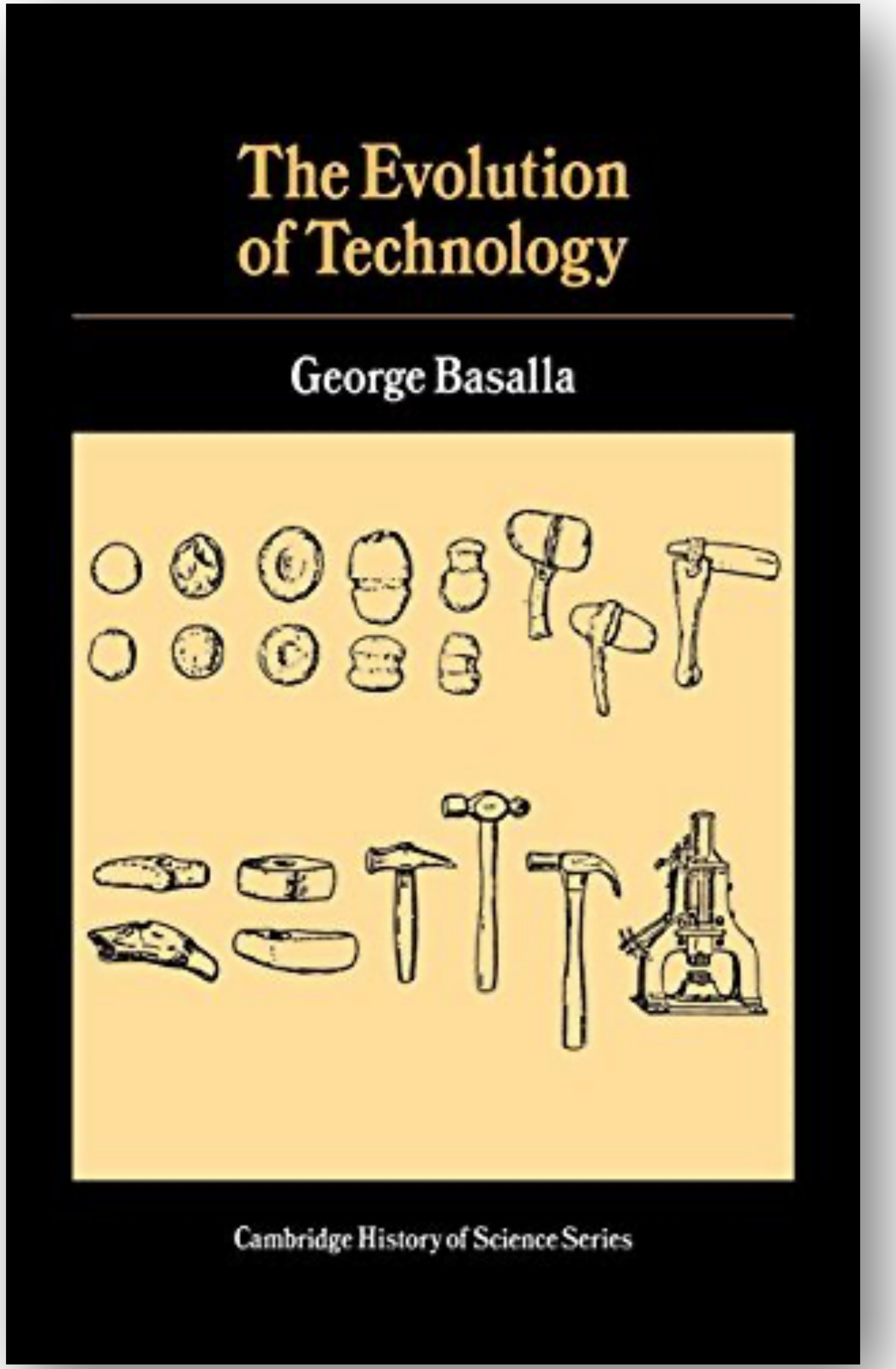
NETWORKS AND THE CITY

'Cities need to change to survive. As living beings that are constantly replacing their cells, rebuilding their veins and arteries, and pumping energy and matter or producing waste, cities are also growing and evolving as they age.' Just how complex, though, are cities? **Sergi Valverde and Ricard V Solé** of the the ICREA-Complex Systems Lab at the Universitat Pompeu Fabra in Barcelona look at how network theory and emergent dynamics might be bringing us closer to an overarching theory of urban organisation.

Sergi Valverde, Skeleton frame of a virtual skyscraper, ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Barcelona, 2013
The skeleton of a building forms a uniform grid of horizontal layers. This highly regular organisation is the fingerprint of design and conscious planning.

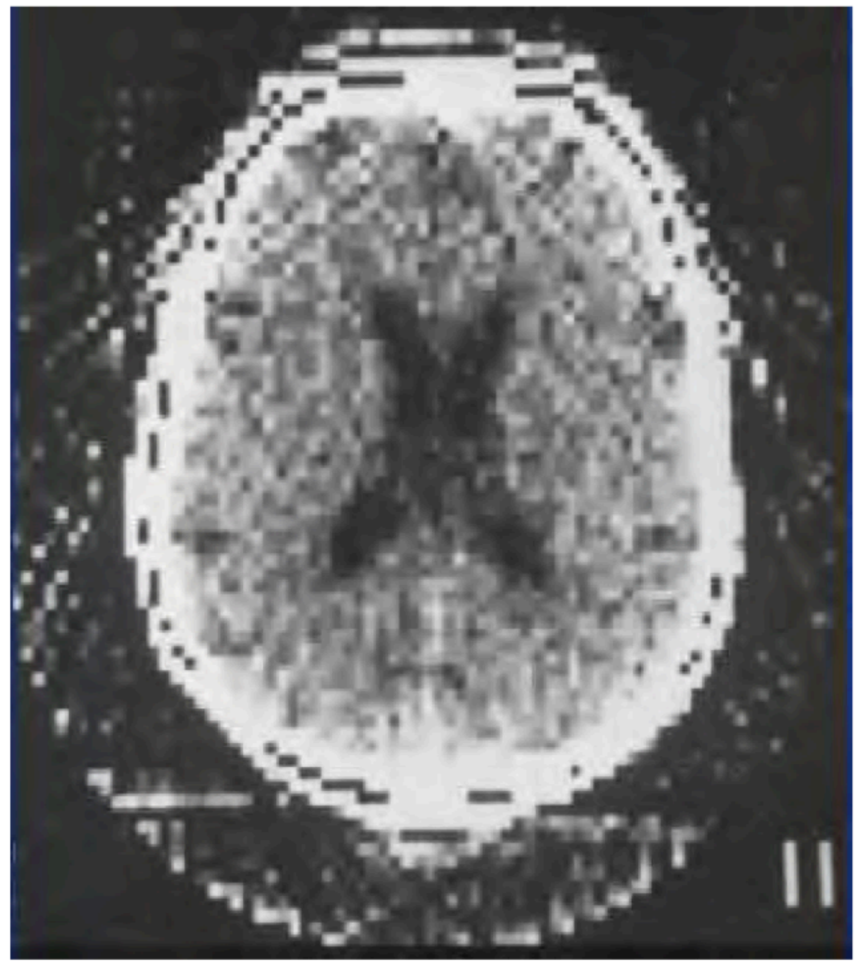
Evolution of Technology



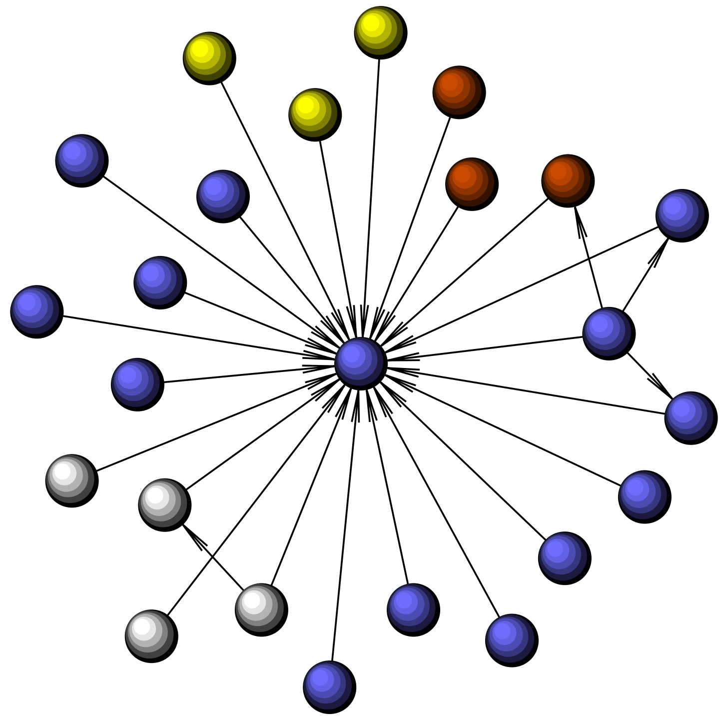


Growth: Patent Networks

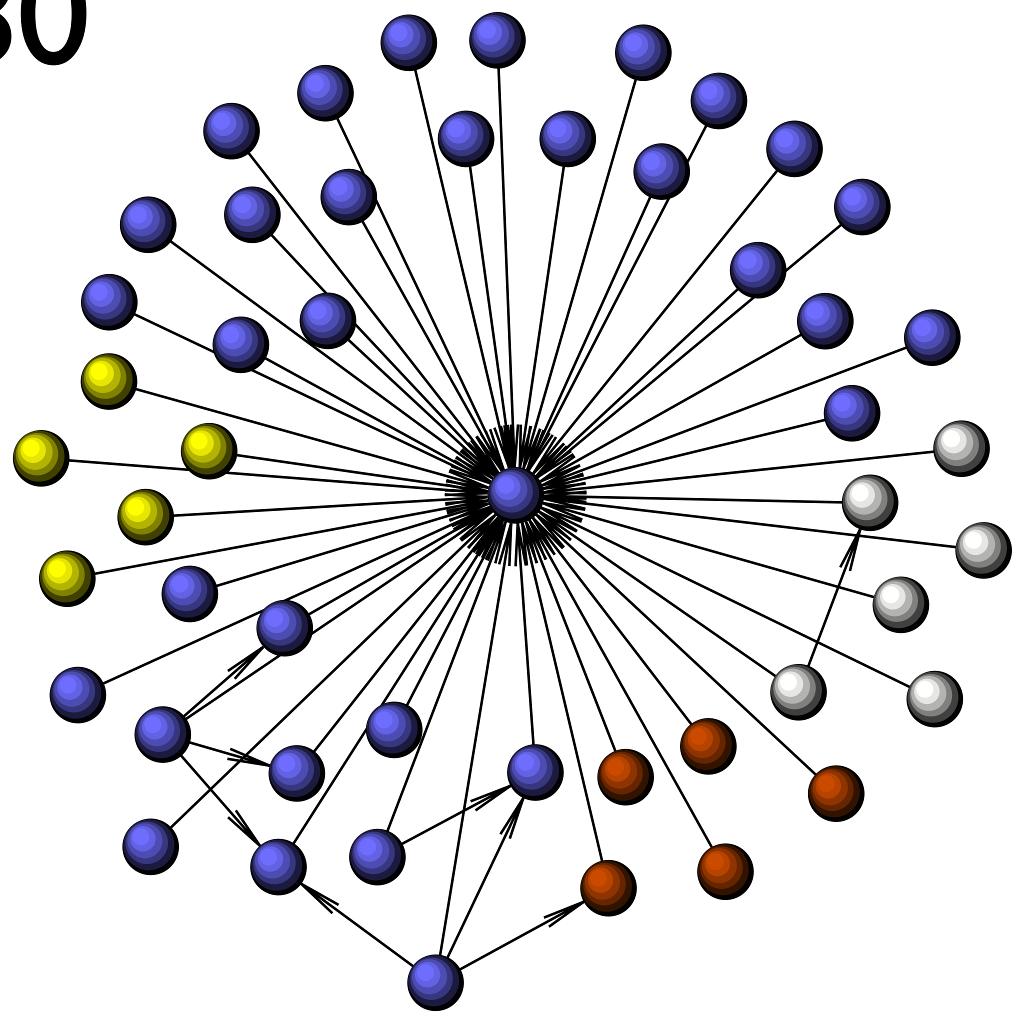
1974



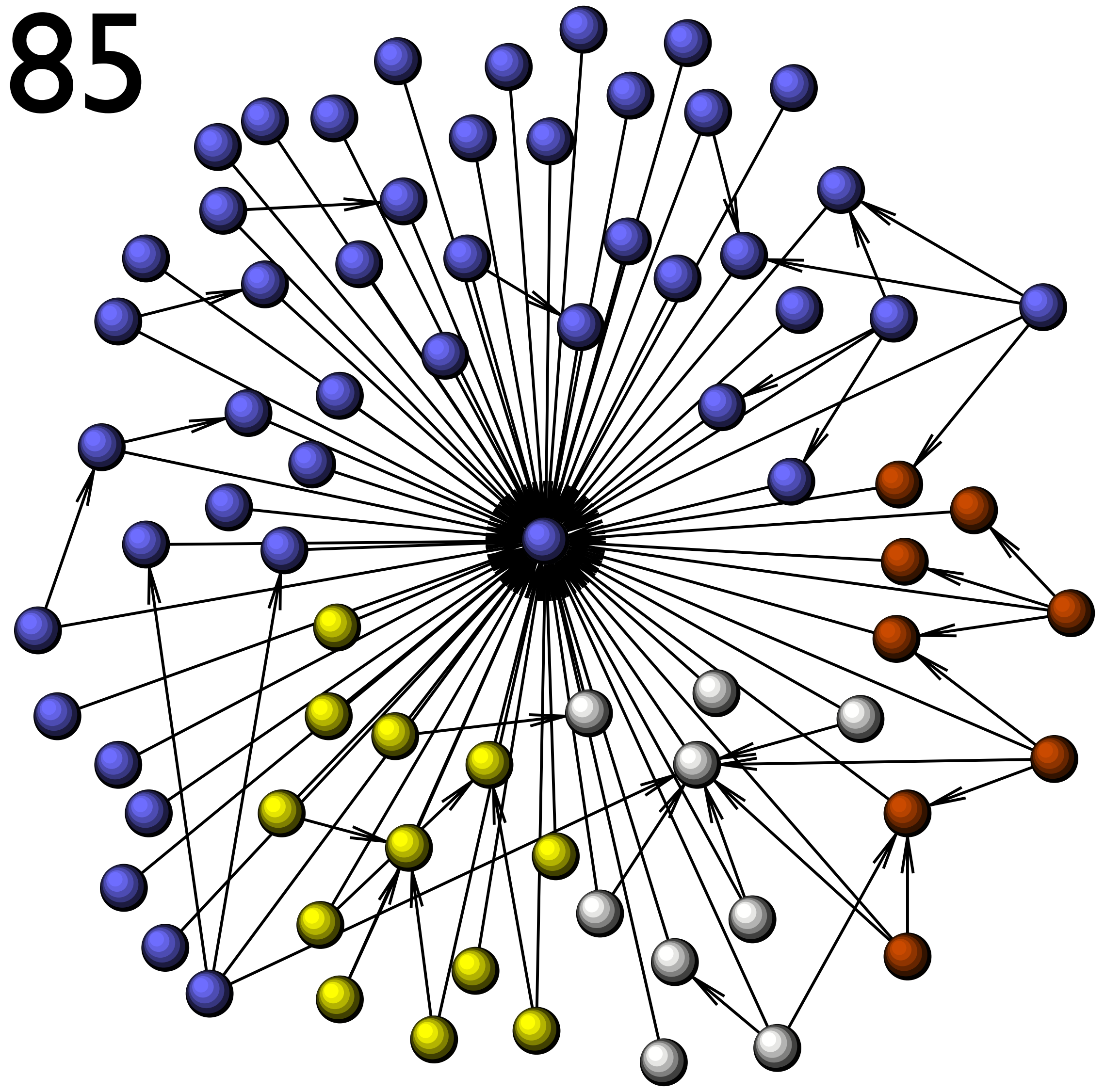
1973



1980

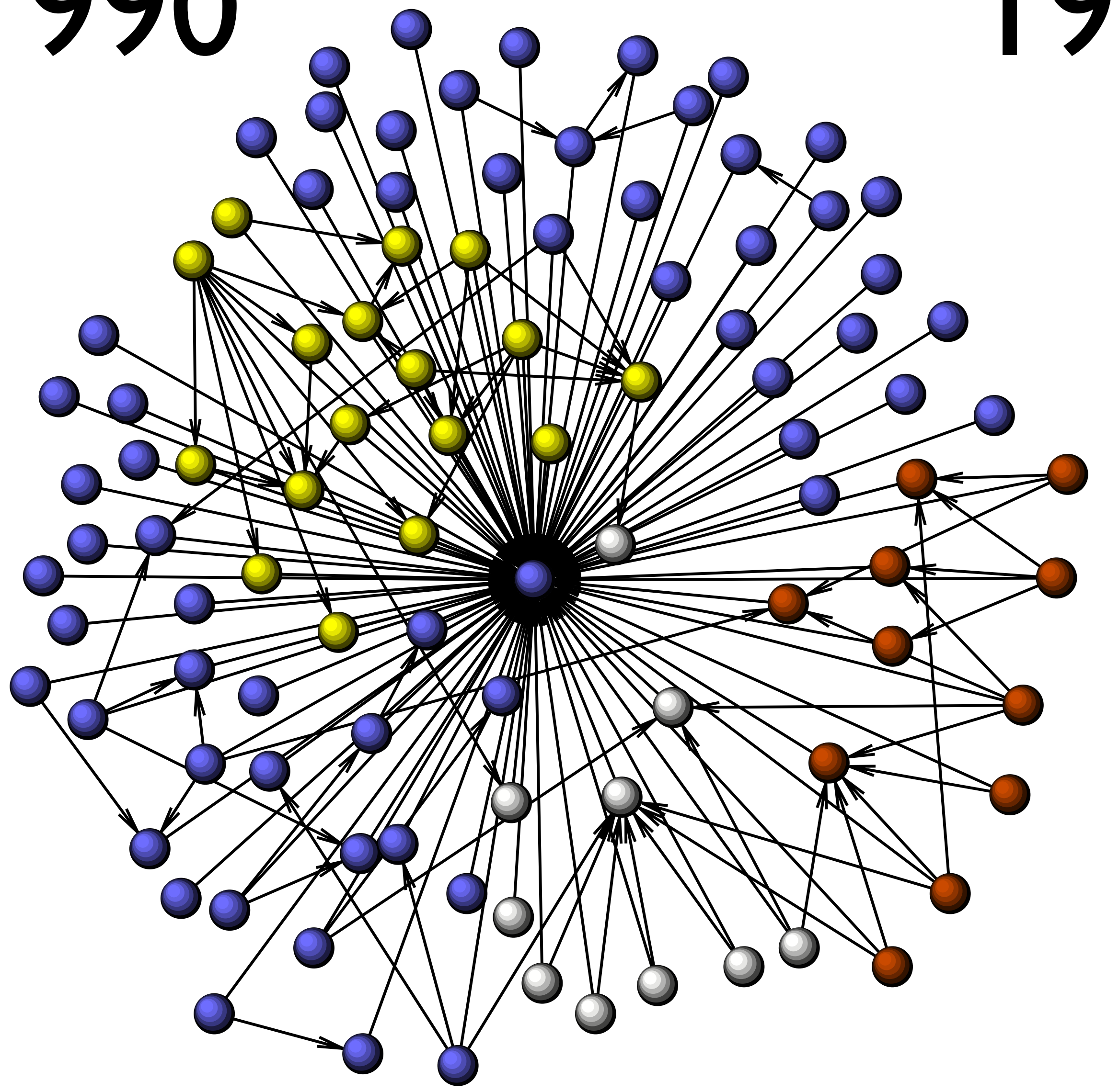


1985

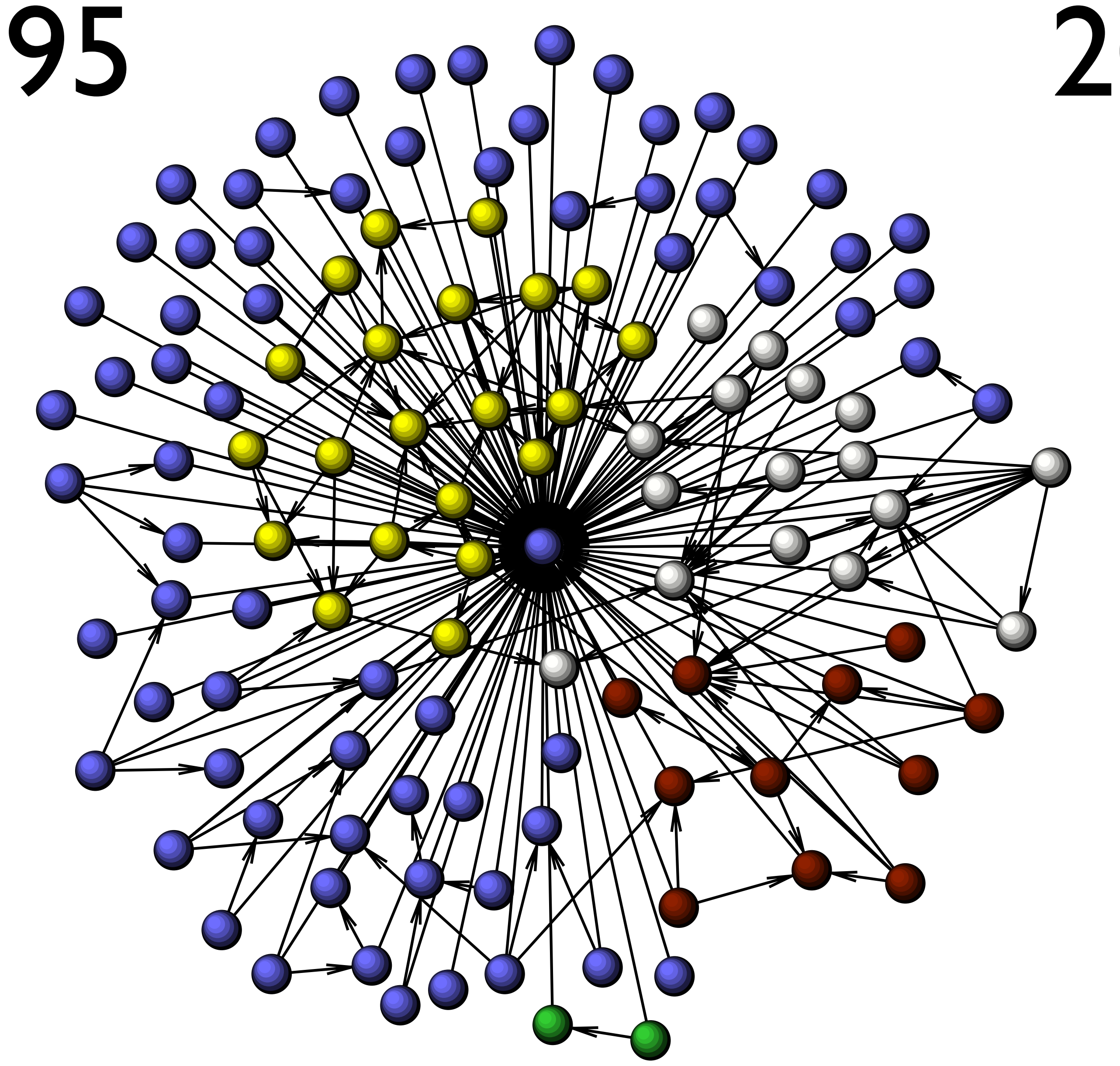


1994

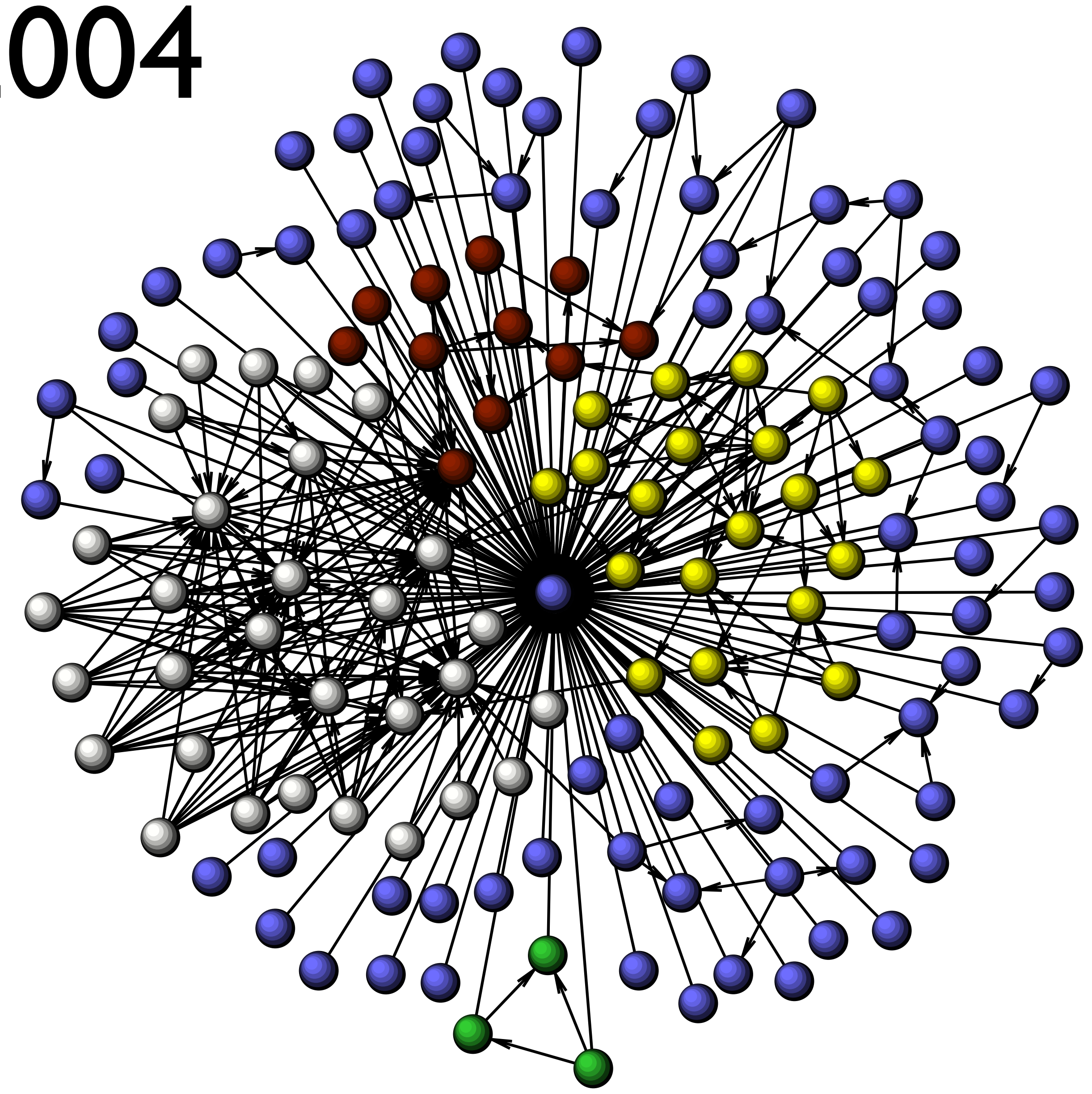
1990



1995

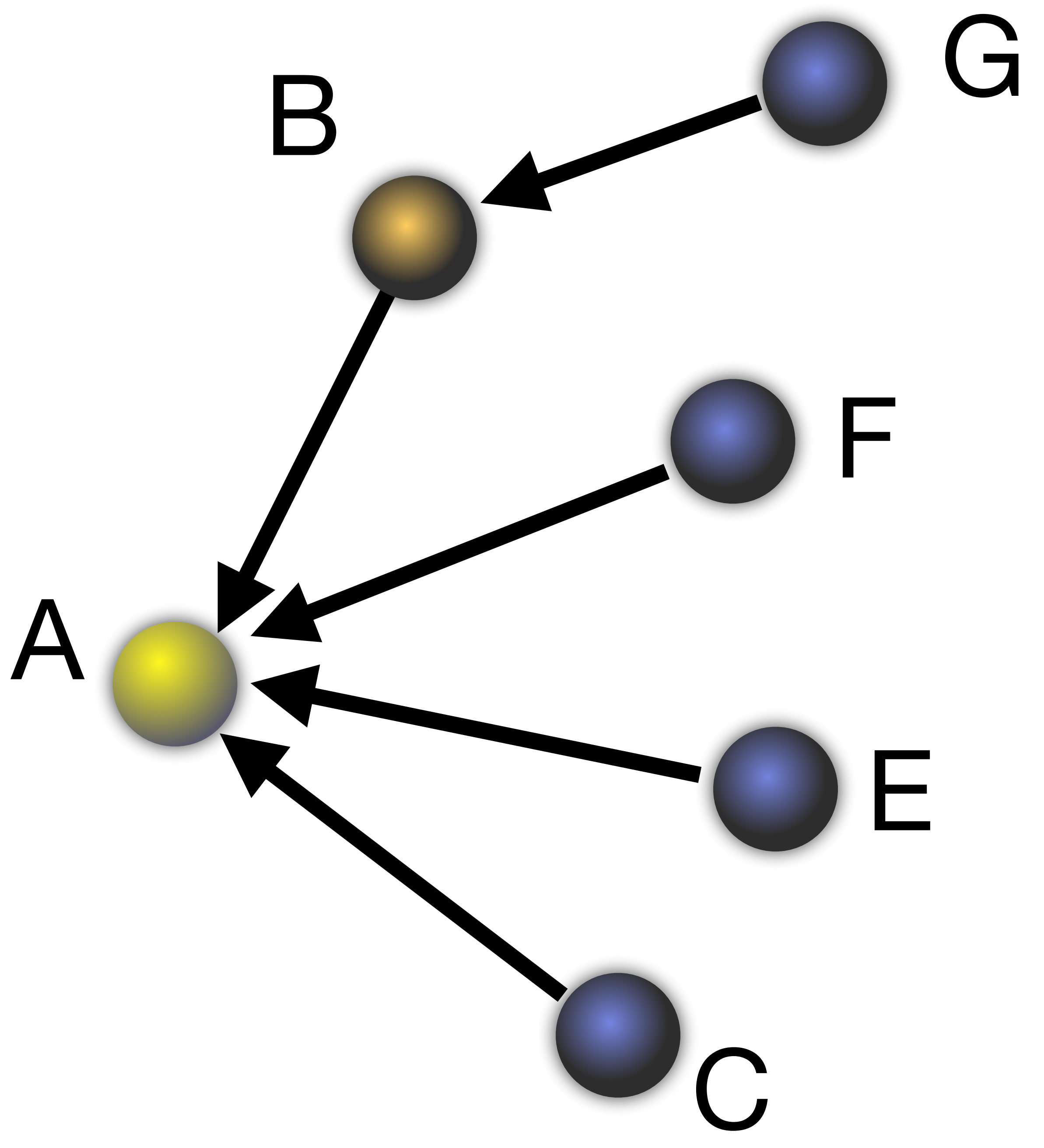


2004

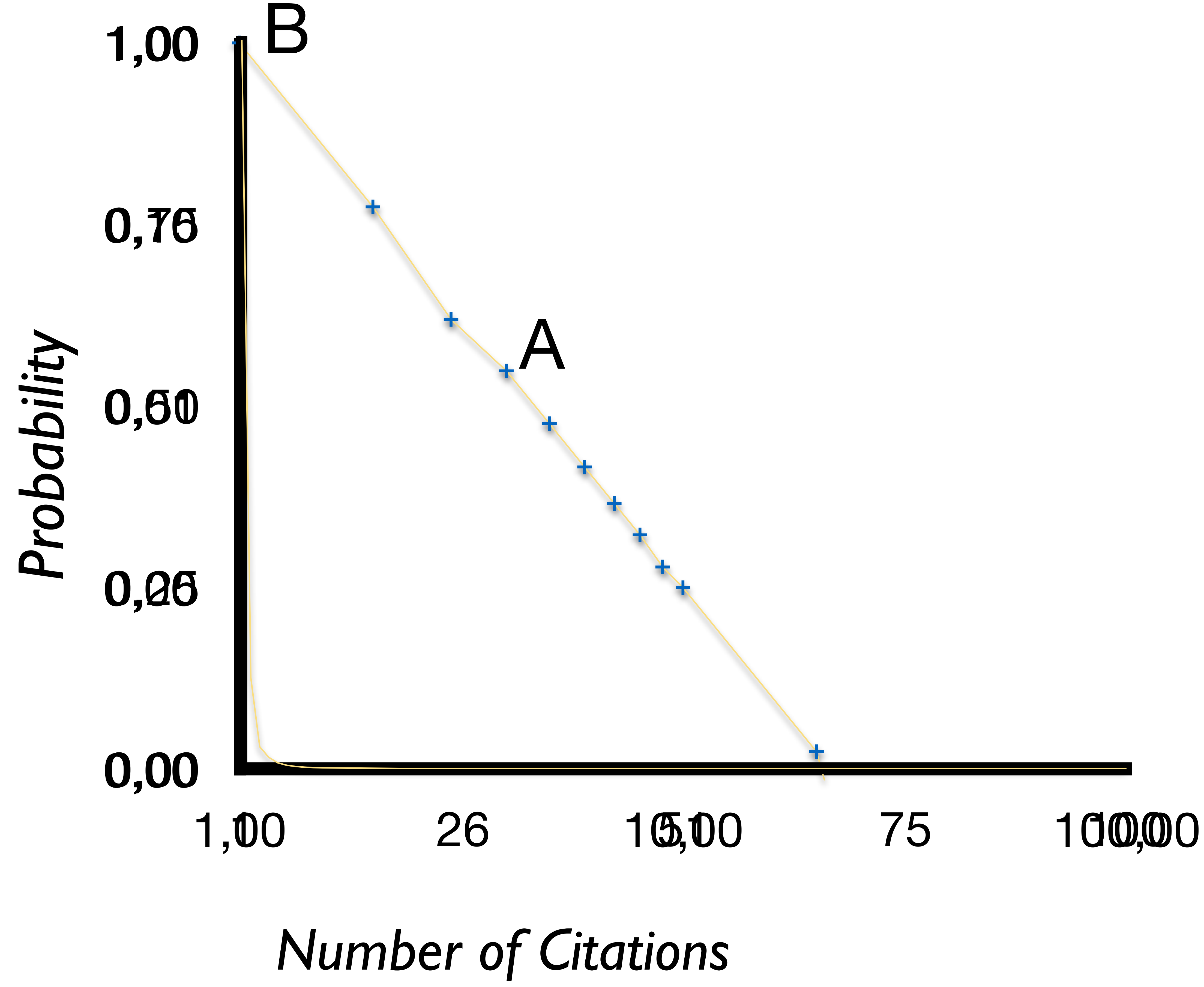


Growth: Preferential Attachment

$$\Pi(k) \sim k^{\beta} \rightarrow P(k) = Uk^{-\gamma}$$



(Price, 1965) & (Price, 1976)



Derek de Solla Price (1922-1983)

Cumulative degree distribution

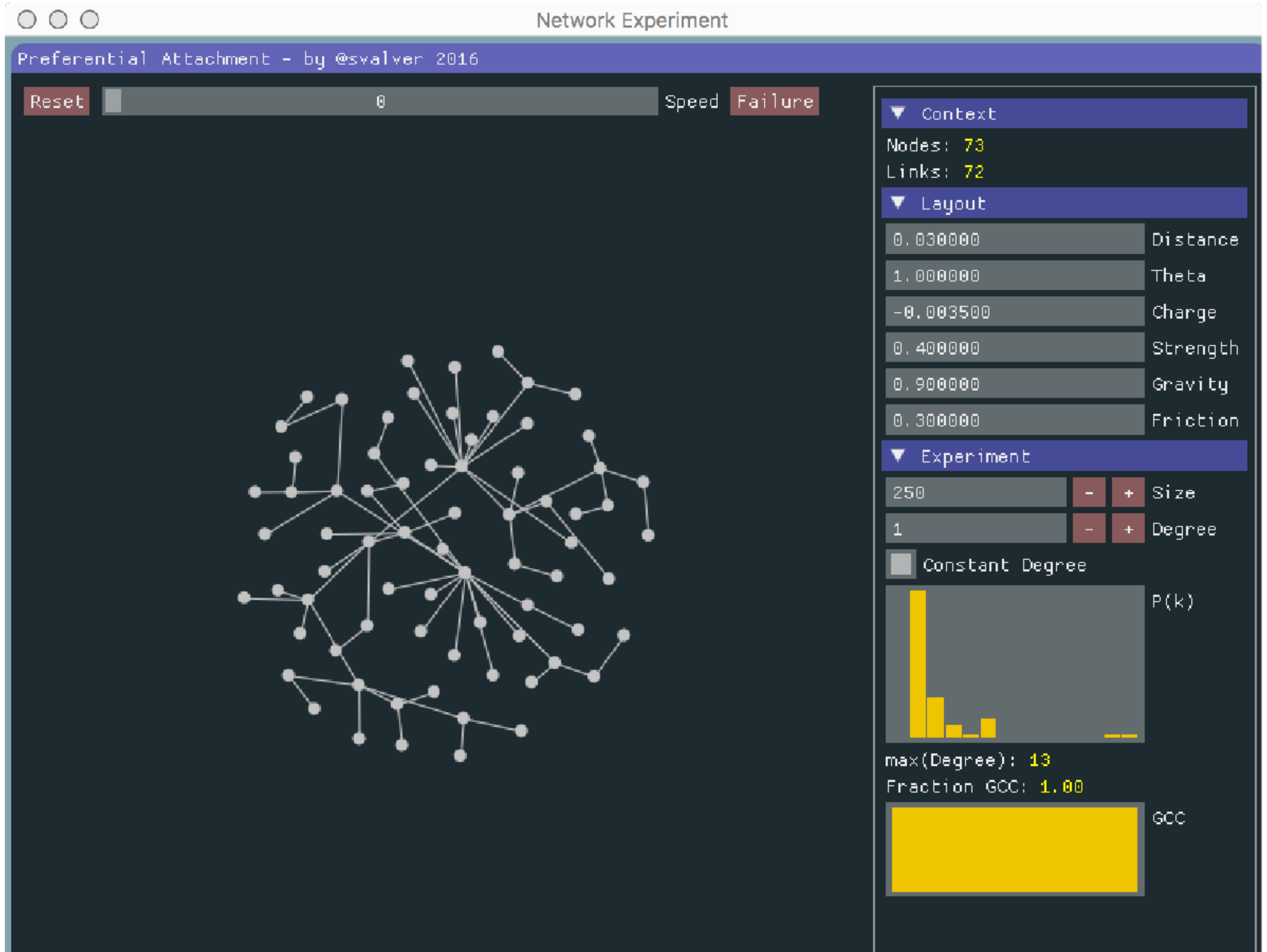
$$P_{>k} = \sum_{k'=k}^{\infty} P(k')$$

$$P_{>k} = U \sum_{j=k}^{\infty} j^{-\gamma} \approx U \int_{k}^{\infty} j^{-\gamma} dj = \frac{U}{\gamma - 1} k^{-(\gamma-1)}$$

Activity: Preferential Attachment

How history and reinforcement influence network architecture?

https://tinyurl.com/3ttchcep



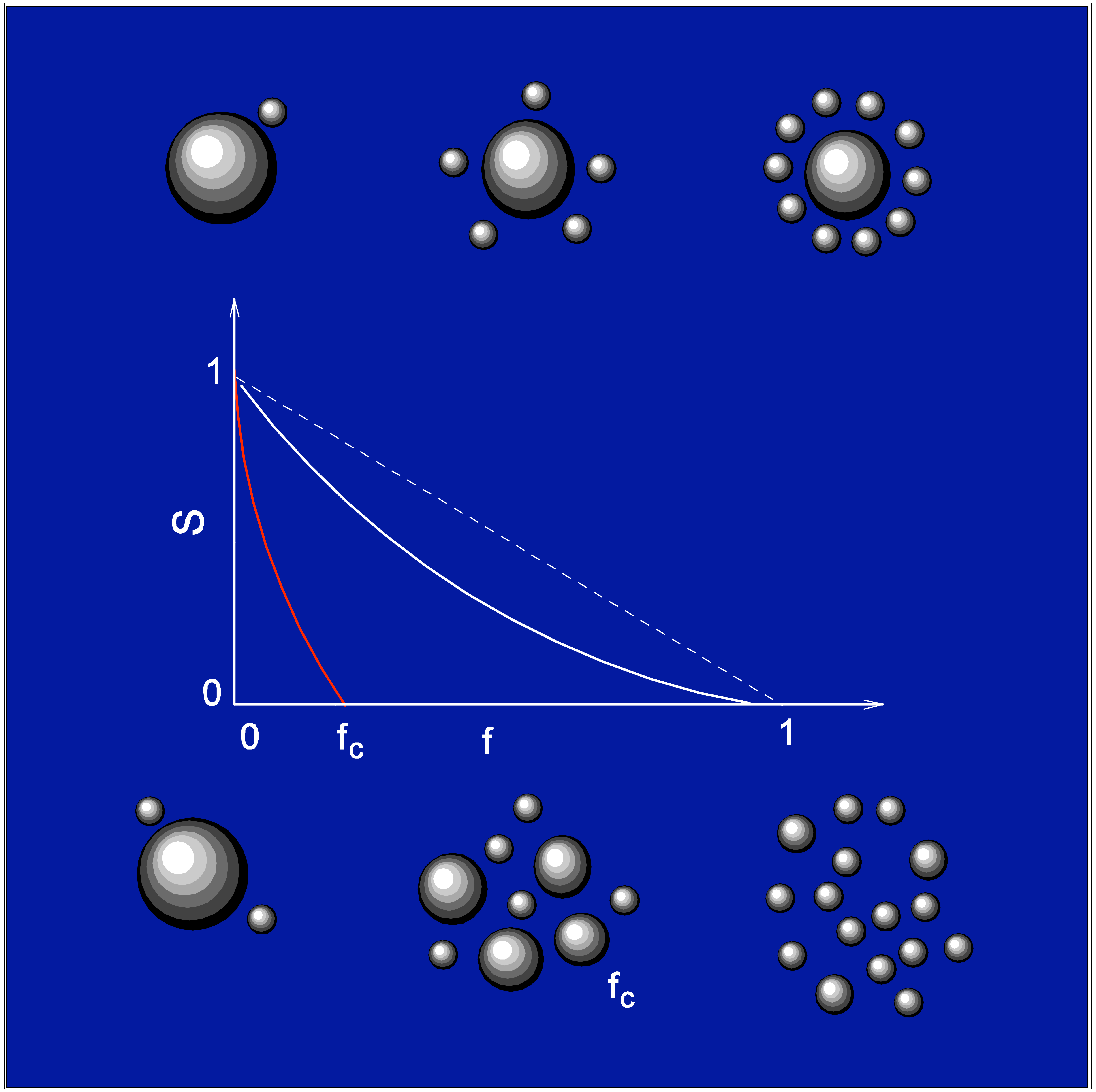
5. *How many nodes are "hubs"?*

6. *How many nodes have only a few links?*

7. *Does some low k node ever become a hub? How often?*

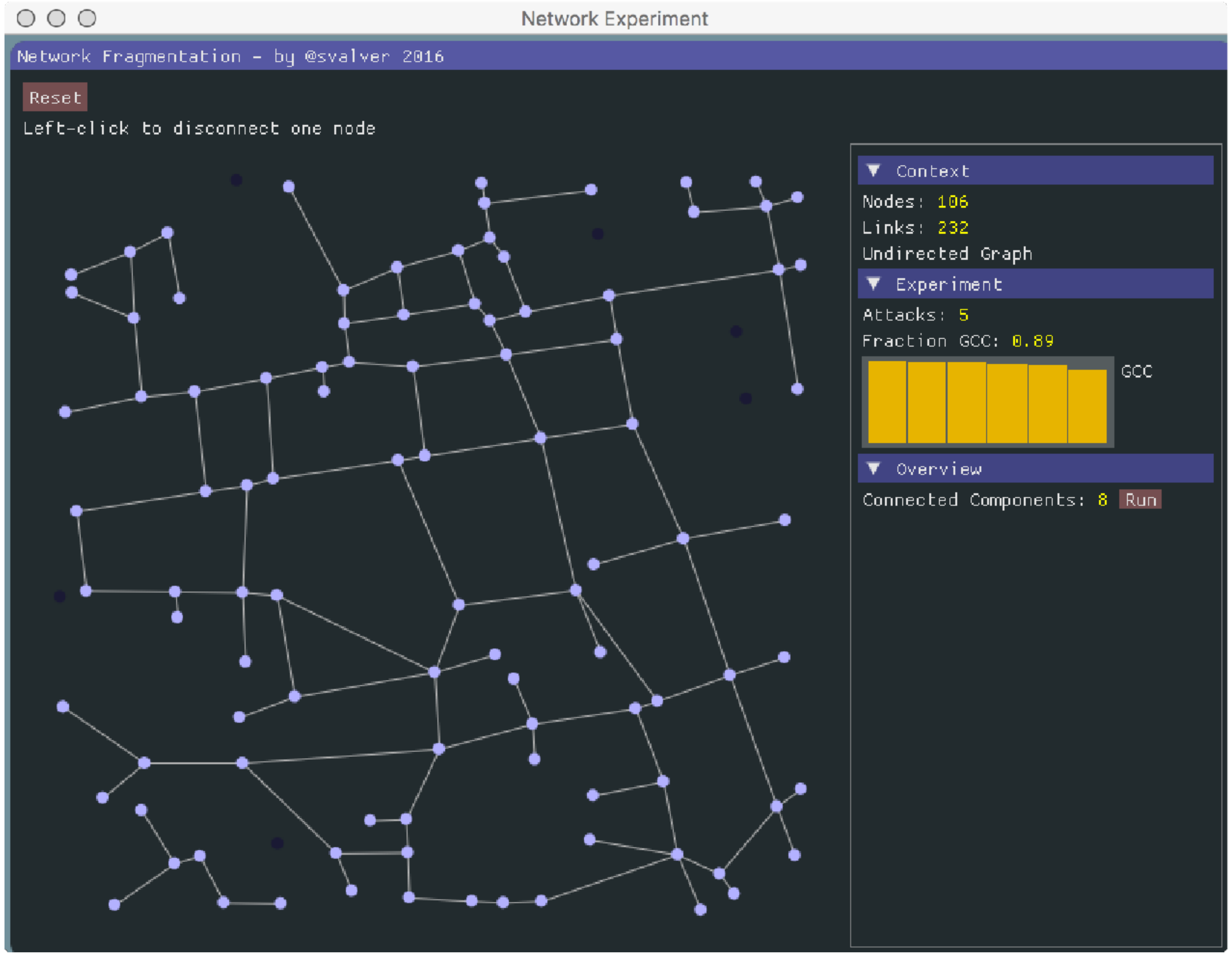
Network Robustness

"Error and attack tolerance of complex networks"
R. Albert, H. Jeong & L-A Barabási
Nature **406** (2000) 378-382

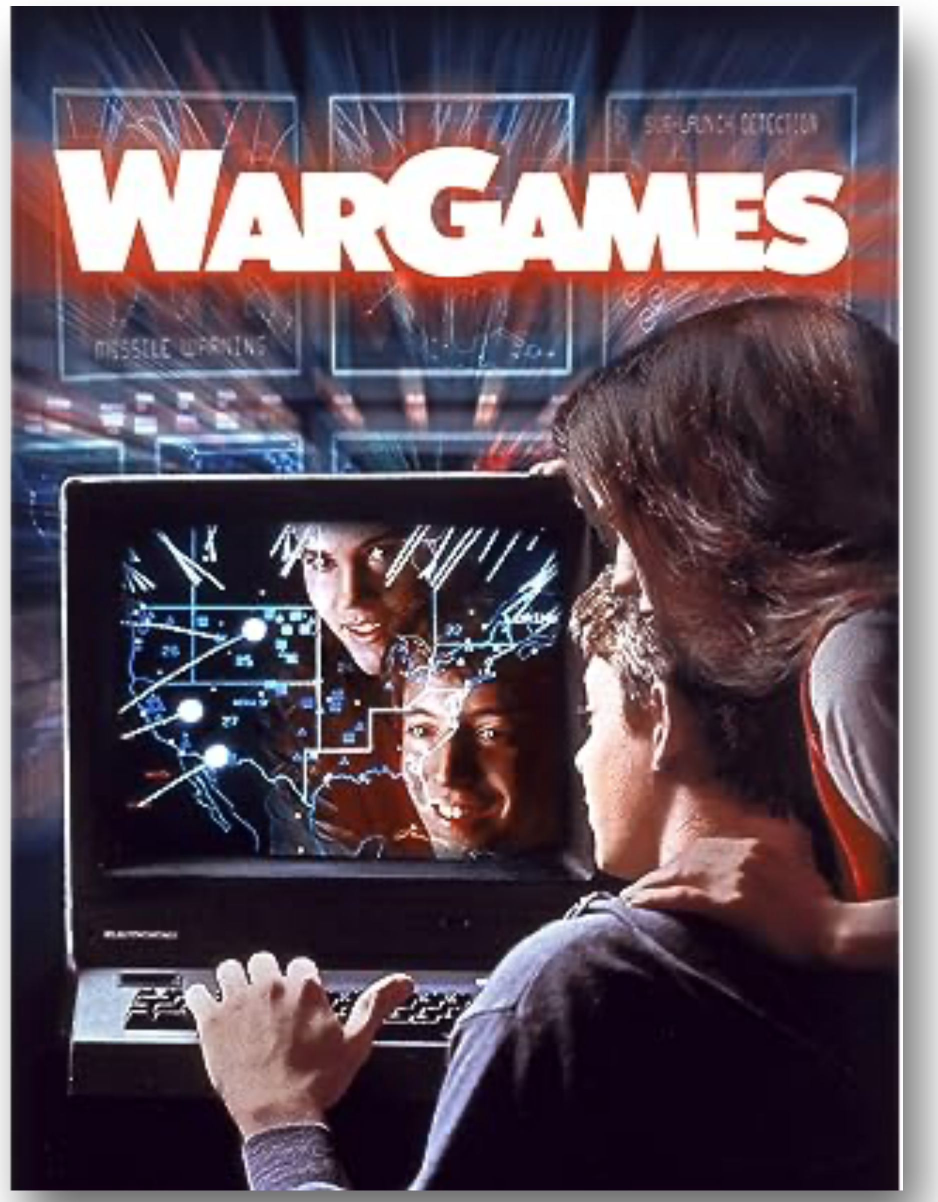


Activity: Directed Attacks

https://tinyurl.com/3jkubj8j



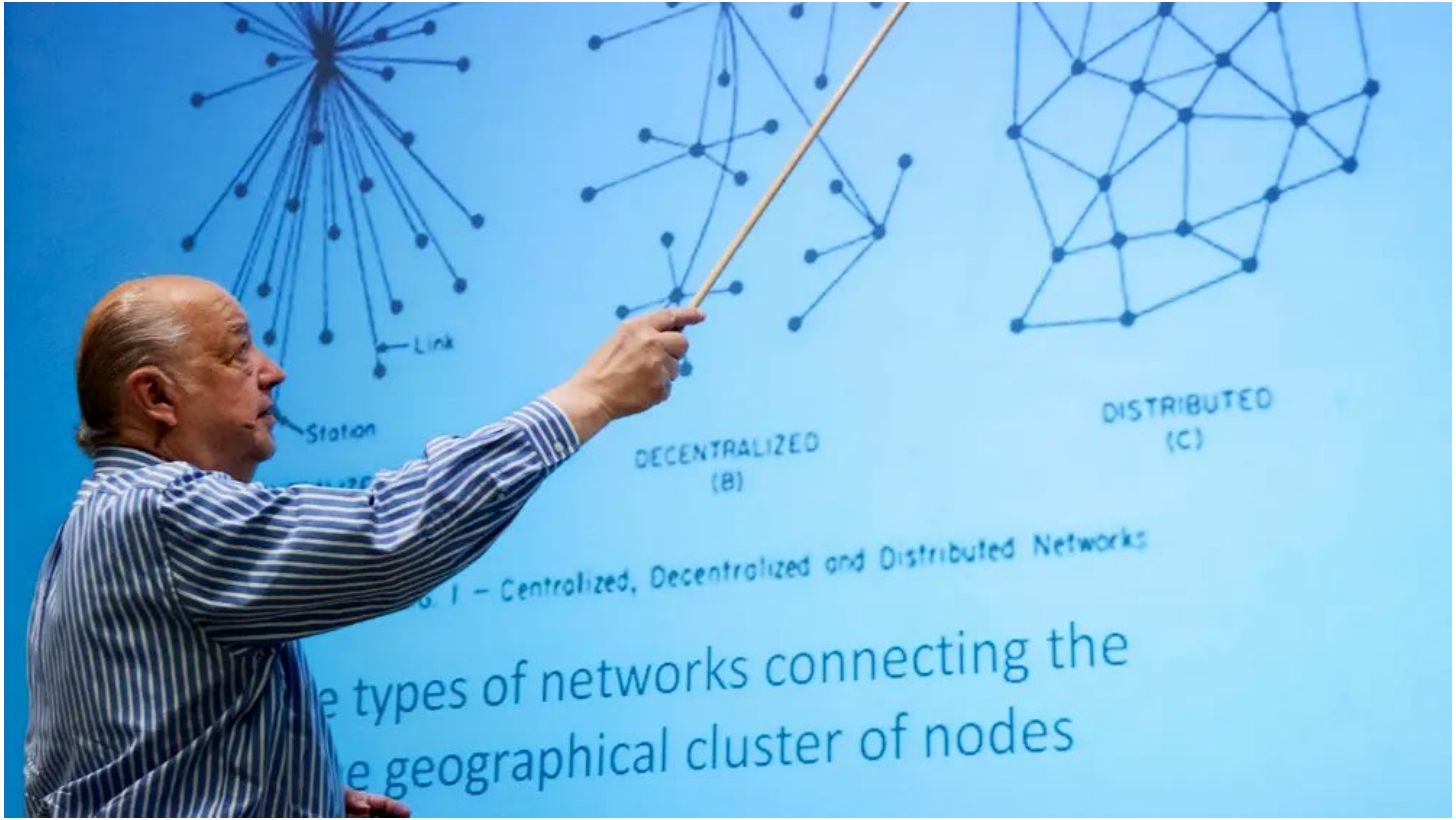
8. If you wanted to shut down the network, how many nodes would you have to take out?



9. Are collapses quick or gradual?

10. Can you predict the breaking point? Is this network fragile or robust? Why?

Origins of Internet



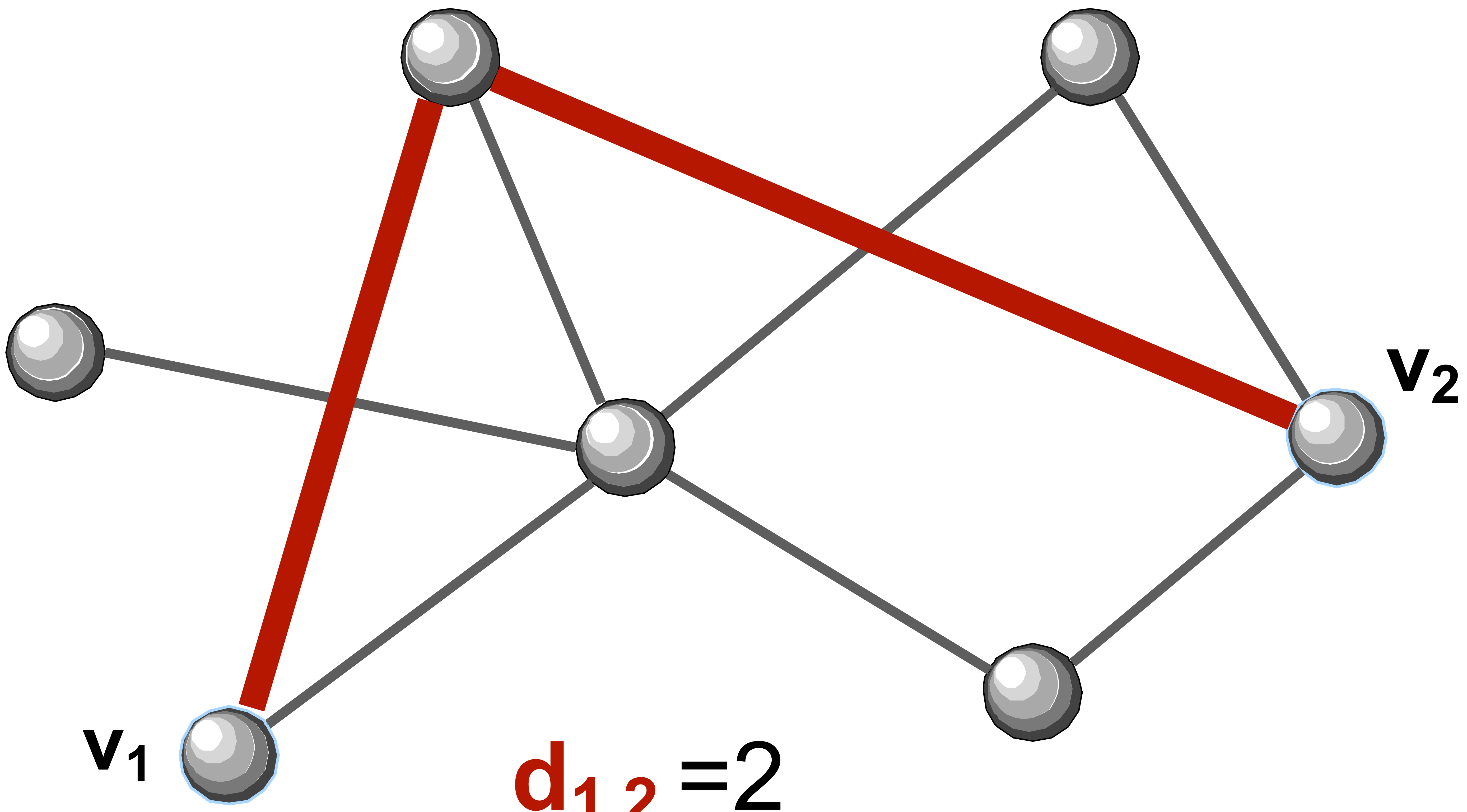
Paul Baran presents his work at a RAND Alumni Association event on July 25, 2009

Network Efficiency:
Hubs, Connectors & Paths

Definitions

- Path Length
- Power of Matrices
- Geodesic Path
- Diameter
- Components
- Global Efficiency

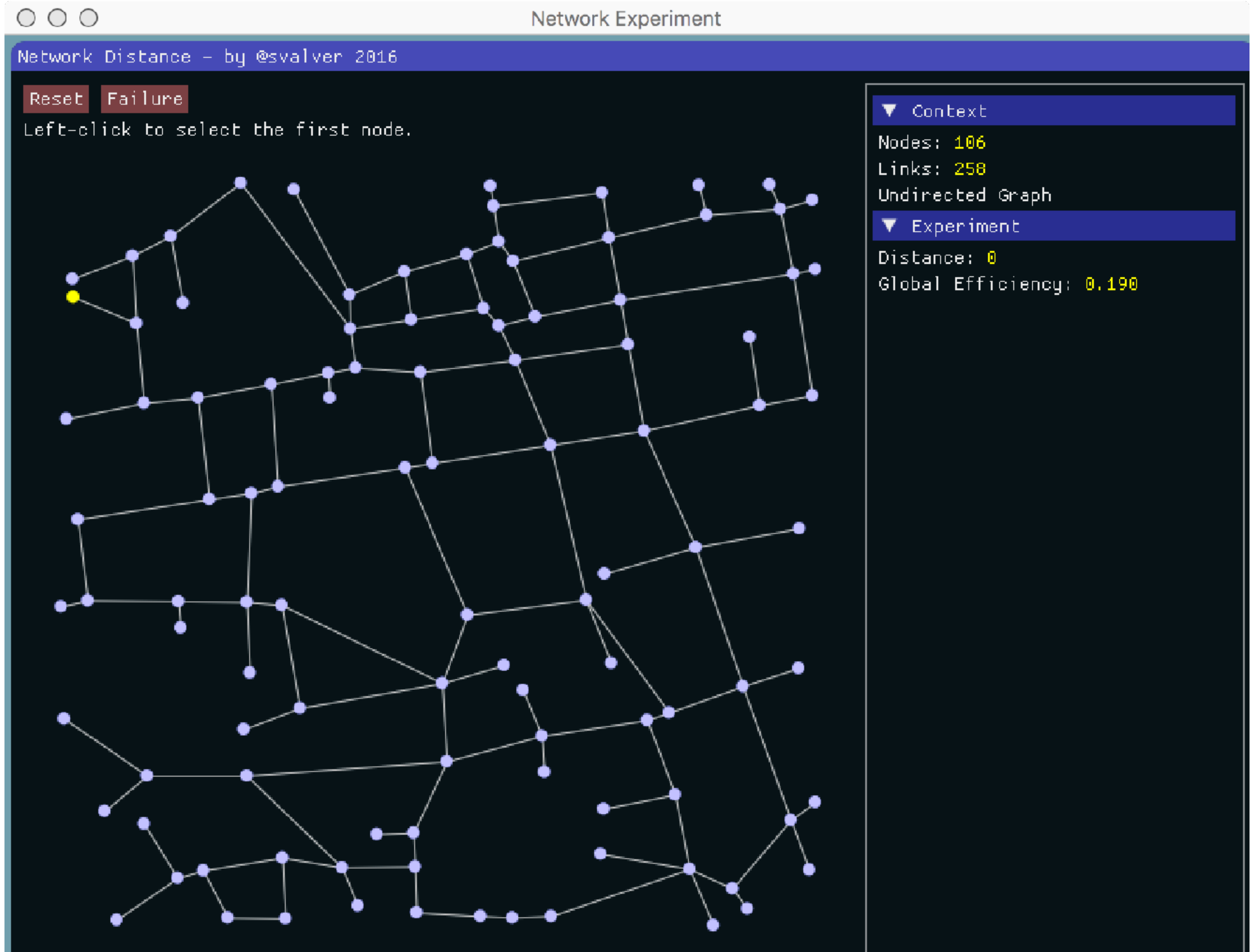
Path Length



$d_{1,2} = 2$

Activity: Shortest Paths

https://tinyurl.com/587wsvwj



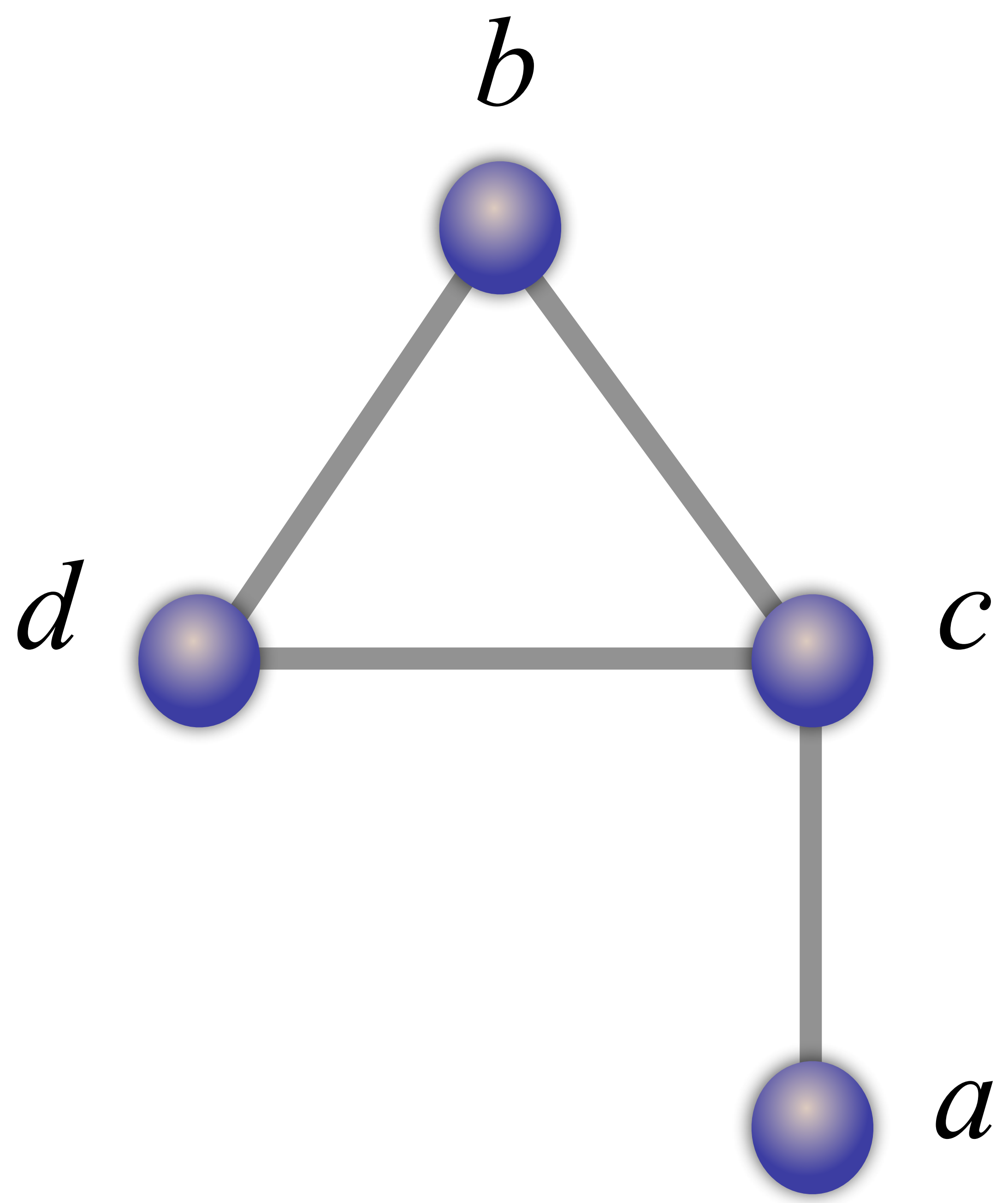
Click on a pair of nodes to see the shortest path connecting them.

Click the 'Failure' button repeatedly to remove nodes at random.

Describe the dynamical evolution of the shortest path under random failures.

Network Distance

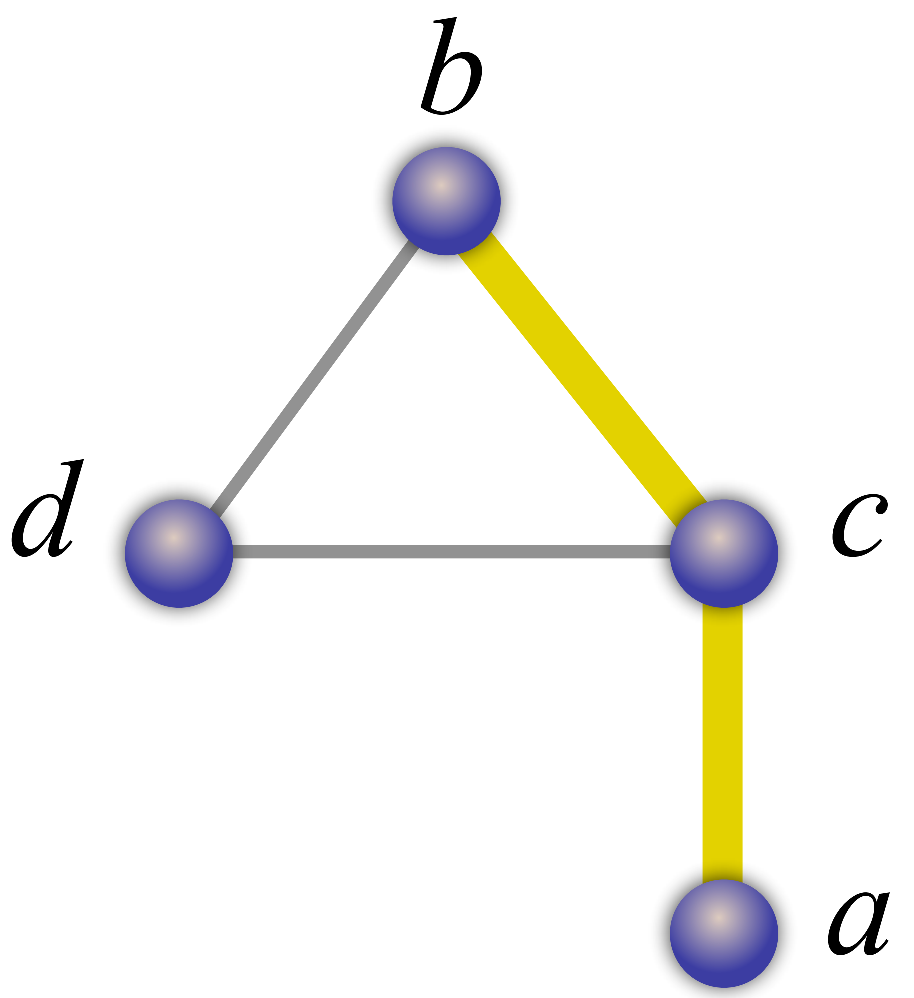
Length of a path is the number of edges traversed along a path (not the nodes).



$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix}$$

$$\begin{matrix} a & b & c & d \end{matrix}$$

Network Distance



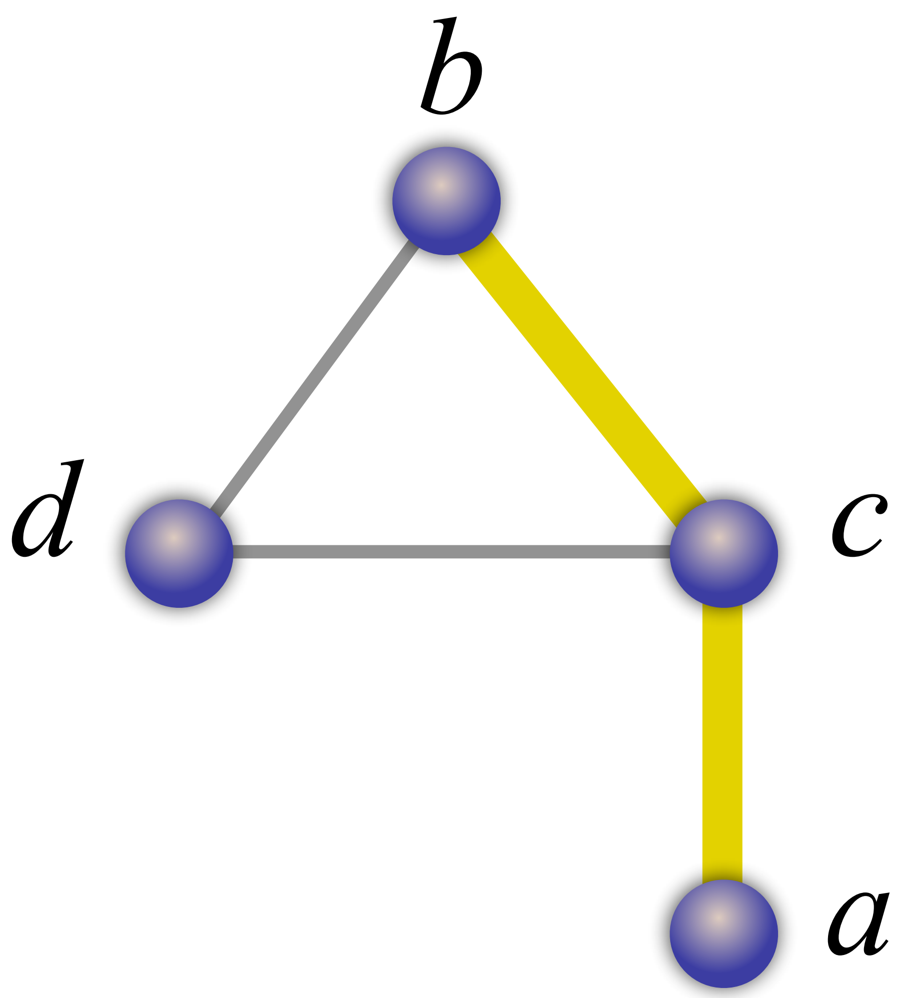
Power Matrices

$$A^2 = AA$$

$$A^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 3 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$$

(Row a of the first matrix, column b of the second matrix, and entry N_{ab}^2 of the result are highlighted.)

Network Distance



Power Matrices

$$A^2 = AA$$

$$A^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 3 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$$

(row a of the first matrix, column b of the second matrix, entry N_{ab}^2 of the result)

Network Distance

Number of paths of given length

Number of paths of length 2: $$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

Number of paths of length 3: $$N_{ij}^{(3)} = \sum_{k=1}^{N}\sum_{l=1}^{N} A_{ik}A_{kl}A_{lj} = [A^3]_{ij}$$

Number of paths of length *r* : $$N_{ij}^{(r)} = [A^r]_{ij}$$

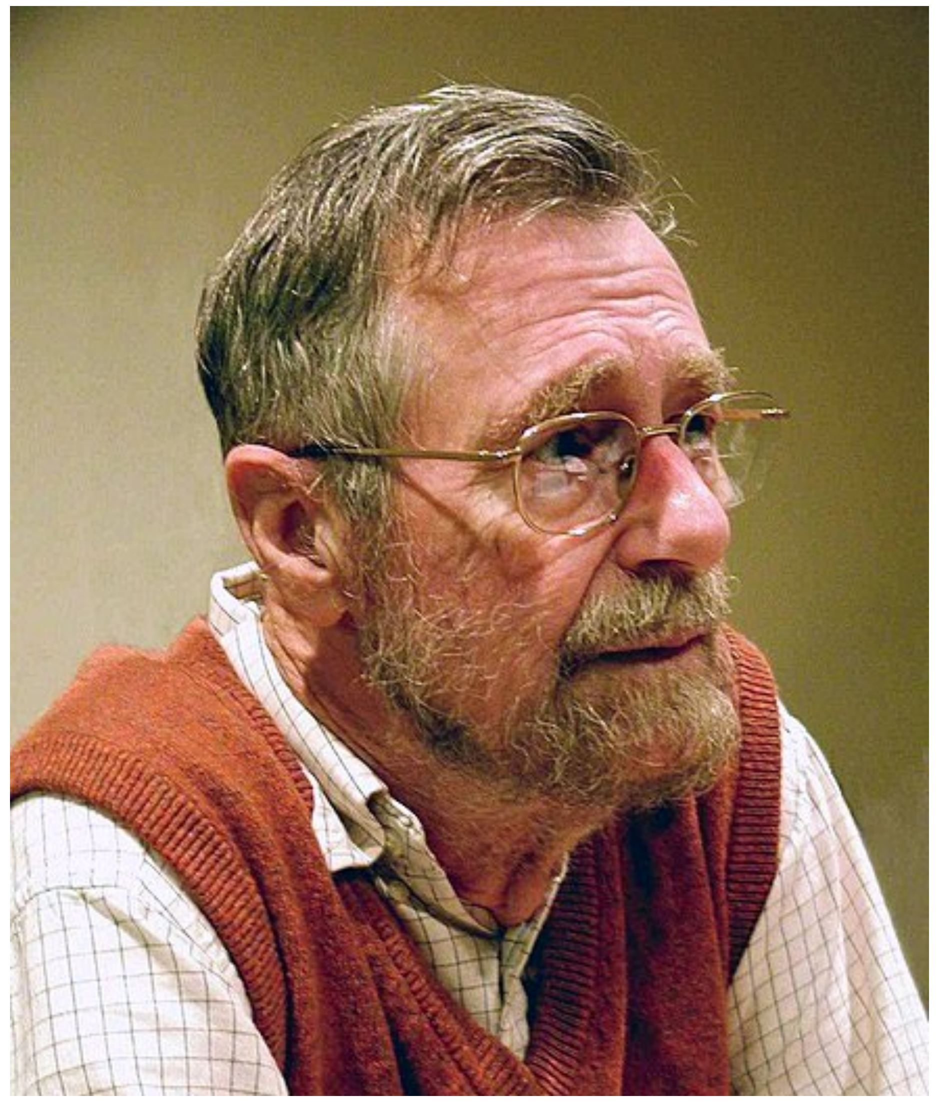
Network Distance

A geodesic path (or **shortest path**) is a path through a network between two vertices such that no shortest path exists.

The **shortest path distance** is the length of the shortest path, i.e., the smallest value of ***r*** such that:

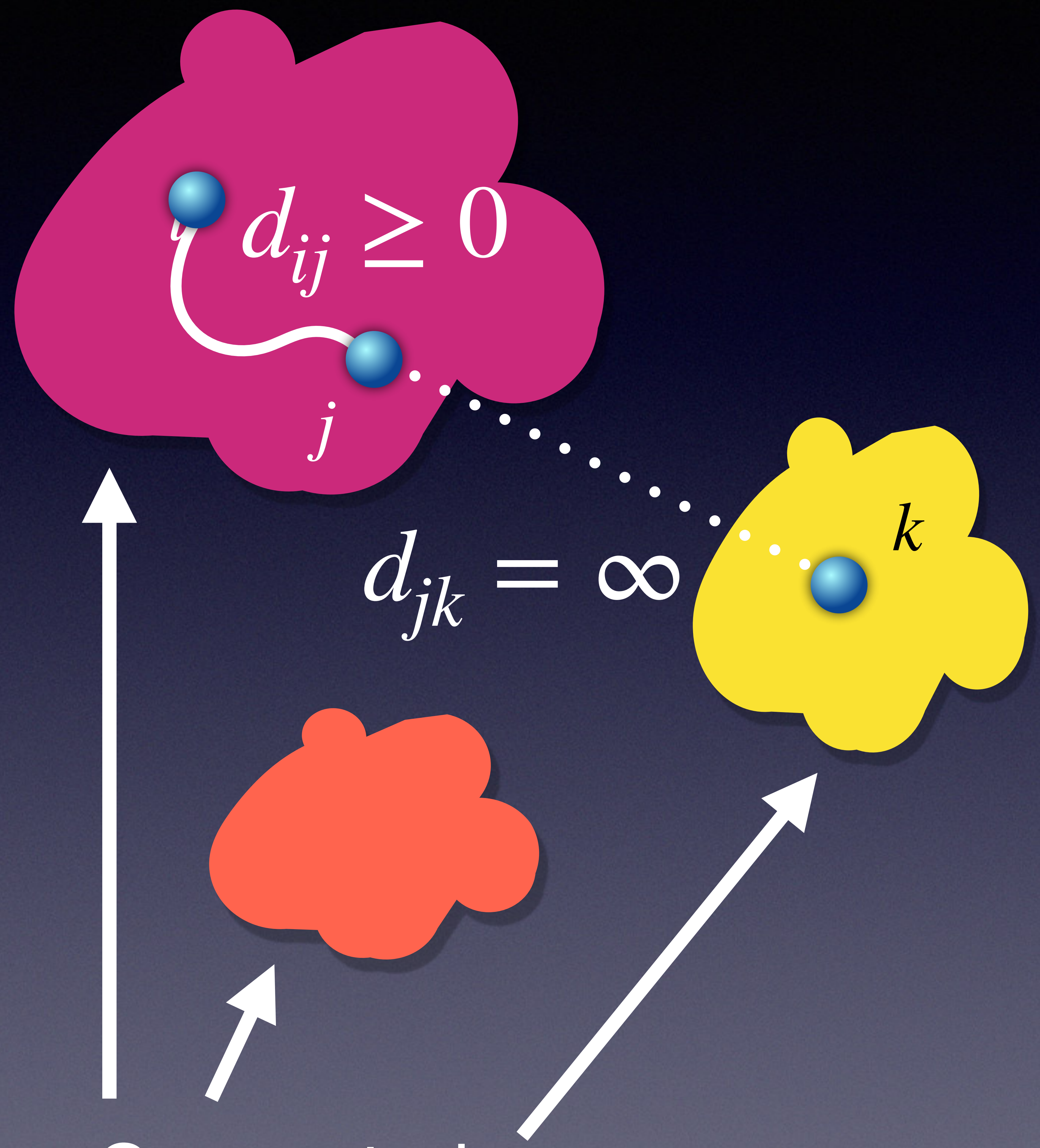
$$[A^r]_{ij} > 0$$

In practice, there are more efficient ways of calculating shortest distances in a graph (e.g., **Dijkstra's Algorithm**).

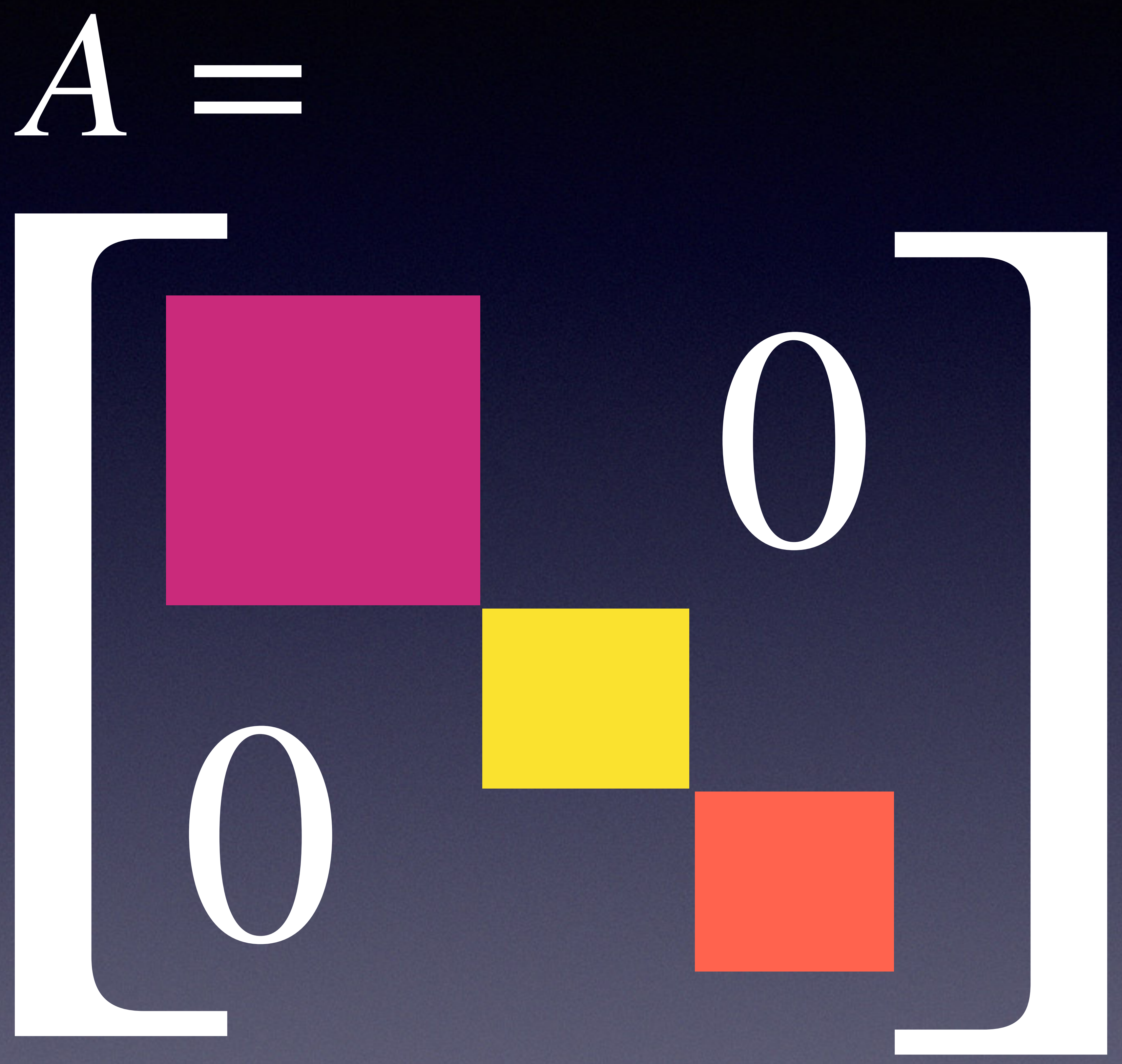


Edsger W. Dijkstra
(1930-2002)
Turing Award (1972)

Network Distance



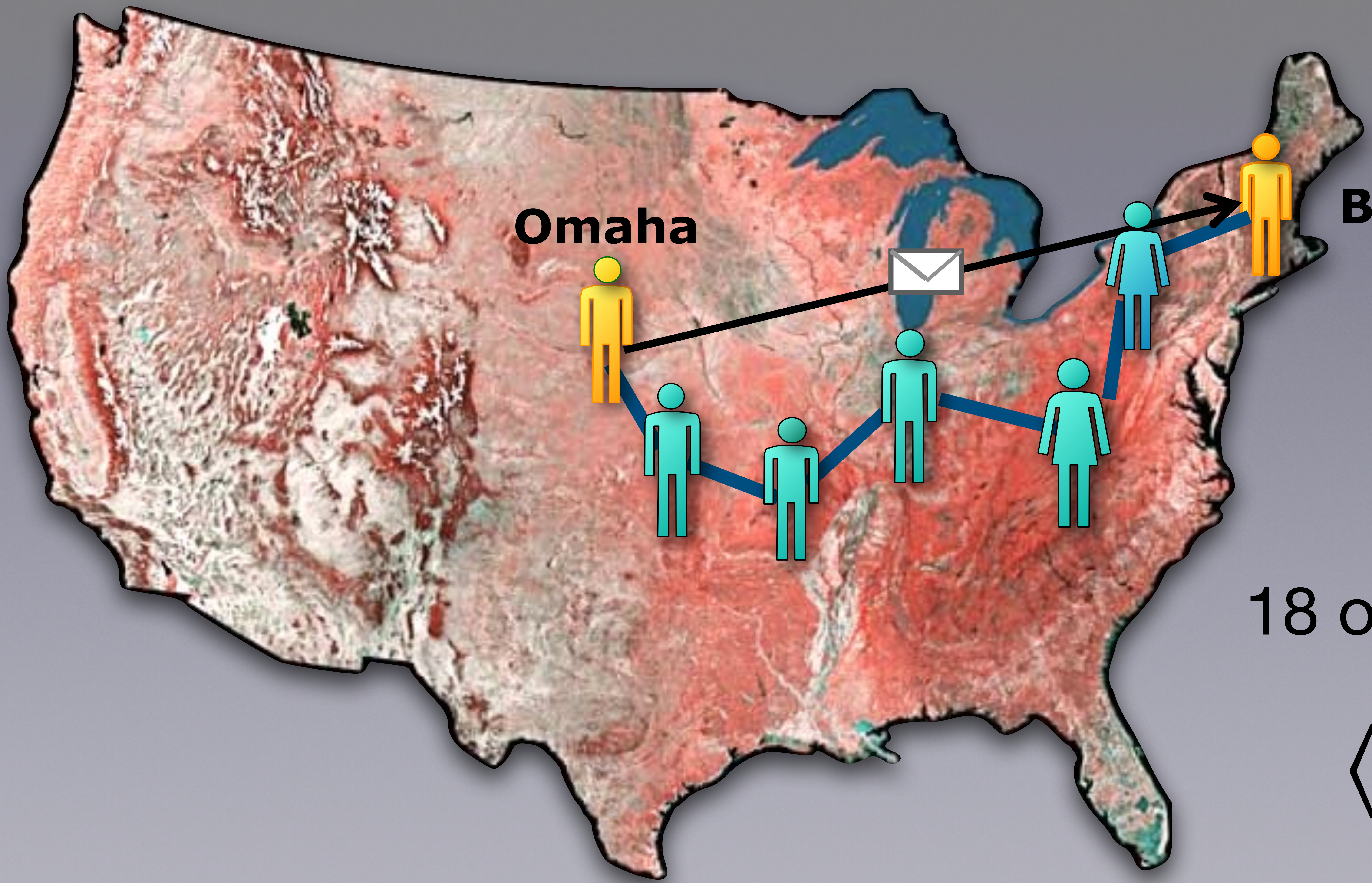
Connected
Components

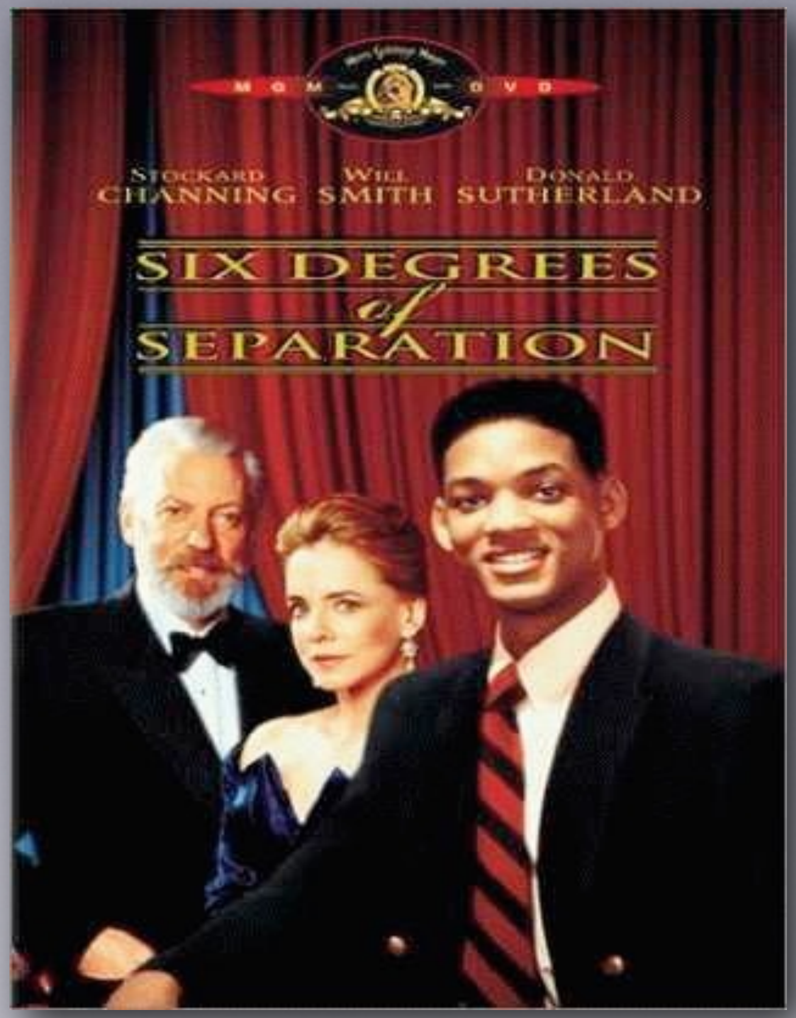


Block diagonal form

Network Distance

Is your Network Large or Small?





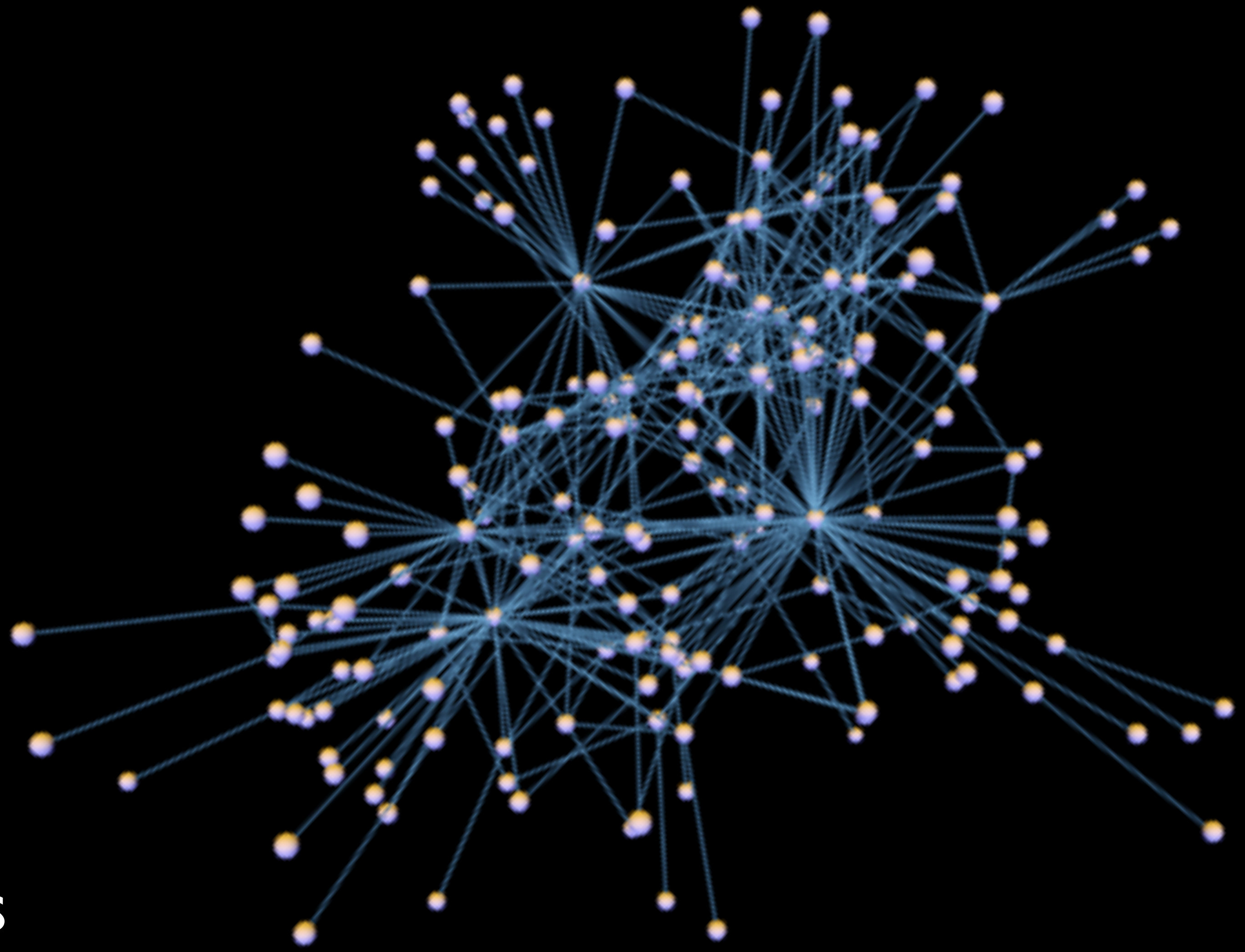
18 out 96 received

$$\langle d \rangle = 5.9$$

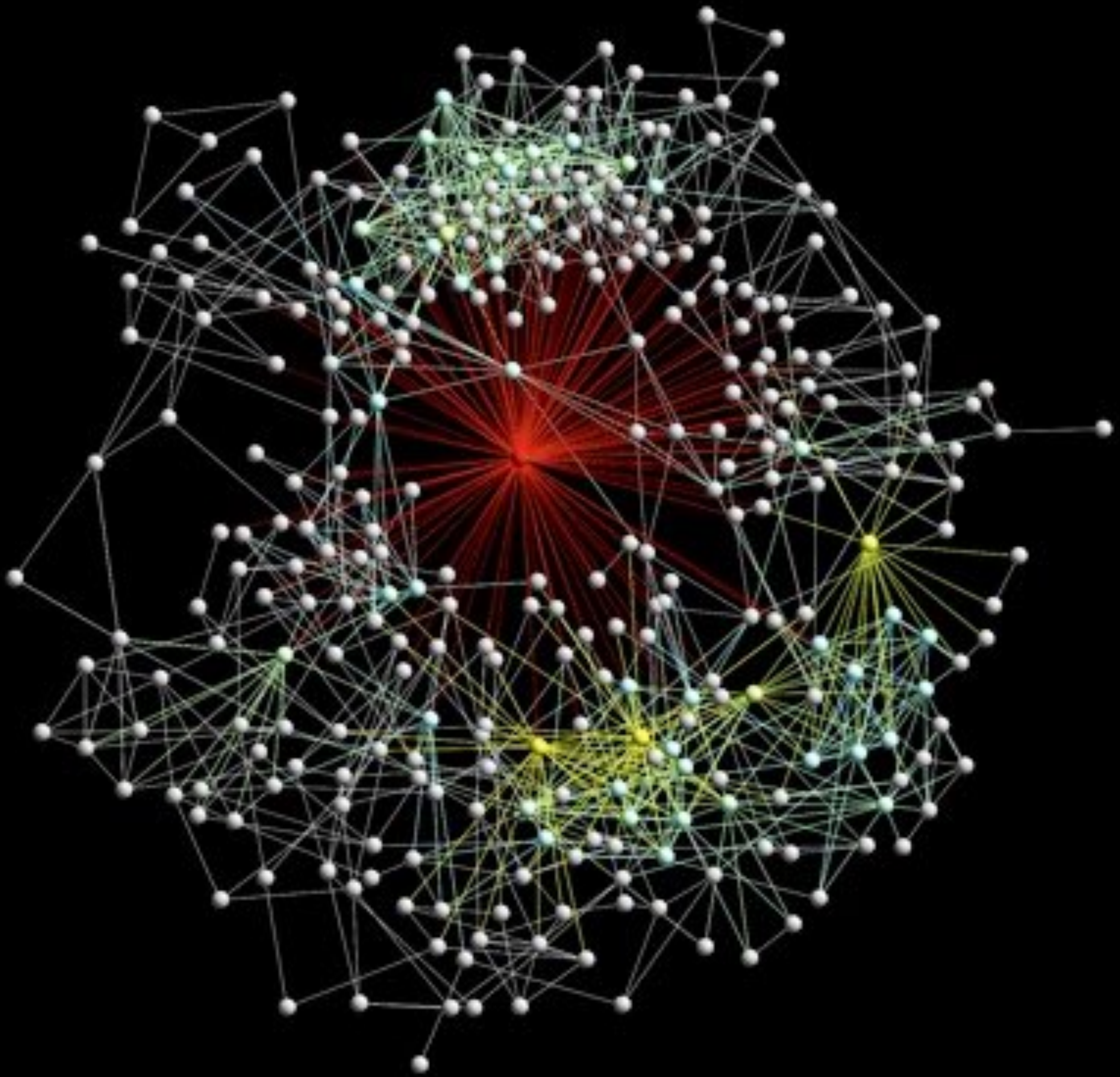
Stanley Milgram (1967)

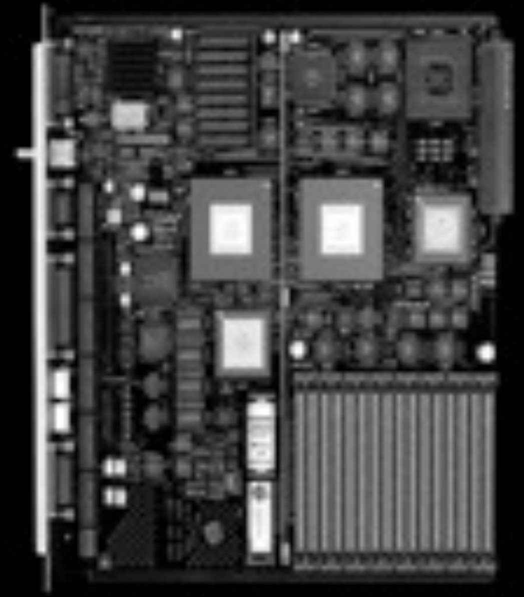
Between Order and Randomness

Why Many Networks are Small and yet Clustered?

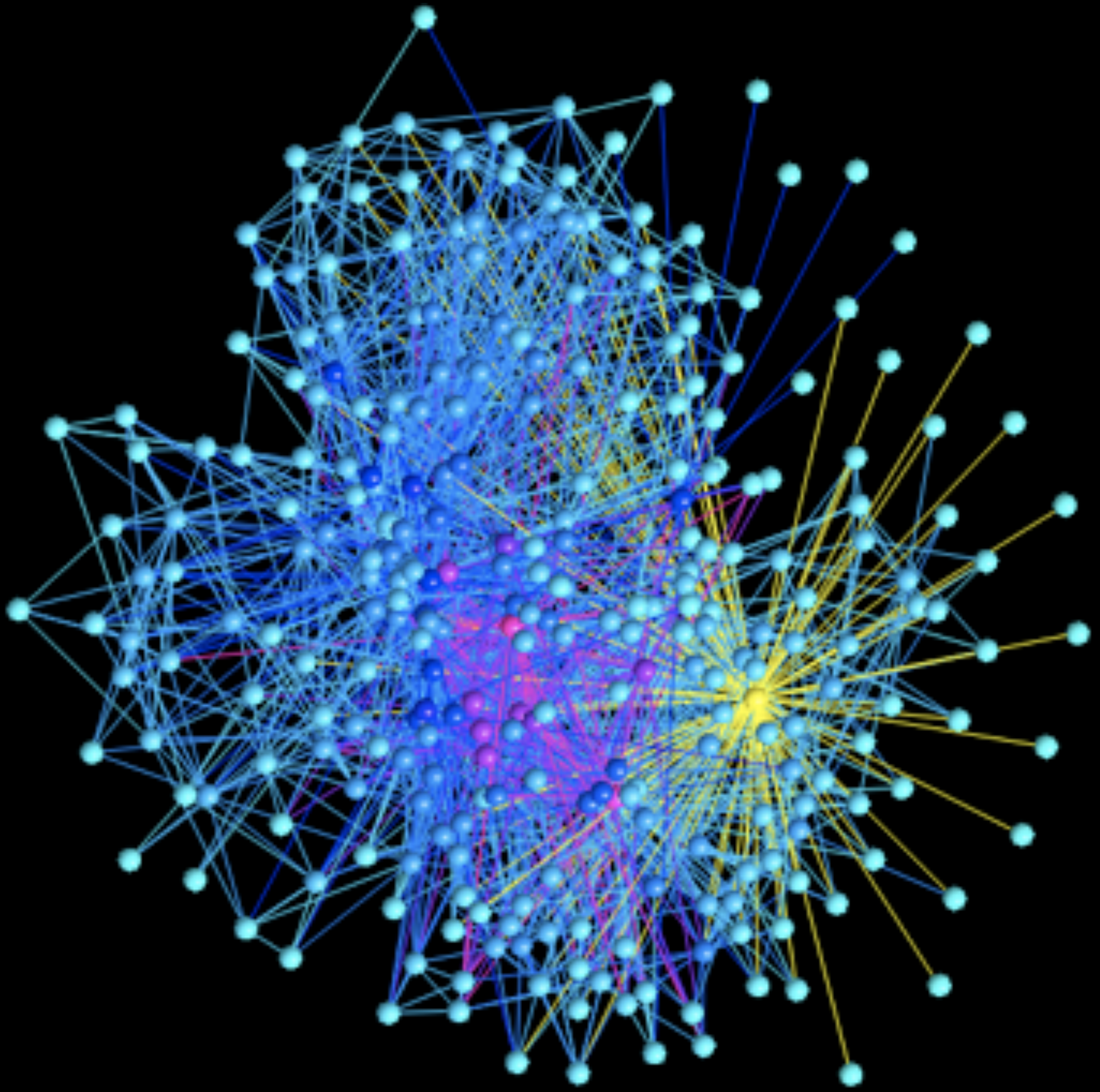


Linguistic Networks





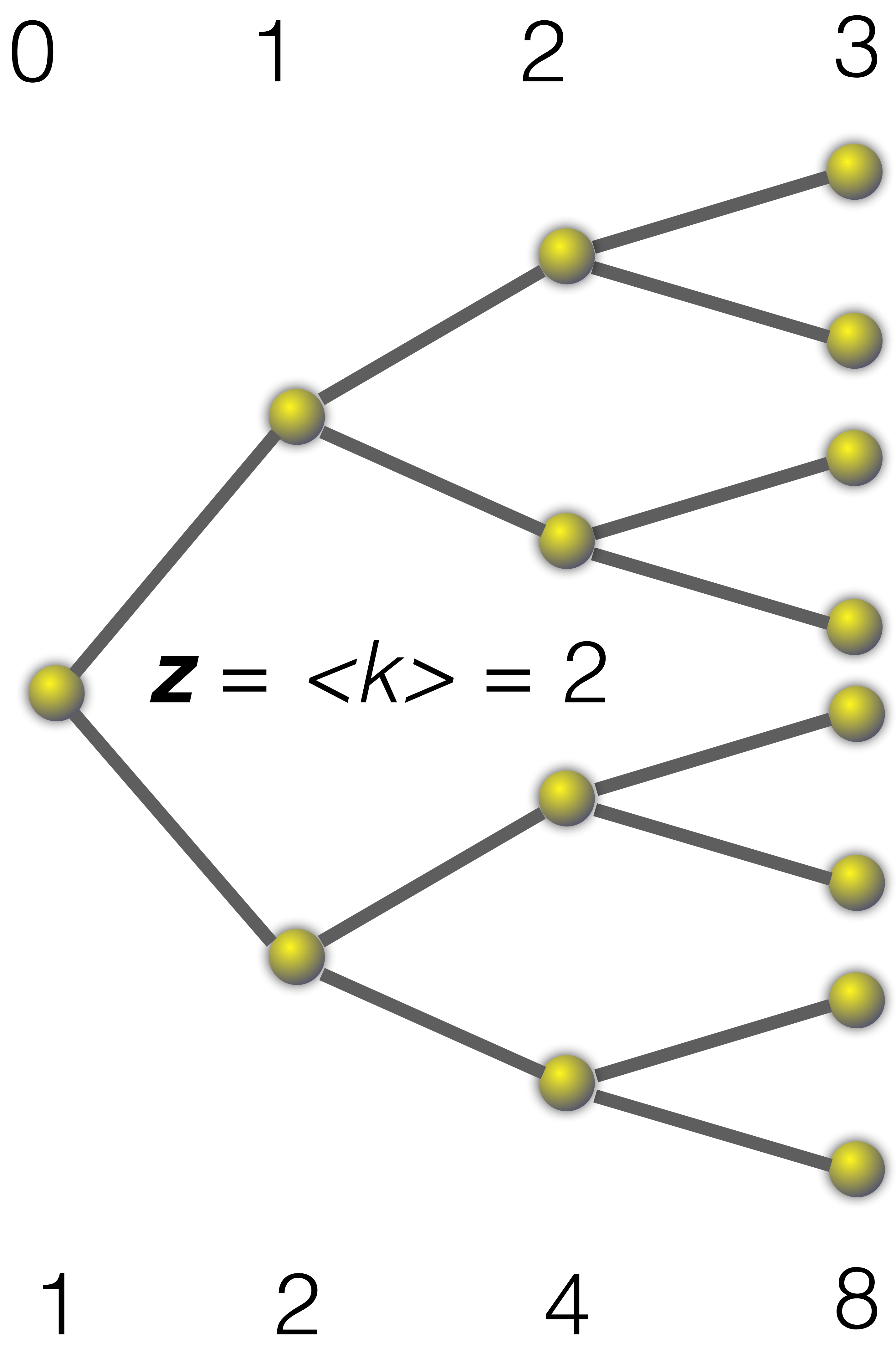
Electronic Circuits



Brain of a worm (*C. Elegans*)

Power grids

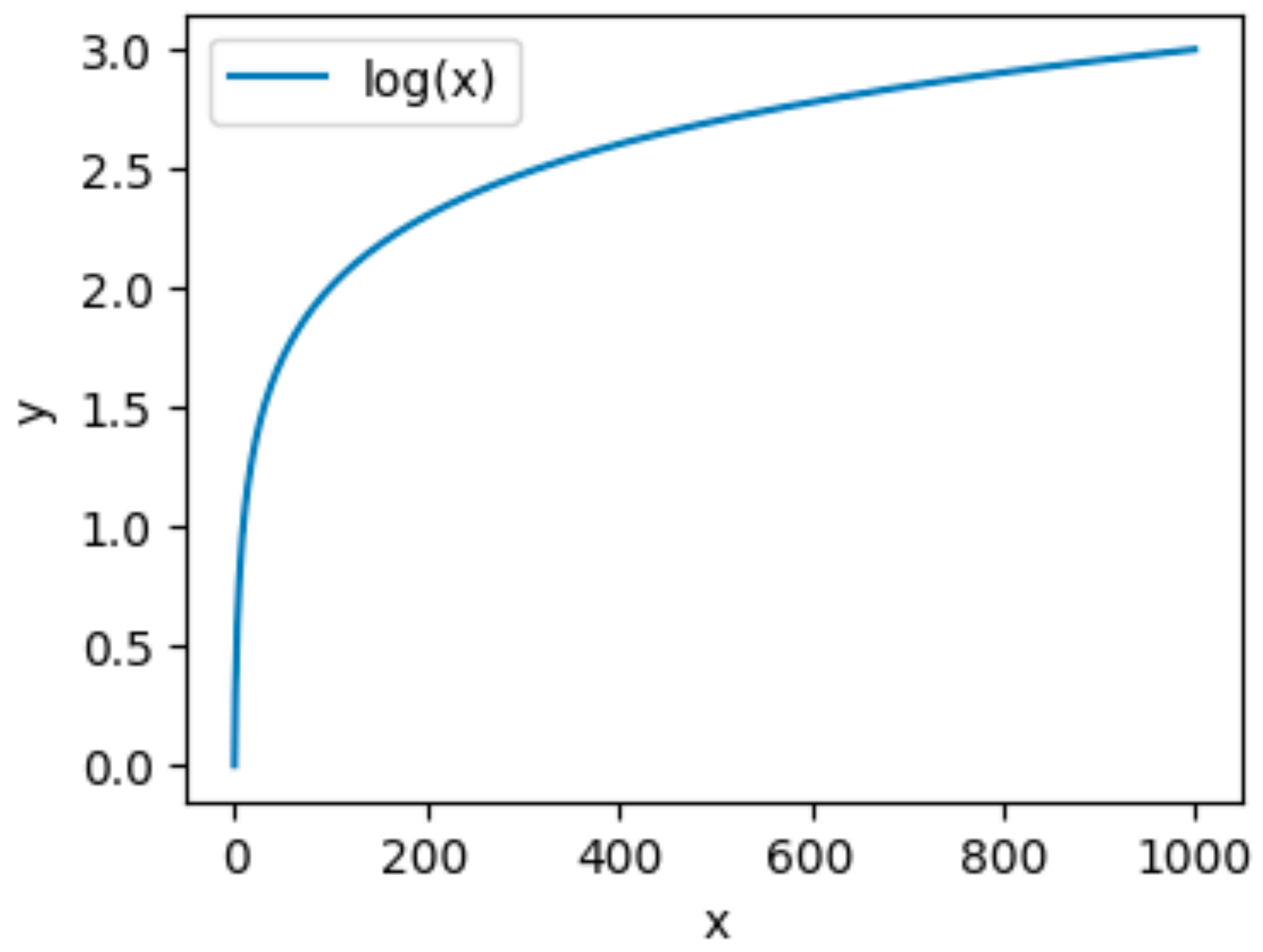
Average Path Length



$$N_d = z^d$$

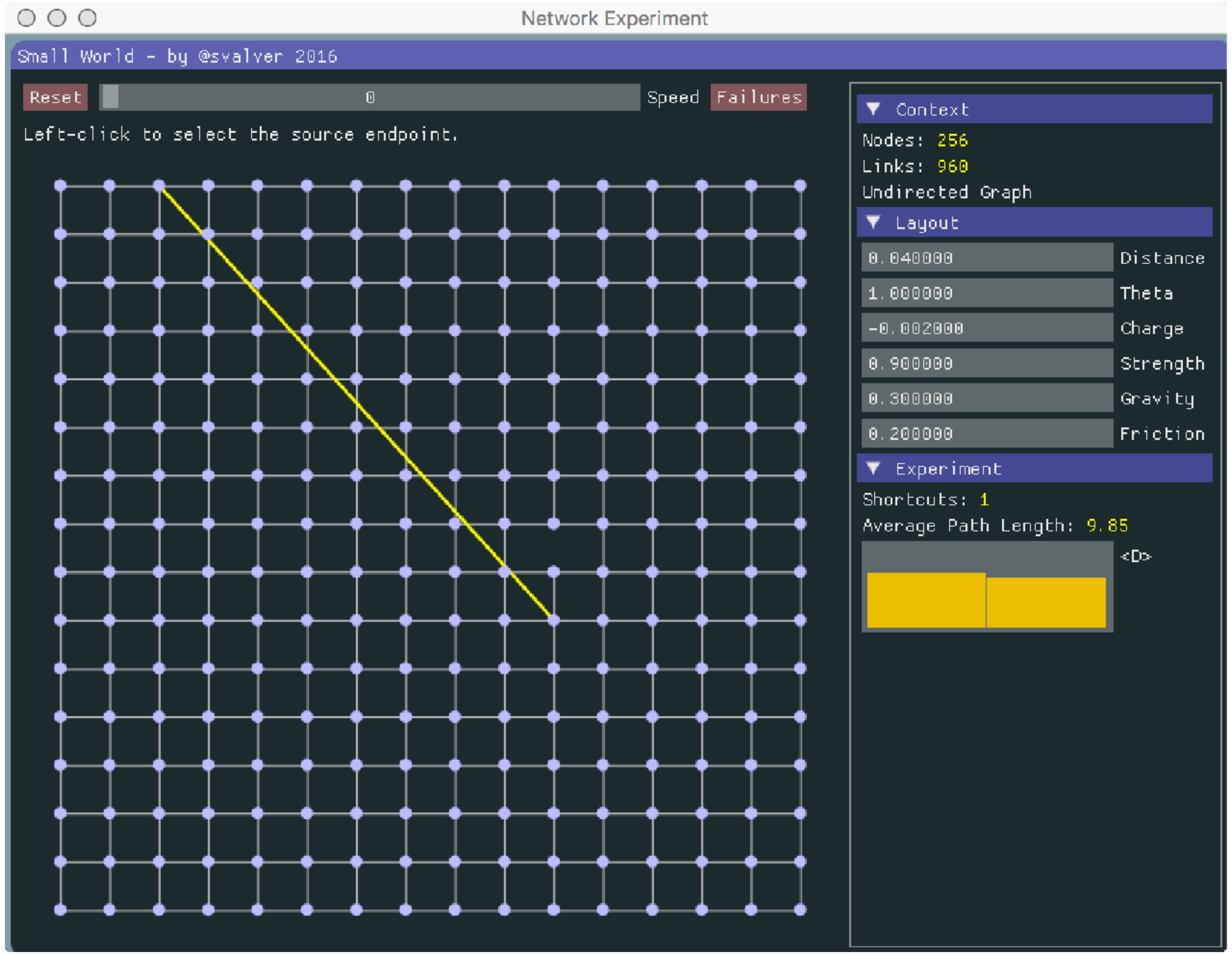
$$\log(N) = d\log(z)$$

$$\langle d \rangle \approx \frac{\log(N)}{\log(z)}$$

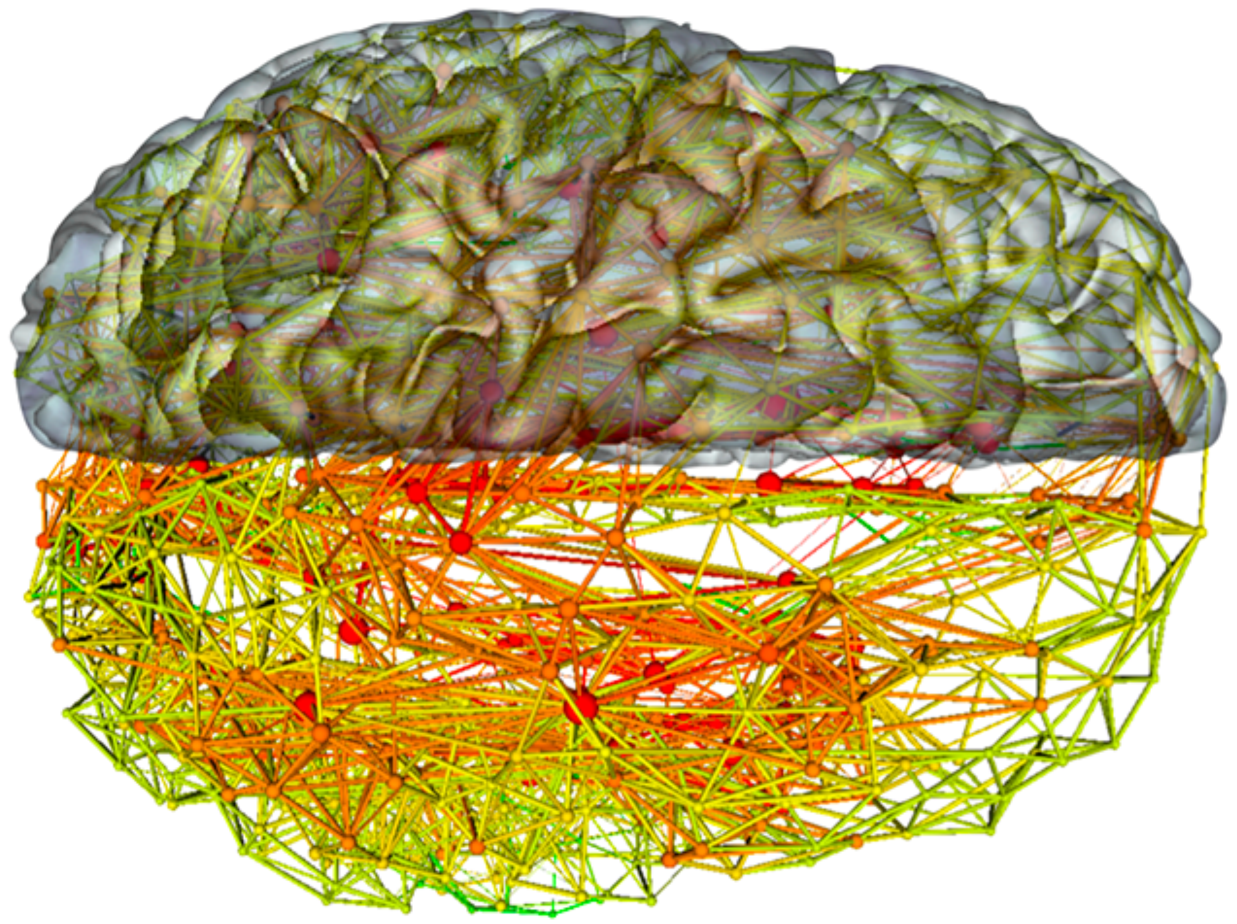


Activity: Small Worlds

https://tinyurl.com/587wsvwj



11. Which shortcuts reduce the average distance ?

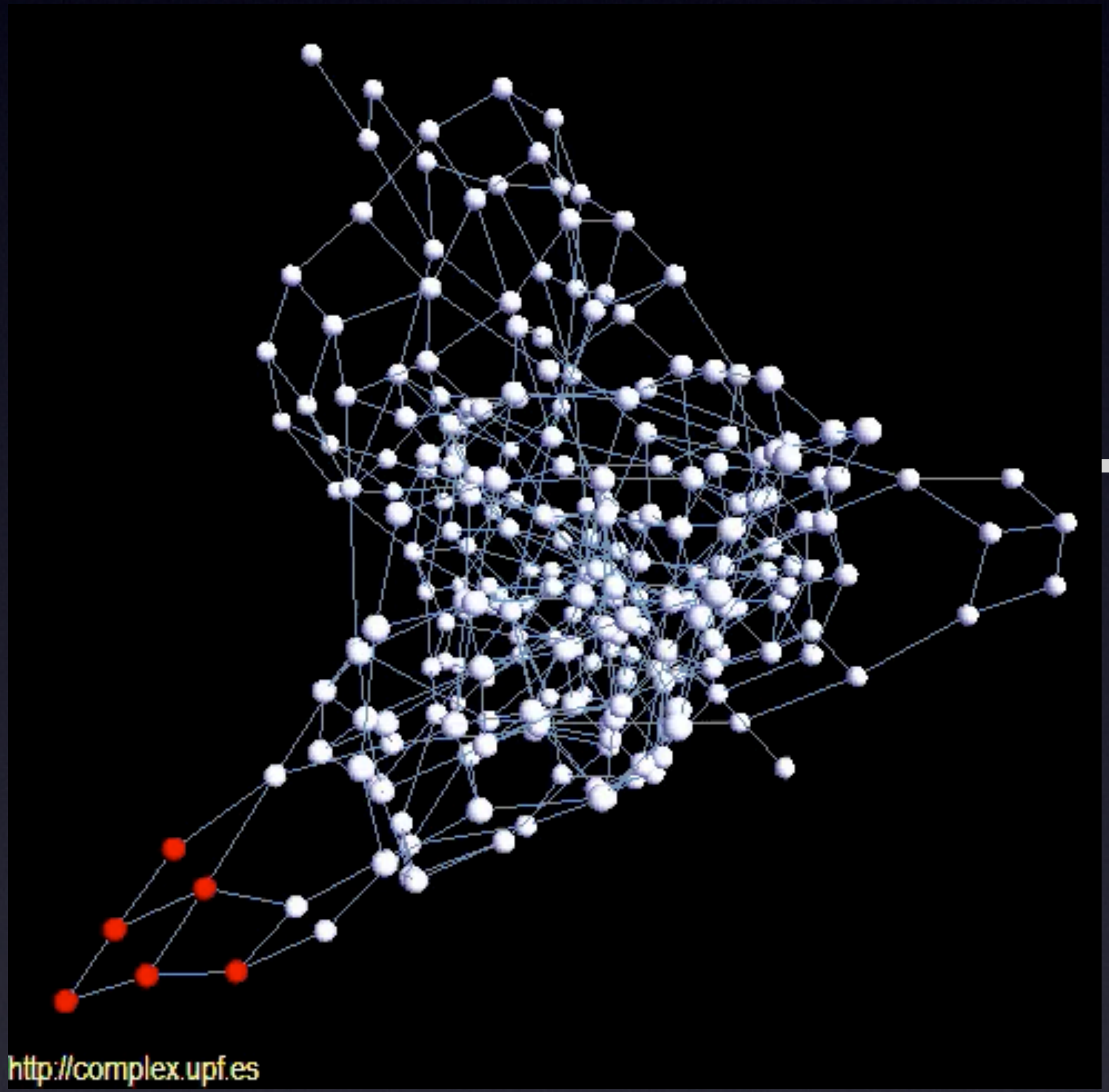


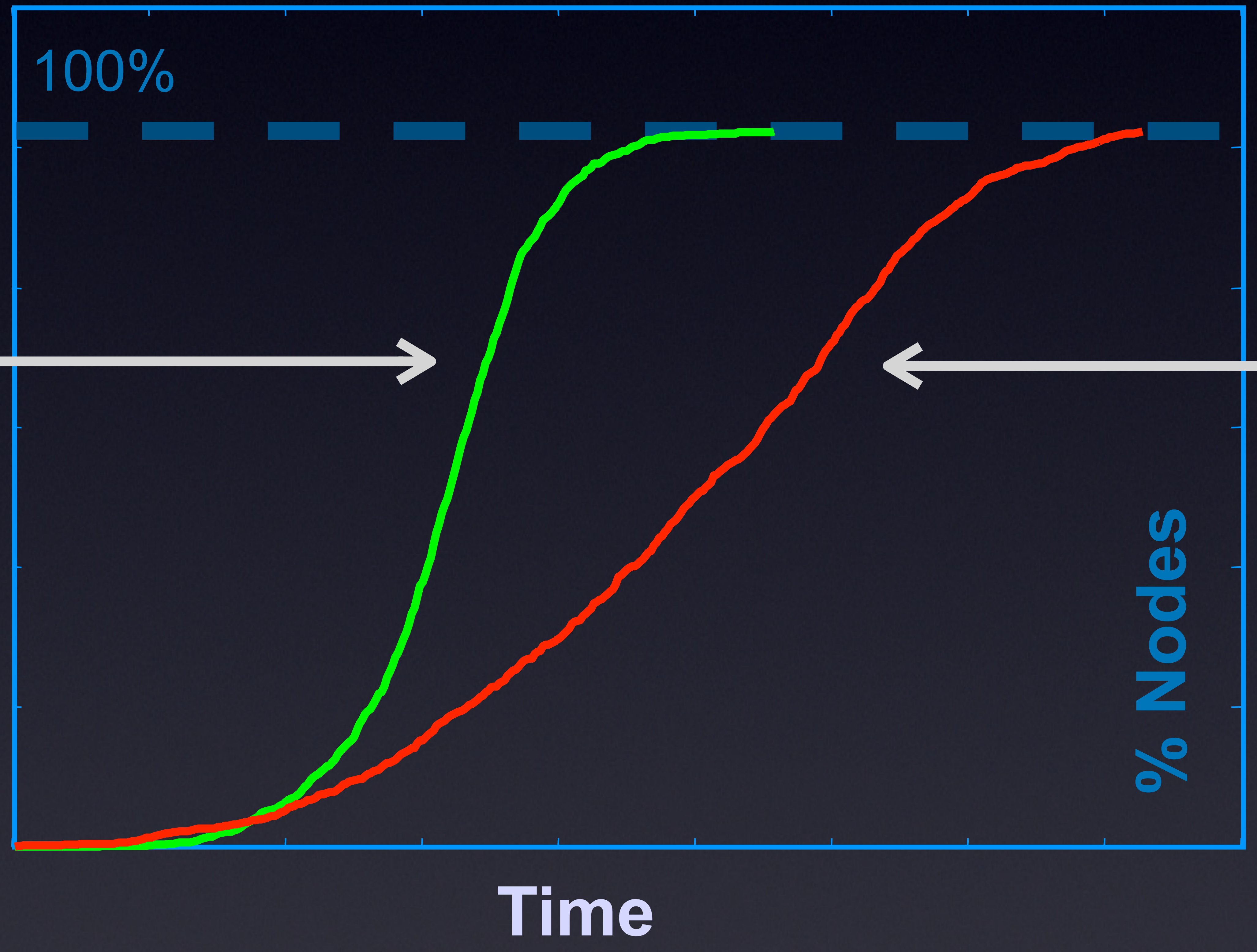
12. After completing 10 experiments, plot the (shortcuts, mean path length) curve. Can the distinction between good and poor networks be made?

Diffusion Processes

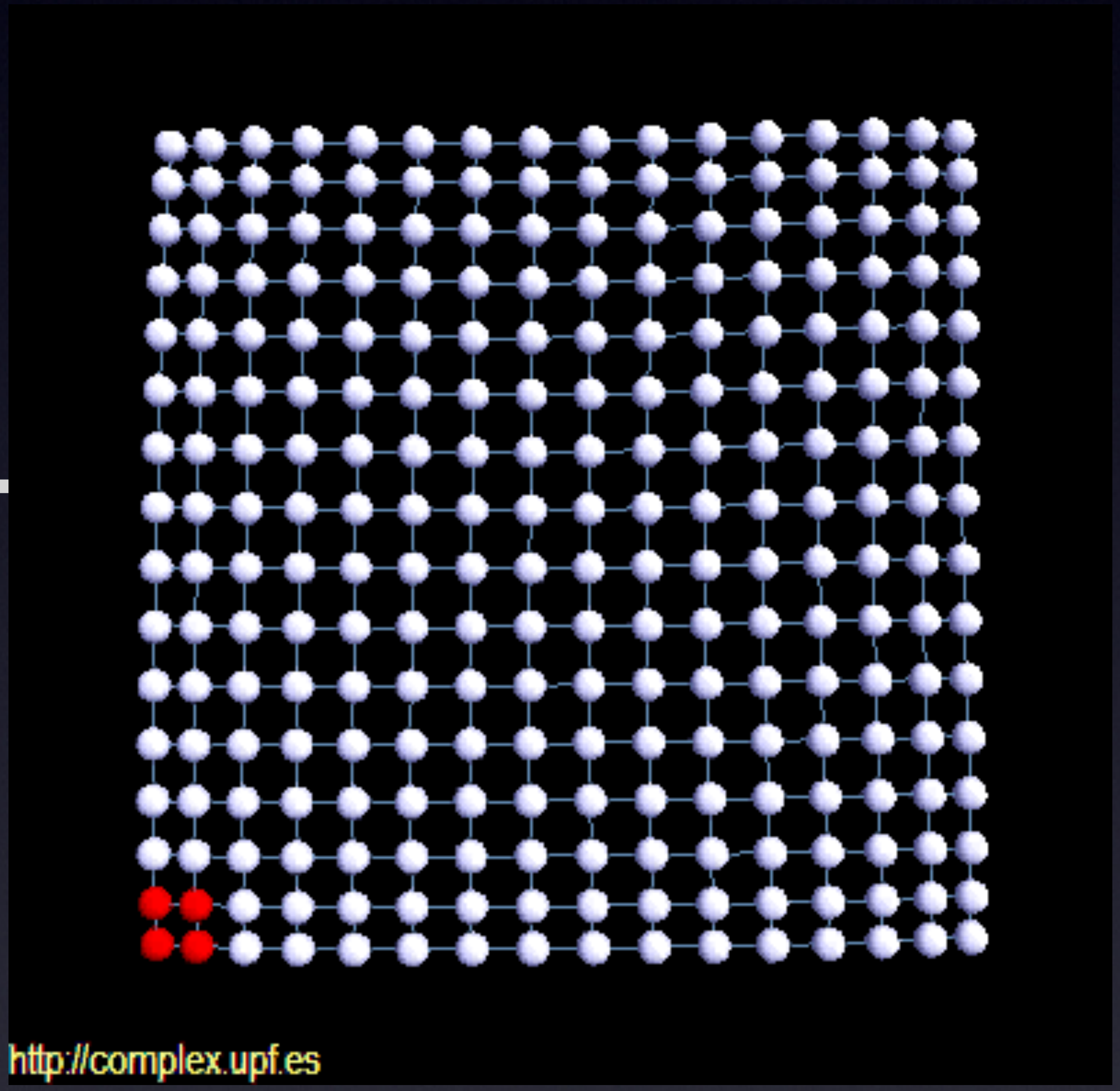
By defining a few long-distance links, diffusion may be accelerated

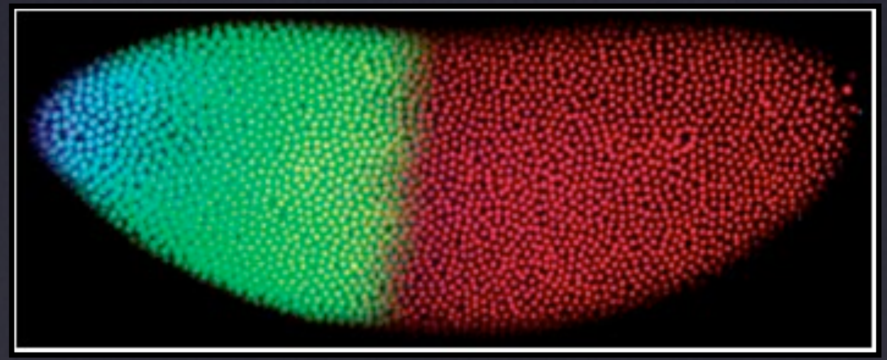
Small-World



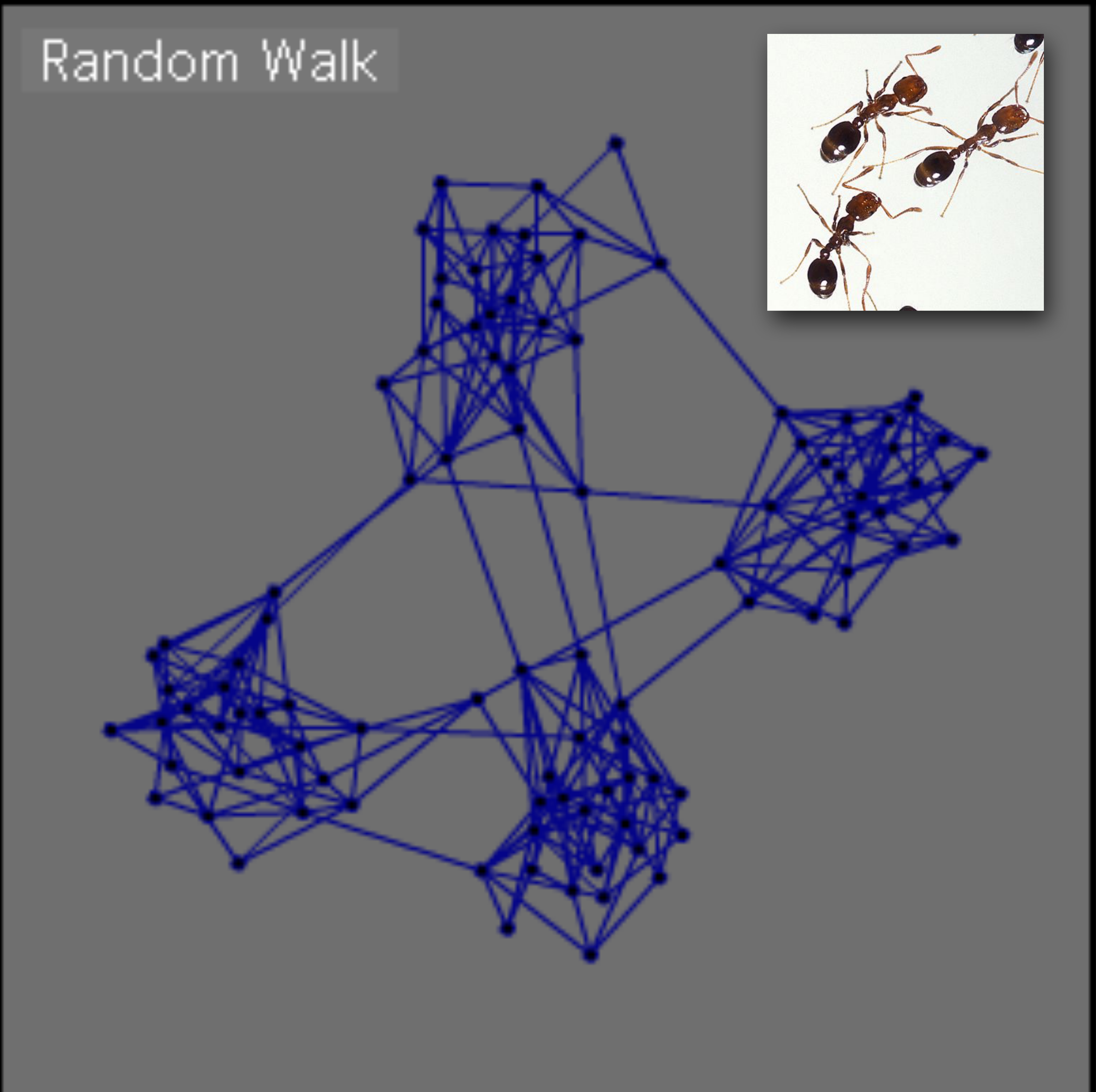


Lattice





Structure-Function Relationship

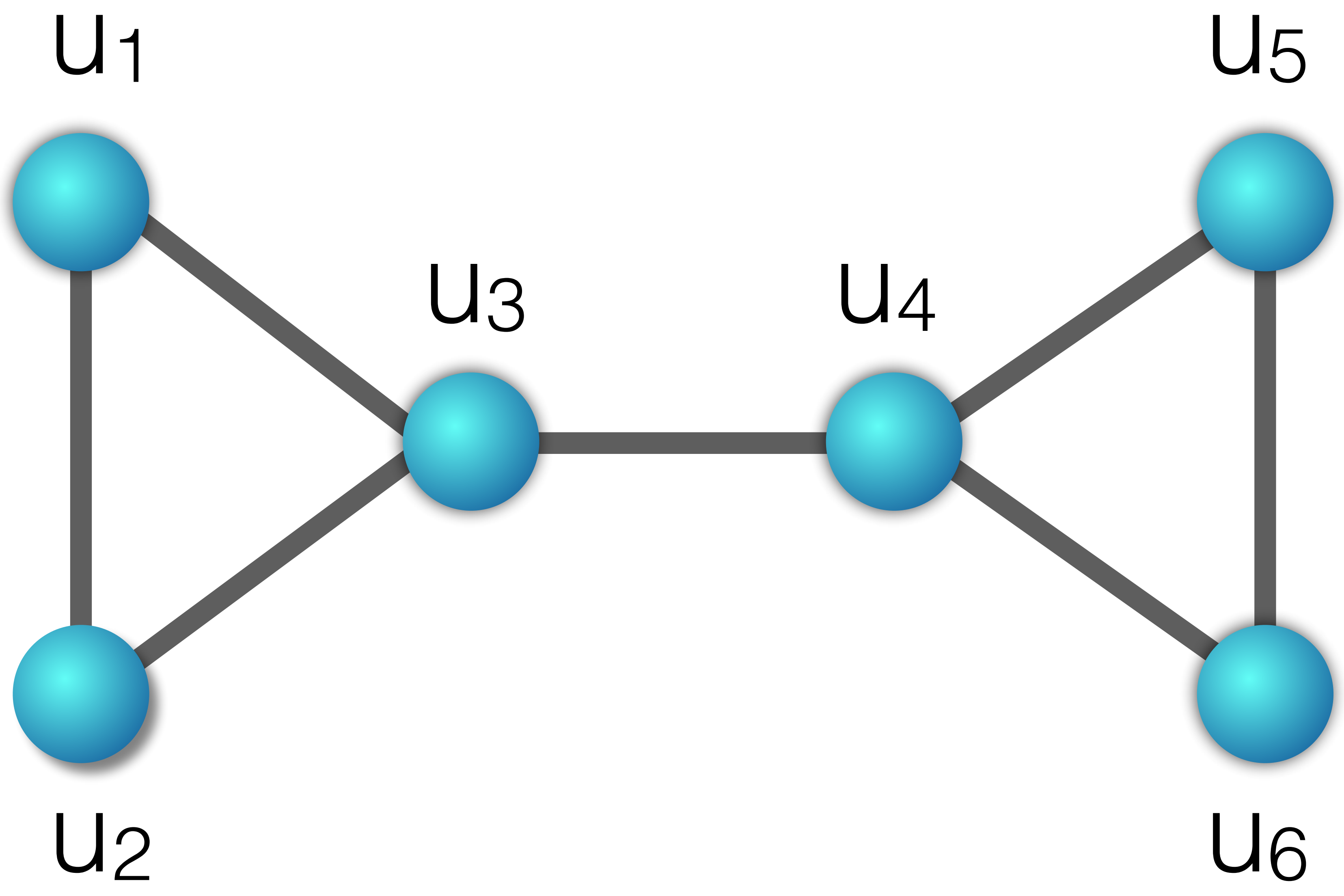


Modularity
Evolution & Tinkering

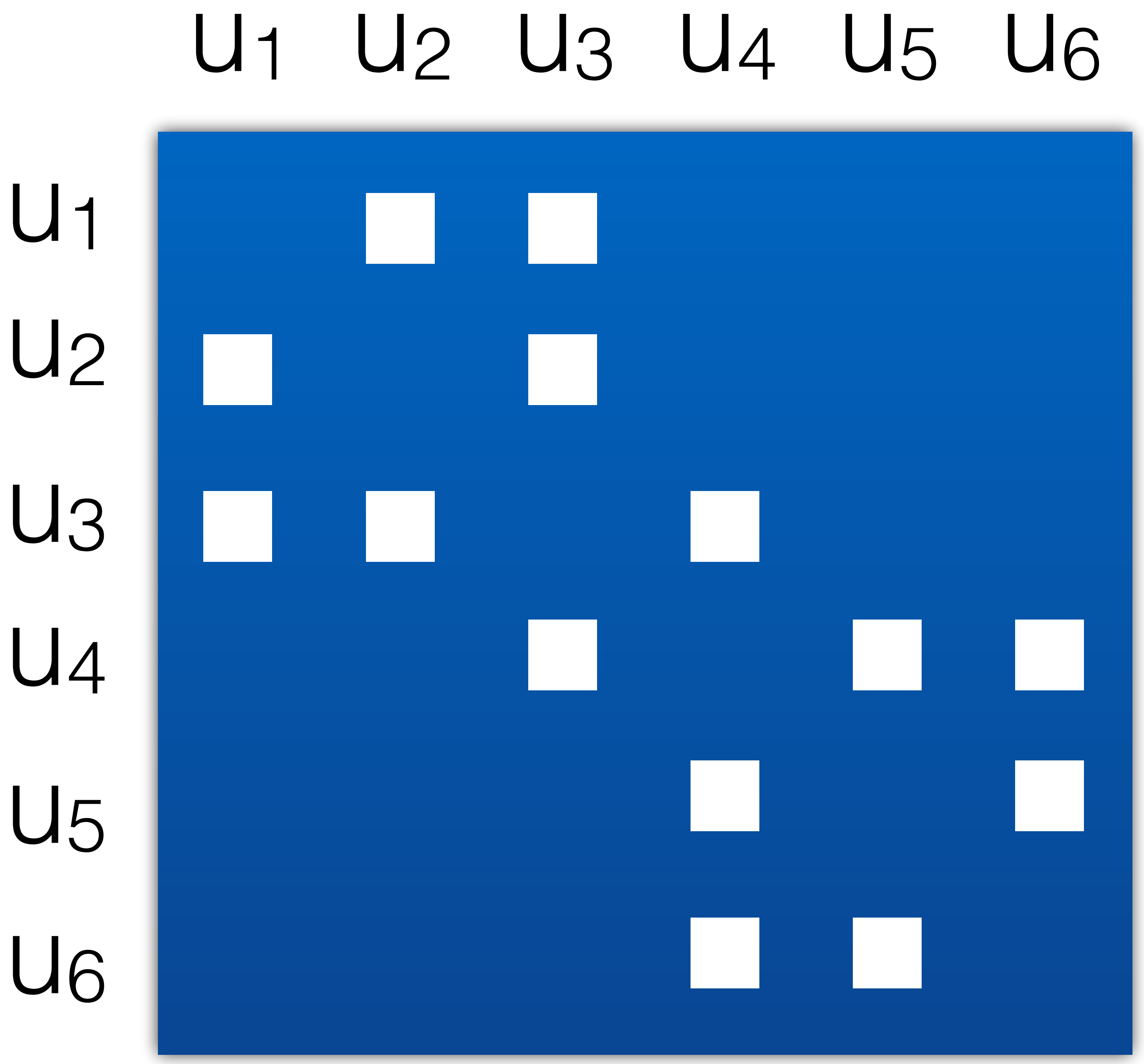
Definition

Modularity quantifies the degree to which nodes are grouped together and dependent on one another.

How species coexist in a competitive world?



Network



Adjacency Matrix

Newman & Girvan **Phys Rev E** 69, 026113 (2004)

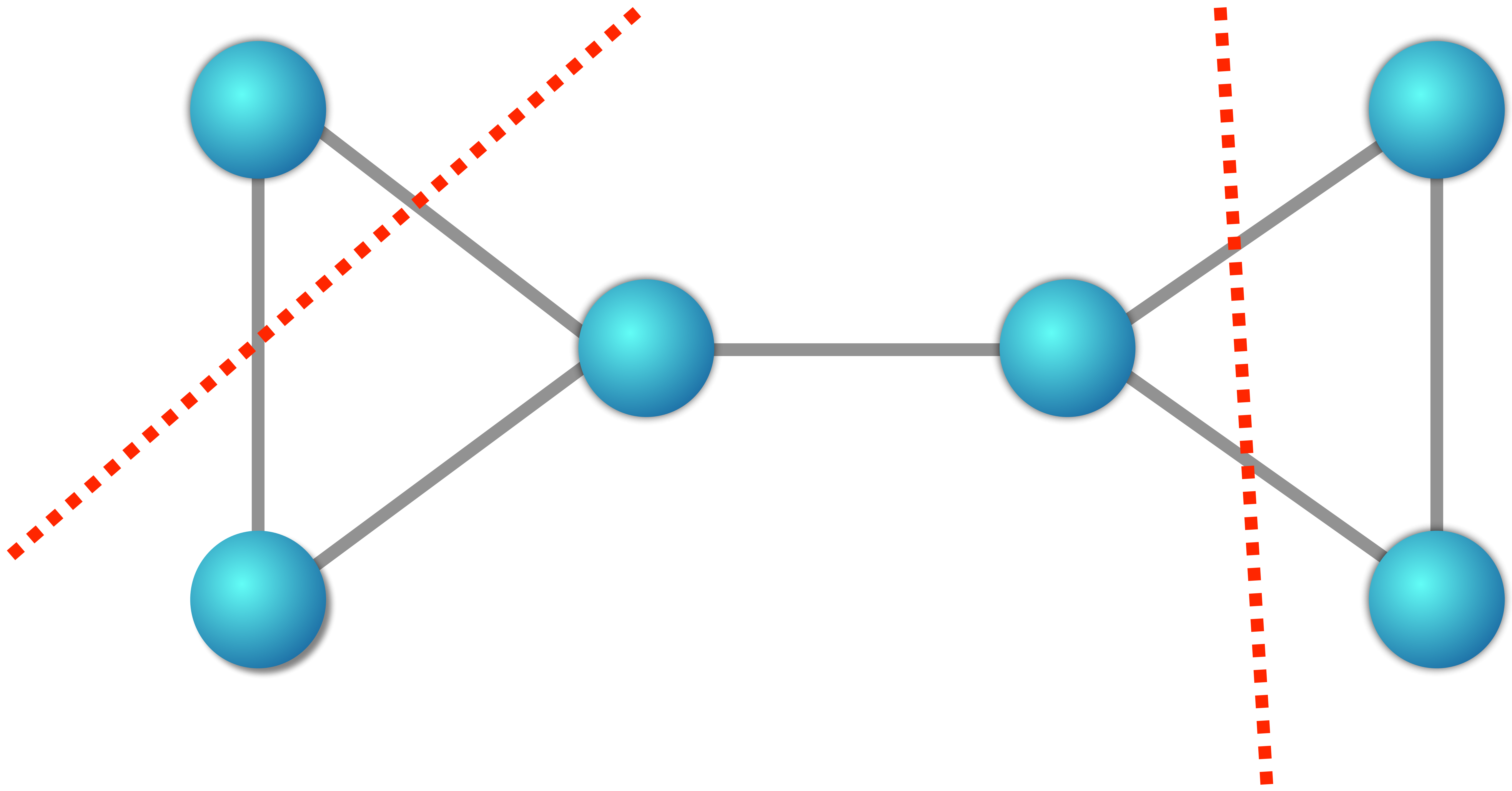
Community Detection

(1) Divide up the network

(2) Calculate the modularity value (Q)

(3) Repeat until a solution is optimised

(1) Divide up the network



(2) Calculate the modularity value (Q)

$$Q = \sum \left[\text{\textbf{Observed} fraction of links in group} - \text{\textbf{Expected} fraction of links in group} \right]$$

For each of the modules

(2) Calculate the modularity value (Q)

$$Q = \sum_{s=1}^{N_m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$

Number of Modules

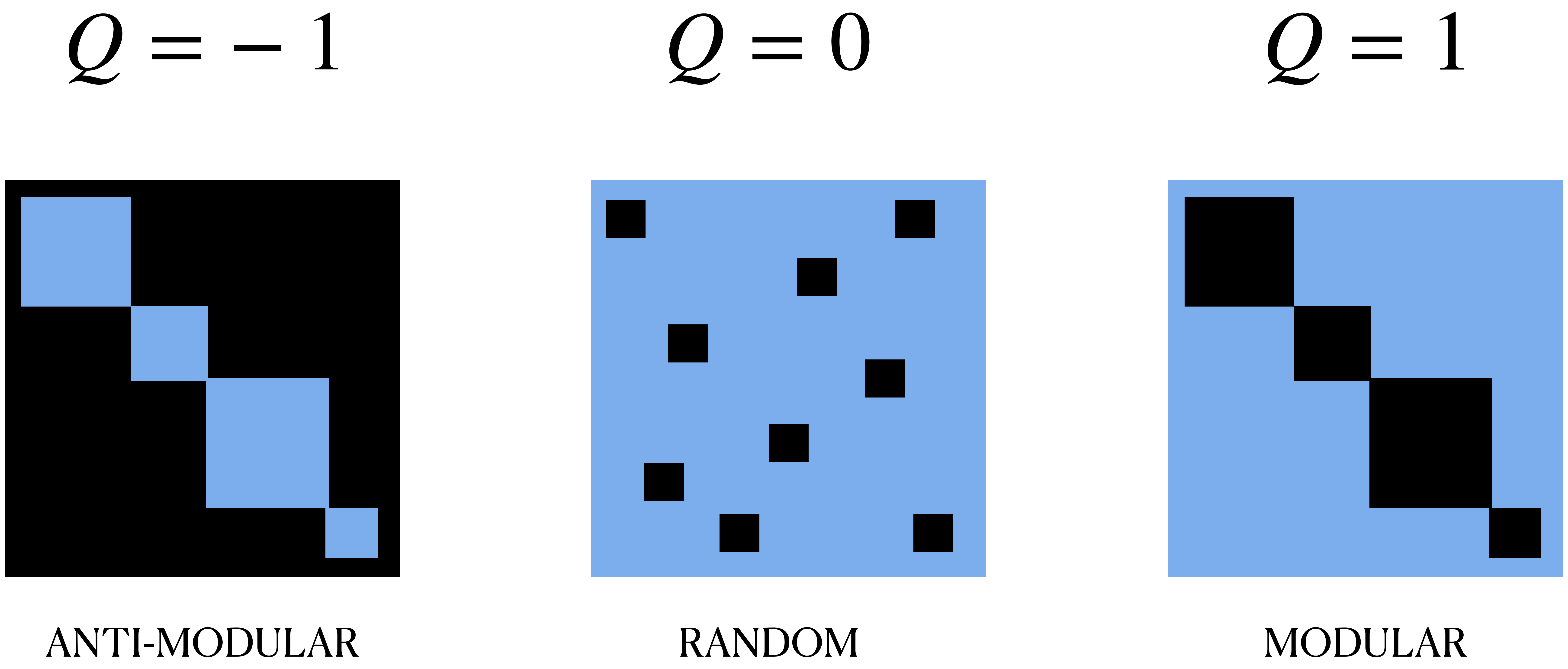
Number of links between nodes in module 's'

Sum of degrees of nodes in module 's'

Taking square to obtain link probability

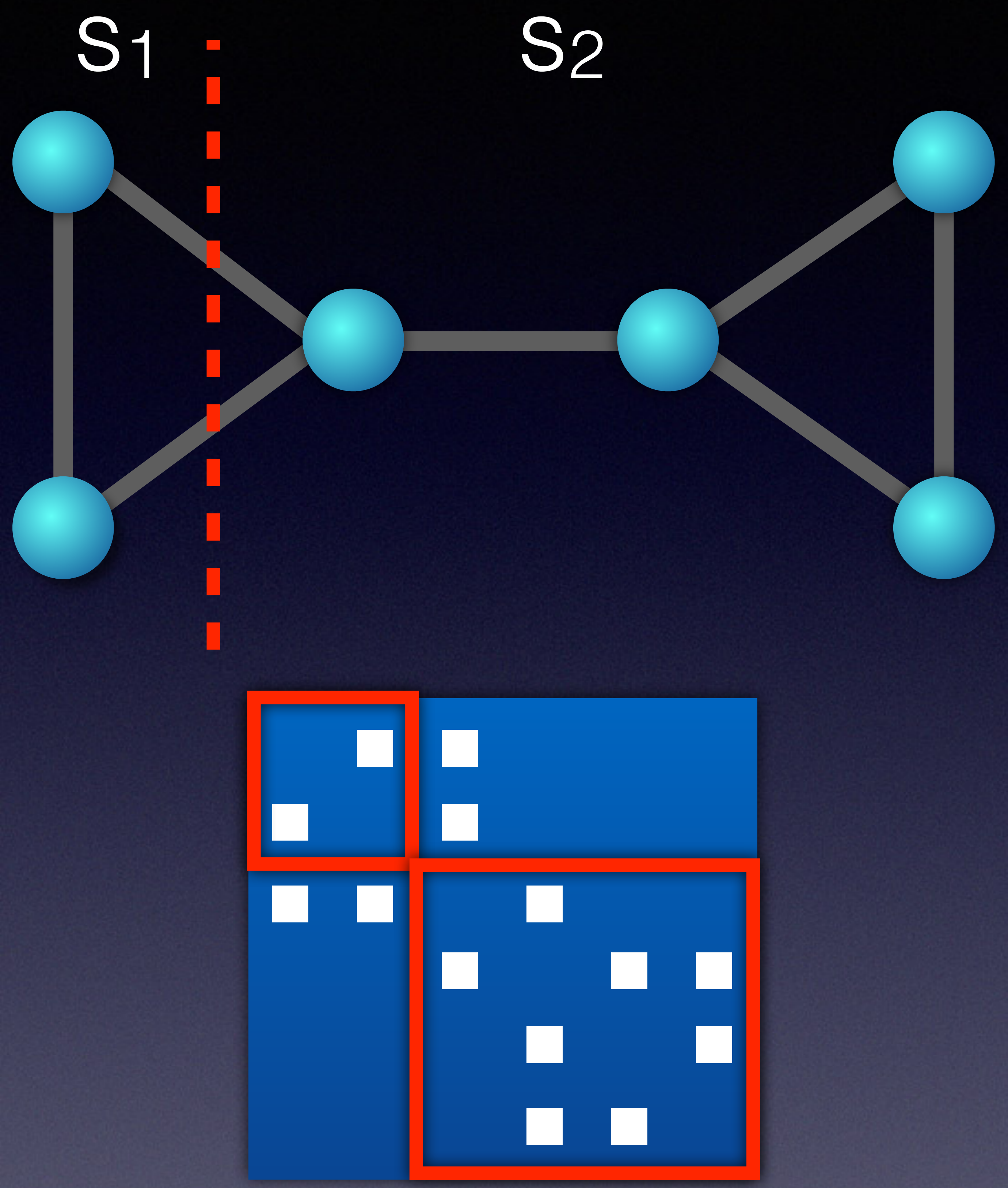
Number of links in the network

Girvan and Newman **PNAS** 99:7821 (2002)



$$Q = \sum_{s=1_m^N} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$

Example (1/2)



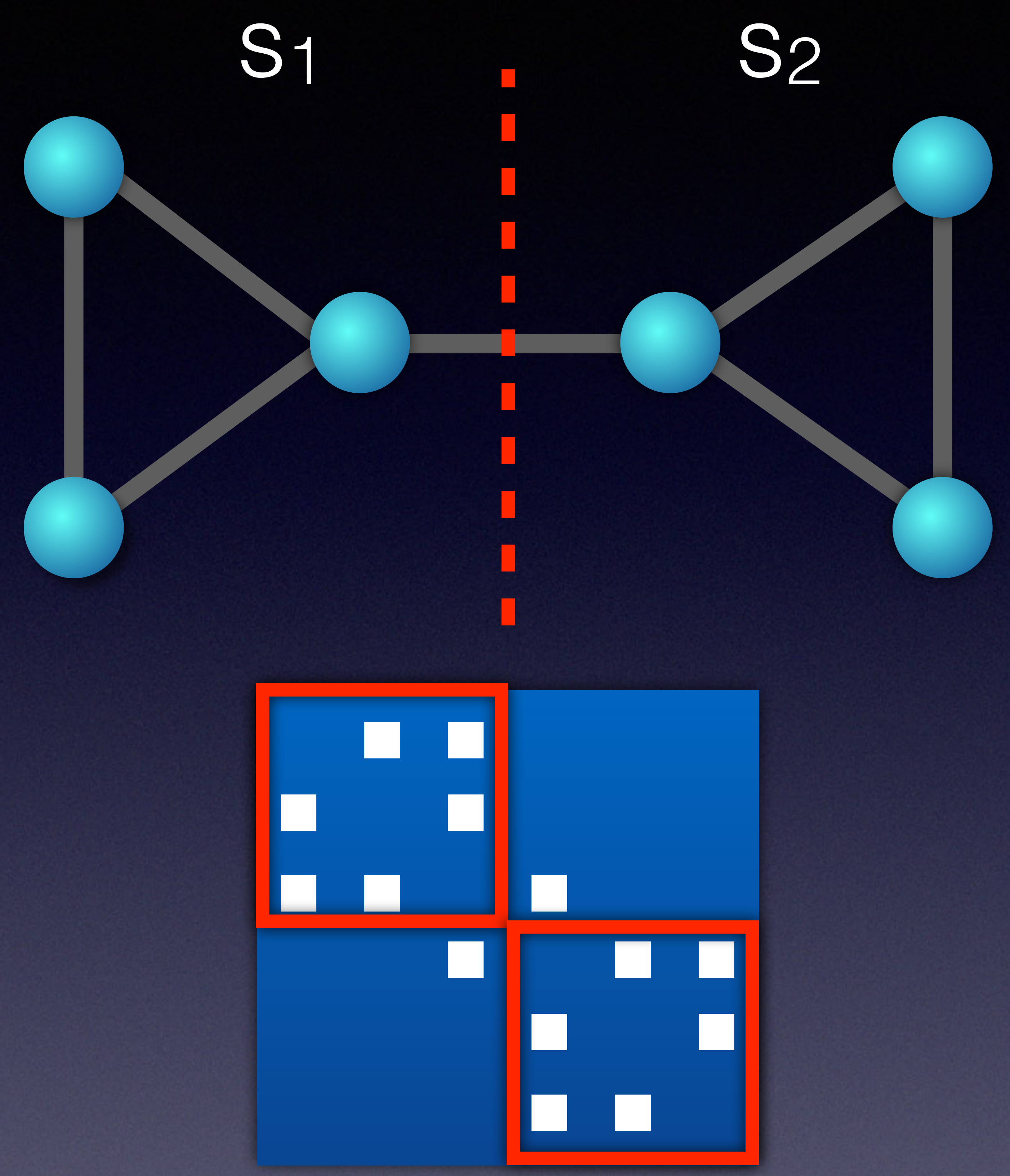
$$Q = \sum_{s=1}^{N_m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$

$$Q_{s_1} = \frac{1}{7} - \left(\frac{4}{14} \right)^2 = 0.06$$

$$Q_{s_2} = \frac{4}{7} - \left(\frac{10}{14} \right)^2 = 0.06$$

$$Q = Q_{s_1} + Q_{s_2} = 0.12$$

Example (2/2)



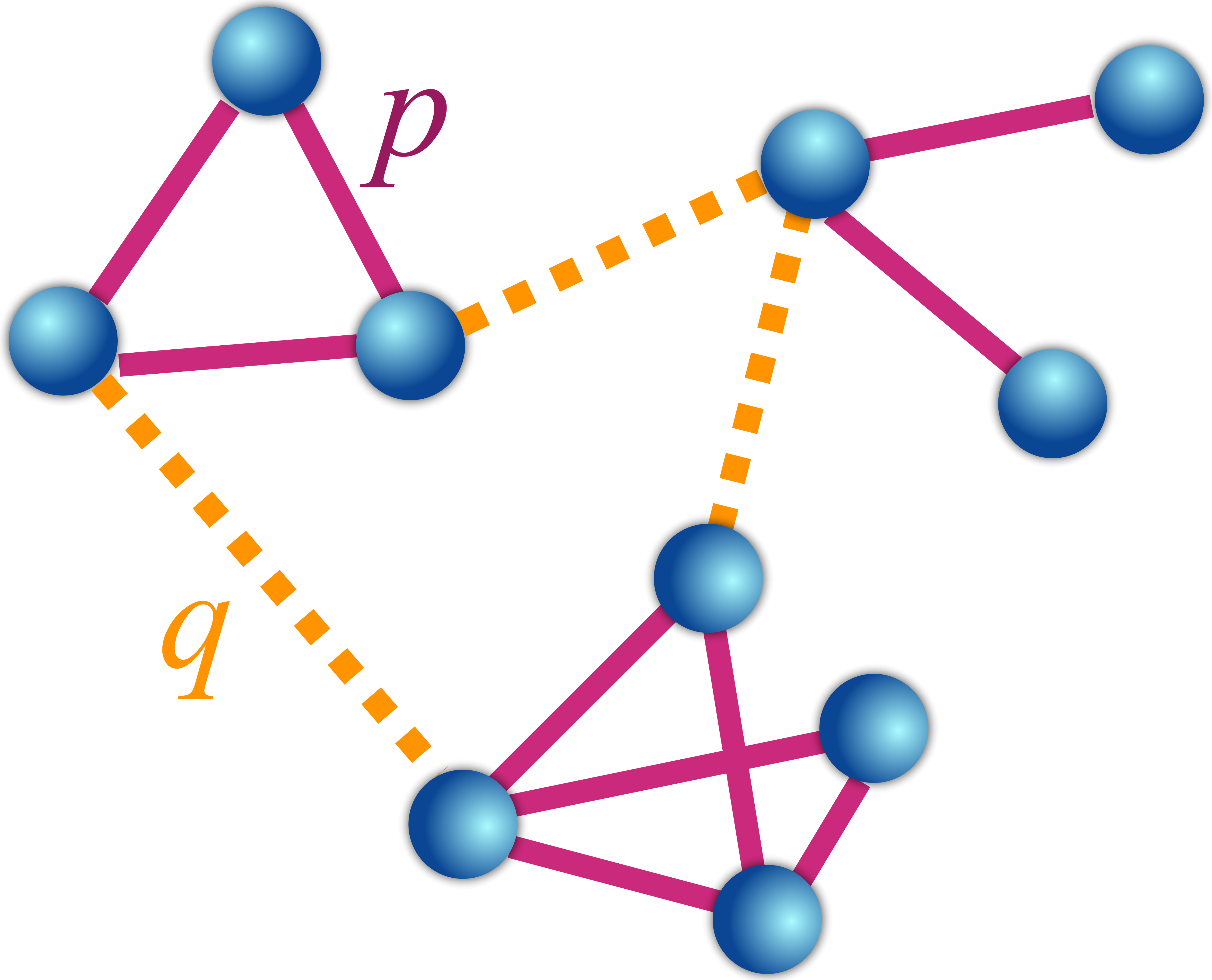
$$Q = \sum_{s=1}^{N_m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$

$$Q_{s_1} = \frac{3}{7} - \left(\frac{7}{14} \right)^2 = 0.18$$

$$Q_{s_2} = Q_{s_1} = 0.18$$

$$Q = Q_{s_1} + Q_{s_2} = 0.36 > 0.12$$

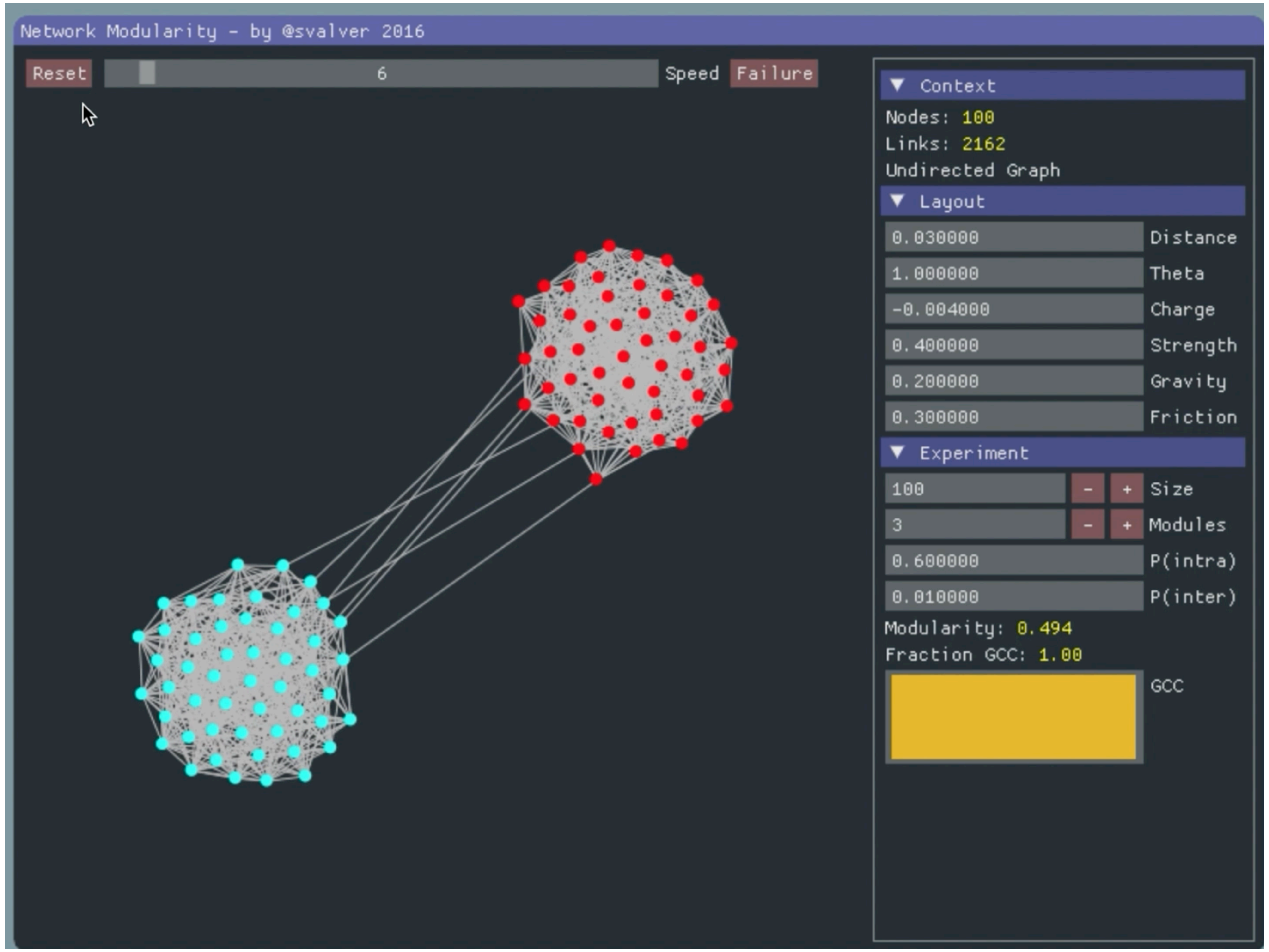
Random Modular Networks



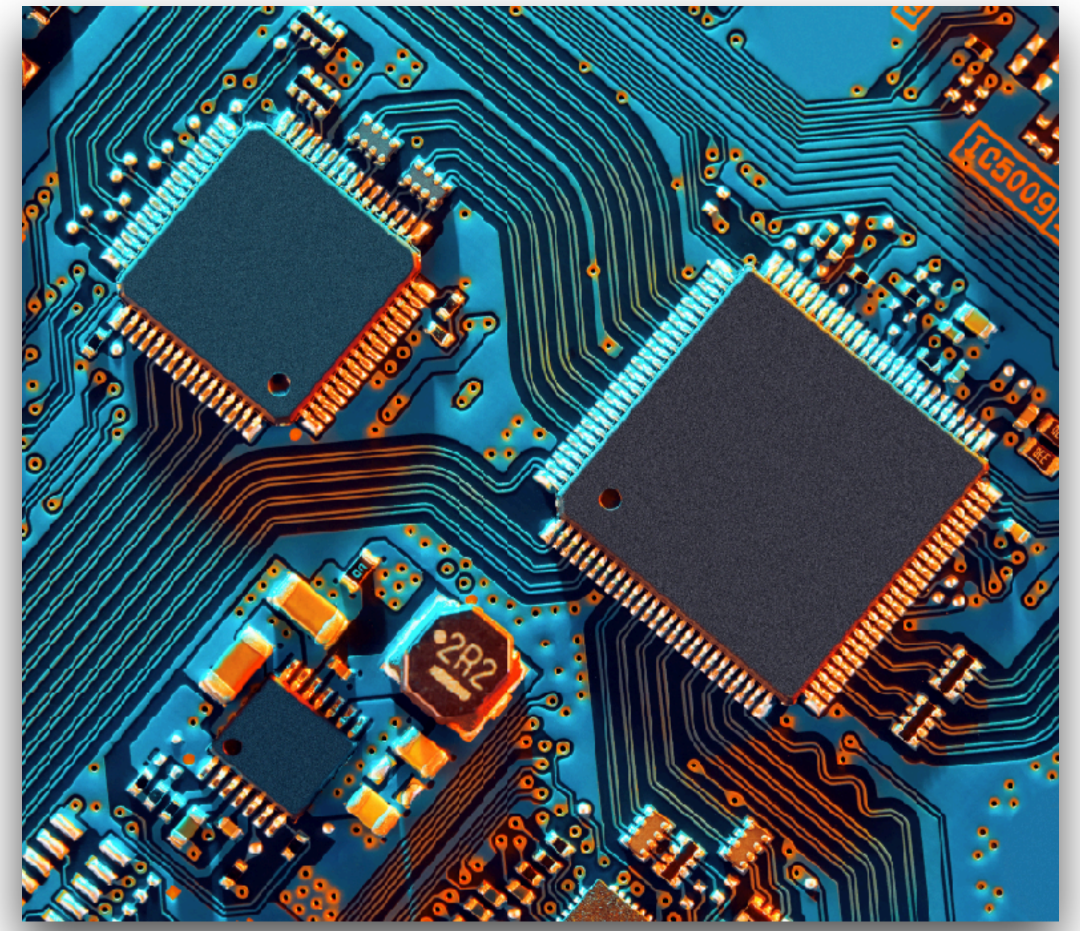
$$RMG(p, q)$$

Activity: Random Modular Networks

https://tinyurl.com/4a7syzuk



13. Can you use this model to generate a random graph? How?



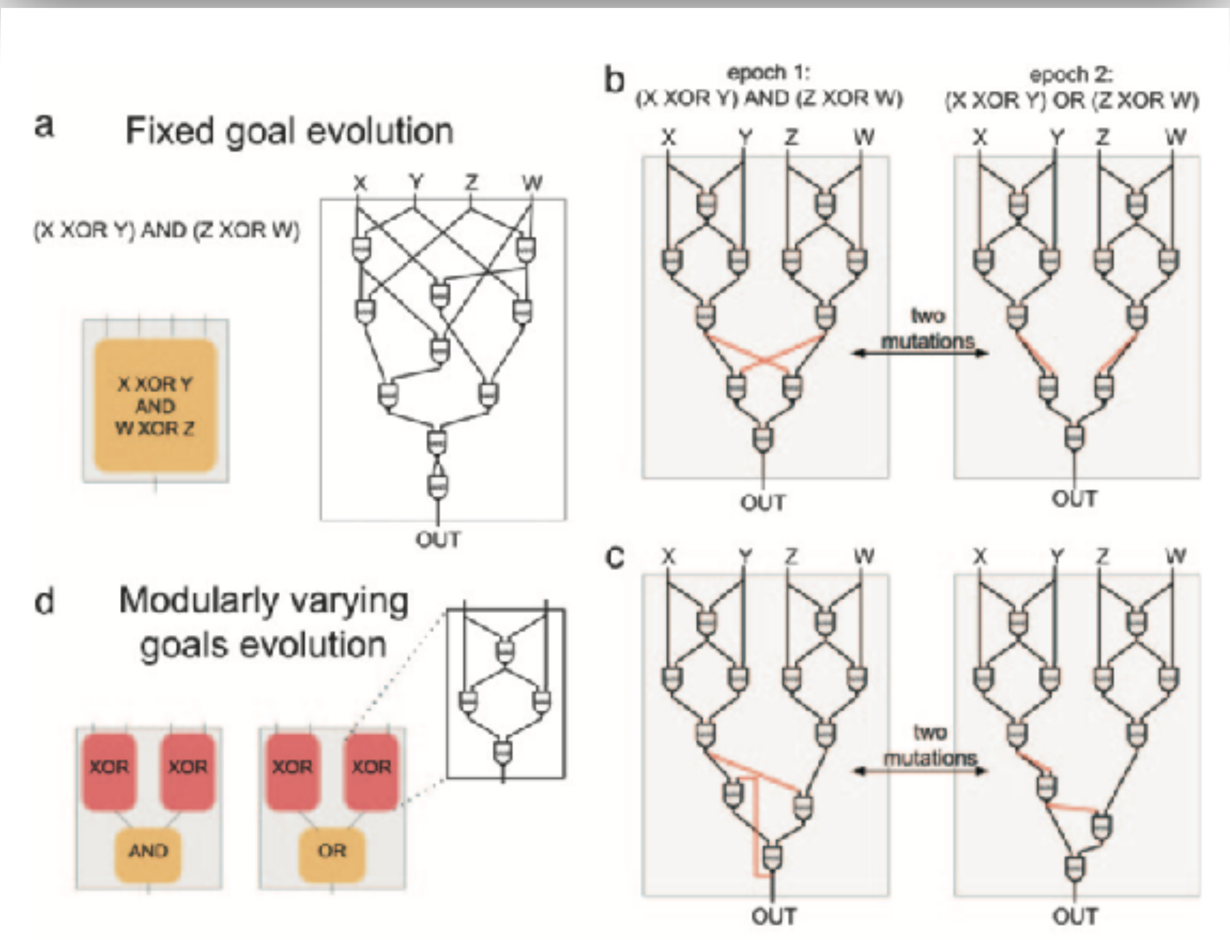
14. Which network has more linkages, RMG (p,q) or RMG (q,p)? Which one is more modular? Why?

Evolution of Modularity

Understanding the contributions of multiples forces in the evolutionary origins of modularity

Spontaneous evolution of modularity and network motifs

Nadav Kashtan and Uri Alon*



It has been suggested that networks evolved under "modularly varying goals" must be modular. However, it is unclear how many biological environments change in a modular way and if they change frequently enough.

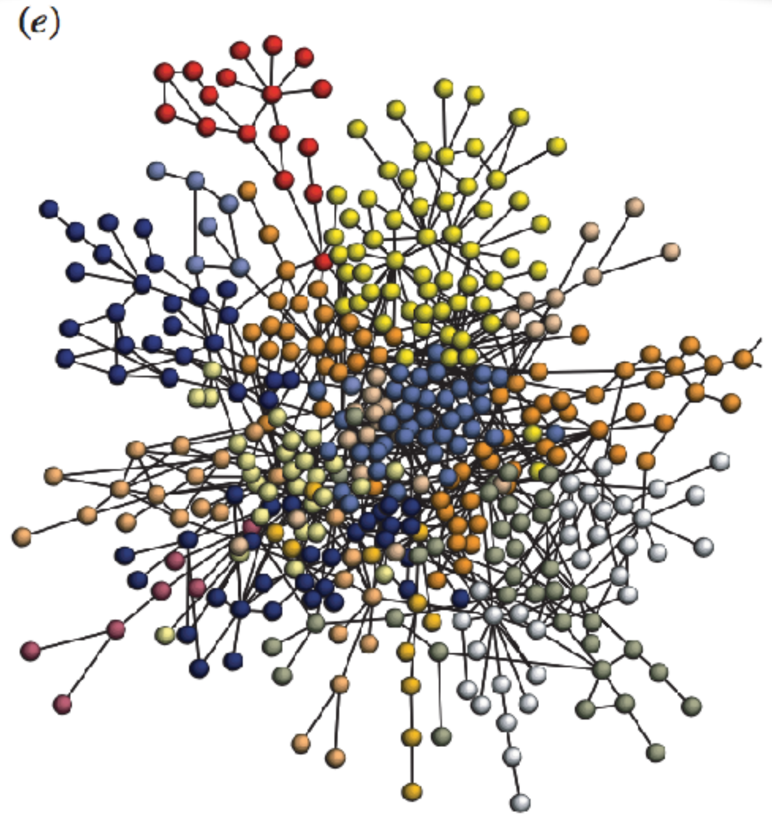
The evolutionary origins of modularity

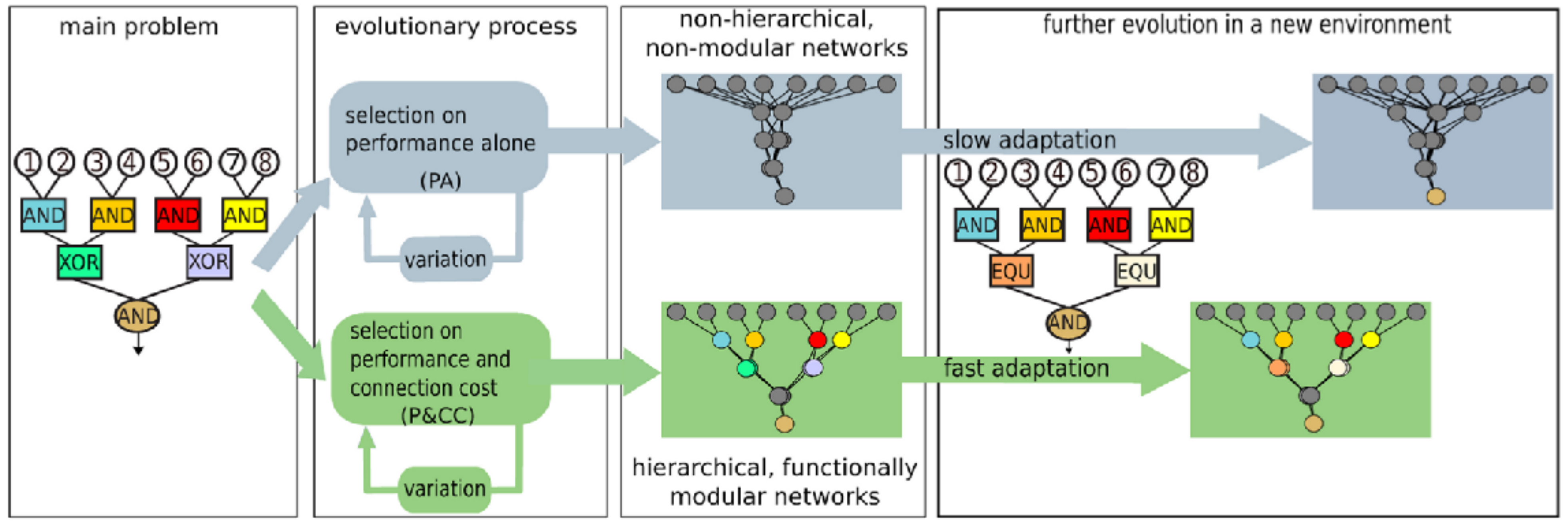
Jeff Clune[1,2,†], Jean-Baptiste Mouret[3,†] and Hod Lipson[1]
[1]Cornell University, Ithaca, NY, USA
[2]University of Wyoming, Laramie, WY, USA
[3]ISIR, Université Pierre et Marie Curie-Paris 6, CNRS UMR 7222, Paris, France

Most hypotheses of the emergence of modularity assume indirect selection for evolvability, but a direct selection pressure to reduce the cost of links causes the emergence of modular networks.

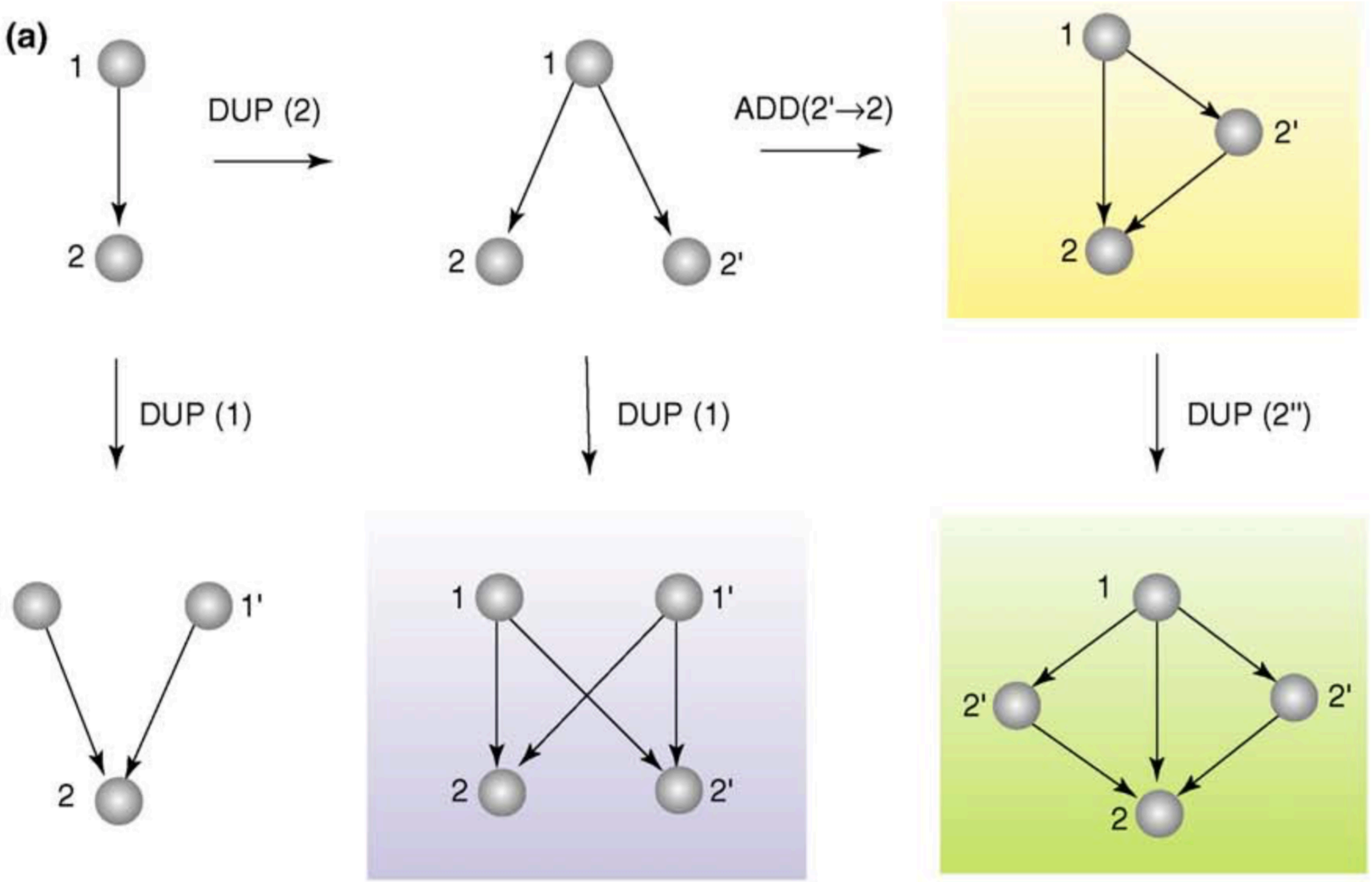
Spontaneous emergence of modularity in cellular networks

Ricard V. Solé[1,2,*] and Sergi Valverde[1,2]



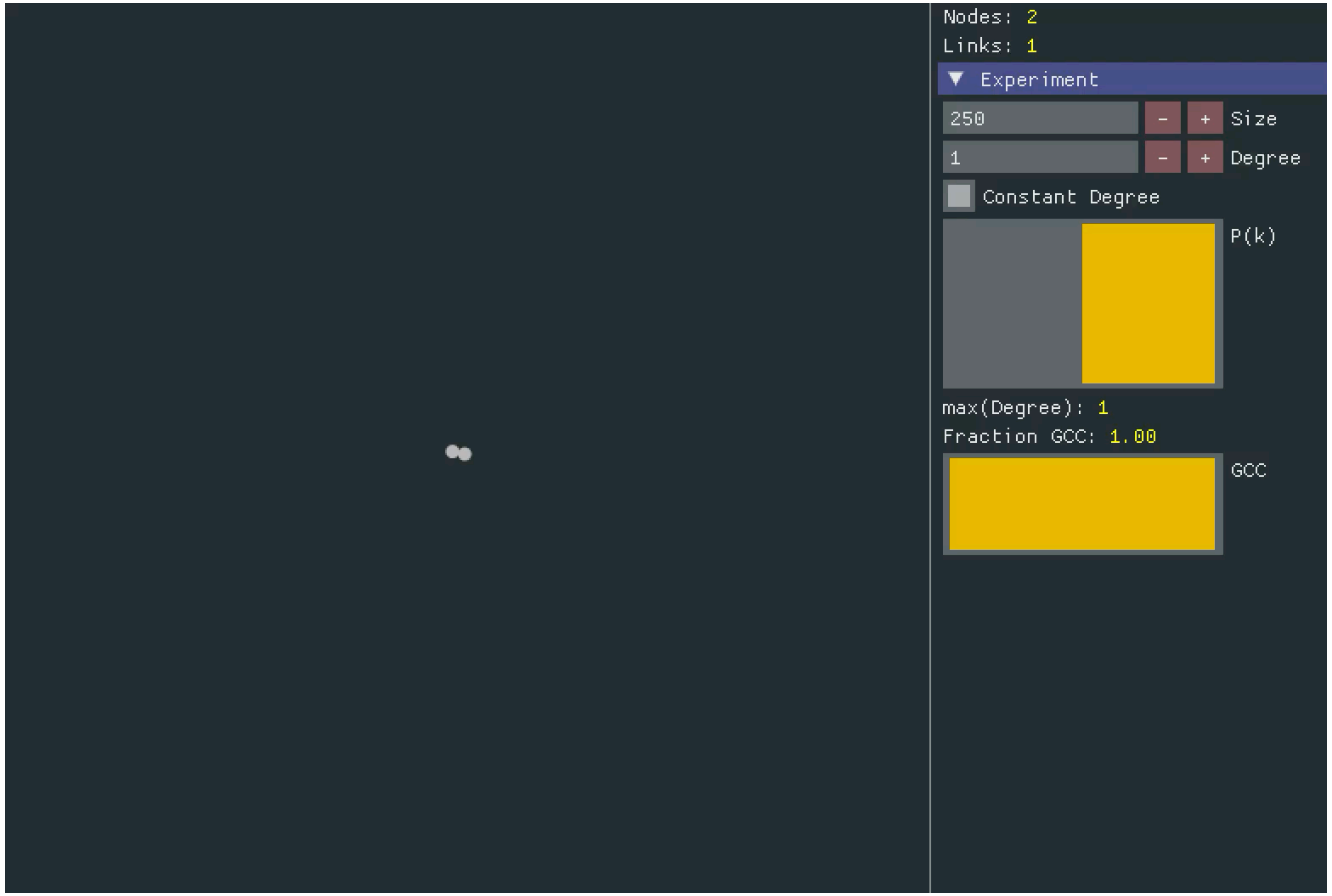


Diversity from Structural Rules

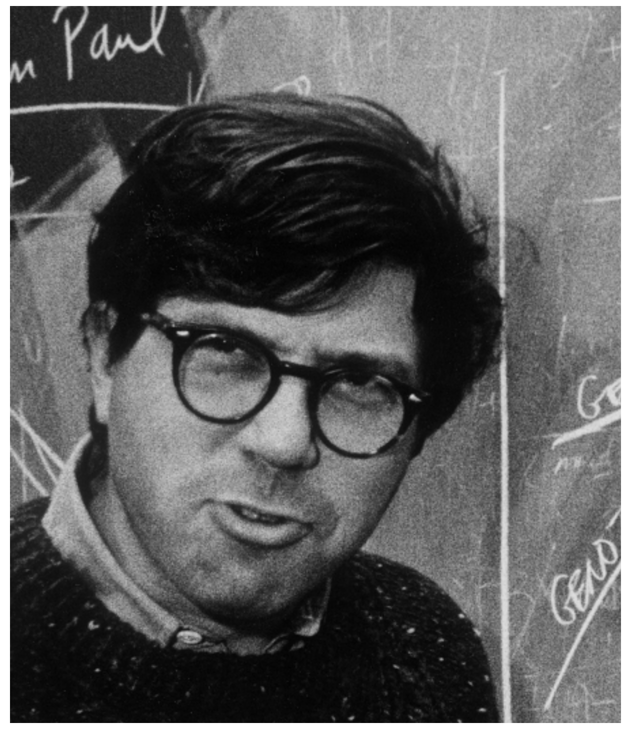


TRENDS in Ecology & Evolution

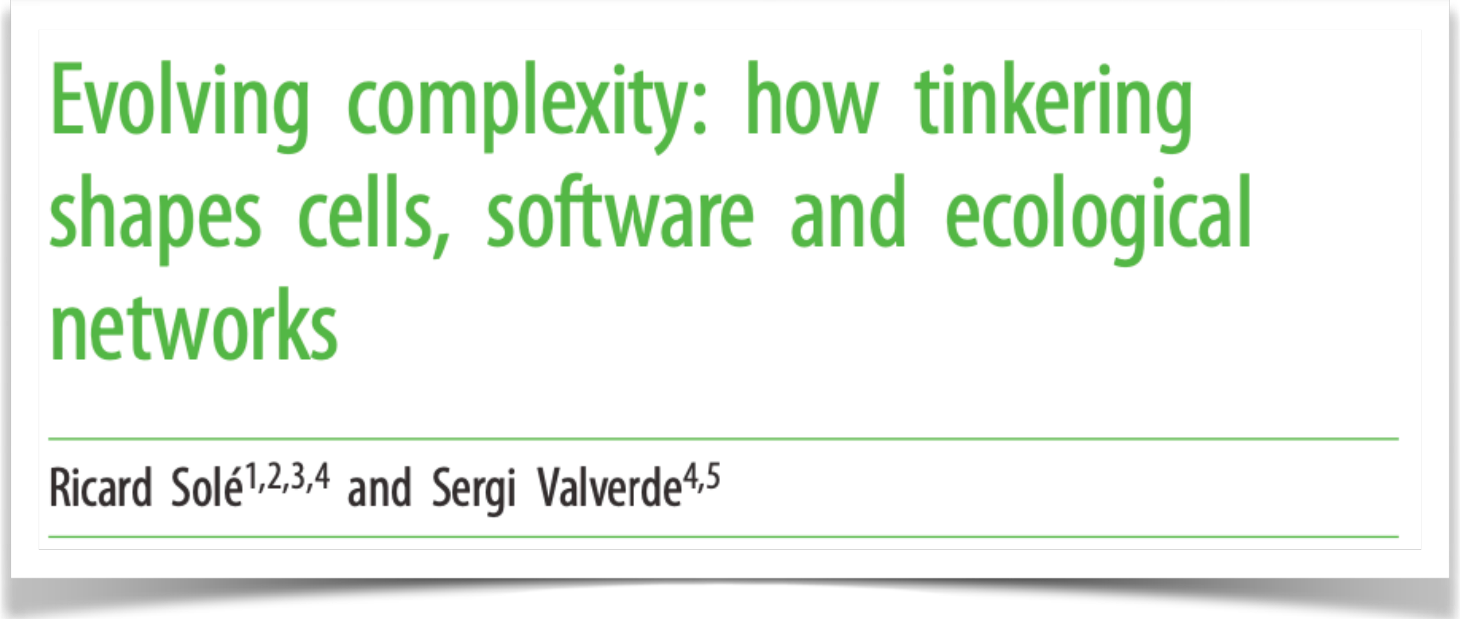
Tinkered Evolution of Networks



Stephen Jay Gould



Richard Lewontin



Evolving complexity: how tinkering shapes cells, software and ecological networks

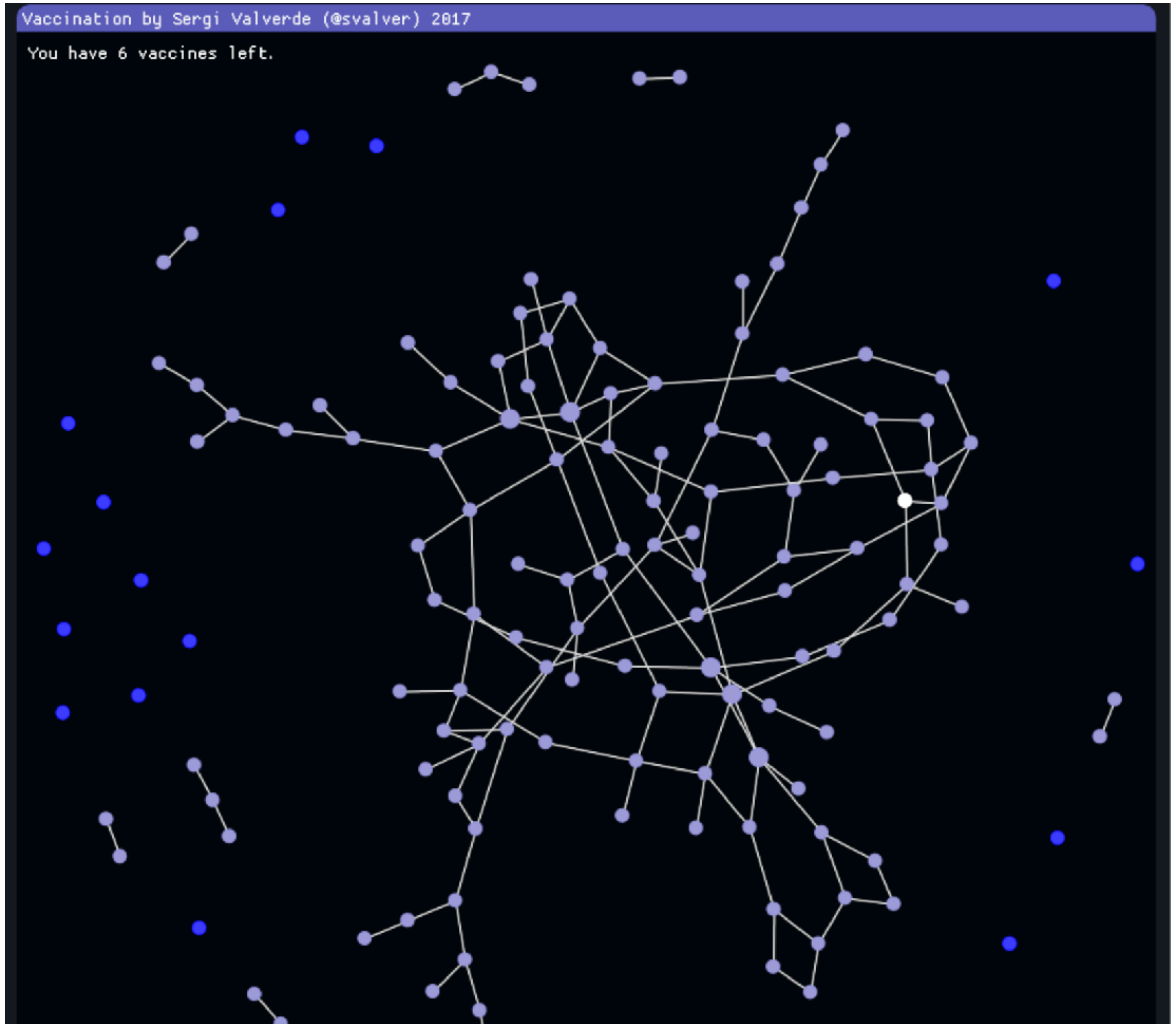
Ricard Solé[1,2,3,4] and Sergi Valverde[4,5]

Valverde and Solé, **Physical Review E** (2005)

Solé and Valverde, **Trends Eco Evol** (2006)

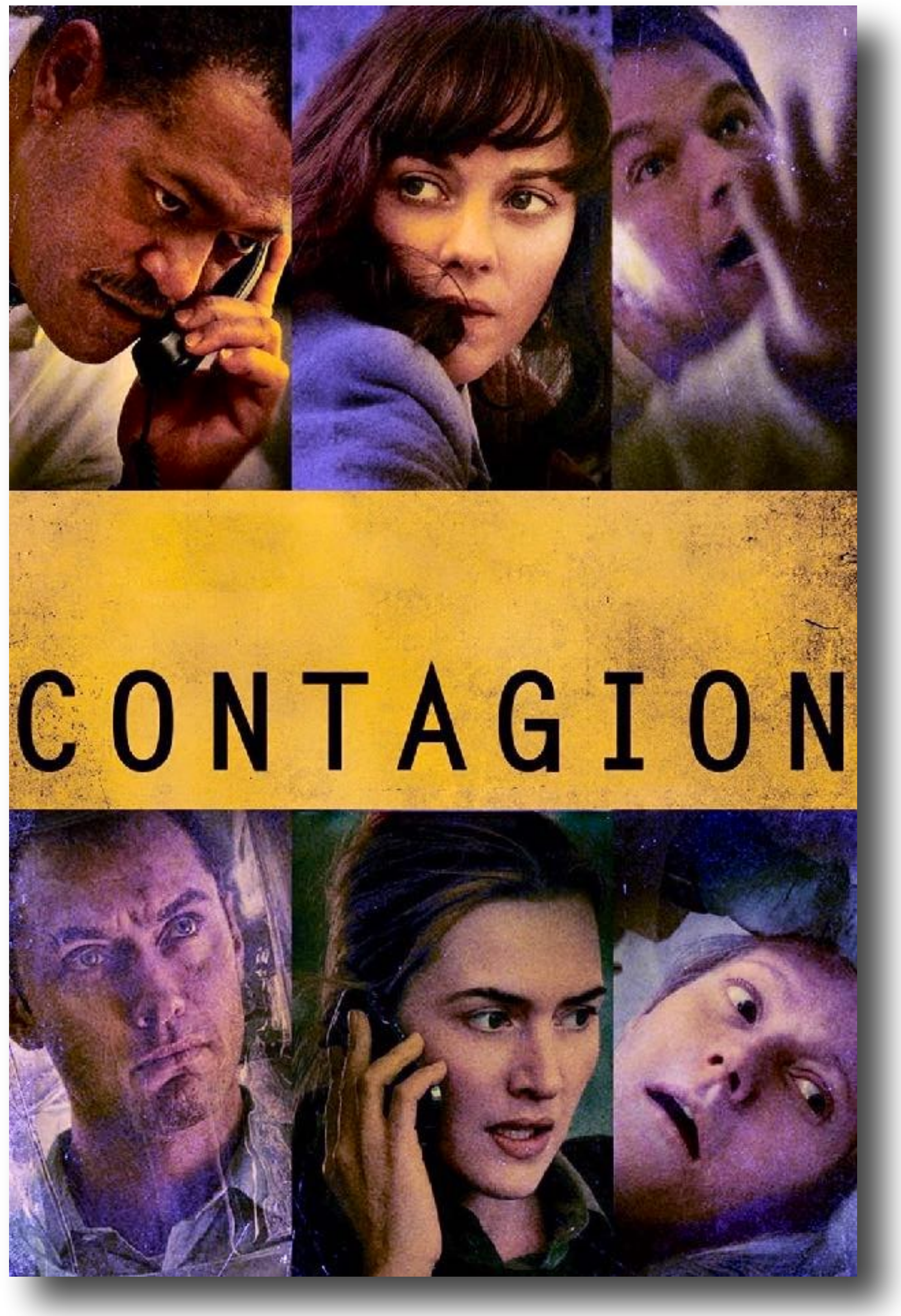
Vaccination Game

https://tinyurl.com/c42yx3pc



Can you control an epidemic?

Take action to prevent the spread of illness in various urban settings. After a small amount of vaccinations have been distributed, the epidemic continues to spread, and the players must act quickly to isolate everybody who could be sick.



NOTE: This game was designed in 2017.

Summary

Networks are the language of complexity.

Many real systems are close to the percolation transition.

Networks evidence multiple evolutionary mechanisms.

A good model explains multi-scale network features.

Complexity emerges from simplicity.

“The future cannot be predicted, but futures can be invented”

–Dennis Gabor (Hungarian physicist)