系统生物学

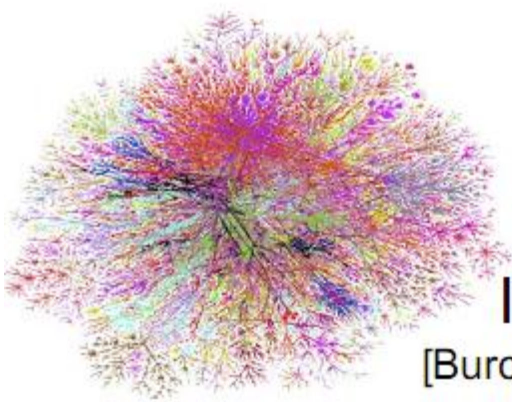
(Systems Biology)

马彬广

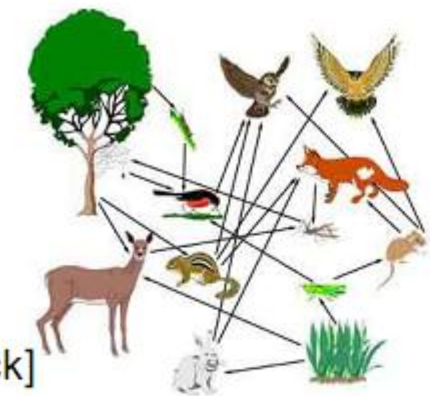
网络模型基础

（第三讲）

Networks as a universal language

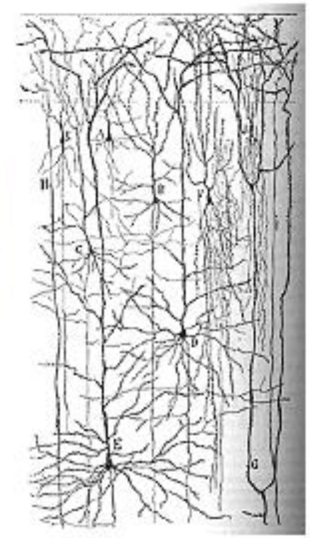


Internet
[Burch & Cheswick]

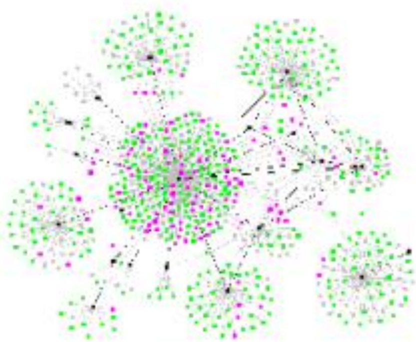


Food Web

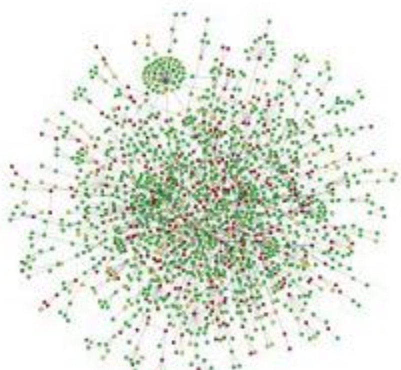
Electronic
Circuit



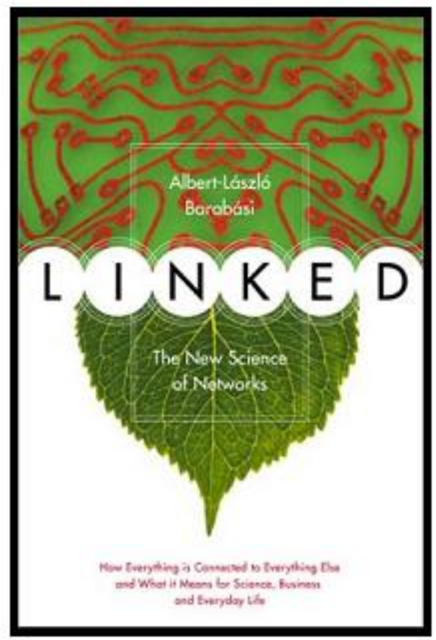
Neural Network
[Cajal]

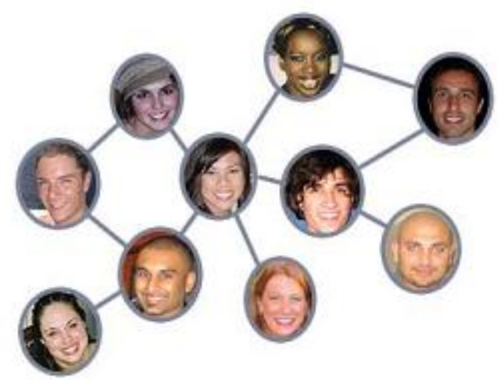


Disease
Spread
[Krebs]

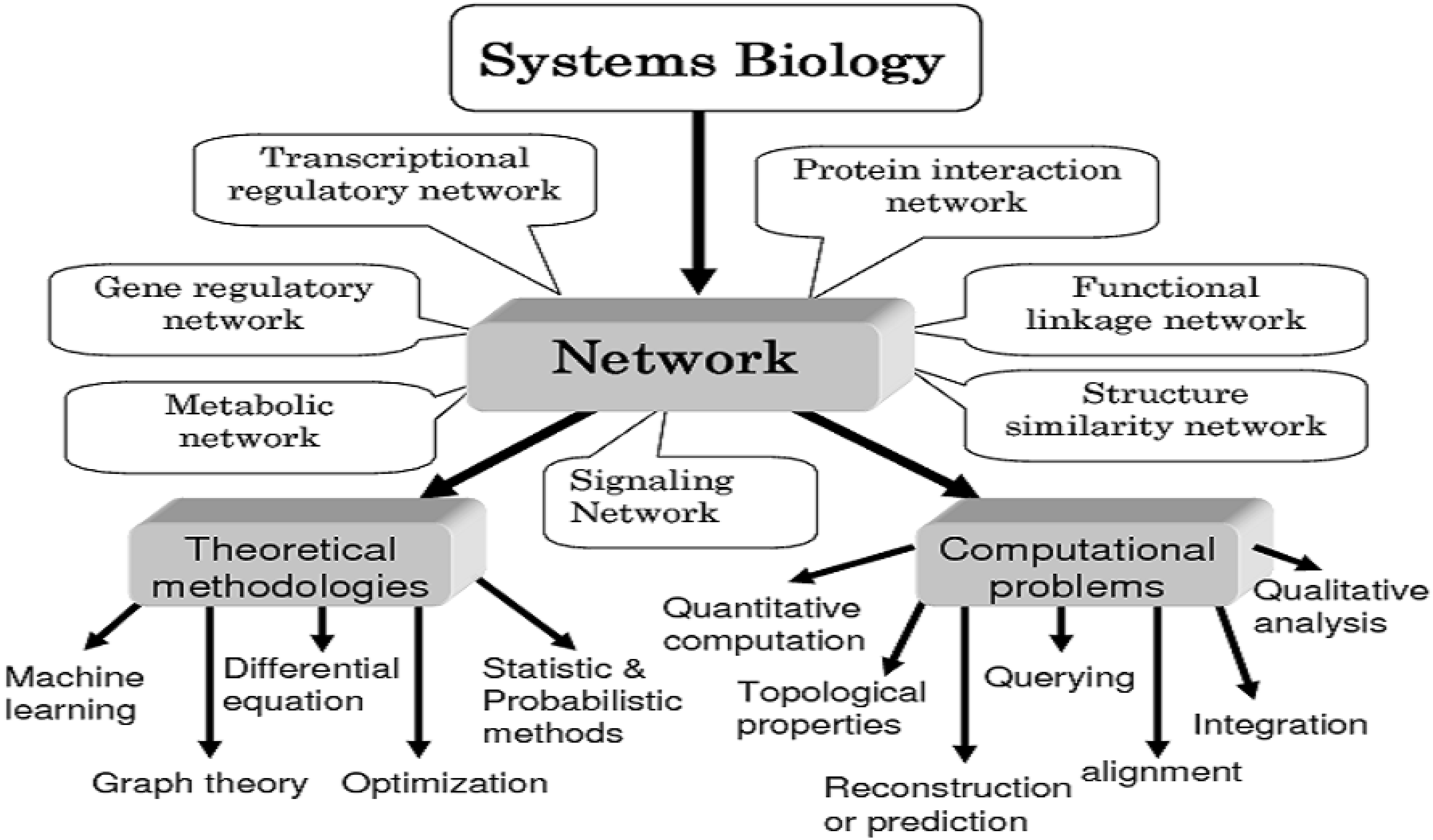


Protein
Interactions
[Barabasi]



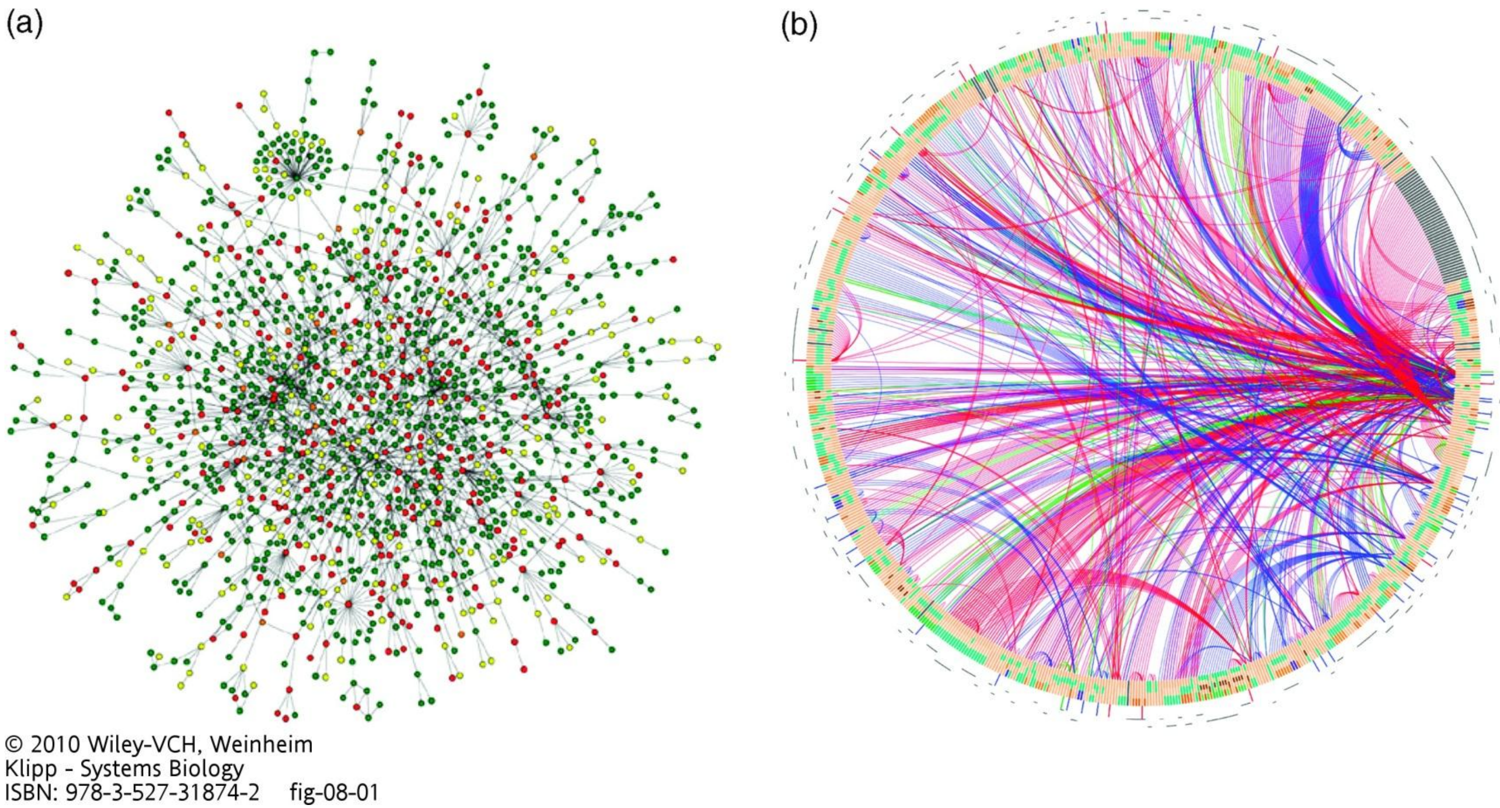


Social Network



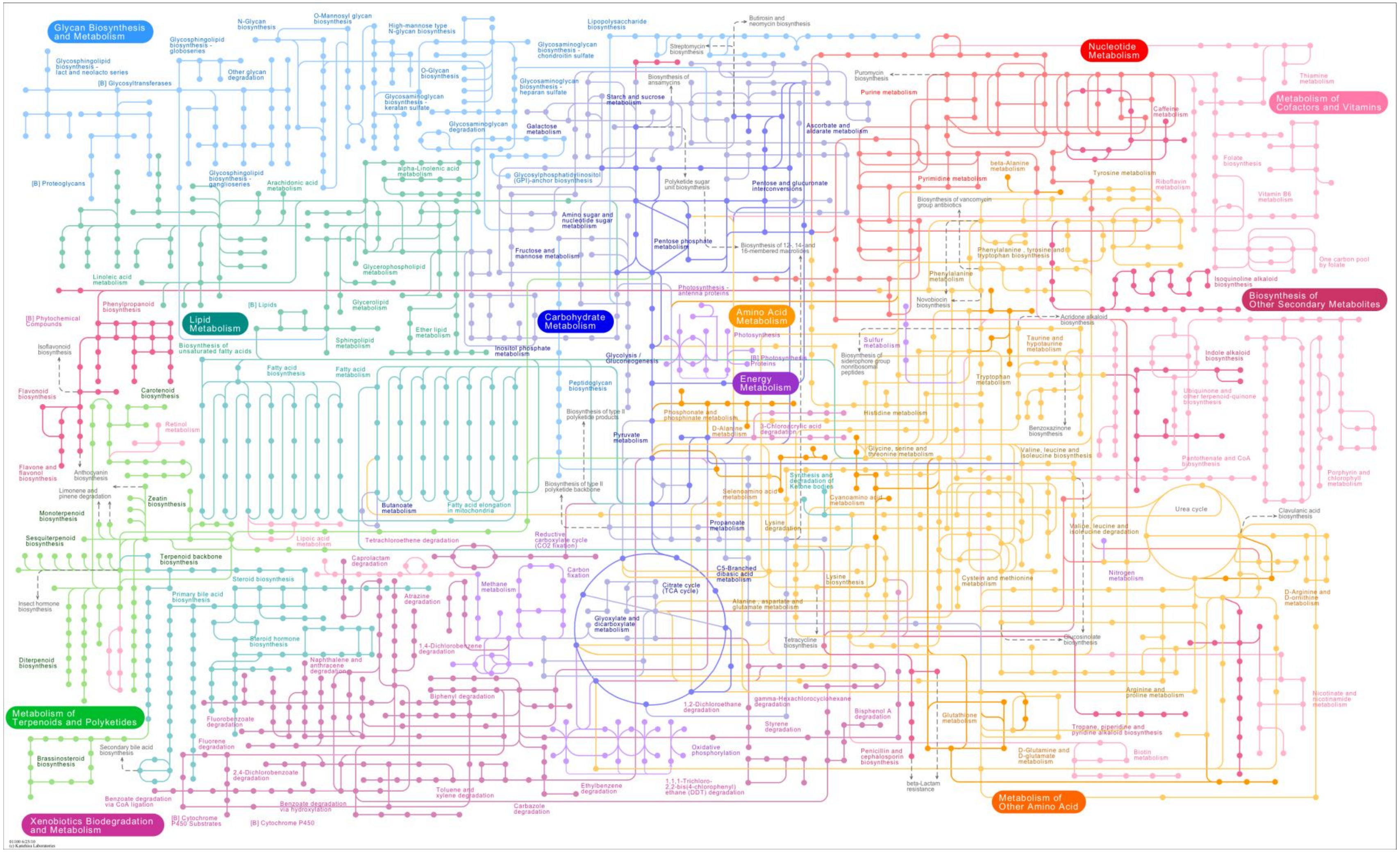
Network Systems Biology

蛋白互作网络与基因调控网络



(a) Network of protein–protein interactions in yeast. From Jeong et al. (b) Regulatory interactions between E. coli genes. Courtesy of S. Ortiz, L. Rico, and A. Valencia.

代谢网络



Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid
Citrate cycle (TCA cycle)
Glycolysis / Gluconeogenesis
Pentose phosphate metabolism
Urea cycle

图论

格尼斯堡七桥问题

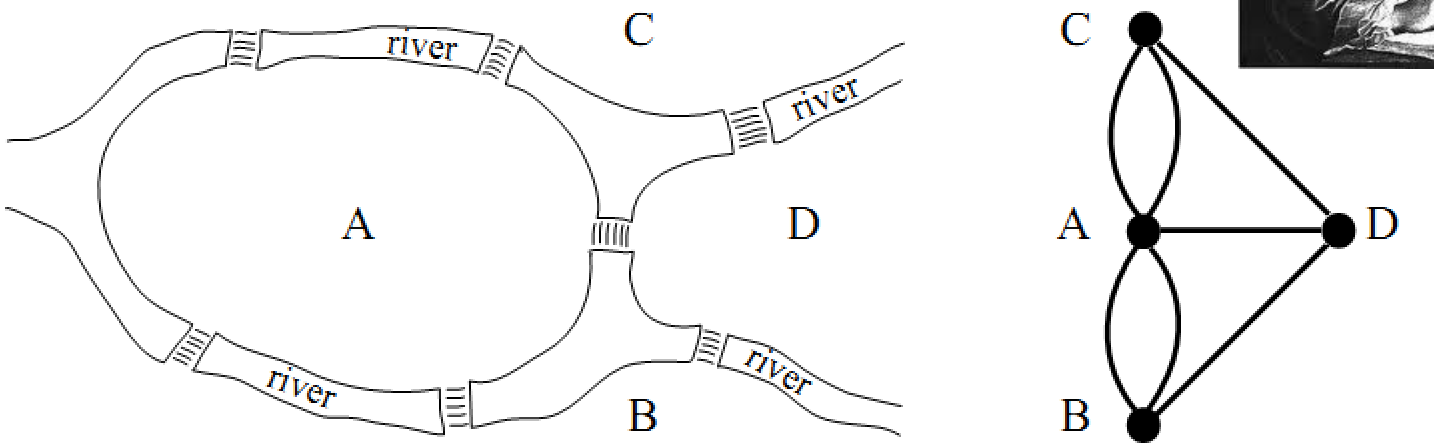
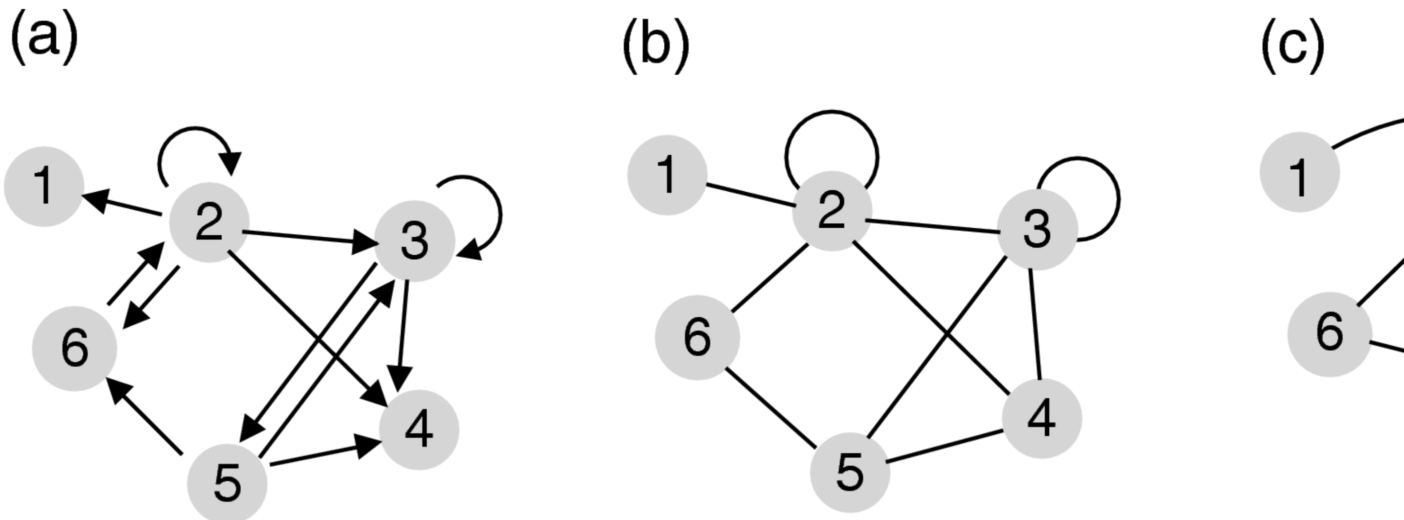


FIGURE 2.1 A map of Königsberg with the river Pregel and the representation of the "Königsberg bridge problem" problem as a graph.

Taken from the book "analysis of biological networks"

Mathematical Graph



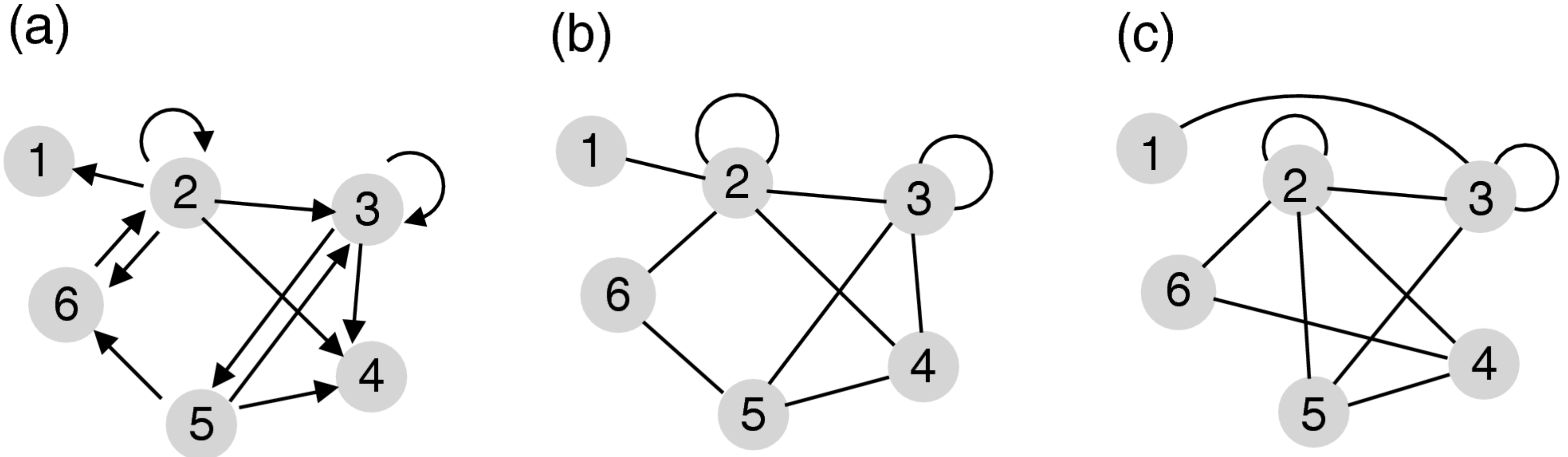
Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-08-02

(a) Directed graph with 6 nodes and 9 edges. (b) An undirected graph with similar topology. (c) By rewiring, we can obtain a new graph without changing the degrees k_i.

Basic Notions for Graph and Network

- Graph: G = {V, E}, V, a set of vertices (nodes), E, a set of edges. (used in Math Theory)
- Network: directed graph with weights for edges. (used in practice)
- Usually: the two concepts Graph and Network are used interchangeably.
- Directed or Undirected: a edge is a ordered or order-less pair of nodes.
- Neighbors: the nodes directly connected to the current node.
- Order and Size: the number of nodes and the number of edges.
- Degree: for undirected, the degree for a node is the number its edges; for directed graph, two kinds of degree, the incoming degree is the number of edges from other nodes, the outgoing degree is the number of edges to other nodes.

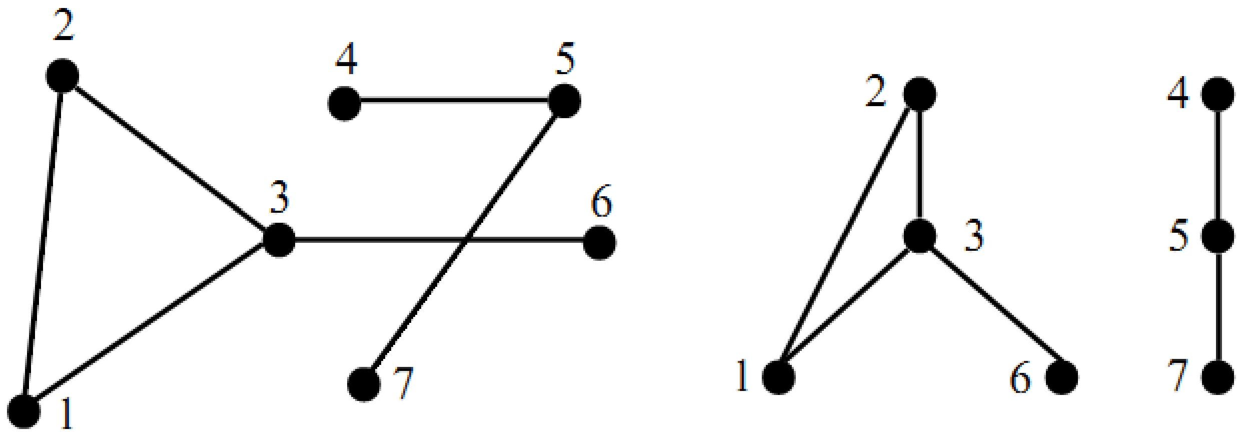
Basic Notions for Graph and Network



Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-08-02

- Loop: an edge start from and end with the same node.
- Path: a sequence of edges leads one node to another node.
- Acyclic graph: a directed graph without cycles (a path starting and ending at the same node).

例图：



邻接矩阵表示：

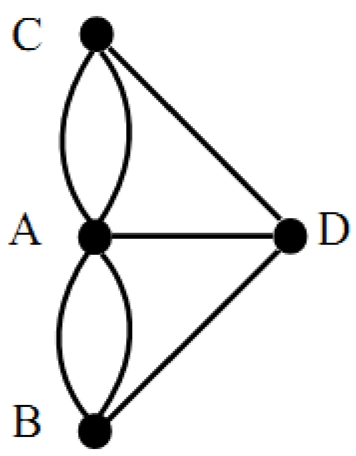
vertex	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	0	1	1	0	0	0	0
2	1	0	1	0	0	0	0
3	1	1	0	0	0	1	0
4	0	0	0	0	1	0	0
5	0	0	0	1	0	0	1
6	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0

邻接表：

L_1: ({1, 2}, {1, 3})
L_2: ({2, 1}, {2, 3})
L_3: ({3, 1}, {3, 2}, {3, 6})
L_4: ({4, 5})
L_5: ({5, 4}, {5, 7})
L_6: ({6, 3})
L_7: ({7, 5})

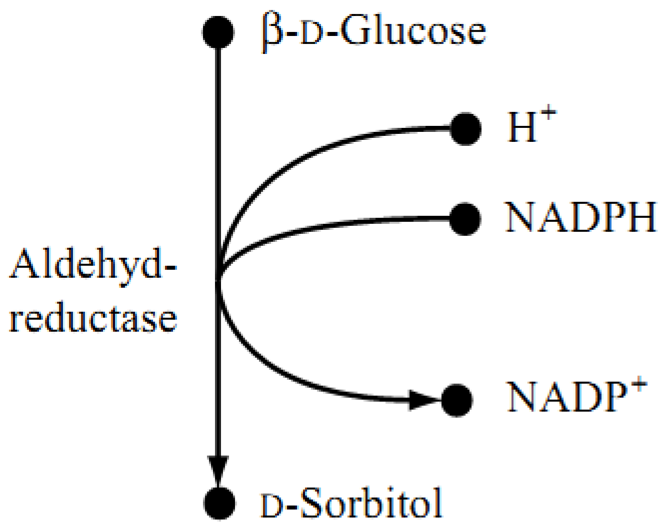
多图、超图、两分图

多图



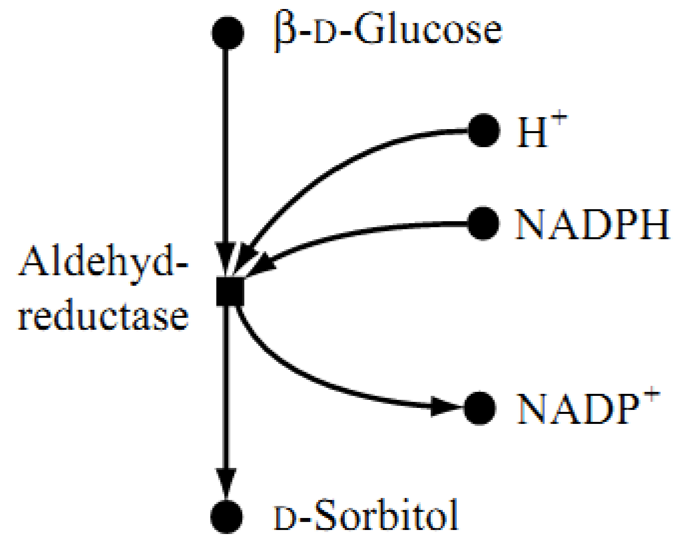
Multigraph

超图



Hypergraph

两分图



Bipartite graph

度量网络特征的三个常用指标

- 度分布（Degree Distribution）: the number (or probability) of nodes with the degree *k*.

$$N(k) \sim k \quad \text{or} \quad p(k) \sim k.$$

- 簇系数（Clustering Coefficient）: for a node, the ratio of the numbers of the connections between its neighbors to the number of all possible connections. Average CC over all nodes is a measurement of clustering.

$$C_i = \frac{e_i}{\frac{k_i(k_i-1)}{2}} = \frac{2e_i}{k_i(k_i-1)}; \quad C = \frac{1}{N}\sum_{i=1}^{N} C_i$$

- 平均路径长度（Average Path Length）： path length is the minimum numbers (sum of weights) of edges from one node to another nodes. （网络直径，longest path）

$$L = \frac{1}{N(N-1)}\sum_{i\neq j} d_{ij}, \quad d_{ij} \text{ is the path length from node i to node j.}$$

中心性（centrality）指标

点度中心性（degree centrality）：节点的连接数，衡量一个节点的连接能力。

$$C_D(v) = \frac{\deg(v)}{n-1}, \quad \text{n is the number of nodes.}$$

介度中心性（betweenness centrality）：通过节点的最短路径的个数，衡量节点承载流量的能力。

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

接近中心性（closeness centrality）：节点到其它节点最短路径的平均值，衡量传递速度。

$$C_C(v) = \frac{\sum_{t \in V \setminus v} d(v,t)}{n-1}$$

基本网络举例

- 规则网络（Regular Networks）
- 随机网络（Random Networks）
- 小世界网络（Small-World Networks）
- 无标度网络（Scale-Free Networks）
- 层级网络（Hierarchical Networks）

http://networkx.lanl.gov/

NetworkX

NetworkX Home | Download | Developer Zone | Documentation | Blog »

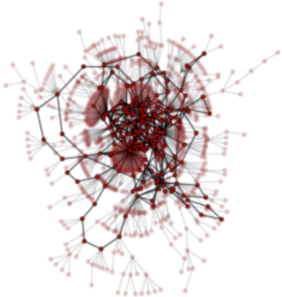
High productivity software for complex networks

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Quick Example

```
>>> import networkx as nx

>>> G=nx.Graph()
>>> G.add_node("spam")
>>> G.add_edge(1,2)
>>> print(G.nodes())
[1, 2, 'spam']
>>> print(G.edges())
[(1, 2)]
```



Documentation

Tutorial
start here

Reference
guide to all functions and classes

Examples
using the library

Gallery
network drawings

Contents
a complete overview

Search Page
search the documentation

General Index
all functions, classes, terms

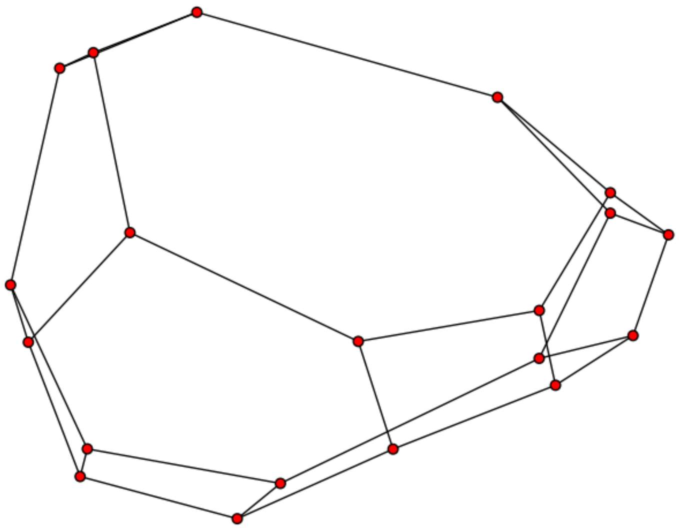
Module Index
quick access to all documented modules

Features

- Standard graph-theoretic and statistical physics functions
- Easy exchange of network algorithms between applications, disciplines, and platforms
- Many classic graphs and synthetic networks
- Nodes and edges can be "anything" (e.g. time-series, text, images, XML records)

Current version: **1.8**

Regular Network



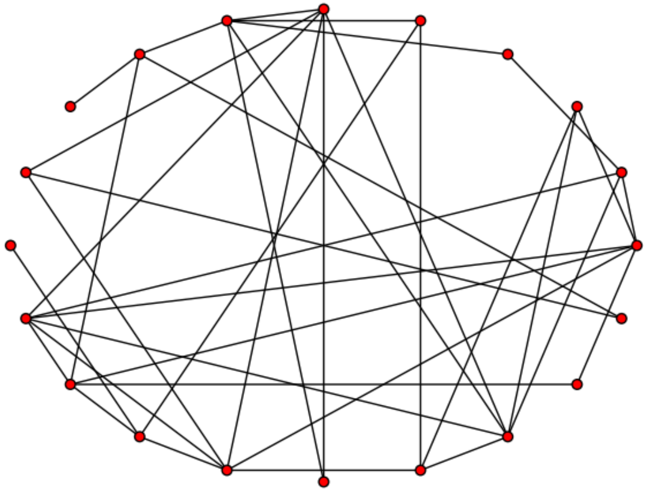
每个节点与固定数目的K个节点连接，度分布为：

$$P(k)=\delta(k-K)$$

集聚程度高，平均路径长。

在NetworkX中，用random_graphs.random_regular_graph(d, n)方法可以生成一个含有n个节点，每个节点有d个邻居节点的规则图。

ER (Erdös-Rényi) Network



由N个节点构成的图，可以存在

$$C_N^2 = \frac{N(N-1)}{2}$$

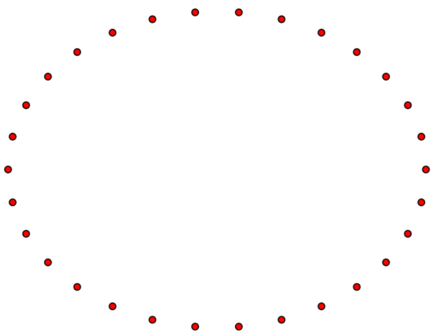
从中随机连接M条边所构成的网络就是随机网络。ER网络节点的连接度服从泊松分布：

$$P(k) = \frac{<k>^k}{k!} e^{<k>}$$

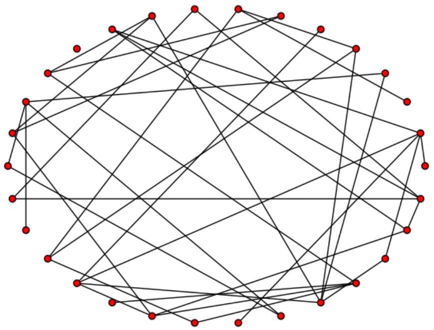
ER随机图是早期研究得比较多的一类“复杂”网络，这个模型的基本思想是以概率p连接N个节点中的每一对节点。在NetworkX中，可以用random_graphs.erdos_renyi_graph(n,p)方法生成一个含有n个节点、以概率p连接的ER随机图。

Dependency on p

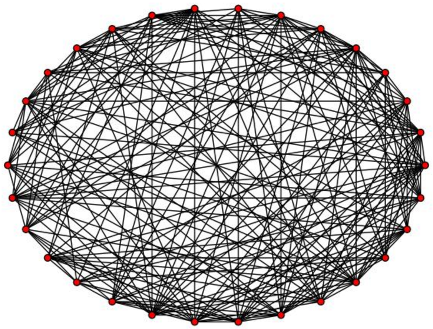
$p = 0.0$



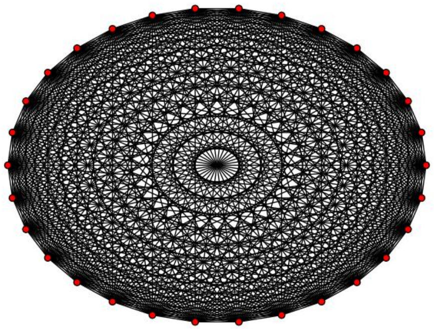
$p = 0.1$



$p = 0.5$

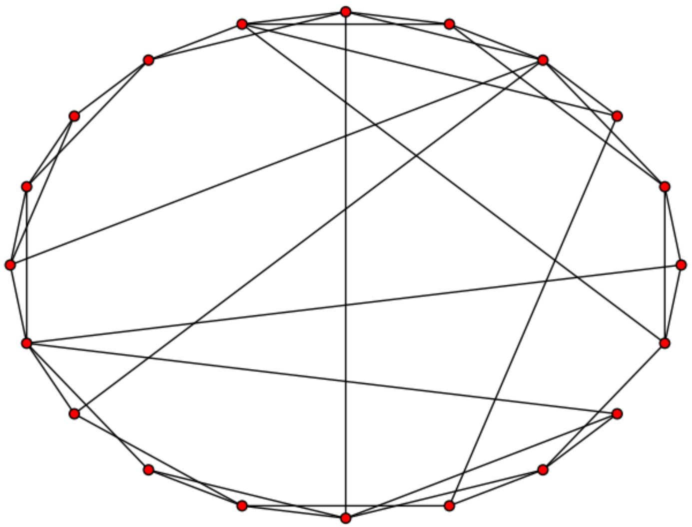


$p = 1.0$



WS (Watts-Strogatz) Small World Network

小世界概念是哈佛大学心理学家米尔格伦提出的六度分离（six degrees of separation）概念的推广。

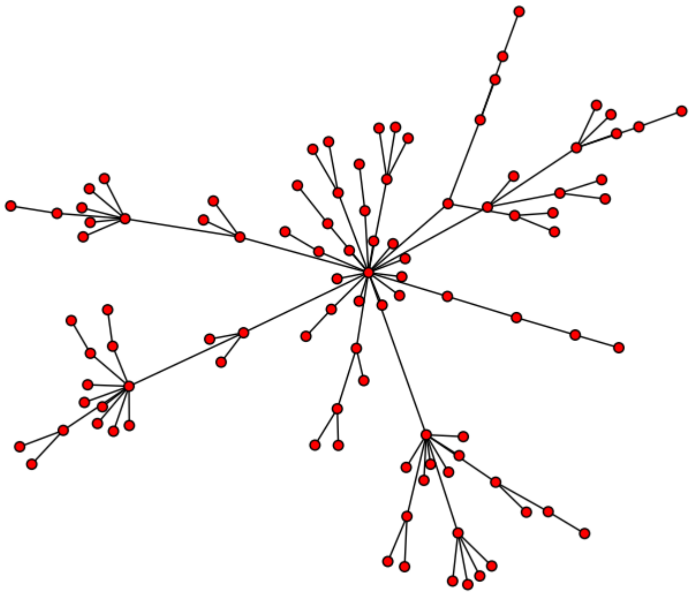


从N节点环开始，环上每个节点，与两侧各有m条边相连，然后，对每条边以概率p随机进行重新连接（自我连接和重复连接除外），这些重新连接的边称为“长程连接”。长程连接大大减小了网络的平均路径长度。

在NetworkX中，可以用random_graphs.watts_strogatz_graph(n, k, p)方法生成一个含有n个节点、每个节点有k个邻居、以概率p随机化重连边的WS小世界网络.

小世界概念描述了，现实世界中，有的网络尽管规模很大，但任意两个节点之间却存在一条很短的路径这样一个事实。

BA (Barabasi-Albert) Scale Free Network



无特征标度，连接度满足幂分布（power law），

$$P(k) \sim k^{\gamma} \text{ or } P(k) = ck^{-\gamma}$$

无标度的含义：

比较幂函数 $y(x) = cx^{a}$ 和 指数函数 $z(x) = ce^{-x}$

改变测量单位，即乘以因子λ， 有

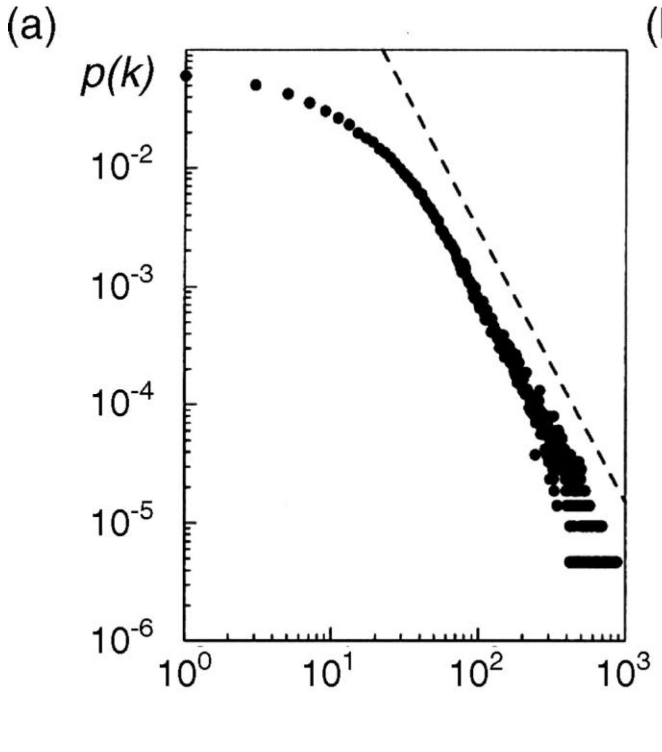
$y(\lambda x) = c(\lambda x)^{a} = \lambda^{a} cx^{a} = \lambda^{a} y(x)$

$z(\lambda x) = ce^{-\lambda x} = c(e^{\lambda})^{-x}$

前者函数形式不变，后者改变。

在NetworkX中，可以用random_graphs.barabasi_albert_graph(n, m)方法生成一个含有n个节点、每次加入m条边的BA无标度网络.

无标度网络的度分布举例



Klipp - Systems Biology
ISBN: 978-3-527-31874-2 fig-08-03

Scale-free degree distribution in different networks . (a) Movie actors (n =212250, $\langle k \rangle = 28.78$, $\gamma = 2.3$). (b) World wide web (n =325729, $\langle k \rangle = 5.46$, $\gamma = 2.1$). (c) Power grid ($n = 4941$, $\langle k \rangle = 2.67$, $\gamma = 4$). From Barabási and Albert (Science, 1999, 286: 509) .

随机网络与无标度网络的区别

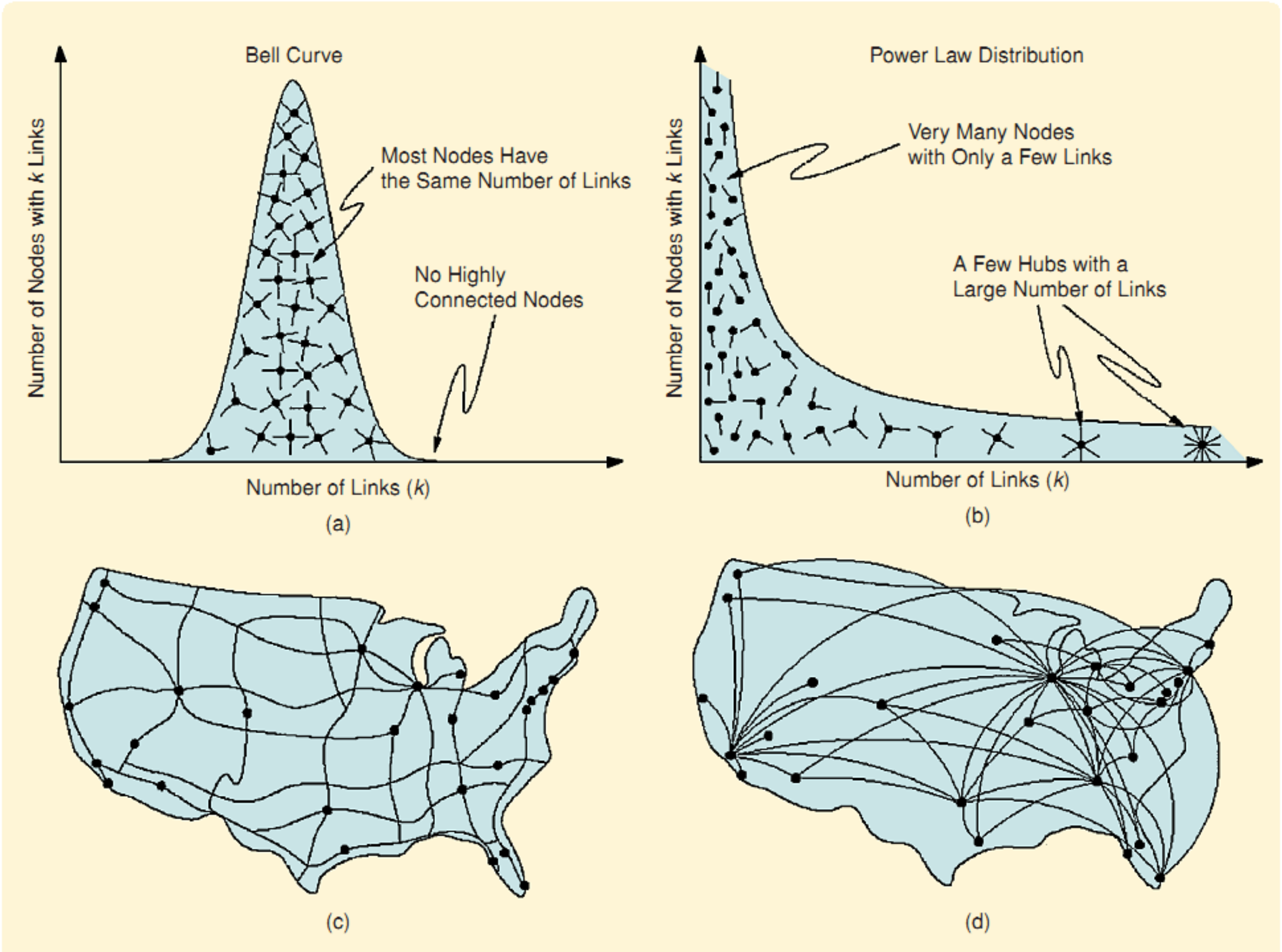


FIGURE 1 Random and scale-free networks. The degree distribution of a random network follows a Poisson distribution close in shape to the bell curve, telling us that most nodes have the same number of links, and that nodes with a large number of links don't exist (a). Thus, a random network is similar to a national highway network in which the nodes are the cities and the links are the major highways connecting them. Indeed, most cities are served by roughly the same number of highways (c). In contrast, the power-law degree distribution of a scale-free network predicts that most nodes have only a few links held together by a few highly connected hubs (b). Such a network is similar to the air traffic system, in which a large number of small airports are connected to each other by means of a few major hubs (d). After [1].

生成基本网络的函数

GenerateGraphs.py - E:/mbg/files/reports/系统生物学讲义/Refs/GenerateGraphs.py

File Edit Format Run Options Windows Help

```
#生成常用的网络（规则网络、ER随机网络、WS小世界网络、BA无标度网络）

import networkx as nx
import matplotlib.pyplot as plt

#生成规则网络
RG = nx.random_graphs.random_regular_graph(3,20)   #生成包含20个节点、每个节点有3个邻居
pos = nx.spectral_layout(RG)            #定义一个布局，此处采用了spectral布局方式，后变还会
nx.draw(RG,pos,with_labels=False,node_size = 30)   #绘制规则图的图形，with_labels决定
plt.show()   #显示图形

#生成ER随机网络
ER = nx.random_graphs.erdos_renyi_graph(20,0.2)   #生成包含20个节点、以概率0.2连接的随机
pos = nx.shell_layout(ER)             #定义一个布局，此处采用了shell布局方式
nx.draw(ER,pos,with_labels=False,node_size = 30)
plt.show()

#生成WS小世界网络
WS = nx.random_graphs.watts_strogatz_graph(20,4,0.3)   #生成包含20个节点、每个节点4个近
pos = nx.circular_layout(WS)            #定义一个布局，此处采用了circular布局方式
nx.draw(WS,pos,with_labels=False,node_size = 30)   #绘制图形
plt.show()

#生成BA无标度网络
BA= nx.random_graphs.barabasi_albert_graph(100,1)   #生成n=20、m=1的BA无标度网络
pos = nx.spring_layout(BA)           #定义一个布局，此处采用了spring布局方式
nx.draw(BA,pos,with_labels=False,node_size = 30)   #绘制图形
plt.show()
```

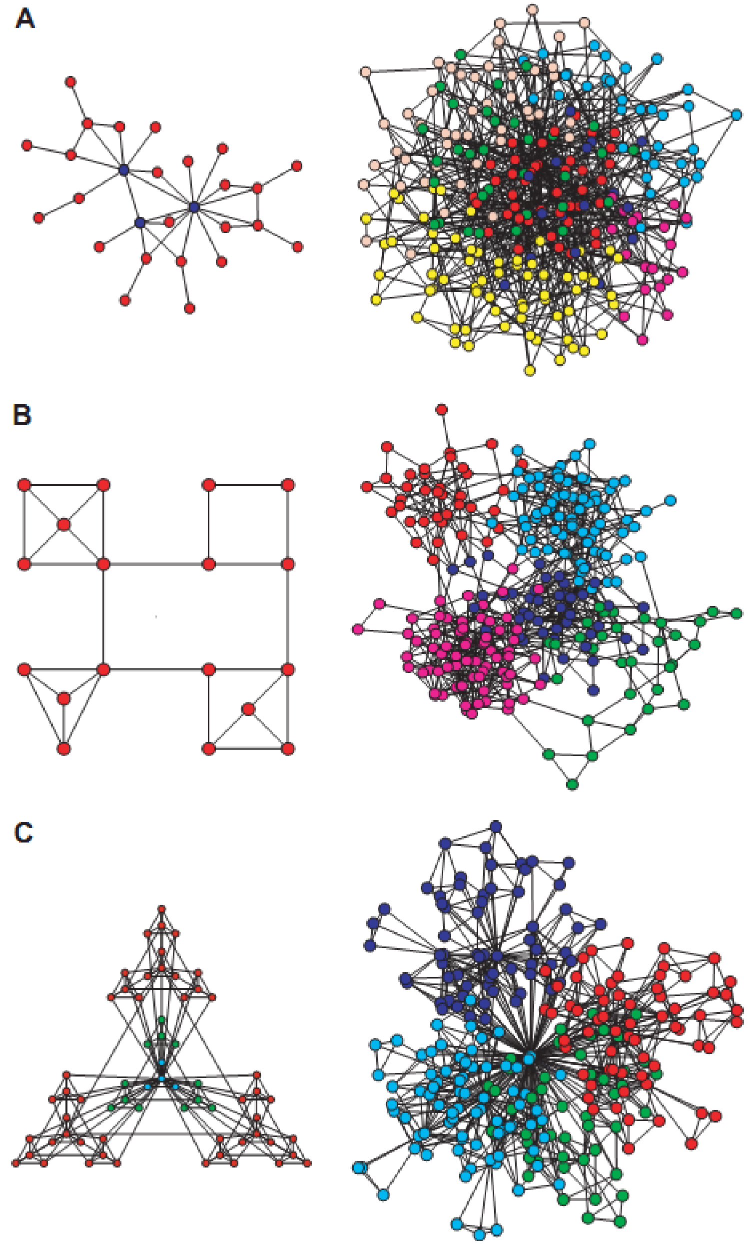
Ln: 31 Col: 0

层级网络

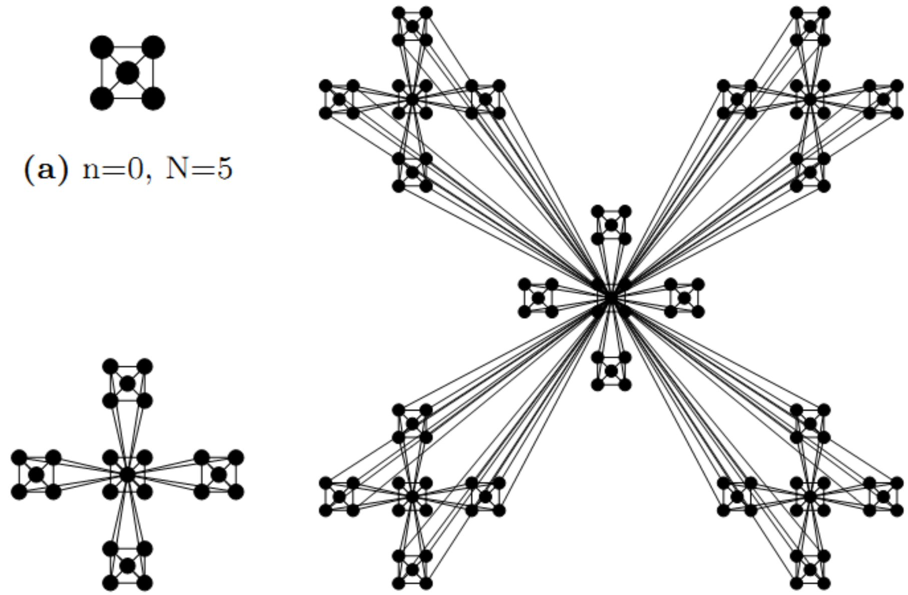
无标度性与模块性的复合

Science (2002) 297: 1551

Fig. 1. Complex network models. (**A**) A schematic illustration (left) of a scale-free network, whose degree distribution follows a power law. In such a network, a few highly connected nodes, or hubs (blue circles), play an important role in keeping the whole network together. A typical configuration (right) of a scale-free network with 256 nodes is also shown, obtained using the scale-free model, which requires the addition of a new node at each time such that existing nodes with higher degrees of connectivity have a higher chance of being linked to the new nodes (*12*). The nodes are arranged in space with a standard clustering algorithm (*30*) to illustrate the absence of an underlying modularity. (**B**) Schematic illustration (left) of a manifestly modular network made of four highly interlinked modules connected to each other by a few links. This intuitive topology does not have a scale-free degree distribution, as most of its nodes have a similar number of links, and hubs are absent. A standard clustering algorithm uncovers the network's inherent modularity (right) by partitioning a modular network of $N = 256$ nodes into the four isolated structures built into the system. (**C**) The hierarchical network (left) has a scale-free topology with embedded modularity. The hierarchical levels are represented in increasing order from blue to green to red. Standard clustering algorithms (right) are less successful in uncovering the network's underlying modularity. A detailed quantitative characterization of the three network models is available in (*16*).



层级网络的生成



(a) n=0, N=5

(b) n=1, N=25

(c) n=2, N=125

FIG. 1: The iterative construction leading to a hierarchical network. Starting from a fully connected cluster of five nodes shown in **(a)** (note that the diagonal nodes are also connected – links not visible), we create four identical replicas, connecting the peripheral nodes of each cluster to the central node of the original cluster, obtaining a network of $N = 25$ nodes **(b)**. In the next step we create four replicas of the obtained cluster, and connect the peripheral nodes again, as shown in **(c)**, to the central node of the original module, obtaining a $N = 125$ node network. This process can be continued indefinitely.

层级网络的度分布与尺度变换特性

$$C(k) \sim k^{-1}$$

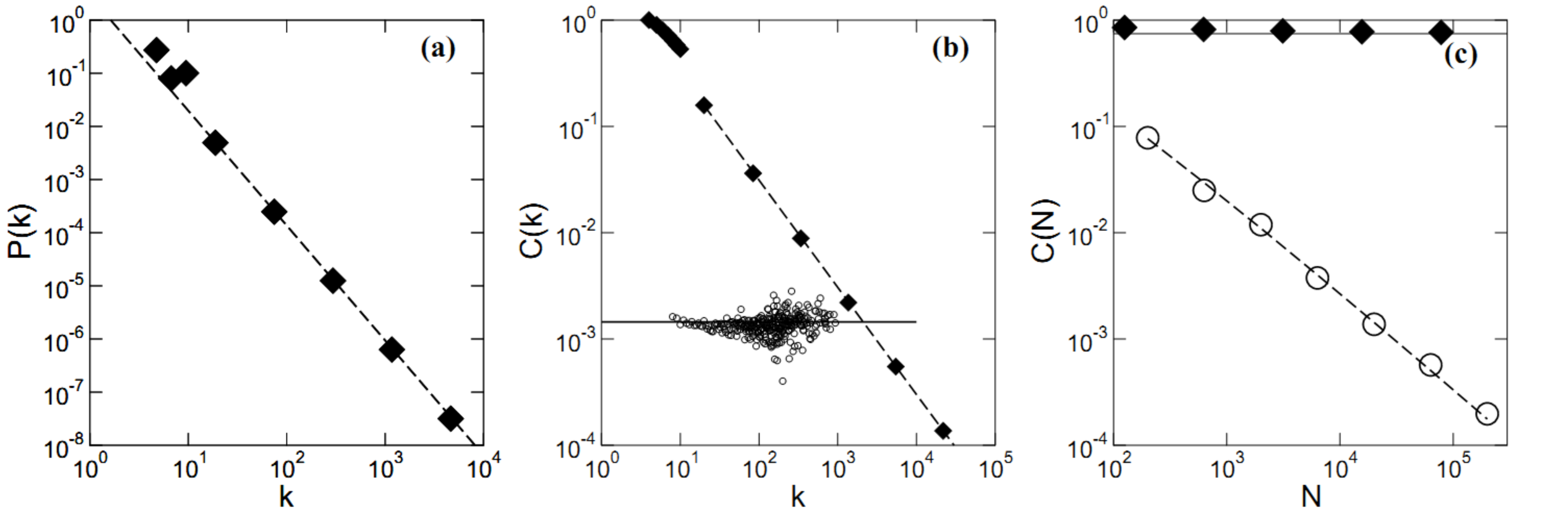


FIG. 2: Scaling properties of the hierarchical model shown in Fig. 1 ($N = 5^7$). **(a)** The numerically determined degree distribution. The assymptotic scaling, with slope $\gamma = 1 + \ln 5/\ln 4$, is shown as a dashed line. **(b)** The $C(k)$ curve for the model, demonstrating that it follows Eq. (1). The open circles show $C(k)$ for a scale-free model [12] of the same size, illustrating that it does not have a hierarchical architecture. **(c)** The dependence of the clustering coefficient, C, on the size of the network N. While for the hierarchical model C is independent of N (♦), for the scale-free model $C(N)$ decreases rapidly (○).

arXiv:cond-mat/0206130v2