WS小世界与BA无标度

网络建模：从规则到随机到规则的随机

Statistical mechanics of complex networks, Réka Albert and Albert-László Barabási
Network Science: An Introduction, Xiaofan Wang, Xiang Li and Guanrong Chen

2015-04-08 王佳晟

目录

- 几个网络的基本拓扑性质
- 规则图与ER随机图
- WS小世界网络
- BA无标度网络
- 小结
- 编程实现中的发现

拓扑性质

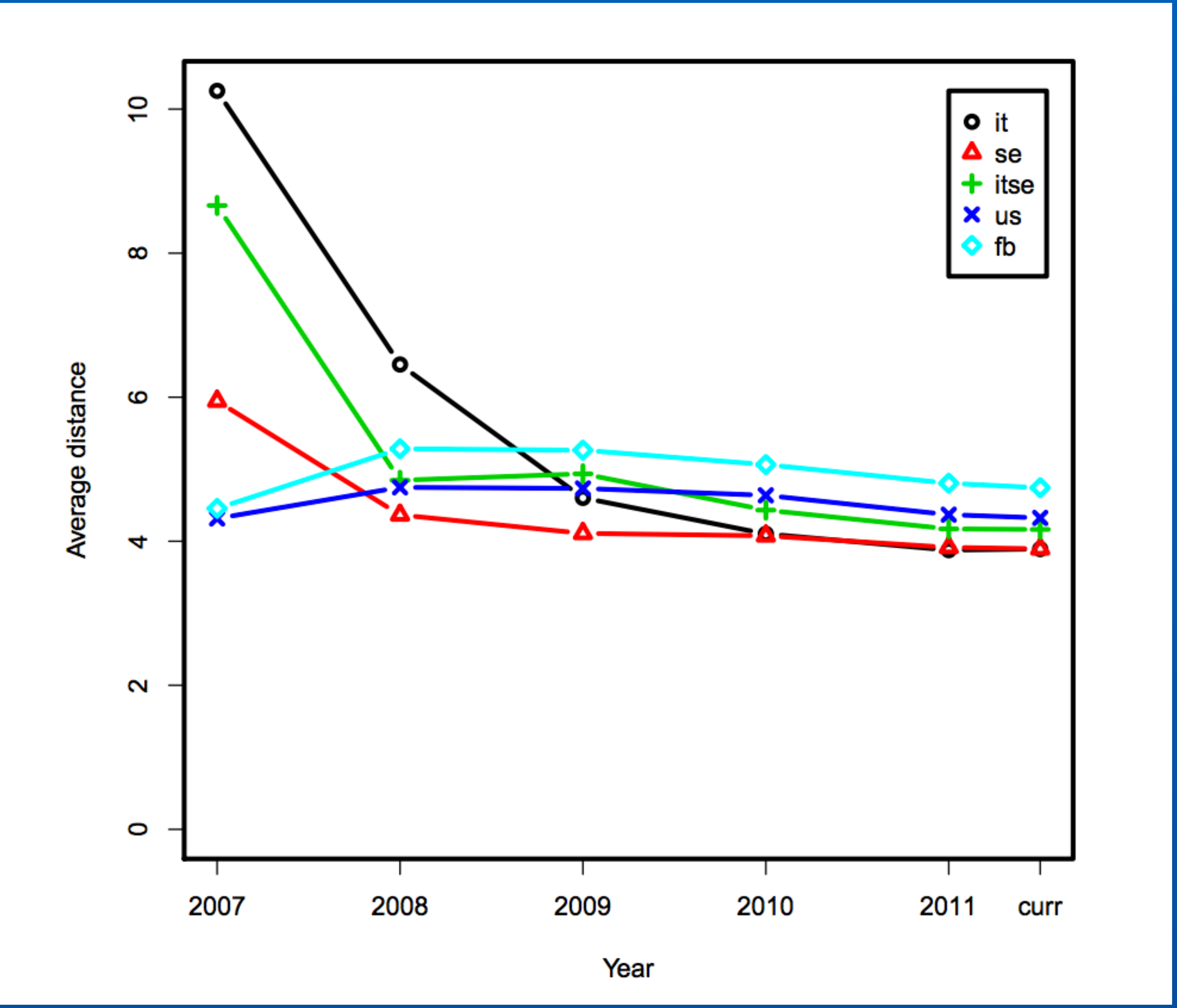
- 平均路径长度（Average Path Length）
- 聚类系数（Clustering Coefficient）
- 度分布（Degree Distribution）

网络中两个节点i和j之间的距离定义为连接这两个节点的最短路径上的边的数目。

网络的平均路径长度L定义为任意两个节点之间的距离的平均值，即:

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i>j} d_{ij}$$

拓扑性质▷平均路径长度



Backstrom, Lars, et al. "Four degrees of separation." Proceedings of the 4th Annual ACM Web Science Conference. ACM, 2012.

在朋友关系网络中，你的两个朋友很可能彼此也是朋友，这种可能性的大小反映了你的朋友圈的紧密程度。我们可以使用聚类系数类定量刻画你的任意两个朋友也互为朋友的概率。

网络中一个度为k_i的节点i的聚类系数C_i定义为：

$$C_i = \frac{E_i}{(k_i(k_i-1))/2} = \frac{2E_i}{k_i(k_i-1)}$$

其中E_i是节点i的k_i个邻节点之间实际存在的边数。

从另一个角度来阐述定义

$$C_i = \frac{E_i}{(k_i(k_i - 1))/2}$$

E_i：包含节点i的三角形数目

$(k_i(k_i - 1))/2$：以节点i为中心的连通三元组的数目

拓扑性质▷聚类系数

$$A=(a_{ij})_{N\times N}$$

$$E_i=\frac{1}{2}\sum_{j,k}a_{ij}a_{jk}a_{ki}=\sum_{k>j}a_{ij}a_{jk}a_{ki}\Rightarrow C_i=\frac{\sum\limits_{j\neq i,k\neq j,k\neq i}a_{ij}a_{jk}a_{ki}}{\sum\limits_{j\neq i,k\neq j,k\neq i}a_{ij}a_{ik}}$$

一个网络的聚类系数C定义为网络中所有节点的聚类系数的平均值，即

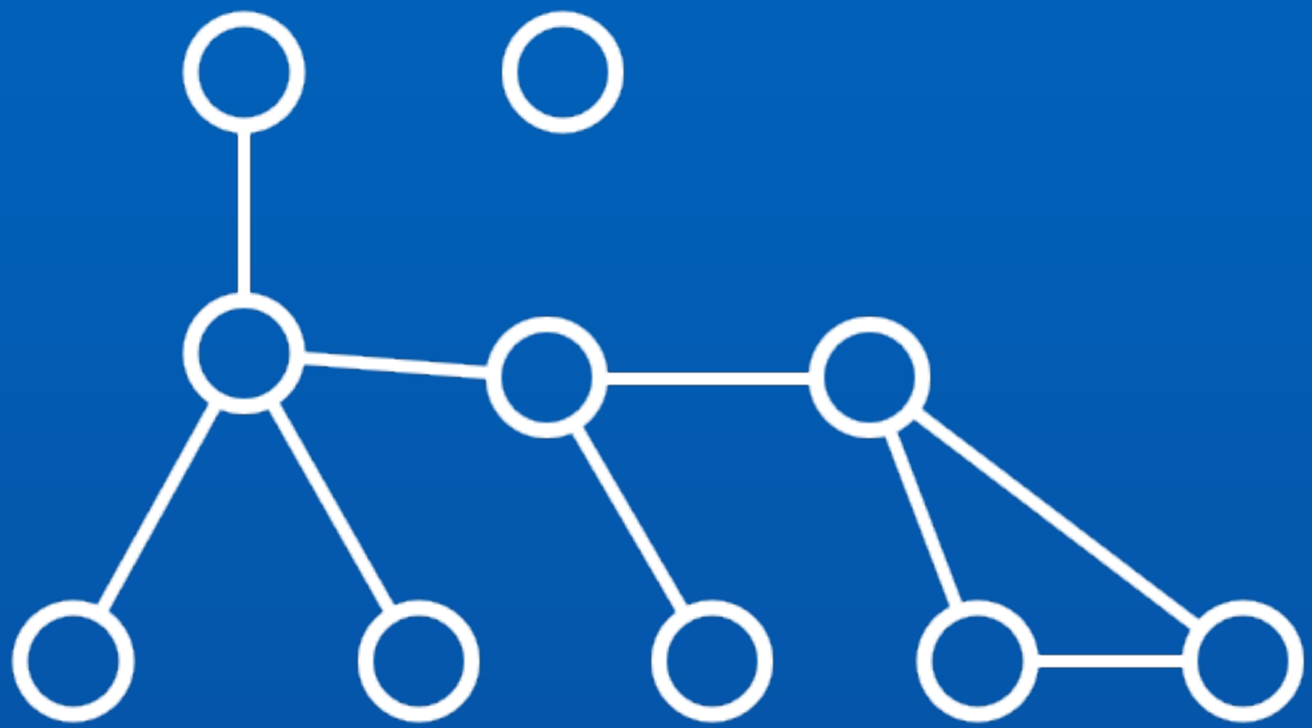
$$C=\frac{1}{N}\sum_{i=1}^{N}C_i$$

三角形的社会意义就是你的两位朋友也是朋友的话，那么3人就形成了一个三角形，所以可以用网络中三角形的相对数量来刻画网络的聚类特性，即

$$C=\frac{\text{网络中三角形的数目}}{\text{网络中连通三元组的数目}/3}$$

这里除以3是因为每个三角形对应于三个不同的连通三元组，他们分别以三角形的三个顶点为中心。

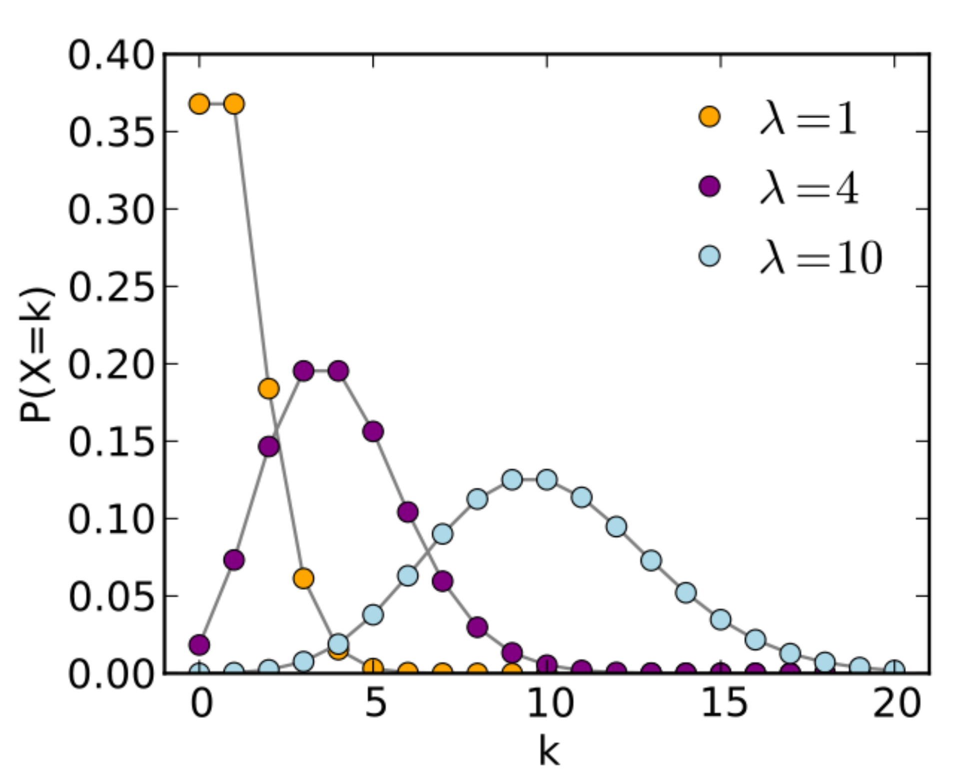
拓扑性质▷度分布

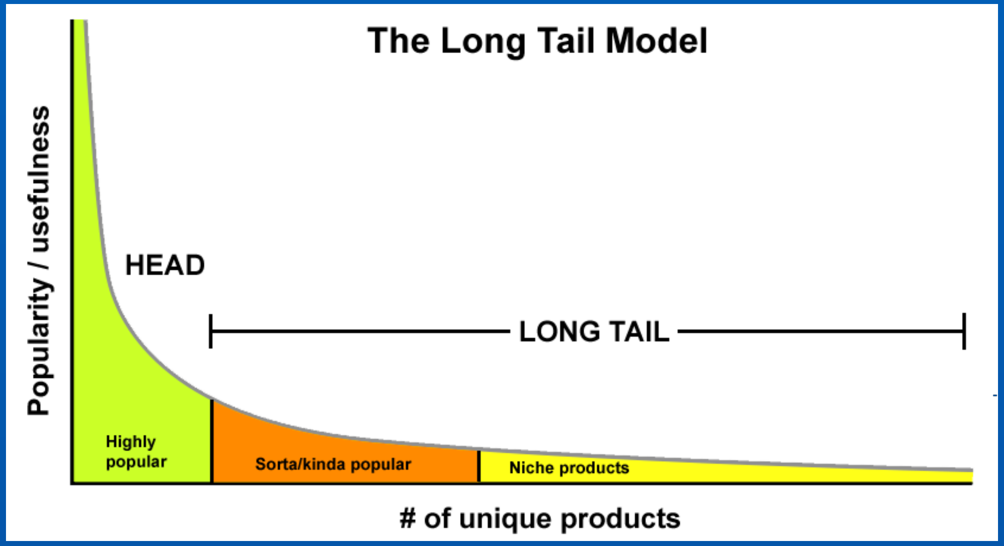


k	P(k)
0	1/10
1	2/10
2	4/10
3	2/10
4	1/10
>4	0

从概率统计的角度看，P(k)也可以视为网络中一个随机选择的节点的度为k的概率。

拓扑性质▷度分布



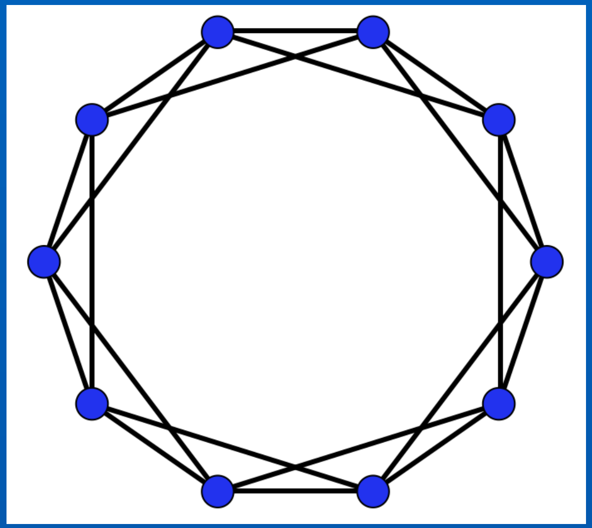


泊松分布

有标度

长尾分布

无标度



N=10, K=4

- 聚类系数

$$C=\frac{3\times N\times\binom{K/2}{2}}{N\times\binom{K}{2}}=\frac{3(K-2)}{4(K-1)}$$

- 平均路径长度

$$L\approx\frac{1}{N/2}\sum_{m}^{N/2}\lceil 2m/K\rceil\approx\frac{N}{2K}$$

- Erdös和Rényi于20世纪50年代末开始研究
- 具有固定边数的ER随机图G(N,M)
- 具有固定连边概率的ER随机图G(N,p)

1. 初始化：给定N个节点以及连边概率p∈[0, 1]
2. 随机连边：
   (1) 选择一对没有边相连的不同的节点
   (2) 生成一个随机数r∈(0, 1)
   (3) 如果r<p，那么在这对节点之间添加一条边
   (4) 重复步骤(1)~(3)，直至所有的节点对都被选择过一次

- 聚类系数

$$C=p=\langle k\rangle/(N-1)$$

- 平均路径长度

$$L \le D \sim \frac{\ln N}{\ln\langle k\rangle}$$

这种平均路径长度为网络规模的对数增长的特性
就是典型的小世界性质

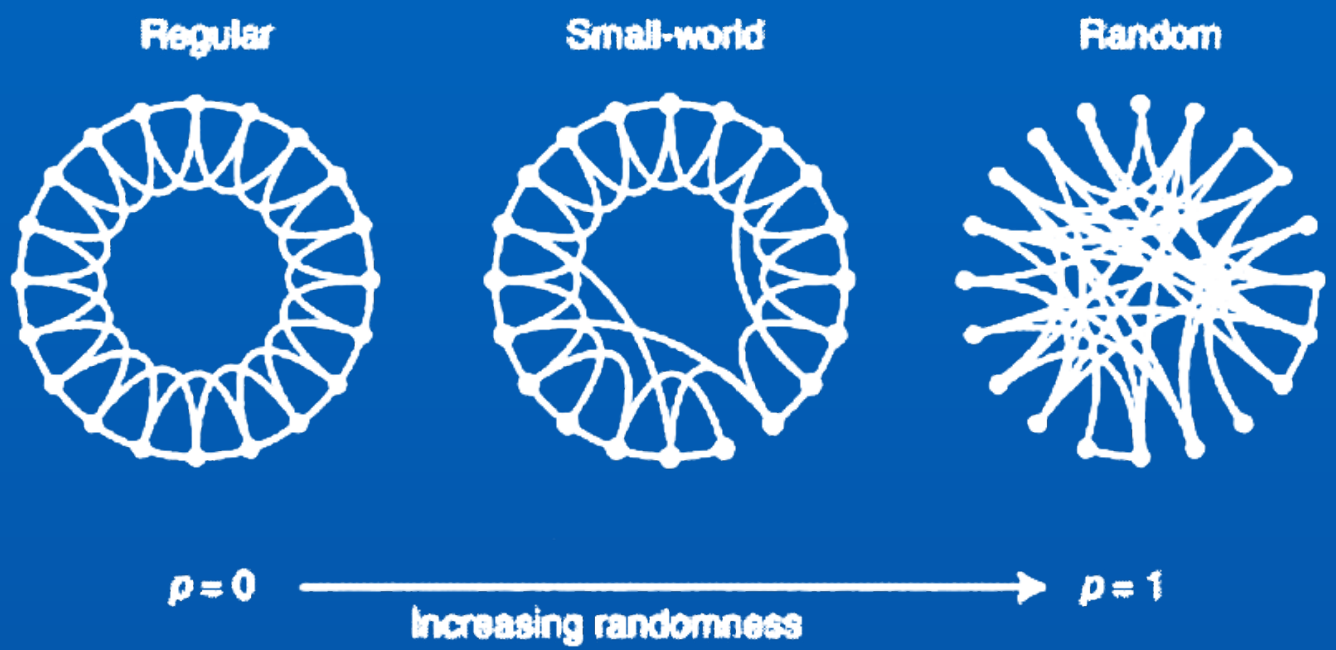
	最近邻耦合网络	ER随机图
聚类系数	$\geq 3/8$	$N \to \infty$, C很小
平均路径长度	$N \to \infty$, $L \to \infty$	小世界特性

WS小世界网络

- 现实社会中的网络并非完全随机也非完全规则，同时又具有高聚类特性和小世界特征。而无论是最近耦合网络还是ER随机网络都无法同时再现现实网络的特征。
- 目标：建立一种网络模型，其同时具有较大的聚类特性又具有较短的平均距离。

WS小世界网络

- Watts和Strogatz发现：作为从完全规则网络想完全随机网络的过渡，只要在规则网络中引入少许的随机性就可以产生具有小世界特征的网络模型。

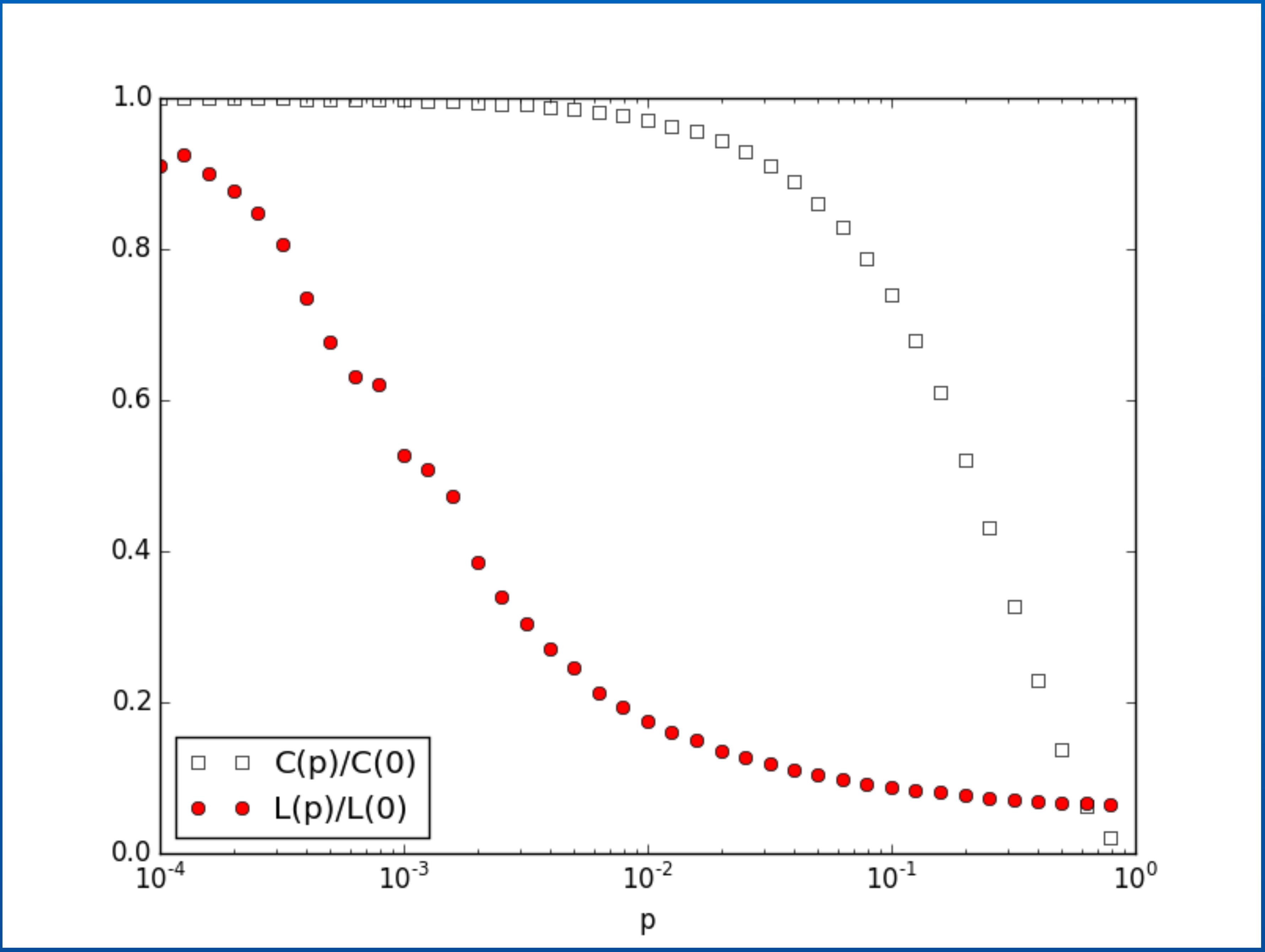


WS小世界模型的构造算法：

(1) 从规则图开始：给定一个含有N个点的环状最近邻耦合网络，其中每个节点都与在它左右的各K/2个节点相连，K是偶数。

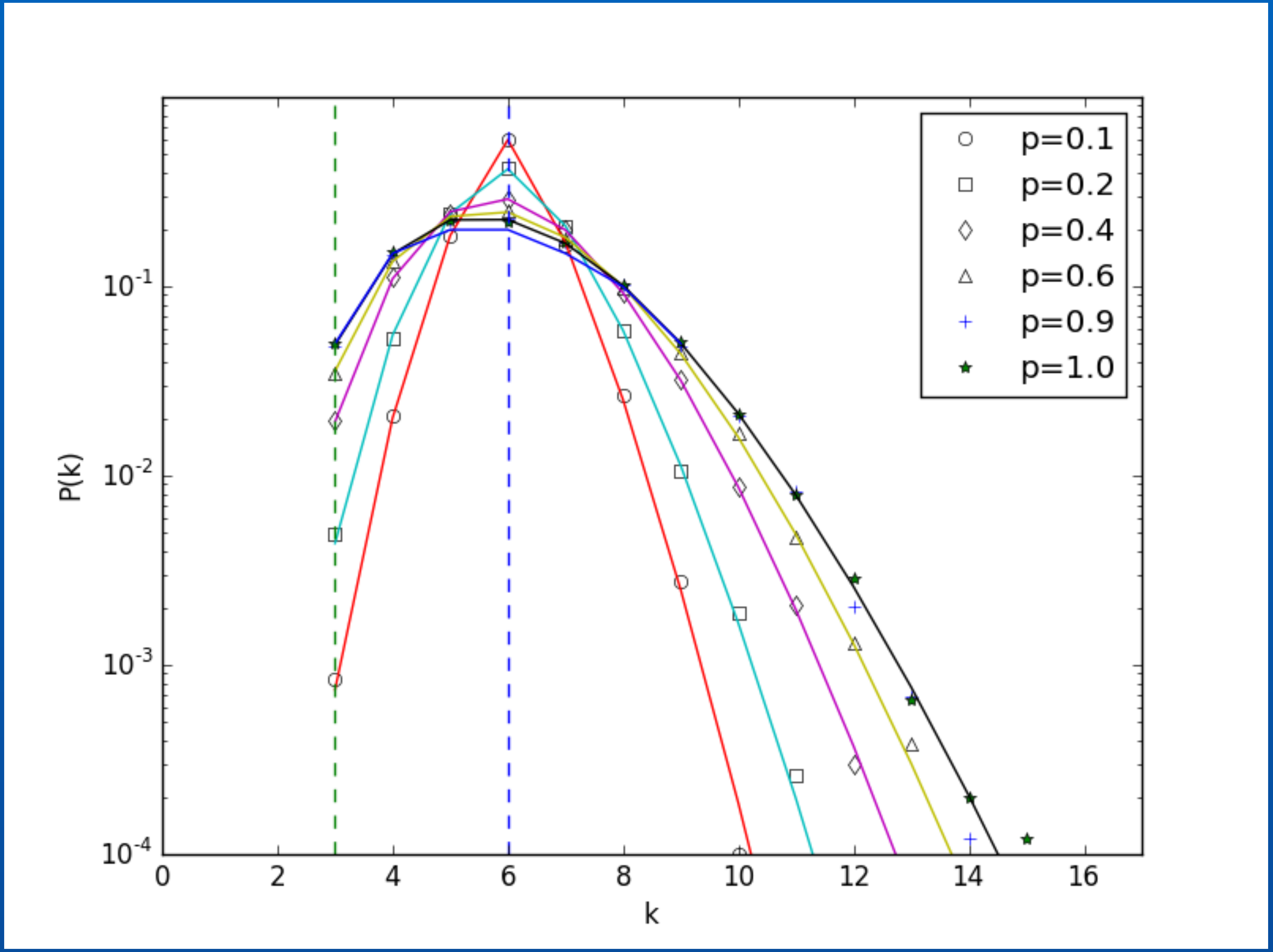
(2) 随机化重连：以概率p随机地重新连接网络中的每条边，即把每条边的一个端点保持不变，另一个端点改取为网络中随机选择的一个节点。其中规定不得有重边和自环。

WS小世界网络▷仿真分析



N=1000, K=10, 20轮

WS小世界网络▷仿真分析



N=1000, K=6

BA无标度网络

- Barabási 和 Albert
- WS小世界网络的度分布近似于泊松分布，在度平均值<k>处有一峰值，然后呈指数快速衰减。
- 包括Internet、WWW、科研合作网络以及蛋白质交互网络等网络的度分布都可以用适当的幂律形式来较好地描述。这类网络的节点的度没有明显的特征值，故称为无标度网络。
- 目的：建立网络模型，使其能刻画实际网络的幂律度分布特性。

BA无标度网络

- 切入点：ER随机图和WS小世界模型忽略了实际网络的两个特性：
  - 增长特性
  - 优先连接特性

构造算法

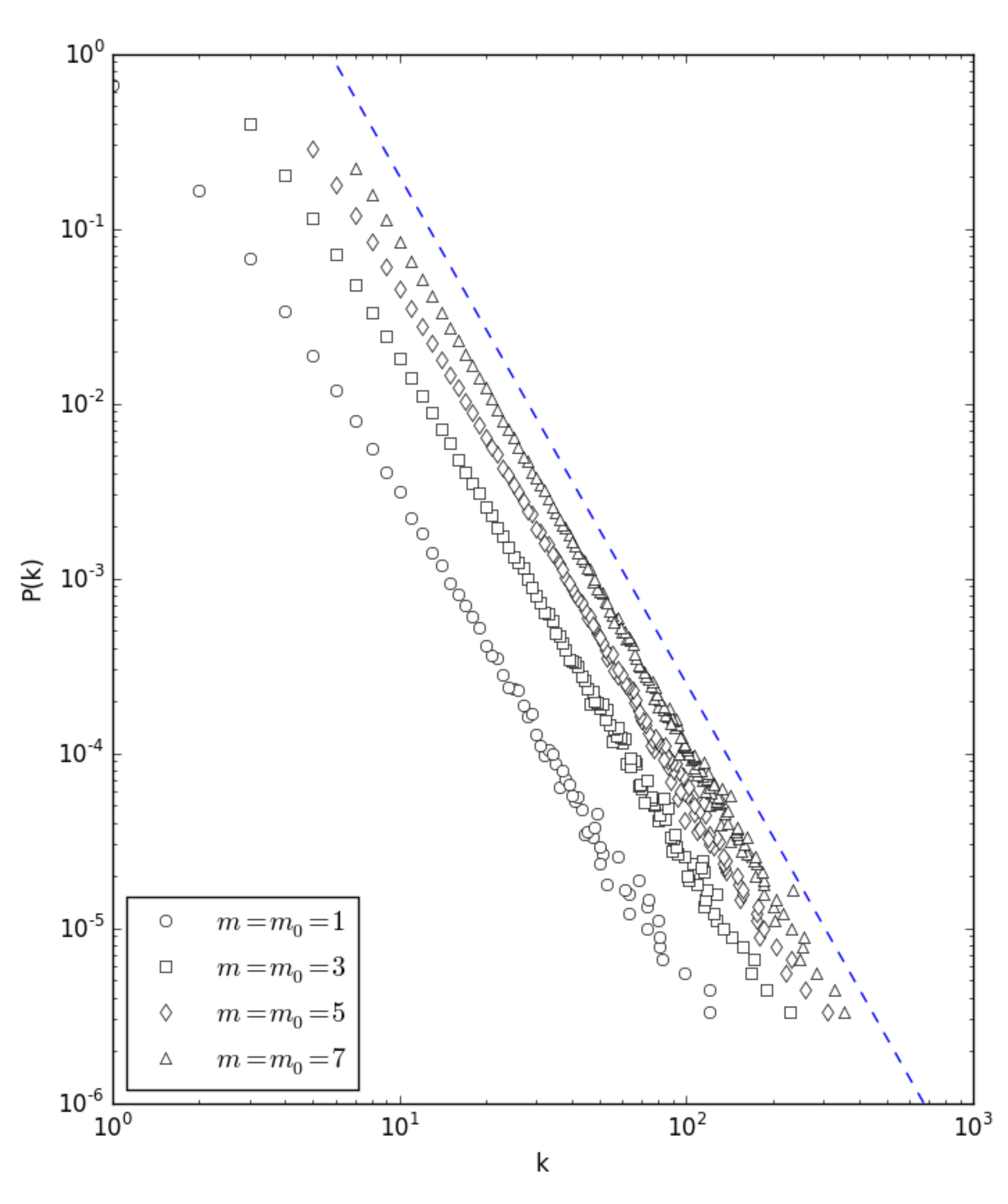
(1) 增长：从一个具有m_0个节点的连通网络开始，每次引入一个新的节点并且连接到m个已存在的节点上，这里$m \le m_0$。

(2) 优先连接：一个新节点与一个已经存在的节点i相连的概率π_i与节点i的度k_i之间满足如下关系：

$$\Pi_i = \frac{k_i}{\sum_j k_j}$$

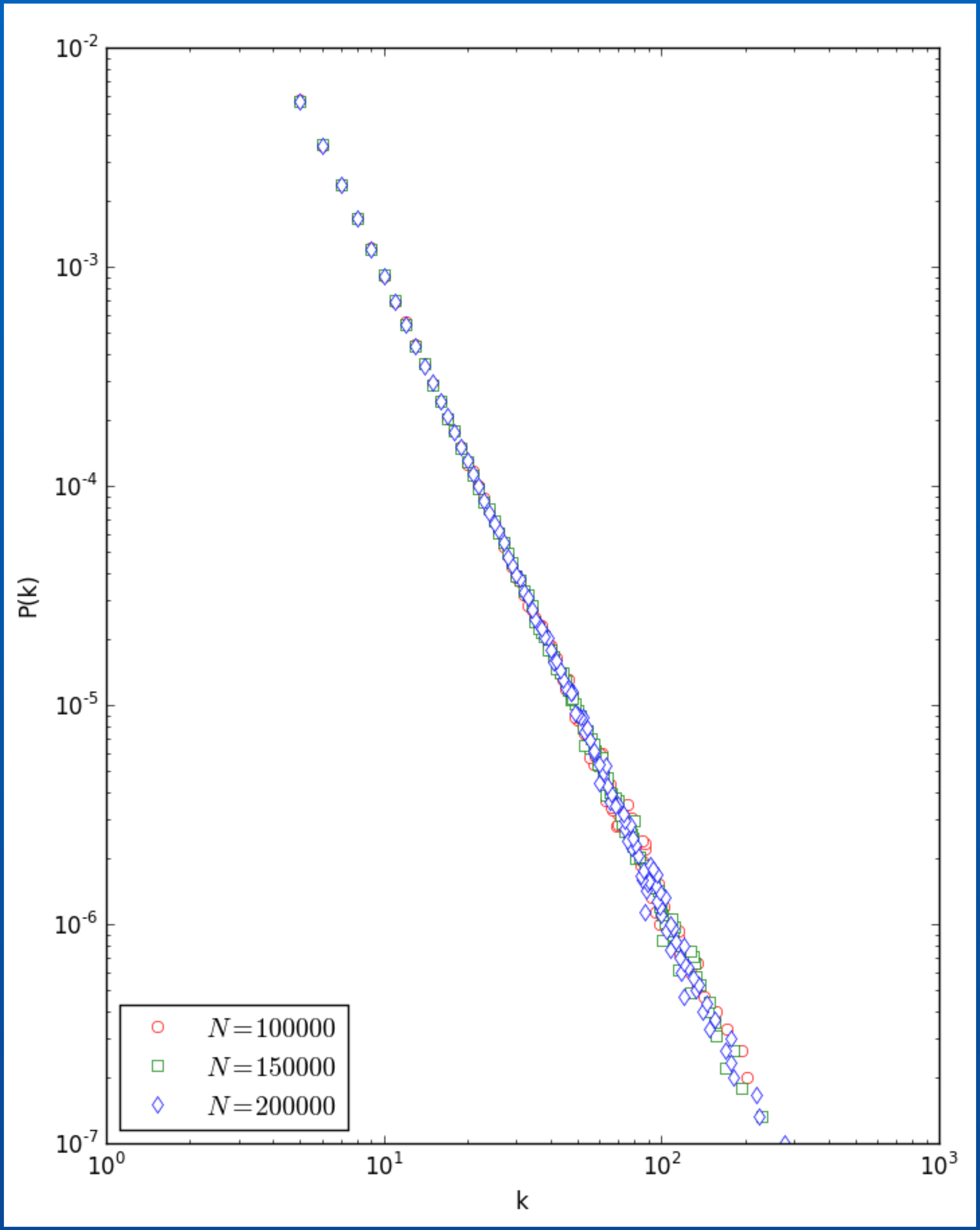
- http://localhost:9000/

BA无标度网络▷仿真实验▷幂律度分布



N=300k

BA无标度网络▷仿真实验▷幂律度分布



$m=m_0=5$

小结

	WS小世界	BA无标度	
建模目的	兼顾高聚类特性与低平均路径长度两个特性	刻画实际网络的幂律度分布特性	明确
构建模型	从规则图开始随机重连	基于增长和优先连接机制	简洁
仿真分析	满足期望特征	满足期望特征	

并行计算

编程实现中的发现▷并行计算

BA无标度，$m=m_0=7$

加速比：3.37

效率：42.13%

加速比：3.71

效率：46.37%

考虑因素

可并行的计算块

共享资源的分配

Thanks