“集智俱乐部热力学与进化论小组”第四次读书活动

时间：2009.4.5
地点：三号会所
人物：Jake, Miner, Wu Lingfei, Cao xudong, Tan hongchao, ..
内容：研讨 2 篇文章：Banavar:《Size and form in efficient transportation networks》和 O. Dreyer 的《Allometric Scaling and Central Source Systems》。

Wuling fei 的读书笔记：
Size and form in efficient transportation networks 读书报告

J．R. Banavar, Amos Maritan & Andrea Rinaldo. Nature VOL399 13MAY 1999

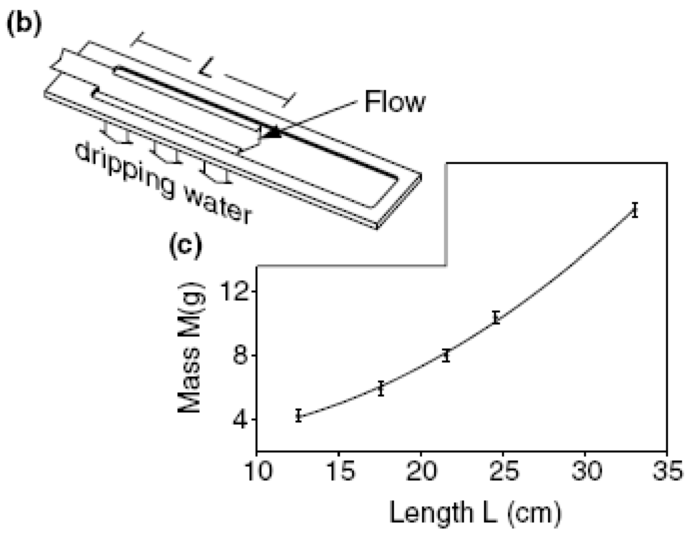
1.该研究的目标：解释生物学幂律 $B = B_0 M^{3/4}$ ，以及更广泛的**流与存**幂律里 $D/D+1$ 幂指数存在的原因

有趣的大自然幂律

$B = B_0 M^{3/4}$ B 代表新陈代谢率，M 代表生物体的体积

$A = A_0 M^{2/3}$ A 代表河流的流域盆地面积，M 代表河中水的总量

$L = L_0 M^{1/2}$ L 代表 Olaf Dreyer[1]设计的实验中水渗透的长度， M 代表实验中水的总量



2. Banavar 的解释思路

[1] Olaf Dreyer. Allometric Scaling and Central Source Systerms Physical Review Letters V87, N3 16 JULY 2001

（1）概念

L 一维量纲长度（可用系统相邻两点之间的平均长度 l 来测量，不过实际上单位是无所谓的，我们可以将 l 设为 1 而不失一般性）

L^D（使用一维量纲来刻画的欧式几何的）D 维空间的体积 V

B 生物体的新陈代谢率

C 给定生物体内总的血液量（因其取决于生物体内输送网络的结构，我们在此假设**最有效**的输送网络中意味着血液量最小），与生物体的体积 M 成正比

$|I_b|$ 第 b 个 link 上流的大小，对于中心点则只有 out flow,相当于网络中所有 site 上 inflow 的加总。b 是从第 X 个 site 到第 Y 个 site 的 link

F_x 第 x 个 site 血液的供应量，等于 x 个 site 上的 infow 减去 outflow，总为正值，其大小介于 $F_{\min}$ 与 $F_{\max}$ 之间（$F_{\min}$ 与 $F_{\max}$ 都取决于 l）

$B=\sum_x F_x$ 与 $N=L^D$ 成正比，可以认为 $B=L^D$

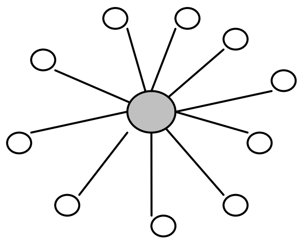
对于任何 D 维有向网络有

$\sum_b I_b\left(L_Y-L_x\right)=\sum_x L_x F_x$ 最大有向网上所有 link 的存之和等于所有 site 的存之和

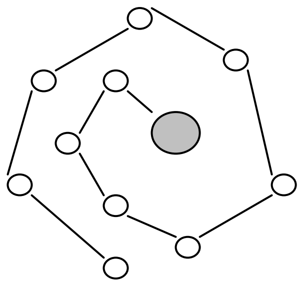
$\sum_b I_b\left(L_Y-L_x\right)=\sum_b^{directed} I_b$ 因为当 x 到中心点的距离和 y 相等时 L_Y-L_x=0

$C=\sum_b|I_b|\geq\sum_b^{directed}|I_b|\geq\sum_b^{directed} I_b=\sum_b I_b\left(L_Y-L_x\right)=\sum_x L_x F_x$（中间步骤不重要，大概就是说一下网络中的流向和网络结构的问题，为把全连接网 Full connected 和最大有向网 Maximal directed network 简化为单向树和盘旋蛇做铺垫）

C= L^D *节点到中心的平均距离,又因为后者介于 L 和 L^D 之间



单向树
Directed spanning tree



盘旋蛇
Space-filling Spiral

因此有 $L^D * L$ <C< $L^D * L^D$

即 $L^{D+1} \le \mathrm{C} \le L^{2D}$

又因 $B = L^D$

因此 $C^{\frac{D}{D+1}} \le B \le C^{\frac{1}{2}}$

当网络是最有效率结构(比如单向树)时，B 最小为 $C^{\frac{D}{D+1}} = M^{\frac{D}{D+1}}$

2.算法的思路分析

这套算法后面的思路是什么呢？我们可以这样看：生物体是一个有唯一中心源的网络结构，在这个结构中，新陈代谢率 B 其实就是网络中的流，生物体中的血液总量 C 就是网络中的存。生物体中，流的作用是什么呢？就是输送血液。我们假设中心点不断往各个节点上输送血液，Link 上不损失但要保留一定的“存”，那么维持一个这样的系统最少要多少血液呢？

因为离中心点越远的点所需要的血液越多，所以我们可以认为在输送血液的问题上，最关键的是点到中心源的距离[2]。每个点的存（或者每个 link 的存，这两个是相等的，在数学上二者都可以转化为距离的问题）与该点到中心源的距离相乘的值的加总，就是总的血液需求量。把问题进一步简化，通过比较两种极端情况后我们发现，在单向树这样的网络里，点到中心源的平均距离为 L，因此需要的血的存最小；而在盘旋蛇这样的结构里，点到中心源的平均距离为 L^D，因此需要的血的存最小。

文中没有详细回答的是，网络系统如何就实现了这样一个最有效率的结构呢？或者说，在一个复杂网络里，随着节点的迅速增加，节点到中心源的平均距离为是如何被维持在 L 一次方这个量纲级的？

对复杂网络比较熟悉的人都会觉得这个问题似曾相识。因为这正是 Barabasi 的经典研究中所揭示的现象。Barabsi 在研究互联网网络的拓扑结构时发现，大部分的复杂网络都有**平均路**

[2] 在这里“距离”（distance）特指图论中两点之间的最短路径 path，即捷径 geodesics 的长度

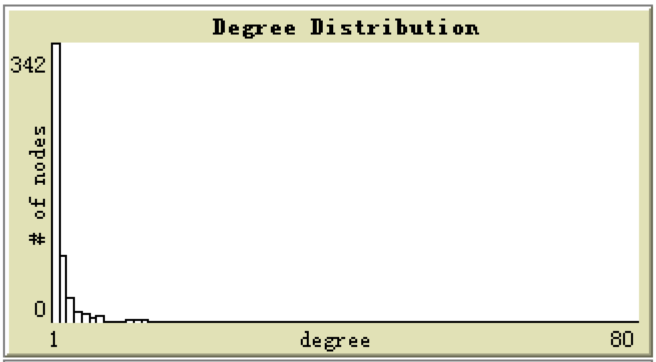
径的幂律现象和**节点度的幂律**分布，使得它们成为无标度 scale-free 网络。

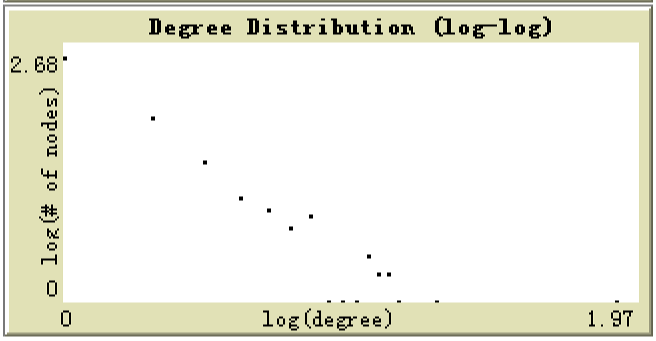
所谓**平均路径幂律现象**的问题，就是在网络中的节点变为 N 的时候，网络的平均路径变为 $\log N$（实际上大部分观察结果往往还要小于这个数）。实际上，1998 年 Barabasi 和 Albert 通过分析互联网数据拟合的平均路径与节点数之间的关系是 $d = 0.35 + 2\log N$，当时的网页数接近 8 亿，因此平均路径为 18.59。

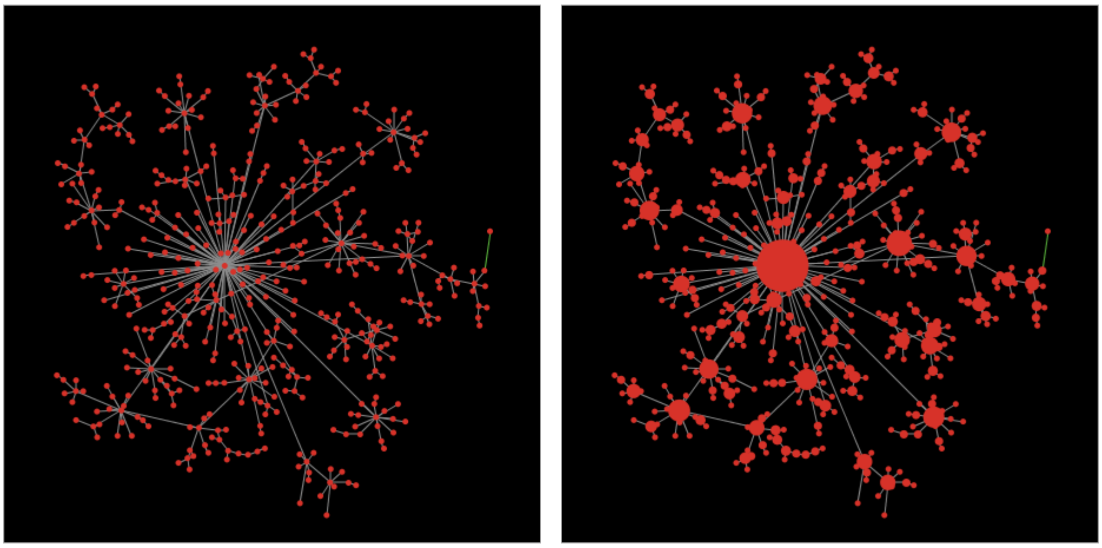
现在我们要接着问，这种差别悬殊的增长率是怎么造成的呢？按照 barabasi 的解释，这是每个节点都有很多 link 造成的。什么意思呢。就是说如果平均每个节点有 k 个 link，那么，从任一个节点出发走一步我们就能到达 k 个节点，走两步到达 k^2 个节点，•••走 d 步我们就可达到 k^d 个节点。把这个思路反过来，设总节点数为 N=k^d，则有平均路径 $d = \log_k^N = \frac{\log N}{\log k}$，易见，当 K 稍微比较大时（比如 k=10），只需要节点数众多的网络的平均路径是很小的。或者说，对节点数进行量纲级的增加，并不会导致平均路径量纲级的增加。

那么，我们还要进一步追问，这个“平均 link 数”是怎么实现的呢？在网络的所有节点上是一个什么样的分布呢？Barabasi 发现，节点的“度”（即 link 数）是一个幂律分布（为什么网络不以状态分布的方式实现这种平均 link 数，值得思考）。Barabasi 和 Albert 实验中分析的互联网数据的“出度”的分布是 $N(k) \sim k^{-2.5}$,入度的分布是 $N(k) \sim k^{-2.1}$。最后，Barabasi 最具创造性的发现是只要引入**两条规则：增长与优先选择**（马太效应）[3]就可以模拟出幂律分布。下图是使用 Netlogo 模拟的 BA 网络（Preferential attachment 规则）500 个节点的例子。

[3] 也有材料指出优先连接并非 Barabasi 原创——尽管在 *Linked* 的回忆中他表示自己是独立想到的，1925 年 Yule 在解释植物种类分布的时候已经提出类似思想（Yule G U.Philos. Trans.R.Soc. London B,1925,213:21），1955 年 Simon 对优先连接做出了进一步研究，还列举了五种可以用他理论解释的幂律：文献中的单词——经典 Zipf 律，科学家撰写文献的数量，城市人口、收入，及每类生物中的物种数(Simon H A. Biometrika,1955,42:425)。实际上在 Simon 之前 Champernowne 就提出了类似于“乘法过程”的模型来解释个人收入的分布(Champernowne D.Economic Journal,1953,63:318)。







（右图是 resized 以后，也就是根据节点度的大小来绘制节点大小的样子，这个对于我们理解生态系统中动物的 Body size 及其捕食策略，生育策略等情况有很大的启发意义（包括一个地区内各个 body size 数量级上动物的分布和总数量））。

也就是只要按照 BA 网络的两个原则：增长与优先连接（马太效应）就可以制造出具备生物体新陈代谢幂律的网络。

3.思考

（1）无中心化的网络：需要一个中心源假设么？

（2)度分布幂律对于平均路径与节点数之间的异速增长是不是必须的？实现平均为 k 个 link 的分布可以是任意的么？

（3）在一个具体的线路上，link 上的存与 site 上的存究竟是什么关系？ Link 与 site 消耗存吗？

（4）对于流来自于生物体外界，如氧气的情况，幂律仍然成立吗？为什么？如果成立的话，生物体的边界意义何在？

（5）幂律的讨论：幂律分布是自组织系统在临界状态（或者说混沌边缘），即从稳态到混沌态过度的标志，可以用来预测系统的相位及相变（Bak，Wiesenfielf 等人的沙堆模型，自组织的减熵有序化过程）。如果是这样的话，优先连接与自组织有没有什么关系？相变与通用计算、重整化群与小波分析。

主要讨论问题：

1、 Banavar 网络与 West 网络的主要区别：Banavar 模型将体积作为正比于新陈代谢的流，而把对流动的积分作为 Body mass。这与 West 将边界作为流，将体积当作存不同
2、 Banavar 网络与 Scale-free 网络之间的区别和联系，正如 lingfei 的读书笔记所述

Cao xudong 关于《Allometric Scaling and Central Source Systems》的报告：

从一个简单的数学问题说起：

【问题】

长度为 L 的线段，线段最左端有一个点源，线段上均匀分布强度为常数 c 的汇，线段最右端的流量密度为 0.

问：

1) 流量密度 j 在线段上的分布 j(x)
2) j(x)在线段上的积分$\int_0^L j(x)\,\mathrm{d}x$和线段长度 L 之间的函数关系

<求解>

微分方程及边界条件：

$$\mathrm{div}(j(x)) = c$$
$$j(0) = 0\text{（右端点为坐标系原点）}$$

<j(x)在线段上的分布>

$$j(x) = cx$$

<$\int_0^L j(x)\,\mathrm{d}x$ 和 L 之间的关系>

$$\int_0^L j(x)\,\mathrm{d}x = \frac{c}{2}L^2$$

如果将流密度在去整个物体上的积分记为 X，物体的体积记为 Y，那么上面的结果可以表示为：

$$X = \frac{c}{2}Y^2$$

【2 维和 3 维的结果】

我们可以建立二维和三维上的类似问题：

- 2d 问题：半径为 R 的圆，中心为点源，圆面上均匀分布强度为常数 c 的汇，圆的边界上流量密度为 0.
- 3d 问题：半径为 R 的球，中心为点源，球内均匀分布强度为常数 c 的汇，球面边界上的流量密度为 0。

这里只给出 2，3 维中 X，Y 之间的关系，求解方法类似，不再具体说明求解过程。（求解过程中有点细节问题没有搞清楚，见面的时候再讨论吧）

- 对于 2d 问题有：$X = \frac{c}{3\sqrt{\pi}}Y^{\frac{3}{2}}$
- 对于 3d 问题有：$X = \frac{c}{3\sqrt[3]{\pi}}Y^{4/3}$

【规律】

考虑 1，2，3 维情况中的 X，Y 之间的关系$X = \frac{c}{2}Y^2$，$X = \frac{c}{3\sqrt{\pi}}Y^{\frac{3}{2}}$，$X = \frac{c}{3\sqrt[3]{\pi}}Y^{4/3}$，可以发现如下规律：

$$X \sim Y^{\frac{D+1}{D}}$$

但是这个规律是在特殊几何体，汇的强度为常数的记为特殊的情况下推导出来的，如何才能推广到一般的的情况？

【启发】

还是回到最初的 1 维问题上，将线段长度扩大为原来的 2 倍，也就是 2L，汇的强度保

持不变。那么扩大后的线段和原来的线段相似。原线段$x = L/2$处，对应的扩大 2 倍之后$x = L$的位置。对比这两个位置的流发现，流量密度在扩大之后变为原来的 2 倍。这给我们一个很好的启示：

将物体扩大 s 倍，如果保持相似位置汇的强度不变，那么相似位置的流量密度 J 将扩大为原来的 s 倍

这样当物体扩大 s 倍时，体积变为$s^D Y$，流量密度的体积分变为 $s^{D+1}X$，所以就有了

$$X \sim Y^{\frac{D+1}{D}}$$

这里没有给出这个重要启示的推导，原文献中有，挺简单就不写了。

【现象。解释。质疑】

这里不太同意文献中的解释，所以直接引文献。

We start with our first example of animals and plants. Let V be the part of the animal or plant that contributes to the metabolism and that is supplied with nutrients. The metabolism B is then proportional to the volume of V, i.e., to our quantity Y. It is important to note here that Y is generally smaller than the total volume of the animal or plant.

Next we have to identify the quantity X. It is known that the mass M of an animal is proportional to the amount of blood contained in the animal [4]. Our quantity X, which is the integral of ρ over V, is exactly the mass of the blood. X is thus proportional to the mass M of the animal.

For a river the size A of the drainage basin is the volume of the system and thus equals our quantity Y. By an argument similar to the one given for animals and plants we see that X corresponds to the amount of water contained within the river. The assumption that the velocity of the water within the basin is independent of the size of the basin has been checked experimentally [7,8]. Setting $D = 2$ in Eq. (6) then gives Eq. (3).

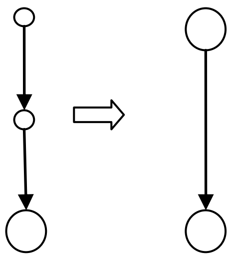
这两处都直接将 X 视为存量，但是之前的讨论中我们知道 X 并非存量，而是流量密度的体积分，只有速度保持不变的时候存量和流量密度之间才存在正相关，但是这个条件显然不满足，因为前面的讨论中流体密度为常数，流量密度并不是常数，也就是说速度不是常数。

讨论问题：

1、 与Banavar的网络相同，为何将流量的体积分定义为整个生物体的存？

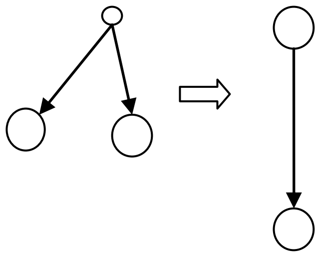
2、 关于$C = \sum_b |I_b| \geq \sum_b^{directed} |I_b| \geq \sum_b^{directed} I_b = \sum_b I_b \left(L_Y - L_x \right) = \sum_x L_x F_x$ 的直观解释

3、 一维的线性网络可看作是整个输运网络的建筑块，这样可以开发各种方法合并这些线性得到更复杂的网络。关于合并简单的线性网络有两种方式：



将不均匀的线性网络转换成均匀的线性网络

将并联的网络转化为单向的线性网络



这样，任意一个复杂的输运网络也可以拆解成多条线性的网络。