

基于复杂网络理论的代谢网络结构研究进展*

王冰^{1,2} 修志龙³ 唐焕文^{1,2**}

(1 大连理工大学应用数学系 大连 116024)

(2 大连理工大学计算生物学和生物信息学研究所 大连 116024)

(3 大连理工大学环境与生命学院 大连 116024)

摘要 后基因组时代研究的一个重要内容就是在系统生物学的基础上对多种分子和基因相互作用网络进行分析,理解生物系统是如何从单个构造模块的基础上组织起来的。近几年,复杂网络理论得到了迅速发展,其理论方法在生物网络的研究中得到了广泛应用。一些学者运用该理论方法研究了大量有关代谢网络的结构组成以及网络中子集团的层次组成关系,并获得了一些有意义的结果。这些结果对生物功能的预测具有一定的指导作用。对近几年来有关这方面的进展作简要综述。

关键词 代谢网络 小世界网络 无尺度网络 集聚系数 平均路径长度

一个多世纪以来,还原论一直在生物学研究中占有重要地位。人们已经获得了大量有关细胞组分和组分功能方面的知识。但是,经过大量的研究,人们发现大多数生物功能是由许多细胞分子相互作用共同完成的。因此,21世纪生物学的一个重要挑战就是从系统水平上对多种分子和基因相互作用网络进行分析,理解细胞内复杂的相互作用网络的结构和动力学对生物功能的影响^[1]。复杂网络理论的发展为在系统水平上研究生物网络提供了新的平台。

任何系统都可以用个体和个体之间相互作用构成的网络来描述,如细胞代谢网络^[2]、万维网^[3]、科学引文网络^[4]等。经过大量的实证研究,人们发现这些属于不同领域的网络遵循共同的普适规律。

自2000年Jeong等^[2]在Nature上发表利用复杂网络理论研究代谢网络的结构组成的论文以来,许多学者引用该文献并对代谢网络的拓扑结构进行了深入的研究。这些研究不仅提供了代谢网络的结构组成信息,而且提出了生物体代谢网络的演化过程和生命进化的猜想。本文对近几年来这方面的研究进展做简要综述,主要分为三个部分:第一部分简要介绍复杂网络理论的基本知识;第二部

分介绍应用复杂网络理论研究代谢网络取得的进展;第三部分为小结。

1 复杂网络理论的基本知识

1.1 描述网络的静态几何量

网络是一个包含大量个体和个体之间相互作用的系统^[5],可以用结点和结点之间作用关系构成的图 $G=(V,E)$ 来表示,其中 V 代表顶点集合, E 代表边集合。按照图中的边是否有方向,可以把图分为有向图和无向图。描述网络拓扑性质的常用几何量包括度分布、平均集聚系数、平均路径长度、介数等^[5-8]。下面做简要说明。

对于无向图,顶点的度是指与此顶点连接的边的数量。网络的度分布是指随机地选择一个顶点具有度为 k 的概率 $p(k)$,它是度量网络属性的一个重要几何量。

集聚系数描述了顶点的邻接点之间连接的可能性。在社会网络中描述为你的朋友,他们之间认识的可能性。集聚系数描述了网络的集团化的程度。

最短路径 l_{ij} 定义为所有连通 (i,j) 的通路中,所经过其它顶点最少的一条(或几条)路径,其长度称为最短路径长。平均路径长是对网络中任意一对顶点的最短路径长求平均,描述了网络中分离任意两个顶点所需的平均步。

介数(betweenness)包括顶点的介数和边的介数。顶点 u 的介数定义为网络中所有的最短路径之中, 经过 u 的数量, 反映了顶点 u 的影响力。类似地, 边 l 的介数是指网络中所有的最短路径中, 经过边 l 的数量, 反映了边 l 的影响力。Girvan 等^[8]发现, 利用边的介数可以对网络中的顶点进行分类, 获得网络的社区结构。

1.2 基本的网络机制模型

网络的静态几何量的研究给出实际网络中存在的一些普适规律, 那么从理论上构造机制模型来解释这些网络的形成与发展是研究网络上的动力学模型的基础。目前最常用的网络演化模型有规则网络模型、随机网络模型、小世界网络模型和无尺度网络模型。

规则网络(regular network)的特征是每一个顶点的度都相等, 集聚系数大, 平均路径长较大, 而随机网络(random network)的特征是度分布服从泊松

分布, 平均集聚系数小, 平均路径长度小。规则网络和随机网络模型是研究网络演化的基本模型, 这两个模型代表了网络演化的两个极端情况。

现实世界中的许多网络都具有‘小世界’特性, 即网络的平均路径长 l 很小, 与网络的规模成对数增长, 网络的平均集聚系数远大于相同规模的随机网络的集聚系数, 如电影演员合作网络、电力网、万维网和细胞代谢网络等^[7]。1998年, Watts 等^[9]给出了小世界网络的产生模型, 简称 WS 模型。其产生过程是从一个 k 阶规则网络开始, 以一定的概率 p 断开某一个端点, 重新连接, 连接的新的端点从网络的其它顶点中随机地选择。当 $p = 0$ 和 $p = 1$ 分别对应着规则网络和随机网络两个极限情形。当 $0 < p < 1$ 时, 存在一个很大的 p 的区域, 这时网络既具有规则网络的局部高集聚系数又具有随机网络平均路径短的特性。规则网络、小世界网络、随机网络之间的关系见图 1。

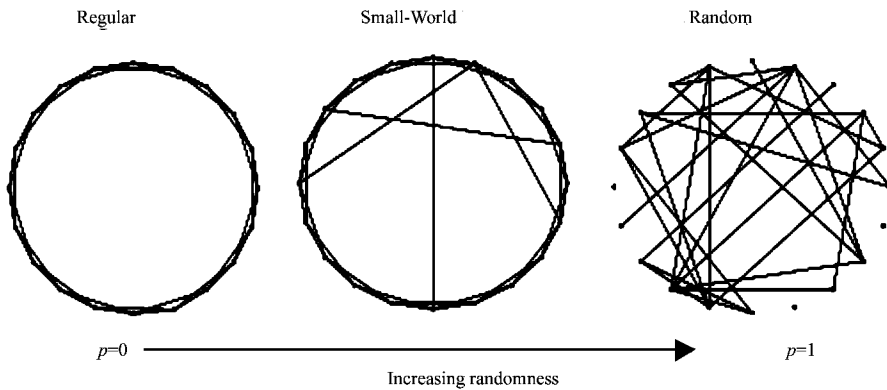


图 1 规则网络、小世界网络、随机网络之间的演化

Fig.1 The evolution of regular network, small-world network and random network

1999年, Barabási 等^[10]在对万维网的网页之间的超级链接关系做统计分析时发现, 网络的度分布并没有呈现预想中的随机网络的度分布, 而是服从幂率分布, 即 $P(k) \sim k^{-\gamma}$, 在双对数坐标系下是一条下降的直线。这意味着网络中大多数结点的度很小, 但存在少数结点的度远远高于网络的平均度, 这些连接度很大的结点称为集散结点(hub node), 具有这种幂率分布的网络称为无尺度网络(scale-free network), 无尺度网络的一个范例见图 2。

Barabási 和 Albert 根据万维网的特点, 建立了无尺度网络的演化模型, 简称 BA 模型。该模型由如下两个原则产生 (1) 网络规模的扩增, 即新结点

的增加 (2) 偏好依附, 新结点更倾向于和已有结点的度较高的结点连接, 即‘富者越富’的原则。

经过大量的实证研究, 人们发现属于生命领域、技术领域、社会领域的不同网络如代谢网络、蛋白质相互作用网络、万维网、科学家合作网络、电影演员合作网络等都呈现了共同的特征, 即它们都是无尺度网络, 网络的度分布服从幂率分布, 幂指数 γ 介于 2 到 3 之间^[4]。

2 代谢网络结构组成的研究进展

2.1 代谢网络的无尺度网络性质和‘小世界’特性

复杂网络理论的迅速发展为在系统水平上研

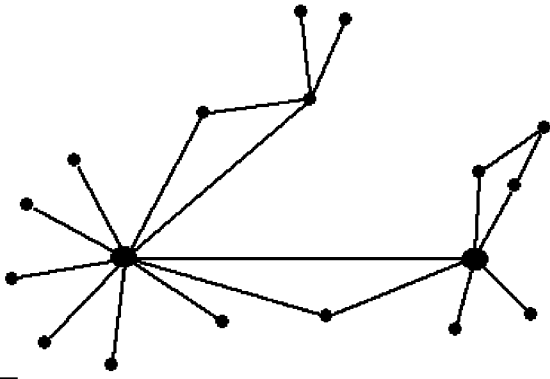


图2 无尺度网络

大的椭圆形黑圈代表网络的集散结点,集散结点控制整个网络

Fig.2 Scale-free network

The big elliptical black circles represent hub nodes.

These nodes control the whole network

究生物网络的结构组成和网络的功能提供了新的平台。近几年,网络生物学方面取得了惊人的进展^[11~19]。网络生物学的研究表明细胞网络同样遵循一定的普适规律。

代谢网络是生物体内发生的多个化学反应的总和。根据研究目的的不同,代谢网络可以表达成三种不同形式的图:代谢物图、反应图、二分图。代谢物图是把代谢网络中的代谢物看成图中的结点,连接代谢物之间的反应看成是图的边;反应图是把酶看成图中的结点,连接两个连续反应的中间复合物看成是图的边;二分图(bipartite graph)是把代谢物和酶看成两种不同类型的结点。这样,代谢网络就描述为一个数学意义上的图。

Jeong 等^[2]对分别属于古细菌、细菌、真菌的共43种生物的中央代谢网络进行了统计分析,发现随机地选择一个底物参加 k 个反应的概率服从幂率分布,即 $p(k) \sim k^{-\gamma}$ ($\gamma \approx 2.2$)。这表明代谢网络中存在少数参加多个反应的代谢物。同时,Jeong 等比较了43种生物的中央代谢网络的网络半径。网络半径定义为所有最短路径长度中的最大值。作者发现所有生物的网络半径都很小,呈现“小世界”特性,这意味着代谢网络中的大多数代谢物之间,只需三到四步反应二者就可以相互转化,代谢物浓度的变化会很快传播到整个网络。

代谢网络同其它技术领域、社会领域的网络一样遵循共同的普适规律,即都是无尺度网络。生物代谢网络是随着其代谢和生物合成能力的不断增强而发展到现在的状态的。因此,一些学者对代谢

网络的演化过程做出了如下两个猜想:第一,代谢网络的结构能否说明代谢网络演化的历史呢?按照 BA 模型的产生机理,连接度高的结点是在网络形成过程中较早出现的结点。类似地,对于代谢网络,如果生命的代谢网络的演化是按照添加新的代谢物而扩大规模的话,那么具有最大连接数的代谢物应该是系统中最早出现的代谢物。文献[11]中列出了大肠杆菌 *E. coli* 代谢网络中高度连接的代谢物,如被看作是 RNA 世界遗留的复合物:辅酶 A (coenzyme A)、乙酰辅酶 A (acetyl-CoA)、早期的氨基酸如谷氨酸 (glutamate)、谷氨酰胺 (glutamine)、丝氨酸 (serine) 等;早期的代谢物如丙酮酸 (pyruvate) 等。糖酵解 (glycolysis) 和三羧酸循环 (TCA) 被看作是最早出现的代谢途径。第二,代谢网络的“小世界”特性优化了代谢功能的哪些方面呢?代谢物的浓度或酶的浓度的变化通过网络的集散结点传播到整个网络而对代谢网络产生影响,代谢网络需要对这些扰动做出相应的反映。代谢网络的“小世界”特性使得其对外界环境的变化能够做出迅速的反应^[11]。

无尺度网络的最大特性是其高度的容错能力,即随机地去除网络中的某些结点,网络半径变化不明显^[5~7]。这也是为什么代谢网络中编码某些酶的基因发生突变,并不影响生物的正常代谢和生理过程。但同时无尺度网络的致命弱点就是对网络中集散结点的蓄意攻击会导致整个网络的瘫痪。无尺度网络中存在一些高度连接的结点称为集散结点 (hub node),这些结点控制着整个网络的结构。当去除这些集散结点时,网络的拓扑结构会发生巨大的改变。Ma 等^[12]对80种生物的代谢网络,按照每种生物中连接度最高的20种基本代谢物在80种生物中出现的次数多少排序,列出了20种基本代谢物。这些代谢物包括糖酵解途径中的丙酮酸 (pyruvate)、3-磷酸甘油 (glycerate-3-phosphate)、乙酰辅酶 A (acetyl-CoA)、6-磷酸果糖 (D-fructose 6-phosphate) 等。在不同生物体中,这些代谢物控制着整个代谢网络的结构。

2.2 代谢网络中反应物和可逆反应的处理

代谢网络自身具有不同于其它网络的一些特点,因此研究代谢网络时需要注意如下两个方面:第一,对于代谢网络中处处存在的反应物的处理。这些反应物包括 ATP、ADP、NADH 等参加多个生化反应的反应物。为了建立更具有生物学意义的网

络连接,这些反应物不能作为反应之间的有效中间物。因此,应该去除这些结点所参加的反应的连接,否则将得到从葡萄糖到丙酮酸只需2步反应的错误结论。第二,对于可逆反应的处理。根据研究目的的不同,对待可逆反应的处理方式也不同。如为了研究酶浓度变化(如变异)或底物浓度变化(如消耗)对整个网络产生的影响,就不能把代谢网络看成是有向图,因为有向图无法反映顶点的变化在网络中的传播。因此应把代谢网络看作无向图,即生化反应均看成是可逆反应。相反,为了研究不同生物的代谢网络之间的组成关系,用静态几何量描述代谢网络的拓扑组成,则应按照代谢网络内发生的真实生化反应将代谢网络表达成有向图的形式^[11,12]。基于这两点假设, Ma等^[12]从更具有生物学意义的角度改进了 Jeong 等的工作。作者利用基因组信息重建了80种生物的代谢网络,并得到了古细菌、细菌、真菌这3类生物具有不同的代谢网络结构,即细菌的平均路径长度要小于古细菌和真菌的平均路径长度,而不是3类物种具有相同的平均路径长度的结论。

2.3 代谢网络的子集团及其功能

用网络的几何量描述代谢网络的结构不是目的,最终目的是为了能够更好地发掘和预测网络的潜在功能。Ravasz等^[13]发现代谢网络除具有“小世界”和无尺度特性之外,代谢网络的平均集聚系数大于同规模的无尺度网络的平均集聚系数,表现出潜在的模块化特征,但是模块化网络的度分布一般不服从幂率分布,而且不存在无尺度网络中的集散结点(hub node)。为了解决这二者之间的矛盾,作者提出了代谢网络是嵌入模块的层次网络结构模型。该模型既具有高的集聚系数,同时度分布服从幂率分布。作者研究了该模型与代谢网络功能组成之间的关系,发现属于同一模块内的代谢物更有可能是功能相似的代谢物。小的模块不断组成较大的模块,然后组成更大的模块,自组织形成层次结构。因此代谢网络具有继承的自相似的性质。

近几年来,随着用复杂网络理论描述代谢网络结构的陆续完成,人们逐渐转入对代谢网络的分解和子网络之间的组成关系的研究,即如何把一个错综复杂的网络分解成具有特定功能的多个子集团。Schuster等^[14]依据几何量——顶点的度把代谢网络分解成若干具有相对独立功能的子集团。其方法

是计算所有顶点的度,设定一个度阈值 k_c 。如果 $k > k_c$,则顶点作为外部代谢物,否则将该顶点看作内部代谢物,是子集团的一部分。

Holme等^[15]利用几何量——介数对生物库中43种微生物的代谢网络、全细胞网络(信号转导网络、膜运输网络、电子运输网络等)进行分解,获得了网络的自然分类,构建了层次树,清晰地再现了子集团之间的层次组成关系。

代谢网络极其庞大而复杂。Ma等^[16]研究发现代谢网络并不是一个充分连接的网络,但存在高度连接的连通子网络。通过寻找强连通组分(giant strong component),比较其和全网络的平均路径长度,发现强连通组分的平均路径长度略小于整个网络的平均路径长度,二者呈线性关系。因此,整个网络的平均路径长度主要由其强连通组分的平均路径长度决定。这样,分析复杂代谢网络的结构就可以简化为分析其强连通组分的拓扑结构,极大地简化了问题的复杂性。

3 结束语

复杂网络的理论方法已经在代谢网络的研究中得到了广泛的应用,研究表明代谢网络和其它不同领域的网络如万维网、科学引文网络、电影演员合作网络等具有共同的普适性质,如它们都是无尺度网络,都具有“小世界”特性。对这些网络的结构性质进行研究有助于更进一步了解这些网络的形成机理和功能。代谢网络极其复杂,用网络分析的方法将代谢网络分解成若干具有相对独立功能的子网络,不仅能够简化问题的复杂度,同时有助于网络功能的预测。

利用复杂网络理论研究代谢网络还只停留于利用网络的静态几何量来描述代谢网络的结构性质,而利用网络的动力学性质研究代谢网络内部复杂的动力学将带来更具有建设性意义的成果。

参考文献

- [1] 孙之荣. 后基因组信息学. 北京: 清华大学出版社, 2002
Sun ZH R. Post-Genome Informatics. Beijing: Tsinghua University Press, 2002
- [2] Jeong H, Tombor B, Albert R. The large-scale organization of metabolic networks. *Nature*, 2000, 407: 651 ~ 654
- [3] Albert R, Jeong H, Barabasi A L. Internet: diameter of the world-wide web. *Nature*, 1999, 401: 130 ~ 131
- [4] Redner S. An empirical study of the citation distribution. *The*

- European Physical Journal B, 1998, 2(2):131 ~ 134
- [5] 吴金闪, 狄增如. 从统计物理学看复杂网络研究. 物理学进展, 2004, 24(1): 18 ~ 46
- Wu J S, Di Z R. Progress in Physics, 2004, 24(1): 18 ~ 46
- [6] Drogovtsev S N, Mendes J F F. Evolution of networks. Advances in Physics, 2002, 51: 1079 ~ 1187
- [7] Albert R, Barabasi A L. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002, 74: 47 ~ 97
- [8] Girvan M, Newman M E J. Community structure in social and biological networks. Proc Natl Acad Sci USA, 2002, 99(12): 7821 ~ 7826
- [9] Watts D J, Strogatz S H. Collective dynamics of small world networks. Nature, 1998, 393: 440 ~ 442
- [10] Barabasi A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509 ~ 512
- [11] Wagner A, Fell D. The small world inside large metabolic networks. Proceedings of Royal Society London Ser B, 2001, 280: 1803 ~ 1810
- [12] Ma H W, Zeng A P. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 2003, 19(2): 270 ~ 277
- [13] Ravasz E, Somera A L, Mongru D A, et al. Hierarchical organization of modularity in metabolic networks. Science, 2002, 297: 1551 ~ 1555
- [14] Schuster S, Pfeiffer T, Moldenhauer F, et al. Exploring the pathway structure of metabolism: Decomposition into subnetworks and application to *Mycoplasma pneumoniae*. Bioinformatics, 2002, 18: 351 ~ 361
- [15] Holme P, Huss M, Jeong H. Subnetwork hierarchies of biochemical pathways. Bioinformatics, 2003, 19(4): 532 ~ 538
- [16] Ma H W, Zeng A P. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics, 2003, 19(11): 1423 ~ 1430
- [17] Wagner A. The large-scale structure of metabolic networks: a glimpse at life's origin. Complexity, 2003, 8(1): 15 ~ 19
- [18] Podani J, Oltvai Z N, Jeong H, et al. Comparable system-level organization of Archaea and Eukaryotes. Nature Genetics, 2001, 29: 54 ~ 56
- [19] Wolf Y I, Karev G, Koonin E V. Scale-free networks in biology: new insights into the fundamentals of evolution. BioEssays, 2002, 24(2): 105 ~ 109

Progress of Metabolic Network Structure Based on Complex Network Theory

WANG Bing^{1,2} XIU Zhi-long³ TANG Huan-wen^{1,2}

(1 Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China)

(2 Institute of Computational Biology and Bioinformatics, Dalian University of Technology, Dalian 116024, China)

(3 School of Environmental and Biological Science and Technology, Dalian University of Technology, Dalian 116024, China)

Abstract One of the most important impacts in the postgenomic times is to analyze the interactions among molecules and genes based on systematic biology, to understand how the biology system is organized with the singular modular. Recently, a complex network theory has been developed fast and applied in biologic network widely. Some researchers have applied this theory to study metabolic network's structure organization, hierarchical structure of its clusters and got some instructing results. These results will play an important guiding role in biological functional prediction. The progress in these fields were reviewed.

Key words Metabolic network Small-world network Scale-free network Clustering coefficient Average path length