

转录调控网络模块和模体识别算法研究进展*

邹青宇, 刘 富, 侯 涛

(吉林大学 通信工程学院, 长春 130022)

摘要: 转录调控网络是生物体遗传信息传递的整体表示, 是人们理解基因表达过程的重要内容。识别转录调控网络的模块和模体是分析网络拓扑结构和组织方式的重要方法, 是揭示转录调控机制、生物发育与进化过程的重要环节之一。通过分析比较近年来用于转录调控网络模块识别的三类典型算法, 阐述了它们各自的优势和不足。介绍了一种被广泛使用的转录调控网络模体识别算法。以此为基础, 提出了转录调控网络模块和模体识别算法未来的研究方向。

关键词: 转录调控网络; 模块; 模体; 识别; 算法

中图分类号: TP391; O235

文献标志码: A

文章编号: 1001-3695(2012)11-4006-05

doi:10.3969/j.issn.1001-3695.2012.11.002

Advances in algorithms for modular and motif identification of transcriptional regulatory networks research

ZOU Qing-yu, LIU Fu, HOU Tao

(College of Communication Engineering, Jilin University, Changchun 130022, China)

Abstract: Passing transcriptional regulatory network as a whole of genetic information transmit is important for researchers to understand the gene expression. Identifying modules and motifs of them is a significant mean of analysis of topology and organization of network, and is one of the most important procedure to research the principle of transcription and the process of developmental and evolution of organisms. This paper described the advantages and disadvantages of three algorithms for network modules identification in recent years by analysing and comparing. Then it introduced an algorithm that widely used for motif identification. Based on these problems, it presented some future research trends.

Key words: transcriptional regulatory network; module; motif; identify; algorithm

0 引言

存在于所有生物细胞中的基因调控网络在系统水平上调控生物体发育和进化过程^[1]。转录调控网络是基因调控网络最重要的组成部分^[2,3], 是由携带转录调控信号的转录因子和被转录因子激活或抑制的靶基因构成的典型调控网络^[4]。随着 ChIP-chip^[5]、PBM^[6]和 DamID^[7]等实验技术的发展, 大量的转录因子与靶基因调控关系数据的迅速出现, 使人们能够从整体和局部上更加深入地理解转录调控网络^[8~10]。在过去的几年中, 大量的方法从网络拓扑结构和组织方式两个方面对转录调控网络进行了深入的研究^[11]。在网络拓扑结构方面, 转录调控网络是具有层次性^[12]、模块性^[13]、小世界性^[14]和无标度性^[15]的动态复杂定向网络。在网络组织方式方面, 转录调控网络可以分为基本元素、模体、模块和全网络四层结构^[16]。其中, 基本元素(网络节点)包括转录因子和靶基因两个部分^[17]。模体是由3~5个基本元素构成的具有一定拓扑结构的网络最小功能单元, 它在真实网络中出现的次数远大于在对应的随机网络中出现的次数^[18]; 模块是由多个模体构成的具有一定生物学功能和动力学特性的组织结构^[16]。模体和模块

是分析网络组织结构特征以及研究转录调控机制的重要部分。

本文对近年来典型的转录调控网络模体和模块识别算法进行了总结, 并探讨了各种识别算法的优势和存在的问题, 最后阐述了未来的发展方向。

1 网络模块识别算法

20世纪70年代出现了很多解决图分割问题的算法, 尽管图分割问题与网络模块识别问题不完全相同, 但两种问题的相似性还是使一些图分割领域的算法被成功地用于解决网络模块识别问题, 如谱平分算法和Kernighan-Lin算法等。现代的网络模块识别算法大多是由其发展而来的, 本文将现有的转录调控网络模块识别算法分为三类, 即聚类算法、基于中心度的算法、不属于以上两类的其他算法。

1.1 聚类算法

聚类算法被广泛地应用于社会网络、计算机网络和生物网络等各种复杂网络分析领域。用于转录调控网络模块识别的聚类算法分为凝聚层次聚类算法和启发式分裂聚类算法两类。无论哪一类算法都是以网络节点或链接(边)的特征度量值为

收稿日期: 2012-04-28; **修回日期:** 2012-06-09 **基金项目:** 国家自然科学基金资助项目(51105170); 吉林省科技发展计划资助项目(10100505)

作者简介: 邹青宇(1983-), 男, 吉林长春人, 博士研究生, 主要研究方向为系统生物学、模式识别(qyzou11@mails.jlu.edu.cn); 刘富(1968-), 男, 吉林长春人, 教授, 博导, 主要研究方向为生物信息学关键技术、生物识别技术、计算机视觉及模式识别方法及应用; 侯涛(1980-), 女, 贵州人, 博士研究生, 主要研究方向为生物信息学、系统生物学等。

基础,而特征度量值的计算方法有很多种,不同的计算方法将会产生不同的结果。

1.1.1 凝聚层次聚类算法

凝聚层次聚类算法采用自底向上的原则,首先将网络中每一个节点作为单独的一个类,然后根据各类的特征度量值将所有的类依次合并,最终成为一个类。常用层次聚类方法包括单连锁、全连锁和平均连锁三种方法。单连锁是将特征度量值最小的类合并,该方法不考虑类结构,在节点数量较大的情况下可能将存在中间节点的两个类合并成一类而产生散乱的分类。相反,全连锁是将特征度量值最大的类合并,该方法同样不考虑类结构,但对于紧密类的识别更加有效。平均连锁是依据各类特征度量值的平均值进行聚类划分,该方法考虑了类的结构,产生的分类相对稳定,倾向于合并差异小的两个类。层次聚类的过程可以用树状图表示。在树状图中,树叶代表网络节点,叶高度则代表节点的特征度量值大小,每一条树枝代表一个类。依据表示网络中节点聚集程度的聚类系数切割树枝识别网络模块,常用的聚类系数包括全局聚类系数、局部聚类系数和平局聚类系数等。

文献[19]对由 55 个大肠杆菌转录调控因子构成的调控网络进行模块识别,其方法如下:

a) 使用 Floyd-Warshall 算法计算出任意两节点间的最短路径,并计算节点 i 和 j 间的特征度量值(相似度) S_{ij} , 得到相似度矩阵。

$$S_{ij} = \frac{1}{d_{ij}^2} \quad (1)$$

其中: d_{ij} 为节点 i 和 j 间的最短路径。

b) 使用平均连锁凝聚层次聚类算法对其进行层次聚类,构建如图 1^[19]所示的树状图。

节点的聚类系数 C_i 为

$$C_i = \frac{2n_i}{k_i(k_i - 1)} \quad k_i > 1 \quad (2)$$

其中: k_i 为节点 i 的邻接节点数; n_i 为在节点 i 的 k_i 个邻接节点间所有链接的数量。依据平均聚类系数判断网络节点的聚集程度,文献[19]最终识别出八个功能模块,但此方法忽略了转录调控网络的定向性,将其视为非定向网络进行模块识别。

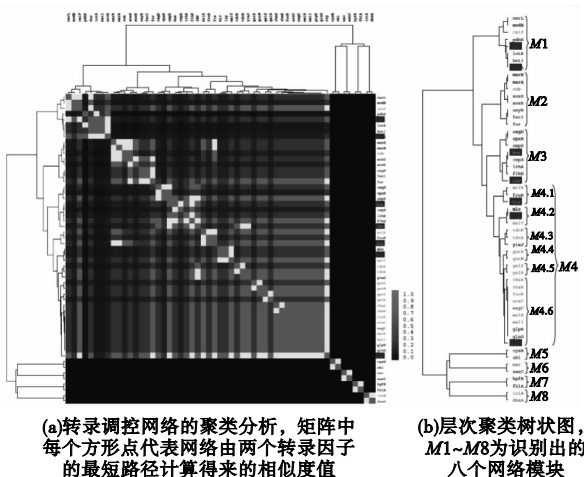


图1 大肠杆菌网络模块识别

文献[20]利用 ASAP 数据库中的基因表达数据,按照调控关系的方向将某节点及该节点下游的所有节点的集合定义为子网。用基于秩统计量的 SAM 算法计算得到统计数据矩阵,

用基于信息论的算法计算文献[21]中的表达数据得到时间过程数据矩阵;然后用曼哈顿距离算法和离差平方和聚类法对统计数据矩阵和时间过程矩阵进行层次聚类;最后将聚类结果与实际转录调控网络进行比较分析。该方法对转录调控机制有了一个更深入的理解。

尽管凝聚层次聚类算法便于理解和应用,但是通过该算法将类合并后,就无法再将其分离到合并前的状态,并以合并后得到的新类为基础进行下一步的运算。因此,节点特征度量值的算法及合并点的选择非常关键。用不同的节点特征度量标准和聚类标准得到的聚类结果有很大的差异,而且该算法趋向于聚集特征明显的节点,却忽略了特征模糊的节点,所以将其用于识别没有鲜明层次结构的转录调控网络时可能得到的模块并不准确。

1.1.2 启发式分裂聚类算法

文献[13]是一种被广泛用于网络模块识别的启发式分裂聚类算法,该算法计算网络中每条链接的边介数。由于模块间链接的边介数大于模块内链接的边介数,因此通过不断地移除边介数最高的链接及相关节点识别出网络模块。不同的边介数计算方法会产生不同的识别结果,常用的方法有最短路径法和最小二乘法等。

为了在模块数量未知的情况下识别网络模块,采用模块度来定量评估网络划分的质量^[22]。例如评估某一识别结果,将网络划分为 k 个模块,构建一个 $k \times k$ 维的对称矩阵 E ,用下式来计算模块度的值 Q :

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tre} - \|e^2\| \quad (3)$$

其中, Tre 为矩阵中对角线上各元素之和,即

$$\text{Tre} = \sum_i e_{ii} \quad (4)$$

e_{ii} 为模块 i 内部各节点的链接数与网络中所有链接数的比例, a_i 为模块 i 与网络中其他模块间的链接数与网络中所有链接数的比例,即

$$a_i = \sum_j e_{ij} \quad (5)$$

其中, e_{ij} 为模块 i 和模块 j 间的链接数与网络中所有链接数的比例。

模块度 Q 是同类型节点实际链接数量与随机连接情况下同类节点链接数量差的期望值。 Q 值越接近 1, 模块结构越明显, 实际应用中 Q 值通常在 0.3 ~ 0.7 间。文献[23]将网络拓扑结构与高通量表达数据相结合识别网络模块。首先构建一个概率网络模型, 然后利用高通量基因表达数据计算出模型中节点度量值, 不断移除度量值最高的节点及相关链接识别出网络模块。

尽管启发式分裂聚类算法效率高、用途广, 能够有效地识别出网络模块, 但是用该算法识别出的网络模块大小总是约等于全网络大小的平方根。因此, 该算法识别结果并不是十分准确。

1.2 基于中心度的算法

中心度是衡量网络节点或链接在网络中重要程度的指标, 不同的中心度算法含义不同。基于中心度的模块识别算法首先根据中心度算法计算每个节点或链接的重要程度, 识别出网络中心节点或中心链接, 然后以中心节点或中心链接为核心进行扩展, 识别出网络模块。文献[24]将五种不同的中心度算法: 度中心算法、接近中心算法、相关整合算法、卡茨状态指数及页面等级算法和基于模体的中心度算法, 应用在大肠

杆菌基因调控网络上,并对结果进行了分析比较。其中,度中心算法以节点的邻接节点数量衡量节点的重要程度,分为入度和出度两种。入度表示调控该节点的邻接节点数量,出度表示被该节点调控的邻接节点数量,度数最大的节点为中心节点。接近中心算法以节点接近度 C_{clo} 衡量节点的重要程度,节点 u 的接近度为

$$C_{clo} = \frac{1}{\sum_{v \in V} \text{dist}(u, v)} \quad (6)$$

其中: V 为所有节点的集合, $\text{dist}(u, v)$ 为节点 u 到节点 v 的距离。相关整合算法与接近算法相似,基于两节点间的距离来计算相关中心度和整合中心度。

扭转距离矩阵中的元素 RD_{ij} 为

$$RD_{ij} = \text{diameter}(G) + 1 - D_{ij} \quad (7)$$

其中, $\text{diameter}(G)$ 为网络中所有节点间最短距离的最大值。

相关中心度 $C_{rad}(i)$ 为

$$C_{rad}(i) = \frac{\sum_{j \neq i} RD_{ij}}{n-1} \quad (8)$$

整合中心度 $C_{int}(i)$ 为

$$C_{int}(i) = \frac{\sum_{j \neq i} RD_{ji}}{n-1} \quad (9)$$

其中 D_{ij} 为 $\text{dist}(i, j)$ 。

相关中心度高的节点容易接近其他节点,整合中心度高的节点容易被其他节点接近。节点 i 基于模体 M 的中心度 C_m 是网络中与模体 M 匹配并包含节点 i 的模体数量。如图 2^[24] 所示, (a) 中所有与 (b) 所示的 FFL 模体匹配并包含节点 $v01$ 的模体共有两个, 因此 $C_m(v01) = 2$ 。随着转录调控网络模体的发现, 基于模体的中心度算法能够有效地识别出网络模块。

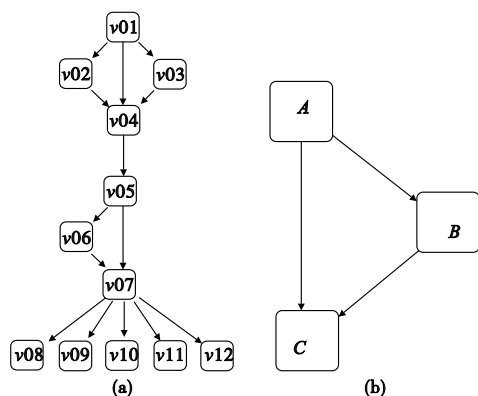


图2 基于模体的中心度算法实例

文献[25]对基因表达数据、模体信息数据和 ChIP-chip 数据进行分析,利用直观算法识别出拥有最多共调控因子和共调控模体的基因表达模块作为网络中心模块;通过计算模块的平均表达谱将剩余的基因按照其与中心模块的相关性排序,并依据模块饱和值进行划分识别出网络模块。文献[26]利用大肠杆菌和枯草芽孢杆菌的基因表达数据构建基因临界值矩阵。将临界值最大的模块作为中心模块,然后逐渐减小临界值建立它们各自的模块树。将两个模块树进行对比找到两个物种间的相似模块,通过同源性基因图识别出相似模块中的核心部分和可变因素。通过此方法找到两个物种间最相关的保守模块。

1.3 其他算法

文献[27]用 Pajek 网络分析软件按照高层为调控节点、低层为被调控节点的原则,将大肠杆菌转录调控网络分为 5 层;然后移除 3~5 层的所有节点及相关链接,对剩余的节点子集

按照子集中链接的数量进行分类识别网络模块。最终利用这种从上到下的方法识别出了 39 个功能模块。文献[28]提出了一种完全不切割网络调控关系的模块识别算法,该算法首先找到一个只受网络外部信号调控的节点作为起点,将该信号沿着调控关系顺向传递过程中,路经的所有节点构成的集合作为网络模块。文献[29]将转录调控网络模块定义为被一组相同转录因子控制的所有共表达基因所组成的子网,并利用全基因组位置数据和基因表达数据识别共表达基因集合。文献[30]用网络节点的聚类系数计算出切割阈值,将节点分为模块节点和层次节点两类,用模块节点识别网络模块,用层次节点划分网络层次。用此方法对由 4 570 个转录因子与靶基因调控关系构建的大肠杆菌转录调控网络进行了模块识别。

2 网络模体识别算法

网络模体的概念由文献[31,32]提出,他们将网络模体定义为:在真实网络中出现的次数远大于在对应的随机网络中出现的次数的子网。目前,大多数模体识别算法仍采用这个概念来判断一个子图是否为模体。因此,如图 3^[31] 所示,典型的网络模体识别算法包括以下三个步骤:

a) 依据真实网络的性质构建随机网络。随机网络模型构建是否得当,与能否有效地识别出网络模体息息相关。因此,要求随机网络必须与真实网络具有相似的统计特性。因为度分布^[33]是复杂网络中最重要的全局统计特性,所以模体识别算法通常使用与真实网络具有相同度分布的随机网络模型。

b) 在真实网络和随机网络中搜索特定规模的子网,并按照子网同构性进行归类。典型子网搜索算法有穷尽递归搜索算法^[31]、ESA 算法^[34]和 ESU 算法^[35]三种。穷尽递归搜索算法通过枚举 N 节点网络的 $N \times N$ 邻接矩阵中所有 $n \times n$ 子矩阵得到对应的子网,由于计算量较大,该算法只能有效地识别小规模模体。ESA 算法首先在网络中随机地选取一条链接,并识别此链接的两个关联节点;然后分别搜索这两个节点的邻接节点,依次重复搜索邻接节点的邻接节点直至搜索到指定规模的子网。此方法更适合于对大规模网络的分析,并且能够识别 8 个节点的模体。但是,ESA 算法不能保证识别出所有的子网,而且同一个子网可能被重复识别。ESU 算法从输入网络的一个节点 n 开始,搜索在 n 点下游的节点,搜索过程中排除已被搜索节点的邻接节点,直至满足指定的子网规格。由于在搜索过程中有了条件限制,因此保证了一个子网仅被搜索一次。此方法更适合用于对稀疏网络的分析,但由于问题本身的复杂性,搜索的子网规模最大也只能达到 6 个节点。

c) 比较每一类子网在真实网络和随机网络中出现的次数,根据其统计意义确定网络模体。根据网络模体的性质,利用 Z_score 判断子网 n_i 是否为网络模体,即

$$Z_score = (N_{rel} - N_{ran}) / S_{ran} \quad (10)$$

其中: N_{rel} 为 n_i 在真实网络中的出现次数; N_{ran} 和 S_{ran} 分别为在所有随机网络中的平均出现次数和标准差^[31]。

文献[19]用文献[31]的模体识别方法,对转录调控网络中单一网络模块和两个相邻的网络模块进行模体识别,得到了前馈模体和双边模体在模块内及模块间的分布情况。文献[36]用边缘交换算法、边缘匹配算法和蒙特卡罗算法构建了大肠杆菌和酵母转录调控随机网络模型,然后利用穷尽递归搜

索算法和 Z_score 值对网络中的 3 和 4 节点模体进行了识别,并发现了模体间的相互依存关系。为了更有效地搜索非树型的子网,文献[37]改进了 ESU 子网搜索算法。用改进后的 ESU 算法对蛋白质相互作用网络和转录调控网络进行了子网搜索,用层次聚类算法对搜索到的子网进行聚类分析,识别出网络模体。

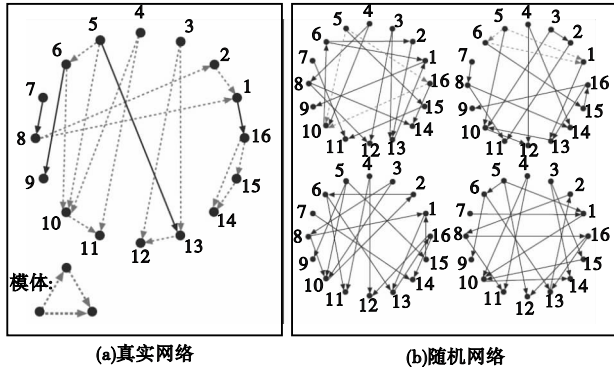


图3 网络模体识别算法(其中虚线表示的前馈环模体在真实网络中出现了5次,在随机网络中平均出现0.5次)

3 结束语

本文对典型的转录调控网络模块和模体识别算法进行了总结和评价,指出了每种算法的优势及其存在的问题。除上述算法外还有很多其他的转录调控网络分析算法。文献[38]将转录网络中的一组紧密连接且共表达的差分表达基因定义为响应基因模块的基础标签,通过查找标签来识别网络模块。文献[39]构建了一个酵母转录调控网络模型,用贝叶斯算法分析此网络,识别出了 27 个网络模块。文献[40]用 Monte-Carlo 算法对转录调控网络模型中的模体进行了统计,讨论了转录调控网络中模体分布规律。该算法是通过随机搜索得到模体出现的频率,用此频率代替概率计算出网络中模体的数量。文献[41,42]将网络模体分为桥模体和砖模体。桥模体是只由弱连接构成,并且与其他模体既不交互也不重叠的模体;砖模体是只由强连接构成,并且在全拓扑结构中起重要作用的模体。该算法将任意两个节点间链接的超几何系数定义为两个节点链接的权重值,依据此权重值判断链接的强弱,将所有链接分为强链接和弱链接两种,然后依据其 Z_score 值和 SP 值识别出桥模体和砖模体。由于篇幅有限,不能将所有算法一一介绍。

转录调控网络是近年来系统生物学的研究热点之一。鉴于转录调控网络的复杂性、模块性、无标度性和动态性等性质,未来的转录调控网络分析算法趋向于以下几个方面:

a) 重叠的网络模块。传统的转录调控网络模块识别算法按照一一对应的原则,将节点划分到对应的网络模块中。而事实上转录调控网络的模块是普遍重叠的,很多节点都不只属于一个网络模块,发展能够识别重叠网络模块的算法非常重要。目前,模糊聚类法^[43,44]和中心发现法^[45,46]等算法可以用于识别转录调控网络的重叠模块。

b) 网络生物学属性。转录调控网络是生物网络,网络的节点具有生物学意义,如转录因子和靶基因等。节点的相互作用关系除具有方向性外还包括增强、抑制和既增强又抑制三种生物学属性。而以往的算法大多没有考虑网络节点和链接的生物学属性,这样就没有准确地说明网络模块的生物学意义。文献[32,47]通过比较大肠杆菌和啤酒酵母的转录调控网络,

认为不同生物转录调控网络的拓扑结构相近,它们的功能差别是因其转录因子和靶基因比例不同而造成的。另外,基于图论模块识别方法能够在一定程度上考虑到转录调控网络生物学属性^[48,49]。

c) 转录调控的动态性。以往的研究都是根据网络的拓扑结构来识别网络模块,随着对转录调控机制的深入研究,人们提出了网络动态性的概念^[50]。在不同的时间、不同的环境下网络的调控关系并不完全相同,动态网络更能体现基因之间及基因和蛋白质之间的调控关系,更加精确的调控关系更加有利于分析调控网络的拓扑结构和组织方式。因此,识别网络模块和模体不能忽略网络的动态性^[51]。当前,已经有一些基于网络动态性的子网识别技术用于转录调控网络的模块和模体识别中^[52-55]。

综上所述,识别转录调控网络的模块和模体结构有很大的理论和实际意义。经过多年的发展,该领域的研究已取得了长足的进步,但在一些复杂的方面还有待进一步研究。通过对该领域的研究,人们将更加深入地理解转录过程的调控机制和生命过程。

参考文献:

- [1] DAVIDSON E H. Emerging properties of animal gene regulatory networks[J]. *Nature*, 2010, 468(7326): 911-920.
- [2] PTASHNE M. Regulation of transcription: from lambda to eukaryotes[J]. *Trends in Biochemical Sciences*, 2005, 30(6): 275-279.
- [3] BABU M M. Structure, evolution and dynamics of transcriptional regulatory networks[J]. *Biochemical Society Transactions*, 2010, 38: 1155-1178.
- [4] KIM T M, PARK P J. Advances in analysis of transcriptional regulatory networks[J]. *Wiley Interdisciplinary Reviews-Systems Biology and Medicine*, 2011, 3(1): 21-35.
- [5] BUCK M J, LIEB J D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments[J]. *Genomics*, 2004, 83(3): 349-360.
- [6] BULYK M L. DNA microarray technologies for measuring protein-DNA interactions[J]. *Current Opinion in Biotechnology*, 2006, 17(4): 422-430.
- [7] GREIL F, MOORMAN C, Van STEENSEL B. DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase[J]. *Methods in Enzymology*, 2006, 410: 342-359.
- [8] KASOWSKI M, GRUBERT F, HEFFELFINGER C, et al. Variation in transcription factor binding among humans[J]. *Science*, 2010, 328(5975): 232-235.
- [9] ZHENG Wei, ZHAO Hong-yu, MANCERA E, et al. Genetic analysis of variation in transcription factor binding in yeast[J]. *Nature*, 2010, 464(7292): 1187-1191.
- [10] KIM J, CHU Jian-lin, SHEN Xiao-hua, et al. An extended transcriptional network for pluripotency of embryonic stem cells[J]. *Cell*, 2008, 132(6): 1049-1061.
- [11] MACISAAC K D, FRAENKEL E. Practical strategies for discovering regulatory DNA sequence motifs[J]. *PLOS Computational Biology*, 2006, 2(4): 201-210.
- [12] RAVASZ E, BARABASI A L. Hierarchical organization in complex networks[J]. *Physical Review E*, 2003, 67(2): 026112.
- [13] GIRVAN M, NEWMAN M E J. Community structure in social and

- biological networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821-7826.
- [14] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440-442.
- [15] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [16] BABU M M, LUSCOMBE N M, ARAVIND L, *et al.* Structure and evolution of transcriptional regulatory networks[J]. *Current Opinion in Structural Biology*, 2004, 14(3): 283-291.
- [17] MASTON G A, EVANS S K, GREEN M R. Transcriptional regulatory elements in the human genome[J]. *Annual Review of Genomics and Human Genetics*, 2006, 7: 29-59.
- [18] ALON U. Network motifs: theory and experimental approaches[J]. *Nature Reviews Genetics*, 2007, 8(6): 450-461.
- [19] RESENDIS-ANTONIO O, FREYRE-GONZALEZ J A, MENCHACAMENDEZ R, *et al.* Modular analysis of the transcriptional regulatory network of E-coli[J]. *Trends in Genetics*, 2005, 21(1): 16-20.
- [20] MARR C, THEIS F J, LIEBOVITCH L S, *et al.* Patterns of subnet usage reveal distinct scales of regulation in the transcriptional regulatory network of escherichia coli[J]. *PLOS Computational Biology*, 2010, 6(7): e1000836.
- [21] SANGURDEKAR D P, SRIENC F, KHODURSKY A B. A classification based framework for quantitative description of large-scale microarray data[J]. *Genome Biology*, 2006, 7(4): R32.
- [22] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2): 1-16.
- [23] ULITSKY I, SHAMIR R. Identification of functional modules using network topology and high-throughput data[J]. *BMC Systems Biology*, 2007, 1: 8.
- [24] KOSCHUTZKI D, SCHREIBER F. Centrality analysis methods for biological networks and their application to gene regulatory networks[J]. *Gene Regulation and System Biology*, 2008, 2: 193-201.
- [25] LEMMENS K, DHOLLANDER T, De BIE T, *et al.* Inferring transcriptional modules from ChIP-chip, motif and microarray data[J]. *Genome Biology*, 2006, 7(5): R37.
- [26] ZARRINEH P, FIERRO A C, SANCHEZ-RODRIGUEZ A, *et al.* COMODO: an adaptive coclustering strategy to identify conserved co-expression modules between organisms[J]. *Nucleic Acids Research*, 2011, 39(7): E41-U45.
- [27] MA Hong-wu, BUER J, ZENG An-ping. Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach[J]. *BMC Bioinformatics*, 2004, 5: 199.
- [28] BALAZSI G, BARABASI A L, OLTVAI Z N. Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(22): 7841-7846.
- [29] BAR-JOSEPH Z, GERBER G K, LEE T I, *et al.* Computational discovery of gene modules and regulatory networks[J]. *Nature Biotechnology*, 2003, 21(11): 1337-1342.
- [30] ALONSO-PAVON J A, TREVINO-QUINTANILLA L G, *et al.* Functional architecture of Escherichia coli: new insights provided by a natural decomposition approach[J]. *Genome Biology*, 2008, 9(10): R154.
- [31] MILO R, SHEN-ORR S, ITZKOVITZ S, *et al.* Network motifs: simple building blocks of complex networks[J]. *Science*, 2002, 298(5594): 824-827.
- [32] SHEN-ORR S S, MILO R, MANGAN S, *et al.* Network motifs in the transcriptional regulation network of Escherichia coli[J]. *Nature Genetics*, 2002, 31(1): 64-68.
- [33] DENG Wei-bing, LI Wei, CAI Xu, *et al.* The exponential degree distribution in complex networks: non-equilibrium network theory, numerical simulation and empirical data[J]. *Physica A: Statistical Mechanics and Its Applications*, 2011, 390(8): 1481-1485.
- [34] KASHTAN N, ITZKOVITZ S, MILO R, *et al.* Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs[J]. *Bioinformatics*, 2004, 20(11): 1746-1758.
- [35] WERNICKE S. A faster algorithm for detecting network motifs[C]// *Proc of the 5th International Conference on Algorithms in Bioinformatics*. Berlin: Springer-Verlag, 2005: 165-177.
- [36] GINOZA R, MUGLER A. Network motifs come in sets: Correlations in the randomization process[J]. *Physical Review E*, 2010, 82(1): 011921.
- [37] QIN Gui-min, GAO Lin. An algorithm for network motif discovery in biological networks[J]. *International Journal of Data Mining and Bioinformatics*, 2012, 6(1): 1-16.
- [38] GU Jin, CHEN Yang, LI Shao, *et al.* Identification of responsive gene modules by network-based gene clustering and extending: application to inflammation and angiogenesis[J]. *BMC Systems Biology*, 2010, 4: 47.
- [39] HOU Lin, WANG Lin, QIAN Min-ping, *et al.* Modular analysis of the probabilistic genetic interaction network[J]. *Bioinformatics*, 2011, 27(6): 853-859.
- [40] BURDA Z, KRZYWICKI A, MARTIN O C, *et al.* Motifs emerge from function in model gene regulatory networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(42): 17263-17268.
- [41] CHENG C Y, HUANG C Y, SUN C T. Mining bridge and brick motifs from complex biological networks for functionally and statistically significant discovery[J]. *IEEE Trans on Systems, Man, and Cybernetics Part B: Cybernetics*, 2008, 38(1): 17-24.
- [42] HUANG C Y, CHENG C Y, SUN C T. Bridge and brick network motifs: identifying significant building blocks from complex biological systems[J]. *Artificial Intelligence in Medicine*, 2007, 41(2): 117-127.
- [43] SUN Peng-gang, GAO Lin, HAN Shan-shan. Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks[J]. *Information Sciences*, 2011, 181(6): 1060-1071.
- [44] GREGORY S. Fuzzy overlapping communities in networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2011: P02017.
- [45] LEE C, REID F, MCDAID A, *et al.* Seeding for pervasively overlapping communities[J]. *Physical Review E*, 2011, 83(6): 066107.
- [46] SHANG Ming-sheng, CHEN Duan-bing, ZHOU Tao. Detecting overlapping communities based on community cores in complex networks[J]. *Chinese Physics Letters*, 2010, 27(5): 058901.
- [47] GUZMAN-VARGAS L, SANTILLAN M. Comparative analysis of the transcription-factor gene regulatory networks of E. coli and S. cerevisiae[J]. *BMC Systems Biology*, 2008, 2: 13. (下转第 4016 页)

- Techniques. Washington DC: IEEE Computer Society, 2009: 227-236.
- [16] CHO S, JIN Lei. Managing distributed, shared L2 caches through OS-level page allocation [C]//Proc of the 39th Annual IEEE/ACM International Symposium on Microarchitecture. Washington DC: IEEE Computer Society, 2006: 455-468.
- [17] CHANG Ji-chuan, SOHI G S. Cooperative caching for chip multiprocessors [C]//Proc of the 33rd Annual International Symposium on Computer Architecture. Washington DC: IEEE Computer Society, 2006: 264-276.
- [18] KIM C, BURGER D, KECKLER S W. An adaptative, non-uniform cache structure for wire-delay dominated on-chip caches [C]//Proc of the 10th International Conference on Architectural Support for Programming Languages and Operating Systems. New York: ACM, 2002: 211-222.
- [19] LIRA J, MOLINA C, GONZALEZ A. Analysis of non-uniform cache architecture policies for chip-multiprocessors using the parsec benchmark suite [C]//Proc of the 2nd Workshop on Managed Multi-Core Systems. 2009: 1-8.
- [20] HERRERO E, GONZÁLEZ J, CANA R. Elastic cooperative caching: An autonomous dynamically adaptive memory hierarchy for chip multiprocessors [C]//Proc of the 37th Annual International Symposium on Computer Architecture. New York: ACM, 2010: 419-428.
- [21] LEE H, CHO S, CHILDERS B R. CloudCache: expanding and shrinking private caches [C]//Proc of the 17th International Symposium on High Performance Computer Architecture. Washington DC: IEEE Computer Society, 2011: 219-230.
- [22] MENG Jia-yuan, SKADRON K. Avoiding cache thrashing due to private data placement in last-level cache for manycore scaling [C]//Proc of the 27th International Conference on Computer Design. Washington DC: IEEE Computer Society, 2009: 282-288.
- [23] LU Qing-da, ALIAS C, BONDHUGULA U, *et al.* Data layout transformation for enhancing data locality on NUCA chip multiprocessors [C]//Proc of the 18th International Conference on Parallel Architectures and Compilation Techniques. Washington DC: IEEE Computer Society, 2009: 348-357.
- [24] GARCÍA-GUIRADO A, FERNÁNDEZ-PASCUAL R, ROS A, *et al.* DAPSCO: distance-aware partially shared cache organization [J]. *ACM Trans on Architecture and Code Optimization*, 2011, 8(4): 25(1-19).
- [25] CHAIKEN D, FIELDS C, KURIHARA K, *et al.* Directory-based cache coherence in large-scale multiprocessors [J]. *Computer*, 1990, 23(6): 49-58.
- [26] ZEBCHUK J, SRINIVASAN V, QURESHI M K, *et al.* A tagless coherence directory [C]//Proc of the 42nd International Symposium on Microarchitecture. New York: ACM, 2009: 423-434.
- [27] BARROSO L A, GHARACHORLOO K, MCNAMARA R, *et al.* Piranha: a scalable architecture based on single-chip multiprocessing [C]//Proc of the 27th Annual International Symposium on Computer Architecture. New York: ACM, 2000: 282-293.
- [28] GOLLA R. Niagara2: a highly threaded server-on-a-chip [C]//Proc of the 18th IEEE HotChips Symposium on High-Performance Chips. 2006.
- [29] MARTU M R, HILL M D. Virtual hierarchies to support server consolidation [C]//Proc of the 35th Annual International Symposium on Computer Architecture. New York: ACM, 2007: 46-56.
- [30] FERDMAN M, LOTFI-KAMRAN P, BALET K, *et al.* Cuckoo directory: a scalable directory for many-core systems [C]//Proc of the 17th International Symposium on High Performance Computer Architecture. New York: ACM, 2011: 169-180.
- [31] BAER J, WANG W. On the inclusion properties for multi-level cache hierarchies [C]//Proc of the 15th Annual International Symposium on Computer Architecture. New York: ACM, 1988: 345-352.
- [32] LEE H, CHO S, CHILDERS B R. Perfector: a fault-tolerant directory memory architecture [J]. *IEEE Trans on Computers*, 2010, 59(5): 638-650.
- [33] WENTZLAFF D, GRIFFIN P, HOFFMANN H, *et al.* On-chip interconnection architecture of the tile processor [J]. *IEEE Micro*, 2007, 27(5): 15-31.
- [34] BROWN E. 64-way chip gains Linux IDE, dev cards, design wins [EB/OL]. (2008-04-28). <http://www.linuxfordevices.com/c/a/News/64way-chip-gains-Linux-IDE-dev-cards-design-wins/>.
- [35] ROS A, ACACIO M E, GARCIA J M. A direct coherence protocol for many-core chip multiprocessors [J]. *IEEE Trans on Parallel and Distributed Systems*, 2010, 21(12): 1779-1792.
- [36] PUGSLEY S H, SPJUT J B, NELLANS D W, *et al.* SWEL: hardware cache coherence protocols to map shared data onto shared caches [C]//Proc of the 19th International Conference on Parallel Architectures and Compilation Techniques. New York: ACM, 2010: 465-476.
- [37] CHOI I, ZHAO M-S, YANG X, *et al.* Experience with improving distributed shared cache performance on tilera's tile processor [J]. *IEEE Computer Architecture Letters*, 2011, 10(2): 45-48.
- [38] LI Yong, A BOUSAMRA A, MELHEM R, *et al.* Compiler-assisted data distribution for chip multiprocessors [C]//Proc of the 19th International Conference on Parallel Architectures and Compilation Techniques. New York: ACM, 2010: 501-512.
- (上接第 4010 页)
- [48] MA' AYAN A. Insights into the organization of biochemical regulatory networks using graph theory analyses [J]. *Journal of Biological Chemistry*, 2009, 284(9): 5451-5455.
- [49] MASON O, VERWOERD M. Graph theory and networks in biology [J]. *IET Systems Biology*, 2007, 1(2): 89-119.
- [50] LUSCOMBE N M, BABU M M, YU H Y, *et al.* Genomic analysis of regulatory network dynamics reveals large topological changes [J]. *Nature*, 2004, 431(7006): 308-312.
- [51] ALEXANDER R P, KIM P M, EMONET T, *et al.* Understanding modularity in molecular networks requires dynamics [J]. *Science Signaling*, 2009, 81(2): 44.
- [52] IRONS D J, MONK N A M. Identifying dynamical modules from genetic regulatory systems: applications to the segment polarity network [J]. *BMC Bioinformatics*, 2007, 8: 413.
- [53] SORANZO N, RAMEZANI F, IACONO G, *et al.* Decompositions of large-scale biological systems based on dynamical properties [J]. *Bioinformatics*, 2012, 28(1): 76-83.
- [54] GRAMMATICOS B, CARSTEA A S, RAMANI A. On the dynamics of a gene regulatory network [J]. *Journal of Physics A: Mathematical and General*, 2006, 39(12): 2965-2971.
- [55] BABU M M, TEICHMANN S A, ARAVIND L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks [J]. *Journal of Molecular Biology*, 2006, 358(2): 614-633.