

doi: 10.3969/j.issn.1672-5565.2013.04.04

比较图元向量和点的聚类系数对差异网络的研究

肖碧玉, 李先彬, 沈良忠, 刘文斌*

(温州大学物理与电子信息工程学院, 浙江 温州, 325035)

摘要: 差异分析对于揭示生命体的生长、发育和衰老过程及疾病发生具有重大的意义, 基于网络的差异分析方法已经成为系统生物学的一个研究热点。网络节点往往通过与局部结构作用实现某种功能, 其与局部结构的关系变化, 很可能影响其功能。本文利用仿真实验的方法比较了图元向量和点的聚类系数两种局部结构测度的性能, 并且利用他们分别设计算法挖掘差异网络中模块化变化的基因簇。应用 AGEMAP 数据库中小鼠 12 个组织基因表达数据进行实验, 大部分聚类簇都高度显著富集与衰老相关的 GO 项。

关键词: 差异网络; 点的聚类系数; 图元向量; 小鼠衰老

中图分类号: R978.1+6 **文献标识码:** A **文章编号:** 1672-5565(2013)-04-264-07

Comparing graphlet orbit and clustering coefficient in differentially network

XIAO Bi-yu, LI Xian-bin, SHEN Liang-zhong, LIU Wen-bin*

(Department of Physics and Electronic information engineering, Wenzhou University, Wenzhou 325035, China)

Abstract: Differential analysis is very important for understanding the process of biological evolution and the progress of diseases. Recently, graph based differential analysis has turned to be a hot area in system biology. One node often works with its neighboring nodes, so that their functions will change if their connections change. In this paper, we compared graphlet orbit and clustering coefficient by random networks. We used them to mine within module differential co-expression clusters. Application to data for mice showed that most of the clusters are significant enriched in some GO terms related to aging.

Keywords: Differentially Network; Clustering Coefficient; Graphlet Orbit; Aging

系统生物研究表明: 生命体的生长、发育、衰老过程以及疾病的发生与基因之间相互作用的变化密切相关。随着以微阵列技术为代表的各种高通量技术的飞速发展, 人们可以同时观察到一个细胞中成千上万个基因的活动状况。如何利用这些数据挖掘出其中潜在变化, 对于揭示衰老及疾病的发生发展, 具有重要的生物学意义。因此, 基于基因表达谱数据的差异分析成为系统生物学的一个研究热点。基因差异表达的分析大致可以分为以下三个层次:

(1) 生物学中, 有些关键基因表达水平的高低往往是导致疾病产生的关键因素。因此, 最早的差异分析主要集中在单个基因表达水平的变化, 如

Jacob 等发现小鼠衰老期各组织的基因表达水平变化存在较大的差异, 可归类为神经组织、血管组织、类固醇反应组织三种衰老类型^[1]

(2) 基因之间往往是通过相互作用来实现某种功能, 因而挖掘基因作用关系的变化能够揭示基因(簇)与功能之间的关系。Remondini 通过比对基因时间序列网络的度的分布, 发现 c-myc 致癌基因的活性与基因的网络结构有密切关系^[2]。Voy 等通过研究网络边的变化, 确定小鼠受辐射影响的基因簇^[3]。Oldham 等比对人类和黑猩猩差异网络的拓扑重叠, 挖掘与生物驱动进化有关的基因集合^[4]。Omar Odiba 等在差异网络分析中将网络中介性及

收稿日期: 2012-04-28; 修回日期: 2012-07-23.

基金项目: 国家自然科学基金(61272018; 60970065; 61174162,); 浙江省自然科学基金(R1110261; Y1080227,); 浙江省新苗人才计划项目(2012R424051)资助。

作者简介: 肖碧玉, 女, 研究生, 主要研究方向为生物信息学、数据挖掘; E-mail: 616961187@qq.com.

* 通讯作者: 刘文斌, 男, 教授, 博士, 主要研究领域为生物信息学、数据挖掘、DNA 计算等; E-mail: wbliu6910@126.com.

中心性结合起来,挖掘出了生物标记(Biomarker)^[5]。Zhang 基于条件概率提出差异相关子网络(DDN),检测两个差异转录网络的拓扑变化显著性^[6]。Southworth 等构建基因带权差异网络,挖掘与小鼠衰老密切相关的基因模块^[7]。

(3)虽然基因模块往往与某种功能相关,但是这种功能往往需要其他基因模块的协调与配合,基因模块之间关系的变化也往往与功能的变化密切相关。Tesson 等提出了 DiffCoEx 算法,可以同时挖掘基因模块及基因模块间的变化^[8]。此外,在差异表达分析方面,Fang 研究了二组样本中表达水平高低的变化,提出了三种基因表达谱差异模式及其关系,这种模式特征的研究对于差异模式的挖掘具有重要的指导意义^[9-10]在生物网络中,节点往往通过与局部结构作用实现功能,其与局部结构的关系发生变化,可能会影响其功能。因此,如何理解和刻画网络节点的局部结构特征及变化,对于认识生物系统的结构和功能的关系具有非常重要的意义。点的聚类系数是比较经典的网络节点局部结构描述测度,人们利用它做出了许多有意义的结果^[16]。2003 年,Przulj 提出利用图元及图元向量刻画网络中节点邻域关系,并且基于它们研究了生物网络与随机网络的拓扑结构差异^[11-12]、癌症相关基因^[13]、生物网络的进化树^[14]等。本文利用仿真实验的方法分析和比较图元向量和点的聚类系数两个测度的性能,并且分别利用它们设计算法挖掘差异网络中模块化变化的节点簇,最后利用 AGEMAP 数据库中小鼠的基因表达数据进行实验,分析比较基于图元向量和基于点的聚类系数挖掘小鼠基因时序差异网络的差异模块与小鼠衰老的关系。

1 基本概念

点的聚类系数定义为:

$$c_v = \frac{2e_v}{d_v(d_v-1)} \quad (1)$$

其中 d_v 表示节点 v 的度,这 d_v 个节点之间实际存在的边数目 e_v 表示节点 v 及其相邻节点之间的边数。图元就是包含一定节点的非同构子图,其非同构位置唯一标号构成的向量称为图元向量。图 1 列出了包含 2,3,4 个节点的图元和其 15 维图元向量。由于随着图元节点规模的增大,图元向量的规模及计算的复杂度将指数级增加,本文采用图 1 中的 15 维的图元向量来研究差异网络。此外,由于标号大的图元向量可能包括标号小的维度信息(如标号 14 与 3,7 与 2),Milenkovi 按照标号大的图元向量包括

标号小的图元向量的个数对每一维图元向量定义一个权值 $w_k(k=0, \dots, 14)$ 。

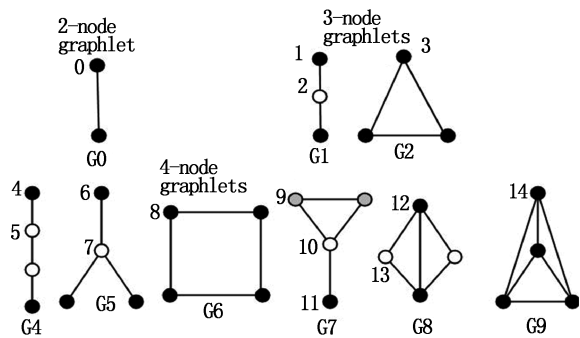


图 1 图元结构

Fig. 1 Graphlets and orbits

给定两个差异网络 $G(V, E)$ 和 $G(V, E')$, 具有相同的节点集不同边集,我们定义节点 $v \in V$ 的图元向量差异度为:

$$D_v = \frac{\sum_{k=0}^{14} w_k (\log(v_k + 1) - \log(v'_k + 1))}{w \log(\max\{v_k, v'_k\} + 2)} \quad (2)$$

其中 v_k 和 v'_k 分别表示节点 v 在两个差异网络 $G(V, E)$ 和 $G(V, E')$ 中的第 k 维图元向量。 w_k 表示图元向量第 k 维的图元向量的权值, $w = \sum_{k=0}^{14} w_k$ 度是反映网络节点邻接结构的一个非常重要的指标,结合图元向量和点的聚类系数两个测度,我们定义节点 $v \in V$ 的两种局部结构差异度为:

$$D_v^1 = d_v * D_v \quad (3)$$

$$D_v^2 = d_v * c_v - d'_v * c'_v \quad (4)$$

其中 d_v, d'_v 分别表示节点 v 在两个差异网络 $G(V, E)$ 和 $G(V, E')$ 中的度, c_v, c'_v 分别表示节点 v 在两个差异网络 $G(V, E)$ 和 $G(V, E')$ 中的点的聚类系数。直观上,网络中同一个簇的节点变化,应该具有一定的相似性,即对于一个簇中的节点 $u, v \in V'$, 他们的局部结构差异度可能基本相当。因此,我们给出衡量节点局部结构变化差距的定义如下:

$$L_{uv} = |D_u^x - D_v^x| \quad (5)$$

其中 D_u^x 对应 D_u^1 或 D_u^2, D_v^x 同上。

2 比较图元向量和点的聚类系数的性能

为了定量证明图元向量和点的聚类系数两个测度的性能,本文采用计算机产生随机网络作仿真数据,该实验方法已被广泛采用^[15-16]本仿真实验产生的每一个随机网络,包含 128 个节点,每个节点的度 $p=16$,这些点被划分成 4 个社团,每个点随机的与所在社团内部的 p_{in} 点相连,与其它的 $p_{out} = p - p_{in}$ 个

点相连。显然,随着 p_{out} 的增大,网络中的社团结构越来越模糊;反之,社团结构则越来越明显。一共产生了 3 组随机网络,分别对应 $p_{out} = 1, 5, 8$, 每组共 100 个网络。为了验证比较图元向量和点的聚类系数的性能,我们对这每一组的 100 个网络随机加(减) 10、20、30、40、50、60、70、80、90、100 条边进行分析。

日前,由于实验平台、实验技术等限制,生物数据还存在一定的误差,这对差异信息的研究带来了一定的困难,所以,现在人们主要是对较明显的差异信息进行研究,而往往把较小的变化当做噪声处理。本实验以加/减少量 10(1%)的边仿真噪声,以加/减较多的边 100(10%)仿真差异变化,用噪声鲁棒性衡量测度容忍噪声的能力,也就是计算在有噪

声干扰下测度度量时的变化量;用差异灵敏度度量测度发现差异变化的能力,也就是计算发生差异变化时测度发生的变化量。

为了分析和比较图元向量、点的聚类系数的噪声鲁棒性和差异灵敏度,我们分别对 3 类网络加(减) 10 条边和 100 条边,各节点的点的聚类系数和图元向量的变化量做统计(见图 2)。图中右侧 3 个图中,3 类网络中加(减) 100 条边时图元向量的变化量都高于加(减) 10 条边时的变化,图元向量测度能在容忍噪声的同时,有效提取差异信息,而图中左侧 3 类网络中点的聚类系数测度则不然。可见,相对点的聚类系数,图元向量的差异灵敏度和噪声鲁棒性比较好。

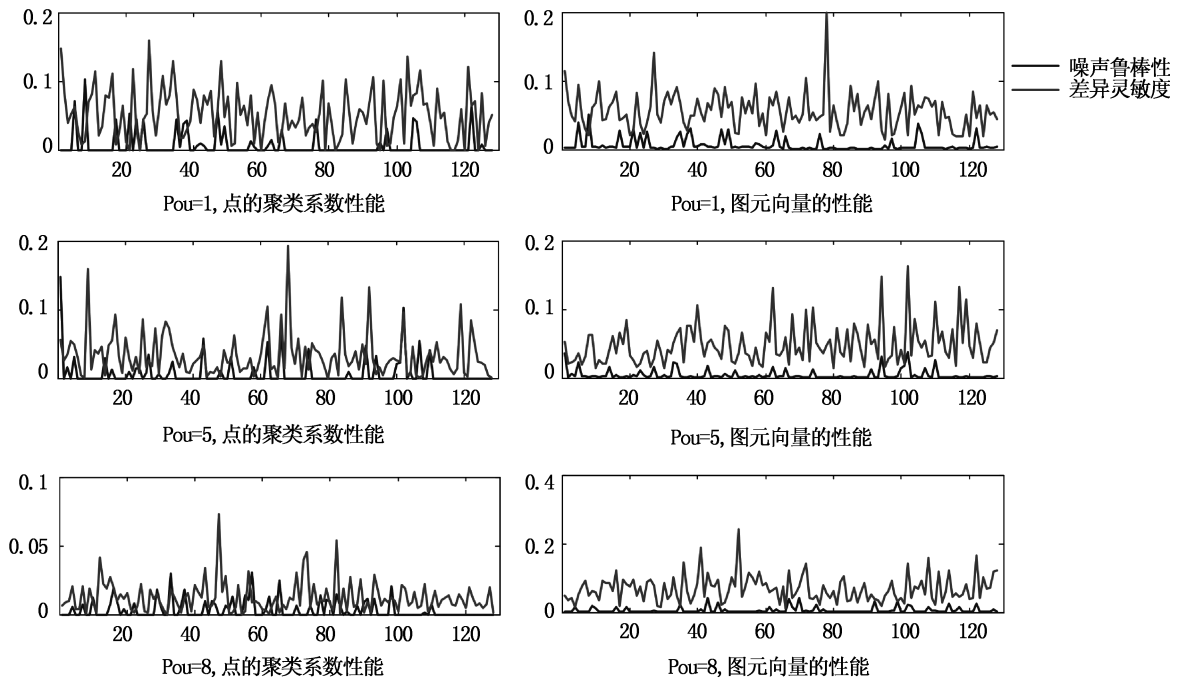


图 2 比较点的聚类系数和图元向量的噪声鲁棒性、差异灵敏度

Fig. 2 The comparison of robustness and difference sensitivity of clustering coefficient and orbits under various noises

点的聚类系数(图元向量)在各类网络中的噪声鲁棒性、差异灵敏度是否相同呢?我们统计 3 类网络加(减) 10 条边、100 条边时,各节点的点聚类系数、图元向量变化(见图 3)。图 3 右侧两个图中,在加(减) 10 条边的噪声鲁棒性研究,和加(减) 100 条边的差异灵敏性分析,图元向量在 3 类网络($P_{out} = 1, 5, 8$)中都相似。从图 3 左侧两个图中不难看出,在加(减) 10 条边的噪声鲁棒性研究,和加(减) 100 条边的差异灵敏性分析,各节点的点聚类系数的变化,在 $P_{out} = 1$ 的网络的噪声鲁棒性最差、差灵敏度最好,在 $P_{out} = 5$ 的网络噪声鲁棒性次之、差灵敏度也次之,在 $P_{out} = 8$ 的网络噪声鲁棒性最好、差灵敏度最差,点聚类系数在 3 类网络($P_{out} = 1, 5, 8$)中的噪声

鲁棒性和差异灵敏度不同。相对点的聚类系数,图元向量的适用性较强,应用范围较广。

我们统计 3 类网络加(减)边 10、20、30、40、50、60、70、80、90、100 条,局部结构变化的节点的数量(见图 4)。分析图 4 不难发现,在 3 类网络中,各种加(减)边时,图元向量变化的节点几乎是整个网络 128 个节点,而点的聚类系数变化的节点则随加(减)边的数量的增加而增多,可见,相对点的聚类系数,图元向量更能细致反应节点局部结构的影响。

综上所述,相对点的聚类系数,图元向量具更好的噪声鲁棒性和差异灵敏度,有较强的适用能力,但是在时间复杂性方面较差。点的聚类系数和图元向量各具优缺点,我们可以根据需要酌情选择。

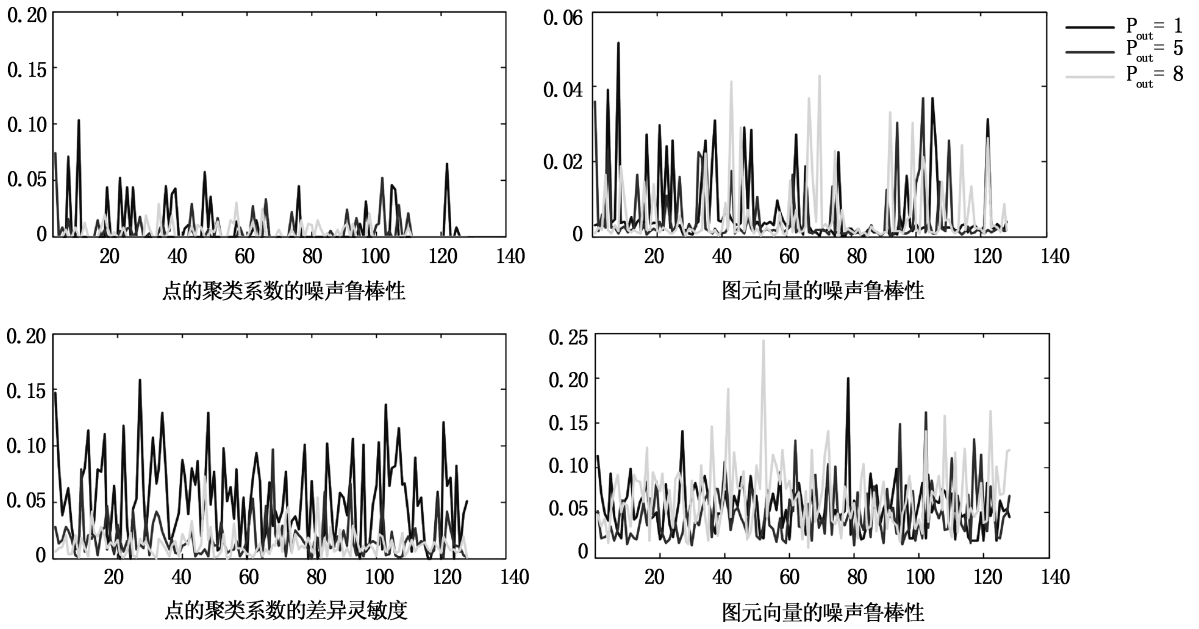


图 3 比较点的聚类系数和图元向量的适用性

Fig. 3 The comparison of clustering coefficient and orbits by using

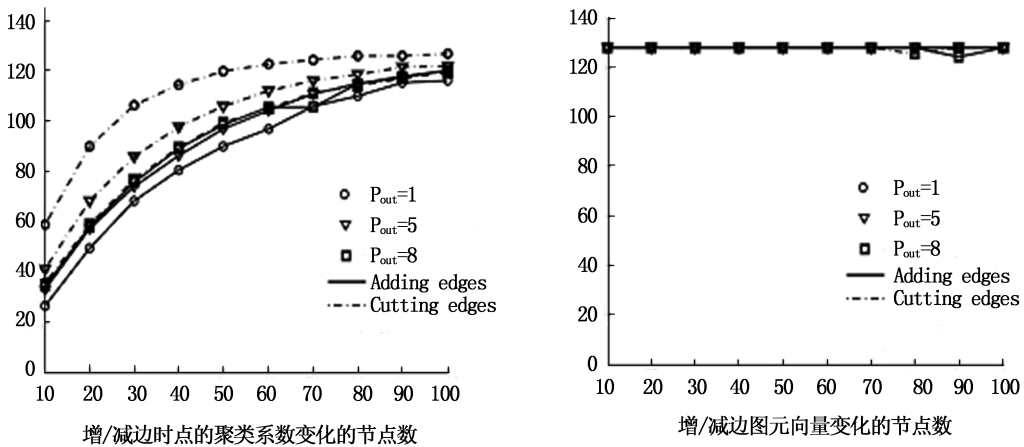


图 4 比较点的聚类系数和图元向量变化的节点数

Fig. 4 The comparison of clustering coefficient and orbits with the number of nodes Changed adding or cutting

3 实验

3.1 材料

本文生物数据来自 AGEMAP 数据库 (http://cmgm.stanford.edu/~kimlab/aging_mouse/), 其中包括 C57BL6 小鼠 16 个组织的 1、6、16、24 个月的 8 932 个基因表达数据, 由于其中 4 个组织数据存在较大噪声和不完整性, 本文使用 Adrenal Glands(1)、Cerebellum(2)、Cerebrum(3)、Eye(4)、Gonads(5)、Heart(6)、Hippocampus(7)、kidney(8)、Lung(9)、Muscle(10)、Spinal Cord(11)、Thymus(12) 等 12 个组织的数据。通常认为 16 月至 24 月为小鼠的衰老

期, 我们利用这两个月基因表达数据研究差异簇与小鼠衰老的关系。共表达网络是一个无向图, 每个节点表示一个基因, 每条边表示两个基因的共表达关系。共表达关系通过 Pearson 相关系数 r 确定

$$r = \frac{N \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{N \sum X_i^2 - (\sum X_i)^2} \sqrt{N \sum Y_i^2 - (\sum Y_i)^2}}$$

然后将 Pearson 系数 r 转化另一变量

$$r' = \sqrt{(n-2)r^2 / (1-r^2)},$$

其中, n 表示计算这两个基因相关系数时所用的数据点的个数。转化后的 r' 值服从自由度为 $n-2$ 的 t 分布。如果两个基因之间的 r' 值大于设定的 p-value 对应的 t 分布表, 则就在这两个基因之间加一条边;

否则在这两个基因之间就不加边。除 Eye 组织的 p-value 取值为 1E-04, Spinal Cord 组织为 1E-07、其他各组织均取 5E-06 构建网络。David 数据库可以对小鼠基因簇的功能进行分析,如果一个基因簇有 50% 以上的基因显著的共享一个或多个 GO 项,通常认为该基因簇具有生物学意义。本文 GO 项的 P-value 取 0.05。Southworth 等验证了 401 条与小鼠衰老相关的 GO 项^[8],因此,同其分析一个基因簇显著富集的 GO 项就可以判断其是否与衰老有关。下面我们通过图元向量和点的聚类系数二种测度,分别设计算法挖掘 16 和 24 月的基因表达数据中变化的基因簇,并对其功能进行分析。

3.2 挖掘差异基因簇

基于图元向量挖掘差异基因簇,首先,提取局部结构发生变化的差异基因;在 16 月(24 月)基因共表达网络中,取 D_v^1 大于等于 5 提取 16 月(24 月)的变化的基因。然后,根据提取出的基因多少和 D_v^1 分布,利用 L_m 分别在 16 月和 24 月基因共表达网络中进行 K-means 聚类,最后,对聚类簇进行 GO 注释和分析。基于点的聚类系数挖掘差异基因簇方法类似。

3.3 实验结果分析

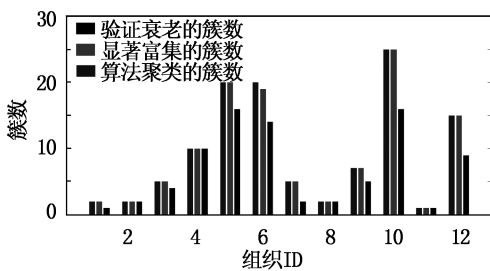
基于图元向量和点的聚类系数挖掘差异基因簇的实验结果如图 5。从图中可以看出,基于图元向量和点的聚类系数挖掘的差异基因簇,几乎每个基因簇都显著的富集一个或多个 GO 项,并且大部分的

基因簇都与衰老相关。二种方法挖掘聚类簇显著富集的 GO 项几乎相似,主要相似 GO 项见表 1,这些功能的差异与衰老具有密切关系。有些基因簇按照 Southworth 验证的 401 条衰老 GO 项与衰老无关,这是由于目前有关衰老相关的 GO 项还不完善,这方面的研究工作还在不断进行中,如 Chunxiao Fu^[17]、Jamie L. Barger^[18] 等也验证其它一些与小鼠衰老有关的 GO 项。

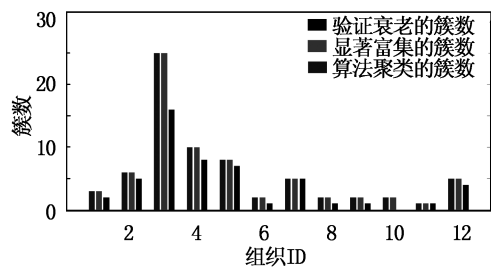
表 1 二种方法挖掘的相似 GO 项

Table 1 Similar GO terms from two algorithms

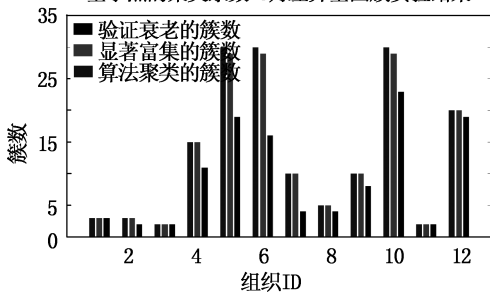
GO:0005488
GO:0005515
GO:0005622
GO:0005623
GO:0005737
GO:0008152
GO:0009987
GO:0043226
GO:0043227
GO:0043229
GO:0043231
GO:0044424
GO:0044464
GO:0044237
GO:0044238



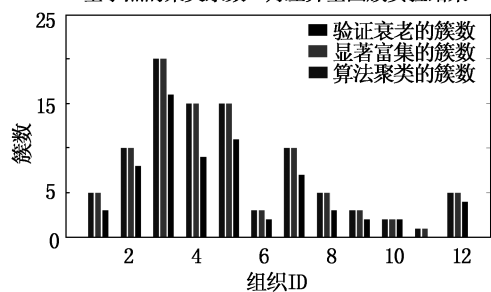
基于点的聚类系数16月差异基因簇实验结果



基于点的聚类系数24月差异基因簇实验结果



基于图元向量16月差异基因簇实验结果



基于图元向量24月差异基因簇实验结果

图 5 基于图元向量和点的聚类系数的实验结果

Fig. 5 Experimental results of the two algorithms based on Graphlet and Clustering Coefficient

二种方法取相同的阈值提取差异基因,同组织中基于图元向量提取的差异基因和基于点的聚类系数提取的差异基因大部分相同,且基于图元向量提取的差异基因相对较多(见表2)。

表2 比较各组织基于图元向量和点聚类系数提取的基因数

Table 2 Comparing the number of genes mined by the two algorithms

月份	组织	图元向量	聚类系数	交集
16个月	adrenal	479	254	227
	carebrum	958	882	754
	cerebellum	574	356	299
	eye	2 588	2 171	1 911
	gonads	3 335	3 169	3 057
	hearts	3 240	2 863	2 751
	hippocampus	1 167	797	759
	kindey	550	385	368
	lung	1 880	1 151	1 100
	mucle	3 634	3 315	3 262
	spinal	170	116	61
	thymus	2 178	2 133	572
24个月	adrenal	663	467	395
	carebrum	3 292	3 211	2 978
	cerebellum	1 488	951	877
	eye	2 169	1 846	1 635
	gonads	1 637	1 483	1 381
	hearts	518	324	270
	hippocampus	1 687	896	890
	kindey	790	390	383
	lung	774	457	416
	mucle	610	377	356
	spinal	85	64	47
	thymus	742	611	489

这与前面图元向量和点的聚类系数的仿真实验结论一致,图元向量更能反映局部结构的信息。相对点的聚类系数,图元向量的噪声鲁棒性和差异灵敏度较好,而基于图元向量和点的聚类系数的实验却得到了相似较好的实验结果,从而可知 AGEMAP 数据库提供的数据质量较高,而这点在其他的实验中也得到证明^[1,7]。此外,我们发现基于图元向量和点的聚类系数挖掘的有些基因簇在 David 数据库中不能注释到与衰老相关的 GO 项,但是在 AGEMAP 数据库中发现它们具有这些功能。如 Thymus 组织利用基于图元向量挖掘的一个基因簇(其中包括 196 个基因),在 David 数据库中注释衰老 GO:0005488~binding 项的富集度为 52.6%,显著性为 0.027。此簇中的 Mm.20935 基因不含 GO:0005488~binding 项功能,但在 AGEMAP 提供的数据中则显示此组织基因 Mm.20935 具有 GO:0005488~binding 功能。另外,Thymus 组织利用基于点的聚类系数挖掘的一个基因簇(其中包括 90

个基因)在 David 数据库中注释衰老 GO:0005488~binding 项的富集度为 60%,显著性为 0.013。此簇中的 Mm.383175 基因不含 GO:0005488~binding 项功能,但在 AGEMAP 提供的数据中则显示此组织基因 Mm.383175 具有 GO:0005488~binding 功能。因此,本文那些与衰老无关的基因(簇)存在潜在的小鼠衰老研究价值。

4 小结

生物分子往往是通过与局部结构相互作用发挥功能,其与局部结构关系发生变化可能引起其功能变化,所以基于局部结构测度进行差异研究对于认识生命的进化、发育、衰老过程及疾病的产生等生物问题具有重要的意义。本文首先分析和比较了图元向量和点的聚类系数两个测度的性能,并基于图元向量和点的聚类系数分别设计了挖掘模块化变化簇的算法。利用 AGEMAP 数据库中小鼠 12 个组织的基因表达数据进行实验,本文设计的二个算法挖掘的差异簇都显著的富集于一些 GO 条目,而且其中大部分都与衰老有关。

图元向量和点的聚类系数作为刻画网络局部拓扑结构的二种参数,各具优缺点,在差异分析中具有广阔的应用前景。本文是基于局部结构变化挖掘模块化变化基因簇,而基于模块基因簇识别变化模块基因簇将是本文后续工作。

参考文献(References)

- [1] Jacob M. Zahn, Suresh Poosala, Art. B Owen. AGEM AP: A Gene Expression Database for Aging in Mice [J]. PLOS Genetics, 2007, 3(11): e201.
- [2] Remondini D, O'Connell B, Intrator N, Sedivy JM, Neretti N, Castellani GC, Cooper LN. Targeting c-Myc-activated genes with a correlation method: Detection of global changes in large gene expression network dynamics [J]. Proc Natl Acad Sci U S A, 2005, 102(19): 6902-6906.
- [3] Voy BH, Scharff JA, Perkins AD, Saxton AM, Borate B, Chesler EJ, Branstetter LK, Langston MA. Extracting gene networks for low-dose radiation using graph theoretical algorithms [J]. PLoS computational biology, 2006, 2(7): 757-768.
- [4] Oldham Michael C, Horvath Steve, Geschwind Daniel H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains [J]. Proceedings of the National Academy of Sciences, 2006, 103(47): 17973-17978.
- [5] Omar Odibat. Ranking Differential Genes in Co-expression Networks [J]. Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine, 2011, 350-354.
- [6] Zhang B, Li H, Riggins RB, zhan M, Xuan J, Zhang Z, Hoffman EP, Charke R, Wang Y. Differential dependency network analysis to identify condition-specific topological changes in biological

- networks[J]. *Bioinformatics*, 2009, 25(4):526–532.
- [7] Lucinda K. Southworth, Art B. Owen, Stuart K. Kim. Aging Mice Show a Decreasing Correlation of Gene Expression within Genetic Modules[J]. *PLOS Genetics*, 2009, (5):e1000776.
- [8] Bruno M Tesson, Rainer Breiling and Ritsert C Jansen. Diffcoex; a simple and sensitive method to find differentially coexpressed gene modules[J]. *BMC Bioinformatics*, 2010, 11(1):497.
- [9] Gang Fang, Gaurav Pandey. Mining Low-Support Discriminative Patterns from Dense and High-Dimensional Data. *Knowledge and Data Engineering [J]. IEEE Transactions*, 2012, 24(2):279–294.
- [10] Gang Fang, Rui Kuang, Gaurav Pandey, Michael Steinbach, Chad L. Myers, Vipin Kumar. Subspace differential coexpression analysis: Problem Definition and a General Approach[J]. *Pacific Symposium on Biocomputing*, 2010, 15: 145–156.
- [11] Natasa Przulj. Biological network comparison using graphlet degree distribution [J]. *Cancer Inform*, 2008, 6: 257–273.
- [12] Przulj N, Corneil D G, Jurisica I. Modeling Interactome: Scale-Free or Geometric [J]. *Bioinformatics*, 2004, 20(18): 3508–3515.
- [13] Tijana Milenkovic and Natasa Przulj. Uncovering Biological Network Function via Graphlet Degree Signatures [J]. *Cancer Inform*, 2008, 6:257–273.
- [14] Oleksii Kuchaiev, Tijana Milenkovic, Vesna Memisevic, Wayne Hayes, Natasa Przulj. Topological network alignment uncovers biological function and phylogeny[J]. *Journal of the Royal Society Interface*, 2011, 5(17):1341–1354.
- [15] Yang B., Liu D. Y., Liu J. M.. Complex Network Clustering Algorithms[J]. *Journal of Software*, 2009, 20(1):54–66.
- [16] Radicchi F., Castellano C., Cecconi F. Defining and identifying communities in networks[J]. *PNAS*, 2004, 101(9):2658–2663.
- [17] Fu Chunxiao, Hickey Morgen, Morrison Melissa, McCarter Roger, Han Eun-Soo. Tissue specific and non-specific changes in gene expression by aging and by early stage CR [J]. *Mechanisms of Ageing and Development*, 2006, (127):905–916.
- [18] Barger Jamie L, Kayo Tsuyoshi, Vann James M, Arias Edward B, Wang Jelai, Hacker Timothy A, Wang Ying, Raederstorff Daniel, Morrow Jason D, Leeuwenburgh Christiaan, Allison David B, Saupé Kurt W, Cartee Gregory D, Weindruch Richard, Prolla Tomas A. A Low Dose of Dietary Resveratrol Partially Mimics Caloric Restriction and Retards Aging Parameters in Mice [J]. *PLoS ONE*, 2008, (3):e2264.