

# 基于模体的复杂网络测度量研究\*

韩华 刘婉璐 吴翎燕<sup>†</sup>

(武汉理工大学理学院, 武汉 430070)

(2012年3月30日收到; 2013年4月28日收到修改稿)

针对复杂网络拓扑结构中模体的存在性, 在传统的顶点度和边聚类系数定义的基础上, 提出了基于模体的顶点度和边度来衡量网络中顶点和边的重要性. 用 Rand-ESU 算法对不同规模的 8 个网络进行模体检测, 验证了网络中模体的存在性, 重点分析了 Karate 网络和 Dolphin 网络中模体的结构和特征. 用 Pearson 相关系数衡量基于模体的顶点度与传统顶点度、基于模体的边度与边聚类系数的相关性, 仿真分析结果表明相关性大小与模体种类有关, 基于模体的顶点度和边度是对原定义的一种改进和拓展, 更全面地刻画了顶点和边在网络中的重要性.

**关键词:** 模体, 顶点度, 边度, Pearson 相关系数

**PACS:** 89.75.-k, 89.75.Fb

**DOI:** 10.7498/aps.62.168904

## 1 引言

复杂网络已成为当前最重要的多学科交叉研究领域之一. 要理解网络结构和网络行为之间的关系, 并进而考虑改善网络的行为, 就需要对实际网络的拓扑结构特征有很好的了解<sup>[1]</sup>. 在研究网络拓扑结构和动态特性的过程中, 很重要的一步就是识别网络的模体 (motif) 和社团 (community). 模体是网络的基本拓扑结构之一, 它的大小介于网络个体和社团之间, 一般由少数几个节点连接构成, 模体也揭示了网络的演化规律, 它是社团内部成员之间基本的连接模式<sup>[2]</sup>.

Milo 等<sup>[3,4]</sup> 首先提出了网络模体的概念. 网络模体是网络的子图, 在实际网络中模体出现的可能性比随机网络的期望值要高, 他们具体研究了生物神经网络、食物链网络以及万维网等实际网络, 发现不同的网络中存在的模体结构也不同. 基于这一点, Barabasi 和 Oltvai<sup>[5]</sup> 也对模体进行了研究, 认为模体在实际网络演化中扮演了重要的角色. 随后, Milo 等<sup>[6]</sup> 又提出了一种衡量模体重要性的概念——Z 得分, 利用 Z 得分还可以进一步提出网络的形象特征的概念来描述网络中各个模体

的重要程度. Onnela 等<sup>[7]</sup> 研究了加权网络中的模体, 提出了加权网络中基于模体的子图密度和重要性的测度量. Luciano 等<sup>[8]</sup> 也指出模体是网络拓扑结构演化过程中的重要结构. Krumov<sup>[9]</sup> 研究了模体在 Internet 网络中 P2P 技术优化中的应用, 从动态角度首次提出了使用网络模体来优化网络拓扑结构, 并提出了模体签名、新的 Kademia 协议算法. 发现新的基于模体的 P2P 网络拥有许多优化效果, 新的网络拓扑结构更加稳定, 网络传输、共享速度更快, 为 BT 客户端的隐私信息提供了更高的安全保证, 使对网络的恶意攻击变得几乎不可能. Kotorowicz 和 Kozitsky<sup>[10]</sup> 用模体结构研究了 Ising 模型的临界点, 使其拓扑性质更加明显. Squartini 和 Gahschelli<sup>[11]</sup> 研究了世界贸易网络的模体结构, 描述了 13 种三节点的模体的结构特征及其在网络中的重要性. 国内董昭和李翔<sup>[12]</sup> 研究了离散时间序列中网络模体的分布特征, 他们将离散时间序列转换成复杂网络的形式, 利用模体来研究时间序列的时域和相空间特征. 缪莉莉等<sup>[13]</sup> 研究了科学家合作网的模体结构, 提供了基于模体的科学家合作网的结构辨识方法. 刘亮等<sup>[14]</sup> 基于模体和超家族辨识方法, 研究了国家关键交通网的基元结构特性,

\* 国家自然科学基金 (批准号: 71140015) 和中央高校基本科研业务费 (批准号: 2013-Ia-007, 2013-Ia-040) 资助的课题.

<sup>†</sup> 通讯作者. E-mail: 815007651@qq.com

可用于定义交通系统并区分社会或生态系统.

随着复杂网络小世界性和无标度性的研究发现, 人们对存在于不同领域的大量实际网络的拓扑性质进行了广泛的实证性研究. 在此基础上, 人们从不同的角度出发提出了各种各样的网络拓扑结构模型, 通过平均路径长度、聚类系数、度分布等刻画网络拓扑特征. 传统的聚类系数、顶点度等定义没有考虑网络中模体的结构特征, 导致在网络的拓扑研究及演化规律方面存在欠缺. 本文首先验证网络中模体的存在性, 在此基础上研究基于模体的测度理论, 提出基于模体的顶点度、基于模体的边度等衡量网络拓扑性质的测度量, 并通过基于模体的顶点度与传统的顶点度、基于模体的边度与边聚类系数的相关性分析验证新的测度理论的合理性和优越性.

## 2 模体的基本理论

### 2.1 模体的定义

模体是满足下列条件的网络子图: 1) 该子图在与真实网络对应的随机网络中出现的次数大于它在真实网络中出现的概率是很小的, 通常要求这个概率小于某个阈值  $P$ , 如  $P = 0.01$ ; 2) 该子图在真实网络中出现的次数  $N_{\text{real}}$  不小于某个下限  $U$ , 如  $U = 4$ ; 3) 该子图在真实网络中出现的次数  $N_{\text{real}}$  明显高于它在随机网络中出现的次数  $N_{\text{rand}}$ , 一般要求  $(N_{\text{real}} - N_{\text{rand}}) > 0.1N_{\text{rand}}$ .

### 2.2 模体的基本统计特征

#### 1) 模体的频率

对于给定的具有  $n$  节点的子图  $V$ , 它在实际网络中出现的次数为  $n(V)$ , 所有具有  $n$  节点的子图出现的总次数为  $N$ , 那么子图  $V$  出现的频率为

$$f(V) = \frac{n(V)}{N}, \quad (1)$$

如果子图  $V$  是网络的模体, 那么该频率称为模体的频率.

#### 2) 模体的 $P$ 值

对于模体  $M$ , 它在随机网络中出现的次数大于在实际网络中出现的次数的概率即为  $P$  值.  $P$  值越小, 说明该模体在网络中越重要.

#### 3) 模体的 $Z$ 得分

对于模体  $M_i$ , 它在真实网络中出现的次数记为  $N_{\text{real}_i}$ , 它在随机网络中出现的次数记为  $N_{\text{rand}_i}$ ,  $N_{\text{rand}_i}$  的平均值记为  $\langle N_{\text{rand}_i} \rangle$ , 标准差为  $\sigma_{\text{rand}_i}$ , 那么模体  $M_i$  在该真实网络中的  $Z$  得分为

$$Z_i = \frac{N_{\text{real}_i} - \langle N_{\text{rand}_i} \rangle}{\sigma_{\text{rand}_i}}. \quad (2)$$

将  $Z_i$  归一化, 有

$$SP_i = \frac{Z_i}{\sqrt{\sum_i Z_i^2}}. \quad (3)$$

$Z$  得分用来衡量模体的重要性,  $Z$  得分越大表明模体在网络中越重要.  $Z$  得分的归一化处理是为了强调子图的相对重要性, 这对于比较不同规模的网络是重要的, 因为大型网络中的模体比小型网络中的模体具有更高的  $Z$  得分.

### 2.3 模体的结构

模体一般由少数几个节点连接构成, 在大多数网络中, 三个节点和四个节点构成的模体较为常见. 在有向网络中, 三个节点之间可以形成 13 种不同的连通关系, 四个节点之间可以形成 199 种不同的连通关系<sup>[15]</sup>. 无向网络中的模体结构是有向网络中模体结构的特殊情况, 对于无向网络, 三个节点构成的模体有两种结构, 如图 1 所示. 无向网络中四个节点构成的模体结构(部分)如图 2 所示.



图 1 三节点的模体

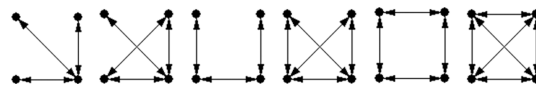


图 2 四节点的模体(部分)

## 3 基于模体的测度量


顶点度和边聚类系数分别用来衡量网络中顶点和边的重要性, 但其定义的提出并没有考虑网络中模体结构的存在, 而模体是网络中重要的拓扑结构, 故在此提出基于模体的测度量来定量分析顶点和边的性质.

### 3.1 基于模体的顶点度

度是单独节点的属性中简单而又重要的概念. 节点  $i$  的度  $k_i$  定义为与该节点连接的其他节点的数目. 直观上看, 一个节点的度越大就意味着这个节点在某种意义上越“重要”. 网络中所有节点  $i$  的度  $k_i$  的平均值成为网络的 (节点) 平均度, 记为  $\langle k \rangle$ . 网络中节点的度的分布情况可用分布函数  $P(k)$  来描述 [16].

传统的顶点度的定义没有考虑到网络中不同结构的模体的存在, 用顶点度来衡量网络拓扑性质时忽略了网络中存在模体这一事实. 基于此我们提出基于模体的顶点度的定义.

对于任意一个网络  $G$ ,  $M$  为它的模体, 定义包含节点  $i$  的模体  $M$  的个数为节点  $i$  的模体顶点度, 记作  $k_i^M$ .

事实上, 我们可以把两个节点和其连边看成一个两节点的模体 (), 则传统顶点度的定义是指包含节点  $i$  的两节点模体的个数, 是我们提出的基于模体顶点度的特殊情况.

### 3.2 基于模体的边度

边的性质在复杂网络拓扑研究中起着至关重要的作用, 如边聚类系数, 它体现了网络的聚类特性. 边聚类系数定义为实际包括该边的三角形的数目与所有可能包括该边的三角形的数目之比 [17]. 对于连接节点  $i$  和  $j$  的边, 其聚类系数定义为

$$C_{ij}^{(3)} = \frac{z_{ij}^{(3)} + 1}{\min[k_i - 1, k_j - 1]}, \quad (4)$$

其中  $z_{ij}^{(3)}$  表示包含该边的三角形的数目.

边聚类系数是用来分析网络中某条边周围的连接紧密程度的测度量, 本质上刻画了该边在周围的一种特殊模体结构 (由三个节点构成的完全图, 也就是三角形) 中的重要程度. 然而, 不同网络中模体的结构不同, 如层次树结构网络中不存在三角形结构的模体, 若用边聚类系数来衡量该网络中边的重要性则没有意义, 所以需要提出新的测度量来定量分析网络中边的性质.

对于任意一个网络  $G$ ,  $M$  为它的模体, 定义包含边  $e$  的模体  $M$  的个数为边  $e$  的模体边度, 记作  $C_e^M$ .

事实上, 边聚类系数的定义即是包含该边的三角形模体的个数与所有可能包括该边的三角形模体的数目之比, 故基于模体的边度的定义是对边聚类系数定义的一种延伸和拓展.

## 4 仿真分析

针对文中提出的基于模体的顶点度、边度的定义, 下面将用实际网络进行仿真分析. 研究基于模体的顶点度与传统顶点度、基于模体的边度与边聚类系数的相关性, 以此验证新提出的测度量的合理性.

### 4.1 网络模体检测分析

为了验证基于模体的顶点度和边度的合理性, 首先要验证模体的确存在于网络中, 同时要研究模体的结构和特征, 即模体的种类、数量以及模体的重要程度等特性.

#### 4.1.1 模体的存在性

选取规模大小不同的 8 个网络进行模体检测分析, 分别为 Karate 网络 [18]、Dolphin 网络 [19]、Political books 网络 [20]、Football 网络 [21]、Jazz 网络 [22]、Email 网络 [23]、Newman science coauthors 网络 [24] 和 PGP 网络 [25]. 在此, 我们只验证三角形模体的存在性, 其他类型模体的存在性可类似验证; 在此基础上, 对 Karate 网络和 Dolphin 网络进行详细的模体分析.

目前, 检测模体的软件主要有 Mfinder 和 Fanmod. Fanmod 软件运用 Rand-ESU 算法进行模体检测 [26], 该算法速度较快, 检测的模体的种类更多 (可检测出三个节点到八个节点的模体), 且适用于检测各种规模网络的模体. 故选取 Rand-ESU 算法对网络进行模体检测分析, 结果见表 1.

从表 1 中可以看出: 在这些网络中, 三角形子图在真实网络中出现的频率远高于其在随机网络中出现的频率, 且它们的  $Z$  得分相对较高,  $P$  值均为 0, 小于 0.01, 故三角形子图是网络中的模体.

通过以上研究可以认为模体是网络中真实存在的结构, 下面将选取 Karate 网络和 Dolphin 网络做进一步的研究, 即研究模体的种类和结构特征等.

表 1 网络中三角形模体的存在性

网络名称	顶点数	边数	频率/%	平均频率/%	标准差	Z 得分	P 值
Karate	34	78	10.274	0.59849	0.0024305	39.808	0
Dolphin	62	159	12.96	0.19837	0.0014193	89.921	0
Political books	105	441	15.127	0.044314	0.0014656	102.91	0
Football	115	613	18.634	0.040642	0.00083655	222.26	0
Jazz	198	2742	26.551	0.37816	0.0046466	56.326	0
Email	1133	5451	6.2324	0.0095806	0.00033402	186.3	0
Newman science coauthors	1589	2742	42.988	0.0021198	$7.4405 \times 10^{-5}$	5777.3	0
PGP	10680	24316	16.846	0.0016534	$6.4492 \times 10^{-5}$	2611.9	0

4.1.2 Karate 网络

Karate 网络是 Zachary 根据美国一所大学的空手道俱乐部内部关系所构造的, 它反映了该俱乐部内部管理者与管理者, 管理者与学员以及学员与学员之间的相互人际关系. 其网络拓扑结构图如图 3 所示.

使用 Rand-ESU 算法对 Karate 网络进行 3 个节点和 4 个节点的模体检测分析得到表 2 和表 3 的

结果.

从表 2 中可以看出: 编号为 238 的子图即三角形子图的  $P$  值小于 0.01, 它在真实网络中出现的频率远高于其在随机网络中出现的频率, 它的  $Z$  得分最高, 为 39.808, 因此三角形子图是 Zachary's karate club 网络的模体. 编号为 78 的子图的  $P$  值为 1, 远大于 0.01, 且它在随机网络中出现的频率高于其在真实网络中出现的频率, 故它不是该网络的模体.

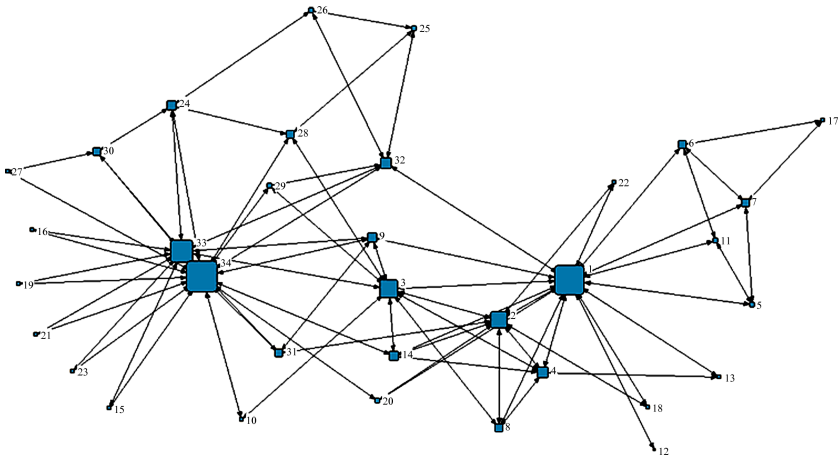


图 3 Karate 网络拓扑结构图







表 2 Karate 网络的 3 节点模体检测分析

模体编号	形状	频率%	平均频率%	标准差	Z 得分	P 值
78		89.726	99.402	0.0024305	-39.808	1
238		10.274	0.59849	0.0024305	39.808	0

从表 3 中可以看出: 编号为 13278, 4958 和 31710 的子图的  $P$  值均为 0, 小于 0.01, 且它们在真实网络中出现的频率远高于在随机网络中出现的频率, 故这些子图即可称之为模体. 另外, 从模体的重要程度来看, 编号为 31710 的模体的  $Z$  得分最

高, 为 421.17, 故认为该种模体在该网络中最重要. 而编号为 8598, 27030, 4382 的子图的  $P$  值均远大于 0.01, 且它们在随机网络中出现的频率高于其在真实网络中出现的频率, 它们的  $Z$  得分也均为负值, 故这些子图不是该网络的模体.

表 3 Karate 网络的 4 节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 得分	P 值
13278		3.5971	0.010251	0.0003562	100.7	0
8598		28.819	43.911	0.0035818	-42.134	1
4958		19.128	1.4001	0.0034275	51.723	0
27030		1.5235	6.5154	0.0023433	-21.303	1
4382		46.466	48.163	0.0027549	-6.1598	0.998
31710		0.46551	$1.3857 \times 10^{-5}$	$1.1053 \times 10^{-5}$	421.17	0

4.1.3 Dolphin 网络

Dolphin 网络是 Lusseau 通过观察生活在新西兰 Doubtful Sound 的 62 只海豚之间的联系情况构建的网络, 网络中每个节点代表一只海豚, 连边表示两只海豚之间有频繁的接触. 其网络拓扑结构如图 4 所示.

使用 Rand-ESU 算法对 Dolphin 网络进行 3 个节点和 4 个节点的模体检测分析得到表 4 和表 5

的结果.

从表 4 和表 5 中可以看出: 对于 3 节点的子图, 编号为 238 的子图 (三角形) 是该网络中的模体; 对于 4 节点的子图, 编号为 13278, 4958, 31710 的子图是该网络中存在的模体, 其中编号为 31710 的模体的 Z 得分最高, 为 3036.7, 故该模体在 Dolphin 网络中最重要.

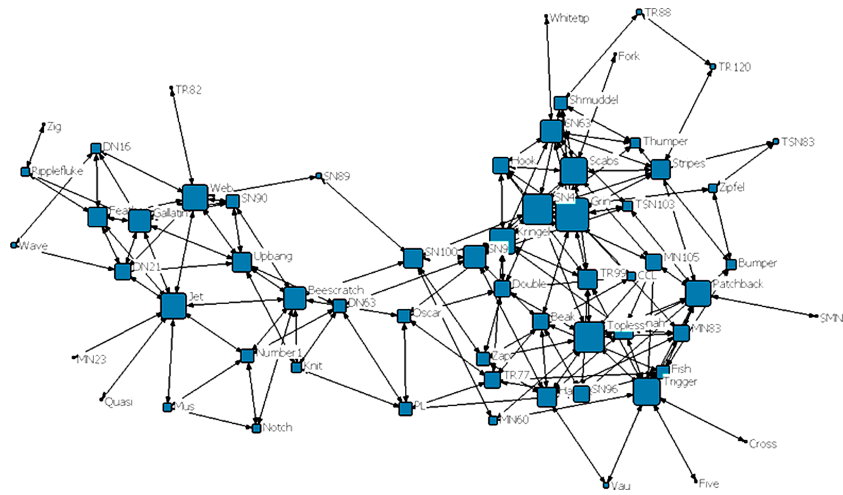








图 4 Dolphin 网络拓扑结构图

表 4 Dolphin 网络的 3 节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 得分	P 值
78		87.04	99.802	0.0014193	-89.921	1
238		12.96	0.19837	0.0014193	89.921	0

表 5 Dolphin 网络的 4 节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 得分	P 值
13278		3.6316	0.02335	0.00036295	99.415	0
8598		55.237	65.569	0.0038687	-26.708	1
4958		20.211	0.45376	0.0031707	62.311	0
27030		1.5526	4.5662	0.0031404	-9.5963	1
4382		18.658	29.387	0.0029719	-36.103	1
31710		0.71053	$3.309 \times 10^{-6}$	$2.3398 \times 10^{-6}$	3036.7	0

由 Karate 网络和 Dolphin 网络的模体检测结果, 我们发现模体的确是真实存在于网络中的, 且模体有多种种类和结构, 这也为进一步的研究奠定了基础, 提供了必要的依据.

4.2 基于模体的顶点度和边度分析

根据网络拓扑性质中顶点度和边聚类系数的

定义, 并结合网络中模体的概念, 我们提出了基于模体的顶点度和边度. 下面用 Karate 网络和 Dolphin 网络进行仿真分析, 将基于模体的顶点度和边度与传统的顶点度和边聚类系数进行对比分析, 研究其相关性.

图 5 和图 6 展示了 Karate 网络和 Dolphin 网络中基于三节点和四节点的模体顶点度和传统顶点

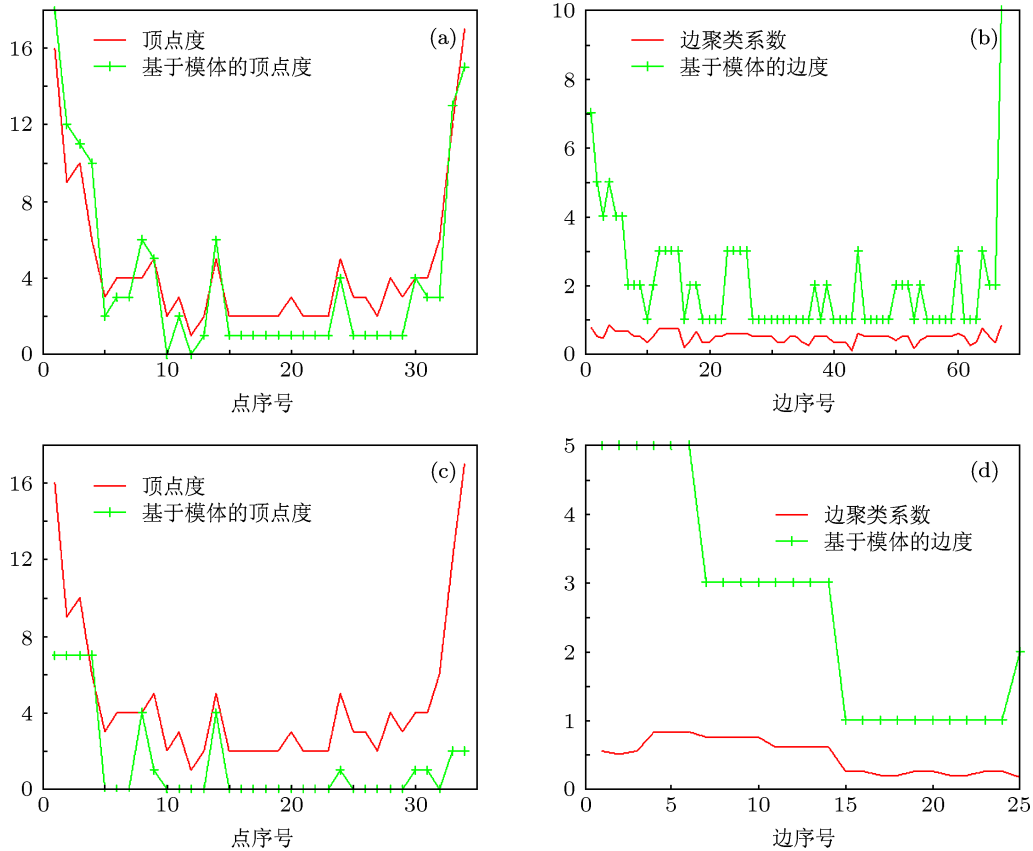


图 5 Karate 网络相关性分析 (a) 基于三节点的模体顶点度与传统顶点度; (b) 基于三节点的模体边度与边聚类系数; (c) 基于四节点的模体顶点度与传统顶点度; (d) 基于四节点的模体边度与边聚类系数



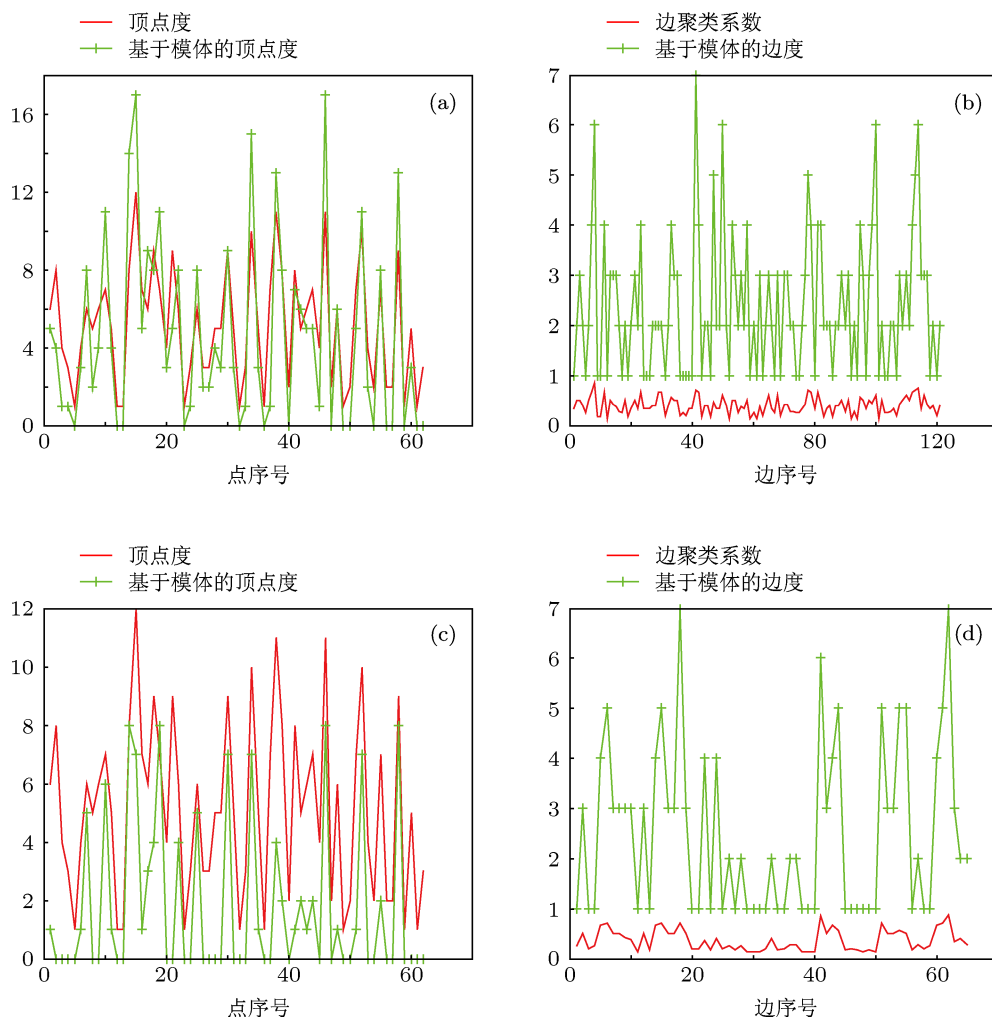


图6 Dolphin网络相关性分析 (a) 基于三节点的模体顶点度与传统顶点度; (b) 基于三节点的模体边度与边聚类系数; (c) 基于四节点的模体顶点度与传统顶点度; (d) 基于四节点的模体边度与边聚类系数

度的对比、边度与边聚类系数的对比情况. 其中, 三节点模体是编号为 238 的三角形模体, 四节点模体是编号为 31710 的模体, 是 Karate 网络和 Dolphin 网络中最重要的四节点模体. 由图中可以看出, 基于模体的顶点度和传统顶点度、基于模体的边度和边聚类系数的走势整体是呈正相关性. 为了更加明确其相关程度, 我们研究其 Pearson 相关系数.

Pearson 相关系数用来度量两个变量  $X$  和  $Y$  之间的相互关系, 它在学术研究中被广泛应用来度量两个变量线性相关性的强弱, 一般用  $r^{[27]}$  表示:

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}}. \quad (5)$$

Karate 网络和 Dolphin 网络中基于模体的顶点度与传统顶点度的相关性、基于模体的边度和边

聚类系数的相关性如表 6 和表 7 所示.

从 Karate 网络和 Dolphin 网络中都可以看出, 基于编号为 238, 13278, 4958 的模体顶点度与传统顶点度的相关系数很高, 这说明基于模体的顶点度与传统顶点度有很强的相关性, 此时, 认为传统顶点度与基于模体的顶点度所表示的顶点的重要性很相似. 而基于编号为 31710 的模体顶点度与传统顶点度的相关系数较低, 这说明基于该模体的顶点度与传统顶点度相关性不高, 此时, 传统顶点度与基于模体的顶点度所表示的顶点的重要性存在一定偏差. 传统顶点度事实上是基于模体的顶点度的一种特殊情况, 即传统顶点度是基于两节点的模体的顶点度. 故我们提出的基于模体的顶点度是对传统顶点度定义的改进和拓展, 更加全面地刻画了顶点在网络中的重要性.

表 6 Karate 网络相关系数

模体编号	基于模体顶点度与传 统顶点度的相关系数	基于模体的边度与边 聚类系数的相关系数
238	0.9514	0.6550
13278	0.8815	0.5077
4958	0.9469	0.4405
31710	0.6540	0.8167

表 7 Dolphin 网络相关系数

模体编号	基于模体顶点度与传 统顶点度的相关系数	基于模体的边度与边 聚类系数的相关系数
238	0.8878	0.8264
13278	0.8240	0.8489
4958	0.9068	0.4059
31710	0.7192	0.9359

Karate 网络中基于编号为 31710 的模体的边度与边聚类系数的相关系数较高, Dolphin 网络中基于编号为 238, 13278 和 31710 的模体的边度与边聚类系数的相关系数较高, 这些情况下认为边聚类系数与基于模体的边度所表示的边的重要性很相似. 而 Karate 网络中基于编号为 238, 13278, 4958 的模体的边度与边聚类系数的相关系数较低, Dolphin 网络中基于编号为 4958 的模体的边度与边聚类系数的相关系数较低, 此时认为边聚类系数与基于模体的边度所表示的边的重要性存在一定偏差. 边聚类系数本质上是刻画了该边在周围的一种特殊模

体结构 (即三角形) 中的重要程度, 它也是基于模体边度的一种特殊情况, 而且对于不存在三角形模体的网络来说, 用边聚类系数来刻画边的重要性没有意义. 故我们提出的基于模体的边度是对传统边聚类系数定义的一种改进和拓展, 更加全面地刻画了边在网络中的重要性.

## 5 结 论

模体结构是复杂网络甚至复杂系统的基本构成模式, 本文通过对 8 个大小规模不同的网络进行模体检测分析, 验证了网络中模体的存在性, 重点分析了 Karate 网络和 Dolphin 网络的三节点模体和四节点模体的结构、性质等. 在此基础上提出了新的衡量网络拓扑性质的测度量, 即基于模体的顶点度、基于模体的边度. 事实上, 传统顶点度与边聚类系数分别是基于模体的顶点度与边度的一种特殊情况, 故新的基于模体的测度量是在原测度量的基础上改进拓展得到的. 通过基于模体的顶点度与传统顶点度、基于模体的边度与边聚类系数的相关性分析, 验证了基于模体顶点度与边度的合理性和优越性. 在以后的研究中, 可以将新提出的基于模体的测度量与其他衡量网络拓扑特征的测度量 (如平均最短路径、边介数等) 进行相关性分析, 更加全面地验证基于模体的测度量的优越性. 另外, 可以继续探索基于模体的其他测度量, 如基于模体的模块度, 以此来研究基于模体的社团结构划分等.

- [1] Wang X F, Li X, Chen G R 2006 *Thory and Application of Complex Networks* (Beijing: Tsinghua University Press) p18 (in Chinese) [汪小帆, 李翔, 陈关荣 2006 复杂网络理论及其应用 (北京: 清华大学出版社) 第 18 页]
- [2] Antiqueira L, Luciano D, Costa F 2009 *New J. Phys.* **11** 013058
- [3] Shen-Orr S, Milo R, Mangan S, Alon U 2002 *Nat. Genet.* **31** 64
- [4] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U 2002 *Science* **298** 824
- [5] Barabasi A L, Oltvai Z N 2004 *Nature* **5** 101
- [6] Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U 2004 *Science* **303** 1538
- [7] Onnela J P, Saramki J, Kertsz J, Kaski K 2005 *Phys. Rev. E* **71** 065103(R)
- [8] Luciano D, Francisco A R, Travieso G, Villas P R B 2007 *Adv. Phys.* **56** 167
- [9] Krumov L 2010 *Local Structures Determine Performance within Complex Network* (Darmstadt: Suedwestdeutscher Verlag fuer Hochschulschriften) p54
- [10] Kotorowicz M, Kozitsky Y 2011 *Condens. Matter Phys.* **14** 13801
- [11] Squartini T, Garlaschelli D 2012 *Lect. Notes. Comput. Sci.* **7166** 24
- [12] Dong Z, Li X 2010 *Acta Phys. Sin.* **59** 1600 (in Chinese) [董昭, 李翔 2010 物理学报 **59** 1600]
- [13] Miao L L, Han C F, Liu L, Cao J M 2012 *Stud. Sci. Sci.* **30** 1468 (in Chinese) [缪莉莉, 韩传峰, 刘亮, 曹吉鸣 2012 科学学研究 **30** 1468]
- [14] Liu L, Xu W S, Han C F 2013 *J. Tongji Univ. (Nat. Sci)* **41** 53 (in Chinese) [刘亮, 许维胜, 韩传峰 2013 同济大学学报 (自然科学版) **41** 53]
- [15] Zhang L, Qian G Q, Zhang L 2010 *Syst. Eng-Theory Pract.* **30** 361 (in Chinese) [张林, 钱冠群, 张莉 2010 系统工程理论与实践 **30** 361]
- [16] Shi D H 2010 *Complex Syst. Complexity Sci.* **7** 16 (in Chinese) [史定华 2010 复杂系统与复杂性科学 **7** 16]
- [17] Hu J, Yang B R 2009 *Appl. Res. Comput.* **26** 858 (in Chinese) [胡健, 杨炳儒 2009 计算机应用研究 **26** 858]
- [18] Zachary W W 1977 *J. Anthropol. Res.* **33** 452
- [19] Lusseau D, Schneider K, Boisseau O J 2003 *Behav. Ecol. Sociobiol.* **54** 396
- [20] Gunes I, Bingol H 2006 arXiv: cs/0610129 [cs.MA]
- [21] Girvan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821



[22] Gleiser P, Danon L 2003 *Adv. Complex Syst.* **6** 565

Rev. E **70** 056122

[23] Ebel H, Mielsch L I, Bornholdt S 2002 *Phys. Rev. E* **66** 035103(R)

[26] Wernicke S, Rasche F 2006 *Bioinformatics* **22** 1152

[24] Newman M E J 2001 *Phys. Rev. E* **64** 016131

[25] Boguna M, Pastor-Satorras R, Diaz-Guilera A, Arenas A 2004 *Phys.*

[27] Egghe L, Leydesdorff L 2009 *J. Am. Soc. Inf. Sci. Tech.* **60** 1027

# The measurement of complex network based on motif\*

Han Hua   Liu Wan-Lu   Wu Ling-Yan<sup>†</sup>

(School of Science, Wuhan University of Technology, Wuhan 430070, China)

(Received 30 March 2012; revised manuscript received 28 April 2013)

## Abstract

According to the existence of motif in complex network topology structure, the motif-based node degree and edge degree are proposed to measure the importance of node and edge in the network on the basis of the traditional node degree and edge clustering coefficient. The Rand-ESU algorithm is used for motif detection of eight different scale networks, and the result demonstrates the existence of motif. The Rand-ESU algorithm is also used for analyzing the motif structures and characteristics in Karate network and Dolphin network. The Pearson correlation coefficient is used to measure the correlations of motif-based node degree and traditional node degree, motif-based edge degree and edge clustering coefficient. The results of simulation analysis show that the correlations are related to the motif species. The definitions of motif-based node degree and edge degree are the improvement and development of original definitions, and they comprehensively depict the importance of node and edge in the network.

**Keywords:** motif, node degree, edge degree, Pearson correlation coefficient

**PACS:** 89.75.-k, 89.75.Fb

**DOI:** 10.7498/aps.62.168904

---

\* Project supported by the National Natural Science Foundation of China (Grant No. 71140015) and the Fundamental Research Fund for the Central Universities of China (Grant Nos. 2013-Ia-007, 2013-Ia-040).

<sup>†</sup> Corresponding author. E-mail: 815007651@qq.com