一种基于最大流的网络结构熵*

蔡萌 1)2) 杜海峰 2)† 费尔德曼 2)3)

1) (西安交通大学管理学院, 西安 710049)

2) (西安交通大学公共管理与复杂性科学研究中心, 西安 710049)

3) (斯坦福大学莫里森人口与资源研究所, 斯坦福 94305)

(2013 年 10 月 8 日收到; 2013 年 11 月 20 日收到修改稿)

熵是可用来反映网络结构异质性的指标. 针对传统熵指标不能很好反映网络全局异构性的不足, 本文引入网络流的概念, 综合考虑径向测度和中间测度, 提出一种新的网络结构熵. 特殊网络 (如公用数据集 Dolphins 网络) 的分析结果表明, 本文提出的熵指标在一定程度上克服了其他网络熵指标的不足, 更能够反映网络的真实拓扑结构; 对随机网络、最近邻耦合网络、星型网络、无标度网络、Benchmark 网络和小世界网络等典型网络的理论分析和仿真实验, 进一步证明本文提出的熵指标在刻画一般复杂网络结构特征上的有效性和适用性.

关键词: 复杂网络, 异构性, 熵

PACS: 05.90.+m, 01.75.+m **DOI:** 10.7498/aps.63.060504

1 引 言

自然界的复杂系统大多可以通过复杂网络的方式来表达, 这些网络在其连接模式上都体现出高度的异构性. 随着复杂网络研究的兴起, 异构性得到了学者们的高度关注, 无论是信息扩散 [1,2]、病毒传播 [3,4] 还是灾害蔓延 [5,6], 都与网络的异构性不无关系 [7−9]. 熵是测度网络异构性的重要指标 [10]: 一般地, 网络结构越是均匀有序, 则熵值越大; 反之, 熵值越小. 然而采用何种形式的熵来反映网络结构特征, 一直是复杂网络研究关注的焦点.

目前网络熵的度量方法大致可以分为两类: 一类基于信息论, 主要是从信息搜索视角提出的目标熵、搜索信息熵、接受信息熵、隐藏信息熵、交换信息熵等 [11]; 另一类侧重反映网络结构特征, 根据网络中的 “连接” 分布, 提出的度分布熵 [10]、剩余度熵 [10] 和 Wu 结构熵 [12] 等.

度分布熵 [13] 以 “边” 为研究对象或基本元 (熵定义中按照度值加总), 根据拥有特定边数的节点数量的概率分布的不确定性反映网络的异构性——边无序性. 对于所有节点均仅与某一中心节点相连的星型网络, 在度分布熵的测度下显得结构相对均衡 (异构性很弱), 这与一般的看法并不相符 [12]. Wu 结构熵 [12] 以 “节点” 为基本元 (熵定义中按照节点加总), 根据节点拥有边的 (相对) 数量的概率分布的不确定性反映网络的异构性——点无序性. 同样, Wu 结构熵对于星型网络的解释存在问题, 主要体现在难以刻画星型网络的规模效应 1). 度分布熵和 Wu 结构熵是最典型的两种测度网络异构性的熵指标, 然而这两者均仅从 “点” 或 “边” 一方面考虑网络的结构特征, 导致了在刻画星型网络等特殊网络异构性的有效性和准确性方面存在一定的

* 国家社会科学基金重点项目 (批准号: 12AZD110)、国家自然科学基金 (批准号: 71071128)、国家教育部新世纪优秀人才支持计划 (批准号: NCET-08-0451) 和中央高校基本科研业务费专项资金 (批准号: 2011jdgz08) 资助的课题.

† 通讯作者. E-mail: haifengdu@mail.xjtu.edu.cn

1) 该指标认为星型网络是异构性最强的一种网络, 那么随着网络规模增大, 这种异构性应该增强. 事实上, 在该指标的测度下, 随着网络规模增大, 熵值也增大, 异构性在减弱.

问题[14].

熵反映的是网络整体的结构特征, 仅从"点"或"边"的局部测量虽然在一定程度上体现基本元对于网络的贡献, 但是容易造成熵值测度上的偏差, 难以对网络的结构特征进行深入系统的描述和诠释. 本文将网络流的概念引入结构熵的定义, 结合径向测度和中间测度方法, 提出一种新的网络结构熵. 一方面, 通过对特殊网络进行分析, 说明本文提出的熵指标的优越性, 可以在一定程度上克服其他网络熵指标的不足, 更能够反映网络的真实拓扑结构; 另一方面, 通过对Erdös-Rēnyi (ER)随机网络、最近邻耦合网络、星型网络、Barabási-Albert (BA) 无标度网络、Benchmark网络和Watts-Strogatz (WS)小世界网络等典型网络的理论分析和仿真实验, 证明本文提出的熵指标在刻画一般网络上的可用性和有效性.

2 基于网络流的结构熵定义

在最近的研究中, 文献[14]在充分考虑"点"差异性和"边"差异性两方面影响的基础上, 提出了点边差异性(SD)结构熵. 该网络结构熵是一种有效度量网络异构性的指标, 并对稀疏网络和星型网络有很好的解释, 在一定程度上克服了其他方法的缺陷. 然而, 该方法的本质仍是以节点度值为基础, 与度分布熵和Wu结构熵等其他指标一样, 过多强调网络的局部特征, 而忽略了网络的全局拓扑结构.

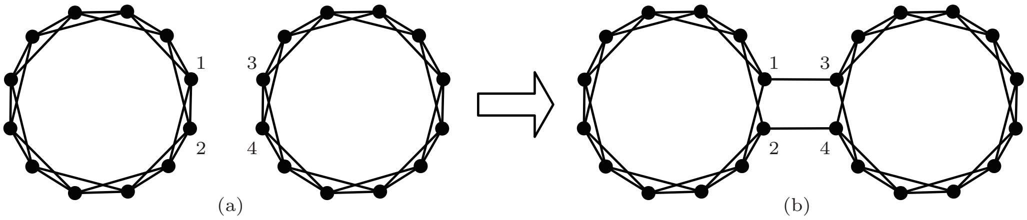

图1 仅改变连接方式的网络变换

图1(a)所示的网络为两个最近邻耦合网络的组合, 网络规模为24, 节点度值均为4. 图1(a)网络的变形网络如图1(b)所示, 即将原网络的节点对(1, 2), (3, 4)断开后重连为(1, 3), (2, 4). 两个网络的规模、密度、度分布和节点的度值完全相同, 惟一的区别在于节点间的连接方式. 显然, 图1(b)网络中节点1, 2, 3, 4相比其他节点的地位更为特殊, 尤其是在网络传播上更具优势. 然而在SD结构熵的测度下, 图1(b)与图1(a)有着相同的熵值, 是一个没有异构性的均衡网络.

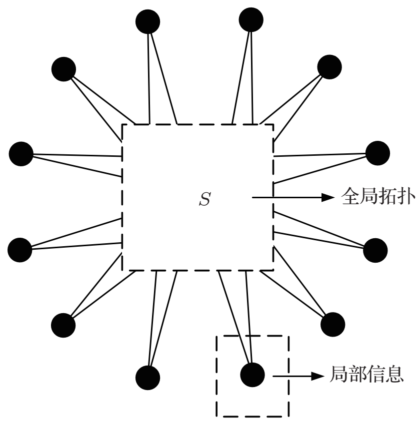

图2 网络熵对于结构的反映

图2所示为熵指标对于网络结构反映时使用的信息. 无论是度分布熵, Wu结构熵, 还是SD结构熵, 在对网络结构异质性的刻画上, 均使用基于节点的局部信息(度值和度分布), 对于整个网络连接模式的信息缺失使得全局拓扑结构成为一个"黑箱", 因此对网络的异构性测度也就不够准确.

网络流为网络全局拓扑结构的刻画提供了新的思路. 供水管道网络、通信数据传输网络、交通运输网络等, 都与网络的最大流有关. 从网络流的角度出发, 可以全面地反映网络连接的结构特征, 并突破网络权值和传播方式上的限制[15] 2). 网络的最大流反映了在限定边集容量的情况下, 两点之间输送流量最大的问题. 对于给定有向网络 $D=(V,\ E,\ C)$, V为节点集, E为边集, C为边的容量集. 对于任意边 $(v_i, v_j)\in E$, 满足 $c(v_i, v_j)\geqslant 0$ (简记为 $c_{ij}\geqslant 0$). 网络的流则是在边集 E 上的实函数 $f: V\times V\rightarrow R$, 定义为 $f=\{f(v_i, v_j)\}$, 其中

2) 以往研究中熵值的构造主要适用于0-1网络, 网络流概念的引入可使其推广至赋权网络; 网络流从全局角度考虑两点间任意可能的传播路径, 而不仅仅限于最短路径, 更具一般性.

$f(v_i, v_j)$ 为边 (v_i, v_j) 上的流量, 简记为 f_{ij}.

给定源点 s 和汇点 t, 最大流问题定义为[16]:

$$\begin{cases} \max v(f) \\ \text{s.t.} \quad 0 \leqslant f_{ij} \leqslant c_{ij} \quad \forall (v_i, v_j) \in E \\ \sum f_{ij} - \sum f_{ji} = \begin{cases} v(f) & i = s \\ 0 & i \neq s, t \\ -v(f) & i = t \end{cases}, \end{cases}$$

式中 $v(f)$ 为可行流 f 的流量, 即源点的净输出量或汇点的净输入量; (v_i, v_j) 表示有向 (或有序) 节点对之间的边, 对于无向网络则表示为 $\{v_i, v_j\}$. 若 $E = \{0, 1\}$, 则 $D = (V,\ E,\ C)$ 为无权网络. 网络中节点 k 的流介数 (flow betweenness) 定义为[15]

$$b_k = \frac{\sum_i^N \sum_j^N m_k(i,j)}{\sum_i^N \sum_j^N m(i,j)} \quad i \neq j \neq k, \qquad (1)$$

其中, N 为网络规模, $m(i,j)$ 为从节点 i 到节点 j 的最大流, $m_k(i,j)$ 为从节点 i 到节点 j 的最大流中通过节点 k 的流量.

由于任意两节点之间的边独立路径并不惟一, 因此流介数并不能通过直接计算路径得到. 在实际操作中, 常采用 (2) 式计算网络节点 k 的流介数:

$$b_k = \frac{\sum_{(i,j \in S(k))} ({}^k\boldsymbol{W}_{i,j} - {}^k\boldsymbol{W}^*_{i,j})}{\sum_{(i,j \in S(k))} {}^k\boldsymbol{W}_{i,j}}, \qquad (2)$$

$$S(k) = \{(i,j) : 1 \leqslant i \leqslant N; \\ 1 \leqslant j \leqslant N; i \neq j \neq k\},$$

其中, $\boldsymbol{W}$ 表示网络中所有节点对之间的最大流矩阵, ${}^k\boldsymbol{W}$ 表示从矩阵 $\boldsymbol{W}$ 中去掉第 k 行和第 k 列. ${}^k\boldsymbol{W}^*$ 表示从原网络 (矩阵) 中去掉第 k 行和第 k 列后, 重新计算的最大流矩阵. 令 b'_k 为网络节点 k 的绝对流介数, 测度当网络中某节点去除或中止传输后, 网络流量的变化:

$$b'_k = \sum_{(i,j \in S(k))} ({}^k\boldsymbol{W}_{i,j} - {}^k\boldsymbol{W}^*_{i,j}). \qquad (3)$$

定义 d'_k 为无向网络中节点 k 的度值, d_k^{in} 和 d_k^{out} 分别表示有向网络中节点 k 的入度值和出度值. a_{ij} 表示网络 D 的邻接矩阵中第 i 行, 第 j 列的元素值.

$$d_k^{\text{in}} = \sum\nolimits_i a_{ik}, \quad d_k^{\text{out}} = \sum\nolimits_j a_{kj}. \qquad (4)$$

b'_k 是典型的中间测度方法 (medial measure), 反映了网络最大流对于节点 k 的依赖性, 测度经过给定节点的路径数量; d'_k 属于径向 (放射) 测度方法 (radial measure), 是一种在网络路径中将给定节点作为起点或终点的中心性测度方法, 本质在于反映给定节点与网络中其他节点间的联系. b'_k 和 d'_k 完整构成了网络游走位置 (walk position) 分类的两部分[17].

径向测度和中间测度反映了节点在网中所扮演的不同角色, 两者共同决定了节点在网络中的参与度. 两者互补构成了给定节点对于网络的总贡献: 径向测度评估团体的成员身份, 中间测度评估桥接能力. 两种测度方式虽然是从自然科学的角度定义, 但在社会科学中也有着明确的含义: 径向测度和中间测度分别反映了内聚社会资本 (bonding social capital) 和外联社会资本 (bridging social capital), 或者说封闭个体网络和开放个体网络[17]. 内聚社会资本从内部联系、团结、信任的观点出发, 认为社会资本存在于组织内部个人之间、小团体之间的联系, 实际中使用径向测度反映其同质群体的联系; 外联社会资本从外部联系、信息不对称、权力利益的观点出发, 认为社会资本的价值源于参与和控制信息的扩散, 实际中使用中间测度反映其异质群体的联系[18]. 内聚社会资本和外联社会资本是个体社会资本的两个方面, 反映了个体社会资本的总量, 应统一考虑而不宜割裂.

从径向测度和中间测度两方面综合考虑, 定义节点 k 在网络中的结构重要性为

$$I'_k = \alpha b'_k + \beta d'_k = \alpha b'_k + \beta(d_k^{\text{in}} + d_k^{\text{out}}), \qquad (5)$$

其中,

$$\alpha + \beta = 1, 0 \leqslant \alpha, \beta \leqslant 1,$$

α 和 β 分别为中间测度和径向测度的权重, $\alpha = 1$, $\beta = 0$ 表示仅考虑中间测度下节点对于网络的贡献, 而不考虑径向测度; $\alpha = 0$, $\beta = 1$ 则表示仅考虑径向测度下节点对于网络的贡献, 此时 I'_k 退化为局部测量, 最终构造的熵指标也等价于 Wu 结构熵. 径向测度主要衡量节点在网络传播中的收发能力, 类似于雷达式辐射的方式, 反映节点在网络中的局部特征, 体现节点在网络中的绝对权力, 是从节点本身属性出发的一种计算方法; 中间测度主要衡量节点在网络传播中的桥接 (中转、中继) 能力, 反映整个网络在全局上的结构特征, 体现网络在传播过程中对节点的依赖性, 是从网络边 (连接) 特征出发的一种计算方法.

因此, 从径向测度和中间测度两方面同时考虑, 可定义节点 k 在网络结构中的相对重要性为

$$
\begin{aligned}
I_k &= \frac{I'_k}{\sum_{n=1}^{N} I'_n} = \frac{\alpha b'_k + \beta d'_k}{\sum_{n=1}^{N}(\alpha b'_n + \beta d'_n)} \\
&= \frac{\alpha \sum_{(i,j\in S(k))}({}^k\boldsymbol{W}_{i,j} - {}^k\boldsymbol{W}^*_{i,j}) + \beta(d_k^{\mathrm{in}} + d_k^{\mathrm{out}})}{\sum_{n=1}^{N}[\alpha \sum_{(i,j\in S(n))}({}^n\boldsymbol{W}_{i,j} - {}^n\boldsymbol{W}^*_{i,j}) + \beta(d_n^{\mathrm{in}} + d_n^{\mathrm{out}})]} \\
&= \frac{e_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \dfrac{\beta}{\alpha}(\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\delta_k + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\delta_k)}{\sum_{n=1}^{N}(e_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1}) + \dfrac{2\beta}{\alpha}\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e}_N},
\end{aligned}
\tag{6}
$$

其中, $\boldsymbol{e}_N$ 为元素为1的 N 维列向量, $\boldsymbol{e}_N^{\mathrm{T}}$ 为 $\boldsymbol{e}_N$ 的转置; $\boldsymbol{A}$ 为网络 D 的邻接矩阵, $\boldsymbol{A}^{\mathrm{T}}$ 为 $\boldsymbol{A}$ 的转置; $\boldsymbol{\delta}_k$ 为 N 维列向量, 其中第 k 位元素为1, 其他为0.

(6) 式中 β/α 为常量, 本文认为径向测度和中间测度对于网络中节点重要性的判断同样重要, 即 $\beta/\alpha = 1$, 故:

$$
I_k \approx \frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\delta}_k + \varDelta}{\sum_{n=1}^{N}(e_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \varDelta) + 2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{e}_N}.
\tag{7}
$$

为避免 $V \neq \phi$ 且 $E = \phi$ 时公式计算无意义, 特引入 $\varDelta$ 项, 满足 $\varDelta \sim O\left(\frac{1}{N^2}\right)$, $N > 1$. 该项的引入并不对网络结构异构性的分析结果产生影响.

基于 (7) 式, 本文的网络结构熵定义为

$$
\begin{aligned}
H =& -\sum_{k=1}^{N} I_k \log I_k \\
=& -\sum_{k=1}^{N} \frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\delta}_k + \varDelta}{\sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \varDelta) + 2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{e}_N} \\
&\times \log \frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k + \boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\delta}_k + \varDelta}{\sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \varDelta) + 2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{e}_N},
\end{aligned}
\tag{8}
$$

(8) 式适用于有向/无向, 赋权/0-1 网络, 对于无向网络, (8) 式可简化为

$$
\begin{aligned}
H =& -\sum_{k=1}^{N} I_k \log I_k \\
=& -\sum_{k=1}^{N} \frac{\sum_{(i,j\in S(k))}({}^k\boldsymbol{W}_{i,j} - {}^k\boldsymbol{W}^*_{i,j}) + d_k + \varDelta}{\sum_{n=1}^{N}[\sum_{(i,j\in S(n))}({}^n\boldsymbol{W}_{i,j} - {}^n\boldsymbol{W}^*_{i,j} + \varDelta) + d_n]} \\
&\times \log \frac{\sum_{(i,j\in S(k))}({}^k\boldsymbol{W}_{i,j} - {}^k\boldsymbol{W}^*_{i,j}) + d_k + \varDelta}{\sum_{n=1}^{N}[\sum_{(i,j\in S(n))}({}^n\boldsymbol{W}_{i,j} - {}^n\boldsymbol{W}^*_{i,j} + \varDelta) + d_n]},
\end{aligned}
\tag{9}
$$

其中,

$$
\begin{aligned}
S(k) = \{(i,j) : 1 \leqslant i \leqslant N; \\
1 \leqslant j < i; \quad i \neq j \neq k\},
\end{aligned}
$$

d_k 为节点 k 的度值.

(8) 式可推广至动态网络研究, 即给定任意时刻 t, 网络的结构熵可表示为

$$
\begin{aligned}
H(t) =& -\sum_{k=1}^{N(t)} I_{k,t} \log I_{k,t} \\
=& -\sum_{k=1}^{N(t)} \frac{\boldsymbol{e}_{N(t)-1}^{\mathrm{T}}\,{}_t^k\boldsymbol{W}\boldsymbol{e}_{N(t)-1} - \boldsymbol{e}_{N(t)-1}\,{}_t^k\boldsymbol{W}^*\boldsymbol{e}_{N(t)-1} + \boldsymbol{e}_{N(t)}^{\mathrm{T}}\boldsymbol{A}(t)\boldsymbol{\delta}_k + \boldsymbol{e}_{N(t)}^{\mathrm{T}}\boldsymbol{A}(t)^{\mathrm{T}}\boldsymbol{\delta}_k + \varDelta}{\sum_{n=1}^{N(t)}(\boldsymbol{e}_{N(t)-1}^{\mathrm{T}}\,{}_t^n\boldsymbol{W}\boldsymbol{e}_{N(t)-1} - \boldsymbol{e}_{N(t)-1}\,{}_t^n\boldsymbol{W}^*\boldsymbol{e}_{N(t)-1} + \varDelta) + 2\boldsymbol{e}_{N(t)}^{\mathrm{T}}\boldsymbol{A}(t)\boldsymbol{e}_{N(t)}} \\
&\times \log \frac{\boldsymbol{e}_{N(t)-1}^{\mathrm{T}}\,{}_t^k\boldsymbol{W}\boldsymbol{e}_{N(t)-1} - \boldsymbol{e}_{N(t)-1}\,{}_t^k\boldsymbol{W}^*\boldsymbol{e}_{N(t)-1} + \boldsymbol{e}_{N(t)}^{\mathrm{T}}\boldsymbol{A}(t)\boldsymbol{\delta}_k^t + \boldsymbol{e}_{N(t)}^{\mathrm{T}}\boldsymbol{A}(t)^{\mathrm{T}}\boldsymbol{\delta}_k^t + \varDelta}{\sum_{n=1}^{N}(\boldsymbol{e}_{N(t)-1}^{\mathrm{T}}\,{}_t^n\boldsymbol{W}\boldsymbol{e}_{N(t)-1} - \boldsymbol{e}_{N(t)-1}\,{}_t^n\boldsymbol{W}^*\boldsymbol{e}_{N(t)-1} + \varDelta) + 2\boldsymbol{e}_{N(t)}^{\mathrm{T}}\boldsymbol{A}(t)\boldsymbol{e}_{N(t)}},
\end{aligned}
\tag{10}
$$

其中, $N(t)$ 为 t 时刻的网络规模; $\boldsymbol{A}(t)$ 为 t 时刻网络D的邻接矩阵; $\boldsymbol{\delta}_k^t$ 为 $N(t)$ 维列向量, 其中第 k 位元素为1, 其他为0; ${}_t^k\boldsymbol{W}$ 表示 t 时刻从最大流矩阵 $\boldsymbol{W}$ 中去掉第 k 行和第 k 列; ${}_t^n\boldsymbol{W}^*$ 表示当节点 k 不存在时, t 时刻网络的最大流矩阵. 对于网络规模的演变, 网络结构熵反映了系统能量的分布异质性, 也表明了新节点添加前的平均不确定性.

本文提出的网络结构熵(flow betweenness structure entropy, FB结构熵)能够从全局角度判断网络的异构性, 更为细致真实地反映网络连接在拓扑结构上的差异. 对于图1所示的网络, 表1给出度分布熵、Wu结构熵、SD结构熵和FB结构熵的计算结果(公式中均使用自然对数). 结果表明, 图1(a)所示网络重连为图1(b)后, 网络异构性增大. 但在度分布熵、Wu结构熵和SD结构熵的测度下, 网络的熵值并没有发生改变, 三种结构熵判别失效. 在FB结构熵的测度下, 重连后的网络熵值出现了明显的下降, 表明网络异构性增大, 准确反映了网络拓扑结构上的差异.

表1 四种结构熵在图1网络上的差异

结构熵	图1(a)	图1(b)
度分布熵	0	0
Wu结构熵	3.1781	3.1781
SD结构熵	3.1781	3.1781
FB结构熵	3.1781	2.882

需要说明的是, FB结构熵具有较强的推广性, (8)式是具有一般意义的熵值测度表达. 例如考虑Freeman介数[19]作为中间测度方式时(虽然执行效率更高, 但无法应用于赋权网络), 只需将最大流矩阵 $\boldsymbol{W}$ 换为最短路径矩阵, 其中 w_{ij} 表示节点 i 到节点 j 之间的最短路径的数量.

3 典型网络结构熵

随机网络、规则网络以及介于这两类之间的网络是三种基本的网络模型[20]. 其中, 随机网络指按照统计规律生成随机图及派生模型; 规则网络指具有规则结构的网络, 包括最近邻耦合网络、全连接网络和星型网络等; 第三类网络兼具前两类网络模型的部分特征, 最具代表性的网络有无标度网络、小世界网络和Benchmark网络等.

本文针对ER随机网络、最近邻耦合网络、星型网络、BA无标度网络、Benchmark网络和WS小世界网络等6种典型网络的结构熵进行讨论.

3.1 ER随机网络(random network)结构熵

ER随机网络[21]是网络研究的主要参考模型, 由Erdös和Rēnyi于1960年提出. 给定网络规模 N, 以概率 $q=\dfrac{2n}{N(N-1)}$ 每次连接任意两个节点(不重复连接)直到网络总边数达到指定数 n, 最终形成的网络即为ER随机网络. 网络的平均度 $\langle k\rangle=q(N-1)$, 度分布服从Poisson分布: $P(k)\approx \mathrm{e}^{-\langle k\rangle}\langle k\rangle^k/k!$, FB结构熵为

$$
\begin{aligned}
H= & -\sum_{k=1}^{N}\frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1}-\boldsymbol{e}_{N-1}{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1}+2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k+\varDelta}{2N\langle k\rangle+\sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1}-\boldsymbol{e}_{N-1}{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1}+\varDelta)} \\
& \times\log\frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1}-\boldsymbol{e}_{N-1}{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1}+2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k+\varDelta}{2N\langle k\rangle+\sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1}-\boldsymbol{e}_{N-1}{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1}+\varDelta)}.
\end{aligned}
\tag{11}
$$

3.2 规则网络结构熵

3.2.1 最近邻耦合网络(nearest-neighbor coupled network)

最近邻耦合网络[20]是规则网络的典型代表, 网络中所有节点只与其周围的 k 个邻居节点相连. 给定网络规模 N 和度值 k, 将所有节点围成一个环并分别连接其两边的 $k/2$ 个邻居节点, 则可构成最近邻耦合网络. 本文提出的网络结构熵, 满足 $H\leqslant\log N$, 当且仅当 $\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1}-\boldsymbol{e}_{N-1}{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1}+\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k+\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\delta}_k$ 对于任意 k, $k\in\{1,2,\cdots N\}$ 相等时取等号. 给定网络规模 N, 最近邻耦合网络中的所有节点在结构上都是对等的, 因此具有最大FB结构熵:

$$
H_{\max}=-\sum\nolimits_{i=1}^{N}\frac{1}{N}\log\frac{1}{N}=\log N. \tag{12}
$$

3.2.2 星型网络 (star network)

星型网络[20]指网络中所有节点均与某一中心节点相连, 且这些节点之间并不相连. 即对于固定的网络规模N, 记中心节点的度值为d_1, 则$d_1 = N-1$, $d_i = 1 (i \neq 1)$. 虽然是从完全不同的角度出发, FB结构熵与文献[14]中的SD结构熵均认为网络熵的最小值出现在星型网络, 且得到的解析表达相同:

$$
\begin{aligned}
H_{\min} = & -\Bigg[\frac{N}{N+2}\log\frac{N}{N+2} + (N-1) \\
& \times \frac{2}{(N-1)(N+2)}\log\frac{2}{(N-1)(N+2)}\Bigg] \\
= & \log(N+2) + \frac{2}{N+2}\log\frac{N-1}{2} \\
& - \frac{N}{N+2}\log N. \qquad (13)
\end{aligned}
$$

3.3 BA无标度网络(scale-free network)结构熵

BA无标度网络[22]是由Barabasi和Albert在1999年提出的, 网络的初始节点数为m_0, 增长率为m, 在经过时间t后, 所形成的网络是一个具有$N = t + m_0$个节点, mt条边的无标度网络, 度分布满足$p(k) \sim 2m^2k^{-3}$. BA无标度网络中的大部分节点拥有少量连接, 而少数节点拥有大量连接, 其幂率分布特征相比ER随机网路的Poisson分布来说体现出结构的“不均匀”性. t时刻的BA无标度网络演化FB结构熵为

$$
\begin{aligned}
H(t) = & -\sum_{k=1}^{m_0+t} \frac{\boldsymbol{e}_{m_0+t-1}^{\mathrm{T}}\,{}_t^k\boldsymbol{W}\boldsymbol{e}_{m_0+t-1} - \boldsymbol{e}_{m_0+t-1}\,{}_t^k\boldsymbol{W}^*\boldsymbol{e}_{m_0+t-1} + 2\boldsymbol{e}_{m_0+t}^{\mathrm{T}}\boldsymbol{A}(t)\boldsymbol{\delta}_k + \varDelta}{4mt + \sum_{n=1}^{m_0+t}(\boldsymbol{e}_{m_0+t-1}^{\mathrm{T}}\,{}_t^n\boldsymbol{W}\boldsymbol{e}_{m_0+t-1} - \boldsymbol{e}_{m_0+t-1}\,{}_t^n\boldsymbol{W}^*\boldsymbol{e}_{m_0+t-1} + \varDelta)} \\
& \times \log\frac{\boldsymbol{e}_{m_0+t-1}^{\mathrm{T}}\,{}_t^k\boldsymbol{W}\boldsymbol{e}_{m_0+t-1} - \boldsymbol{e}_{m_0+t-1}\,{}_t^k\boldsymbol{W}^*\boldsymbol{e}_{m_0+t-1} + 2\boldsymbol{e}_{m_0+t}^{\mathrm{T}}\boldsymbol{A}(t)\boldsymbol{\delta}_k + \varDelta}{4mt + \sum_{n=1}^{m_0+t}(\boldsymbol{e}_{m_0+t-1}^{\mathrm{T}}\,{}_t^n\boldsymbol{W}\boldsymbol{e}_{m_0+t-1} - \boldsymbol{e}_{m_0+t-1}\,{}_t^n\boldsymbol{W}^*\boldsymbol{e}_{m_0+t-1} + \varDelta)}. \qquad (14)
\end{aligned}
$$

3.4 Girvan-Newman (GN) 基准网络(GN benchmark network)结构熵

GN基准网络[23]由Girvan和Newman于2002年提出, 是著名的用于研究社群结构探测的网络类型系列[24]. 典型的GN基准网络可构造如下: 生成4个社群, 分别包含32个节点(共128个节点), 网络的平均度值为12, 且每个节点含有大致相同的度值; 根据所处社群不同, 节点之间的连接具有倾向性, 且满足随机性和独立性, 即以概率P_{in}连接同一社群的节点, 而以P_{out}连接不同社群的节点. 可通过变量μ分配节点社群内连接与社群外连接:

$$
k_{\text{in}} = \mu k_i, \quad k_{\text{out}} = (1-\mu)k_i, \qquad (15)
$$

其中, k_i为节点i的度值, k_{in}为节点i与其社群内其他节点的连接数量, k_{out}为节点i与社群外节点的连接数量. 当$k_{\text{out}} < k_{\text{in}}$时则意味着节点更倾向于连接同社群的节点, 网络具有社群结构, 且其他方面则体现随机特征, 其FB结构熵为

$$
\begin{aligned}
H = & -\sum_{k=1}^{N} \frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + 2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k + \varDelta}{3072 + \sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \varDelta)} \\
& \times \log\frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + 2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k + \varDelta}{3072 + \sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \varDelta)}. \qquad (16)
\end{aligned}
$$

3.5 WS小世界网络(small-world network)结构熵

WS小世界网络[25]是由Watts和Strogatz于1998年提出, 其生成规则如下: 1) 按照3.2.1节中的方式构造规模为N, 度值为k的最近邻耦合网络; 2) 以概率p随机重连网络中的边, 即对于任意一条边, 保持其一个端点不变而将另一端点随机选择新的节点连接, 且排除自环以及重复连线. 该模型中, $p = 0$对应最近邻耦合网络, $p = 1$对应ER随机网络, $0 < p < 1$时的WS小世界网络是规则网络和随机网络之间的过渡网络, 兼具两者的结构特性, 是一种半结构化、半随机的网络. FB结构熵为

$$
H = -\sum_{k=1}^{N} \frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^k\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^k\boldsymbol{W}^*\boldsymbol{e}_{N-1} + 2\boldsymbol{e}_N^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_k + \varDelta}{2Nk + \sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}\,{}^n\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}\,{}^n\boldsymbol{W}^*\boldsymbol{e}_{N-1} + \varDelta)}
$$

$$\times \log \frac{\boldsymbol{e}_{N-1}^{\mathrm{T}}{}^{k}\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}{}^{k}\boldsymbol{W}^{*}\boldsymbol{e}_{N-1} + 2\boldsymbol{e}_{N}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\delta}_{k} + \Delta}{2Nk + \sum_{n=1}^{N}(\boldsymbol{e}_{N-1}^{\mathrm{T}}{}^{n}\boldsymbol{W}\boldsymbol{e}_{N-1} - \boldsymbol{e}_{N-1}{}^{n}\boldsymbol{W}^{*}\boldsymbol{e}_{N-1} + \Delta)}. \quad (17)$$

WS小世界网络的SD结构熵为

$$H_{\mathrm{SD}} = -\sum_{i=1}^{N} \frac{(k_i+1)[1-p(k_i)+\Delta]}{\sum_{j=1}^{N}\{(k_j+1)[1-p(k_j)]+\Delta\}} \times \log \frac{(k_i+1)[1-p(k_i)+\Delta]}{\sum_{j=1}^{N}\{(k_j+1)[1-p(k_j)+\Delta]\}}, \quad (18)$$

其中, k_i为第i个节点的度值, 当$d < k/2$时, $p(d)=0$; 当$d \geqslant k/2$时,

$$p(d) = \sum_{n=0}^{\min(d-k/2,k/2)} \binom{k/2}{n} (1-p)^n p^{(k/2)-n} \times \frac{(pk/2)^{d-(k/2)-n}}{(d-(k/2)-n)!}\mathrm{e}^{-pk/2}. \quad (19)$$

3.6 典型网络的结构熵仿真实验

分别对ER随机网络、最近邻耦合网络、星型网络、BA无标度网络、Benchmark网络和WS小世界网络等6种典型网络的结构熵进行数值模拟, 并以公用数据集Dolphins为例进行验证, 说明本文所提出网络结构熵的有效性和优越性. 对于ER随机网络、最近邻耦合网络、星型网络和BA无标度网络, 由于文献[14]已验证SD结构熵相比其他网络熵更为有效, 因此仅将本文提出的结构熵与SD结构熵进行对比; 对于Benchmark网络和WS小世界网络, 则额外加入另一经典网络熵——Wu结构熵进行对比, 同时也是对SD结构熵研究的一个补充; 以Dolphins网络为例进行分析时, 则同时对比FB结构熵、SD结构熵和Wu结构熵对于网络中节点重要性的评估.

实验均是在Intel$^{\mathrm{R}}$ Core$^{\mathrm{TM}}$ 2, 2.40 GHz的PC上采用Matlab7.0编程实现, 其中每个指标独立进行20次计算后取平均值. 为便于和其他典型的网络熵进行对比, 实验所选用的典型网络均为无向无权网络; 为便于不同网络熵之间的比较, 实验中所有熵的数值计算均采用自然对数.

图3所示为四种典型网络的熵值. BA无标度网络所示曲线表示按照文献[22]所述过程演化的结构熵, 网络增长率$m=10$, 最终的网络演化规模为200. 最近邻耦合网络和星型网络曲线为不同规模下的最近邻耦合网络和星型网络的熵. ER随机网络则按照演化过程中BA无标度网络的规模和密度构造.

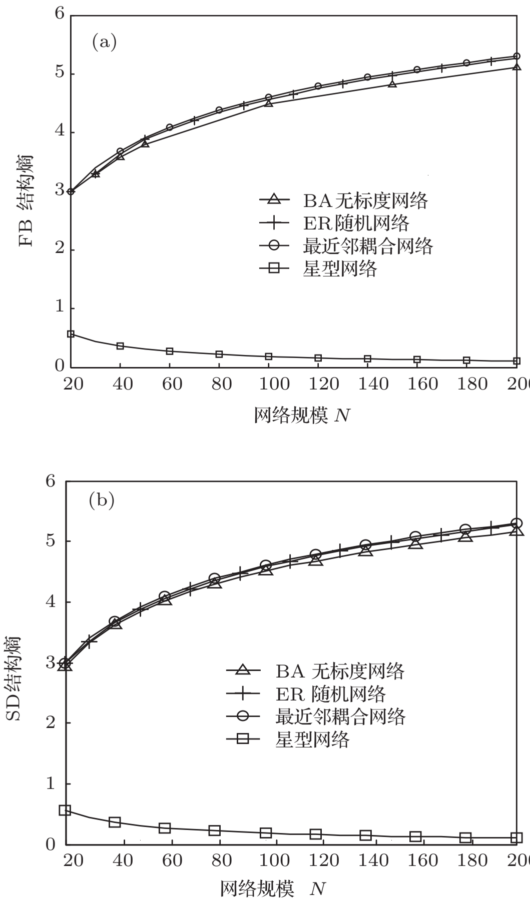

图3 四种典型网络结构熵 (a) 四种典型网络的FB结构熵; (b) 四种典型网络的SD结构熵

对比图3(a)和(b), 可以发现FB结构熵和SD结构熵对于四种典型网络的结构熵刻画上基本一致. 两种结构熵均认为网络熵值的边界(最大值与最小值)分别对应最近邻耦合网络和星型网络, 3.2节中的分析也表明两者具有完全相同的解析表达(即拥有相同的值域). 对于星型网络的准确刻画, 一直是SD结构熵优越性的体现, 而本文提出的FB结构熵也很好地解决了这一问题. ER随机网络的结构熵几乎等于(略低于)同规模下的最近邻耦合网络结构熵. 对比BA无标度网络和ER随机网络, BA无标度网络大量节点拥有少量连接, 而少数节点拥有大量连接, 其异构性显然应强于ER随机网络. 因此, 图3(a)和(b)的相应曲线说明, 相比SD结构熵, FB结构熵能够更好地区分BA无标度网

络和ER随机网络在异构性上的差异.

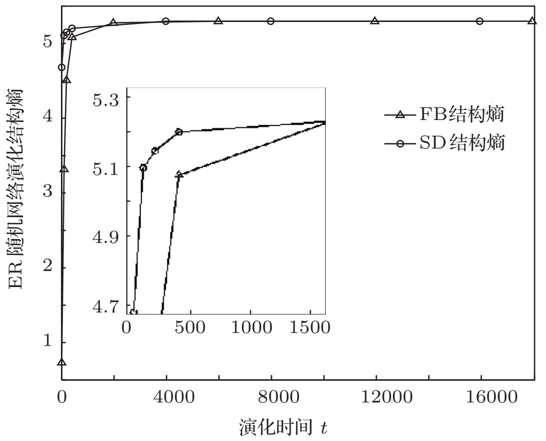

图4 ER随机网络演化结构熵

图4为按照文献[21]演化的ER随机网络结构熵(也反映了不同密度下ER随机网络的结构熵),网络规模为200,最终演化密度为0.9. 与BA无标度网络类似,ER随机网络的演化过程在两种网络熵指标下均体现出明显的熵增效应. 虽然在两种结构熵指标的测度下,网络均在较短时间内演化为稳定状态,然而本文提出的FB结构熵进入稳态的时间更长. 这是由于FB结构熵相比SD结构熵,对于网络结构均衡的要求更严格. SD结构熵仅从网络局部特征出发,以节点度值和度分布作为网络异构性大小的判断依据,而无法识别网络拓扑连接上的异构性; FB结构熵考虑网络全局特征,其测度反映了随着ER随机网络的演化,首先实现网络局部的均衡,进而逐步进入全局均衡(即在演化初期,当不同节点的度值大致相当时,网络在连接上的异构性仍然很强; 随着网络中边的增加,网络拓扑连接的异构性才继续减弱).

图5所示为GN基准网络随社群外连度$k_{\rm out}$变化时的三种结构熵,网络规模为128,社群数为4,网络平均度为12,$k_{\rm out} = 0, 1, \cdots, 12$. 对于$k_{\rm out} = 0$,网络相当于4个规模为32的随机网络的组合,在三种结构熵的测度下异构性均较低. 随着$k_{\rm out}$的增加,不同社群间的连接增多,网络异构性增强; 而当$k_{\rm out} > 9$时,熵值上升,异构性减弱. 一般地,给定网络规模N,节点平均度k,将节点均分为M个社群,构造GN基准网络,当满足$\dfrac{k_{\rm out}}{k} = \dfrac{M-1}{M}$时,GN基准网络等价于同规模同密度的ER随机网络. 图5所示$k_{\rm out} = 9$时出现的拐点表明ER随机网络相比具有明显社群特征的同规模同密度GN基准网络,体现出更强的异构性(虽然从数值比较上来看,这种差异并不十分显著). 与ER随机网络的演化特征相同,GN基准网络的FB结构熵比其他网络熵的熵值更低,这同样是因为FB结构熵从网络全局拓扑结构出发,评判结构均衡的要求更高.

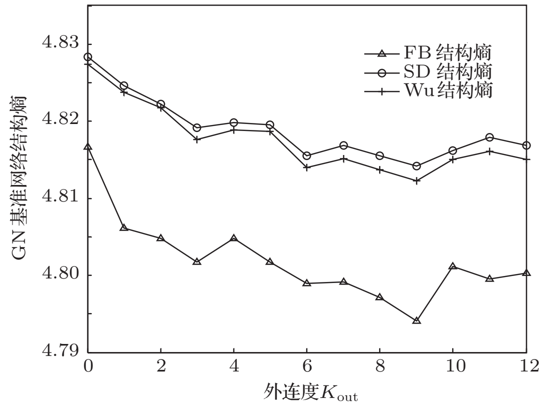

图5 GN基准网络结构熵

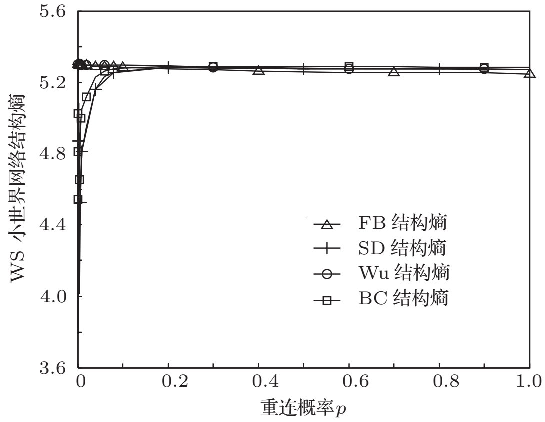

图6 WS小世界网络结构熵

图6所示为WS小世界网络随重连概率p变化的四种结构熵,其中网络规模为200,平均度为10. 该过程同样可以理解为按照文献[25]构造的WS小世界网络演化,反映最近邻耦合网络向随机网络的变化过程. FB结构熵和Wu结构熵均认为在演化过程中,网络的异构性变化不大; 而在SD结构熵的测度下,网络的异构性出现了较大变化. 对于$p=0$时的最近邻耦合网络和$p=1$时的随机网络,其结构异构性都是较低的. 当$p \in (0,\ 0.1]$时,SD结构熵测度体现出了明显的V字型,这也在一定程度上验证了WS小世界网络的小世界特征主要体现在重连概率为(0, 0.1]的区间内[25]. 小世界特征主要体现为高聚类系数和短路径长度,而根据文献[15]的理解,网络中的流并不总按照最短路径传播. 因

此, 从该角度出发, FB结构熵认为小世界特征并不会对网络的异构性带来太大改变. 当需要反映该特点时, 可如本文第2节所述, 将(8)式中的最大流矩阵 $\boldsymbol{W}$ 换为最短路径矩阵即可, 即图6中的BC结构熵.

为进一步验证本文提出的网络结构熵在评估网络异构性能上的优越性, 以公用数据集Dolphins为例进行说明[26]. 该数据集描述了62只宽吻海豚彼此之间结交的情况, 网络中的每个节点代表一只海豚. FB结构熵、SD结构熵和Wu结构熵对于网络异构性的判定均基于节点重要性的评估, 结果如下:

FB结构熵为3.728, 节点重要性排序为(升序)

5 12 13 23 32 36 49 59 61 42 56 24 53 22 17 25 54 4 57 33 51 27 19 26 6 45 47 50 60 62 1 7 9 11 28 3 48 35 40 10 16 43 39 14 34 38 21 46 30 20 29 15 55 44 41 52 58 18 8 31 37 2;

SD结构熵为3.9959, 节点重要性排序为(升序)

5 12 13 23 32 36 49 59 61 40 47 50 54 56 57 4 24 26 27 33 62 3 6 20 45 53 8 11 28 29 31 35 42 60 1 7 9 17 22 25 43 48 10 16 19 37 44 51 55 2 14 39 41 18 21 30 58 34 52 38 46 15;

Wu结构熵为3.9513, 节点重要性排序为(升序)

5 12 13 23 32 36 49 59 61 40 47 50 54 56 57 4 24 26 27 33 62 3 6 20 45 53 8 11 28 29 31 35 42 60 1 7 9 17 22 25 43 48 10 16 19 37 44 51 55 2 14 39 41 18 21 30 58 34 52 38 46 15.

事实上, Wu结构熵对于节点重要性的排序本质上等同于度序列, 而SD结构熵在排序上则与Wu结构熵相同. 这是由于SD结构熵在熵值计算上对于不同的度值进行了权值处理, 以反映度值差异带来的网络异构性, 而具有相同度值的节点之间的差异(异构性)则无法体现. 这两种网络结构熵均是以径向测度为基础的, 而在此基础上还综合考虑了中间测度的FB结构熵, 则给出了不同的排序结果. 图7是利用GN算法[23]得到的Dolphins网络社群结构. 可以发现, FB结构熵测度下的重要节点往往也是连接网络中不同社群的桥梁节点(如节点2, 37, 31, 8), 这些节点不但具有相对较大的度值, 同时也是影响范围较大的信息流控制节点[27]. 以节点8和节点11为例进行对比, 虽然两者的度值相同, 然而节点11仅仅为网络中的局部中心, 而节点8则表现为划分两个社群的关键节点, FB结构熵相比其他两种结构熵能够更好地体现这种差异. 因此, 引入了最大流的FB结构熵在综合考虑径向测度和中间测度的基础上, 相比SD结构熵和Wu结构熵对于网络拓扑结构异质性的刻画更加细致, 不再局限于局部异构而同时考虑了全局异构性.

上述典型网络结构仿真实验表明, FB结构熵可以有效地反映网络的异构性. 本文提出的FB结构熵在ER随机网络、最近邻耦合网络、星型网络、BA无标度网络、Benchmark网络和WS小世界网

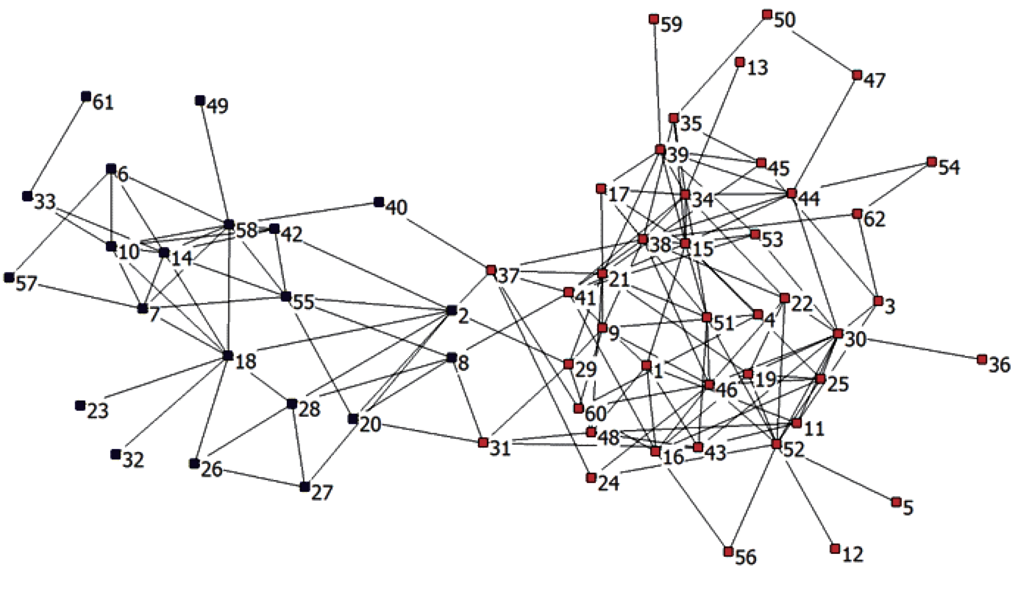

图7 Dolphins网络社群结构

络等6种典型网络的刻画, 与SD结构熵都能较为准确地反映其异构性, 且计算结果近似; 而对于如图1所示的特殊网络和公用数据集Dolphins网络, 由于本文提出的网络结构熵考虑了网络的全局拓扑特性, 在一定程度上克服了SD结构熵的不足, 拥有更好的解释性.

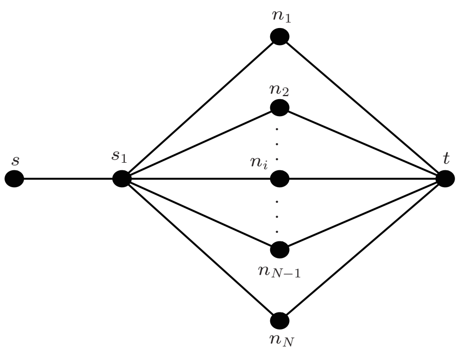

图8 网络流的传播依赖

4 结论与展望

本文引入网络流的概念, 从径向测度和中间测度两方面综合考虑, 提出一种新的网络结构熵定义. 该结构熵克服了SD结构熵(以及Wu结构熵和度分布熵)在反映网络连接模式上的不足, 能够从结构全局上刻画网络的异构性; 通过对ER随机网络、最近邻耦合网络、星型网络、BA无标度网络、Benchmark网络和WS小世界网络等6种典型网络的结构熵分析及对比, 说明本文提出的结构熵的适用性, 并得到了BA无标度网络、ER随机网络和WS小世界网络结构熵演变规律. 本文提出的网络结构熵可以刻画不同网络结构的异构性, 丰富了网络结构指标体系, 并有利于揭示网络的结构特性.

本文的主要工作集中在网络结构熵测度的设计和其特性的实验分析, 更为深入的理论分析是后续研究的主要内容之一, 本文的后续研究还将关注:

一方面, 通过对非惟一最大流依赖性的进一步刻画, 反映网络的异构性. 如图8所示, 给定源点s和汇点t, 当节点s和s_1的容量c_{s,s_1}满足$c_{s,s_1} \leqslant \min(\min(c_{s_1,n_i}, c_{n_i,t}))$, $i = 1, 2, \cdots, N$时, 路径(s, s_1)成为s到t最大流的瓶颈, 而该流对于路径集合$\{(s_1, n_i, t)\}, \forall i \in \{1, 2, \cdots, N\}$有完全的依赖, 却对其中任意一条路径不存在依赖性, 这是因为该集合中的每一条路径均为一条可能的最大流, 随着N的变化, 网络的拓扑结构发生改变, 与其他结构熵一样, FB结构熵也无法很好地体现;

另一方面, 在FB结构熵关注节点特征的基础上, 将网络边的影响一并考虑, 构造新的全局结构熵指标, FB结构熵突出了节点在网络中的重要程度, 然而在中间测度中节点的传播能力并不依赖于其全部的边, 如何进一步考虑网络中冗余边对其结构影响的变化, 是构造新的结构熵指标值得探讨的内容.

参考文献

[1] Hu Q C, Yin Y S, Ma P F, Gao Y, Zhang Y, Xing C X 2013 *Acta Phys. Sin.* **62** 140101 (in Chinese)[胡庆成, 尹龑燊, 马鹏斐, 高旸, 张勇, 邢春晓 2013 物理学报 **62** 140101]

[2] Wang H, Han J H, Deng L, Cheng K Q 2013 *Acta Phys. Sin.* **62** 110505 (in Chinese)[王辉, 韩江洪, 邓林, 程克勤 2013 物理学报 **62** 110505]

[3] Li R Q, Tang M, Hui P M 2013 *Acta Phys. Sin.* **62** 168903 (in Chinese)[李睿琪, 唐明, 许伯铭 2013 物理学报 **62** 168903]

[4] Huang B, Zhao X Y, Qi K, Tang M, Du Y H 2013 *Acta Phys. Sin.* **62** 218902 (in Chinese)[黄斌, 赵翔宇, 齐凯, 唐明, 都永海 2013 物理学报 **62** 218902]

[5] Weng W G, Ni S J, Shen S F, Yuan H Y 2007 *Acta Phys. Sin.* **56** 1938 (in Chinese)[翁文国, 倪顺江, 申世飞, 袁宏永 2007 物理学报 **56** 1938]

[6] Ou Y M, Fei Q, Yu M H 2008 *Acta Phys. Sin.* **57** 6763 (in Chinese)[欧阳敏, 费奇, 余明晖 2008 物理学报 **57** 6763]

[7] Song Y R, Jiang G P 2010 *Acta Phys. Sin.* **59** 7546 (in Chinese)[宋玉蓉, 蒋国平 2010 物理学报 **59** 7546]

[8] Wang Y Q, Yang X Y 2013 *Chin. Phys. B* **22** 010509

[9] Song Y R, Jiang G P, Gong Y W 2013 *Chin. Phys. B* **22** 040205

[10] Solé R V, Valverde S 2004 *Lect. Notes Phys.* **650** 189

[11] Costa L F, Rodrigues F A, Travieso G, Boas P R V 2007 *Adv. Phys.* **56** 167

[12] Wu J, Tan Y J, Deng H Z, Zhu D Z 2007 *Systems Engineer Theory & Practice* **27** 101 (in Chinese)[吴俊, 谭跃进, 郑宏钟, 朱大智 2007 系统工程理论与实践 **27** 101]

[13] Wang B, Tang H W, Guo C H, Xiu Z L 2006 *Physica A* **363** 591

[14] Cai M, Du H F, Ren Y K, Feldman M 2011 *Acta Phys. Sin.* **60** 110513 (in Chinese)[蔡萌, 杜海峰, 任义科, 费尔德曼 2011 物理学报 **60** 110513]

[15] Freeman L C, Borgatti S P, White D R 1991 *Social Networks* **13** 141

[16] Winston W L 1994 *Operations Research: Applications and Algorithms* (Belmont: Duxbury Press) p15

[17] Borgatti S P, Everett M G 2006 *Social Networks* **28** 466

[18] Putnam R D 2000 *Bowling Alone: The Collapse and Revival of American Community* (New York: Simon & Schuster) p65

[19] Freeman L C 1980 *Quality and Quantity* **14** 585

[20] Wang X F, Li X, Chen G R 2006 *Complex Network Theory and Application* (Vol. 1) (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]

[21] Erods P, Renyi A 1960 *Publ. Math. Inst. Hungary Acd. Sci.* **5** 17

[22] Barabasi A L, Albert R 1999 *Science* **286** 509

[23] Girvan M, Newman M E J 2002 *PNAS* **99** 7821

[24] Lancichinetti A, Fortunato S, Radicchi F 2008 *Phys. Rev. E* **78** 046110

[25] Watts D J, Strogatz S H 1998 *Nature* **393** 440

[26] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E, Dawson S M 2003 *Behav. Ecol. Sociobiol.* **54** 396

[27] Liu J G, Ren Z M, Guo Q, Wang B H 2013 *Acta Phys. Sin.* **62** 178901 (in Chinese)[刘建国, 任卓明, 郭强, 汪秉宏 2013 物理学报 **62** 178901]

A new network structure entropy based on maximum flow*

Cai Meng[1)2)] Du Hai-Feng[2)†] Marcus W Feldman[2)3)]

1) (*School of Management, Xi'an Jiaotong University, Xi'an 710049, China*)

2) (*Center for Administration and Complexity Science, Xi'an Jiaotong University, Xi'an 710049, China*)

3) (*Morrison Institute for Population and Resource Studies, Stanford University, Stanford 94305, USA*)

(Received 8 October 2013; revised manuscript received 20 November 2013)

Abstract

Entropy is an index to reflect the heterogeneity of network structure. By introducing the concept of network flow which comprehensively considers radial measurement and betweenness measurement, we define a new network structure entropy index to solve the problem that classical entropy indices cannot effectively reflect heterogeneity of the global network. Analysis results concerning specific network (e.g. public data set Dolphins network) indicate that this new entropy index can reflect the real topological structure of network, and effectively overcome the shortcomings of other network entropy indices to some extent. The theoretical analyses and simulation experiments on Erdös-Rēnyi random network, nearest-neighbor coupled network, star network, Barabási-Albert scale-free network, Benchmark network, and the Watts-Strogatz small-world network further prove the effectiveness and applicability of this new network structure entropy index to describe the characteristics of ordinary complex network structures.

Keywords: complex network, heterogeneity, entropy

PACS: 05.90.+m, 01.75.+m **DOI:** 10.7498/aps.63.060504

* Project supported by the National Social Science Foundation of China (Grant No. 12AZD110), the National Natural Science Foundation of China (Grant No. 71071128), the Program for New Century Excellent Talents in University of Ministry of Education of China (Grant No. NCET-08-0451) and the Fundamental Research Funds for the Central Universities, China (Grant No. 2011jdgz08).

† Corresponding author. E-mail: haifengdu@mail.xjtu.edu.cn