

一种基于点和边差异性的网络结构熵*

蔡 萌¹⁾²⁾ 杜海峰^{2)†} 任义科³⁾ 费尔德曼²⁾⁴⁾

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)(山西师范大学管理学院, 临汾 041004)

4)(斯坦福大学莫里森人口与资源研究所, 斯坦福 94305)

(2011 年 1 月 5 日收到; 2011 年 2 月 21 日收到修改稿)

熵是反映网络异构性的重要指标. 由于只是关注网络结构中“点”或“边”的单一作用, 基于度分布和度相对值的两种传统熵在刻画网络结构特征时均存在缺陷. 文章综合考虑“点”和“边”差异性, 定义一种新的网络结构熵, 并对规则网络、随机网络和无标度网络等结构熵进行理论分析和仿真实验. 结果表明, 这种新网络结构熵可以更加有效地反映网络的结构特征, 尤其是对于稀疏网络及星型网络的结构差异解释更为合理.

关键词: 均匀网络, 无标度网络, 熵

PACS: 05. 90. + m, 01. 75. + m

1. 引言

描述自然界复杂系统的多数网络都体现出高度的异构性——网络结构不均衡性. 随着复杂网络研究的兴起, 网络中的传播问题得到了学者们的高度关注, 无论是病毒传播^[1-6]还是灾害蔓延^[7,8], 都与网络的异构性不无关系^[9]. 熵作为系统无序程度的度量, 凭借其独特的内涵已被应用于统计物理学、信息论以及其他广义系统, 成为了复杂系统研究的重要工具, 是目前测度网络异构性的重要指标^[10]. 采用何种形式的熵来反映网络特征, 一直是复杂网络研究者关注的焦点.

目前有关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]、网络结构熵^[12]以及剩余度熵等^[10].

度分布熵(下文简称 DD 度分布熵) $H = - \sum_{k=1}^{N-1} p(k) \log p(k)$ 是一种有效度量网络随机故障

鲁棒性的指标^[13], 其中 $p(k)$ 为网络节点度为 k 的分布函数, N 为网络中节点数目. 文献[10]指出度分布熵可以测度网络异构性, 对于各节点度值相同的规则网络, 有 $H_{\min} = 0$; 对于节点度值各不相同的网络, 即 $p(k) = 1/(N-1)$, 有 $H_{\max} = \log(N-1)$. 对于所有节点均仅与某一中心节点相连的星型网络, 其度分布熵值与各节点度值相同的规则网络近似, 表明网络结构相对均衡, 事实上, 一般认为星型网络异构性很强^[12]. 因此, 度分布熵对星型网络等特殊网络异构性测度的有效性和准确性值得探讨.

Wu 等提出的 Wu 结构熵^[12]是最典型的采用节点相对度值作为测度网络异构性的指标, 其定义为

$$H = - \sum_{i=1}^N \frac{k_i}{\sum_{i=1}^N k_i} \log \left(\frac{k_i}{\sum_{i=1}^N k_i} \right). \quad \text{显然, 对于星型网络,}$$

Wu 结构熵取得最小值 $H_{\min} = \frac{\log 4(N-1)}{2}$; 而对于

各节点度值相同的规则网络, 取得最大值 $H_{\max} = \log N$. Wu 结构熵将度值大小作为异构性衡量标准, 以节点为对象, 关注网络连接的分布定义节点重要性, 而忽略了节点本身的特性. 例如在网络演化过程或是在稀疏网络中, Wu 结构熵只考虑网络中边

* 国家教育部新世纪优秀人才支持计划(批准号: NCET-07-0668, NCET-08-0451)、国家自然科学基金(批准号: 71071128)、西安交通大学“985 工程”三期、中央高校基本科研业务费专项资金和教育部长江学者和创新团队发展计划(批准号: IRT0855)资助的课题.

† 通讯联系人. E-mail: haifengdu@mail.xjtu.edu.cn

的分布,忽略了孤立节点的存在对于网络结构的影响^[14]. 如图1, Wu 结构熵的计算是将图1(a)中的网络简化为图1(b),这本质上是改变了网络的规模和密度. 虽然通过归一化 $\bar{H} = \frac{H - H_{\min}}{H_{\max} - H_{\min}}$, 将网络规模重新纳入结构熵定义,但是仍存在这样的问题: 归一化后的标准结构熵适用于表达同等规模的网络,而在解释不同规模网络的时候不够恰当. 例如规模为10的星型网络和规模为1000的星型网络具有相同的标准结构熵是不太合理的. 如果考虑网络规模的影响,则需要使用非标准化的结构熵. 故采用 Wu 结构熵对网络异构性进行比较的时候,对于稀疏网络和星型网络的解释存在问题.

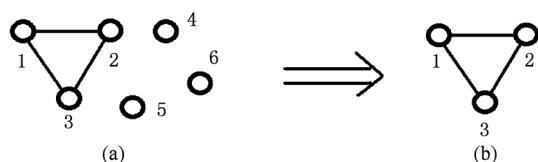


图1 Wu 结构熵对于孤立节点的处理

度分布熵和 Wu 结构熵在测度网络异构性时存在缺陷的本质原因在于,均仅从一方面考虑了网络的结构特征:

度分布熵: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.

度相对值熵,也称 Wu 结构熵:以“节点”为研究对象或基本元(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.

节点和边是构成网络的最基本的两种元素,熵作为对网络系统结构分布度量的宏观指标,应将网络中节点与边的因素统一起来作为测度标准,而不应片面的仅考虑其中一种因素.

因此,本文基于“点”差异性和“边”差异性,定义网络中节点的结构重要性(下文简称重要性),进而提出一种新的网络结构熵. 通过对规则网络、ER 随机网络和 BA 无标度网络等结构熵进行理论分析和仿真实验,验证本文提出的网络结构熵的有效性及优越性,证实其能够充分反映网络的结构特征并在一定程度上克服度分布熵和度相对值熵(Wu 结构熵)的缺陷.

2. 基于点和边差异性的网络结构熵定义

与热力学中熵的概念不同,网络熵的定义围绕节点或边的特性展开,一般将节点度取值的大小或者分布作为“微观状态”,进而计算网络的熵. 以度分布为基础的“度分布熵”关注按照度值大小分类的边在网络中的特殊地位或重要程度,一种度值在网络中的结构地位取决于具有该度值的节点在网络中的相对数量. 将节点度值作为衡量节点在网络中重要程度标准的“Wu 结构熵”,关注节点重要性在网络结构中的分布,一个节点在网络中的结构地位取决于该节点拥有的相对边数.

为避免片面采用“点”或“边”作为网络异构性的测度标准,本文将结构重要性作为引入熵概念的基本思想. 而熵表示网络系统内的混乱程度,即“非一致性”,也就是“异构性”. 这种“异构性”应同时反映网络的两种基本构成:网络中的节点及网络中的边. 为从更为全面的角度定义一个网络结构熵,首先需要分别定义网络中“点”的差异性以及“边”的差异性.

“点”差异性 S_i : k_i 为网络中节点 i 的度, $p(k_i)$ 表示网络中度为 k_i 的节点的概率. 如果将具有的度值为 k_i 的节点 i 作为基,那么网络的异构性可以理解为网络中与该度值不同的节点的个数. 若网络中的其他节点的度值均与节点 i 相同,即服从度值为 k_i 的基,所构成的网络就是一个所有节点具有相同度值的规则网络,是一个不具备异构性的均匀网络,差异程度最低. 若网络规模为 N ,记节点 i 的“点”差异性为

$$S_i = [1 - p(k_i)]N. \quad (1)$$

“边”差异性 D_i : 网络中任意边所连接的节点的度值应该服从 $k p(k)$, 即一个拥有度值为 k 的节点被选中的概率应为度值为 1 的节点的 k 倍^[15]. 与一般的结构熵(如 Wu 结构熵)定义相同,网络中的孤立节点永远不会被选中. 同样,将节点 i 具有的度值 k_i 作为网络的基,节点 i 的“边”差异性定义为

$$D_i = k_i [1 - p(k_i)]N. \quad (2)$$

综合考虑网络节点 i 的“点”差异性和“边”差异性,其结构重要性可定义为

$$I'_i = \alpha S_i + \beta D_i = \alpha [1 - p(k_i)]N + \beta k_i [1 - p(k_i)]N, \quad (3)$$

其中, $\alpha + \beta = 1, 0 \leq \alpha, \beta \leq 1$.

α 和 β 分别为点差异性和边差异性的权重. $\alpha = 1, \beta = 0$ 表示仅考虑点差异性对于网络异构性的影响, 而不考虑边差异性; $\alpha = 0, \beta = 1$ 则表示仅考虑边差异性对于网络异构性的影响.

因此, 节点 i 在网络中的相对重要性可定义为

$$I_i = \frac{I'_i}{\sum_{j=1}^N I'_j} = \frac{\alpha[1 - p(k_i)]N + \beta k_i[1 - p(k_i)]N}{\sum_{j=1}^N \{\alpha[1 - p(k_j)]N + \beta k_j[1 - p(k_j)]N\}} = \frac{[1 - p(k_i)]N + \frac{\beta}{\alpha} k_i[1 - p(k_i)]N}{\sum_{j=1}^N \{[1 - p(k_j)]N + \frac{\beta}{\alpha} k_j[1 - p(k_j)]N\}}. \quad (4)$$

(4) 式中 β/α 为常量, 本文认为点差异和边差异对网络的异构具有同等的影响力, 即 $\beta/\alpha = 1$, 并不影响后文的分析, 故

$$I_i \approx \frac{(k_i + 1)[1 - p(k_i) + \Delta]}{\sum_{j=1}^N \{(k_j + 1)[1 - p(k_j) + \Delta]\}}. \quad (5)$$

物理系统中, (5) 式代表了节点在网络中的结构重要性, 表明了网络中节点重要程度的分布, 也可以理解为网络系统能量的分布. 式中 $\Delta \sim O(\frac{1}{N^2})$, $N > 1$. Δ 项的引入是为了在使用公式计算时, 避免 $p(k_i) \equiv 1$ 的规则网络计算无意义, 其存在并不影响网络结构异构性的分析结果.

(5) 式最右端的近似表达具有明确的社会意义: 一方面, 基于社会资本理论, 如果将一个节点拥有连接的数量作为一种资源的占有, 那么显然拥有较大度值的节点占据的网络资源也就越多, 其重要程度也就越大; 另一方面, 依据权力依赖理论, 对于节点 i , 其度值出现的概率 $p(k_i)$ 越大, 说明网络中与其具有相同地位或拥有相同资源的节点越多, 该节点的不可替代性越弱, 其重要程度也就越低. 相反, $p(k_i)$ 越小说明节点 i 的网络中的地位越特殊, 越不可替代, 也就越重要, 即节点所拥有度值在系统中出现的概率 (反映了节点的特殊程度) 也应作为重要程度的度量标准之一.

基于 (5) 式, 本文的网络结构熵定义为

$$H = - \sum_{i=1}^N I_i \log I_i = - \sum_{i=1}^N \frac{(k_i + 1)[1 - p(k_i) + \Delta]}{\sum_{j=1}^N \{(k_j + 1)[1 - p(k_j) + \Delta]\}} \times \log \frac{(k_i + 1)[1 - p(k_i) + \Delta]}{\sum_{j=1}^N \{(k_j + 1)[1 - p(k_j) + \Delta]\}}. \quad (6)$$

如果将其推广至动态网络中, 网络在 t 时刻的结构

熵可表达如下:

$$H(t) = - \sum_{i=1}^{N(t)} I_{i,t} \log I_{i,t} = - \sum_{i=1}^{N(t)} \frac{(k_{i,t} + 1)[1 - p(k_{i,t}) + \Delta]}{\sum_{j=1}^{N(t)} \{(k_{j,t} + 1)[1 - p(k_{j,t}) + \Delta]\}} \log \frac{(k_{i,t} + 1)[1 - p(k_{i,t}) + \Delta]}{\sum_{j=1}^{N(t)} \{(k_{j,t} + 1)[1 - p(k_{j,t}) + \Delta]\}}. \quad (7)$$

其中, $N(t)$ 为 t 时刻的网络规模, $k_{i,t}$ 为 t 时刻节点 i 的度值. 对于固定的网络结构, 网络结构熵反映了网络结构的平均信息量; 而当新的节点添加进网络时, 则反映添加前的平均不确定性.

对于本文提出的网络结构熵, 有 3 点需要说明:

1) 本文所提出的网络结构熵并不是和网络结构一一对应的, 该指标只是作为网络结构的一种宏观度量, 不同的网络结构可能具有相同的熵值. 图 2 为几种简单结构的网络熵示意, 图 2(a) 的结构熵为 $\log 2$, 为均匀网络, 如果将节点的度看作该节点对资

源的占有程度,那么该网络中两节点地位对等.在图 2(a)中引入新的孤立节点 3 得到图 2(b),该节点虽然对资源毫无掌控,但在网络中的地位却较为特殊.作为潜在资源会成为节点 1,2 争夺的焦点,节点 3 起到平衡节点 1,2 重要性的作用,该网络结构也是一种均衡的结构,熵值为 $\log 3$.由于在规模越大的网络中均衡的难度越大,但也更稳定,故图 2(b)比图 2(a)的熵值更大.图 2(c)同样是网络规模为 3 时的另一种均衡模型,熵值也为 $\log 3$.

在 DD 度分布熵测度下,图 2 三种网络的熵值从左到右依次为 0, $\left(\frac{2}{3}\log \frac{3}{2} + \frac{1}{3}\log 3\right)$, 0(为保证熵计算时对于各个度值分布的完备性,将 DD 度分布熵略作变化为: $H = -\sum_{k=0}^{N-1} p(k)\log p(k)$).显然,认为图 2(b)比图 2(a)和图 2(c)的网络更为均衡是不合理的.

而在 Wu 结构熵测度下,图 2 三种网络的熵值从左到右依次为 $\log 2, \log 2, \log 3$.显然,认为图 2(b)和图 2(a)的熵值相同是不太合理的,这样的结果在于忽略了孤立节点的作用,进而无法反映网络的规模效应.

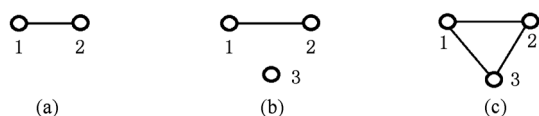


图 2 网络结构熵示例

2) 本文所定义的结构熵,两处求和均是对节点 i 而非对度值 k ,故可以将 $p(k_i)$ 看做是 k_i 的权重因子,是节点 i 结构特殊性的体现.对于一个确定的网络, $p(k_i)$ 恒不为零.

3) 本文所定义的结构熵, $p(k_i)$ 和 k_i 共同作用反映节点 i 的重要性,两者互为补充.即使在网络

中拥有较大度数的节点,要获得网络结构中重要的地位,不仅要求与其具有相同的度数的节点要少,而且与其具有相似度数的节点也要少.而一个度值并不非常大的节点,也能通过其在网络中拥有的“与众不同”的地位而显得较为重要.

3. 典型网络结构熵

一般来说,可以将网络模型按照其特征分为以下 3 类^[16]:第 1 类为随机网络,包括按照统计规律生成随机图及其派生模型;第 2 类为规则网络,包括最近邻耦合网络、全连接网络、环形网络、星型网络等规则结构的网络模型;介于上述两者之间的网络结构统归第 3 类,这些网络兼具规则网络和随机网络的部分特征,包括无标度网络和小世界网络等.

本文将针对 ER 随机网络、最近邻耦合网络、星型网络和 BA 无标度网络 4 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

1960 年由 Erdos 和 Renyi 提出的 ER 随机网络模型^[17],是网络研究的主要参考模型之一.构造方法一般如下:

1) 确定网络规模 N . 2) 每次任选两个节点,以概率 $q = \frac{2n}{N(N-1)}$ 将其相连,其中 n 为给定的总边数. 3) 当边数达到 n 时停止演化.该模型可能出现 $\left(\frac{N(N-1)}{2}\right)$ 个等概率的网络,其平均度 $\langle k \rangle = q(N-1)$,且服从 Poisson 分布: $P(k) \approx e^{-\langle k \rangle} \langle k \rangle^k / k!$.将 $P(k) \approx e^{-\langle k \rangle} \langle k \rangle^k / k!$ 代入(5)式中,有

$$\begin{aligned}
 I_i &= \frac{(k_i + 1)[1 - p(k_i) + \Delta]}{\sum_{j=1}^N \{(k_j + 1)[1 - p(k_j) + \Delta]\}} \\
 &= \frac{(k_i + 1)(1 - e^{-\langle k \rangle} \langle k \rangle^{k_i} / k_i! + \Delta)}{\sum_{j=1}^N [(k_j + 1)(1 - e^{-\langle k \rangle} \langle k \rangle^{k_j} / k_j! + \Delta)]} \\
 &= \frac{(k_i + 1)(1 - e^{-\langle k \rangle} \langle k \rangle^{k_i} / k_i! + \Delta)}{\sum_{k=0}^{N-1} [N \cdot P(k) \cdot (k + 1) \cdot (1 - e^{-\langle k \rangle} \langle k \rangle^k / k! + \Delta)]} \\
 &= \frac{(k_i + 1)[1 - e^{-\langle k \rangle} \langle k \rangle^{k_i} / k_i! + \Delta]}{\sum_{k=0}^{N-1} [N \cdot (k + 1) \cdot (e^{-\langle k \rangle} \langle k \rangle^k / k!) \cdot (1 - e^{-\langle k \rangle} \langle k \rangle^k / k! + \Delta)]}. \quad (8)
 \end{aligned}$$

故 ER 随机网络的结构熵为

$$\begin{aligned}
 H = & - \sum_{d=0}^{N-1} N \cdot (e^{-\langle k \rangle} \langle k \rangle^d / d!) \\
 & \times \left\{ \frac{(d+1)(1 - e^{-\langle k \rangle} \langle k \rangle^d / d! + \Delta)}{\sum_{k=0}^{N-1} [N \cdot (k+1) \cdot (e^{-\langle k \rangle} \langle k \rangle^k / k!) \cdot (1 - e^{-\langle k \rangle} \langle k \rangle^k / k! + \Delta)]} \right\} \\
 & \times \left\{ \log \frac{(d+1)(1 - e^{-\langle k \rangle} \langle k \rangle^d / d! + \Delta)}{\sum_{k=0}^{N-1} [N \cdot (k+1) \cdot (e^{-\langle k \rangle} \langle k \rangle^k / k!) \cdot (1 - e^{-\langle k \rangle} \langle k \rangle^k / k! + \Delta)]} \right\}. \quad (9)
 \end{aligned}$$

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^[16]是一种典型的规则网络,即网络中每一个节点只和它周围的邻居节点相连. 对于一个给定的 k 值 (k 为偶数), 如果将网络中的 N 个节点围成一个环, 其中每个节点只与它两边的 $k/2$ 个邻居节点相连, 当 k 足够大时具有较大的聚类系数 $C = 3(k-2)/4(k-1) \approx 0.75$, 而当 N 足够大时则具有较大的平均路径长度 $L \approx N/2k$.

对于本文提出的结构熵, 有 $H \leq \log N$, 当且仅当 $(k_i + 1)[1 - p(k_i) + \Delta]$ 对于任意 $i, i \in \{1, 2, \dots, N\}$ 均相等时取等号. 由于最近邻耦合网络中所有节点具有相同的度值, 因此具有同等规模下网络可能出现的最大熵值:

$$H_{\max} = - \sum_{i=1}^N \frac{1}{N} \log \frac{1}{N} = \log N. \quad (10)$$

3.2.2. 星型网络 (star network)

对于固定的网络规模 N , 网络结构熵的最小值出现在星型网络^[16], 即网络中所有节点均仅与某一中心节点相连, 不妨设该中心节点为 k_1 , 则 $k_1 = N-1, k_i = 1 (i \neq 1)$.

$$\begin{aligned}
 H_{\min} = & - \left[\frac{N}{N+2} \log \frac{N}{N+2} + (N-1) \right. \\
 & \times \left. \frac{2}{(N-1)(N+2)} \log \frac{2}{(N-1)(N+2)} \right] \\
 = & \log(N+2) + \frac{2}{N+2} \\
 & \times \log \frac{N-1}{2} - \frac{N}{N+2} \log N. \quad (11)
 \end{aligned}$$

3.3. BA 无标度网络 (scale-free network) 结构熵

尽管无标度网络已发展出多种扩展模式^[18,19], 然而其基本思想是由 Barabasi 和 Albert 提出的, 称为 BA 无标度网络^[20]. 他们追踪万维网络的动态演化过程, 从动态、增长的观点研究了具有幂律分布的网络的形成机理. 通过增长和择优连接, BA 无标度网络模型最终服从度分布函数 $p(k) \propto 2m^2 k^{-3}$, 其中 m 为网络演化增长率, 即每步的新增节点数. 相对于随机网络的 Poisson 分布来说, 幂率分布的无标度网络最显著的特点就是其结构的“不均匀”性. 无标度网络的度分布函数没有峰值, 大多数节点仅有少量连接, 而少数节点拥有大量连接.

根据文献[21], 将 $p(k) = 2m^2 k^{-3}$ 代入(5)式中, 有

$$\begin{aligned}
 I_i = & \frac{(k_i + 1)[1 - p(k_i) + \Delta]}{\sum_{j=1}^N \{(k_j + 1)[1 - p(k_j) + \Delta]\}} \\
 = & \frac{(k_i + 1)[1 - 2m^2 k_i^{-3} + \Delta]}{\sum_{j=1}^N \{(k_j + 1)[1 - 2m^2 k_j^{-3} + \Delta]\}}. \quad (12)
 \end{aligned}$$

由于每个节点 i 都是在特定时间 t_i 加入到网络中的, 故对节点 i 求和等同于对网络的演化时间 t_i 求积分,

根据文献[21], 又有 $k_i = m \left(\frac{t}{t_i} \right)^{\frac{1}{2}}$, 故

$$\begin{aligned}
 I_i \approx & \frac{(k_i + 1)[1 - 2m^2 k_i^{-3}]}{\int_0^t (k_i + 1)(1 - 2m^2 k_i^{-3}) dt_i} \\
 = & \frac{m \left(\frac{t}{t_i} \right)^{\frac{1}{2}} + 1 - 2m^{-1} \left(\frac{t}{t_i} \right)^{-\frac{3}{2}} - 2 \left(\frac{t}{t_i} \right)^{-1}}{2mt - \frac{4}{5} m^{-1} t}. \quad (13)
 \end{aligned}$$

BA 无标度网络在 t 时刻的演化结构熵可表示为

$$H(t) = - \sum_{i=1}^N \frac{m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} + 1 - 2m^{-1}\left(\frac{t}{t_i}\right)^{-\frac{3}{2}} - 2\left(\frac{t}{t_i}\right)^{-1}}{2mt - \frac{4}{5}m^{-1}t} \times \log \frac{m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} + 1 - 2m^{-1}\left(\frac{t}{t_i}\right)^{-\frac{3}{2}} - 2\left(\frac{t}{t_i}\right)^{-1}}{2mt - \frac{4}{5}m^{-1}t}. \quad (14)$$

3.4. 四种典型网络的结构熵仿真模拟

分别对最近邻耦合网络、星型网络、ER 随机网络和 BA 无标度网络 4 种典型网络进行数值模拟, 并对 3 种网络熵进行比较, 说明本文所提出网络结构熵的有效性和优越性. 实验均是在 Intel^R CoreTM 22.40 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. 记 SD 为采用本文提出的网络结构熵计算网络指标, Wu 为 Wu 结构熵, DD 为度分布熵. 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 4 种典型网络的结构熵. BA 无标度网络所示曲线表示按照文献[20]中网络演化过程的熵, 最终的网络演化规模定为 1000. 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结构熵. 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.

实验发现 SD 结构熵与 Wu 结构熵指标随 BA 无标度网络增长率 m 的变化具有相同的趋势, 且变化不大(随 m 增大熵值略有提升), 故图 3(a) 和图 3(b) 仅给出变化率 $m=10$ 时的情况作为示例. 图 3(c) 中的最近邻耦合网络和 ER 随机网络也仅对应 BA 无标度网络 $m=10$ 时的情况.

图 3(a) 和图 3(b) 中相同规模下结构熵的最大值和最小值分别为最近邻耦合网络和星型网络. ER 随机网络的结构熵几乎与同规模下的最近邻耦合网络结构熵相等(实际上 ER 随机网络的结构熵略低于同规模下的最近邻耦合网络结构熵). 对比发现本文提出的网络结构熵与 Wu 的网络结构熵在刻画最近邻耦合网络、BA 无标度网络以及 ER 随机网络是基本一致的. 但在刻画星型网络时, 两者的主要区别在于 SD 结构熵随网络规模的增加而减小, Wu 结构熵则相反. 按照文献[12]的理解, 在 Wu 结构熵指标下, 星型网络作为异构性最强的

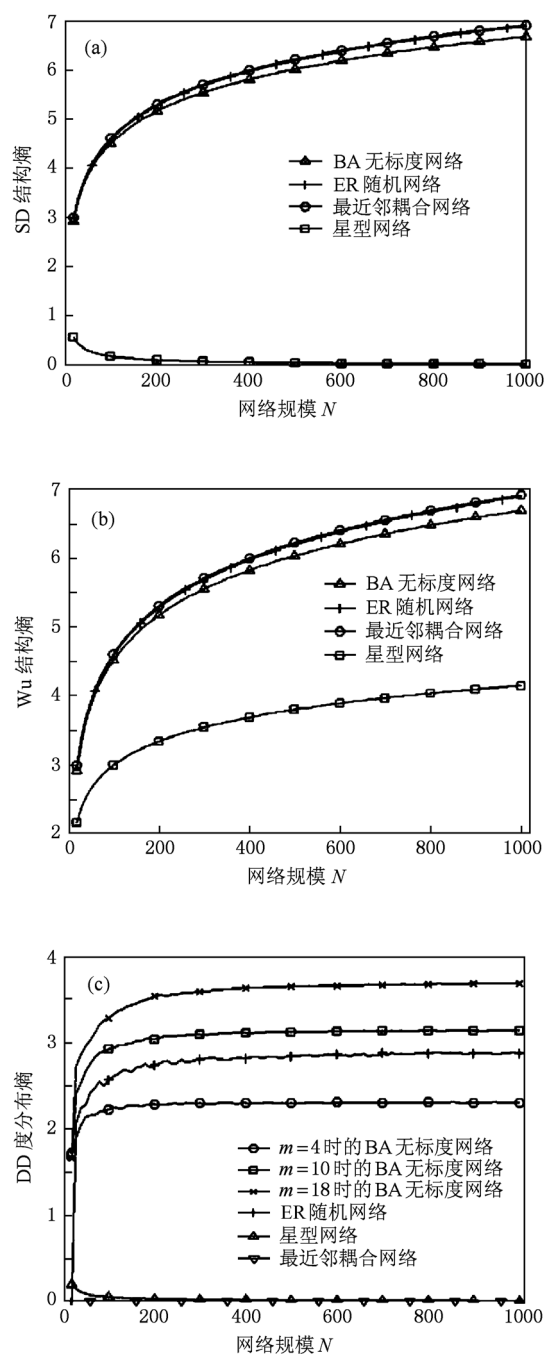


图3 (a)四种典型网络的SD结构熵;(b)四种典型网络的Wu结构熵;(c)四种典型网络的DD度分布熵

代表,那么随着网络规模的增加,这种异构性应该得到增强. 相较于 Wu 网络结构熵,本文提出的 SD 网络结构熵指标能够更好地反映这种理解.

图 3(c) 为 DD 度分布熵指标. 根据 BA 无标度网络增长率 $m = 4$, $m = 10$ 和 $m = 18$ 时的度分布熵,可以发现增长率的变化对于度分布熵有较大影响,且在网络演化过程中当规模达到一定程度时,度分布熵不再随网络规模的增加而变化. 最近邻耦合网络的 DD 度分布熵恒为零,星型网络与其较为接近. 对比 $m = 10$ 时的 BA 无标度网络与 ER 随机网络,发现随机网络的熵值相对较低,这与一般的理解不太一致^[12].

图 4 为按照 3.1 节中方式演化的 ER 随机网络结构熵,网络规模为 1000,最终演化密度为 0.06. 与 BA 无标度网络类似,ER 随机网络的演化过程在 3 种网络熵指标下均体现出明显的熵增效应. SD 结构熵和 Wu 结构熵指标下,网络在较短时间内演化为稳定状态(相比 Wu 结构熵,在 SD 结构熵的测度下,网络进入稳态的时间更短),其熵值与相同规模下的规则网近似,约为 $\ln 1000$. 需要注意的是,虽然这两种网络熵演化过程极为相似,而表示的含义却大不相同,尤其是在网络稀疏时(网络演化进入稳态前的阶段),如在 t_1 时刻,即在网络中加入第一条边后, Wu 结构熵指标仅考虑拥有连接两个节点,此时熵值取得最大值 $\ln 2$,等价于仅由两个节点构成的全连网. 在整个网络演化过程中这个熵值显得较小的原因是由于规模效应,即等价后的网络最大熵值为 $\ln M$, M 为等价后的网络规模. 而 SD 结构熵指标则是将整个网络的结构纳入考虑,如 t_1 时刻熵值在整个网络演化过程中显得较小的原因是由于在整个网络规模为 1000 的情况下,仅有一对节点相连

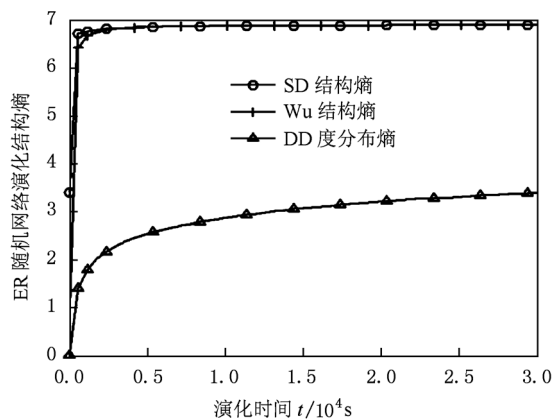


图 4 ER 随机网络演化结构熵

的网络异构性是很强的,节点间的差异情况很大.

在 DD 度分布熵指标下,网络需要较长时间才能到达稳态,熵值随网络演化时间的增长速度明显慢于其他两种网络结构熵. SD 和 Wu 结构熵当密度达到一定程度后,不随密度而变化,而 DD 则受密度影响.

图 5 为网络规模 1000,不同密度时的 BA 无标度网络与 ER 随机网络 SD 熵值对比,可以发现,无论网络密度如何,ER 随机网络的熵值始终高于 BA 无标度网络,这也验证了 BA 无标度网络的异构性相对更强. 同时还发现,网络密度小于 0.1 时,两种网络的熵值与密度正相关,当密度大于 0.1 时,密度对熵值的影响则不大.

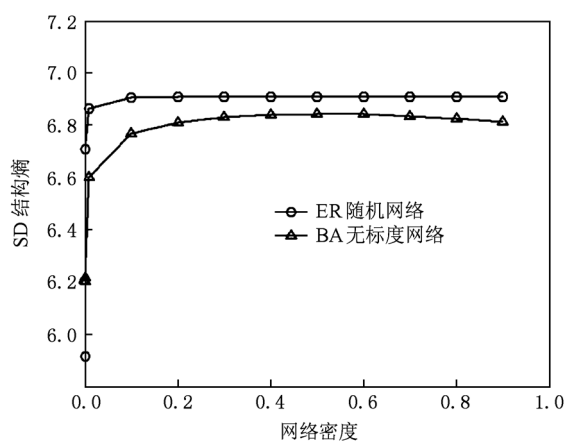


图 5 BA 无标度网络与 ER 随机网络 SD 熵值对比

典型网络结构仿真实验表明,与采用度相对值的 Wu 网络结构熵和 DD 度分布熵相比,本文提出的网络结构熵可以有效地反映网络结构的异质性. 网络熵作为网络异质性的指标,一般认为采用 Wu 结构熵是比较合适的,而 DD 度分布熵在刻画网络的异质性方面则有欠缺^[22]. 本文提出的网络结构熵对于最近邻耦合网络、ER 随机网络以及 BA 无标度网络的刻画,与 Wu 结构熵都能较为准确的反映其异构性,且计算结果近似;在刻画星型网络以及稀疏网络时,本文提出的网络结构熵在一定程度上克服了 Wu 结构熵的不足,拥有更好的解释性.

4. 结 论

本文从“点”差异性和“边”差异性两方面综合考虑度值大小和度分布作为结构重要性的度量,进而提出一种新的网络结构熵定义. 通过对最近邻耦

合网络、星型网络、BA 无标度网络及 ER 随机网络等四种典型网络的结构熵分析及对比,验证本文所定义的网络结构熵能够较好的反映网络的异构性,克服了采用度相对值的网络结构熵和度分布熵的缺陷,并得到了该结构熵定义下的 BA 无标度网络和 ER 随机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实验分析,更为深入的理论分析和应用研究是后续研究的主要内容,具体地:

本文提出的网络结构熵是以度中心性作为节点重要程度的标准,而节点重要程度应该具有多种度量方式. 如流中心性^[23]可以刻画节点对于信息传播的控制能力,采用流中心性定义的结构熵是否也能有效刻画网络的异构性? 如图 6 所示,图中所有的节点均拥有相同的度值,在目前的结构熵测度下应该是均匀网络. 然而节点 1,2,3,4 显然比其他节点在网络传播中更具优势.

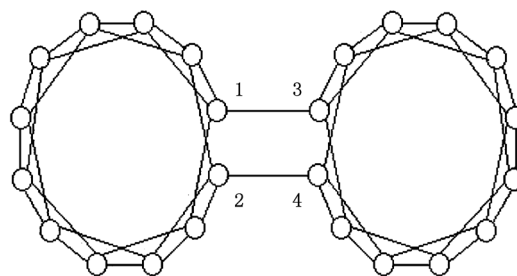


图6 网络中的传播优势

依据节点重要程度计算的熵值反映了网络权利的集散程度,采用现实生活中的网络^[24,25],研究这种结构上的特点对于企业、政府等组织的影响具有现实意义.

BA 无标度网络只是无标度网络中的一个特例,研究无标度网络参数($\alpha k^{-\lambda}$)与熵值之间的对应关系,对于无标度网络的理解和刻画具有积极的意义.

对于典型网络结构熵的更简洁数学解析表达和定量描述.

- [1] Bose T, Trimper S 2009 *Phys. Rev. E* **79** 051903
- [2] Gumel A B, McCluskey C C, Vanden Priessche P 2006 *Bull. Math. Biol.* **68** 2105
- [3] Wang Y, Zheng Z G 2009 *Acta Phys. Sin.* **58** 4421 (in Chinese) [王 延、郑志刚 2009 物理学报 **58** 4421]
- [4] Wang Y Q, Jiang G P 2010 *Acta Phys. Sin.* **59** 6734 (in Chinese) [王亚奇、蒋国平 2010 物理学报 **59** 6734]
- [5] Song Y R, Jiang G P 2009 *Acta Phys. Sin.* **58** 5911 (in Chinese) [宋玉蓉、蒋国平 2009 物理学报 **58** 5911]
- [6] Song Y R, Jiang G P 2010 *Acta Phys. Sin.* **59** 705 (in Chinese) [宋玉蓉、蒋国平 2010 物理学报 **59** 705]
- [7] Piot P, Bartos M, Ghys P D, Walker N, Schwartl B 2001 *Nature* **410** 968
- [8] Helbing D, Molini R 1995 *Phys. Lett. A* **212** 130
- [9] 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]
- [10] Sole' R V, Valverde S 2004 *Lect. Notes Phys.* **650** 189
- [11] Costa L F, Rodrigues F A, Traverso G, Boas P R V 2007 *Adv. in Phy.* **56** 167
- [12] Wu J, Tan Y J, Deng H Z, Zhu D Z 2007 *Sys. Engin. Theo. Pract.* **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 *Phys. A* **363** 591
- [14] Tan Y J, Wu J 2004 *Syst. Engin. Theo. Prac.* **24** 1 (in Chinese) [谭跃进、吴 俊 2004 系统工程理论与实践 **24** 1]
- [15] Newman M E J, Park J 2003 *Phys. Rev. E* **68** 036122
- [16] 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]
- [17] Eros P, Renyi A 1960 *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17
- [18] Li J, Xiao J H, Jiang P Q, Zhou T, Wang W X 2006 *Acta Phys. Sin.* **55** 4051 (in Chinese) [李 季、肖井华、蒋品群、周 涛、王文旭 2006 物理学报 **55** 4051]
- [19] Pan Z F, Wang X F 2006 *Acta Phys. Sin.* **55** 4058 (in Chinese) [潘灶烽、汪小帆 2006 物理学报 **55** 4058]
- [20] Barabasi A L, Albert R 1999 *Science* **286** 509
- [21] He D R, Liu Z H, Wang B H 2009 *Complex Systems and Complex Networks* (Vol. 1) (Beijing: Higher Education Press) p141 (in Chinese) [何大韧、刘宗华、汪秉宏 2009 复杂系统与复杂网络 (北京:高等教育出版社) 第 141 页]
- [22] Wang L, Dai G Z 2009 *Scale-free Characteristic, Scale-free Phenomenon and Their Control of Complex Networks* (Vol. 1) (Beijing: Science Press) p38 (in Chinese) [王林、戴冠中 2009 复杂网络的 Scale-free 性、Scale-free 现象及其控制 (北京:科学出版社) 第 38 页]
- [23] Newman M E J 2005 *Social Network* **27** 39
- [24] Fischback K, Gloor P A, Schoder D 2009 *Busin. Inform. Sys. Engin.* **1** 140
- [25] Salathé M, Kazandjieva M, Lee J W, Leris P, Feldman M W, Jones J H 2010 *PNSA* **107** 22020

A new network structure entropy based node difference and edge difference^{*}

Cai Meng¹⁾²⁾ Du Hai-Feng^{2)†} Ren Yi-Ke³⁾ Marcus W. Feldman²⁾⁴⁾

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) (School of Management, Shanxi Normal University, Linfen 041004, China)

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

(Received 5 January 2011; revised manuscript received 21 February 2011)

Abstract

Entropy is a very important indicator which describes the heterogeneity of the networks. Two kinds of structure entropies, which are based on degree distribution and relative degree respectively, have some problems to describe the characteristics of network structure. Because they focus only on one of “node” or “edge” factors. In this paper, We consider the difference between “node” and “edge” to define a new network structure entropy. The theoretical analyses and the simulation experiments on regular network, random network and scale-free network prove that this new network structure entropy is more effective to reflect the characteristics of network structure, especially more reasonable to describe the sparse network and star network.

Keywords: homogeneous network, scale-free network, entropy

PACS: 05.90.+m, 01.75.+m

^{*} Project supported by the Program for New Century Excellent Talents in University of Ministry of Education of China (Grant Nos. NCET-07-0668, NCET-08-0451), the National Natural Science Foundation of China (Grant No. 71071128), the 3rd Period of the National 985 Project of the Ministry of Education and Treasury Department of China, the Fundamental Research Funds for the Central Universities and the Program for Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (Grant No. IRT0855).

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