

确定度分布条件下可变同配系数的算法构造与影响分析

李静 张洪欣 王小娟 金磊

Algorithm design and influence analysis of assortativity changing in given degree distribution

Li Jing Zhang Hong-Xin Wang Xiao-Juan Jin Lei

引用信息 Citation: *Acta Physica Sinica*, 65, 094503 (2016) DOI: 10.7498/aps.65.094503

在线阅读 View online: <http://dx.doi.org/10.7498/aps.65.094503>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2016/V65/I9>

您可能感兴趣的其他文章

Articles you may be interested in

考虑自适应巡航车辆影响的上匝道系统混合交通流模型

A hybrid traffic flow model with considering the influence of adaptive cruise control vehicles and on-ramps

物理学报.2016, 65(8): 084503 <http://dx.doi.org/10.7498/aps.65.084503>

雨天高速公路车辆换道模型研究

Study of highway lane-changing model under rain weather

物理学报.2016, 65(6): 064501 <http://dx.doi.org/10.7498/aps.65.064501>

受碾区域内颗粒轴向流动特性的离散元模拟

Granular axial flow characteristics in a grinding area studied by discrete element method

物理学报.2015, 64(23): 234502 <http://dx.doi.org/10.7498/aps.64.234502>

驾驶员预估效应下车流能耗演化机理研究

Analysis of evolution mechanism of traffic energy dissipation by considering driver's forecast effect

物理学报.2015, 64(15): 154503 <http://dx.doi.org/10.7498/aps.64.154503>

部分道路关闭引起的交通激波特性研究

Study of the shock wave induced by closing partial road in traffic flow

物理学报.2015, 64(11): 114502 <http://dx.doi.org/10.7498/aps.64.114502>

确定度分布条件下可变同配系数的算法构造 与影响分析*

李静¹⁾²⁾ 张洪欣¹⁾³⁾ 王小娟^{1)†} 金磊¹⁾

1)(北京邮电大学电子工程学院, 北京 100876)

2)(中国信息安全测评中心, 北京 100085)

3)(安全生产智能监控北京市重点实验室(北京邮电大学), 北京 100876)

(2015年12月13日收到; 2016年1月19日收到修改稿)

复杂网络是现实中大量节点和边的抽象拓扑, 如何揭示网络内部拓扑对网络连通性、脆弱性等特征的影响是当前研究的热点. 本文在确定度分布的条件下, 根据 Newman 提出的同配系数的定义分析其影响因素. 首先在可变同配系数下分别提出了基于度分布的确定算法和基于概率分布的不确定算法, 并分别在三种不同类型的网络 (Erdős-Rényi 网络, Barabási-Albert 网络, Email 真实网络) 中验证. 实验结果表明: 当网络规模达到一定程度时, 确定算法优于贪婪算法. 以此为基础, 分析了同配系数改变时聚类系数的变化, 发现两者之间存在关联性, 并从网络的微观结构变化中揭示了聚类系数变化的原因.

关键词: 同配系数, 度分布, 构造算法, 聚类系数

PACS: 45.70.Vn, 42.65.Sf

DOI: 10.7498/aps.65.094503

1 引言

复杂网络是对现实世界的抽象, 其应用范围横跨生物、物理、通信等各个学科, 通过刻画网络内部的拓扑有助于理解网络的整体特性. 目前对于复杂网络结构的研究有两个重要的方面. 一方面是对度分布的研究, 是网络中每个节点邻接节点数目的统计概率分布. 有关度分布条件下网络连通性、脆弱性、传播特性的研究已广泛展开^[1-4]: 如 Schwarte 等^[5]分析了出度、入度对渗流过程的影响, 并得到了在幂律分布下分布参数 λ_{in} , λ_{out} 是不同的; 在不同网络中的表现也是不同的; Dorogovtsev 等^[6]将对度分布的研究扩展到有向网络, 并给出网络中巨分支的计算方法用以体现网络的连通性; Newman 等^[7]在给定量度分布的条件下分析了网络巨分支的大小, 通过比较计算机网络、科学文献引用网络以

及其他经验网络, 得出仅通过度分布不能准确地认识网络的内在属性. 因此必须提出其他指标来衡量网络中的其他属性, 从而开始了对网络同配系数^[8]的研究. 同配系数是衡量网络内在连通性的重要指标, 是网络中度值大的节点趋向于彼此相连的特性. 如同在社交网络中, 具有相同年龄、地位、财富和教育背景的人倾向于彼此连接. 单独地研究一个节点或者一条边的同配性是无意义的, 因为一个同配网络仍然会存在部分边的连接是异配的.

为了理解同配系数对网络的影响, 应当保持度分布不变. Winterbach 等^[9]提出了一种贪婪算法, 在给定量度分布的条件下来提高网络的同配系数, 通过随机选择节点和边互换来增加同配系数, 并在不同的网络中测试算法的性能. 但是这种算法并不高效, 并且在许多网络中无法大幅度提高同配系数, 这样也就无法全面了解同配系数对网络的影响. Zhang 等^[10]分析了社交网络的结构, 并将社交

* 国家自然科学基金 (批准号: 61571063, 61472357, 61501100, 61571059) 资助的课题.

† 通信作者. E-mail: wj2718@163.com

网络的构造模型分为两种,一种是个体为中心的模型,一种是小组为中心的社团模型,并在小规模网络上进行了实验. Yang^[11]提出了一种改变同配系数的算法,并在真实网络中进行了实验,但是通用性不强. Musmann^[12]通过在相等同配系数条件下不同度分布对于网络中其他参数的影响.然而这些算法都是针对小规模网络,并没有揭示同配系数和网络性能之间的理论关系.因此,本文主要在给定度分布的条件下,从理论和算法两个方面分析同配系数的影响.

本文余下的部分安排如下:第2节主要介绍同配系数的定义,并推导出同配系数的理论上限;第3节提出了两种基于度分布的确定性构造算法和基于概率分布的不确定性构造算法,并将这三种算法在三个网络(Erdős-Rényi (ER)网络、Barabási-Albert (BA)网络、Email真实网络)上进行仿真分析,比较在不同网络中三种算法的性能;第4节由基于概率分布的不确定性算法得到的同配系数增长图,计算这一过程中的聚类系数的变化,通过分析网络中微观结构的变化,加深对算法的理解,并得到同配系数与聚类系数之间的联系;第5节总结本文的研究成果.

2 同配系数的定义与分析

2.1 同配系数的定义

度分布仅仅度量节点属性,对于边的属性没有涉及.对于一条给定的边,两个端点的度值并不总是独立的.这种度-度相关性可以通过条件概率 $p(j|k)$ 来衡量,即一个度为 k 的节点在连接度为 j 时节点的概率.

对于给定度序列 $D = [d_1, d_2, \dots, d_n]^T$, 其中 d_i 表示节点 i 邻居节点的数目,对于给定的 q^k ,可以得到两个端点的额外度值的联合概率分布 $e_{j,k}$.对于无向网络, $e_{j,k}$ 是对称的,可以据此给出同配系数^[13]的定义:

$$\rho = \frac{1}{\sigma_q^2} \left[\sum_{i,j} d_i d_j e_{i,j} - \mu_q^2 \right] = \frac{E[D_i D_j] - \mu_{D_i}^2}{E[D_i^2] - \mu_{D_i}^2}, \quad (1)$$

其中 σ_q 代表每条边两端端点的度值的方差; d_i, d_j 分别代表一条边两端节点的度值; μ_q 代表单侧节点

度值的均值;等式第二个式子是第一个式子的向量表示形式.该同配系数的定义计算的是随机选择一条边,两端节点度值的皮尔逊(Pearson)相关系数,取值范围是 $[-1, 1]$.当网络为规则网络时,有 $\rho = 1$;当网络为随机网络时, $\rho = -1$;当网络为完全二分网时, $\rho = 0$.此处我们可以用矩阵来表示同配系数,有

$$E(D_i, D_j) = \frac{\mathbf{d}^T \mathbf{A} \mathbf{d}}{2L}, \quad (2)$$

其中 \mathbf{A} 为网络的邻接矩阵, \mathbf{d} 为度值向量, L 为网络中边的数量.为了规范化表达式,引入

$$N_k = \mathbf{u}^T \mathbf{A}^k \mathbf{u} \quad (3)$$

表示网络中长度为 k 的路径数量(不是自回避路径,同一个节点可以在路径中出现多次),其中 \mathbf{u} 是元素全为1的向量,那么有

$$\mu_{D_i} = \frac{\mathbf{d}^T \mathbf{d}}{2L} = \frac{N_2}{N_1}, \quad (4)$$

其中 \mathbf{d} 为度序列向量,以及

$$E(D_i, D_j) = \frac{\mathbf{d}^T \mathbf{A} \mathbf{d}}{2L} = \frac{N_3}{N_1}. \quad (5)$$

同时有

$$\begin{aligned} \sigma_{D_i}^2 &= E[D_i^2] - \mu_{D_i}^2 = \frac{1}{2L} \sum_i d_i^3 - \frac{N_2^2}{N_1^2} \\ &= \frac{1}{N_1} \sum_i d_i^3 - \frac{N_2^2}{N_1^2}, \end{aligned} \quad (6)$$

那么可以得到

$$\rho(D_i, D_j) = \frac{N_1 N_3 - N_2^2}{N_1 \sum_i d_i^3 - N_2^2}. \quad (7)$$

(7)式为同配系数的矩阵向量表示形式,其中 N_k 为网络中长度为 k 的路径数量,可以分别用邻接矩阵 \mathbf{A} 和度向量 \mathbf{d} 来表示, d_i 代表节点 i 的度值.

2.2 指定度分布条件下的上限分析

谱半径的定义是图 G 的邻接矩阵 \mathbf{A} 的最大特征值.为了深入理解谱半径的意义,我们找出谱半径与同配系数间的关系^[14].

为了得到 ρ 与邻接矩阵最大特征值 λ_1 的关系,首先有不等式

$$\lambda_1 \geq \left(\frac{N_{2k}}{N} \right)^{1/(2k)} \geq \left(\frac{N_k}{N} \right)^{1/k}. \quad (8)$$

证明如下.

因为正交矩阵 \mathbf{A} 是对称矩阵; 那么可以找到 \mathbf{A} 的一组规范化正交量 u_1, u_2, \dots, u_N 使得 $u_k = a_1 u_1 + a_2 u_2 + \dots + a_N u_N$

$$\left(\frac{N_{2k}}{N}\right)^{1/2k} = \left(\frac{\lambda_1^{2k} a_1 + \lambda_2^{2k} a_2 + \dots + \lambda_N^{2k} a_N}{N}\right)^{1/(2k)} \quad (9)$$

而 λ_1 是 \mathbf{A} 的最大特征值, 且任取 $|u| = \sum_i a_i^2 = N$, 那么

$$\left(\frac{\lambda_1^{2k} a_1 + \lambda_2^{2k} a_2 + \dots + \lambda_N^{2k} a_N}{N}\right)^{1/(2k)} \leq \left(\frac{N \lambda_1^{2k}}{N}\right)^{1/(2k)} = \lambda_1 \quad (10)$$

接下来证明

$$\left(\frac{N_{2k}}{N}\right)^{1/(2k)} \geq \left(\frac{N_k}{N}\right)^{1/k} \quad (11)$$

等式两边同时取 $2k$ 次幂, 两边同时乘以 N^2 , 不等式即变为

$$(a_1^2 + a_2^2 + \dots + a_N^2) \times ((a_1 \lambda_1^k)^2 + (a_2 \lambda_2^k)^2 + \dots + (a_N \lambda_N^k)^2) \leq (\lambda_1^k a_1^2 + \lambda_2^k a_2^2 + \dots + \lambda_N^k a_N^2)^2 \quad (12)$$

利用柯西-施瓦茨不等式显然得证, 那么利用 λ_1 与 N_k 的关系可以得到

$$\lambda_1 \geq \frac{N_1}{N} = \frac{2L}{N} = \mathbf{E}[D], \quad (13)$$

$$\lambda_1 \geq \sqrt{\frac{N_2}{N}} = \sqrt{\frac{\sum_k d_k^2}{kN}} = \frac{2L}{N} \sqrt{1 + \frac{\text{Var}[D]}{(\mathbf{E}[D])^2}}, \quad (14)$$

$$\lambda_1 \geq \left(\frac{N_3}{N}\right)^{1/3} = \left[\rho_D \left(\frac{\sum_i d_i^3}{N} - \frac{N_2^2}{NN_1}\right) + \frac{N_2^2}{NN_1}\right]^{1/3} \quad (15)$$

$n_i = n_i - 1 (n_{\text{MAX}} = i).$

通过上述公式推导我们得到了同配系数与谱半径之间的关系, 同时也可以得到同配系数的上限, 但是这个上限并不能真实地反映网络中的变化, 并且一个不等式的关系也未能给出精确上限. 因此我们希望找到一种新的方法来在真实网络中得到较大的同配系数. 目前在给定度分布的条件下

增加同配系数的算法主要是贪婪算法, 如图 1 所示. 节点 u_1 的度值为 3, 节点 v_1 的度值为 3, 节点 u_2 的度值为 3, 节点 v_2 的度值为 2. 在重连之前同配系数的贡献值为 $3 \times 3 + 2 \times 2 = 13$; 重连之后同配系数的贡献值为 $3 \times 2 + 3 \times 2 = 12$, 也就是说重连之后同配系数下降了. 通过改变边的连接关系可以构造同配系数不同的网络, 这些网络的度序列、节点数、边数等都相同. 这里还需要判定互换后的节点之间不能存在边, 即不允许有重边的存在. 由 (1) 和 (2) 式可知, 当如图 1 改变节点间连接时, 只有 \mathbf{A} 发生了变化, \mathbf{A} 是网络的邻接矩阵. 因此当给定度序列时, 可以通过两条边重连的方式, 在不改变度序列的前提下, 改变同配系数, 这是因为重连改变的只是 $ije_{i,j}$.

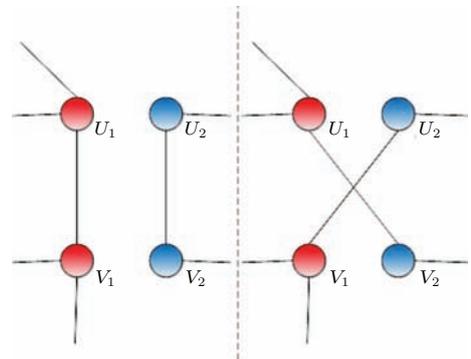


图 1 边重连过程

Fig. 1. The process of link dynamics.

这一算法可以提高同配系数, 但是离真实上限仍有一段差距, 因此我们采用新的思路来改进原有算法. 本文中提出了两种新算法: 一种是基于度分布的确定性算法, 可以构造出一个确定的同配网络; 一种是在贪婪算法的基础上基于概率分布的不确定性构造算法, 构造出一个随机网络并有效提高同配系数. 下一节中将在三种网络中比较三种算法的性能.

3 可变同配系数的算法构造

3.1 基于度分布的确定性构造算法

对于同配网络, 网络中有较大度值的节点倾向于彼此相连, 因此考虑到若给定度分布, 我们选择网络中度值高的节点与度值高的节点优先连接, 在这一原则下可以构造出一个确定的网络, 并计算这个网络的同配系数, 在这一过程中必须避免重边和

自边的产生. 对于给定度分布的条件下, 结合可图序列构造无向图的方法, 构造出一个同配网络.

定义 1 设 $N_k = \{n_1, n_2, \dots, n_{k-1}, n_k$ 为非负整数序列, 为了说明方便, 不如设任意 i, j , 若有 $i < j$, 则 $n_i \leq n_j$, 那么一个正整数序列若为可图序列必须首先满足如下条件:

$$n_k \leq k - 1 \quad \text{and} \quad \sum_{i=1}^k n_i = 2h, \quad h \in N. \quad (16)$$

定义 2 分配-减点法, 对序列 $N_k = \{n_1, n_2, \dots, n_{k-1}, n_k\}$ 进行分配减点, 令 $n_1 = 0, n_k = n_k - 1, n_{k-1} = n_{k-1} - 1, \dots, n_{k-n_1} = n_{k-n_1} - 1, n_{k-n_1+1} = n_{k-n_1+1} - 1$, 然后对 N_k 从大到小重新排列并记为 $N_k^1 = \{n'_1, n'_2, \dots, n'_k\}$, 则 N_k^1 为 N_k 的分配减点序列, 更一般的 N_k^i 为 N_k 的 i 次分配减点序列.

引理 正整数序列 $N_k = \{n_1, n_2, \dots, n_{k-1}, n_k\}$ 与其任意次分配减点序列有相同的可图性质. 则设 N_k^i 为 N_k 的 k 次分配减点序列, 那么 N_k 为可图序列的充要条件是 N_k^i 是可图序列.

因此对于确定性构造算法, 一定要保证构造出的图没有自边和重边. 我们将上面阐述的分配减点法进行变换即可得到一个新的构造算法. 首先, 计算网络中所有节点的度值并从大到小排序, 将度值最大的节点的度值 k 置为 0, 并依次在最大度值节点之后的 k 个节点的度值上分别减 1, 这一过程对应着最大度值节点与其他度值较大的节点分别相连. 然后再依据现有度值进行重排并重复上述过程, 最终当节点对应的所有度值均为 0 即可结束, 即无向图构造完成, 算法表示如表 1.

表 1 确定构造算法
Table 1. Construction Algorithm.

Input: degree $N_k = [n_1, n_2, n_3, \dots, n_k]$	
Output: $\{G', N_k^{i+1} = \{n'_1, n'_2, \dots, n'_k\}\}$	
1	Input degree $N_k = [n_1, n_2, n_3, \dots, n_k]$
2	sort degree N_k
3	$n_{\text{MAX}} = 0$
4	$n_i = n_i - 1 (n_{\text{MAX}} = i)$
5	go to step 2 until $n_k == 0$
6	calculate ρ
7	return $\{G', \rho\}$

这一构造过程可以避免重边和自边的发生, 因为我们输入的度分布均为已有的无向图的度分布,

因此均为可图序列, 则必定可以构造出一个同配系数较高的网络.

3.2 基于贪婪算法的不确定性构造算法

贪婪算法 [9] 是在给定度分布的条件下为提高同配系数而设计的一种算法. 其基本原理如图 1 所示, 在边重连的过程中并不改变网络的度分布. 其基本过程如表 2.

表 2 贪婪构造算法
Table 2. Greedy algorithm for generating network.

Input: $\{G, A\}$	
Output: $\{\rho(\rho_0, \rho_1, \rho_2, \dots, \rho_i), G'\}$	
1	Input $\{G, A\}$
2	edges(l), nodes (k)
3	randomly choose nodes a, c
4	randomly choose b, d linked with a, c
5	if $(a - d)(c - b) > 0$ exchange edges
6	calculate ρ_i
7	obtain and ρ_i go to 6 for i times
8	return $\{G', \rho(\rho_0, \rho_1, \rho_2, \dots, \rho_i)\}$

贪婪算法的特点就在于分别随机选择节点并在节点的邻接节点中也随机选取构成边, 这一过程中构成出的网络必然是随机的, 但是这也是贪婪算法无法得到较大的同配系数的原因, 因为算法的随机性较大, 当互换过一定数量的边时, 选择出可互换边的概率变得极低, 导致同配系数收敛, 这一现象在本节中的仿真分析中呈现. 这里可以给出一些公式上的证明: 在每一个重连过程中, 我们选择边 $u_1 \sim v_1$ 和 $u_2 \sim v_2$. 对于这四个点 u_1, v_1, u_2, v_2, ρ 在数值上的改变为

$$\begin{aligned} \Delta\rho = & [d(u_1)d(v_1) + d(u_2)d(v_2) - d(u_1)d(v_2) \\ & - d(u_2)d(v_1)] \times \{E[D_i^2] - \mu_{D_i}^2\}^{-1} \\ = & \frac{[d(u_1) - d(u_2)][d(v_1) - d(v_2)]}{l}, \end{aligned} \quad (17)$$

其中 $d(u_i)$ 或 $d(v_i)$ 代表 u_i 或 v_i 节点的度值, $E[D_i^2]$ 为节点度值平方的均值, μ_{D_i} 为节点的度值总和. $E[D_i^2] - \mu_{D_i}^2$ 没有改变, 我们用 l 来表示. 忽略两条边的顺序, 如果 $d(u_1) \geq d(u_2)$ 和 $d(v_1) \geq d(v_2)$, 那么同配系数就会增加. 我们可以构建网络来测试贪婪算法. 目前有两种主要的构建模型, 一个是 ER

随机网络其中节点以独立的概率 P 相互连接, 以此我们可以计算选中的两条边满足互换条件的概率:

$$\begin{aligned} & \mathbb{P}(\Delta\rho \geq 0) \\ &= 2 \left(\sum_{i=0}^{\infty} \sum_{j=0}^i q_i q_j p \right)^2 \\ &= 2 \left(\sum_{i=0}^{\infty} \sum_{j=0}^i (n-2i)(n-2j) p^{i+j+1} \right. \\ & \quad \left. \times (1-p)^{2n-4-(i+j)} \right)^2, \end{aligned} \quad (18)$$

其中 $\binom{n-2}{i}$ 代表排列组合即 C_{n-2}^i . 另一种是 BA 网络生成模型, 其度分布为 $p_k = k^{-a}/\zeta(a)$, $q_k = k^{-a+1}/\zeta(a-1)$, 因此选中可互换边的概率为

$$\begin{aligned} & \mathbb{P}(\Delta\rho \geq 0) \\ &= \left(\sum_{i=0}^{\infty} \sum_{j=0}^i q_i q_j \frac{ij}{2m} \right)^2 \\ &= \left(\frac{1}{l\zeta(a-1)} \sum_{i=0}^{\infty} \sum_{j=0}^i (ij)^{-a+2} \right)^2, \end{aligned} \quad (19)$$

其中 ζ 是 Riemann zeta 函数.

从 (18) 和 (19) 式可以看出, 当我们使用贪婪算法来实现同配系数增加时, 得到可互换边即 $\Delta\rho > 0$ 的概率十分小, 特别是在完成了部分边互换以后. 由此我们可以得出 ρ 会收敛更快并且拥有更短的跨度. 这是因为更多的边互换以后, 我们通过随机选边可能无法选中正确的可互换边. 这一过程表现为同配系数保持不变, 我们将会在之后的仿真中展现这一过程.

由此, 我们将设计随机构造算法, 它必须比贪婪算法拥有更大的同配系数跨度, 并且产生的图必须保证随机性. 也就是说我们可以通过同样的度分布和同样的同配系数得到一系列不同的图, 我们可以利用这些图进行深入的研究.

3.3 基于概率分布的不确定性构造算法

对于一个同配网络, 拥有较高度值的节点倾向于彼此连接. 因此如果我们想提高同配系数, 就应当选择有较高度值的节点与其他有较高度值的节点彼此相连.

首先计算每个节点的度值 $N_k = [n_1, n_2, n_3, \dots, n_k]$. 我们可以利用 $p_i = n_i / \sum N_k$ 来计

算每个节点的概率分布密度. 在基于概率分布的随机构造算法中依据概率 P_k 来选择节点 a, c , 因此有较高度值的节点被选择的概率较大. 然后在选中节点的邻接节点中随机选择两个节点 b, d 来组成两条边. 我们将两条边 $a-b, c-d$, 改变为 $a-c, b-d$, 这会有更大概率满足 $\Delta\rho \geq 0$. 在这一过程中应该避免重边和自边. 我们选择有较高度值的节点 a, c 彼此连接, 原因在于这可以明显得出只有 $E(D_i, D_j)$ 发生了变化, 通过边重连在 (1) 式中, 希望通过将两个度值较大的节点彼此相连得到较大的 $E(D_i, D_j)$, 这一算法有较大的概率选中可以使得同配系数增加的两条边.

对于任何给定的概率分布函数, 其累计分布函数满足均匀分布. 由此我们设计算法如表 3.

表 3 随机构造算法

Table 3. Random algorithm for generating network.

Input: $\{G, A\}$	
Output: $\{\rho(\rho_0, \rho_1, \rho_2, \dots, \rho_i), G'\}$	
1	Input $\{G, A\}$
2	edges(l), nodes (k)
3	degree $N_k = [n_1, n_2, n_3, \dots, n_k]$
4	$p_k = [p_1, p_2, p_3, \dots, p_k]$
5	calculate ρ_0
6	use distr. rvs choose a, c refer to p_k
7	obtain b, d linked with a, c
8	if $(a-d)(c-b) > 0$ exchange edges
9	obtainand ρ_0 go to 6 for i times
10	return $\{G', \rho(\rho_0, \rho_1, \rho_2, \dots, \rho_i)\}$

3.4 仿真实验与分析

本文仿真采用的软件是基于 python 语言下的 Spyder 平台, Spyder 具有较快的计算速度. 我们主要用 python 中的 Networkx 模块来生成并计算网络中的参量, 并使用 Origin 绘制图表. 对于 ER 网络, 我们采用 Networkx 模块中的生成函数 `nx.random_graphs.erdos_renyi_graph(n, p)`, 并设定参量 (包括节点数和节点间连接概率), 可以直接生成 ER 网络. BA 网络也采用 Networkx 模块中自带的生成函数 `nx.random_graphs.barabasi_albert_graph(n, m)`, 设定参量 (节点数和参量 P), 生成网络. Email 网络通过网络爬虫得到真实通信网络的拓扑信息.

对于上面提到的三种算法, 分别用python实现其对网络中节点以及边的操作, 利用Networkx模块中的网络参数函数计算每一步的同配系数变化, 因此我们只需要导入已经生成好的网络即可对网络进行操作.

对于贪婪算法, 我们实现的是随机选取边, 判断是否满足互换条件(同配系数是否增加), 若满足然后在网络中的边进行互换. 对于基于概率分布的不确定性构造算法, 我们在选择边的时候是依据度分布的概率密度, 然后再判断同配系数是否增加, 若增加对网络中的边进行交换. 而确定性构造算法, 是先计算整个网络的度分布, 然后重新构造出一个确定的网络, 最终计算网络的同配系数.

我们发现对于贪婪算法和基于概率分布的不确定性构造算法, 当网络规模较大时, 每次计算同配系数将会十分繁琐. 为了提高计算速度, 使其在大规模节点下可以观测同配系数, 在实验方法上从两个方面进行设计:

1) 在数据结构的存储中, 对于每个生成的网络只保存其邻接列表, 因为网络为无向图, 因而每条边只保存一次, 当需要进行仿真时, 直接读取列表文件即可;

2) 通过分析(1)式可以发现, 在算法改变网络的过程中只有 d_i, d_j 是变化的, 而 μ_q^2, σ_q^2 只和度序列相关, 在互换过程中是不变的, 因此在每次互换完成后, 只需要计算表达式中 $\sum_{i,j} ije_{i,j}$ 变化量即可, 同配系数的变化也只与此变化量有关.

我们分别在ER网络、BA网络和真实Email网络上测试三种算法的性能, ER网络中有30000个节点, 每个节点间相互连接的概率是0.0005, 并且有230000条边在ER网络中. BA网络有50000个节点, BA网络的参数 P 为1. 对于真实Email网络, 有265214个节点和364481条边. 我们选择真实网络来测试所提算法是为了证明我们的算法可以处理大型网络中的大规模数据. 图2—图4中的X轴代表迭代次数, Y轴代表同配系数. 这意味着如果指数在相同的迭代次数下上升得越高, 算法的性能越好. 图中GA代表贪婪算法, SSA代表基于概率分布的不确定性构造算法, SA代表确定性构造算法.

为了保证实验的可靠性, 在三个网络中测试三种算法时, 每组均进行50次实验, 然后取各次实

验的平均值再进行绘图, 以防止单次实验造成的误差.

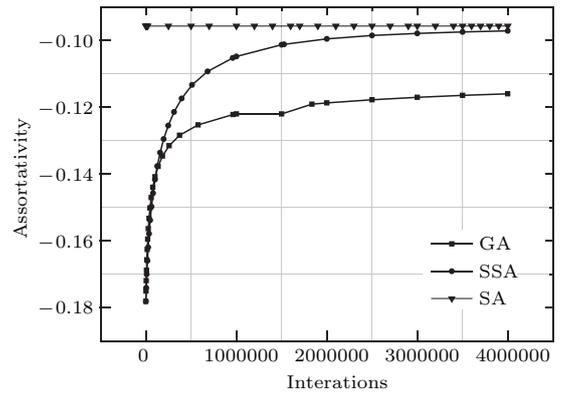


图2 真实网络同配系数算法仿真

Fig. 2. Simulation with the assortativity coefficient in the real network.

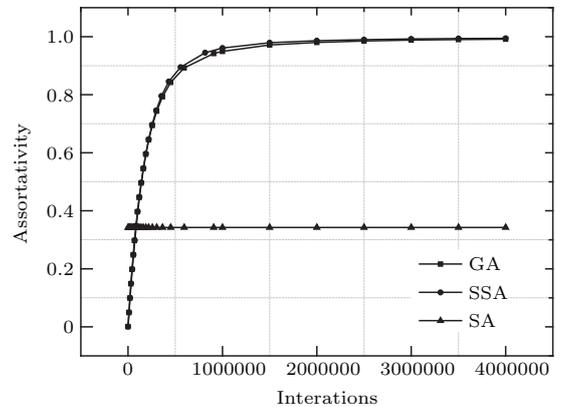


图3 ER网络同配系数算法仿真

Fig. 3. Simulation with the assortativity coefficient in ER network.

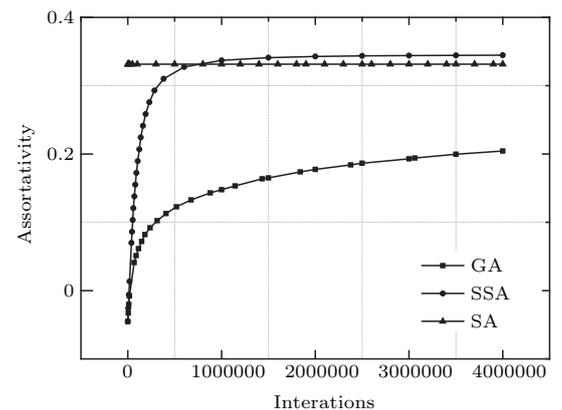


图4 BA网络同配系数算法仿真

Fig. 4. Simulation with the assortativity coefficient in BA network.

从图2真实网络中可以看到确定性构造算法取得最大的同配系数, 概率分布的不确定算法在

经过 4000000 次迭代之后逐渐上升到 -0.097 . 由此可以看出 Email 网络是一个异配网络, 因为经常给别人写邮件的人也会给一些不爱给别人写邮件的人发邮件. Email 网络拥有大量的度值很小的节点, 可以看作一个稀疏网络, 这会表现出稀疏网络的特性. 而贪婪算法上升缓慢并且最终的同配系数要小于另外两种算法, 其对应的同配系数的跨度也相对较小. 因此确定性算法和基于概率分布的不确定性构造算法在真实网络中有一定的价值.

图 3 为在 ER 网络中进行的测试. 两种不确定性算法均在 4000000 次迭代以后提升同配系数到接近于 1, 因为 ER 网络中的每个节点的地位相同并且有相同的概率彼此连接. 通过边重连以后它会变得规则而又同配, 发现增加同配系数的算法使得网络变得更加规律. 在 ER 网络使用概率选点的方法比贪婪算法性能要优越, 但是确定性算法由于构造过程中导致的节点连边的冲突使其同配系数只能达到 0.34, 具有一定的局限性. 因此对于 ER 网络确定性算法只是提供了一种构造同配网络的一种方法. 贪婪算法和概率分布算法都可以实现同配系数的大幅度变化增长.

图 4 为三种算法在 BA 网络中的性能测试结果. 对于 BA 网络, 概率分布构造算法明显优越于贪婪算法. 概率分布构造算法最终到达 0.34 而贪婪算法只达到 0.21, 并且概率分布构造算法上升得更快. 在 BA 网络中不能如同 ER 网络一样使得同配系数上升至 1 的原因在于 BA 网络中的节点不是等同的. 我们无法通过边重连的形式使得网络完全规则, 它拥有上限并且从 (13), (14), (15) 式可以得到证明. 但是使用构造算法能够达到更高的同配系数, 在 BA 网络中, 确定性构造算法依然可以达到很高的同配系数, 结果显示优于贪婪算法, 这是因为分配减点的策略更加适合度分布比较差的网络, 能够有效地降低边重新分配带来的冲突.

通过上面的仿真图可以发现, 概率分布的不确定性算法明显要优于贪婪算法, 并且概率分布算法同时拥有随机性可以保证结果的实用性. 而确定性算法在某些网络中比如 Email 网络拥有其他两种算法所没有的优势, 可以达到更高的同配系数, 同时其在 BA 网络中的表现依然优于贪婪算法.

4 可变同配系数下的网络拓扑参数变化分析

Thechanamoorthy 等 [15] 定义并分析了无向图中节点的同配性的计算, 发现所有的真实网络都表现出一定的异配性, 他们强调中心的同配性与网络的鲁棒性之间的联系. Xin 等 [16] 提出了一个数学模型, 可以使用不同的函数来衡量同配系数的影响, 他们扩展了同配系数并利用网络中其他参数来发现隐藏的社团. Dwivedi [17] 研究了遗传算法导致的网络中异配结构的改变, 并找到真实网络中存在大量的异配结构的原因, 通过分析证实, 在一些特定情况下无标度网络比相应的随机网络更加稳定. 因此本节对聚类系数和微观结构 (巨分支等) 的影响进行分析.

4.1 聚类系数的变化

聚类系数是度量某顶点的两个邻居顶点也互为邻居的平均概率. 实际上, 该参量度量的是网络中三角形结构的密度. 大多数数值的数量级都是 10%, 即一个顶点的两个邻居顶点相互之间也是邻居顶点的概率在 10%—60% 之间. 可以将聚类系数的计算式表示如下:

$$C = \frac{1}{n} \frac{[\langle k^2 \rangle - \langle k \rangle]^2}{\langle k \rangle^3}, \quad (20)$$

其中 n 为网络中节点数, $\langle k \rangle$ 代表网络中的平均度, $\langle k^2 \rangle$ 代表网络中节点度值平方的均值. 网络中 $\langle k^2 \rangle$ 和 $\langle k \rangle$ 都有固定的有限值, 当 n 趋于无穷时, 上式中的值变小, 因此可以推断对于大规模网络, 聚类系数会很小.

我们可以推导同配系数的公式来得到同配系数与聚类系数之间的联系.

$$\begin{aligned} \rho &= \frac{\langle k|A|k \rangle - \frac{1}{2m} (\langle 1|E|1 \rangle)^2}{\langle 1|E^2|1 \rangle - \langle k|A|k \rangle - \frac{1}{2m} (\langle 1|E|1 \rangle)^2} \\ &= \frac{|P_2|(|P_{3/2}| + C - |P_{2/1}|)}{3|S_{1,3}| - |P_2|(|P_{2/1}| - 1)}, \end{aligned} \quad (21)$$

其中 \mathbf{k} 为全 1 向量, $\mathbf{E} = \nabla^T \nabla$, 为 $n \times m$ 的矩阵; n 为网络中节点数 m 为边数, \mathbf{E} 代表边与边之间的相

连矩阵, 即若边 i 和边 j 相连那么对应1 否则为0;

$$|P_1| = \frac{1}{2} \sum_{i=1}^n k_i, |P_2| = \sum_{i=1}^n \binom{k_i}{2}, |C_3| = \frac{1}{6} \text{tr}(A^3),$$

$$|P_3| = \sum_{i,j \in E} (k_i - 1)(k_j - 1) - 3|C_3|,$$

$$\text{和 } |S_{1,3}| = \sum_{i=1}^n \binom{k_i}{3}, \text{ 其中的表示法 } |P_{r/s}| = |P_r|/|P_s|.$$

因此在公式上同配系数与聚类系数有一定联系, 但在不同网络中, 它们并不成一定的比例联系, 需要通过仿真来发现其中内在的变化.

通过上一节的仿真结果, 我们得到了一系列具有相同度分布但不同同配系数的网络. 可以观察在同配系数变化时网络中其他参数的变化, 以此来分析同配系数对网络的影响, 同时也可以得到概率分布构造算法对网络中微观结构的影响. 我们可以计算在 Email 网络中当同配系数增加时聚类系数的变化. 均匀选择十个节点, 当同配系数增加时输出 10 个图, 然后计算 10 个图的聚类系数. 为了防止出现误差, 重复上一步骤 5 次, 发现每次的变化都十分相似, 并取平均值.

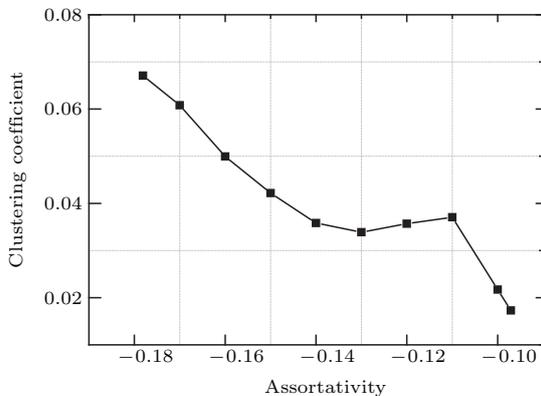


图5 同配系数的变化

Fig. 5. Variation of the assortativity coefficient.

整体上, 聚类系数随着同配系数的增加而减少. 现在, 我们研究同配系数增加时聚类系数减小的原因. 聚类系数是衡量网络中形成一个环形结构的概率, 有较高的同配系数意味着有相同度的节点相互连接的概率较大. 在真实网络中, 有许多的节点度值十分小, 通过重连以后这些节点彼此相连从而形成环形结构的概率变低. 聚类系数是网络的统计特性, 仅仅分析聚类系数的走向, 并不能很好地

分析同配系数对网络性能的影响机理, 因此我们需要选定微观结构来得到更加准确的结论.

4.2 微观影响分析

为了了解不确定性构造算法对于网络内部微观结构的影响, 我们计算了网络中的巨分支的节点数变化以及网络中分支的数目随着同配系数的变化, 如图 6 和图 7 所示.

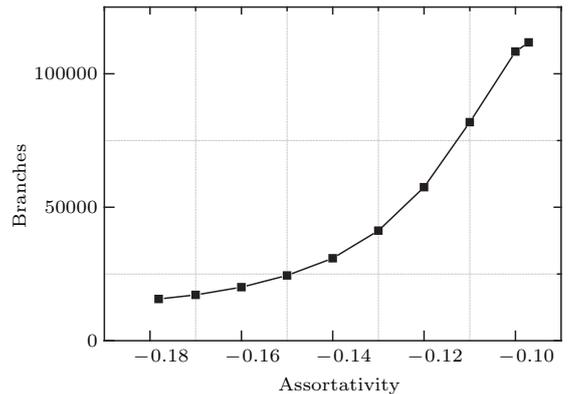


图6 分支数目变化

Fig. 6. The hanges in the number of branches.

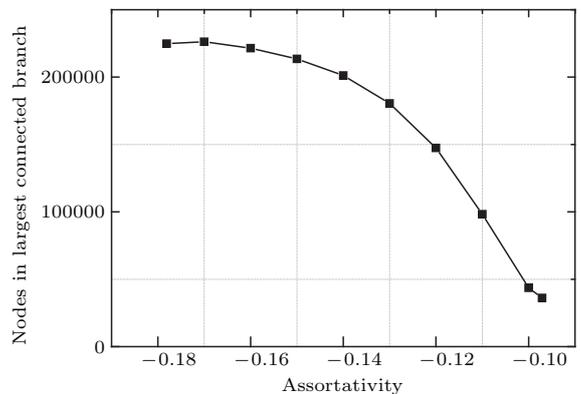


图7 巨分支中节点数变化

Fig. 7. Variation of the number of nodes in the giant branch.

从图 6 和图 7 中可以得到随着网络中的同配系数不断增加, 网络中的分支数不断增加, 巨分支中的节点数不断减小. 这说明网络在同配系数增加的过程中逐渐变得破碎, 我们得到除去巨分支外的小分支中的节点数一般为两个, 这意味着那些原来的度值为 1 的节点经过概率分布算法之后彼此相连. 巨分支中的节点数逐渐减小, 这一过程是将巨分支中度值较低的节点 (一般为 1) 逐渐分离破碎为小分支, 而巨分支不断变得聚合并且十分紧密. 还有一

个值得注意的现象是巨分支中减少的节点数大约是小分支增加的数目的2倍,这也印证了其破碎为两个度值为1的节点彼此相连成为小分支这一过程.因此这一微观过程也说明了图5中聚类系数整体趋向于变小的原因,因为网络逐渐破碎成线形节点连接,因此形成环状结构的概率就变低了,聚类系数因此变小.我们也注意到在图5中有两步聚类系数有小范围上升,这也对应于相应步骤的巨分支分离可能形成了一定数量的三角环状结构,注意到每一步的巨分支节点减少数目并不正好是小分支增加数目的两倍,因此聚类系数有小范围上升.

通过这一节的仿真实验,我们更加清楚了不确定性构造算法对于网络的微观影响,这也让我们全面认识网络以及同配系数的特性,同时也了解到不确定性构造算法可以使巨分支不断聚合,这也是一种构造密集同配网络的方法.

5 总 结

本文主要研究给定度分布条件下同配系数变化的影响.在贪婪算法的基础上,提出两种新的同配系数构造算法,一种是确定性算法,仿真验证了其在BA网络和Email网络中的有效性,另一种是基于概率分布的不确定性构造算法,并仿真验证了其性能优于贪婪算法.在同配系数变化的同时,观察网络中其他参数的变化,从宏观角度分析了聚类系数的变化,并以此从微观角度了解了网络中巨分

支以及小分支的变化,在微观变化中理解了宏观上聚类系数变化的原因,也加深了对不确定性构造算法对网络微观结构影响的理解.

参考文献

- [1] Newman M E J 2007 *Phys. Rev. E* **76** 045101
- [2] Juyong P, Newman M E J 2003 *Phys. Rev. E* **68** 026112
- [3] Wu Zh X, Guan J Y, Wang Y H, Feng C F 2010 *Chin. Phys. B* **19** 060203
- [4] Quintana R, Kopcow L, Marconi G, Sueldo C, Speranza G, Baraño R I 2001 *Human. Rep.* **16** 1814
- [5] Schwarte N, Cohen R, Ben-Avraham D, Barabási A L, Havlin S 2002 *Phys. Rev. E* **66** 015104
- [6] Dorogovtsev S N, Mendes J F F, Samukhin A N 2001 *Phys. Rev. E* **64** 1107
- [7] Newman M E J, Strogatz S H, D J 2001 *Phys. Rev. E* **64** 359
- [8] Litvak N 2013 *Phys. Rev. E* **87** 522
- [9] Winterbach W, Ridder D D, Wang H J, Reinders M, Mieghem P V 2012 *Europ. Phys. J. B* **85** 151
- [10] Zhang D Y, Bi-Feia H E, Chen P 2013 *Complex Systems & Complexity Science* **10** 45
- [11] Yang R 2014 *Acm Southeast Regional Conference* **1** 5
- [12] Mussmann S, Moore J 2015 *Proceedings of the 29th AAAI Conference on Artificial Intelligence* **238** 25
- [13] Newman M E J 2002 *Phys. Rev. Lett.* **89** 111
- [14] Mieghem P V 2011 *Performance Analysis of Communications Networks and Systems* (Cambridge: Cambridge University Press) p20
- [15] Thedchanamoorthy G, Piraveenan M, Kasthuriratna D, Senanayake U 2014 *Proc. Comput. Sci.* **29** 2449
- [16] Xin L, Tsuyoshi M, Ken W 2014 *Phys. Rev. E* **90** 012806
- [17] Dwivedi S 2015 *Phys. Rev. E* **92** 022802

Algorithm design and influence analysis of assortativity changing in given degree distribution*

Li Jing¹⁾²⁾ Zhang Hong-Xin¹⁾³⁾ Wang Xiao-Juan^{1)†} Jin Lei¹⁾

1) (*Institute of Electrical Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*)

2) (*China Information Technology Security Evaluation Center, Beijing 100085, China*)

3) (*Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications, Beijing 100876, China*)

(Received 13 December 2015; revised manuscript received 19 January 2016)

Abstract

Complex network is the abstract topology of a large number of nodes and edges in reality. How to reveal the influences of internal network topology on network connectivity and vulnerability characteristics is a hotspot of current research. In this paper, we analyze the influence of assortativity according to Newman's definition of assortativity in a given degree distribution. To fully understand the influence of assortativity we should change the assortativity to see how the topology of network changes. But we find the existing greedy algorithm cannot improve assortativity effectively. First we put forward a deterministic algorithm based on degree distribution and an uncertain algorithm based on probability distribution to increase assortativity. The deterministic algorithm can create a certain network which has a large assortativity without changing node degree. The uncertain algorithm can increase the assortativity continuously by changing the connection of edges. And the uncertain algorithm creates different graphs each time, so the result of the algorithm is uncertain. Then we test our algorithms on three networks (ER network, BA network, Email network) and compare with greedy algorithm, and the experimental results show that the uncertain algorithm performs better than greedy algorithm in three networks which have a large span of assortativity. And our deterministic algorithm performs well in a real world network. We find that we can increase assortativity coefficient up to 1 in ER network. This is because nodes in the ER network are peer to peer. We can also show that that the assortativity cannot increase up to 1 in some networks because nodes in these networks are not in the same status. Because we obtain a large span of assortativity, we can fully understand the change of network topology. On this basis, we analyze the changes of clustering coefficient when using the uncertain algorithm based on a probability distribution to increase the assortativity. We find that there is a certain correlation between assortativity and clustering. And we study the micro influence of uncertain algorithm on network, by which the reason of the change of clustering coefficient is explained. We calculate the changes of giant branches and small branches. The changes of the number of nodes in giant branches and the number of small branches show that the scale of giant branches becomes smaller, which means that the connection between nodes in giant branches becomes closer. The increase of the number of small branches means that the network as a whole becomes more fragile. So we can show how the uncertain algorithm changes the topology of the network without changing the degree of nodes in the network. Then we can use this algorithm to change the network to obtain a larger span of assortativity for further study.

Keywords: assortativity, degree distribution, structural algorithm, clustering

PACS: 45.70.Vn, 42.65.Sf

DOI: 10.7498/aps.65.094503

* Project supported by the National Natural Science Foundation of China (Grant Nos. 61571063, 61472357, 61501100, 61571059).

† Corresponding author. E-mail: wj2718@163.com