

基于社团结构的负载传输优化策略研究^{*}

邵 斐^{1)2)†} 蒋国平¹⁾

1) (南京邮电大学控制与智能技术研究中心, 南京 210003)

2) (金陵科技学院信息技术学院, 南京 211169)

(2010 年 9 月 27 日收到; 2011 年 2 月 25 日收到修改稿)

研究表明网络社团结构特征对负载传输有影响, 明显社团结构特征会降低网络的承载能力. 由于最短路由策略在选择路由时有一定的随机性, 本文提出了一种基于社团结构的负载传输策略, 减少最短路由经过的社团数量, 从而降低社团边缘节点的介数. 实验结果显示, 该策略在保证最短路由小世界特性的同时, 提升了网络的承载能力, 社团划分得越准确传输优化策略效果越显著.

关键词: 优化路由策略, 社团结构, 复杂网络, 负载传输

PACS: 89.75.Da, 89.75.Fb, 89.40.-a

1. 引言

现代通信业的发展速度很快, 移动通信网、因特网等诸多大规模复杂网络的相继出现, 网络上的负载越来越多, 现有网络的承载能力受到了极大的挑战, 如何提高网络承载能力受到了研究人员的普遍关注. 人们提出了各种负载传输策略, 主要分为两大类, 一类是基于大尺度全局信息的传输策略, 例如基于跳数和最小的最短路由策略^[1], 基于节点度数和最小的有效路由策略^[2], 基于节点排队长度和最小的全局动态路由策略^[3], 基于节点度数连乘积最小的优化路由策略^[4]等; 第二类是基于小尺度局部信息的传输策略, 例如随机行走和队列长度结合起来的自适应路由策略^[5], 以一定的概率将负载发送给邻居的最近邻策略^[6]和次近邻策略^[7], 基于蚁群算法的局域信息素浓度路由策略^[8]等; 当然也有这两者的有机结合, 比如基于局域路由和全局最短路径的混合导航策略^[9]. 现实网络的边可能会有不同的权重, 因此加权网络上的路由研究也受到人们的重视^[10,11].

随着对复杂网络结构复杂性及其与网络行为之间关系的深入研究, 人们发现现实网络在中尺度

上显示出所谓“物以类聚, 人以群分”的特性^[12–17], 即社团结构. Danon 等人研究了网络内在结构对信息传输的影响^[18], 指出网络的社团结构会降低最短路由策略的传输能力, 将社团结构信息提供给数据包可以提高网络的承载能力. 受其启发, 我们研究了网络中社团结构对最短路由策略和有效路由策略的影响, 给出基于社团结构的负载传输优化策略, 减少最短路由经过的社团数量, 以降低社团边缘节点的介数, 提升网络的承载能力, 同时也发现这种方法对有效路由几乎没有作用.

2. 负载传输模型

首先给出网络负载传输的简化模型: 每个时刻 t , 在网络上随机产生 R 个信息包, 每个信息包随机选择一对源节点和目标节点 (两者不相同), 如果源节点已有积压的信息包, 则将信息包放入源节点排队队列最后等待传输, 每个节点传输信息包的规则为先进先出; 对于已经产生的信息包, 路由器将为其按照预定的传输策略选择一条路径以传输到目的节点; 任意节点 i 的处理能力为 C_i , 即任意时刻 t , 节点 i 最多能够传递 C_i 个信息包出去, 在本模型中, 假设各节点的处理能力相同, 不失一般性, 令 C_i

^{*} 国家自然科学基金 (批准号: 60874091), 江苏省“六大人才高峰”高层次人才计划 (批准号: SJ209006), 高等学校博士学科点专项科研基金 (博导类) (批准号: 20103223110003) 和金陵科技学院科研基金 (批准号: JIT-N-201010) 资助的课题.

[†] E-mail: shaofei@jit.edu.cn

$=1$; 每个节点的存储能力没有限制, 可以容纳任意长的信息包等待队列; 一旦信息包到达目标节点, 则离开网络, 不再在网络上继续传输; 所有节点都兼有终端和路由器的功能, 既能产生和接收信息包, 又能转发信息包.

不同的传输策略下, 网络有不同的吞吐量, 随着单位时间内产生的信息包数量 R 的增大, 网络将由自由状态进入阻塞状态, 其间存在一个临界值 R_c , 当 $R < R_c$ 时, 产生的信息包和传输的信息包基本相同, 网络处于稳定的自由状态; 当 $R > R_c$ 时, 部分节点传输的信息包将小于其所接收的信息包, 开始出现阻塞, 且随着运行时间的增加, 阻塞规模不断增大. 我们使用有序参数 η 与 R 的关系来刻画网络的上述状态变化^[19]:

$$\eta = \lim_{t \rightarrow \infty} \frac{\langle \Delta \Theta \rangle}{R \Delta t}, \quad (1)$$

其中 $\Delta \Theta = \Theta(t + \Delta t) - \Theta(t)$, $\Theta(t)$ 是 t 时刻网络中的信息包的总数, $\langle \dots \rangle$ 表示基于时间 Δt 的平均. 当网络处在自由状态时, $\Delta \Theta \leq 0$, 所以 $\eta \leq 0$, 处在阻塞状态时, $\Delta \Theta$ 逐步增大, η 也随之逐步增大, 所以使 $\eta \leq 0$ 的最大的 R 就是临界值 R_c , R_c 越大, 网络的承载能力也就越强.

在给定传输策略下, 节点 i 的介数 b_i 定义如下:

$$b_i = \sum_{s,t} \frac{\sigma(s,i,t)}{\sigma(s,t)}, \quad (2)$$

其中 $\sigma(s,t)$ 是在指定传输策略下, 节点 s, t 之间的总路径数, $\sigma(s,i,t)$ 在指定传输策略下, 节点 s, t 之间经过节点 i 的总路径数. 在自由状态, 任意一个节点 i 收到的信息包为 $Rb_i/N(N-1)$ (N 为网络中节点总数), 它的处理能力 $C_i = 1$, 当收到的信息包大于其处理能力, 即 $Rb_i/N(N-1) > 1$ 时, 进入阻塞状态. R_c 可以用以下公式进行估计:

$$R_c = \min(N(N-1)/b_i). \quad (3)$$

不断地从一个网络中移除介数最大的边, 可以将网络分裂成一个个社团, 也就是说, 两个社团之间的边的边介数相对较大, 这些边的两端的节点 (也就是社团的边缘节点) 的介数也比较大, 受 (3) 式的启发, 如果降低这些节点介数较大的节点的 b_i 就可以增加 R_c , 从而增加网络承载能力. 由于最短路由策略在选择路由时有一定的随机性, 即如果有多条同时满足条件的路径, 从中随机选择一条, 我们考虑将满足条件的路径进行约束, 使其经过的社团数量最少, 从而减少社团边缘节点的介数, 达到增加 R_c 的目的; 而有效路由策略则不同, 其在选择

路由时已经尽量避开介数较高的节点, 所以即使对其进行约束, 效果也不会十分明显. 基于社团结构的路由策略具体如下:

1) 首先采用某种社团划分算法将网络划分为 mod 个社团, 节点 i 所在社团记为 $G(i)$, 社团 $G(i)$ 中所有节点集记为 $N(G(i))$, 节点集 $N(G(i))$ 对应原始网络的子图记为 $\text{SUB}(N(G(i)))$, 若社团 $G(i)$ 和 $G(j)$ 之间有边连接, 则 $\text{neigh}(G(i), G(j)) = 1$, 否则为 0;

2) 若 $G(i) = G(j)$, 则 $T = \text{SUB}(N(G(i)))$;

3) 若 $G(i) \neq G(j)$ 且 $\text{neigh}(G(i), G(j)) = 1$, 则 $T = \text{SUB}(N(G(i)) \cup N(G(j)))$;

4) 若 $G(i) \neq G(j)$ 且 $\text{neigh}(G(i), G(j)) = 0$, 则 $T = \text{SUB}(N(G(i)) \cup N(m_1) \cup N(m_2) \cup \dots \cup N(m_l) \cup N(G(j)))$, $m_1, m_2, \dots, m_l$ 为连接两个社团 $G(i)$ 和 $G(j)$ 且经过社团数量最少的中间社团序列;

5) 在源节点 i 和目标节点 j 对应原始网络的子图 T 上采用相应的路由策略得到路由, 若采用最短路由, 得到的就是基于社团结构的最短路由, 记为 CSHT; 若采用有效路由, 则为基于社团结构的有效路由, 记为 CEFF (采用有效路由时, 各节点的度数为其在原始网络中的度数).

为了研究划分社团结构的准确性对于路由策略的影响, 我们采用 Newman 等人提出的模块度来衡量网络社团划分的质量^[12,13]: 网络被划分成为 m 个社团, 定义一个 $m \times m$ 的对称矩阵 $E = (e_{ij})$, e_{ij} 表示第 i 个社团和第 j 个社团之间的边数占原始网络总边数的比例, 则模块度 Q 可以定义为

$$Q = \sum_i (e_{ii} - a_i^2), \quad (4)$$

其中 $a_i = \sum_j e_{ij}$, Q 值越大, 社团结构特征越明显, Q 的上限为 1, 实际网络中, Q 一般位于 0.3—0.7 之间^[20].

3. 仿真与分析

首先研究社团结构对最短路由^[1]和有效路由^[2]的影响. 基于 ER 模型生成一系列的伪随机网络^[13] (即除了社团结构, 其他诸如节点度、聚类系数等特性和完全随机网络都是一样); 其中每一个网络的节点数 n 都为 128, 把它分成 4 个社团, 每个社团的节点数均为 32 个. 相同社团内的节点以概率 p_{in} 连接, 不同社团内的节点以概率 p_{out} 连接, 通过调

节 p_{in} 和 p_{out} 的大小, 保证整个网络的节点平均度 $\langle k \rangle$ 为 16. 一旦相同社团内的节点连接概率 p_{in} 确定了, 节点和它所在社团内部其他节点的之间边数的平均值 Z_{in} 就确定了. 通过调节 Z_{in} 的大小, 可以得到不同的伪随机网络 (整个网络的节点平均度保持 $\langle k \rangle = 16$ 不变), Z_{in} 越大, 社团结构特征越明显. 在 $\text{mod} = 1$ (整个网络划分成一个社团即未划分社团) 时对不同的 Z_{in} 产生的伪随机网络分别采用 CSHT 和 CEFF 策略进行仿真, 得到如图 1 所示结果 (本文中仿真结果数值都是进行 10 次仿真后的平均值).

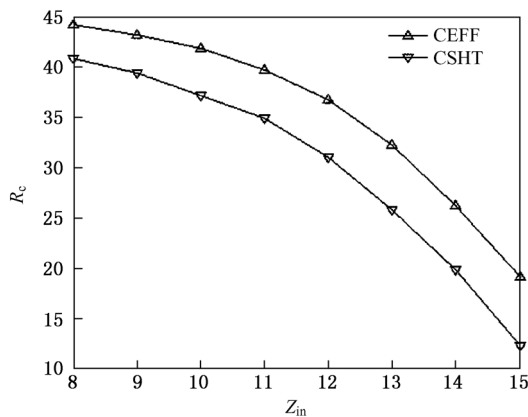
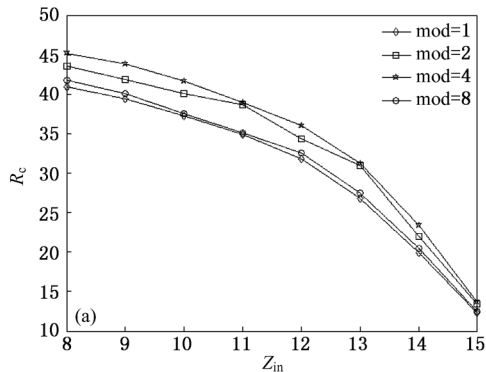


图 1 Z_{in} 与 R_c 的关系 ($\text{mod} = 1$)

从图 1 中可以看出, 当 Z_{in} 为 8 时, $Z_{in} = Z_{out} = 8$, 为一个完全随机网络, 随着 Z_{in} 的增加, 社团结构特



征变得明显, 采用最短路由和有效路由对应的网络负载传输临界值 R_c 都逐渐减小, 这就表明, 不论采用最短路由还是有效路由, 网络社团结构特征对负载传输都有影响, 明显社团结构特征会降低网络负载传输的性能.

接着, 我们研究社团划分的准确性对基于社团结构的路由策略的影响.

我们将上述伪随机网络分别划分成为 2, 4, 8 个社团, 对应的模块度 Q 如图 2 所示. 采用 CSHT 和 CEFF 策略得到路由表, 然后进行仿真得到对应的网络负载传输临界值 R_c 如图 3 所示, 平均路由长度如图 4 所示.

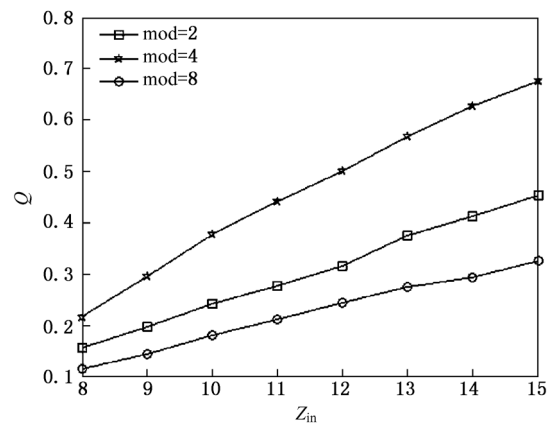


图 2 Z_{in} 和 Q 的关系

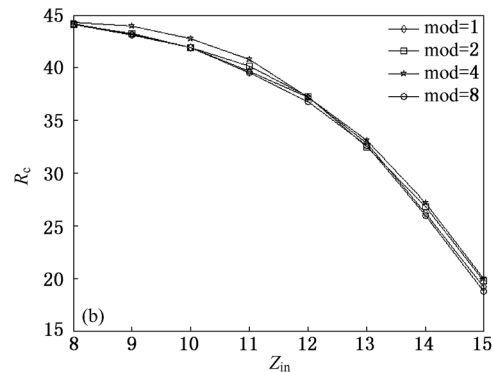
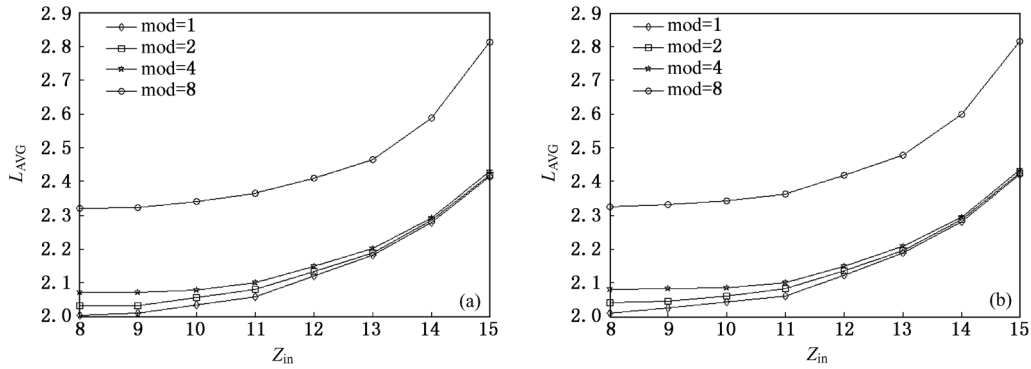


图 3 划分成 $\text{mod} = 1, 2, 4, 8$ 个社团, Z_{in} 和 R_c 的关系 (a) CSHT; (b) CEFF

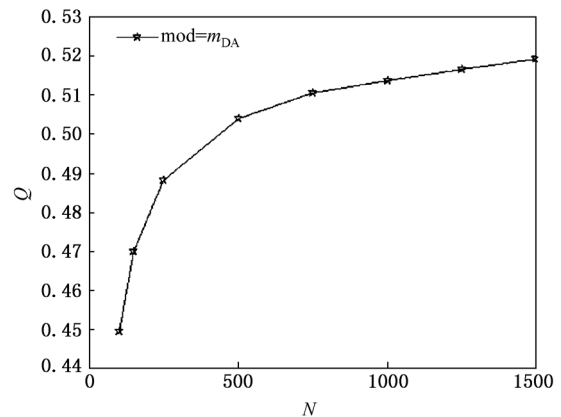
对照图 2 和图 3 可以看出, 采用 CSHT 策略能提高网络的承载能力; 当网络划分成不同社团时, 得到的模块度 Q 不相同, 网络承载能力也不相同, Q 值越大, 网络承载能力越强, 当划分成 4 个社团时, 模块度 Q 达到最大, 采用 CSHT 策略也最有效; 而采用 CEFF 策略, 网络承载能力几乎不受模块度的影响, 这和前文分析一致. 图 4 表明采用 CSHT 和

CEFF 策略的平均路由长度在未划分社团 ($\text{mod} = 1$) 和划分成得到最大模块度 Q 的社团 ($\text{mod} = 4$) 时差别不大.

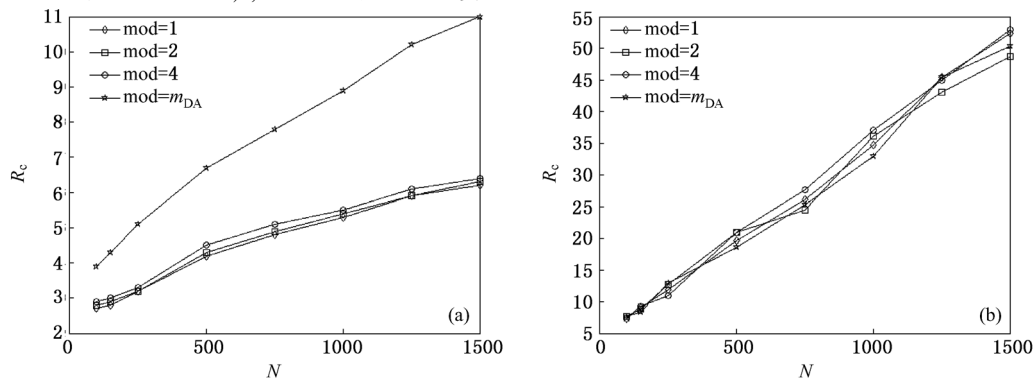
由于随机网络不能反映现实复杂网络的异构特征, Barabási 与 Albert 提出了一种具有无标度特征的 BA 网络模型^[21], 下面再研究 CSHT 和 CEFF 策略对 BA 网络模型承载能力的影响. 如上文所述,

图4 划分成 mod = 1, 2, 4, 8 个社团, Z_{in} 和 L_{AVG} 的关系 (a) CSHT; (b) CEFF

我们同样考虑对 BA 网络模型进行划分, 以得到最大的模块度 Q . 时间复杂度和准确性是复杂网络社团结构分析算法面临的两个主要问题, 早期 Girvan 和 Newman 提出的不断从网络中移除介数较大的边的 GN 算法^[12]时间复杂度为 $O(m^2n)$ (m 为网络中边数, n 为节点数, 下同), 目前, 最快的是 Wu 和 Huberman 提出的一种基于电阻网络电压谱的快速谱分割法^[22], 时间复杂度为 $O(m+n)$, 但要求事先知道社团数量. 在不知道社团数量的情况下, Clauset, Newman 和 Moore 在 Newman 快速算法^[14]的基础上提出了用堆的数据结构来计算的算法^[16], 时间复杂度只有 $O(n \log^2 n)$, 但准确性不是太高. Duch 和 Arenas 提出了一种基于启发式搜索的极值优化算法^[23], 通过调整局部极值来优化全局的变量, 提高了运算效率, 采用堆数据结构时时间复杂度为 $O(n^2 \log n)$, 准确性得到了很好的保证. 人们对如何找到快速而且可靠的网络社团结构分析算法进行了研究和对比^[24, 25], 针对本模型的特点, 网络规模较小或者适中 (对于大规模网络也可通过将其划分为多个子网来实施), 对社团划分的准确性要求较高 (以得到较高的模块度), 本模型采用 DA 算

图5 划分成 mod = m_{DA} 个社团, N 和 Q 的关系 (BA 网络参数 $m = 2$, N 为节点数, 下同)

法^[23]对 BA 网络进行划分以得到相对较准确的社团结构, 得到的社团数量记为 m_{DA} , 同时采用 GN 算法^[12]分别将不同节点数的 BA 网络 (BA 网络参数 $m = 2$) 划分成 2, 4 个社团以进行对比, 对应的模块度 Q 如图 5 所示 (当不同节点数的 BA 网络划分成 2, 4 个社团时, 得到的模块度 Q 波动较大, 故只画出 m_{DA} 个社团对应的 Q), 然后进行仿真得到对应的网络负载传输临界值 R_c 如图 6 所示.

图6 划分成 mod = 1, 2, 4, m_{DA} 个社团, N 和 R_c 的关系 (a) CSHT; (b) CEFF

可以看出,对于异构的 BA 网络,采用 CSHT 策略也能提高网络的承载能力;同样当我们划分社团得到的模块度最大时,采用 CSHT 策略最有效,而且比基于 ER 模型的伪随机网络更明显(如图 3 所示后者网络承载能力提高最多达到 16.4%,而在 BA 网络中,节点数为 100 时,网络承载能力提高了 44.4%,节点数为 150 时,网络承载能力提高了 53.57%,且随着节点数的增加,模块度也越大,网络承载能力提高得越多);而采用 CEFF 策略仍然几乎不受模块度的影响。

由于负载传输时经过的路径长度直接影响到

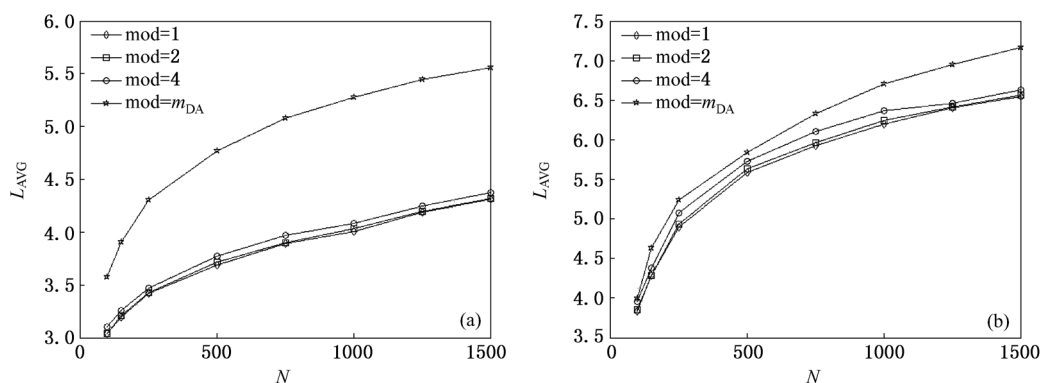


图 7 划分成 $\text{mod} = 1, 2, 4, m_{\text{DA}}$ 个社团, N 和 L_{AVG} 的关系 (a) CSHT; (b) CEFF

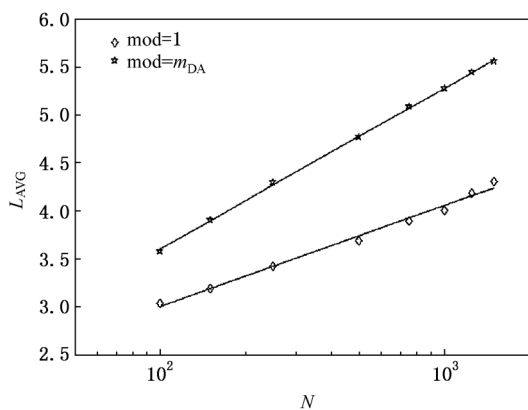


图 8 采用 CSHT 策略 N 和 L_{AVG} 的关系 ($\text{mod} = 1$ 和 $\text{mod} = m_{\text{DA}}$)

最后,采用两个实际网络来验证 CSHT 策略的有效性. 分别选择具有 198 个节点的 Jazz 网络^[26]和 1133 个节点的 E-mail 网络^[27],采用 CSHT 策略在未划分社团时的网络负载传输临界值 R_c 分别为 6.8 和 26.2,平均路由长度为 2.235 和 3.606;采用 DA 算法分别划分成 5 个和 15 个社团,此时的模块度最大,分别为 0.445 和 0.574,采用 CSHT 策略的网络负载传输临界值 R_c 分别为 7.2 和 32.9,平均路由长度为 2.255 和 3.928,分别以 0.895% 和 8.93% 的

传输的效率和成本,我们对平均路径长度进行了研究得到结果图 7 所示,发现采用 CSHT 策略的平均路由长度在未划分社团 ($\text{mod} = 1$) 和划分成得到最大模块度 Q 的社团 ($\text{mod} = 4$) 时有一定的增加(增幅在 17.7%—31.7% 之间),但是由此带来的网络负载传输临界值的增加更为显著(增幅在 44.4%—77.4% 之间);而采用 CEFF 策略的平均路由长度受模块度的影响仍然很小. 接着对采用 CSHT 策略的平均路径长度与网络规模的关系进行了研究,发现两者之间也呈对数关系,如图 8 所示,表明该策略具有较高的实际应用价值。

平均路由长度的增长换来了 5.88% 和 25.57% 的传输临界值的增加,表明 CSHT 策略在一些实际网络中也有效。

该策略的时间复杂度主要由划分社团的算法决定,本文采用的 DA 算法时间复杂度为 $O(n^2 \log n)$,随后进行的最短路径算法时间复杂度为 $O(n^3)$,时间复杂度变化不大,但是本策略求最短路径时,是在相对较小的社团内部进行的,此时的节点数比整个网络节点数小了近一个数量级,所以总体耗时和直接采用最短路径算法(复杂度为 $O(n^3)$)相差无几. 因此,如果网络规模较小或者适中,则生成固定路由表的时间复杂度是完全可以接受的。

4. 结 论

研究表明中尺度上的网络社团结构特征对网络的承载能力有影响,社团结构特征的存在意味着在社团的边缘存在着负载传输的瓶颈,明显的社团结构特征会降低网络的承载能力. 最短路径策略在网络工程中有广泛的应用,一方面是因为沿着最

短路径进行信息传输可以提高传输效率,另一方面是因为最短路径策略在网络上容易实现. 我们提出了基于社团结构的负载传输策略,尽量保证节点对之间的最短路径经过的社团数量尽可能的少,从而降低社区边缘节点的介数,达到提高网络承载能力的目的. 实验表明,采用 CSHT 策略在划分社团前后效果明显,而 CEFF 几乎没有作用;而且网络社团划分得越准确,得到的模块度也越大,采用 CSHT 策略

提高网络承载能力也越明显. 通过对基于 ER 模型的伪随机网络、BA 无标度网络和两种实际网络的分析,在采用 CSHT 策略时,能够以略增加平均路由长度换来网络承载能力的较大幅提升. 因此,对于中小规模网络,本策略具有较高的实用价值. 即使对于大规模网络,也可将其划分成不同层次结构的子网来实现.

-
- [1] Zhou T 2008 *Physica A* **387** 3025
 - [2] Yan G, Zhou T, Hu B, Fu Z Q, Wang B H 2006 *Phys. Rev. E* **73** 046108
 - [3] Ling X, Hu M B, Jiang R, Wu Q S 2010 *Phys. Rev. E* **81** 016113
 - [4] Li T, Pei W J, Wang S P 2009 *Acta Phys. Sin.* **58** 5903 (in Chinese) [李涛、裴文江、王少平 2009 物理学报 **58** 5903]
 - [5] Wang W X, Yin C Y, Yan G, Wang B H 2006 *Phys. Rev. E* **74** 016101
 - [6] Wang W X, Wang B H, Yin C Y, Xie Y B, Zhou T 2006 *Phys. Rev. E* **73** 026111
 - [7] Yin C Y, Wang B H, Wang W X, Yan G, Yang H J 2006 *Eur. Phys. J. B* **49** 205
 - [8] Ling X, Hu M B, Jiang R, Wang R L, Cao X B, Wu Q S 2009 *Phys. Rev. E* **80** 066110
 - [9] Ling X, Jiang R, Wang X, Hu M B, Wu Q S 2008 *Physica A* **387** 4709
 - [10] Chen H L, Liu Z X, Chen Z Q, Yuan Z Z 2009 *Acta Phys. Sin.* **58** 6068 (in Chinese) [陈华良、刘忠信、陈增强、袁著祉 2009 物理学报 **58** 6068]
 - [11] Pu C L, Pei W J 2010 *Acta Phys. Sin.* **59** 3841 (in Chinese) [濮存来、裴文江 2010 物理学报 **59** 3841]
 - [12] Girvan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821
 - [13] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
 - [14] Newman M E J 2004 *Phys. Rev. E* **69** 066133
 - [15] Shen H W, Cheng X Q, Cai K, Hu M B 2009 *Physica A* **388** 1706
 - [16] Clauset A, Newman M E J, Moore C 2004 *Phys. Rev. E* **70** 066111
 - [17] Zou S R, Peng Y J, Liu A F, Xu X L, He D R 2011 *Chin. Phys. B* **20** 018902
 - [18] Danon L, Arenas A, Díaz – Guilera A 2008 *Phys. Rev. E* **77** 036103
 - [19] Arenas A, Díaz – Guilera A, Guimerà R 2001 *Phys. Rev. Lett.* **86** 3196
 - [20] Wang X F, Li X, Cheng G R 2005 *Theory and Application of Complex Networks* (Beijing: Tsinghua University Press) p174 (in Chinese) [汪小帆、李翔、陈关荣 2005 复杂网络理论及其应用(北京:清华大学出版社) 第174页]
 - [21] Barabási A L, Albert R 1999 *Science* **286** 509
 - [22] Wu F, Huberman B A 2004 *Eur. Phys. J. B* **38** 331
 - [23] Duch J, Arenas A 2005 *Phys. Rev. E* **72** 027104
 - [24] Danon L, Díaz – Guilera A, Duch J, Arenas A 2005 *Stat. Mech.* **09** 09008
 - [25] Fortunato S 2010 *Phys. Rep.* **486** 75
 - [26] Gleiser P, Danon L 2003 *Adv. Complex Syst.* **6** 565
 - [27] Guimerà R, Danon L, Díaz – Guilera A, Giralt F, Arenas A 2003 *Phys. Rev. E* **68** 065103

Optimal traffic routing strategy based on community structure^{*}

Shao Fei^{1)2)†} Jiang Guo-Ping¹⁾

1) (*Center for Control and Intelligence Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*)

2) (*Department of Information Technology, Jinling Institute of Technology, Nanjing 211169, China*)

(Received 27 Septmeber 2010; revised manuscript received 25 February 2011)

Abstract

It is shown that community structure has great influence on traffic transportation. Networks with pronounced community structure are less efficient in terms of packet delivery. While the shortest path is chosen at random in the shortest path routing strategy, a routing strategy based on community structure is proposed in this paper which can reduce the betweenness centrality of the nodes on the edge of the community by minimizing the number of the communities that the shortest path passes through. Simulations show that the new strategy can enhance the packet delivery capability with the small – world character and that the more accurately the community is identified, the more efficient the new strategy is.

Keywords: optimal routing strategy, community structure, complex network, traffic transportation

PACS: 89.75. Da, 89.75. Fb, 89.40.-a

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 60874091), the Six Projects Sponsoring Talent Summits of Jiangsu Province (Grant No. SJ209006), the Specialized Research Fund for the Doctoral Program of Higher Education of China (SRFDP) (Grant No. 20103223110003), and the Research Foundation of Jinling Institute of Technology of China (Grant No. JIT-N-201010).

[†] E-mail: shaofei@jit.edu.cn