

基于度与集聚系数的网络节点重要性度量方法研究*

任卓明 邵凤 刘建国[†] 郭强 汪秉宏

(上海理工大学复杂系统科学研究中心, 上海 200093)

(2013年1月16日收到; 2013年2月6日收到修改稿)

网络中节点重要性度量对于研究网络的鲁棒性具有十分重要的意义. 研究者们普遍运用度或集聚系数来度量节点的重要程度, 然而度指标只考虑节点自身邻居个数而忽略了其邻居之间的信息, 集聚系数只考虑节点邻居之间的紧密程度而忽略了其邻居的规模. 本文综合考虑节点的邻居个数, 以及其邻居之间的连接紧密程度, 提出了一种基于邻居信息与集聚系数的节点重要性评价方法. 对美国航空网络和美国西部电力网进行的选择性攻击实验表明, 采用该方法的效果较 k -shell 指标可以分别提高 24% 和 112%. 本文的节点重要性度量方法只需要考虑网络局部信息, 因此非常适合于对大规模网络的节点重要性进行有效分析.

关键词: 网络科学, 鲁棒性, 节点重要性, 集聚系数

PACS: 89.75.Hc, 89.75.Fb

DOI: 10.7498/aps.62.128901

1 引言

现实世界的许多系统都可以用网络模型进行刻画^[1,2], 对网络中的节点重要性进行评估和度量对于提高实际系统的鲁棒性^[3-5]、设计高效的系统结构具有十分重要的意义. 准确评估节点重要性之后, 一方面可以通过重点保护这些节点来提高整个网络的可靠性与抗毁性^[6,7], 例如在由通航城市和航线所构成的航空网络中, 如果某通航城市受到突发事件影响陷入瘫痪, 也就意味着同时取消了与该通航城市相连的所有的航线, 从而有可能使得航空网络中其他通航城市之间的一些运输路径中断. 于是发现航空网络中的重要节点并进行有效保护, 可以避免其在受到外界干扰时造成航线大面积延误甚至导致网络瘫痪, 保证航空运输安全高效地运营. 另外一方面, 也可以通过攻击这些重要节点达到摧毁整个网络的目的, 例如 2011 年 1 月, 由美国和以色列精心研制的“震网 (Stuxnet)”病毒破坏伊朗核电网络的关键设施, 成功攻击和控制了伊朗核电网络, 造成伊朗核电站推迟发电. 大量的

研究表明, 分析节点的度信息是度量节点重要性的一个重要方法. 例如 Jeong 等^[8]研究了蛋白质网络, Dunne 等^[9]研究了食物链网络, Newman 等^[10]研究了电子邮件网络, Magoni^[11]研究了英特网, Samant 和 Bhattacharyya^[12]研究了 P2P 网络, 这些研究结果表明去掉网络中度最大的节点后, 网络变得非常脆弱. 然而, 度指标只利用了节点自身信息, 并没有考虑节点在网络中的位置, 也无法对其邻居之间的连接状况进行刻画. Freeman^[13]则考虑节点到达网络中所有其他节点的最短距离来衡量节点的重要性, 并将其定义为紧密度. Holme 等^[14]则从全局信息的角度对网络鲁棒性进行了研究, 发现介数较大的节点对网络鲁棒性具有重要的作用. 因为紧密度与介数都需要计算网络中任意一对节点之间的最短路径, 所以算法的时间复杂度非常高, 对于大规模网络并不适用. 最近, 也有学者基于信息扩散和疾病传播的视角挖掘网络中的重要节点. 例如 Lü 等^[15]提出的 LeaderRank 解决了经典的 PageRank 算法^[16]排序的惟一性的问题, 按 LeaderRank 算法对网络节点重要性进行排序比

* 国家自然科学基金 (批准号: 11275186, 71071098, 71171136, 91024026)、上海市青年科技启明星计划 (A 类)(批准号: 11QA1404500)、上海市教委科研创新项目 (批准号: 11ZZ135, 11YZ110)、教育部科学技术研究重大项目 (批准号: 211057)、上海市一流学科 (系统科学) 建设项目 (批准号: XTKX2012) 和上海市研究生创新基金 (批准号: 5411115004) 资助的课题.

[†] 通讯作者. E-mail: liujg004@ustc.edu.cn

PageRank 算法排序更精准, 而且对网络噪音 (节点随机加边或删边) 有更好的容忍性. Kitsak 等^[17] 提出节点在网络中的位置是影响其重要性的主要因素, 并利用 k -shell (ks) 分解来识别网络中最重要节点, 认为 ks 指标越大的节点越重要. 然而, ks 指标赋予大量节点相同的值, 例如 Barabasi-Albert (BA) 网络^[18] 模型的所有节点的 ks 值都相等, 从而导致 ks 指标无法衡量其节点的重要性^[19]. Chen 等^[20] 考虑节点多级邻居的度信息, 定义了一个多级邻居信息指标来识别复杂网络中最重要节点, 然而该指标没有考虑目标节点与其邻居之间的紧密程度. Centola^[21] 邀请 1540 名志愿者, 随机地一对一地分派具有小世界特性的随机网络和高集聚^[22] 的规则网络, 通过 12 次独立实验, 研究在线社会网络的传播行为, 发现传播行为在高集聚类网络中传播得更快, 节点的传播影响力与该节点的集聚系数有关. Ugander 等^[23] 通过对 Facebook 上一个用户收到某个邮件联系人的邀请信而成为 Facebook 用户情况的分析, 发现邻居节点的绝对数目不是影响节点重要性的决定性因素, 起决定作用的是邻居节点之间形成的联通子图的数目. 由此可见, 节点的度和集聚系数信息对刻画其重要性都具有十分重要的意义. 本文利用节点的邻居信息, 并考虑节点邻居之间的紧密程度, 提出一种基于邻居信息与集聚系数的节点重要性评价指标. 对不同参数的 BA 理论模型网络和美国航空网络、美国西部电力网络等实际网络的实验结果表明, 该指标较度指标、基于节点度和其邻居度的指标、 ks 指标更能准确度量节点重要性.

2 基于度和集聚系数的节点重要性度量方法

假设网络 $G = (V, E)$ 是由 $|V| = N$ 个节点和 $|E| = M$ 条边所组成的一个无向网络. 度指标描述了一个节点的邻居节点的个数, 表示为

$$k_i = \sum_{j \in G} \delta_{ij}, \quad (1)$$

其中, $\delta_{ij} = \begin{cases} 1 & i \text{ 到 } j \text{ 有连接} \\ 0 & i \text{ 到 } j \text{ 没有连接} \end{cases}$. 度指标体现了该节点与周围节点之间建立直接联系的能力, 但不能反映该节点的邻居节点的连边情况.

集聚系数^[22] 描述了网络中节点的邻居之间互为邻居的比例, 表示为

$$c_i = \frac{2e_i}{k_i(k_i - 1)}, \quad (2)$$

其中 e_i 表示节点 i 与其任意两个邻居节点之间所形成的三角形的个数. 与度指标相反, 集聚系数虽然在一定程度上能够反映邻居节点的连边情况, 但不能反映邻居节点的规模. 于是我们利用节点邻居信息, 并考虑集聚系数, 提出一种新的节点重要性评价指标 p_i , 表示为

$$p_i = \frac{f_i}{\sqrt{\sum_{j=1}^N f_j^2}} + \frac{g_i}{\sqrt{\sum_{j=1}^N g_j^2}}, \quad (3)$$

其中, f_i 为节点 i 自身度与其邻居度之和, 表示为

$$f_i = k_i + \sum_{w \in I_i} k_w, \quad (4)$$

其中, k_w 表示节点 w 的度, I_i 表示节点 i 的邻居节点集合. 函数 g_i 表示为

$$g_i = \frac{\max_{j=1}^N \left\{ \frac{c_j}{f_j} \right\} - \frac{c_i}{f_i}}{\max_{j=1}^N \left\{ \frac{c_j}{f_j} \right\} - \min_{j=1}^N \left\{ \frac{c_j}{f_j} \right\}}, \quad (5)$$

其中, c_i 为节点 i 的集聚系数. 集聚系数仅能反映邻居节点之间的紧密程度, 而不能反映邻居节点的规模, 我们对 c_i/f_i 的结果进行如 (5) 式所示的归一化处理^[24]. 由于 f_i 反映的是节点自身度和邻居的度信息, g_i 反映的是节点邻居之间的紧密程度, 采用同趋化函数^[24] $u(x) = x/\sqrt{\sum x^2}$ 同时对 f_i 和 g_i 进行处理, 使得 p_i 正确反映 f_i 和 g_i 不同作用力的综合结果.

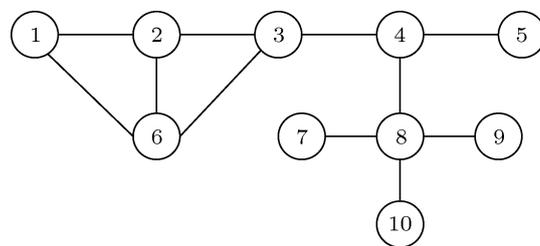


图 1 10 个节点的网络图

例如在图 1 中, 显然删除节点 4 对该小网络造成的破坏性大于删除节点 2. 然而当仅考虑节点度及其邻居节点的度时, $k_2 = k_4 = 3$, $f_2 = f_4 = 11$. 进一步考虑节点的邻居节点之间的关系时, 通过 (3)

式计算可知 $p_2 = 0.59$, $p_4 = 0.77$, 此时 $p_4 > p_2$, 表明节点的邻居之间的紧密程度对节点重要性评价是有影响的, 因此衡量一个节点的重要性需要综合考虑节点邻居信息与其集聚系数的综合影响.

表 1 图 1 所示的网络中各节点属性

节点	k	f_i	c_i	g_i	p_i
1	2	8	1	0	0.29
2	3	11	0.67	0.19	0.59
3	3	12	0.33	0.29	0.73
4	3	11	0	0.37	0.77
5	1	4	0	0.37	0.52
6	3	11	0.67	0.19	0.59
7	1	5	0	0.37	0.56
8	3	10	0	0.37	0.74
9	1	5	0	0.37	0.56
10	1	5	0	0.37	0.56

3 网络效率

网络效率^[25,26]是用来表示网络连通性好坏的指标, 网络的连通性越好, 则网络效率越高. 假设在一个网络中, 在某个节点遭受到网络攻击后, 移走该节点, 这也就意味着同时移走了与该节点相连的所有边, 从而可能使得网络中其他节点之间的一些路径被中断. 如果在节点 i 和节点 j 之间有多条路径, 中断其中的一些路径就可能使这两个节点之间的最短路径 d_{ij} 增大, 进而整个网络的平均路径长度也会增大, 从而使得网络的连通性变差. 两节点 i 和 j 之间的效率为 $1/d_{ij}$, 则网络效率表示为

$$\varepsilon = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}, \quad (6)$$

ε 的取值范围为 $[0, 1]$, $\varepsilon = 1$ 表示网络连通性最好, $\varepsilon = 0$ 则表示网络是由孤立的节点组成的.

本文通过选择性删除网络中一定比例的节点进行网络的蓄意攻击仿真实验, 计算网络攻击前、后的网络效率下降比例来定量刻画各种节点重要性指标的准确性^[27,28]. 假设网络在未遭受到网络攻击时, 网络效率为 ε_0 , 选择性删除网络中一定比例 p (p 的取值范围为 $[0,1]$) 的节点后网络效率为 ε , 网络效率下降比例记为

$$e = 1 - \frac{\varepsilon}{\varepsilon_0}, \quad (7)$$

其中 e 的取值范围为 $[0, 1]$, $e = 1$, 表示网络在遭受攻击后, 网络效率下降到了 0, 即此时网络由孤立的节点组成; $e = 0$, 表示网络在遭受攻击后, 整个网

络的效率没有变化. 由 (7) 式可以看出, 采用不同的节点重要性指标分别选择性删除网络中一定比例 p 的节点后, e 值越大, 表明摘除这些节点后网络效率变得越差, 采用该方法更能准确度量节点的重要性.

4 实例验证

为了进一步验证指标 p_i 度量节点重要性的效果, 本文选用 p_i 、度 (k)、基于节点度和其邻居度的指标 (f_i) 和 ks 指标四个指标分别对美国航空网^[29]、美国西部电力网^[30] 以及不同参数的 BA 网络进行节点的选择性摘除的蓄意攻击仿真实验. 美国航空网络的节点数为 332, 边数为 2126, 选择性删除按各种指标排名的前 $[10\%, 20\%]$ 的节点; 美国西部电力网络节点数为 4940, 边数为 6591, 选择性删除按各种指标排名的前 $[1\%, 10\%]$ 节点; 9 个不同新连边数的 BA 网络节点数都为 10000, 新连边数 m 值分别为 2, 3, 4, 5, 6, 7, 8, 9, 10, 选择性地删除各种指标排名前 5% 的节点.

对美国航空网和美国西部电力网分别采用 p_i 指标、度指标 k 、指标 f_i 以及 ks 指标删除一定比例 (p) 的排名靠前的节点. 实验结果如图 2(a), (b) 所示, 采用 p_i 指标删除节点导致网络效率下降的幅度最大. 例如在图 2(a) 美国航空网中, $p = 0.1$ 时, 指标 p_i 、度指标、指标 f_i 以及 ks 指标的 e 值分别为 0.83, 0.74, 0.72, 0.67, 这表明采用 p_i 指标删去排名前 10% 的节点, 与采用度指标, 指标 f_i 以及 ks 指标删去排名前 10% 的节点相比, 网络的效率变得最差. p_i 的攻击效果较 ks 指标提高了 24%. 同样例如在图 2(b) 美国西部电力网中, $p = 0.01$ 时, 指标 p_i 、度指标、 f_i 指标以及 ks 指标的 e 值分别为 0.34, 0.20, 0.14, 0.16, 这表明采用 p_i 指标删去排名前 1% 的节点, 与采用度指标, 指标 f_i 以及 ks 指标删去排名前 1% 的节点相比, 网络效率同样变得最差. p_i 的攻击效果较 ks 指标提高了 112%. 这表明 p_i 指标在两个实证网络中较度指标、 f_i 指标以及 ks 指标更能准确识别网络中的最重要节点. 在不同 m 值的 BA 网络模型中, 分别删除按指标 p_i 、度指标 k 和指标 f_i 进行排名的前 5% 的节点. 实验结果如图 2(c) 所示, 随 m 值的增大, 虽然 p_i 指标、度指标、 f_i 指标的 e 值都逐渐减小, 但 p_i 指标比度指标和 f_i 指标的结果好, 而 ks 指标不适用于 BA 网络. 因此, p_i 指标比度指标和 f_i 指标能更为准确地度量节点的重要性.

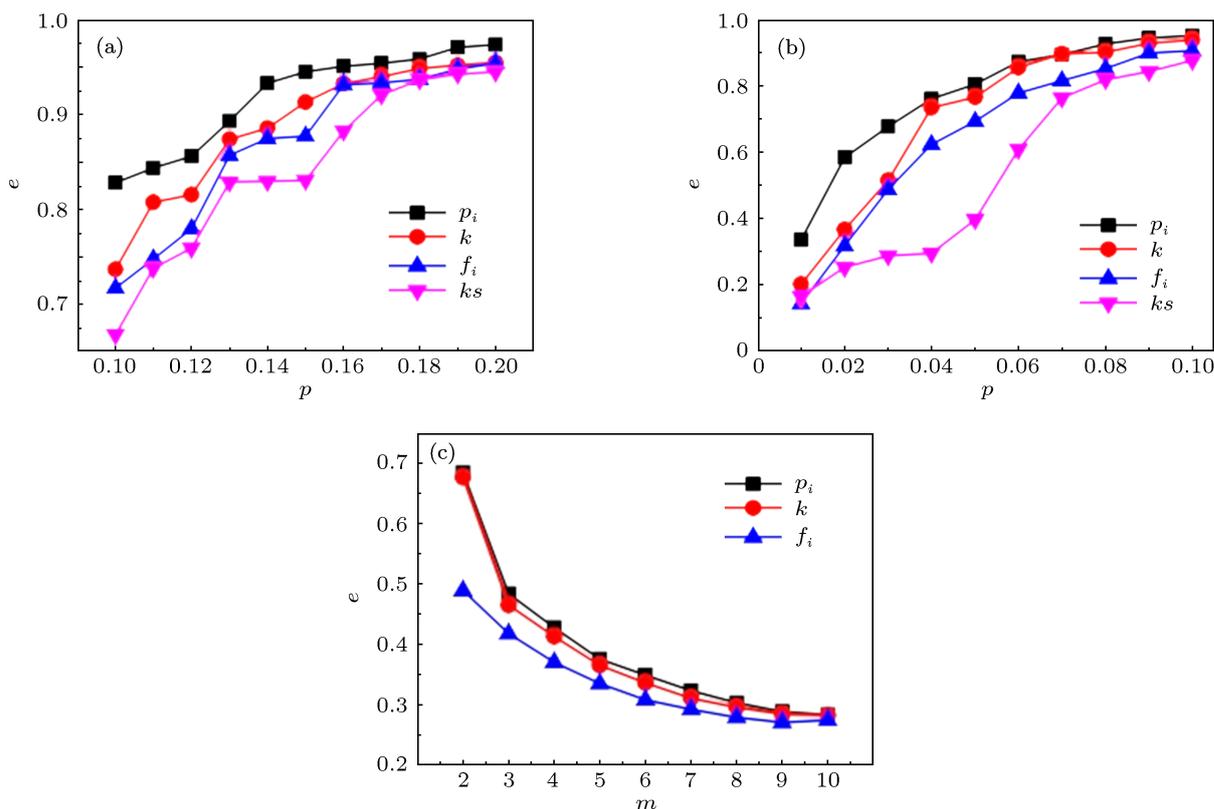


图2 美国航空网、美国西部电力网以及不同 m 值的 BA 网络分别采用节点重要性指标删除排名靠前的节点后网络 e 值的变化情况 (a) 美国航空网; (b) 美国西部电力网; (c) BA 网络

5 总结

本文提出了一种基于节点邻居信息以及聚集系数的节点重要性评价指标 p_i . 该指标只需要计算节点的邻居信息及其聚集系数等局部信息, 因此对刻画大规模网络的可靠性与抗毁性具有十分重要的意义. 对美国航空网、美国西部电力网络和 BA 网络的鲁棒性仿真研究结果表明, 采用 p_i 指标对网络节点进行重要性排序, 蓄意攻击其排名靠前的节点后, 网络效率的下降幅度比根据度或考虑节点自身度和邻居度信息指标、 ks 等指标更大. 由于指标 p_i 涉及聚类系数, 在社会网络等高聚集性网络中,

其节点重要性评估的效果会有所降低, 我们将会进一步探讨在社会网络中指标 p_i 的度量的准确性与度、聚集系数之间的关系.

另外, 也有学者发现节点的重要性不仅与网络结构相关, 还与传播机制有关. Klemm 等^[31] 的研究表明集群动力学中节点的重要性是由网络拓扑结构和集群动力学机制决定的. Borge-Holthoefer 和 Moreno^[32] 认为在真实网络的谣言传播过程中, 节点的重要性不是由该节点的 ks 位置决定的, 而是由谣言传播扩散机制决定的. 本文从结构的角度对节点的重要性进行了分析, 未来将重点结合动力学特性和网络结构对节点的重要性度量进行研究.

[1] Newman M E J 2003 *SIAM Rev.* **45** 167
 [2] Barabasi A L 2009 *Science* **325** 412
 [3] Zhou T, Bai W J, Wang B H, Liu Z J, Yan G 2005 *Physics* **34** 31 (in Chinese) [周涛, 柏文洁, 汪秉宏, 刘之景, 严钢 2005 物理 **34** 31]
 [4] Liu J G, Wang Z T, Dang Y Z 2005 *Mod. Phys. Lett. B* **19** 785
 [5] Zhou X, Zhang F M, Zhou W P, Zhou W, Yang F 2012 *Acta Phys. Sin.* **61** 190201 (in Chinese) [周漩, 张凤鸣, 周卫平, 邹伟, 杨帆 2012 物理学报 **61** 190201]
 [6] Tan Y J, Wu J, Deng H Z, Zhu D Z 2006 *Syst. Eng.* **24** 1 (in Chinese) [谭跃进, 吴俊, 邓宏钟, 朱大智 2006 系统工程 **24** 1]
 [7] Liu H K, Zhou T 2007 *Acta Phys. Sin.* **56** 106 (in Chinese) [刘宏鲲, 周涛 2007 物理学报 **56** 106]
 [8] Jeong H, Mason S, Barabasi A L 2001 *Nature* **411** 41
 [9] Dunne J A, Williams R J, Martinez N D 2002 *Ecol. Lett.* **5** 558
 [10] Newman M E J, Forrest S, Balthrop A 2002 *J. Phys. Rev. E* **66** 035101
 [11] Magoni D 2003 *IEEE J. Sel. Area. Commun.* **21** 949
 [12] Samant K, Bhattacharyya S 2004 *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* Washington, USA, January 5-8, 2004 p289
 [13] Freeman L C 1977 *Sociometry* **40** 35

- [14] Holme P, Kim B J, Yoon C N, Han S K 2002 *Phys. Rev. E* **65** 056109
 [15] Lü L Y, Zhang Y C, Yeung C H, Zhou T 2011 *PLoS One* **6** e21202
 [16] Bryan K, Leise T 2006 *SIAM Rev.* **48** 569
 [17] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
 [18] Barabasi A L, Albert R 1999 *Science* **286** 509
 [19] Zeng A, Zhang C J 2013 *Phys. Lett. A* **377** 1031
 [20] Chen D B, Lv L Y, Shang M S, Zhang Y C, Zhou T 2011 *Physica A* **391** 1777
 [21] Centola D 2010 *Science* **329** 1194
 [22] Watts D J, Strogatz S H 1998 *Nature* **393** 440
 [23] Ugander J, Backstrom L, Marlow C, Kleinberg J 2012 *Proc. Natl. Acad. Sci.* **109** 5962
 [24] Codd E F 1970 *Commun. ACM* **13** 377
 [25] Vragovic I, Louis E, Diaz-Guilera A 2005 *Phys. Rev. E* **71** 036122
 [26] Latora V, Marchiori M 2007 *New J. Phys.* **9** 188
 [27] Li P X, Ren Y Q, Xi Y M, 2004 *Syst. Eng.* **22** 13 (in Chinese) [李鹏翔, 任玉晴, 席酉民系统工程 2004 **22** 13]
 [28] He N, Li D, Gan W Y, Zhu X 2007 *Comput. Sci.* **34** 1 (in Chinese) [赫南, 李德, 淦文燕, 朱熙 2007 计算机科学 **34** 1]
 [29] Batagelj V, Mrvar A 1998 *Connections* **21** 47
 [30] Strogatz S H 1998 *Nature* **393** 440
 [31] Klemm K, Serrano M, Eguiluz V, Miguel M 2012 *Sci. Rep.* **2** 292
 [32] Borge-Holthoefer J, Moreno Y 2012 *Phys. Rev. E* **85** 026116

Node importance measurement based on the degree and clustering coefficient information*

Ren Zhuo-Ming Shao Feng Liu Jian-Guo[†] Guo Qiang Wang Bing-Hong

(Complex Systems Science Research Center, University of Shanghai for Science and Technology, Shanghai 200093, China)

(Received 16 January 2013; revised manuscript received 6 February 2013)

Abstract

The node importance measurement plays an important role in analyzing the robustness of the network. Most researchers use the degree or clustering coefficient to measure the node importance. However, the degree can only take into account the neighbor size, regardless of the clustering property of the neighbors. The clustering coefficient could only measure the closeness among the neighbors and neglect the activity of the target node. In this paper, we present a new method to measure the node importance by combining neighbor and clustering coefficient information. The robustness results measured by the network efficiency through removing the important nodes for the US Air network, the power grid of the western United States and Barabasi-Albert networks show that the new method can more accurately evaluate the node importance than the degree, neighbor information and k -shell indices.

Keywords: network science, robustness, node importance, clustering coefficient

PACS: 89.75.Hc, 89.75.Fb

DOI: 10.7498/aps.62.128901

* Project supported by the National Natural Science Foundation of China (Grant Nos. 11275186, 71071098, 71171136, 91024026), the Shanghai Rising-Star Program, China (Grant No. 11QA1404500), the Innovation Program of Shanghai Municipal Education Commission, China (Grant Nos. 11ZZ135, 11YZ110), the Key Project of Chinese Ministry of Education (Grant No. 211057), the Shanghai Leading Academic Discipline Project (Systems Science) (Grant No. XTKX2012), and the Innovation Fund For Graduate Student of Shanghai, China (Grant No. 54-11-115-004).

[†] Corresponding author. E-mail: liujg004@ustc.edu.cn