

演化博弈与自组织合作^{*}

王 龙 伏 锋 陈 小 杰 楚 天 广 谢 广 明

(北京大学系统与控制研究中心智能控制实验室, 北京 100871)

摘要 主要研究复杂网络上的演化博弈. 首先研究具有社团结构的无标度网络上的演化囚徒困境博弈及 Newman-Watts 小世界网络中异质性对合作演化的影响. 然后考察了在不同合作者和作弊者初始分布配置情况下, 不同的初始比例条件对合作水平的影响, 且在社会网络上研究了雪堆博弈中的合作演化. 进一步地, 讨论了网络拓扑和博弈动力学的共同演化问题和网络上演化囚徒困境中的强化学习问题. 最后给出了复杂网络上演化博弈论的未来发展方向与应用前景.

关键词 群体行为, 合作, 演化博弈论, 复杂网络, 囚徒困境博弈, 雪堆博弈, 强化学习, 自组织, 拓扑, 动力学.

MR(2000) 主题分类号 91A22, 91D30, 91A43

1 引 言

合作现象在自然界非常普遍, 从单细胞的微生物到高级哺乳动物, 从社会性昆虫到人类社会. 生物个体之间的合作, 增加了整个种群的生存机会. 然而, 依据达尔文主义, 自然选择是基于竞争的, 个体将会自私地最大化自身的利益, 这显然不能解释处处可见的合作行为. 为了解合作行为是如何从自私的个体之间演化产生的, 近年来, 一些可能的合作机理, 如群体选择 (group selection)、亲缘选择 (kin selection)、直接 (间接) 互惠 (direct/indirect reciprocity)、空间互惠 (spatial reciprocity)、声望与惩罚 (reputation and punishment) 等得到了一定的探讨^[1]. 自从 Axelrod 利用迭代囚徒困境博弈研究合作的演化以来^[2], 演化博弈论为研究合作演化提供了方便的数学框架, 囚徒困境博弈 (Prisoner's dilemma) 和雪堆博弈 (Snowdrift game) 是被研究者广泛采用的研究合作行为的两个范例. 在原始的囚徒困境博弈中, 每个纯策略的个体都有两种选择: 合作 (cooperation, C) 与作弊 (defection, D). D 策略个体利用 C 策略个体, 获得 T 收益, 而 C 获得 S . 双方都合作则获得 R , 都作弊则获得 P , 其中 $T > R > P > S$, $2R > T + S$. 在单轮博弈情况下, 无论对手采取何种策略, 个体的最佳策略总是作弊. 然而, 在双方都采取合作策略的情况下, 二者总的收益才是最大的. 这一现象说明了社会两难 (social dilemma) 问题的实质. 而在雪堆博弈中, 与囚徒困境博弈不同的是 P 和 S 的大小顺序变了, 即 $T > R > S > P$. 此时个体的最佳策略取决于对手的选择:

^{*} 国家自然科学基金 (60674050, 60528007), 国家 973 项目 (2002CB312200), 国家 863 项目 (2006AA04Z258) 和国防科工委 11-5 项目 (A2120061303) 资助课题.

收稿日期: 2007-05-10.

当对手合作时则选择作弊, 当对手作弊时则选择合作. 依据无限混合均匀人口中的复制动力学, 合作者的频率将会收敛到 $1-r$, 其中 r 为双方合作时的损益比 (cost-to-benefit ratio) [3]. 雪堆博弈被广泛地用于研究生物之间的合作行为 [4].

复杂网络理论为描述博弈个体之间的博弈关系提供了方便的框架 [5-8]. 网络上的节点代表博弈个体, 边代表个体与其邻居的博弈关系. 这样就可以利用复杂网络拓扑关系, 来研究一些复杂博弈关系下的博弈. 例如, 以前博弈理论中的混合均匀 (well-mixed) 假设就可以看成是在全连通图上进行的博弈. 在空间二维格子 (lattice) 或一维格子 (ring) 上的博弈就可转化为规则网络上的博弈. 然而, 真实世界的网络大多是异质的 (heterogeneous), 节点的邻居数目存在差异, 甚至呈幂率分布. 因此, 研究接触网络 (network of contacts) 的异质性对其上的博弈动力学的影响是很有意义的. 在具有小世界、无标度等特性的复杂网络上的演化博弈得到了广泛的研究 [9-14], 如 Santos 等人发现异质性 (无标度) 能够提升合作水平, 促进合作产生. 另外, 网络拓扑和博弈动力学的共同演化也是合作产生的一个重要机理 [15-17].

2 复杂网络上的自组织合作

真实社会的网络拓扑除了具有小世界、无标度等特性外, 还具有社团结构 (community structure) 这一重要特性. 社团结构是指整个网络是由若干个“群 (group)”或“团 (cluster)”构成的. 每个群内部的节点之间的连接相对比较紧密, 但是各个群之间的连接却比较稀疏. 研究社团结构对合作水平的影响是很有意义的. 考虑具有社团结构的无标度网络上的囚徒困境博弈问题 [18]. 不失囚徒困境博弈的一般特性, 博弈矩阵 M 取为

$$M = \begin{pmatrix} 1 & 0 \\ b & 0 \end{pmatrix}, \quad (2.1)$$

其中 $1 < b < 2$. 采用文 [19] 中具有社团结构的无标度网络模型, 生成节点总数 $N = 6000$ 、具有 3 个相同大小群的社团结构的博弈关系网络. 用二维向量表示个体的策略类型: 合作 (C) 与作弊 (D):

$$s = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

个体 x 的收益为它跟所有邻居博弈一次的收益的总和: $P_x = \sum_{y \in \Omega_x} s_x^T M s_y$, 其中 s_x 、 s_y 表示节点的状态 (策略), Ω_x 表示 x 的所有邻居. 我们采用同步更新规则 (synchronous update rule), 在每一时间步长, 节点 x 从其邻居中随机选取节点 y 进行策略更新, 若 $P_y > P_x$, 则以概率

$$W_{s_x \leftarrow s_y} = \frac{P_y - P_x}{bk_{>}} \quad (2.2)$$

采用节点 y 所用的策略 s_y , 其中 $k_{>}$ 为节点 x 和 y 的度中的较大值. 初始时刻, 合作者与作弊者等比例随机分布在网络顶点上. 系统演化 10000 时间步长后, 再取 1000 步时间步长上合作者比例的平均数, 得到平衡态时合作者的比例. 每个数据点对应于 40 次不同的网络实现和初始分布条件下合作者比例的平均值. 图 1 显示了相同网络规模, 但不同平均度 $m+n$ 及不同社团内外连接数之比 $\frac{m}{n}$ 时的合作频率对参数 b 的变化情况. 我们发现, 在具有社团结构的无标度网络上, 随着平均度 $m+n$ 的增加, 合作水平相应地减小. 同时, 在保持平均

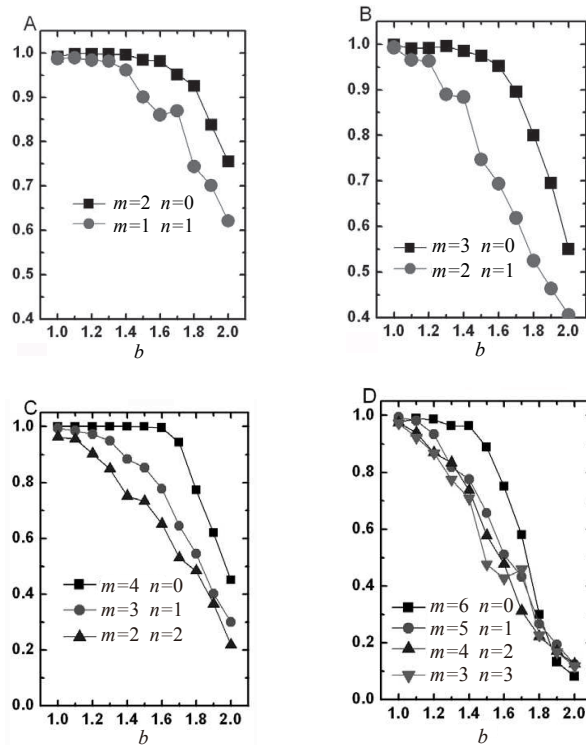


图 1 对应于不同 $m+n$ 与 $\frac{m}{n}$ 时合作频率对参数 b 的变化情况

度 $m+n$ 不变, 改变内外连接数之比 $\frac{m}{n}$ 时, 合作水平随着 $\frac{m}{n}$ 减小而降低. 另外, 在没有外部连接时 (对应于 $n=0$), 合作水平总是最优的. 此时对应于 3 个 Barabasi-Albert (BA) 无标度网络, 而无标度网络是有利于合作的产生的 [20], 因此此时合作水平最高. 随着外部连接数的增加、内部连接的减少, 网络结构中的一些 hub (网络中度较大的节点) 并不直接相连, 并且网络中回路 (loop) 减少了, 这些因素影响了合作水平 [18].

文 [20-22] 指出复杂网络的异质性是影响合作水平的重要因素. 但复杂网络的异质程度大小会对合作水平产生什么影响呢? 考虑异质 Newman-Watts 小世界网络上的演化囚徒困境博弈问题 [23]. 与 Watts-Strogatz 小世界模型中断边重连机制不同, 我们采用改进的 Newman-Watts 小世界模型, 即在低维规则环上添加 m 条长程边形成小世界网络. 首先随机地从 N 个节点中选出 N_h 个节点作为 hub. 然后使添加的长程边至少保证每条边的一个端点随机地与所选出的 N_h 个 hub 中的一个相连, 另一个端点则随机地与 N 个节点中的一个相连 (但避免自连和重边). 可以看出, 参数 $\frac{N_h}{N}$ 可反映拓扑异质性的. $\frac{N_h}{N} = 1$ 时, 网络退化到一般均质小世界的情形, 而 $\frac{N_h}{N} = \frac{1}{N}$ 对应于网络最异质的情形, 即所有的长程边都与唯一的 hub 相连. 博弈矩阵采用 (2.1) 式, 更新规则采用 (2.2) 式. 网络节点规模 $N = 2001$, 长程边数 $m = 1000$. 所有的数据点对于 100 次运行取平均, 即 10 次网络拓扑实现对应于 10 次独立的初始条件分布. 初始时刻, 合作者和作弊者等比例随机分布在网络上, 在经过 10000 步演化后, 再取 2000 步结果作平均, 作为平衡时合作者的比例. 图 2 给出了对应于不同参数 b 合作水平随着网络异质程度 $\frac{N_h}{N}$ 的变化情况. 我们发现, 对于固定的 b , 总存在适当的网络

异质程度 $\frac{N_h}{N}$ ，使得合作水平最高。也就是说，合作水平在 $\frac{N_h}{N}$ 取某个值时达到最大值，此时网络的异质程度既不是最大的也不是最小的。图 3 显示了合作水平在参数空间 $(b, \frac{N_h}{N})$ 中的变化情况。合作水平的大小由右边的色带标记。同样地，我们发现，对于固定的 b ，合作水平在 $\frac{N_h}{N} \approx 0.1$ 时达到最优。另外，更新规则使用节点的平均收益，而不是累积收益时，会削弱网络异质性对合作水平的影响 [23]。

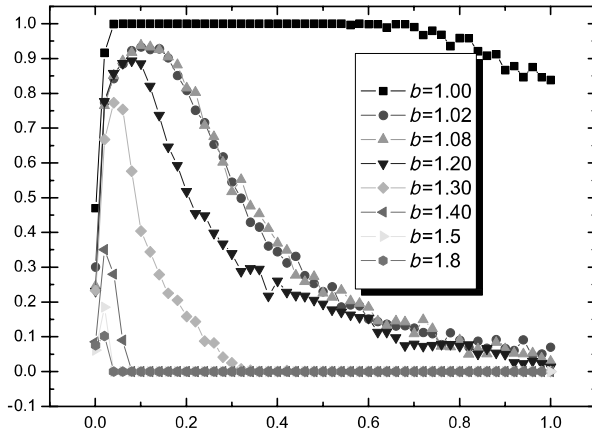


图 2 对应于不同参数 b 时合作水平随着网络异质程度 $\frac{N_h}{N}$ 的变化情况

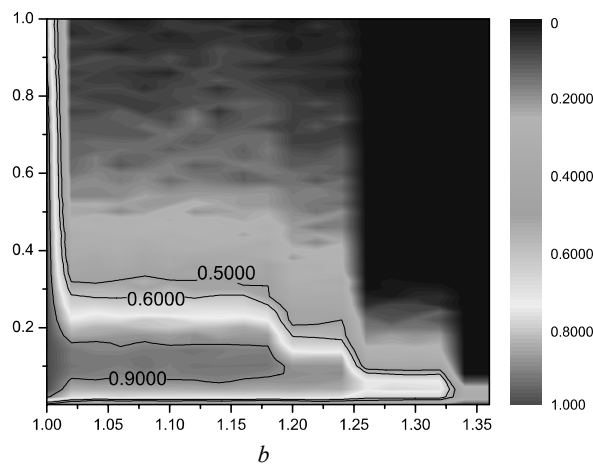


图 3 合作水平在参数空间 $(b, \frac{N_h}{N})$ 中的变化情况

一般在考虑复杂网络上的演化博弈时，初始条件大都设定为合作者 (C) 和作弊者 (D) 等比例地随机分布在网络的顶点上，并认为初始条件中合作者与作弊者比例的变化不会定性地影响结果。然而，当初始条件不是随机分布，而是故意使得一些节点状态为合作者或作弊者，将会影响合作的产生和平衡态时的合作水平。考虑无标度网络上三种初始条件的分布配置情况：(1) C 和 D 随机分布；(2) D 占据网络中度较大的节点；(3) D 占据网络中度较小的节点 [24]。我们首先采用 Barabasi-Albert 无标度网络模型，生成节点数为 $N = 3000$ ，平均度为 4 的网络。然后考虑此网络上的演化囚徒困境博弈问题，更新规则采用

$$W_{s_x \leftarrow s_y} = \frac{1}{1 + \exp[\frac{P_x - P_y}{K}]}, \tag{2.3}$$

其中 P_x, P_y 分别代表节点 x 、节点 y 与邻居博弈所得的总收益. 参数 K 表示节点的理性程度, 这里取 $K = 0.125$. 平衡时合作者的比例取经过 10000 步演化后 1000 步结果的平均值. 所有的数据点对应于 30 次运行结果的平均值. 我们研究了在三种初始条件配置情况下, 平衡时合作水平如何受不同初始条件的影 响问题, 结果如图 4 所示. 可以看出, 配置 (2) 即初始时 D 占据度较大节点, 合作水平受到初始条件 D 的比例大小的显著影响. 同时, 配置 (3) 即初始时 D 占据度较小节点, 合作水平对初始 D 的比例大小最不敏感 (即使在参数 b 较大的情况下). 换言之, 配置 (3) 呈现出了鲁棒合作 (robust cooperation). 图 5 显示了三种配置情况下, 使得合作者最终消失的临界 D 初始比例值随着参数 b 的变化情形. 我们发现对应于同一参数 b , 配置 (3) 的临界 D 的初始比例大于其它两种配置情况, 且配置 (2) 的临界 D 的初始比例最小. 这进一步说明了配置 (3) 最利于合作的产生和维持. 另外, 我们的结果表明, 无标度网络中 hub 及其相互之间的连接对合作的演化起着重要作用. 如果初始时刻度较大的节点都被 D 占据, 合作水平将会受到较大抑制. 相反地, 在配置 (3) 中, 合作者占据了网络中度较大的节点, 由于 BA 无标度网络中度大的节点之间相互连接比较紧密, 合作者可以形成簇结构, 起到相互保护的作用. 同时由于 C 占据 hub, 可以通过策略更新影响周围小节点的策略选择, 因此在这种配置情况下, 合作水平受初始 D 比例大小的影响较小, 并且合作水平较高, 为鲁棒合作提供了有利的条件. 此外, 在配置 (3) 情况下, 以相同的参数 b 和 D 的初始比例, 系统演化到全 C 的速度最快. 我们的结果说明了无标度网络中 hub 结构及 hub 之间紧密连接是合作产生的重要因素.

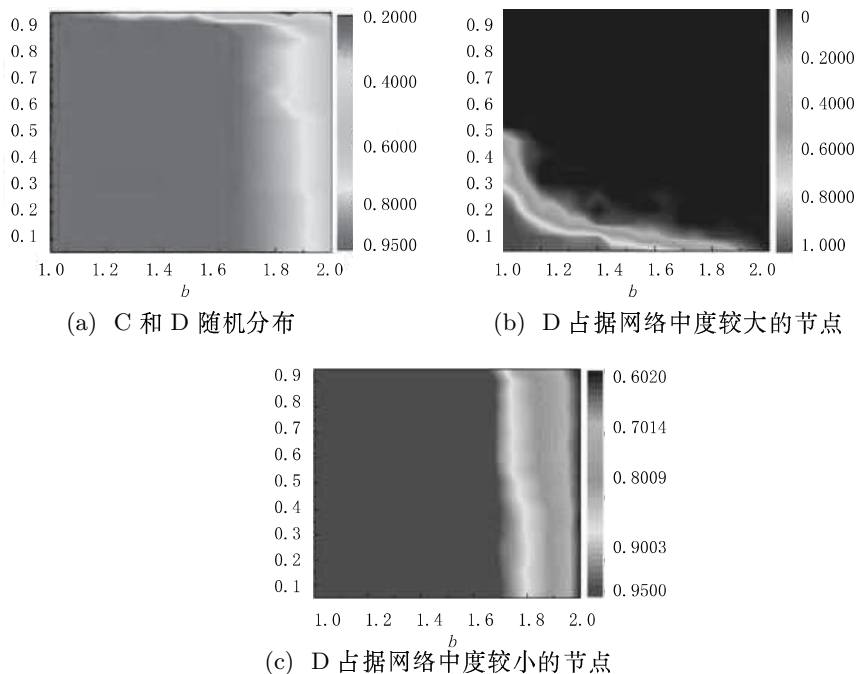


图 4 合作水平在参数空间 (b, f_{1D}) 中变化情形

以上研究都是基于模型网络上的演化博弈. 由于真实的社会网络远比一些模型网络复杂, 而后者通常只是前者的一个特性或某些特性的反映, 因此研究真实社会网络上的演化博弈对揭示合作的演化具有重要意义. 考虑在一个在线好友网络上的演化雪堆博弈问题^[25].

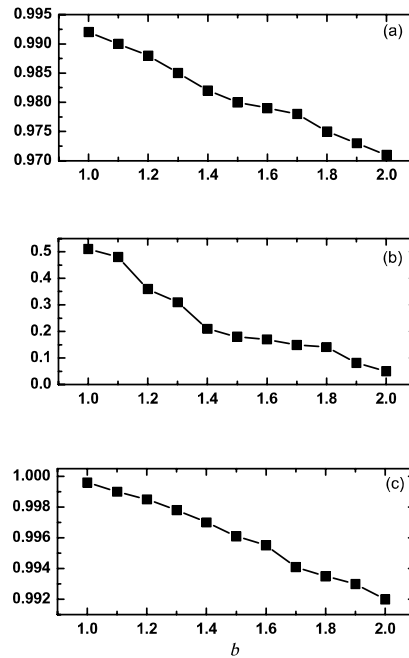


图 5 三种初始配置情况下, 使得合作者消失的临界 D 初始比例值随着参数 b 的变化情形

网络规模为 $N = 9590$, 此网络具有小世界、无标度、层次结构、同配性 (assortative) 等性质. 博弈矩阵可以简化为以相互合作时损益比 (cost-to-benefit ratio) 为单参数的形式:

$$M = \begin{pmatrix} 1 & 1-r \\ 1+r & 0 \end{pmatrix},$$

其中 $0 < r < 1$. 更新规则采用 (2.3) 式, 其中 $K = 0.1$. 初始时刻, 合作者和作弊者等比例地随机分布在网络上, 在经过 10000 步演化后, 再取最后 2000 步结果作平均, 作为平衡时合作者的比例. 所有的数据点对应于 100 次运行结果的平均. 图 6 显示了合作水平随着参数 r 变化的情况, 其中直线代表在无限混合均匀人口中, 由复制动力学预测的平衡时合作者的比例 $1-r$. 可以看出, 真实社会网络中平衡时合作者的比例都明显大于混合均匀时的情形. 但是, 此时的合作水平不如在模型网络 (如 BA 无标度网络) 上高. 一方面, 由于我们考虑的网络是异质的, 能促进合作; 另一方面, 因为真实社会关系网络具有很多模型网络不具有的拓扑性质, 而这些拓扑性质对合作水平的影响是可加的, 因此在这些可加的拓扑性质的影响下, 合作水平不如模型网络上高是可以理解的. 同时, 我们在图 7 中给出了从初始时刻等比例 C 和 D 随机分布的 100 次独立仿真中系统演化到全 C 的比例随着参数 r 变化的情况. 可以发现, 在 $r < 0.5$ 时, 仿真演化到全 C 的比例大约是 50%, 而 $r > 0.5$ 时, 多数仿真结果都演化到大部分节点的状态都是 D 的情形. 这个结果表明, 由于系统有可能演化到节点的状态大部分都是 D 的情形, 尽管在 $r < 0.5$ 的时候, 仿真演化到全 C 的比例大约是 50%, 但经过多次数据平均, 使得总体的合作水平较低. 为了解真实社会网络上合作的产生, 我们应该从一个综合的角度去看待各种网络因素对合作的影响. 我们所考虑的社会网络具有小世界性, 这个性质加强了合作策略在网络上的传播, 因此有利于合作. 另外, 我们的网络具有层

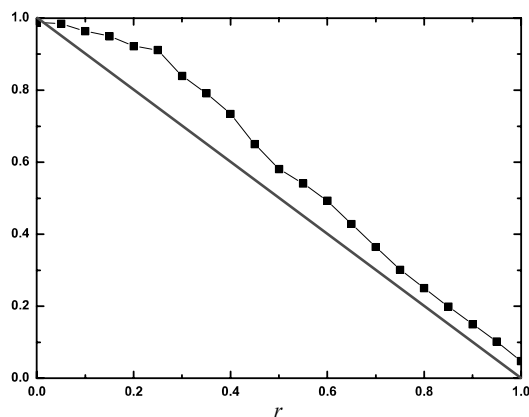


图 6 合作水平随参数 r 的变化情况, 其中直线对应于 $1-r$

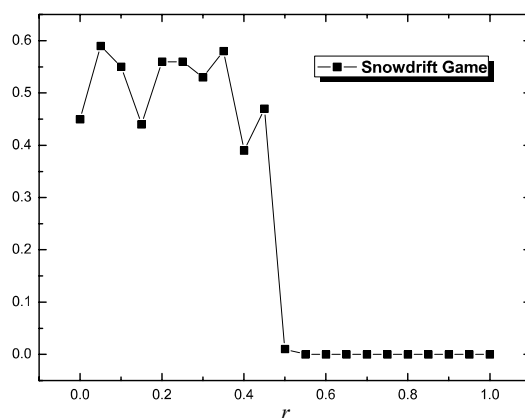


图 7 演化到全 C 的仿真次数比例随参数 r 的变化情形. 每个数据点对应于从初始时刻等比例 C 和 D 随机分布的 100 次独立仿真中系统演化到全 C 的比例

次结构, 即度小的节点以某些度大的节点为中心相互之间紧密连接, 这容易形成合作者的簇结构 (cluster), 合作者与合作者相互帮助, 抵御作弊者的入侵, 从而有利于合作. 但由于我们的网络是同配的, 而非相关网络中合作水平最好, 因此同配性削弱了合作水平. 我们认为理解真实社会网络上的合作演化, 应该综合考虑真实网络的各种拓扑性质对合作水平的影响. 我们的研究结果对理解人类群体社会中的合作行为有一定帮助.

3 网络拓扑与博弈的共同演化

大多数复杂网络上的演化博弈研究都是基于静态网络的, 即网络拓扑从博弈一开始就固定不变了. 而实际上真实网络是动态演化的, 因此所考虑的静态网络只相当于真实网络的一张快照. 如果网络拓扑影响博弈结果, 而博弈结果反过来作用于网络拓扑、调整网络拓扑 (或社会关系), 这种情形更符合实际.

考虑个体带简单记忆的网络拓扑与博弈共同演化的简单模型^[26]. 初始网络从规则随机图开始, 每个节点与其所有邻居连续进行囚徒困境博弈 n 轮, 在每一轮, 节点可依据 (2.3)

式进行策略调整, 同时记下邻居作弊次数. 博弈完 n 轮后, 随机选择 m 个个体进行邻居关系调整. 被选中的个体将把连到作弊次数最多的邻居的边断开, 然后随机重连到该邻居的一个邻居. 参数 n, m 可以看成是博弈动力学和拓扑调整的时间尺度. 在我们的模型中, 策略更新采用同步方式, 拓扑调整是异步的, 因此, 拓扑调整要比博弈动力学缓慢很多, 这与现实是符合的.

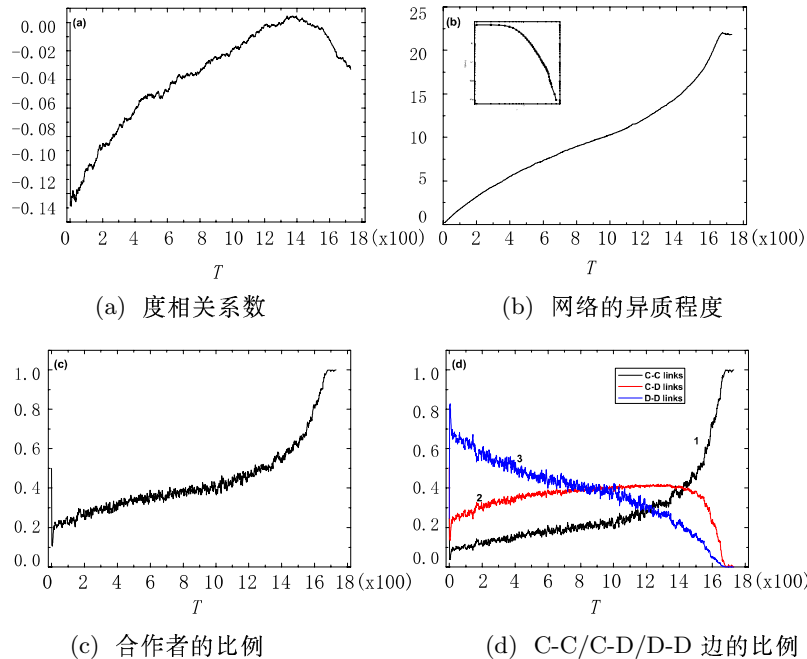


图 8 网络结构随着个体邻居关系调整而变化的情况. (b) 中插图是终态时网络的累积度分布图. $N = 10^4$, $\langle k \rangle = 8$, $b = 1.2$, $n = 6$, $K = 0.02$, $m = 100$

我们在图 8 中给出了网络拓扑随着个体调整邻居关系而变化的过程. 从图 8(a) 可以看出, 演化的网络是异配的, 即度大的节点倾向于与度小的节点相连. 由于我们的拓扑调整规则是断开重连到邻居的邻居, 在拓扑调整中度大的节点易于一般的节点被其它节点搜索连接上, 因此网络呈现出异配性. 同时, 拓扑调整也造成了网络的异质性, 图 8(b) 显示了网络度的方差变化情况. 可以看出, 随着网络的演化, 网络变得越来越异质, 而异质性是利于合作产生的. 因此在拓扑和博弈共同作用下, 合作水平会慢慢增强, 如图 8(c) 所示. 另外, 图 8(d) 给出了网络中 C-C/C-D/D-D 边的比例变化情况. C-C 边数不断增多, 而 C-D 和 D-D 边最终受到抑制而消失. 这说明拓扑调整加强了合作者和合作者之间的同配连接, 削弱了 C-D 和 D-D 之间的连接, 从而使得整个网络向有利于合作者的方向进行演化, 最终使得合作者占上风. 进一步地, 我们在图 9 中给出了对应于不同 b 时, 合作者的比例随着调整拓扑次数 m 的变化结果. 我们发现, 在保持平均度、博弈轮数不变的情况下, 对于固定的 b , 存在调整拓扑次数的临界值 m_c , 当 $m > m_c$ 时, 合作者的比例将会演化到 100%. 同时, 图 9 中的插图给出了固定其它参数时, m_c 随着 b 的变化情形. 我们发现, 随着作弊收益 b 的增加, 必须使调整拓扑次数相应地增加, 才能保证合作者占上风. 我们的结果表明了拓扑和博

弈动力学共同演化是促进合作水平提高的一个重要机理. 此外, 我们还研究了模型中其它参数变化对合作演化的影响 [26].

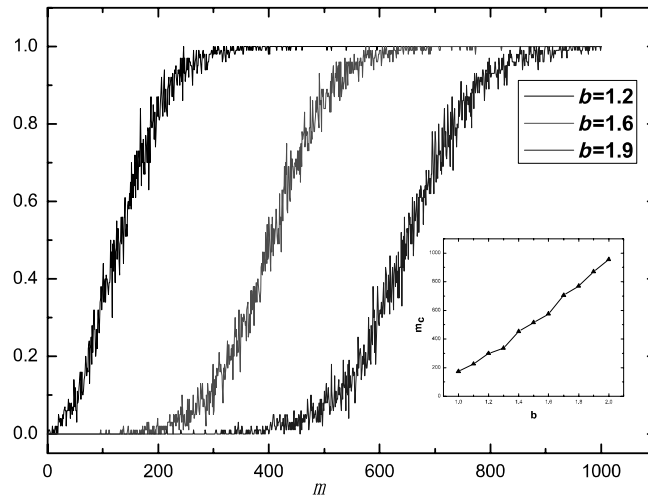


图 9 对应于不同 b 时演化到全 C 的仿真次数比例随着 m 增加的变化情况. 每个数据点对应于从初始条件 C 和 D 等比例随机分布开始的 100 次仿真结果. 插图显示临界值 m_c 随着 b 的变化情形. $N = 10^3$, $\langle k \rangle = 4$, $n=6$, $K = 100$

上面模型中假设个体断开与作弊次数最多的邻居相连的边, 再重连到此邻居的邻居. 事实上, 更为合理的情形是, 个体断边重连时, 既可以与邻居的邻居形成新边, 也可以与除邻居之外的节点相连. 因此基于以上模型, 我们假设个体断边重连时, 以概率 p 连到邻居的邻居, 反之, 以 $1-p$ 的概率随机选择除邻居之外的节点相连. 这里参数 p 的大小表明个体与个体之间产生新边时的“有序性”(order)与“随机性”(randomness)的对比. 当 $p \rightarrow 0$ 时, 个体随机选择除邻居之外的节点产生新边(完全随机性); 当 $p \rightarrow 1$ 时, 个体选择邻居的邻居产生新边(有序性); 当 $0 < p < 1$ 时, 个体重连的新边介于完全随机性与有序性之间. 因此这个假设很好地反映了现实情形中社会网络的演化特点: 通常人们可以通过朋友介绍, 认识朋友的朋友; 也可以偶然地不通过朋友介绍结交一个新朋友. 图 10 给出了合作水平随着参数 p 的变化情形. 有趣地是, 我们发现, 在博弈关系和策略更新共同演化的情形下, 断开旧边产生新边中的“有序性”的倾向越大, 越不利于合作的产生. 换言之, 在共同演化情形下, 断边重连到邻居的邻居并不利于合作. 相反地, 如果随机地选择除邻居之外的节点作为新的博弈对象, 将会有助于合作现象的涌现. 我们知道, 在固定拓扑上的演化博弈中(博弈关系不随博弈结果调整), 如果个体与个体之间联系很紧密的话(个体的邻居与邻居之间有较多边存在), 将会形成合作者的簇, 由于这些合作者之间相互帮助, 将有利于维持合作. 然而, 我们的结果表明当博弈关系和策略更新共同演化的时候, 断边重连时的“有序性”(形成社会关系网络中常见的“三角形”结构)在一定程度上并不十分有利于合作的产生. 这是因为, 当重连到邻居的邻居时, 将会加强合作者与合作者之间的同配连接, 这是有利于合作的; 但是另一方面, 如果重连到除邻居之外的节点, 在一定程度上将会加强合作者与作弊者之间的异配连接, 而由于演化过程中产生的网络异质性, 即合作者变成 hub, 此时重连上的作弊者将会通

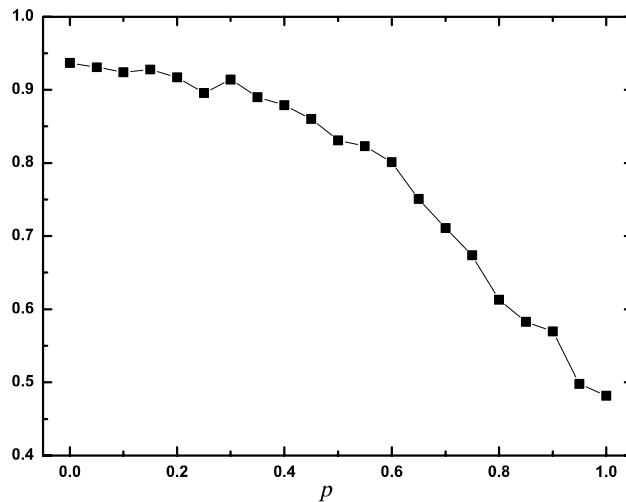


图 10 演化到全 C 的仿真次数比例随着 p 的变化情况. 每个数据点对应于从初始条件 C 和 D 等比例随机分布开始的 10^3 次独立仿真结果. $N = 10^3$, $\langle k \rangle = 4$, $K = 100$, $b = 1.4$, $n = 6$, $m = 50$

过策略更新学习 hub 的策略变成合作者, 因此这进一步加强了合作水平. 在“有序性”与“随机性”相互作用下, “随机性”更能够促进合作的产生和维持较高的合作水平. 我们这一结果对理解人类社会中的合作现象是非常有意义的.

4 囚徒困境中的强化学习

前文所考虑的演化博弈策略更新方式都是个体模仿学习邻居的策略, 而不涉及个体对博弈过程的认知和经验积累. 实际上, 个体可以通过博弈结果不断地调整选择策略的倾向, 即根据历史经验的积累进行归纳学习 (Inductive Learning). 强化学习 (Reinforcement Learning) 是经典二人博弈中常见的个体学习方法^[27]. 其基本思想是个体依据期望水平 (Aspiration Level) 判断采用的策略的结果是奖赏 (Reward) 还是惩罚 (Punishment), 若为奖赏, 则后续行动中采用此策略的倾向增加, 反之, 则减小. 我们考虑在 Newman-Watts 小世界网络上的演化囚徒困境博弈中的强化学习过程. 节点数 $N = 1000$, 平均度 $\langle k \rangle = 42$. 博弈矩阵元素取为

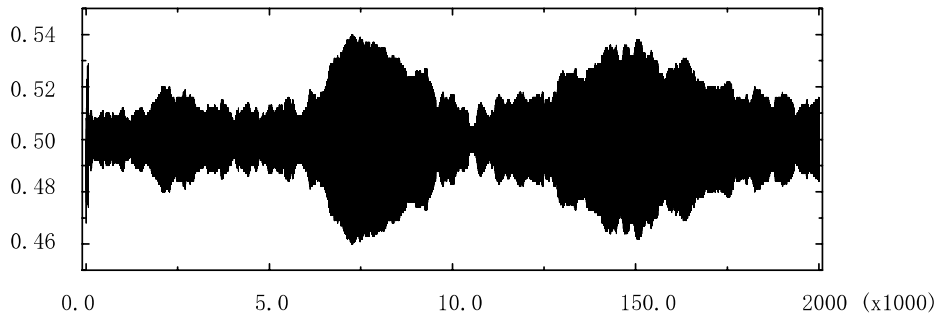
$$T = 5, \quad R = 3, \quad P = 1, \quad S = 0.$$

设个体 i 采用策略 $a \in \{C, D\}$ 的平均收益为 π_a^i , 期望收益为 A_0^i . 从而采用策略 a 进行博弈的满意度为

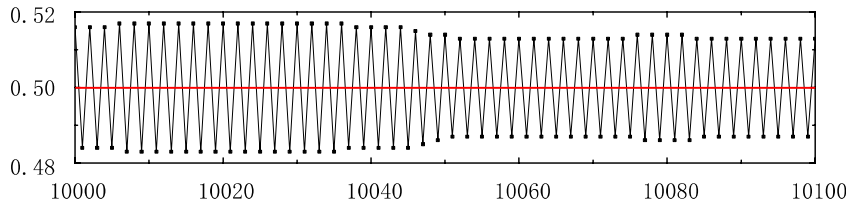
$$s_a^i = \frac{\pi_a^i - A_0^i}{\sup[|T - A_0^i|, |R - A_0^i|, |P - A_0^i|, |S - A_0^i|]}, \quad a \in \{C, D\}, \quad (4.1)$$

其中 $0 \leq s_a^i \leq 1$. 个体根据 s_a^i 进行调整后续行动中采取策略 a 的概率如下

$$p_{a,t+1}^i = \begin{cases} p_{a,t}^i + (1 - p_{a,t}^i) s_{a,t}^i, & \text{若 } s_{a,t}^i \geq 0, \\ p_{a,t}^i + p_{a,t}^i s_{a,t}^i, & \text{若 } s_{a,t}^i < 0, \end{cases} \quad a \in \{C, D\}, \quad (4.2)$$



(a) 0-20000步长上的结果



(b) 10000-10100步长上的结果

图 11 网络中合作者的比例随着时间步长变化的情况, 初始时刻合作者与作弊者等比例随机分布在网络上, $N = 1000, \langle k \rangle = 42, A_0 = 4.5, l = 0.1$

其中 $p_{a,t}^i$ 是个体 i 在 t 时刻选择 a 的概率, 并有

$$p_{a,0}^i = \frac{1}{2},$$

l 是学习率 (learning rate) ($0 < l < 1$), $s_{a,t}^i$ 是个体 i 在 t 时刻采取策略 a 的满意度. 个体选择 $\{C, D\} \setminus a$ 的概率为 $1 - p_{a,t}^i$. 不失一般性, 我们假设所有个体具有相同的期望收益 A_0 . 限于篇幅, 这里我们只讨论 $R < A_0 = 4.5 < T$ 的情况. 这时系统中由于强化学习, 个体之间的博弈过程呈现出比较有趣的结果. 图 11 给出了合作者的比例随着时间演化的结果. 可以看出, 经过较短的时间后, 网络中的合作者的比例将以 0.5 为对称中心振荡, 即个体保持当前策略的概率为零, 将会在策略 C,D 之间来回切换, 出现有趣的拉锯现象 (tug of war). 当个体的期望水平 A_0 较高时, 系统就会出现类似的现象. 下面我们利用平均场理论简单地定性分析这一结果.

由于采用的网络是相对比较均质的 Newman-Watts 小世界网络, 平均度较大, 因此可以用平均场估计进行定性分析. 设 t 时刻系统中合作者的比例为 $\rho_c(t)$, 个体保持当前状态的概率为 $p_a(t)$, 即 C 个体保持当前状态的概率为 $p_c(t)$, D 个体保持当前状态的概率为 $p_d(t)$. C 个体的平均收益为

$$U_c = R\rho_c + (1 - \rho_c)S,$$

D 个体的平均收益为

$$U_d = T\rho_c + (1 - \rho_c)P.$$

由 $T > R > P > S$, 知 $U_d > U_c$. 当 $A_0 = 4.5$, 系统从等比例 C,D 开始演化时, 易知不等式 $A_0 > U_d > U_c$ 总是成立, 于是有 $s_c < s_d < 0$. 按照上文所介绍的强化学习规则 (4.2) 式, 个

体保持当前状态的概率将会趋于零, 这是由于

$$0 \leq \dots < p_a(t+1) < p_a(t) < \dots < p_a(0),$$

即

$$p_a(t) \rightarrow 0, \quad \text{当 } t \rightarrow \infty.$$

另外合作者比例 $\rho_c(t)$ 满足速率方程

$$\frac{d\rho_c(t)}{dt} = (1 - p_d(t))(1 - \rho_c(t)) - (1 - p_c(t))\rho_c(t), \quad (4.3)$$

其中方程右边第一项表示系统中 D→C 的比例, 第二项表示 C→D 的比例. 当 $p_a(t) \rightarrow 0$ 时, 令

$$\frac{d\rho_c(t)}{dt} = 0,$$

系统存在一个稳定平衡点 $\rho_c^* = \frac{1}{2}$. 而在我们的仿真结果中 $\rho_c(t)$ 时而趋向 0.5, 时而又偏离 0.5, 并不是收敛到 0.5. 这是由于我们采用的网络并不是理想地满足平均场效应成立条件, 系统的行为还受到个体策略选择的随机性等不确定因素的影响. 但是以上定性分析在一定程度上与仿真结果能很好的相吻合.

5 结论与展望

复杂网络上的演化博弈研究是近年来随着复杂网络研究兴起而逐渐引起关注的一个重要研究领域. 目前大部分工作都集中在囚徒困境博弈或雪堆博弈研究上, 其它类型的博弈还缺乏系统地研究. 因此有必要进一步考虑多人博弈的情形, 如公用品博弈 (Public Goods game) 或多策略的博弈^[28], 如石头 - 剪刀 - 布 (Rock-Scissors-Paper) 博弈. 近来一些学者开始关注这些问题, 得到了一些有趣的结果.

目前大多数工作只是一些数值仿真结果, 由于数学工具的不足, 对复杂网络上的博弈动力学进行解析分析是非常困难的. 目前的一些近似方法, 如平均场方法、对估计方法对异质程度很大的网络很有可能失效. 因此, 寻求有效的数学工具, 探求更好的理论结果, 将一些数值结果命题化、严格化、一般化, 将是十分有意义的.

在演化网络上的博弈动力学还没有得到充分的研究. 网络拓扑和博弈的共同演化将是一个非常具有前景的研究方向. 另外, 还应对个体的学习、记忆等能力上进行更为合理的描述, 使得模型能更好地反映现实, 用来解决一些社会、经济、工程中的公开问题. 对于合作机制的研究依然是演化博弈研究中的一个重要方向. 目前演化博弈主要集中在对合作行为的研究上^[1]. 除此之外, 还可以考虑复杂网络上的其它动力学行为, 应用演化博弈的思想, 解决一些实际问题, 如在“路由设计”, “病毒扩散”, “生物进化”, “控制系统设计”^[29], “市场经济行为”, “信息传播”^[30]等问题上做进一步的探索将是十分有意义的.

近年来有关复杂系统和复杂性科学的研究蓬勃发展^[31], 这对人们原先的还原论的认识带来革命性的冲击. 物理学家 Hawking 说: 21 世纪将是复杂性科学的世纪. 复杂网络上的演化博弈可作为研究复杂系统、复杂性科学一个可行的切入点, 并将会在生态演化、神经网络、群体智能、认知科学、自组织涌现行为、网络化系统、经济动力学、社会心理学等研究中彰显它的作用^[32].

参 考 文 献

- [1] Nowak M A. Five rules for the evolution of cooperation. *Science*, 2006, **314**: 1560–1563.
- [2] Axelrod R. *The Evolution of Cooperation*. New York: Basic books, 1984.
- [3] Hofbauer J and Sigmund K. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press, 1998.
- [4] Smith J M. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1982.
- [5] Albert R and Barabasi A L. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 2002, **74**: 47–97.
- [6] Szabo G and Fath G. Evolutionary games on graphs. *Phys. Rep.*, in press. [arXiv:cond-mat/0607344].
- [7] Doebeli M and Hauert C. Models of cooperation based on the Prisoner’s Dilemma and the Snowdrift game. *Ecology Letters*, 2005, **8**: 748–766.
- [8] Nowak M A and Sigmund K. Evolutionary dynamics of biological games. *Science*, 2004, **303**: 793–799.
- [9] Nowak M A and May R M. Evolutionary games and spatial chaos. *Nature*, 1992, **359**: 826–829.
- [10] Nowak M A and May R M. The spatial dilemmas of evolution. *International Journal of Bifurcation and Chaos*, 1993, **3**: 35–78.
- [11] Hauert C and Doebeli M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, 2004, **428**: 643–646.
- [12] Szabo G and Hauert C. Phase transitions and volunteering in spatial public goods games. *Phys. Rev. Lett.*, 2002, **89**: 118101.
- [13] Tomassini M and Luthi L, Giacobini M. Hawks and doves on small-world networks. *Phys. Rev. E*, 2006, **73**: 016132.
- [14] Ohtsuki H and Hauert C, Lieberman E and Nowak M A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 2006, **441**: 502–505.
- [15] Zimmermann M G and Eguiluz V M. Cooperation, social networks, and the emergence of leadership in a prisoner’s dilemma with adaptive local interactions. *Phys. Rev. E*, 2005, **72**: 056118.
- [16] Santos F C, Pacheco J M and Lenaerts T. Cooperation prevails when individuals adjust their social ties. *PLOS Computational Biology*, 2006, **2**: 1284–1291.
- [17] Pacheco J M, Traulsen A and Nowak M A. Active linking in evolutionary games. *J. Theor. Biol.*, 2006, **243**: 437–443.
- [18] Chen X J, Fu F and Wang L. Prisoner’s dilemma on community networks. *Physica A*, 2007, **378**: 512–518.
- [19] Li C G and Maini P K. An evolving network model with community structure. *J. Phys. A: Math. Gen.*, 2005, **38**: 9741–9749.
- [20] Santos F C and Pacheco J M. Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.*, 2005, **95**: 098104.
- [21] Santos F C, Rodrigues J F and Pacheco J M. Graph topology plays a determinant role in the evolution of cooperation. *Proc. Royal Soc. London B*, 2006, **273**: 51–55.
- [22] Santos F C, Pacheco J M and Lenaerts T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc. Natl. Acad. Sci. (USA)*, 2006, **103**: 3490–3494.
- [23] Fu F, Liu L H and Wang L. Evolutionary Prisoner’s dilemma game on heterogeneous Newman-Watts small-world network. *Euro. Phys. J. B*, 2007, **56**: 367–372.
- [24] Chen X J, Fu F and Wang L. Influence of initial distributions on robust cooperation in evolutionary Prisoner’s dilemma. arXiv:physics/0701318.
- [25] Fu F, Chen X J, Liu L H and Wang L. Social dilemmas in an online social network: the structure and evolution of cooperation. *Phys. Lett. A*, in press. [arXiv:physics/0701323].
- [26] Fu F, Chen X J, Liu L H and Wang L. Promotion of cooperation induced by the interplay between structure and game dynamics. *Physica A*, 2007, in press. [doi:10.1016/j.physa.2007.04.099].

- [27] Macy M W and Flache A. Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. (USA)*, 2002, **99**: 7229–7236.
- [28] Wang J, Fu F, Wang L and Xie G M. Evolutionary game dynamics with three strategies in finite populations. arXiv:physics/0701315.
- [29] Vincent T L and Vincent T L S. Evolution and control system design. *IEEE Control Systems Magazine*, 2000, **20**: 20–35.
- [30] Liu L H, Fu F and Wang L. Information propagation and collective consensus in blogosphere: a game-theoretical approach. arXiv:physics/0701131.
- [31] Barabasi A L. Taming complexity. *Nature Physics*, 2005, **1**: 68–70.
- [32] Wang L and Fu F. Collective behaviors and self-organizing cooperation. Technical Report 2006-001, Center for Systems and Control, Peking University, Beijing, 2006.

EVOLUTIONARY GAMES AND SELF-ORGANIZING COOPERATION

Wang Long Fu Feng Chen Xiaojie Chu Tianguang Xie Guangming

(*Intelligent Control Laboratory, Center for Systems and Control,
College of Engineering, Peking University, Beijing 100871*)

Abstract Evolutionary games on complex networks are considered in this paper. First, the evolutionary Prisoner's dilemma game on scale-free networks with community structures is investigated. Then the heterogeneity's role in the evolution of cooperation on a variant of Newman-Watts small-world networks is explored. The influence of different initial conditions on the evolution of cooperation corresponding to different initial configurations for cooperators and defectors distributing among the vertices of networks is also studied. Moreover, Snowdrift game on an empirical social network is considered, and the entangled dynamics of the evolution of network structure and strategy is investigated. The reinforcement learning dynamics in networked populations playing Prisoner's dilemma is also explored. Finally, some open problems, future research directions, and possible application areas of evolutionary games on complex networks are presented.

Key words Collective behaviors, cooperation, evolutionary games, complex networks, Prisoner's dilemma, Snowdrift game, reinforcement learning, self-organization, topology, dynamics.