ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, No.12, December 2007, pp.3071−3079
DOI: 10.1360/jos183071

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种网络自组织演化的数学模型*

董 攀+, 朱培栋, 卢锡城

(国防科学技术大学 计算机学院,湖南 长沙 410073)

A Mathematical Model for Network Self-Organized Evolvement

DONG Pan+, ZHU Pei-Dong, LU Xi-Cheng

(School of Computer Science, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: Phn: +86-731-4512376, E-mail: pandong@nudt.edu.cn

Dong P, Zhu PD, Lu XC. A mathematical model for network self-organized evolvement. *Journal of Software*, 2007,18(12):3071−3079. http://www.jos.org.cn/1000-9825/18/3071.htm

Abstract: This paper develops a self-organized dynamic network model based upon the network's self-organization natures. In this new model, network behavior is treated with nodes' trade-off between the value of information and the cost of establishing link, and the evolvement of network is described as a convergent stochastic process. The paper gives a detailed deduction for the possible result of network evolution. It should be pointed out that PGP (pretty good privacy) certificate network is a good example of this model. Furthermore, according to the change of parameters of the model, the self-organized evolvement exhibits multiform possible results. This phenomenon is consistent with self-organized criticality theory. The work provides a new method for the topological model research and self-organization theory in computer network.

Key words: network; self-organization; evolvement; model; PGP (pretty good privacy)

摘 要: 根据计算机网络本身所具有的自组织特点,提出了网络自组织演化的数学模型.在该模型中,网络行为表现为节点对信息价值的追求以及维护网络连接所付出代价的权衡.模型将网络的演化表述为一个收敛的随机过程.对一种简单的信息网络进行了具体建模和演化结果的数学证明,进而给出了这种网络的一种实例——PGP(pretty good privacy)证书网络.针对 PGP 证书网络实例,根据参数的改变对其自组织演化的其他可能结果进行讨论,最后指出这些结果和自组织临界理论是一致的.该模型可以为计算机网络的拓扑模型研究以及网络自组织理论研究提供一种新方法.

关键词: 网络;自组织;演化;建模;PGP(pretty good privacy)
中图法分类号: TP393 **文献标识码**: A

近年来,互联网中一些新规律的发现引起了广泛关注,例如,网络连接的无尺度(scale-free)特性、小世界

* Supported by the National Natural Science Foundation of China under Grant Nos.60573136, 60673169 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2005AA121570 (国家高技术研究发展计划(863)); the National Basic Research Program of China under Grant No.2005CB321801 (国家重点基础研究发展计划(973)); the Fund of National Laboratory for Modern Communications of China under Grant No.51436050605KG0102 (现代通信实验室基金)

Received 2006-02-22; Accepted 2006-08-29

(small world)特性等等.这些规律揭示了网络中的某种“序”,它不受人类的支配和影响.更为有趣的是,这些规律并不仅仅在互联网中存在,它们是在研究人类社会关系时被首先发现的,在自然界也广泛存在.我们认为,这些规律反映了系统演化的“自组织”特性——一种不受主观因素影响的演化特性.

自组织现象在生活中普遍存在.例如自然界中,鱼类自组织为一种结构性非常好的集群游动,萤火虫以同步方式发光.在经济学、人口动力学、心理学以及大脑机理等领域都能发现自组织现象的例子[1].所有这些例子中,参与者建立一种组织结构但不需要任何集中调控,因此它们是自组织的.这些自组织系统一般有很好的灵活性、自适应性、容错性、可扩展性.对于自组织规律的研究被统一归入了自组织理论,该理论起源于 Prigogine 对热力学研究的成果——耗散结构(dissipative structure)理论[2],之后在系统学领域得到了长足的发展,在社会学、生物学、天文学等方面得到广泛应用.

网络作为一个复杂的开放系统,符合自组织系统的典型特征[3].实际上,网络自组织演化反映了网络作为一个复杂系统,在内在机制驱动下,自行从简单向复杂、从粗糙向细致方向发展,不断地以局部优化达到全局优化的过程.因此,研究网络的各种自组织演化规律和自组织特性也是网络发展的必然需求和有力途径.对网络自组织演化规律的揭示不仅有助于把握网络的拓扑结构、演化过程,避免不必要的投入,有助于摆脱当前越来越复杂的网络系统所遇到的困境,必将对网络的发展起到深远影响.本文对计算机网络的自组织演化行为建立了抽象的数学模型.模型将网络行为定义为节点对信息价值和连接代价的权衡,将网络的演化表述为一个收敛的随机过程.我们对可能的演化结果进行了数学推导和证明.针对 PGP(pretty good privacy)证书网络实例,我们根据参数的改变对其自组织演化的其他可能结果进行了讨论,最后指出这些结果与自组织临界理论是一致的.

1 与网络演化相关的研究成果

- 随机图理论和 ER 模型

对网络建模的直观数学工具是图论,鉴于具有复杂拓扑结构的网络通常表现出随机性,随机图理论[4]成为首个对复杂网络研究的有力工具.随机图理论的主要模型是由 Erdos 和 Renyi 共同提出的 ER 模型[5].ER 模型将含有 N 个节点、但具有任意边的图作为概率空间,研究当 $N\to\infty$时图的性质.在数学中,构造随机图的过程通常被称为演化:从 N 个孤立顶点出发,通过随机加边来形成图.随机图理论的主要目标是确定在何种连接概率下,随机图将产生某些特定的性质.临界现象是随机图理论的重要发现,也就是说,从可能具有某种性质到可能不具有某种性质的变化非常突然,这与物理学中的渗流理论极为相似.在随机图理论中还定义了直径、连通性、度分布、集群系数等在各种网络模型中广为使用的概念.

- 小世界(small world)模型

小世界网络模型[6]是 Watts 和 Strogatz 在 1998 年提出的基于人类社会关系的网络模型.模型中主要的“小世界”特征是节点之间的平均距离随远程连接的个数增加呈指数级下降,对于规则网络,平均距离 L 的平方同 N 成比例;而对于小世界网络,$L\sim\ln(N)/\ln(K)$,小世界概念描述了这样一个事实:在大多数网络中,尽管其规模很大,但任意两个节点间总有一条相当短的路径.小世界最为通用的表现形式就是由社会心理学家 Milgram 在 1967 年提出的“六度分离”概念[6].例如,对于一个千万人口的城市,人与人的平均接触距离为 6 左右.另外,小世界网络具有相对较高的集群系数.一些研究表明,,万维网、PGP 证书网络等都很符合小世界模型.

- 无尺度(scale free)网络

网络中各节点一般有不同的边数(节点度).用分布函数 $p(k)$来表示节点度的分布.它给出了一个任意节点正好有 k 条边的概率.在随机图中,度分布服从泊松分布.但研究者发现,许多大型网络(包括万维网、国际互联网)的度分布不是泊松分布,而是服从幂律 $p(k)\sim k^{-r}$,它们被称为无尺度网络[7].网络的幂律度分布现象是由 Baralasi 和 Albert 在 1999 年首先观察得到的,这个发现极大地改变了我们对复杂系统的认识.无尺度直观地体现为网络被少数连接度较大的节点所支配,有些节点甚至具有数不清的连接,而且不存在代表性的节点.研究发现,无尺度特性源于众多实际网络所共有的两种生成机制:增长和择优连接.这种网络还具有一些独特的性质.例如,对意外故障具有惊人的承受力,但面对协同式攻击时则很脆弱.无尺度规律的认识会帮助我们解决一系列重要问

题,包括开发更好的药物、防止黑客侵入互联网、阻止致命流行病的传播等等.

2 网络自组织演化的数学建模

我们建立模型的主要目的是:1) 对真实网络进行合理精炼,反映本质规律;2) 能够运用数学工具对模型进行分析和演化证明;3) 最终对网络的发展情况进行解释、预测或指导.我们的具体作法是,首先根据网络的基本元素特征和基本自组织规律建立通用的模型,再针对每种实际网络的特点对模型进行丰富,最后进行数学推导和证明.

2.1 建模思想

首先,3 种网络演化理论(随机图、小世界和无尺度)中一些思想值得借鉴,例如:

- 随机图理论中网络图的概率空间思想和演化概念;
- 小世界网络模型中连接参数反映了节点选择连接的某些非随机性——或者说选择性.小世界现象不仅由拓扑决定,同时也与节点位置相关;
- 无尺度模型中关于网络的动态特性、增长特性和择优连接特性.

第二,对计算机网络进行合理的抽象和简化.从价值论的观点出发,与某个节点建立连接就是为了获取有价值的信息,但建立连接和维护连接也必须付出代价.

第三,网络演化的“利益驱动论”.我们认为,在任何网络中,节点建立一条连接都是为了追求一定的“利益”(这符合人或其他生物天性自私的一面).建立连接则要付出一定的“成本”.节点只有认为在建立连接后,它所获得的利益高于其付出的成本,才有可能真正地建立这条连接.连接的维护也要付出成本,除非该连接被撤销.网络的演化就表现为节点为追求利益的最大化,而不断增加或撤销连接.

第四,对自组织系统特点的借鉴.自组织系统是指在内在机制的驱动下,自我优化、自我发展的系统.注意,一个自组织系统的前提是在系统级上没有外力的施加,不存在全局的控制,然而系统却能走向全局的优化.通过对典型自组织系统的研究,人们将自组织机制分解为相对独立的若干方面[8],在本节的建模中,我们仅用到以下几个方面:信息共享、单元自律(即网络中的节点具有独立决策的能力)、微观决策(即各节点的决策只关乎它自己的行为)、迭代趋优(网络在反复迭代中不断趋于优化).

考虑以上自组织网络的特性和机制,基本模型建立的主要出发点为:

1) 模型借鉴随机图的思想,也就是说,节点建立连接是有一定随机因素的,网络的演化表现为一定的随机过程;

2) 各节点的随机性相互独立,即没有全局控制者节点能主观影响其他节点;

3) 节点具有主观能动性,并以共同的价值标准衡量建立连接的收益得失,以追求节点自身利益最大化的方式进行决策(建立或撤销连接);

4) 网络的全局优化表现为全网节点利益之和的最大化;

5) 网络迭代就是网络演化,数学表示为随机过程,且此随机过程收敛的结构就是网络演化的最终结构;

6) 网络的自组织目标是在节点追求自身利益最大化的过程中使网络向着全局优化的方向演化.

2.2 模型假设

为了简化讨论,本文在模型中假设网络节点的个数有限,但对节点之间建立连接的能力不作限制,也就是说,节点相关联的边可以很多,并且节点的计算能力可以完全满足通信的需求.

2.3 数学模型

本节首先给出一个尽可能简化的网络模型,这个模型事实上具有一定的广泛性,可作为一些计算机网络(如互联网)的抽象.本节所用到的主要证明方法来源于文献[9],该文献对社会交流网络的自组织现象进行了建模和数学推导.我们将文献[9]的思想移植到计算机网络,并对其中的数学推导和证明进行合理简化和改进.鉴于现在的研究还不够深入,本文中结论与实例网络的符合程度还有待进一步考察,这点将在第 4 节中加以讨论.

设 N 是网络节点集,i 和 j 是其中的节点,节点从 1~n 进行编号.每个节点拥有私有信息,其价值 $V>0$.节点 i 可以通过建立到节点 j 的单向连接(指向节点 i)获得节点 j 的信息.每个连接的建立成本为 $c>0$.每个节点可与任何其他节点建立连接,为简化起见,两个节点之间不允许有重复的有向连接.网络符合下述特点:当节点 i 建立到 j 的连接后,i 可以获取 j 的所有信息,包括 j 通过连接其他节点获得的信息.也就是说,信息可以多级传递并且在传递过程中不会衰减.以上的简化模型主要是出于分析和推导的需要,即使模型已经如此简化,形式化描述和推导仍然非常困难.

- 节点的连接策略:在演化的任何时刻,节点 i 的连接状况可以用向量 $g_i=(g_{i,1},\ldots,g_{i,i-1},g_{i,i+1},g_{i,n})$表示,其中,$g_{i,i}\in\{0,1\}$且 $i\neq j.g_{i,i}=1$ 表示节点 i 建立了与 j 的连接,反之表示该连接没有建立.我们称这个向量为 i 的(连接)策略.节点 i 的策略集合可以表示为 G_i,易知其势为$|G_i|=2^{n-1}$.全网节点的策略空间可以表示为 $G=G_1\times\ldots\times G_n$.某时刻的网络状态可以表示为向量 $g=(g_1,g_2,\ldots,g_n)$.
- 节点行为表现:当 i 建立了到 j 的连接后,i 消耗了大小为 c 的成本,同时从 j 获取一定量的信息.i 会计算它可获得的所有信息的价值总和以及所有连接的成本总和,继而得到当前总收益.此后,i 根据网络的变化计算其他策略下的收益,在一定概率下变更策略以使其利益最大化.
- 网络的演化过程表现为:以一定的时间间隔,节点修正连接策略.在修正时,节点选择一个对于当前网络的最佳策略.在将时间轴划分得足够细后,可以假定在某个时刻最多只有 1 个节点改变策略,其他节点保持惯性.如果最优策略有多个,它将从中随机选择一个.这种修正过程会产生一个网络状态空间中的马尔可夫链.通过分析网络演化过程,可以澄清其中的自组织概念.我们说一个网络演化过程具有自组织性,是指从初始网络开始的马尔可夫链在有限时间内以较大概率收敛到一个拓扑稳定的网络.

为了能使用数学工具对演化进行推导,有必要先对模型进行形式化,下面对用到的定义和符号进行集中说明:

1) $N_d(i;g)=\{k\in N|g_{i,k}=1\}$,即 $N_d(i;g)$表示 i 与之建立连接的节点集;
2) $j\xrightarrow{g}i$:表示在网络 g 中存在从 j 指向 i 的路径;
3) $d(i,j;g)$:表示在 g 中从 j 到 i 的距离,即最短路径的有向边数;
4) $N(i;g)=\{k\in N\mid k\xrightarrow{g}i\}\cup\{i\}$,表示能够到达 i 的节点集合,包括 i 本身;
5) g_i:i 的一种连接策略;
6) g_{-i}:表示在 g 中将节点 i 建立的所有连接删去后得到的网络,$g=g_i\oplus g_{-i}$;
7) g_{-i}:表示将 g_{-i} 中所有 $g_{j,i}=1$ 替换为 $g_{j,i}=0$ 得到的网络;
8) $\Pi_i(g)=|N(i;g)|V-|N_d(i;g)|c$,表示 i 在 g 中的收益;
9) $W(g)=\sum_{i\in N}\Pi_i(g)=\sum_{i\in N}|N(i;g)|V-\sum_{i\in N}|N_d(i;g)|c$,表示网络的整体效益(全部节点收益之和).

定义 1. 称 g_i 是节点 i 对于网络 g_{-i} 的最佳响应策略,若 $\Pi_i(g_{-i}\oplus g_i)\geq\Pi_i(g_{-i}\oplus g_i')$ 对所有 $g_i'\in G_i$ 成立.令 $BR_i(g)$表示 i 对 g 的所有最佳响应组成的集合.

定义 2. 称网络 $g=(g_1,\ldots,g_n)$是强支撑(sustainable)的,如果对每个 $i\in N$ 都有 $g_i\in BR_i(g)$.对于给定的 V 和 c,令 $S(V;c)$表示所有强支撑网络的集合.

定义 3. 称 g 为最大效益网络,如果 $W(g)\geq W(g')$,对所有 $g'\in G$ 成立.

定义 4. 称节点集 $E\subset N$ 为 g 的一个组件,指对所有 $i,j\in E$ 且 $i\neq j$ 都有 $j\xleftrightarrow{g,E}i$.一个组件 E 称为最大的,如果不存在 E 的严格超集 $E'\subset N$ 也是 g 的组件.

定义 5. 称含有有向环子图的网络为超环.

鉴于目前信息网络的特点,我们可以假定网络 g 以极大概率演化为全连通的,且一般情况下 $V>c>0$(此时,节点才具有互联的动机).

为了形式化地描述网络演变,令 G_{-i} 代表除了节点 i 之外的所有节点的策略空间,对于给定集合 A,令 $\Delta(A)$代表在 A 上的所有概率分布.设对于每个节点 i 都存在 $p_i\in(0,1)$和函数 $\phi_i:G\to\Delta(G_i)$满足

$$\phi_i(g)\in Interior\,\Delta(BR_i(g_{-i})),\forall g_{-i}\in G_{-i}.$$

对于$\phi_i(g)$支撑中的$\hat{g}_i$,$\phi_i(g)(\hat{g}_i)$代表在概率测度$\phi_i(g)$下$\hat{g}_i$的概率.若 $t\geq 1$ 时刻的网络表示为$g^t=g^t_{-i}\oplus g^t_i$,则节点 i 在 t+1 时刻的策略可以用下式给出:

$$g_i^{t+1}=\begin{cases}\hat{g}_i, & \text{以概率} p_i\times\phi_i(g)(\hat{g}_i)\\ g_i^t, & \text{以概率} 1-p_i\end{cases}.$$

上式表明,节点 i 以概率 $p_i\in(0,1)$选择最佳响应策略.函数ϕ_i表示了 i 在多于一个最佳响应策略时选择的随机性.i 还以概率 $1-p_i$ 保持了一定惯性,维持当前的连接策略.如果将时间轴分割得足够细,任意两个节点不在同一时刻变动其连接策略.由于节点的行为相互独立,可以从节点的选择导出一个转移矩阵 T,将状态空间 G 映射到 G 上的所有可能的分布集$\Delta(G)$.令$\{X_t\}$为从初始网络 g 出发,按照上述转移矩阵形成的马尔可夫链,则$\{X_t\}$描述了网络基于节点行为演化的动态性.

由此我们可以得到一个网络自组织的形式化定义:

定义 6. 给定一个初始网络 g,如果随机过程$\{X_t\}$以较大概率收敛到一个拓扑形式固定的终态网络,则称$\{X_t\}$是自组织的.

3 网络演化的数学推导和证明

自组织的概念很引人注目,因为它表明:即使节点各自追求私有利益,网络仍然可以在有限时间内达到一种稳定状态.下面给出这种自组织过程收敛性的推导过程.

引理 1. 给定网络 g 和 $i\in N$,则存在唯一的对 $N\backslash\{i\}$的划分,每个划分是 g'_{-i} 的最大组件 $E_1,E_2,\ldots,E_m$,划分得到的集类以 $\boldsymbol{E}$ 表示.

为了方便表达,我们将'$\xrightarrow{g}$'关系由节点扩展到组件之间.给定两个组件 E 和 E',则 $E\xrightarrow{g}E'$ 当且仅当 $\forall j\in E$ 和$\forall j'\in E'$都有 $j\xrightarrow{g}j'$.其实,由组件的概念不难看出,此定义中的∀可换为∃.进而应该指出,组件中的'$\xrightarrow{g}$'是一种偏序关系.我们定义 $T\subset E$ 为偏序意义下的最大组件构成的集类.也就是说,对于 $E\in T$ 不存在 $E'\subset E$,使得 $E\xrightarrow{g}E'$.设 $B_1\subset E$ 包含所有 E 中在偏序'$\xrightarrow{\bar{g}}$'意义下最小的元素,也就是说,在不存在 $E'\in E$ 使得 $E'\xrightarrow{\bar{g}}E$ 时,$E\in B_1$.所以,B_1 包含了"底层"的组件,其中,节点没有建立任何到其他组件节点的连接.利用递归,通过 B_p 可以定义 B_{p+1}:

$B_{p+1}=\{E\in E\backslash E_p\mid \exists E'\in B_p \text{ s.t } E'\xrightarrow{\bar{g}}E,\text{且 不存在}E''\in E\backslash E_p \text{ s.t } E''\xrightarrow{\bar{g}}E\}$,这里,$E_p=\cup_{1\leq q\leq p}B_p$.可以通俗地说,$B_{p+1}$是将其前面所有 B_p 中包含节点删除后所得网络的底层组件集合.

引理 2. 给定网络 g,对于 $i\in N$,设 $g_i\in BR_i(\bar{g})$,这里 $\bar{g}=g'_{-i}$.令 K 是非空节点集并满足对所有 $k\in K$ 有 $g_{i,k}=1$.如果 $\hat{k}$ 是对所有 $k\in K$ 满足 $k\xrightarrow{\bar{g}}\hat{k}$ 的节点,则满足式(1)的策略 $\hat{g}_i$ 也是节点 i 的最佳响应策略.

$$\hat{g}_{i,j}=g_{i,j}\quad \text{for all } j\notin K\cup\{\hat{k}\},\ \hat{g}_{i,k}=0\ \text{ for all } k\in K,\text{且 } \hat{g}_{i,\hat{k}}=1 \tag{1}$$

引理 1 和引理 2 对我们主要结论的证明非常有用,它们的证明过程见文献[9].

为了证明随机过程$\{X_t\}$是收敛的,只需证明网络在有限次的迭代之后会演化到一个稳定的拓扑状态.下面我们给出这个主要结论及其证明.

定理 1. 设 $V>c>0$,网络 g 以大概率演化为有向环形拓扑.

证明:设 n 是组件 $\boldsymbol{T}$ 中的节点.设在组件 $E_1,\ldots,E_r$ 和 E 之间存在'$\longrightarrow$'关系,令 $k_1,\ldots,k_r$ 分别是 $E_1,\ldots,E_r$ 中的节点,则 n 将当前策略中 $g_{n,k_1},\ldots,g_{n,k_r}$ 置为 1,其他连接不变,所得到的策略是其最佳响应策略(如图 1 中粗箭头线表示新的连接).因此,$N(n;g)=N$ 以较大概率出现.

设 $\boldsymbol{B}_1\subset\boldsymbol{E}$ 包含集合$\{B_1^1,\ldots,B_1^{q_1}\}$.根据假设,每个 $B_1^k\in\boldsymbol{B}_1$ 仅包含一个节点.对应节点 B_1^1 为 j_1^1.考虑 j_1^1 的最佳响应.由于 $V>c$,有 $N(n;g^1)=N$,也就是说,节点 n 能得到 N 中的每个节点所拥有的信息.这样,如果 $k\in N\backslash\{j_1^1,n\}$,则 $k\xrightarrow{g^1}n$.事实上,由于 $j_1^1\in B_1^1\in\boldsymbol{B}_1$,必有 $k\xrightarrow{g^1_{-j_1^1}}n$.这是因为网络 g^1 除了 n 相关的连接外都与 g 相同.由于在结构上 j_1^1 是个"底层"节点,无法"看到"g 中的其他节点,在 g^1 中同样如此.所以,从 k 到 n 的任何路径在 g^1 中都存

在与 j_1^1 无关,即 $k \xrightarrow{g^1_{-j_1^1}} n$.由于 k 是任意的,在 $g^1_{-j_1^1}$ 中从任何节点 k 到 n 都存在路径.由引理 2,可以选择节点 j_1^1 对 g^1 的最佳响应 $\hat{g}_{-j_1^1}$ 为简单的令 $\hat{g}_{j_1^1,n}=1$ 且 $\hat{g}_{j_1^1,k}=0$ 对于所有 k 成立(如图 2 所示,粗箭头线表示新的连接).换句话说, j_1^1 只需建立一条与节点 n 的连接就能得到网络的所有信息.设 $g^2=g^1_{-j_1^1}\oplus\hat{g}_{j_1^1}$ 是 j_1^1 通过这种方式选择其最佳响应后形成的网络结构,同时,其他节点保持惯性.按照这种规律,g^2 以一定概率出现.

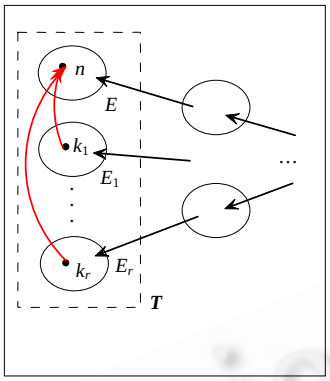

Fig.1 One of n's best response strategy

图 1 节点 n 的最佳响应策略

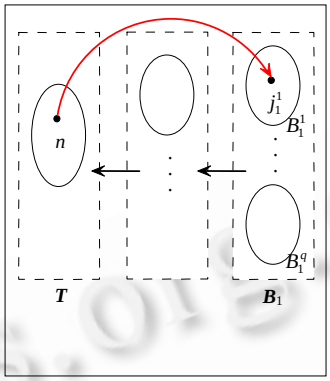

Fig.2 j_1^1's best response strategy

图 2 节点 j_1^1 的最佳响应策略

下面考虑 $B_1^2\in\boldsymbol{B}_1$,将 B_1^2 记为 j_1^2.在 g^2 中除了 j_1^1 的连接之外,g^1 的结构不变,而在 g^1 中除了 n 的连接之外,都保持了 g 的结构.特别地,如果 $k\in N\setminus\{j_1^1,n\}$,则 $k\xrightarrow{g^2_{-j_1^1}} n$.然而,由于 $j_1^2\in B_1^2\in\boldsymbol{B}_1$, j_1^2 无法"看到"g 中的任何节点.于是对于每个 $k\in N\setminus\{j_1^1,n\}$,在 g^2 中都有从 k 到 n 的路径而与 j_1^2 无关,也就是说, $k\xrightarrow{g^2_{-j_1^2}} n$.由引理 2 可知, j_1^2 拥有最佳响应 $\hat{g}_{j_1^2}$ 满足对所有 $k\notin\{j_1^1,n\}$ 有 $\hat{g}_{j_1^2,k}=0$.进而,由于 $g^2_{j_1^1,n}=1$,再次应用引理 2, j_1^2 的最佳响应可以通过令 $\hat{g}_{j_1^2,j_1^1}=1$ 和所有其他节点 k, $\hat{g}_{j_1^2,k}=0$ 获得.令 $g^3=g^2_{-j_1^2}\oplus\hat{g}_{j_1^2}$,从 g^2 将以正的概率演变为 g^3.

现在以同样的方式推进,直到穷举了 $\boldsymbol{B}_1$ 中的所有集合为止.所得的网络(称为 g^4)在 $\boldsymbol{B}_1$ 层之上的所有网络结构都与 g^1 相同,并且 $g^4_{j_1^1,n}=g^4_{j_1^2,j_1^1}=\cdots=g^4_{j_1^{q_1},j_1^{q_1-1}}=1$.现在考虑 $\boldsymbol{B}_2$ 中的集合.令它们为 $\{B_2^1,\ldots,B_2^{q_2}\}$.将 B_2^1 中的节点用 j_2^1 表示.由于 $\boldsymbol{B}_2$ 中所有节点或高层节点具有惯性,网络 g^4 对于 $\boldsymbol{B}_2$ 层之上的节点结构与 g^1 相同.

特别地,给定任意 $k\in E$,其中, $E\in\{B_2^2,\ldots,B_2^{q_2}\}\cup\bigcup_{\{E'\in B_p:\,p\geq 3\}}E'$,我们有 $k\xrightarrow{g^4_{-j_2^1}} n$,原因同上.再次应用引理 2,我们可选择 j_2^1 的最佳响应 $\hat{g}_{j_2^1}$ 满足 $\hat{g}_{j_2^1,j_1^{q_1}}=1$ 且对所有其他的 k,有 $\hat{g}_{j_2^1,k}=0$.新网络(以正的概率出现)记作 $g^5=g^4_{-j_2^1}\oplus\hat{g}_{j_2^1}$,假设其他节点仍保持惯性.

对 $\boldsymbol{B}_2$ 中的所有剩余集合重复此过程,然后是每个集的高一层,直到穷尽所有.所得网络 g^6 满足:

$$g^6_{j_1^1,n}=g^6_{j_1^2,j_1^1}=\cdots=g^6_{j_1^{q_1},j_1^{q_1-1}}=g^6_{j_2^1,j_1^{q_1}}=\cdots=g^6_{j_s^1,j_{s-1}^{q_{s-1}}}=\cdots=g^6_{j_s^{q_s},j_s^{q_s-1}}=1.$$

进而考察 $\boldsymbol{B}_s\subset\boldsymbol{T}$.由于假设节点 n 从 g^1 到 g^6 表现出惯性,且在 g^1 中有 $g^1_{n,j(E)}=1$对每个 $E\in\boldsymbol{T}$ 以及某个 $j(E)\in E$ 成立,特别地,有 $g^6_{n,j_s^{q_s}}=g^1_{n,j_s^{q_s}}=1$.如此网络 g^6 包含环,也就是说,它是超环.

现在有多个冗余连接的节点仅可能有节点 n,现设节点 n 对于网络 g^6 选择最佳响应,根据引理 2,它只需保

留与 g^6 中节点 $j_s^{q_s}$ 的连接即可,所得的网络即是有向环型. □

至此我们证明了即使所有节点都表现出自私性和随机性,网络拓扑仍然会“自组织”演化为有向环的形式.以上证明过程表明,本节开始时建立的网络模型是自组织的.我们简单提及网络整体利益的问题,一般而言,网络的整体利益可以表现为节点的收益之和,也可以表现为节点受益的公平程度.在一个自私的系统中,公平只能通过制度来得到.因此,本文的模型针对的系统中,全网的总体利益可以理解为节点收益的总和.在这个意义上,全网收益也可理解为所有节点对有限信息价值的利用程度.定理 1 的结果还说明了另一个问题:网络的自组织收敛结果不仅使得各节点的私有利益最大化,同时也使全网的总体收益达到了最大化.由于篇幅所限,本文不再给出这个结论的证明.该结论正好与自然界中的自组织系统形成一致,即成员虽然追逐局部利益,但最终会导致全局利益的最大.

4 讨 论

第 3 节的模型看似简单并过于抽象,这主要有几方面因素:(1) 节点拥有同等价值的信息量;(2) 链路建立成本完全相同;(3) 链路的建立成本包含维持通信的开销(否则在建立新链路时仍可以维持老的链路不变);(4) 信息价值的获取量与节点间距离(跳数)的远近无关.下面我们给出一种 PGP 证书网络实例并对其演化进行更为深入的讨论.

4.1 PGP证书网络的简化模型

PGP 证书是由网络中节点(用户)相互签发的(公钥)证书,目的是为了相互认证.关于分布式 PGP 认证系统的详细资料参见文献[10].显然,如果证书的签发者视为节点,证书的签发视为连接,则全部的 PGP 证书组成一个 overlay 网络.从价值论的观点出发,确认了某个节点的证书即是获知此节点是可以信任的,排除了不可靠性,从而获得了一定的价值.一般来说,对应于各个节点的认证价值可认为是相同的,因此,我们定义每个节点都拥有价值为 V 的信息,对一个节点的证书确认,相当于获得了大小为 V 的价值量.节点获取利益的多少最终表现为节点可以认证的节点数的多少.下面考虑连接的建立和维护成本,从前面的论述中已经知道连接的建立和维护就是证书的签发和有效性维护.一般来说,证书的签发是指节点利用一定的算法对其身份等属性生成具有时效性的公钥证书,证书的维护主要是节点每隔一定的周期对证书的时效进行更新.因此,链路的建立和维护成本主要是证书生成和时效更新的计算成本,对于各个节点我们认为大致相同,设此成本为 c.如果模型简化至此,不再考虑建立证书链的计算成本(该成本和跳数有关,在后面还要讨论),则与第 3 节中所描述的模型基本吻合,至此,我们找到了一个可能的实例.

我们将时间轴进行足够细的分割后,仍然可以假定在每个时刻至多只有 1 个节点变更其连接策略,它以一定的概率选择最佳响应策略,而同时,其他节点按照惯性保持连接状态不变.网络结构的变化形成了一个马尔可夫过程$\{X_t\}$.根据第 3 节的结论,基于以上模型的 PGP 证书网络会向着有向环形演化.显然,在这个结果下,每一对节点沿着有向环形方向很容易发现相互认证的证书链,而每个节点只需要签发一个公钥证书即可.节点在最小化自己的证书签发成本的情况下,最大化了自己所能认证的节点个数.当然有人会提出反对意见,因为前面已经说明,我们忽略了一个重要的成本——证书链的建立和验证.在此我们指出,某些情况下,连接的建立(证书的签发和维护)成本要比证书链的建立和验证大很多,这是因为:(1) 证书的签发计算量大于验证;(2) 证书的签发需要带外机制来实现,例如在无线网络中,证书通常利用存储介质或是红外端口传递.基于以上原因,可以认为在某些情况下,节点主要考虑连接的建立成本.所以,上述 PGP 网络模型和演化结果有其合理性.

4.2 考虑验证成本

现在我们加入对证书链验证成本的考虑,显然,这项成本与证书链长度成正比.因此,模型描述为:考虑具有 n 个节点的网络,每个节点拥有私有信息,其价值 $V>0$.节点 i 可以通过建立到节点 j 的单向连接(指向 i)获得 j 的信息.每个连接的建立成本为 $\bar{c}>0$.网络符合下述特点:当 i 建立到 j 的连接后,i 可以获取 j 的所有信息,包括 j 通过建立连接从其他节点获得的信息.但是在获取信息的同时也要付出对应的成本 $\bar{c}$,这个成本与信息的源节点 k

和 i 之间的有向距离成正比.基于这个模型的网络会进行怎样的演化?

事实上,第 2.3 节中的模型可以作为上述模型的一个特例,也就是说,当 $\bar{c} << \vec{c}$ 时,第 2.3 节的结论适用于此.现在考虑另外一种极端情况,即 $\bar{c} >> \vec{c}$,为了方便讨论,我们还假设 $V > \bar{c}$.显然,节点为了最大化所能认证的节点数并最小化成本会建立与其他所有节点的连接,所以,网络的演化结果是全互联有向网络.那么,对于 $\vec{c}$ 和 $\bar{c}$ 的其他取值,网络的演化结果是否还会不同?下面对这种情况进行讨论.由于成本 $\bar{c}$ 的增加(相对于 $\vec{c}$ 不是太大),节点在追求更多信息(对更多节点的认证)的同时必须力争降低与其他节点的(有向)距离,以降低成本 $\bar{c}$,因此,自组织演化的证明思路如下:

1) 首先应确定 $V, \vec{c}$ 和 $\bar{c}$ 的取值范围,显然,如果 V 充分大到 $\bar{c}$ 可忽略,则此模型可以与上一节中的模型等同,演化结果也会一致,而如果 $\bar{c}$ 充分大到 V 可忽略,则网络必然会收敛到全互联情形.而对于正常情况的 $V, \vec{c}$ 和 $\bar{c}$,网络的收敛应在以上两种情形之外,这是因为在有向环型网络中, $\bar{c}$ 的成本太高,节点必然会建立其他的连接;全互联网络的连接成本 $\bar{c}$ 又过高.应该指出的是,确定 3 个值的范围并非一件容易的事,极有可能是 NP 类问题.

2) 证明对于 $V, \vec{c}$ 和 $\bar{c}$ 的某种合适的取值范围(排除两种极端情形),证明星型结构(中心节点和其他节点都有双向连接)既是最大效益网络又是强支撑网络.

3) 证明节点的最佳响应策略中必然包含与出度最大的节点建立连接.

4) 证明(各节点按照自己的最佳响应策略选择连接的)网络最终收敛并演化为星型结构(如图 3 所示,中心节点和其他节点都有双向连接).

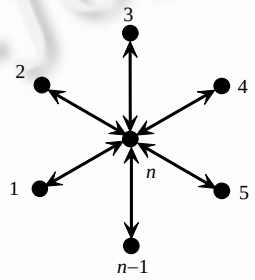

Fig.3 Bidirectional star shape network

图 3 星型有向全连通网络

文献[12]中的统计结果显示,Internet 上较大规模的 PGP 证书网络服从幂律,也就是说,网络由许多星型结构的子网络互联而成,从一定意义上说明本节所希望证明的演化结果具有合理性.

由以上讨论可以看出,随着 $\vec{c}$ 和 $\bar{c}$ 相对大小的变化,网络的自组织演化可能会有 3 种(以上)不同结果,这其实与自组织临界性理论是一致的.自组织临界性理论[11]认为,复杂系统能够自发地向临界状态演化;在这种自组织临界状态,一个小的参数变动会导致突变.自组织临界性理论是一种新的观察自然界的方式.其基本立场是认为自然界总是处于持续的非平衡状态,由于系统内部要素之间的相互作用,它们可以组织成为一种临界稳定的状态,即临界态.典型的自组织临界性现象是沙堆模型.在本文的证书网络模型中,我们认为,随着 $\vec{c}$ 和 $\bar{c}$ 相对大小的改变,网络的自组织演化结果会在有向环形、全连通星型和全互连网络之间发生突变.

5 小 结

本文建立了公钥证书网络自组织演化的数学模型.在此模型中,网络行为表现为节点对信息价值的追求以及维护网络连接所付出代价的权衡.模型将网络的演化表述为一个收敛的随机过程.模型的自组织性表现在:1) 网络没有全局控制者;2) 节点仅追求私有(局部)利益;3) 网络的最终演化结构具有固定拓扑;4) 网络的整体利益可以与节点的私有利益达成一致.文中给出了演化结果的证明.针对 PGP 证书网络实例,我们根据参数的改变对其自组织演化的其他可能结果进行了讨论,最后指出这些结果与自组织临界理论是一致的.这些工作为计算机信息网络的拓扑模型研究以及自组织理论研究提供了一种具有数学基础的新方法.由于自组织网络模型和

理论的研究还刚刚起步,所以模型的完善工作还需要不断深入.

References:

[1] Camazine S, Deneubourg JL, Franks NR, Sneyd J, Theraula G, Bonabeau E. Self-Organization in Biological Systems. Princeton: Princeton University Press, 2003.

[2] Prigogine I, Nicolis G. Self-Organization in Nonequilibrium Systems. John Wiley & Sons, 1977

[3] Zhang JC, Li B, Liu JQ. Application of self-organization theroy in military system. Systems Engineering and Eleceronics, 2002,24(5):11−13 (in Chinese with English abstract).

[4] Albert R, Barabasi AL. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002,74:47−97.

[5] Erdos P, Renyi A. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Science, 1960,5:17−60.

[6] Newman MEJ. Models of the small world. Journal of Statistical Physics, 2000,101: 819−841.

[7] Goh KL, Oh E, Jeong H, Kahng B, Kim D. Classification of scale free networks. Proc. of the National Academy of Sciences of the United States, 2002,99(20):12583−12588.

[8] Alderson D, Willinger W. A contrasting look at self-organization in the Internet and next-generation communication networks. IEEE Communications Magazine, 2005,43(7):94−100.

[9] Bala V. Self-Organization in communication network. 2005. http://www.eur.nl/webdoc/doc/econometrie/ eeb19960111120063.pdf

[10] Yaw D. PGP: An algorithmic overview. 2005. http://davidyaw.com/crypto/index.html

[11] Mei KY. On the self organized criticality and the evolutionary behavior of complex systems. Journal of System Dialectics, 2004,12(4):38−41 (in Chinese with English abstract).

[12] Capkun S, Buttyan L, Hubaux JP. Small worlds in security systems: An analysis of the PGP certificate graph. In: Proc. of the ACM New Security Paradigm Workshop (NSPW), 2002.

附中文参考文献:

[3] 张金春,李彪,刘景权.自组织理论在军事系统中的应用.系统工程与电子技术,2002,24(5):11−13.

[11] 梅可玉.论自组织临界性与复杂系统的演化行为.自然辩证法研究,2004,12(4):38−41.

董攀(1978－),男,河南开封人,博士生,主要研究领域为信息安全,MANET.

卢锡城(1946－),男,教授,博士生导师,工程院院士,CCF 高级会员,主要研究领域为计算机网络,并行与分布处理,高性能计算.

朱培栋(1971－),男,博士,副教授,主要研究领域为路由技术,移动网络,网络安全.