

# 虚拟社区网络的演化过程研究\*

张 立† 刘 云

(北京交通大学通信与信息系统北京市重点实验室,北京 100044)

(2008 年 3 月 30 日收到;2008 年 6 月 12 日收到修改稿)

模拟了虚拟社区网络的演化过程并研究其拓扑结构.发现虚拟社区网络在演化过程中,节点的加入、边的加入、网络中度分布、节点的度与其加入网络时间的关系、平均度随时间的变化等方面与传统的无标度网络有所不同.根据国内某论坛的实际网络数据统计与分析,提出了虚拟社区网络的演化机理——虚拟社区网络构造算法.仿真结果表明,模拟以互联网论坛为代表的虚拟社区网络时,该模型能够得到与真实网络相符的特性.

关键词:复杂网络,虚拟社区,无标度网络

PACC: 0590, 0175

## 1. 引 言

近年来复杂网络的研究使得人们对各种网络结构的复杂性和其网络行为之间的关系产生了浓厚的兴趣.复杂网络中两个比较著名的模型是小世界模型和无标度模型. Watts 和 Strogatz 于 1998 年提出了一个小世界模型(WS 模型)<sup>[1]</sup>,其主要贡献是提出了介于规则网络和随机网络之间的小世界网络,并能通过重连概率  $p$  进行调节,使之在规则网络和随机网络之间转化. WS 模型的构造算法中存在一个随机化重连的过程,此过程有可能破坏网络的连通性. Newman 和 Watts 稍后提出了另一个广为人知的模型<sup>[2]</sup>,称为 NW 小世界模型.该模型采用随机化加边取代了 WS 模型构造中的随机化重连方法.无标度网络是在复杂网络研究上的另一个重大发现<sup>[3]</sup>,包括 Internet, WWW 以及新陈代谢网络等的连接度分布函数具有幂率形式,这些网络节点的连接度没有明显的特征长度.为了解释幂率分布的产生机理, Barabási 和 Albert 提出了一个无标度网络模型(BA 模型),认为以前的许多网络模型都没有考虑到实际网络的两个重要特性:增长特性和优先连接特性.基于以上两个特性, Barabási 和 Albert 提出了 BA 模型的构造算法.随着复杂网络模型的不提出,加之各种大型数据库的出现以及计算机处理能力的提高,

使得对海量网络数据的处理成为可能,针对各个类型网络的实证研究接踵而至,如电影演员合作网<sup>[4]</sup>,科学家合作网络<sup>[5-7]</sup>, WWW 链接关系网络<sup>[8]</sup>, Internet<sup>[9]</sup>,食物链网络<sup>[10]</sup>,新陈代谢网<sup>[11]</sup>,蛋白质相互作用网络<sup>[12]</sup>,电力网<sup>[1]</sup>,手机短信网<sup>[13]</sup>等等.

舆论演进是近年来在物理学领域兴起的另一个方向. Sznajd 等人在提出的 Sznajd 模型中使用了一维媒介模型<sup>[14]</sup>,认为个体是排列在一维空间上的点. Stauffer 等人将原始的一维 Sznajd 模型进行了扩展<sup>[15]</sup>,使其在二维环境中进行观点的演化,认为个体是处在二维网格中的格子.还有一些学者以无标度网络模型为媒介对舆论演进模型进行了研究<sup>[16-19]</sup>.尽管上述研究都得到了部分结果,解释了舆论演进中的一些现象,但从媒介模型上看,简单的一维和二维模型显然无法合理地解释实际中个体之间的交互关系,而采用无标度网络模型是否符合实际舆论演进中的关系也没有得到很好的证实.

舆论传播和虚拟社区都属于复杂的社会现象,存在着众多的随机因素,很难完全重现其演进状况. Bianconi 等人的提出的适应度(Fitness)模型从一定程度上解释了真实社会网络的特性<sup>[20]</sup>,但本文研究的对象是以互联网论坛为代表的虚拟社区网络,社区网络中的衰减、非匀速增长等特性在适应度模型中都没有进行讨论.此类网络恰恰又是国内舆论突发事件的主要集中地.在对实际网络进行统计与分

\* 教育部高等学校科技创新工程重大项目培育基金(批准号:707006),通信与信息系统北京市重点实验室基金和北京市教育委员会共建项目专项资助的课题.

† E-mail: 05111037@bjtu.edu.cn

析时,发现其在网络演化过程中,节点的加入、边的加入、网络中度分布、节点的度与其加入网络时间的关系、平均度随时间的变化等方面均与传统的网络有所不同.本文进一步提出了符合虚拟社区网络的网络模型构造算法,得到了比适应度模型更符合社区网络的结果.仿真结果表明,这一算法能更真实地反应虚拟社区网络的真实情况.

## 2. 无标度网络

BA 无标度网络模型的构造算法分为以下两个步骤:

1) 增长 从一个具有  $m_0$  个节点的网络开始,每次引入一个新的节点,并且连接到  $m$  个已经存在的节点上,此处  $m \leq m_0$ .

2) 优先连接 一个新节点与一个已经存在的节点  $i$  相连接的概率  $\Pi_i$  与节点  $i$  的度  $k_i$ 、节点度之和之间满足如下关系:

$$\Pi_i = k_i / \sum_j k_j. \quad (1)$$

在复杂网络模型中,一个重要的性质是平均路径长度.网络中两个节点  $i$  和  $j$  之间的距离  $d_{ij}$  定义为连接这两个节点的最短路径上的边数,网络中任意两个节点之间的距离的最大值称为网络的直径,记为  $D$ ,即

$$D = \max_{i,j} d_{ij}. \quad (2)$$

网络的平均路径长度  $L$  定义为任意两个节点之间的距离的平均值,即

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i>j} d_{ij}. \quad (3)$$

根据 BA 无标度网络的构造算法生成一个简单的无标度网络,如图 1 所示.

图 1 所对应无标度网络的线性度分布和双对数坐标下的度分布如图 2 所示.

图 1 生成的无标度网络基本数据如表 1 所示.

表 1 图 1 所示网络的基本数据

平均度	最大度	直径	平均路径长度
3.992	290	8	2.934

BA 无标度网络是一种典型的复杂网络,能解释一些实际网络的性质,但本文研究的对象是虚拟社区网络,本文作者发现该网络的某些特征与 BA 无标度网络有很大不同.

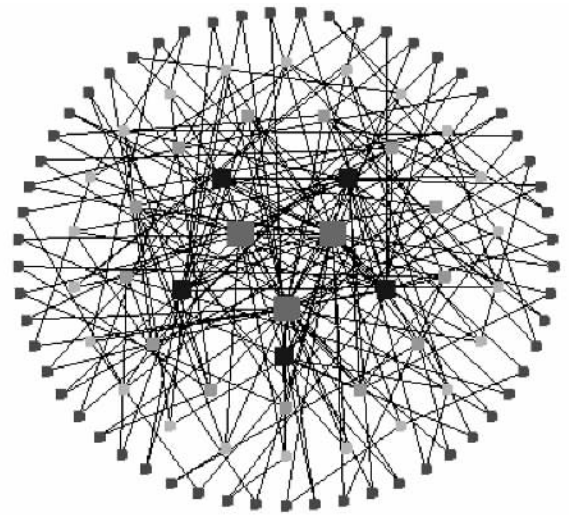


图 1 BA 无标度网络

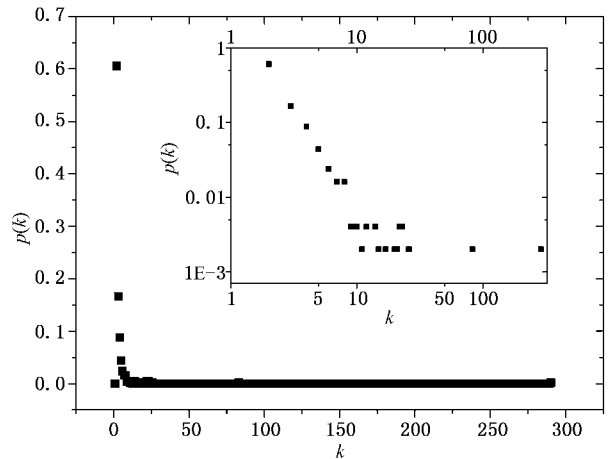


图 2 图 1 所示无标度网络的度分布

## 3. 虚拟社区网络

虚拟社区又被称为虚拟社群、网络社群或电子社群等,是互联网使用者经过互动后,产生的一种社会群体. Rheingold 认为虚拟社群是一群人在网络上从事公众讨论,经过一段时间,彼此形成人际关系的网络<sup>[21]</sup>. Rheingold 认为虚拟社区是一种新型的社会组织,并有以下四种特质:1)表达的自由;2)缺乏集中的控制;3)多对多的传播;4)成员自主的进行互动.虚拟社区的形式包括了早期的电子公告版、讨论区、MUD,以及近年才出现的 Blog, Wiki 等.我国的互联网应用中,互联网论坛有着重要的作用,始终拥有庞大的用户群.一些早期的电子公告版也都相继开发了 Web 页面形式的互联网论坛,这些论坛是我国

互联网虚拟社区的重要组成部分,同时也促成了某些善意或者恶意的网络舆论突发事件。

互联网论坛是一个为广大用户提供张贴特定或一般内容以及进行讨论的 Web 应用。一般而言,论坛指整个讨论区,而版块作为论坛的子区域用于更为细分的话题讨论。论坛中的内容一般可以按照时间顺序或主题模式进行浏览。当前的论坛一般都具备较为丰富的功能,提供了游戏、Blog 等扩展应用,但其基本功能都相同。

版块中的话题一般称为帖子,根据帖子的发布时间和回复关系不同可细分为根贴、回帖等。其组织形式如图 3 所示。

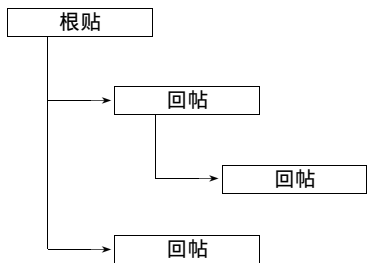


图 3 论坛中帖子组织形式

通过图 3 可以看到,用户之间通过帖子的回复关系建立明显的连接。随着时间的推移、帖子的发布和论坛的建设,用户之间的关系也逐渐复杂,一个庞大的网络将会被建立起来。本文选取了国内某论坛的一个子版块,得到了其 Web 页面中的主要数据,共获得约 15 万帖子信息,5000 个用户信息。

首先定义网络的形成模型如下:

节点 在该版块中发言的不同用户 ID 分别为一个节点;

边 根据表 1 中的数据结构,若两个用户之间存在回复关系,则认为用户之间存在一条边。

模型中假设网络中没有自环,即忽略用户回复自己帖子时形成的自环,不考虑由于用户之间存在多次的回复关系形成的重边,即认为任意两个用户之间最多只存在一条边。

根据上述规则建立了一个网络,其基本数据如表 2 所示。

表 2 虚拟社区网络的基本数据

平均度	最大度	直径	平均路径长度
7.555	290	10	3.902

其度分布如图 4 所示。

在 BA 无标度网络中,度分布符合幂率分布,即

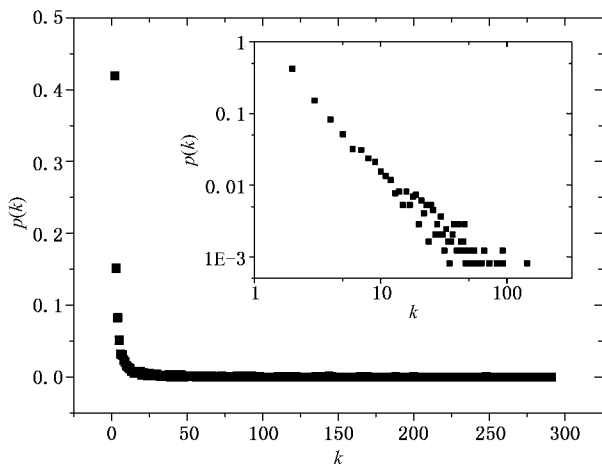


图 4 虚拟社区网络的度分布

$$P(k) \propto k^{-r}, \quad (4)$$

其中  $r = 3$ ,对上图中的曲线进行拟合的结果是  $r = 2.28937 \pm 0.01321$ ,虽然比较好的符合了幂率分布,但指数  $r$  却有较大差别,这说明虚拟社区网络的构造形式与传统的无标度网络可能存在着不同。

对所获得的论坛数据进行了统计,图 5 即网络节点数的日增加数随日期的变化趋势。

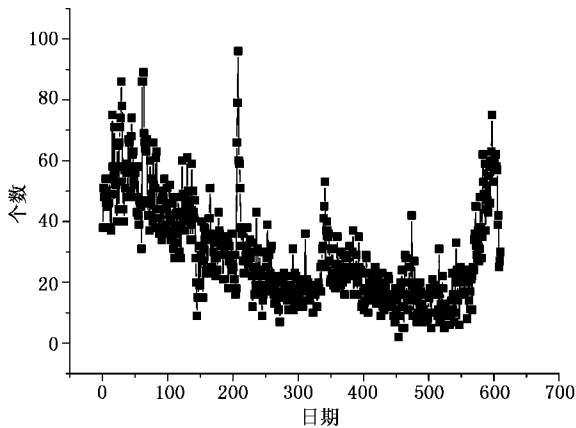


图 5 节点日增加数分布图

从图 5 可以看到,版块中节点个数的增加是不平均的,除了尾部突然出现了一个上升趋势,其基本趋势可以描述为带有随机波动的下降趋势,这点与 BA 无标度网络的演化过程是不同的。

图 6 显示了网络中边的日增加数与时间之间的关系。

图 6 中可以看出,网络中边数增加的大体趋势与节点相比比较平稳。求得日每节点平均增加边数随时间变化如图 7 所示。

从图 7 中可以看出,网络中日每节点平均增加

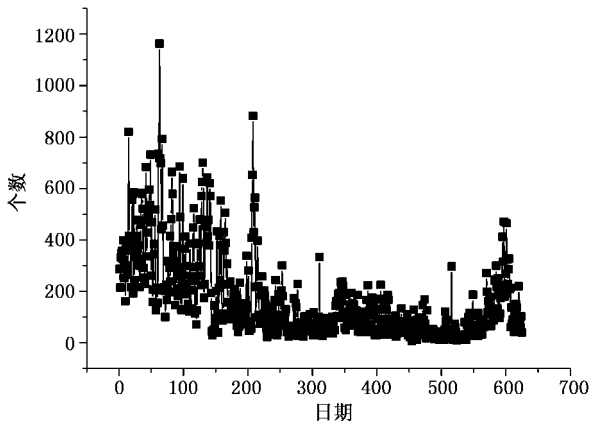


图 6 边日增加数分布图

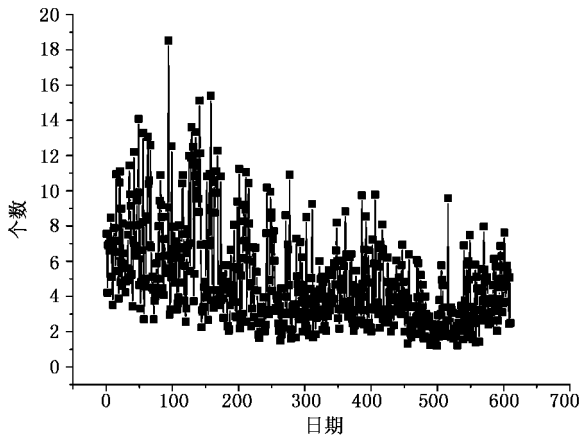


图 7 每节点日平均增加边数图

边数具有很大的波动.虽然 BA 无标度网络构造算法中限定的  $m$  具有平均上的意义,但仍然需要一种更符合实际的表达方法.

网络的平均度随时间的变化如图 8 所示.

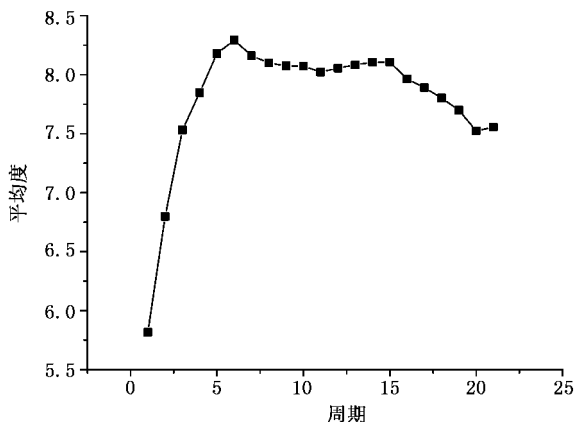


图 8 平均度随时间变化图

图 8 中以月为单位,计算了网络的平均度随网络演化时间的变化情况,从图中可以看出,网络的平

均度除开始的迅速增长外,基本处于一个比较平稳的水平之中.

网络节点的度与其加入到网络的时间,即论坛中的用户与其他用户的回复连接数与时间的关系如图 9 所示.

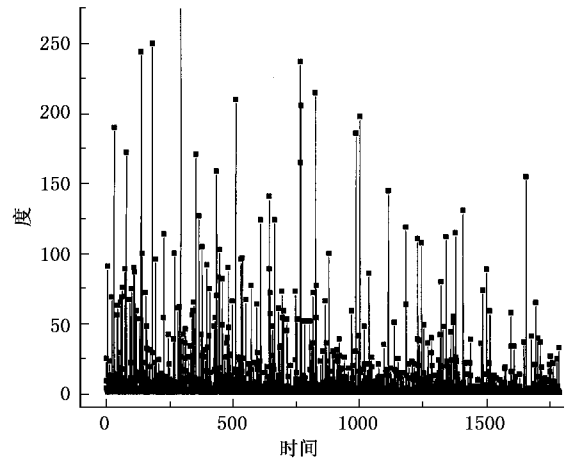


图 9 节点最终度与加入时间的关系

从图 9 中可以看到,在网络随时间演化的各个阶段,都出现了度比较大的节点,在网络演化初期进入网络中的节点也有非常多的处于图 9 的底部,没有成为度较大的节点,这些变化都说明虚拟社区网络的演化过程与 BA 无标度网络存在着不同.

### 4. 虚拟社区网络模型

本文提出一个新的适合于虚拟社区的网络模型构造算法:

1) 增长 从一个具有  $m_0$  个节点的网络开始,  $n_0$  为新加入节点数的上限,每次引入不超过  $n_0$  的  $n$  个新的节点,并且连接到  $m$  个已经存在的节点上.此处  $m \leq m_0, n \leq n_0$ ,设  $n$  和  $m$  的概率密度函数分别为  $\rho(n)$  和  $\rho(m)$ .

2) 优先连接 一个新节点与一个已经存在的节点  $i$  相连接的概率  $\Pi_i$  与节点  $i$  的度  $k_i$ 、节点度以及其优先连接概率  $\eta$  之间满足如下关系:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{5}$$

其中  $\eta$  的概率密度函数为  $\rho(\eta)$ ,  $n, m, \eta$  的概率密度函数可能相同,也可能不同,需要根据不同的应用场景进行不同的限定,如本文第 5 节的仿真中均采用了平均分布进行处理.设节点  $i$  加入网络的时间

为  $t_i$ , 设新节点与节点  $i$  相连接的概率  $\Pi_i$  服从如下的衰减规律：

$$\Pi_i(t) = \frac{\eta_i k_i (t - t_i + 1)^{-\alpha}}{\sum_j \eta_j k_j (t - t_j + 1)^{-\alpha}} \quad (6)$$

3) 加边 在网络的演化过程中, 随机地选择部分节点, 使其按照规则 2) 与已有的节点建立新边。

上述算法中涉及的增长、优先连接和加边三个过程基本涵盖了下面提到的虚拟社区网络的 5 个特性。

以上建立的虚拟社区的网络模型, 主要对应了网络论坛中的一些实际特性, 主要有以下方面：

1) 节点(即用户)在进入网络时, 一般伴随着一个帖子, 这个帖子的内容对节点度的变化有着重要的影响, 即如果该帖子的内容非常吸引人、是热门话题, 则其为节点带来的优先连接概率<sup>[20]</sup>较大。

2) 已经存在于网络中的节点还有可能继续通过发表新的帖子来增加自身的优先连接概率, 使得自身更容易为其他节点所连接。

3) 除了新加入的节点会对已有的节点进行连接外, 网络中已存的节点之间也会添加新边, 即网络中已有用户对同一话题展开讨论。

4) 已经存在于网络中的节点随着网络的演化会逐步的降低自身的优先连接概率, 这是由于网络论坛中的帖子随着时间的增长会逐步丧失吸引力。当然, 这些用户也可能通过发布新帖子来继续增加自身的吸引力。

5) 节点和边的增加并非匀速增加, 存在较大的随机因素(分析见第 3 节)。

### 5. 仿真结果与分析

使用 networkx 包及其他扩展, 对第 4 节中提出的模型进行了数值仿真。其中  $n$  和  $m$  的分布具有很大的随机性, 因此在仿真时采用了平均分布。(6)式中的  $\eta$  按照文献 [19] 中的平均分布方式进行处理, 并限定  $0 < \alpha < 1$ 。在 networkx.generators.random\_graphs.barabasi\_albert\_graph 方法之上进行了改进, 得到的结果如图 10 所示。

从图 10 中可以看出, 仿真结果基本得出了与实际统计结果相类似的结果(见图 4), 图 11 展示了节点最终的度与其加入网络时间之间的关系。

从图 11 中可以看出, 仿真结果与典型的无标度网络有所不同, 同样与实际的统计结果有较好的契

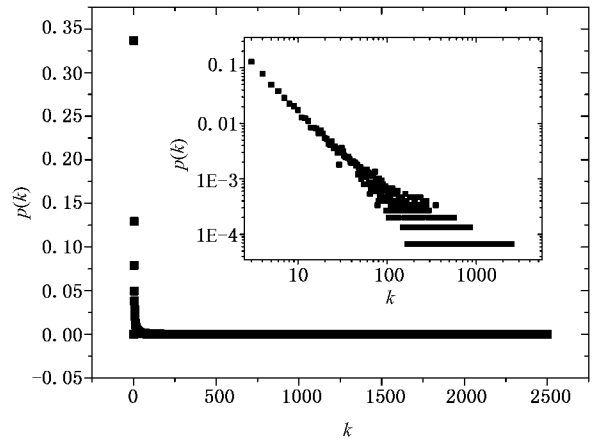


图 10 仿真模型的度分布

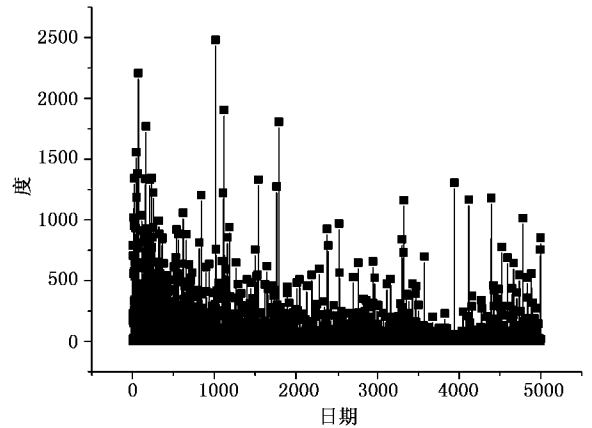


图 11 仿真模型的节点度与加入时间的关系

合(见图 9)。图 12 展示了仿真结果中网络的平均度与网络演化时间之间的关系。

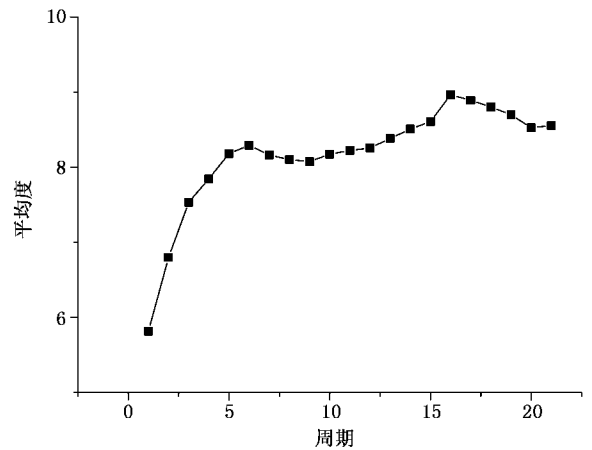


图 12 仿真模型的平均度随时间的变化图

从图 12 可以看出, 仿真结果中平均度初期呈现出较快的上升趋势, 在一定阶段后趋于平稳, 并伴随

着一些随机的波动,这点也与实际的统计结果基本符合(见图8)。

## 6. 结 论

本文研究了以互联网论坛为代表的虚拟社区网络,考察了其某些实际数据与传统无标度网络之间的不同,进一步提出符合虚拟社区网络实际特性的

网络模型构造算法,扩展了适应度模型,使建立的模型更符合社区网络的特点,得到的仿真结果与实际的统计结果基本符合。可以进一步考察虚拟社区网络的基本性质,以本文提出的算法为基础进行舆论演进的仿真实验,以期得到更符合虚拟社区网络中舆论传播的结果。

感谢李勇老师和沈波老师参与讨论。

- [ 1 ] Watts D J , Strogatz S H 1998 *Nature* **393** 440
- [ 2 ] Newman M E , Watts D J 1999 *Phys. Lett. A* **263** 341
- [ 3 ] Barabási A L , Albert R 1999 *Science* **286** 509
- [ 4 ] Albert R , Barabási A L 2000 *Phys. Rev. Lett.* **85** 5234
- [ 5 ] Newman M E 2001 *Phys. Rev. E* **64** 016131
- [ 6 ] Newman M E 2001 *Phys. Rev. E* **64** 016132
- [ 7 ] Newman M E 2001 *Proc. Natl. Acad. Sci.* **98** 404
- [ 8 ] Albert R , Jeong H , Barabási A L 1999 *Nature* **401** 130
- [ 9 ] Chen Q , Chang H , Govindan R *et al* 2002 *INFOCOM* 2
- [ 10 ] Martinez N D 1991 *Ecol. Mono.* **61** 367
- [ 11 ] Jeong H , Tombor B , Albert R *et al* 2000 *Nature* **407** 615
- [ 12 ] Jeong H , Mason S P , Barabasi A L *et al* 2001 *Nature* **411** 41
- [ 13 ] Wu Y , Xiao J H , Wu Z Y *et al* 2007 *Acta Phys. Sin.* **56** 2037 ( in Chinese ) 吴 晔、肖井华、吴智远等 2007 物理学报 **56** 2037 ]
- [ 14 ] Sznajd-weron K , Sznajd J 2000 *Int. J. Mod. Phys. C* **11** 1157
- [ 15 ] Stauffer D , Sousa A O , Oliveira S M 2000 *Int. J. Mod. Phys. C* **11** 1239
- [ 16 ] Elgazzar A S 2001 *Int. J. Mod. Phys. C* **12** 1537
- [ 17 ] Elgazzar A S 2003 *Phys. A* **324** 402
- [ 18 ] Bernardes A T , Stauffer D , Kert J 2002 *Euro. Phys. J. B* **25** 123
- [ 19 ] Bonnekoh J 2003 *Int. J. Mod. Phys. C* **14** 1231
- [ 20 ] Bianconi G , Barabási A L 2001 *Phys. Rev. Lett.* **86** 5632
- [ 21 ] Rheingold H 2000 *The Virtual Community : Homesteading on the Electronic Frontier* ( London : MIT Press )

# Research on the evolution process of virtual community networks<sup>\*</sup>

Zhang Li<sup>†</sup> Liu Yun

( Key Laboratory of Communication & Information Systems-Beijing Jiaotong University , Beijing Municipal Commission of Education , Beijing 100044 , China )

( Received 30 March 2008 ; revised manuscript received 12 June 2008 )

## Abstract

We investigate and simulate the evolution process and topological features of the virtual community networks. We founded that the joint of nodes , joint of edges , networks ' degree distribution , average degree and the correlation between the node 's degree and its time in the network are different from the scale-free network model. With statistics and analysis of some of the characteristics of an actual Internet forum , we propose a new network evolution algorithm , namely the virtual community network construction algorithm. The simulation results show that the networks generated from our algorithm has the same characteristics as the real virtual community networks.

**Keywords :** complex networks , virtual community , scale-free networks

**PACC :** 0590 , 0175

<sup>\*</sup> Project supported by the Cultivation Fund of the Key Scientific and Technical Innovation Project ( Grant No. 707006 ) and the Special Foundation from Key Laboratory of Communication & Information Systems ( Beijing Jiaotong University ) and Beijing Municipal Commission of Education.

<sup>†</sup> E-mail : 05111037@bjtu.edu.cn