个体间相互影响的网络舆情演变模型

聂 哲[1,2],李粤平[1],温晓军[1],何国坤[1],陈 健[1]
NIE Zhe[1,2],LI Yue-ping[1],WEN Xiao-jun[1],HE Guo-kun[1],CHEN Jian[1]

1.深圳职业技术学院 电子信息工程学院,广东 深圳 518055
2.哈尔滨工业大学 深圳研究生院,广东 深圳 518055
1.Shenzhen Polytechnic,Shenzhen,Guangdong 518055,China
2.Shenzhen Graduate School,Harbin Institute of Technology Shenzhen,Guangdong 518055,China
E-mail:niezhe@szpt.net

NIE Zhe,LI Yue-ping,WEN Xiao-jun,et al.Evolution model of network opinion with individual affected probability. Computer Engineering and Applications,2009,45(14):220-222.

Abstract: This paper proposes an evolutionary model of opinion formation through the social network in which individuals are embedded.The model takes explicitly into account the probability that an individual is affected by its neighbor in the social network.It is assumed that each individual A will accept its neighbor's opinions at the probability $\alpha(A)$.And be supposed the affected probability of k-degree individuals are the same,denoted by $\alpha(k)$,where the degree k is the sum of the connections from one individual to the others.It is shown that if the distribution of the function α satisfies the condition $\alpha(k)=kp_kc$ for all k,the fraction of the individuals that hold a given opinion is a martingale,i.e.whose expectation is constant in time.The fraction is one of the most important measures of the popularity of opinions.The result can help to evaluate the speed of the opinion spread in the given social network.

Key words: opinion formation;network opinion;martingale

摘 要:提出了一个社会网络中舆论形成的演化模型,模型考虑了网络中个体受其邻居影响的概率。假设个体 A 受到其他邻居影响的概率为 $\alpha(A)$,并且所有 k 度个体具有相同的受影响概率为 $\alpha(k)$,其中 k 是某个体邻居的个数。证明了如果概率 α 的分布满足对所有 k 满足 $\alpha(k)=kp_kc$,那么持某种舆论个体的人数比例是一个鞅,即数学期望是一个常数。本模型有助于衡量某给定社会网络中舆论传播的快慢程度。

关键词:舆论形成;网络舆情;鞅
DOI:10.3778/j.issn.1002-8331.2009.14.068 **文章编号**:1002-8331(2009)14-0220-03 **文献标识码**:A **中图分类号**:TP393.08

1 引言

因特网用户的急剧增加和网络的膨胀,使得无数的观点、新闻和评论发表在各种网站上,特别是借助电子邮件及网络的便利,使得新闻和观点更容易传播。在现实生活中,由于对他人的依赖,所以一个观点的形成和演变常常受到所信赖和熟悉人的影响,因此有必要去研究舆论的形成与由大量个体组成的社会网络结构之间的关系,舆论可以通过个体之间的相互交流而形成和改变。因为在选举和民意调查中,个体的观点起着很重要的作用, 由此引发很多的机构去研究网络舆情的演变机制[1-6]。

学术界提出了许多舆论形成的模型。在经济学中利用信息重叠的理论来解释社会行为[7-8]。这种模型指出,信息重叠令观点强化,直至成为一种社会舆论。然而这种模型和常识相矛盾,舆论的形成具有群体性而不是广泛存在的。

Wu 和 Huberman 的研究表明,舆论的形成与社会网络的结构有关[9]。他们得出的结论是,经过足够长的时间,持有固定观点的人数的期望值是一个常量。而且在一个足够长的时间内,这个常量在社会群体里的比例既不是 0 也不是 1,而是一个相对稳定的值。他们的结论符合实际——观点局限在一个固定的群体内,而不是传播在整个社会中。

在 Wu 和 Huberman 介绍的模型中,每个个体以固定的概率 1 受其它人影响。本文扩展了他们的模型,认为个体受到影响的概率是不同的,这样更符合现实。研究结果表明如果个体受到影响概率的分布符合特定条件,那么持某一观点的个体比例是一个鞅。这和文献[9]的结论不一样,他们指出持某一观点的个体加权比例是一个鞅。

基金项目:深圳市科技计划项目资助项目(No.07KJce140)。
作者简介:聂哲(1970-),男,研究生,副教授,主要研究领域为网络信息挖掘和网络规划;李粤平(1981-),男,博士,讲师,主要研究领域为图论和数学建模;温晓军(1971-),男,博士,副教授,主要研究领域为量子密码学和信息安全;何国坤(1980-),男,讲师,主要研究领域为网络规划;陈健(1980-),男,讲师,主要研究领域为数据库技术。
收稿日期:2008-11-18 **修回日期**:2009-02-11

2 基本概念

在模型中，用一幅随机关联图来描述社会网络。节点用来替代个体，边来展示它们的社会关系。假设社会关系的结构在舆论形成的过程中被认为是静态的，因为结构的改变相对舆论的传播速度来说慢很多，因此这个假设在舆论改变的短时间内是有效的。

引用文献[9]介绍的术语。术语“黑色”和“白色”用来表示个体的二元观点。假设每个节点初始化（个体）是黑色或者白色的。设 n 为总的节点数，即研究的社会个体的总量。节点的度就是它邻居的数量，n_k 为 k 度的节点数，$p_k=n_k/n$ 就是度的分布情况。设 m 是黑色节点的数量并且 m_k 是拥有 k 度黑色节点的数量。$q=m/n$ 是黑色节点的比例，并且 $q_k=m_k/n_k$。

Wu 和 Huberman 假设节点（个体）根据它邻居的颜色，在时间间隔 dt 内更新它的观点。也就是说，在每段时间内，个体随机的选择它的邻居，并且将颜色改为它邻居的颜色（观点）。值得注意的是，考虑节点 u 以概率 α_u 接受别人的观点，$\alpha_u\in[0,1]$，而 Wu 和 Huberman 研究的是 $\alpha_u=1$ 的情况[9]。

设 u 为需要更新的个体并且 k 为 u 的度。根据假设，u 节点随机从它的邻居选择颜色，记作 v。不失一般性，假设节点 u 在更新前是白色的。重点考察节点 v。v 的概率和 j 节点的度是成正比的。并且 v 是黑色节点的概率为 q_j。所以，节点 u 从白色变成黑色的概率如下：

$$P_{w\to b}(u)=\alpha_u p_k(1-q_k)\frac{\sum_j jp_jq_j}{\sum_j jp_j} \tag{1}$$

假设 k 度白点平均的活跃因子为 $\alpha_W(k)$，那么 k 度白点把邻居从黑变白的概率为：

$$P^k_{w\to b}=\alpha_W(k)\times p_k(1-q_k)\frac{\sum_j jp_jq_j}{\sum_j jp_j} \tag{2}$$

对称地，可以用 $\alpha_B(k)$ 定义白色 k 度节点的平均影响概率。因此，更新 k 度节点从黑色变成白色概率如下：

$$P^k_{b\to w}=\alpha_B(k)\times p_kq_k\frac{\sum_j jp_j(1-q_j)}{\sum_j jp_j} \tag{3}$$

简略地，定义

$$<q>=\frac{\sum_j jp_jq_j}{\sum_j jp_j} \tag{4}$$

则公式（2）和公式（3）可以改写为

$$P^k_{w\to b}=\alpha_W(k)\times p_k(1-q_k)<q> \tag{5}$$

$$P^k_{b\to w}=\alpha_B(k)\times p_kq_k(1-<q>) \tag{6}$$

这里介绍一些本文所需的概率术语。条件概率的期望有两种版本，采用近代的版本。近代条件概率期望是定义在代数 F 上的，定义 $E[X|\mathcal{F}]$描述了在给定 F 信息的情况下，X 事件发生的概率，$\mathcal{F}_B$ 由 B，B^C 是 B 的补，Ω（概率空间）和空集 $\varnothing$ 组成。定义 $Y(\omega)$函数如下：

$$Y(\omega)=\begin{cases}E[X|B] & \text{if } \omega\in B\\ E[X|B^C] & \text{if } \omega\notin B\end{cases} \tag{7}$$

$E[X|B]$是另外一个条件概率期望的解释，它是这样定义的：

$$E[X|B]=\sum_\omega X(\omega)P(\omega|B) \tag{8}$$

$$E[X|B]=\sum_\omega X(\omega)P(\omega|B) \tag{9}$$

如果 $E[X_{t+1}|\mathcal{F}]=X_t$，即期望不随时间的变化而变化，则 X_t 是一个鞅。发现期望值不依赖于时间 t，$E[X_{t+1}]=E[X_t]$。

3 舆论演化模型

在预测和分析中，拥有同样观点的人群数量及其在总人群的比例是非常重要的。重点考察随着时间的流逝此黑点的个体数 m_k 和其所占比例 q_k 的演变。在一个时间间隔内（t，t+dt），m_k 的变化可以通过如下的计算得到

$$\Delta m_k=\begin{cases}+1 & \text{at the probability } \alpha_W(k)\times p_k(1-q_k)<q>\\ -1 & \text{at the probability } \alpha_B(k)\times p_kq_k(1-<q>)\\ 0 & \text{otherwise}\end{cases} \tag{10}$$

当时间 $t\to\infty$，考察 Δm_k 的期望值：

$$\begin{aligned}E[\Delta m_k]=&n(\alpha_W(k)\times p_k(1-q_k)<q>-\alpha_B(k)\times p_kq_k(1-<q>))=\\&\alpha_W(k)\times n_k(1-q_k)<q>-\alpha_B(k)\times n_kq_k(1-<q>)=\\&n_k(\alpha_W(k)\times(1-q_k)<q>-\alpha_B(k)\times q_k(1-<q>))=\\&n_k(\alpha_W(k)<q>-\alpha_W(k)<q>q_k-\alpha_B(k)q_k+\alpha_B(k)q_k<q>)\end{aligned} \tag{11}$$

假设 $\alpha_W(k)=\alpha_B(k)$，即 k 度个体受到同样的概率变成黑色或白色。因此，令 $\alpha(k)=\alpha_W(k)=\alpha_B(k)$，那么期望值为：

$$E[\Delta m_k]=n_k\alpha(k)\times(<q>-q_k) \tag{12}$$

上式可以写成

$$E[\Delta q_k]=\alpha(k)\times(<q>-q_k) \tag{13}$$

经过分析和简化可以获得 q 的期望值

$$E[\Delta q]=\sum_k\alpha(k)<q>-\sum_k\alpha(k)q_k \tag{14}$$

上式的平衡条件是将公式的右边设为 0。这样有

$$<q>=\frac{\sum_k\alpha(k)q_k}{\sum_k\alpha(k)} \tag{15}$$

从$\langle q\rangle=\frac{\sum_j jp_jq_j}{\sum_j jp_j}$，即公式（4）的定义，可以计算出平衡条件是

$$\frac{\sum_j jp_jq_j}{\sum_j jp_j}=\frac{\sum_k\alpha(k)q_k}{\sum_k\alpha(k)} \tag{16}$$

上式表明受影响概率的分布和 kp_k 一致。因此，可以得到一个更强的条件使得均衡条件成立：

$$\alpha(k)=kp_kc\text{, for all } k \tag{17}$$

其中 c 为一正常数使得 $\alpha(k)\leqslant1$ 对所有的 k 成立。当公式（16）成立时，$E[\Delta q]=0$ 即 $E[q_{t+1}]=E[q_t]$对任意时刻 t。这表明 q_t 是一个鞅。

Wu 和 Huberman 说明了黑点个体的加权比例$\langle q\rangle$是一个鞅[9]。而研究结果则说明了在受影响概率作用下，黑点比例 q 是

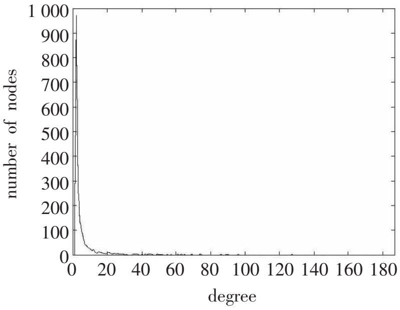

图 1　样例的度分布图

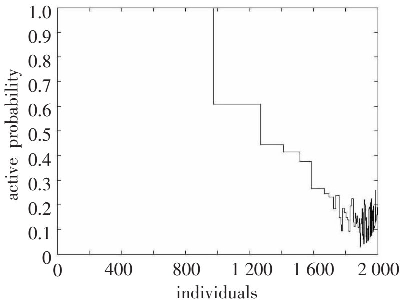

图 2　函数 α 的值

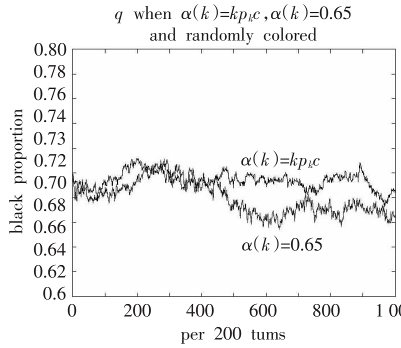

图 3　个体随机染色情况下 q 值对比

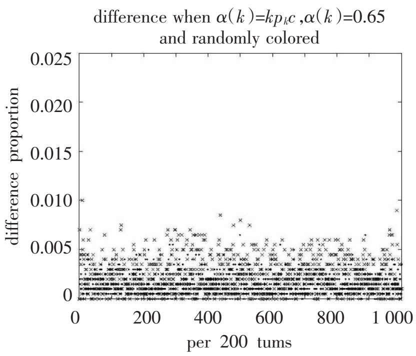

图 4　个体随机染色情况下黑点比例

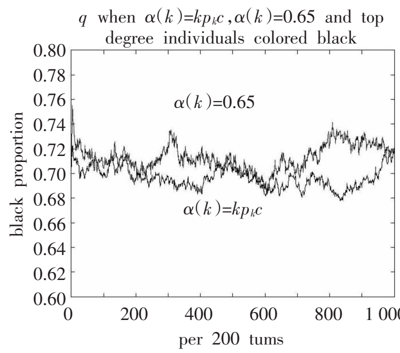

图 5　度数高的点相同颜色情况下 q 值对比

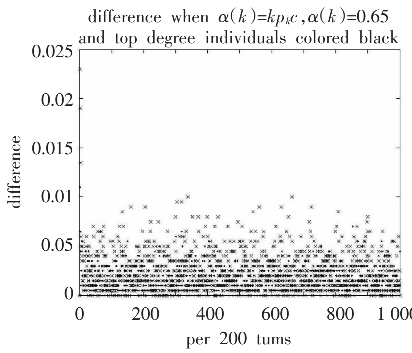

图 6　度数高的点相同颜色情况下黑点比例

一个鞅。由于黑点比例 q 代表拥有相同观点的个体数，比起<q>更为直观，因此本结论对舆论演变的预警和检测有着重要意义。

4　实验结果

随机生成一个网络，个体数 n=2 000 和参数 η=2，其中度分布满足 $p_k=(\eta-1)k^{-\eta}$。如图 1 为一个网络个体度分布的例子。

不妨设 $k_{\max}=\max\{k|kP_k\}$，令 $c=\dfrac{1}{k_{\max}P_{k_{\max}}}$。由此对所有的 k 算出 $\alpha(k)$，分布如图 2 所示，同时得到受影响概率 $\alpha(k)$ 的平均值为 0.681 867。

4.1　第一组模拟

随机选取 70%的个体为黑色，剩余的个体为白色。对于每次信息传递，首先随机选取一个个体；其次，随机选取一个邻居，根据邻居的受影响概率把它的颜色更新邻居上。

每次实验进行信息传递 100×n 次，即平均每个个体向外传递信息 100 次。本组模拟，分受影响概率为 0.65 和受影响概率满足公式(17)进行两轮实验，每次实验运行 1 000 次。黑点比例 q 的平均值变化如图 3 所示。

其中每 200 次更新操作，取样一次。并记录每次黑点所在比例的变化情况如图 4 所示。

其中×点为 $\alpha(k)$=0.65 的变化数据，实心点为 $\alpha(k)=kp_kc$ 的变化数据，由此可见 $\alpha(k)=kp_kc$ 情况下变化得程度较小。表 1 列出这些点的分布情况。

表 1　各概率段×点和实心点分布表

	0.000~0.001	0.001~0.002	0.002~0.003	0.003~0.004	0.004~0.005	0.005~0.006	0.006~0.01
×点	314	216	151	127	125	35	32
实心点	450	273	165	70	35	4	3

4.2　第二组模拟

选取度数最大的 200 个点为黑色，其余随机选取 500 的其他个体也设为黑色，剩余的个体为白色。对于每次信息传递，首先随机选取一个个体；其次，随机选取一个邻居，根据邻居的受影响概率把它的颜色更新邻居上。

每次实验进行信息传递 100×n 次，即平均每个个体向外传递信息 100 次。本组模拟，分受影响概率为 0.65 和受影响概率满足公式(17)进行两轮实验，每次实验运行 1 000 次。黑点比例 q 的平均值变化如图 5 所示。

其中每 200 次更新操作，取样一次。并记录每次黑点所在比例的变化情况如图 6 所示。

其中×点为 $\alpha(k)$=0.65 的变化数据，实心点为 $\alpha(k)=kp_kc$ 的变化数据，由此可见 $\alpha(k)=kp_kc$ 情况下变化得程度较小。表 2 列出了这些点的分布情况。

表 2　各概率段×点和实心点分布表

	0.000~0.001	0.001~0.002	0.002~0.003	0.003~0.004	0.004~0.005	0.005~0.006	0.006~0.01
×点	280	200	141	134	136	48	57
实心点	453	262	147	84	43	6	4

5　结论

扩展了 Wu 和 Huberman 的舆论信息观点[9]，考虑了个体的受影响概率，即对于某一观点，不同个体受他人影响的概率是不同的。研究发现，如果受影响概率对所有的 k，在满足条件 $\alpha(k)=kp_kc$ 下，群体中拥有相同观点的比例将是一个鞅。也就是说，该比例的期望值是一个不随时间变化的常量，从而随时间流逝变化小，实验结果验证了本文的结论。

持有人数比例是评价某舆论热度的最重要指标之一。本模

（下转 227 页）