

基于社会网络的病毒营销传播效率预测*

余力 杨小平 陈明远

(中国人民大学信息学院, 北京 100872)

摘要 病毒营销是随着 Web 2.0 和社会网络发展起来的一种新型社会化网络营销模式, 近年来被许多企业采用。但是, 病毒营销的效果到底如何? 在一定时间内, 究竟会有多少客户被成功营销? 本文采用概率统计分析和网络分析的方法, 对病毒营销的传播效率进行了初步预测研究, 试图回答: “假设有某病毒营销网络 $G(V, E)$, 对于给定的初始营销用户集合 S_0 , 经过一段时间 t 后, 平均能影响的用户数 $E_{S_0}(t)$ 是多少?” 作者通过引入影响响应时间, 先研究了简单的串联、并联影响传播网络图的传播效率情况, 在此基础上研究了一般的社会网络图的病毒营销传播效率, 最后用一个实例进行了展示。

关键词 病毒营销, 网络营销, Web 2.0, 传播效率, 社会商务

中图分类号 TP311

1 引言

Internet 技术的发展极大地促进了人们的交流与沟通, 产生了许多以 Web 为平台的虚拟社会网络系统, 如 Twitter、新浪微博、Facebook、人人网、淘江湖(www.taojianghu.com)等, 为企业进行市场营销提供了巨大的网络平台^[1,2]。病毒营销(viral marketing)就是随社会网络而出现的一种新型的网络营销模式^[3-5], 与传统营销有着本质的区别。传统营销本质上是一种 1-to- n 的营销模式, 营销成功与否在很大程度上取决于商品信息的透明度以及营销员的个人能力; 而病毒营销是一种基于社会网络的 n -to- n 的网络营销模式, 其核心理念是依靠用户去营销用户, 用户不仅是消费者, 更是企业营销的主体。在病毒营销中, 传统营销员的概念被拓展到用户, 并通过用户的“口口相传”(word-of-mouth)进行营销^[6-8]。Hotmail 通过病毒营销在半年时间里就吸引了 1 200 万注册用户, 每天以增加超过 15 万新用户的速度发展, 但营销费用还不到其直接竞争者的 3%。Dierkes 等人研究发现, 口口相传对移动电话市场客户的退出决定(churn decisions)和购买决策分别起到 19.5% 和 8.4% 的作用^[9]。近年来, 许多电子商务企业纷纷尝试使用病毒营销, 如 Amazon、Taobao、Dangdang 等。随着未来社会网络的逐步扩大, 病毒营销将成为未来非常重要的营销模式^[10,11]。

自病毒营销的概念提出以来, 有学者从不同角度对病毒营销进行了研究, 如从案例的角度, 采用实证分析方法对病毒营销的意义、特点、用户动机、实施步骤、关键成功因素等进行研究^[12,15,16]; 有的学者从微观角度, 采用定量方法对社会网络中用户影响关系进行研究, 采用社会网络分析的方法对用户影响力进行研究^[12-14]。但据笔者所知, 还少有学者对病毒营销的效果进行预测和分析; Bampo 等

* 基金项目: 研究得到教育部人文社会科学研究青年基金(No. 11YJC630268)、教育部科技发展中心网络时代的科技论文快速共享专项研究(No. 2010118)、国家自然科学基金(No. 70871115)、中国人民大学科学研究基金(中央高校基本科研业务费专项资金资助)、互联网金融数据挖掘与知识管理研究项目(No. 10XNI008)等资助。

通信作者: 余力(1973—), 副教授, 中国人民大学信息学院, 研究兴趣: 社会化网络营销、社会商务。E-mail: buaayuli@ruc.edu.cn.

研究了社会网络结构对病毒营销的影响^[15], Ralf 等借用疾病传播模型,从宏观上对病毒营销的效果进行了分析^[16],但上述研究都无法回答传播效率问题,即在一定时间范围内的影响。实际上企业在决策是否采用病毒营销时,营销主管一定想知道“企业采用病毒营销策略后,一个月或半年后到底能成功营销到多少用户?”本文把该问题提炼为病毒营销传播效率问题,其形式化描述如下:

假定有某病毒营销网络 $G(V, E)$ (其中, $|V| = N$ 为总用户数), 对给定的初始营销用户集合 S_0 (初始传播源), 经过一段时间 t (传播时间) 后, 平均能影响的用户数 $E_{S_0}(t)$ 是多少?

本文将紧紧围绕此问题展开研究, 通过引入影响响应时间的概念, 分析介绍病毒营销网络的构建, 然后给出本文的研究思路和研究假设; 在分别研究两种特殊的病毒营销网络的传播效率问题的基础上, 研究一般病毒营销网络的传播效率问题; 并通过一个实例进行展示, 最后给出结论, 还针对本文的局限性指出未来研究的方向。

2 带有影响响应时间的病毒营销传播图的构建

2.1 影响传播图的构建

病毒营销的传播都是基于传播网络图进行分析的, 而用户间的影响关系是建立病毒营销传播网络图的关键之一。从理论上讲, 要完全准确地得到两个用户间的影响关系几乎是不可能的。但是, 仍然可以通过用户间的行为关系来估算其影响关系, 这种行为泛指一切互联网行为, 既包括用户购买行为, 也包括浏览、添加好友等行为。例如, 如果 Tom 经常在知道了其好友 John 购买某商品后才购买该商品, 则认为在此商品的购买上, Tom 受 John 的影响, 如果在很多商品的购买上 Tom 受 John 的影响, 则我们很有理由认为 John 对 Tom 的影响很大。如何对这种影响关系进行建模? 我们假设这种影响关系的发生存在两个条件: 第一, Tom 和 John 事先已经是好友关系; 第二, John 的购买时间要早于 Tom。此外, 一般的社会网络系统会为用户提供好友的活动信息, 如购买、浏览信息等, 所以假设用户可以容易地知道其好友的行为。目前有许多社交网都允许用户显式添加其他用户为自己的好友, 如淘宝网、QQ、人人网和 Facebook 等, 如图 1 所示的就是淘宝网邀请好友界面和好友列表界面。根据好友关系, 就可以得到所有用户的社会关系网络图, 如图 2 所示。



图 1 淘宝网上好友关系建立

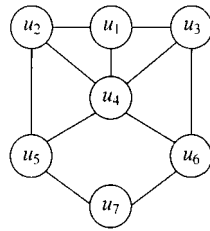


图 2 用户的社会关系图

如果用户 u 在时间 t 购买了商品 g , 则用户的此购买行为可以定义为用户 u 的一个交易(购买)行为, 所有用户的交易行为集合记为 Actions, 如表 1 所示。

表 1 用户的交易集

用户	商品	购买时间	用户	商品	购买时间
u_3	b	13	u_5	b	16
u_2	a	15	u_5	a	18
u_1	b	15	u_4	a	19
u_3	a	16	u_6	a	20

依据用户的社会关系图和交易集, 可以为每个商品 g 定义传播图 $PG(g) = (V(g), E(g))$, 其中 $V(g) = \{u | \exists t: (u, g, t) \in \text{Actions}\}$, 每个节点对应已购买商品 g 的用户, 边表示用户间的影响关系。如果用户 u_i 与用户 u_j 是好友关系, 且都购买了商品 g , 但在购买时间上用户 u_i 比用户 u_j 早 Δt , 则在图 $PG(g)$ 上有一条有向边 $v_i \xrightarrow{\Delta t} v_j$, 其中节点 v_i 对应用户 u_i , 并记作 $\text{prop}(g, u_i, u_j, \Delta t)$, 如图 3 所示。 GS_v 记作用户 v 购买的所有商品 (GS 意为 Goods), $GS_{u \rightarrow v}$ 表示由于受到用户 u 的购买影响, 用户 v 购买的商品集合。至此, 可以构建用户间的影响概率 $p_{u \rightarrow v}$:

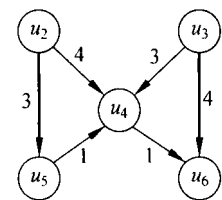


图 3 商品 a 的影响传播

$$p_{u \rightarrow v} = \frac{GS_{u \rightarrow v}}{GS_u \cap GS_v} \tag{1}$$

2.2 影响响应时间的引入

实际上, 在病毒营销传播过程中, 用户间除了影响概率外, 还有影响响应时间问题。影响响应时间反映了一个用户要成功影响另一个用户所需要的平均时间。影响响应时间与影响概率是两个完全不同的概念。影响响应时间不但取决于影响用户的影响力, 可能还更与被影响用户的性格特征有关。比如, 性格内向的消费者可能更理智, 选择购买某产品可能比较犹豫, 其影响响应时间更长; 外向性格的用户可能影响时间较短。此外, 影响时间还与所购产品的类型有关。一般来讲, 对价格高的产品, 如家电、汽车等, 用户间的影响响应时间较长, 而价格低廉的产品, 如图书、CD、电影等, 用户间的影响响应时间较短。

根据上述影响传播图, 为计算用户间影响响应时间, 可以以两用户间购买某商品的时间差作为影响响应时间。具体来讲, 假如对于产品 g , 用户 u 与用户 v 存在影响传播关系 $\text{prop}(g, u, v, \Delta t)$, 则在产品 g 上, 两用户间的影响响应时间 $t_{u \rightarrow v}^g = \Delta t$, 考察用户 u 对用户 v 的所有影响, 则可计算用户 v 对用户 u 的影响响应时间 $t_{u \rightarrow v}$ 为

$$t_{u \rightarrow v} = \frac{\sum_{k \in GS_{u \rightarrow v}} t_{u \rightarrow v}^k}{|GS_{u \rightarrow v}|} \quad (2)$$

有了影响响应时间,即可预测病毒营销传播效率。

3 研究思路与假设

为了回答本文提出的问题,把该病毒营销传播效率问题转化为如下:

假设有某病毒营销网络 $G(V, E)$, 对于给定的初始营销用户集合 S (初始传播源), 经过一段时间 t (传播时间) 后, 能影响 m 个用户的概率 $f_s(t, m)$ 是多少?

在得到上述概率之后, 即可通过如下公式, 得到经过时间 t 后平均影响的用户数 $E_s(t)$:

$$E_s(t) = \sum_{m=0}^{m=|V|-|S|} m \times f_s(t, m) \quad (3)$$

为研究方便, 本文的研究基于以下假设。

假设 1: 影响响应时间等长, 即产品由一个用户成功营销给另一个用户所用时间等长, 理论上不同用户间的营销传播的时间不尽相同。本文为研究方便, 假定传播时间是等长的, 并记为一个单位时间。

假设 2: 传播影响概率相等, 即任两个节点间具有相同的影响概率 p 。

假设 3: 影响的时效性, 即某用户 A 对另一个用户 B 的影响是有一定时效性的。通俗地讲, 如果给足够的时间 t 后, 用户 A 仍然不能影响用户 B , 那么给再多的时间也影响不了。这一点假设还是比较吻合现实生活的, 比如张三向其好友推荐了一本新书, 如果该好友在一个月之内没有购买, 则我们很可能认为他以后也不会购买。

假设 4: 单传播源, 即认为只有一个初始的传播源节点, 而实际上, 病毒营销可以有多个传播源节点, 但我们可以合并为 1 个虚拟的源节点, 并重新构建传播图, 如图 4 所示。

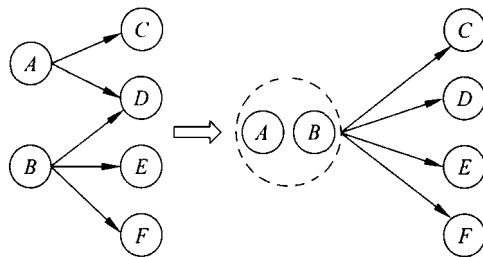


图 4 合并多个初始源节点为一个虚拟源节点

4 串联影响传播图的传播效率

串联影响是指每个用户只受一个用户的影响, 也只能影响一个用户, 用户与用户间形成串联, 这是一种最简单的影响传播图, 每个用户间的影响概率为 p , 如图 5 所示。若以 S 节点为病毒营销的初始传播源, 则接下来我们需要回答: 经过 t 个时间单位的传播, 能影响 m 个用户的概率为多少?



图 5 串联影响传播示意

利用概率论方法,可以较容易计算 t 时间后影响 m 人的概率,如表 2 所示。

表 2 串联影响传播在不同时间段的传播结果

传播时间	$m=0$	$m=1$	$m=2$	$m=3$...	$m=N$	期望人数
1	$1-p$	p	0	0	...	0	p
2	$1-p$	$p(1-p)$	p^2	0	...	0	$p+p^2$
3	$1-p$	$p(1-p)$	$p^2(1-p)$	p^3	...	0	$p+p^2+p^3$
...
t	$1-p$	$p(1-p)$	$p^2(1-p)$	$p^3(1-p)$...	$p^{n-1}(1-p)$	$p+p^2+\dots+p^n$

上述结果可以统一表示为:以 S 作为源节点,经过 t 个单位时间后影响 m 个用户的概率 $f_s(t,m)$ 为

$$f_s(t,m) = \begin{cases} p^m(1-p), & m < t \\ p^m, & m = t \\ 0, & m > t \end{cases} \quad (4)$$

根据上述概率密度函数,可以得到以下结论。

(1) 经过 t 个单位时间后至少影响 m 个用户的概率 $F_s(t,m)$ 为

$$F_s(t,m) = \sum_{i \geq m} f_s(t,i) \quad (5)$$

当 $m < t$ 时,

$$\begin{aligned} F_s(t,m) &= 1 - (1-p)(1+p+p^2+\dots+p^{m-1}) \\ &= 1 - (1-p) \frac{p(p^{m-1}-1)}{p-1} \\ &= 1 + p(p^{m-1}-1) \\ &= 1 - p + p^m \end{aligned}$$

当 $m = t$ 时,

$$F_s(t,m) = p^m$$

当 $m > t$ 时,

$$F_s(t,m) = 0$$

所以

$$f_s(t,m) = \begin{cases} 1-p+p^m, & m < t \\ p^m, & m = t \\ 0, & m > t \end{cases} \quad (6)$$

(2) 经过 t 个单位时间后影响用户的期望人数 $E_s(t)$ 为

$$E_s(t) = p + p^2 + \dots + p^t = \frac{p(p^t-1)}{p-1} \quad (7)$$

5 并联影响传播的传播效率

并联影响表示一个用户可以同时影响多个用户,如图 6 所示,假设初始源节点 A 有 K 个子节点,其中 S^i 表示节点 S 的第 i 个子节点。同样需要回答:经过 t 个时间单位的传播,能影响 m 个用户的

概率为多少?

利用概率知识,容易得到第 1 个单位时间后,影响 m 个人的概率

$$f_s(1, m) = C_K^m p^m (1-p)^{K-m} \quad m = 0, 1, 2, \dots, K \quad (8)$$

根据影响传播的时效性假设,可以知道

$$\forall t > 1 f_s(t, m) = f_s(1, m) = C_K^m p^m (1-p)^{K-m} \quad (9)$$

所以,根据上述公式可以得到以下结论。

(1) 第 t 个单位时间后,至少影响 m 个人的概率

$$\begin{aligned} F_A(t, m) &= \sum_{j=m}^{j=K} f_s(t, j) \\ &= \sum_{j=m}^{j=K} C_K^j p^j (1-p)^{K-j} \end{aligned} \quad (10)$$

(2) 第 t 个单位时间后,平均期望能影响的人数是

$$\begin{aligned} E_s(t) &= \sum_{j=0}^{j=K} j \times f_s(t, j) \\ &= \sum_{j=0}^{j=K} j \times C_K^j p^j (1-p)^{K-j} \end{aligned} \quad (11)$$

在研究了串联和并联传播网络图后,接下来就可以研究一般的传播网络图。

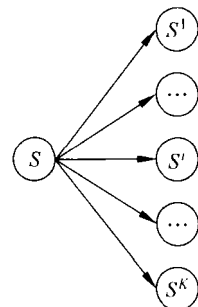


图 6 并联影响传播

6 一般影响传播的传播效率

更一般的影响传播图是包含并联和串联形状的网络图。根据本文假设,我们考虑的是单个传播源,所以一般的传播网络图实际上可以转化成树形的传播图。我们先研究源节点 S 包含两个子节点的情况,如图 7 所示,其中节点 A 和 B 后的虚线框表示其各自影响的子树。

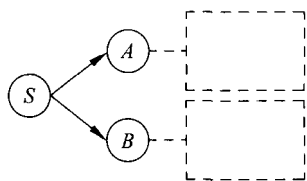


图 7 源节点包含两个节点的传播网络

根据组合概率知识,可以得到

$$f_s(t, m) = \sum_{i=0}^{i=m} \text{from}_A(t-1, i) \times \text{from}_B(t-1, m-i) \quad (12)$$

其中, $\text{from}_X(t-1, i) = \begin{cases} 1-p, & i=0 \\ p \times f_X(t-1, i-1), & i>0 \end{cases}$ 表示来自节点

X 的影响部分,具体指其在时间 $(t-1)$ 内影响 i 个用户的概率。

当源节点 S 包含任意多个子节点时,上述公式可以转化为

$$f_s(t, m) = \sum_{\forall \varphi} \prod_{i=1}^{|\text{son}(s)|=1} \text{from}_{S^i}(t-1, n_{S^i}) \quad (13)$$

式中 φ 为 $n_{S^1}, n_{S^2}, \dots, n_{S^{\text{son}(s)}}$ 任意一组合,组合必须满足

$$n_{S^1} + n_{S^2} + \dots + n_{S^i} + \dots + n_{S^{\text{son}(s)}} = m$$

其中 n_{S^i} 表示 S 的影响人数中来自 S^i 影响的贡献,

$$\text{from}_{S^i}(t-1, n_{S^i}) = \begin{cases} 1-p, & n_{S^i} = 0 \\ p \times f_{S^i}(t-1, n_{S^i}-1), & n_{S^i} > 0 \end{cases} \quad (14)$$

如果概率不均等,则

$$\text{from}_{S^i}(t-1, n_{S^i}) = \begin{cases} 1-p_{S \rightarrow S^i}, & n_{S^i} = 0 \\ p_{S \rightarrow S^i} \times f_{S^i}(t-1, n_{S^i}-1), & n_{S^i} > 0 \end{cases} \quad (15)$$

7 实例

为更好地说明上述方法,本节以一实例说明影响传播效率。假设有一个社区网络,包含有 10 个节点用户,各节点用户间影响概率和影响响应时间如图 8 所示,如节点 A 到节点 D 之间,节点 A 对节点 D 的影响概率是 50%,影响响应时间为 2 天。假设 A 为传播源节点。同时我们把影响阈值定为 0.6,即节点 u 的邻居节点对其影响概率和大于 0.6,则节点 u 就会被影响,否则不然。下面分别考察从节点 A 开始,第 n 天后能够影响到的节点。

第 0 天: 图中的活跃节点只有 A(初始传播节点)。

第 1 天: A 可以影响到 B, C, D, E, F, 其中只有对 B 和 C 的影响概率大于影响阈值,所以 B, C 被成功影响,至此能够影响到的节点为 {A, B, C}。

第 2 天: 活跃节点集合 {A, B, C} 可以影响 D, E, F, G, H, I, 只有 D 被成功影响,所以到第 2 天,能够影响到的节点为 {A, B, C, D}。

第 3 天: 活跃节点集合 {A, B, C, D} 可以影响 E, F, G, H, I, J, 只有 E 被成功影响,所以到第 3 天,能够影响到的节点为 {A, B, C, D, E}。

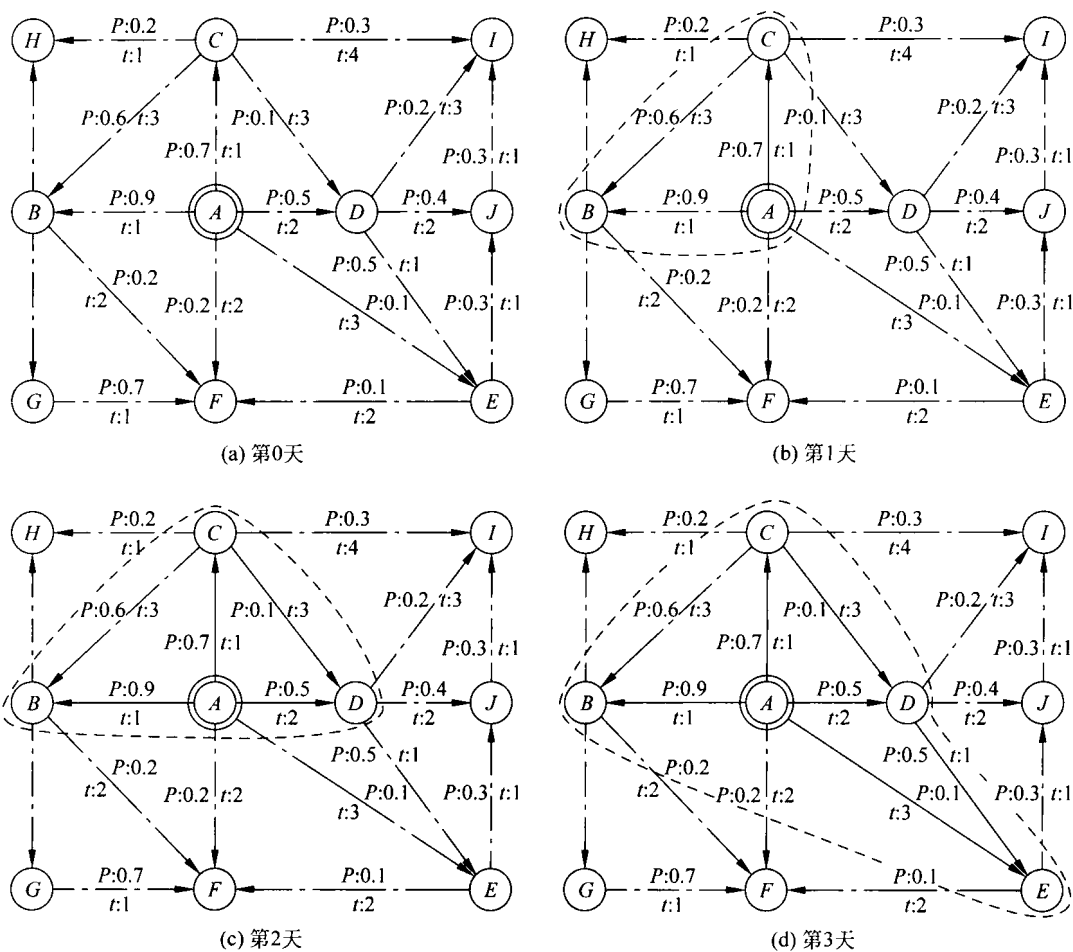


图 8 影响传播实例

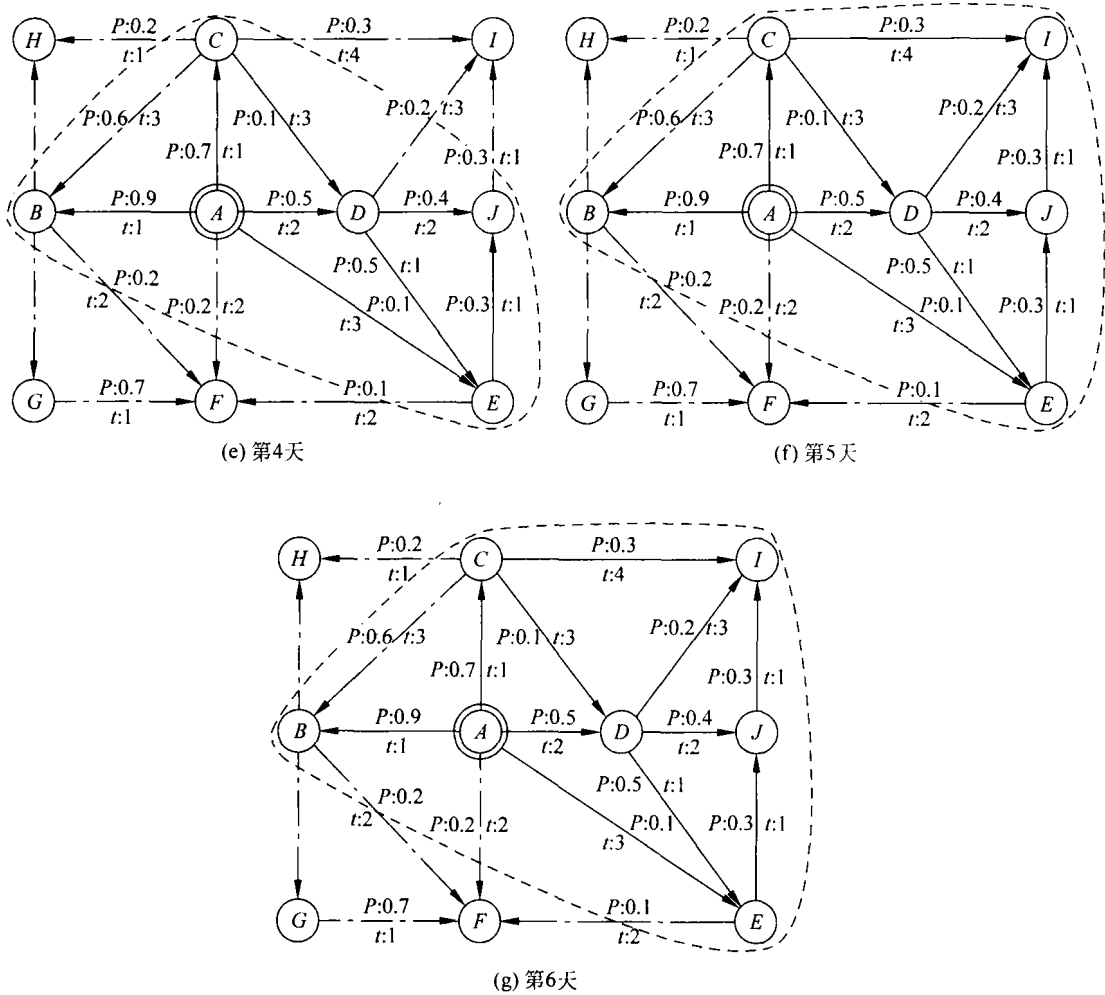


图 8 (续)

第 4 天: 活跃节点集合 $\{A, B, C, D, E\}$ 可以影响 F, G, H, I, J , 只有 J 被成功影响, 所以到第 4 天, 能够影响到的节点为 $\{A, B, C, D, E, J\}$ 。

第 5 天: 活跃节点集合 $\{A, B, C, D, E, J\}$ 可以影响 F, G, H, I , 只有 I 被成功影响, 所以到第 5 天, 能够影响到的节点为 $\{A, B, C, D, E, J, I\}$ 。

从第 6 天开始, 没有新的节点能够被成功影响, 所以从节点 A 开始, 最终影响传播到的节点为 $\{A, B, C, D, E, J, I\}$, 至此, 传播过程结束。

从该例子可以看出, 在不同时间, 传播效率不完全相同, 如在第 1 天, 成功影响了两个节点用户, 而其他时间, 只影响了 1 个节点。相反, 影响传播到一定时间, 如第 6 天, 即使给定足够时间, 也不能继续扩大影响范围, 即并不是所有节点都能被影响到。

8 结论

病毒营销是一种基于社会网络的社会化营销方式, 在社会网络发达的今天, 已经成为一种非常重要的营销方式。本文通过引用影响响应时间, 对基于社会网络的病毒营销传播效率进行了预测。病

毒营销传播效率是病毒营销应用面对的一个重要问题,也是一个非常具有挑战性的问题,本文对此进行了初步探索,希望能抛砖引玉,引起更多学者的兴趣。本文的研究附加了多个假设条件,比如传播等时、影响概率相等,如果去掉这些假设,也许研究更贴近实际,当然研究难度会更大,但这一定是未来研究需要突破的地方。

参考文献

- [1] Andrew S and Toubia O. Deriving value from social commerce networks[J]. *Journal of Marketing Research*, 2010,47(2): 215-228.
- [2] Valck K,Bruggen G H and Wierenga B. Virtual communities: A marketing perspective[J]. *Decision Support Systems*,2009,47 (3): 185-203.
- [3] Sinan A and Dylan W. Creating social contagion through viral product design: A randomized trial of peer influence in networks[J]. *Management Science*,2011,57(9): 1623-1639
- [4] Dellarocas C. The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms[J]. *Management Science*,2003,49(10): 1407-1424.
- [5] Ho J Y C and Dempsey M. Viral marketing: Motivations to forward online content[J]. *Journal of Business Research*,2010,63(9-10): 1000-1006.
- [6] Leskovec J,Adamic L A and Huberman B A. The dynamics of viral marketing[C]. In *Proceedings of the 7th ACM conference on Electronic Commerce*,Ann Arbor,USA,2006: 228-237.
- [7] Ferguson R. Word of mouth and viral marketing: Taking the temperature of the hottest trends in marketing[J]. *Journal of Consumer Marketing*,2008,25(3): 179-182.
- [8] Hogg T. Inferring preference correlations from social networks [J]. *Electronic Commerce Research and Applications*,2011,9(1): 29-37.
- [9] Dierkes T,Bichler M and Krishnan R. Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks[J]. *Decision Support Systems*,2011,3(51): 361-371.
- [10] Pousttchi K and Wiedemann D G. Success factors in mobile viral marketing: A multi-case study approach[C]. In *Proceedings of the 6th International Conference on Mobile Business (ICMB'07)*, Toronto, Ontario, Canada, 2007: 1-8.
- [11] Trusov M,Bodapati A V and Bucklin R E. Determining influential users in internet social networks[J]. *Journal of Marketing Research*,2010,47(4): 643-658.
- [12] Ghosh R and Lerman K. Predicting influential users in online social networks[C]. In *The fourth SNA-KDD Workshop Held in conjunction with ACM SIGKDD2010 (SNAKDD'10)*, May 2010.
- [13] Li Y M,Lin C H and Lai Ch Y. Identifying influential reviewers for word-of-mouth marketing[J]. *Electronic Commerce Research and Applications*,2010 (9): 294-304.
- [14] Kempe D,Kleinberg J and Tardos E. Maximizing the spread of influence through a social network[C]. *ACM SIGKDD international Conference on Knowledge Discovery and Data Mining(KDD'03)*,2003: 137-146.
- [15] Bampo et al.. The effects of the social structure of digital networkson viral marketing performance [J]. *Information Systems Research*,2008,19(3): 273-290.
- [16] Ralf van der Lans,Gerrit van Bruggen,Jehoshua Eliashberg and Berend Wierenga. Aviral branching model for predicting the spread of electronic Word of Mouth[J]. *Marketing Science*,2010,29(2): 348-365.

Prediction on Propagation Efficiency of Viral Marketing Based on Social Network

YU Li, YANG Xiaoping, CHEN Mingyuan

(School of Information, Renmin University of China, Beijing, China 100872)

Abstract Viral marketing model which is a newly network marketing one with the development of Web2.0 and social network, has been employed by lots of companies in the recent years. However, is it really effective and fruitful and how many clients were successfully marketed during a fixed period of time? Probability statistical analysis and social network analysis have been adopted in this essay to make the preliminary research of the propagation speeded of the model and the answer to the question: "Given a directed graph $G(V, E)$, for a initial active set S_0 , what is the number of the average influenced users $E_{S_0}(t)$ after the time t ?" has been explored. Finally, an example is shown to illustrate proposed method. The research of this essay plays an important role as the reference value in applying the viral marketing model for the companies.

Key words Viral Marketing, Network Marketing, Web 2.0, Propagation Efficiency, Social Commerce

作者简介

余力(1973—)男,博士,中国人民大学信息学院副教授、硕士生导师,主要研究方向:电子商务、病毒营销、社会商务、Web 2.0 与社会计算。E-mail: buaayuli@ruc.edu.cn。

杨小平(1956—),男,博士,中国人民大学信息学院教授、博士生导师,主要研究方向:信息系统工程、电子政务、网络安全技术等。E-mail: yang@ruc.edu.cn。

陈明远(1989—),女,中国人民大学在读研究生,主要研究方向:病毒营销。E-mail: cmyruc@126.com。