

# 基于 DSSI 理论的信息易感用户动态提取模型

李 玲<sup>1),2)</sup> 刘 敏<sup>1)</sup> 成国庆<sup>2)</sup>

<sup>1)</sup>(同济大学电子与信息工程学院 上海 201804)

<sup>2)</sup>(景德镇陶瓷大学信息工程学院 江西 景德镇 333403)

**摘 要** 在线社交网络信息扩散研究有助于用户获取信息、企业推广产品、政府调控舆情,其理论及应用价值巨大.该文主要针对在线社交网络中商业信息的扩散问题,在分析用户转发行为随机性和不确定性的基础上,利用通用发生函数(Universal Generating Function, UGF)方法及离散压力-应力(Discrete Stress-Strength Interference, DSSI)理论,提出了一种新颖的信息易感用户动态提取模型.模型中,首先将用户的随机转发行为量化为节点易感性(Node Susceptibility, NS),然后基于 UGF 方法及 DSSI 理论计算出不同用户不同时期关于不同信息的 NS 值,最后通过 NS 值的动态排序提取信息易感用户.模型的决策结果有效地解决了以下三个问题:(1)哪些用户最易受信息影响?(2)他们最易受哪类信息影响?(3)他们何时最易受影响?这三个问题的答案可为有效信息扩散策略的制定提供理论依据.案例分析和模型对比说明了该模型的可行性及有效性.

**关键词** 信息扩散;易感用户;压力-应力干涉理论;在线社交网络;动态提取

中图法分类号 TP393

DOI号 10.11897/SP.J.1016.2017.02812

## A Dynamic Extraction Model for Susceptible Users Based on DSSI Theory

LI Ling<sup>1),2)</sup> LIU Min<sup>1)</sup> CHENG Guo-Qing<sup>2)</sup>

<sup>1)</sup>(College of Electronics and Information Engineering, Tongji University, Shanghai 201804)

<sup>2)</sup>(Department of Information Engineering, Jingdezhen Ceramic Institute, Jingdezhen, Jiangxi 333403)

**Abstract** Fast growing social networks have been integrated into people's daily lives and play an important role, which makes more and more academics study the social networks from different perspectives. The research of information diffusion over online social networks can help users to obtain information, enterprises to promote product, politicians to regulate public opinion, and is with the significant value in theory and application. In the recent years, although there have been a number of significant advancements on information diffusion, most of them have been mainly focusing on optimization algorithm for user extraction, or evolution equations for behavior law. No attempts have been made to quantify user susceptibility through the forwarding action, and dynamically extract susceptible users. In fact, at different times, the influence of users on information diffusion cannot be exactly the same due to the uncertainty and complexity of social networks. In this sense, dynamic analysis and study of user forwarding action is of great importance, and is precisely what we do in this paper. To address dynamically the business information diffusion problem over online social networks, the randomness and uncertainties of user forwarding action are first analyzed, and then a novel dynamic extraction model for the susceptible users is presented, which is based on Universal Generating Function (UGF) method and Discrete Stress-Strength Interference (DSSI) theory. In the model, the random forwarding action of the user is

firstly quantified as Node Susceptibility (NS), and NS is relevant to two random variables of information receiving  $X_{nmt}$  and forwarding  $Y_{nmt}$ . Then, according to UGF method and DSSI theory, the values of NS in regard to different kinds of information at different periods are obtained by deriving the probability distributions of  $X_{nmt}$  and  $Y_{nmt}$ , and UGFs of  $X_{nmt}$  and  $Y_{nmt}$ . Finally, the susceptible users are extracted based on dynamic order of the values of NS. The decision results of the model can effectively address the following three issues: (1) the most susceptible users; (2) the kinds of information that they are most susceptible to; and (3) the period when they are most susceptible. The answers to these three questions can provide theoretical basis for making effective strategy of information diffusion, and the decision results can be updated dynamically with the observation parameters. A case study of online group buying website illustrates the feasibility and practicality of the proposed model. Further, based on the same experimental data set, the proposed model is compared with Influence-Susceptibility-Cynical (ISC) model in literature, and different susceptible users are extracted based on ISC model and our model. The results show that the susceptible users extracted in these two models are roughly the same, and the consistency is more than 70 percent. The consistency also illustrates the validity of our model to some extent. On the other hand, the difference between user extractions in different models is analyzed from both theoretical and practical perspectives. Since the quantification of user susceptibility in our model is based on the statistics characteristics of observation parameters, it is concluded that our model is more scientific and reasonable in quantifying user susceptibility.

**Keywords** information diffusion; susceptible user; stress-strength interference theory; online social networks; dynamic extraction

## 1 引言

随着互联网和移动通信技术的不断发展,在线社交网络已迅速成为最受欢迎的应用,生活中无处不在<sup>[1-2]</sup>. 社交网络由社会角色在信息空间中相互连接而构成,这些社会角色可以是个人、组织、网站,甚至智能设备<sup>①</sup>. 生活中典型的社交网络例子有微信、Twitter、Facebook、微博等. 社交网络最突出的特征是同时共享及交流信息<sup>[3-4]</sup>. 快速发展的社交网络已经融入到人们的日常生活并且发挥着重要的影响,例如,社交网络促进了电子商务的发展,政府机构通过社交网络征集民意信息,消费者通过社交网络购买、评价产品等<sup>[5-6]</sup>. 也正是由于这些影响,越来越多的学者从不同的角度对社交网络进行研究,其中,网络结构建模<sup>[6]</sup>、动态演化建模<sup>[7]</sup>及信息扩散特征分析是该领域的研究热点<sup>[8]</sup>.

当前,在线社交网络日益受到网民及商家的追捧. 2015年8月,Facebook的每月活跃用户数量已达到11.8亿<sup>②</sup>. 每天,海量的用户通过“口口相传”(信息转发)的通信方式分享和交流各种信息. 在这

种背景下,为更好地研究社交网络中的信息扩散,对用户的随机行为(转发或者不转发信息)进行量化和深入分析就显得很重要了<sup>[9]</sup>. 在社交网络中,信息扩散过程的本质是用户转发其收到的信息的过程,而且转发的频率在一定程度上影响信息扩散的速度. 然而,对一个用户而言,转发信息不是必然的,而是受多种不确定因素(如信息的趣味性、可靠性,利己动机、利他动机等)的影响<sup>[10]</sup>. 我们虽然不能确定是哪一种因素导致了信息的转发,但是转发这一结果已经表明用户受影响于相关信息,或者说他易感于相关信息;并且,转发频率越高表明易感程度越大. 因此,对这一随机结果(用户的随机转发行为)进行量化研究,会有助于相关决策者确定易感用户,进而制定最有效的信息扩散策略<sup>[4]</sup>. 另一方面,不良信息(如计算机病毒、谣言等)的扩散对社会危害巨大,及时发现易感用户并阻断信息传播刻不容缓<sup>[11]</sup>. 因此,从信息安全的角度来讲,对用户随机转发行为的研究也有助于阻止有害信息的传播与扩散.

① Wikipedia, Social Network, [Online]. Available: [http://en.wikipedia.org/wiki/Social\\_network](http://en.wikipedia.org/wiki/Social_network)

② Wikipedia, Facebook, [Online]. Available: <https://en.wikipedia.org/wiki/Facebook>

近年来,通信及信息技术的快速发展使我们能够获得和存储现实生活中的各种大数据,使有关信息扩散的研究更有意义,也更具挑战性.追溯到2001年,Pastor-Satorras 和 Vespignani<sup>[12]</sup>首次研究了(计算机)病毒在网络空间的传播,这也表明关于信息扩散的理论研究起源于对(计算机)病毒传播的研究.最早关于信息扩散的研究之一是文献[13],它从两种角度(宏观和微观)研究和分析了网络空间信息传播的动力学模型.随后,出现了许多关于社交网络信息扩散的研究成果.在这里,我们总结与本文相关的最具代表性的研究成果.已有关于社交网络信息扩散的研究成果大致可分为两类.第一类关注于研究网络结构对信息扩散的影响<sup>[14-18]</sup>.在文献[14]中,作者研究了一些大型网络的标度律,指出网络机制的优化设计可以生成无标度结构.吴泓润等人<sup>[15]</sup>以无标度属性为优化目标,类别距离为约束条件,建立了相应的优化模型,揭示了社区特征与类别距离间的因果关系.类似地,Lin 和 Li<sup>[16]</sup>也指出无标度网络可以优化网络性能.最近,文献[17]研究了微博的网络结构如何影响信息扩散,通过分析微博用户粉丝的拓扑结构,作者指出其数量服从一种指数近似为2的幂律形式的分布.同样以微博为研究背景,Park 等人<sup>[18]</sup>对比地探讨了网络结构,用户的地域分布,以及信息交互模式,并指出信息可以由一些具有桥梁功能的核心用户来组织,这些用户可以在不同社区间进行有效的信息扩散.

第二类研究成果是利用不同的数学模型量化和分析社交网络用户(节点)对信息扩散的影响<sup>[19-26]</sup>.Kimura 等人<sup>[19]</sup>研究了从一个社交网络中提取最具影响力节点的优化问题.同样以此优化问题为切入点,Wang 等人<sup>[20]</sup>提出了一种基于社区的贪婪算法,挖掘出移动社交网络中最具影响力的 $k$ 个节点.后来,曹玖新等人<sup>[21]</sup>针对这一问题又提出了基于核数层次特征及影响半径的核覆盖算法,找出最具影响力节点的集合.Ilyas 等人<sup>[22]</sup>则从确定关键信息节点的角度,研究了如何抑制隐私信息扩散的问题.最近,Saito 等人<sup>[23]</sup>提出了在社交网络中筛选一种新类型影响力节点——super-mediators的方法,并研究了该类节点的信息扩散性质.考虑到网络结构的复杂性,一些学者将其与传染病动力学理论结合起来,建立信息扩散的动力学演化方程组<sup>[24-26]</sup>.张彦超等人<sup>[24]</sup>对不同类型节点随着时间的演化关系进行建模,分析了不同类型节点在社交网络中的行为规律.Ma 等人<sup>[25]</sup>和顾亦然等人<sup>[26]</sup>则通过信息传播的

动力学演化模型,提出了抑制谣言传播的控制策略.本文的研究工作也属于第二类,主要着眼于对用户(节点)的随机转发行为进行数学建模,并分析这些随机行为对信息扩散的影响.虽然文献[19-21]分别提出了不同的算法来提取最具影响力的节点,但它们的侧重点在于优化算法的计算成本,而没有考虑决策结果的动态性.而文献[24-26]主要通过从宏观角度建立信息扩散的动力学演化方程组,来估计社交网络信息的扩散规律或用户的行为规律,而这些规律一般也是静态的.

事实上,在不同的时期,由于社交网络的复杂性和不确定性,用户对其信息扩散的影响程度不可能完全一致.例如,社交网络用户的作息时间及上网习惯、信息的产生时间及发布渠道都具有随机性.在这些随机因素的作用下,用户的行为也是随机的,从而导致其对信息扩散影响程度的动态性.另一方面,现有文献大都关注于挖掘和发现社交网络中具有影响力的用户,对于易受影响用户(易感用户)的提取问题的研究并不多.Aral 和 Walker<sup>[27]</sup>提出了基于网络样本随机实验的方法,并结合比例危险率模型来确定其中的易感用户,得出的结论是年轻用户比年长用户更具易感性,而且已婚用户的易感性最低.Lee 和 Lim<sup>[28]</sup>首先分析了潜在的用户因素对网络中带有情感的信息扩散的影响,然后提出了 Influence-Susceptibility-Cynical (ISC)模型,并利用迭代算法进行求解.类似地,Hoang 和 Lim<sup>[29]</sup>也研究了潜在的用户行为因素,建模并分析了用户易感性对于社交网络中病毒扩散的作用及影响.与上面提到的两篇文献不同,Chen 等人<sup>[30]</sup>主要通过网络问卷调查的形式,获取并量化用户对于不同来源的在线品牌信息的态度和产品购买意向,以此来分析不同用户的易感性.

可以看出,在对用户易感性进行量化时,数学模型(文献[27-29]均运用了数学模型)是常用的工具.此外,尽管上述文献都以量化用户易感性为手段,但由于决策目标的不同,其模型的表达式及求解方法也各不相同.一般而言,在模型复杂度相当的情况下,含有较少决策参数的模型更有利于决策者(实践者)在实际中实施.原因在于,求解带有较多输入参数的模型不仅要求决策者掌握较多的专业知识,还需要在决策过程中注入较多的主观判断,这不仅会增加决策过程的难度,也会影响决策结果的客观性.另外,决策结果的动态性也是衡量模型实用性的标准之一.正如前面所提到的,由于社交网络的复杂性

和不确定性,用户对其信息扩散的影响程度不可能完全一样,在不同时期,用户对于不同种类信息的易感性也会有所不同.因此,动态地确定易感用户更有助于有效信息扩散策略的制定.

基于以上考虑,本文拟从以下两个方面为切入点来研究信息易感用户的提取问题:(1)在充分考虑用户转发行为随机性的同时,尽量简化其提取指标与决策过程;(2)决策结果应该是动态的,可在不同时期确定出不同种类信息的易感用户.为此,本文基于通用发生函数(Universal Generating Function, UGF)方法及离散压力-应力(Discrete Stress-Strength Interference, DSSI)理论,提出了一种新颖的信息易感用户动态提取模型.压力-应力(Stress-Strength Interference, SSI)理论是可靠性分析领域处理不确定因素的有力工具.但是,据我们所知,还没有相关文献将其应用于信息扩散的分析与研究.

基于 DSSI 理论,模型首先将用户转发行为量化为节点易感性(Node Susceptibility, NS),NS 只与两个随机变量有关:接收的信息数量和转发的信息数量,NS 被定义为后者(应力)大于前者(压力)的概率.然后利用 UGF 方法推导 NS 的解析表达式,进而计算相应节点的 NS 值.基于 NS 值,本文模型可以解决以下三个问题:(1)哪些用户是信息易感用户?(2)他们易感于哪类信息?(3)他们在何时易感程度最高?这三个问题的回答可以帮助决策者动态地提取信息易感用户,从而制定更为合理有效的信息扩散策略.与已有的信息易感用户提取模型相比,本文模型具有以下三个优点:

(1)唯一且客观的提取指标:将具有高随机性的转发行为量化为唯一的提取指标—用户易感性 NS,即只需通过客观的 NS 值便可提取易感用户,不需注入过多决策者的主观判断.

(2)动态的决策结果:随着信息转发数据的更新,其决策结果(包括易感用户、易感信息类型、易感时段)也会随之更新,更有助于决策者制定有效的信息扩散策略.

(3)简易的实施过程:在现实中,实践者可直接从相应数据库(例如用户日志)调用输入参数进行模型求解,不需过多的专业知识(例如参数估计,优化算法等)来进行优化决策,操作起来相对简易.

本文第 2 节介绍模型的理论背景;在第 3 节,建立基于 DSSI 理论的信息易感用户动态提取模型;第 4 节通过案例分析演示模型的可行性及有效性;第 5 节对全文进行总结及对下一步工作的展望.

## 2 理论背景

在建立数学模型之前,我们先介绍通用发生函数 UGF 方法及离散压力-应力 DSSI 理论的相关定义及符号,它们在第 3 节将会被用到.

### 2.1 UGF 方法

这里,我们只介绍 UGF 方法的基本概念,而不涉及其中的数学原理.UGF 的概念最早是由 Ushakov<sup>[31]</sup>提出的.后来, Lisnianski 等人<sup>[32]</sup>和 Levitin<sup>[33]</sup>将 UGF 方法应用于可靠性分析以及多状态系统的优化.

假设  $X$  是一个离散的随机变量,其概率分布可由包含其可能取值的向量  $\mathbf{x}$  及包含其相应概率值的向量  $\mathbf{p}$  确定,即  $\mathbf{x}=(x_1, x_2, \dots, x_k)$ ,  $\mathbf{p}=(p_1, p_2, \dots, p_k)$ ,其中  $p_i=\Pr(X=x_i)$ ,  $i=1, 2, \dots, k$ .

**定义 1.** 离散随机变量的 UGF.

离散随机变量  $X$  的 UGF 定义为变量  $z$  的多项式函数  $u_X(z)$ ,而且

$$u_X(z) = p_1 z^{x_1} + p_2 z^{x_2} + \dots + p_k z^{x_k} = \sum_{i=1}^k p_i z^{x_i} \quad (1)$$

应该指出的是对于任意一个离散变量,其概率分布与 UGF 之间存在着——对应关系,换句话说,它的 UGF 由概率分布唯一确定.

**定义 2.** 离散随机变量函数的 UGF.

考虑  $n$  个独立的离散随机变量  $X_1, X_2, \dots, X_n$ , 其 UGF 分别为  $u_{X_1}(z), u_{X_2}(z), \dots, u_{X_n}(z)$ . 假设  $f(X_1, X_2, \dots, X_n)$  是变量  $X_1, X_2, \dots, X_n$  一个任意函数,则通过运用复合算子  $\otimes$ , 函数  $f(X_1, X_2, \dots, X_n)$  的 UGF  $u_f(z)$  可表示为

$$u_f(z) = \otimes (u_{X_1}(z), u_{X_2}(z), \dots, u_{X_n}(z)) \quad (2)$$

**定义 3.** 复合算子  $\otimes$ .

根据定义 1,  $u_{X_i}(z) = \sum_{j_i=1}^{k_i} p_{ij_i} z^{x_{ij_i}}$ ,  $i=1, 2, \dots, n$ ,

其中  $k_1, k_2, \dots, k_n$  分别表示每个随机变量可能取值的个数.为得到函数  $f(X_1, X_2, \dots, X_n)$  的 UGF, 复合算子  $\otimes$  定义为

$$\begin{aligned} & \otimes \left( \sum_{j_1=1}^{k_1} p_{1j_1} z^{x_{1j_1}}, \sum_{j_2=1}^{k_2} p_{2j_2} z^{x_{2j_2}}, \dots, \sum_{j_n=1}^{k_n} p_{nj_n} z^{x_{nj_n}} \right) \\ &= \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} \dots \sum_{j_n=1}^{k_n} \left( \prod_{i=1}^n p_{ij_i} z^{f(x_{1j_1}, x_{2j_2}, \dots, x_{nj_n})} \right) \quad (3) \end{aligned}$$

**性质 1.** UGF 的运算满足交换律和结合律:

$$\begin{aligned}
 u_f(z) &= \otimes (u_{X_1}(z), u_{X_2}(z), \dots, u_{X_i}(z), u_{X_{i+1}}(z), \dots, \\
 &\quad u_{X_n}(z)) \\
 &= \otimes (u_{X_1}(z), u_{X_2}(z), \dots, u_{X_{i+1}}(z), u_{X_i}(z), \dots, \\
 &\quad u_{X_n}(z)), \\
 u_f(z) &= \otimes (u_{X_1}(z), u_{X_2}(z), \dots, u_{X_i}(z), u_{X_{i+1}}(z), \dots, \\
 &\quad u_{X_n}(z)) \\
 &= \otimes (\otimes (u_{X_1}(z), u_{X_2}(z), \dots, u_{X_{i+1}}(z)), \\
 &\quad \otimes (u_{X_i}(z), \dots, u_{X_n}(z))).
 \end{aligned}$$

2.2 DSSI 模型

压力-应力 (Stress-Strength Interference, SSI) 模型<sup>[34]</sup>已广泛应用于“部件”的可靠性分析. 这里的“部件”不一定是具体的产品或组件, 也可以是整个系统. 压力-应力分析是可靠性工程领域的一种有效方法.

**定义 4.** 部件可靠性.

假设  $S_1$  和  $S_2$  分别表示作用在部件 (或系统) 上的压力及其自身的应力, 则该部件的可靠性  $R$  可定义为

$$R = \Pr(S_2 > S_1) \quad (4)$$

式(4)是最基本的 SSI 模型表达式, 即部件可靠性定义为应力大于压力的概率.

如果  $S_1$  和  $S_2$  是连续随机变量, 其概率密度函数分别为  $f_1(S_1)$  和  $f_2(S_2)$ , 则式(4)可进一步写成

$$R = \int_{-\infty}^{+\infty} f_1(S_1) \cdot \left[ \int_{S_1}^{+\infty} f_2(S_2) dS_2 \right] dS_1 \quad (5a)$$

$$\text{或者 } R = \int_{-\infty}^{+\infty} f_2(S_2) \cdot \left[ \int_{-\infty}^{S_2} f_1(S_1) dS_1 \right] dS_2 \quad (5b)$$

式(5)被称为连续压力-应力 (Continuous Stress-Strength Interference, CSSI) 模型. 图 1 直观地刻画了 CSSI 模型, 其中连续型随机变量  $S_1$  和  $S_2$  的概率密度函数  $f_1(S_1)$  和  $f_2(S_2)$  位于同一直角坐标系内, 横坐标表示随机变量, 纵坐标表示相应随机变量的概率密度函数. 图中的阴影区域表示两个随机变量

压力和应力的干涉区域.

如果  $S_1$  和  $S_2$  是两个离散随机变量, 它们的概率分布分别如下:

$$S_1 = (S_{11}, S_{12}, \dots, S_{1k_1}), p_1 = (p_{11}, p_{12}, \dots, p_{1k_1}) \quad (6)$$

$$S_2 = (S_{21}, S_{22}, \dots, S_{2k_2}), p_2 = (p_{21}, p_{22}, \dots, p_{2k_2}) \quad (7)$$

其中向量  $(S_{11}, S_{12}, \dots, S_{1k_1})$  和  $(S_{21}, S_{22}, \dots, S_{2k_2})$  的各个分量分别表示  $S_1$  和  $S_2$  的可能取值, 向量  $(p_{11}, p_{12}, \dots, p_{1k_1})$  和  $(p_{21}, p_{22}, \dots, p_{2k_2})$  的各个分量分别表示  $S_1$  和  $S_2$  取相应值的概率, 例如  $S_1$  取值为  $S_{12}$  的概率为  $p_{12}$ .  $k_1$  和  $k_2$  分别表示  $S_1$  和  $S_2$  可能取值的个数. 因此, 根据定义 1,  $S_1$  和  $S_2$  的 UGF 分别为

$$u_{S_1}(z) = \sum_{j_1=1}^{k_1} p_{1j_1} z^{S_{1j_1}}, u_{S_2}(z) = \sum_{j_2=1}^{k_2} p_{2j_2} z^{S_{2j_2}}.$$

如果  $f(S_1, S_2)$  是  $S_1$  和  $S_2$  的一个函数, 则根据上面介绍的 UGF 方法, 可得到函数  $f(S_1, S_2)$  的 UGF 为

$$\begin{aligned}
 u_f(z) &= \otimes (u_{S_1}(z), u_{S_2}(z)) \\
 &= \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} \left( \prod_{i=1}^2 p_{ij_i} z^{f(S_{1j_1}, S_{2j_2})} \right) \\
 &= \sum_{j=1}^K P_j z^{f_j} \quad (8)
 \end{aligned}$$

其中  $f_j$  和  $P_j$  ( $j=1, 2, \dots, K$ ) 分别是函数  $f(S_1, S_2)$  的可能取值和取相应值的概率, 且有  $K \leq k_1 \times k_2$ .

**定义 5.** DSSI 模型.

如果  $f(S_1, S_2) = S_2 - S_1$ , 则部件可靠性可由式(9)计算

$$R = \Pr(f(S_1, S_2) > 0) = \sum_{j=1}^K P_j \sigma(f_j) \quad (9)$$

式(9)被称作 DSSI 模型, 其中  $\sigma(f_j)$  是一个定义在函数  $f(S_1, S_2)$  可能取值集合上的二元值函数, 形如  $\sigma(f_j) = \begin{cases} 1, & f_j > 0 \\ 0, & f_j \leq 0 \end{cases}$ .

需要说明的是, 式(5) (CSSI 模型) 和式(9) (DSSI 模型) 是式(4) (SSI 模型) 的两种不同表现形式. 当压力和应力是连续型随机变量时, 需要用式(5)来计算系统可靠性; 当压力和应力是离散型随机变量时, 则需要用式(9)来计算系统可靠性. 下面通过一个简单的数值例子来说明式(5)和式(9)的应用.

**情形 1.**  $S_1$  和  $S_2$  是两个连续型随机变量.

假设  $S_1$  和  $S_2$  的概率密度函数分别为  $f(x) = \begin{cases} 3e^{-3x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$  和  $f(y) = \begin{cases} e^{-y}, & y > 0 \\ 0, & y \leq 0 \end{cases}$ , 则由式(5a)可

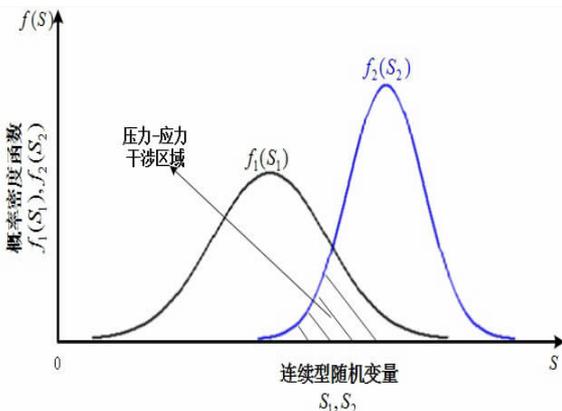


图 1 连续压力-应力干涉 (CSSI) 模型

得系统可靠性

$$R = \int_0^{+\infty} 3e^{-3x} \cdot \left[ \int_x^{+\infty} e^{-y} dy \right] dx$$

$$= \int_0^{+\infty} 3e^{-3x} \cdot e^{-x} dx = 0.75.$$

**情形 2.**  $S_1$  和  $S_2$  是两个离散型随机变量.

假设  $S_1$  和  $S_2$  的概率分布分别为  $S_1 = (12, 15, 18)$ ,  $p_1 = (0.2, 0.5, 0.3)$  和  $S_2 = (14, 16, 20)$ ,  $p_2 = (0.1, 0.4, 0.5)$ . 根据式(9), 可得系统可靠性

$$R = 0.1 \times 0.2 + 0.4 \times 0.2 + 0.4 \times 0.5 +$$

$$0.5 \times 0.2 + 0.5 \times 0.5 + 0.5 \times 0.3$$

$$= 0.8.$$

### 3 基于 DSSI 理论的信息易感用户动态提取模型

在这一节中, 我们将建立基于 DSSI 理论的信息易感用户动态提取模型. 首先, 我们给出模型假设及符号说明. 接着, 基于上节介绍的 DSSI 理论, 用户的随机转发行为被量化为节点易感性(Node Susceptibility, NS), 并给出 NS 的定义. 最后, 制定

决策规则, 提取信息易感用户.

#### 3.1 模型假设及符号说明

考虑一个社交网络及该网络中的多个用户, 这里将第  $n$  个用户记为  $U_n$ . 这些用户可以接收其他对象(包括其他网络用户、商户等)递送的不同种类的信息, 这里的信息是指以产品宣传推广为目的的商业信息. 信息的种类包括休闲娱乐、购物和美食等, 这里将第  $m$  类信息记为  $I_m$ . 为全面地分析用户随机转发行为对商业信息扩散的影响, 该模型将考虑多个时期(这里将第  $t$  个时期记为  $P_t$ ), 以动态地决策以下三个问题: (1) 哪些用户是信息易感用户? (2) 他们易感于哪类信息? (3) 他们在哪个时期易感程度最高? 为建立模型需要, 我们首先给出以下的模型假设.

**假设 1.** 模型考虑多个时期, 且每个时期均包含多个时间节点, 这些节点将相应时期等分为多个时间区间. 图 2 直观地刻画了模型中的时期及时间节点. 模型中共有  $T$  个时期, 每个时期  $P_t$  包含  $L_t$  个时间节点  $Q_{tj}$ , 这些时间节点将  $P_t$  等分为  $L_t$  个时间区间  $[Q_{t,j-1}, Q_{tj}]$ .

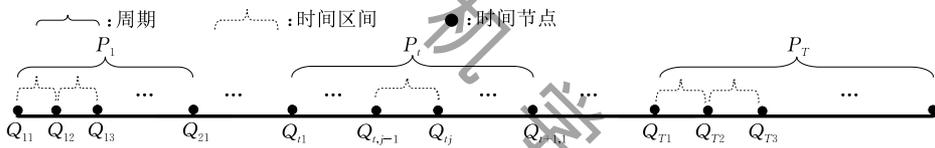


图 2 模型中的时期及时间节点

**假设 2.** 在时期  $P_t$  里, 用户接收或转发某类信息的数量是随机的, 在模型中将其分别设置为两个不同的离散型随机变量.

**假设 3.** 用户在时间区间  $[Q_{t,j-1}, Q_{tj}]$  里接收或转发某类信息的总数量(如果同一信息来自于不同递送者, 或同一信息转发给不同的对象, 则均累计数量)可在时间节点  $Q_{tj}$  通过用户日志(或用户数据)的形式被记录.

为清晰起见, 表 1 给出了上述模型假设中提到的研究对象及相关变量的符号说明. 需要指出的是, 对于离散型随机变量  $X_{nmt}$  和  $Y_{nmt}$ , 需要通过它们的观测参数来刻画其统计特征. 为此, 我们在模型中引入观测参数  $A_{nmtj}$  和  $B_{nmtj}$ , 分别表示用户  $U_n$  在时间区间  $[Q_{t,j-1}, Q_{tj}]$  内接收和转发信息  $I_m$  的总数量. 这两类观测参数是后续决策过程中的主要输入参数.

#### 3.2 基于 DSSI 模型量化用户的随机转发行为

在这一小节中, 首先给出节点易感性 NS 的定义. 然后, 给出 NS 的计算方法.

表 1 符号说明

|                       |   |
|-----------------------|---|
| $n$                   | 用户下标  |
| $m$                   | 信息类型下标  |
| $t$                   | 时期下标  |
| $tj$                  | 时期 $t$ 的时间节点下标  |
| $U_n$                 | 第 $n$ 个用户   |
| $I_m$                 | 第 $m$ 类信息   |
| $P_t$                 | 第 $t$ 个时期   |
| $Q_{tj}$              | 第 $t$ 个时期的第 $j$ 个时间节点   |
| $[Q_{t,j-1}, Q_{tj}]$ | 第 $t$ 个时期的第 $j-1$ 个时间区间   |
| $A_{nmtj}$            | 用户 $U_n$ 在时间区间 $[Q_{t,j-1}, Q_{tj}]$ 内接收信息 $I_m$ 的总数量, 它是在时间节点 $Q_{tj}$ 的观测参数 |
| $B_{nmtj}$            | 用户 $U_n$ 在时间区间 $[Q_{t,j-1}, Q_{tj}]$ 内转发信息 $I_m$ 的总数量, 它是在时间节点 $Q_{tj}$ 的观测参数 |
| $X_{nmt}$             | 用户 $U_n$ 在时期 $P_t$ 接收信息 $I_m$ 的数量, 它是一个随机变量                                   |
| $Y_{nmt}$             | 用户 $U_n$ 在时期 $P_t$ 转发信息 $I_m$ 的数量, 它是一个随机变量                                   |
| $NS_{nmt}$            | 用户 $U_n$ 在时期 $P_t$ 对于信息 $I_m$ 的易感性  |
| $N$                   | 用户的数量   |
| $M$                   | 信息种类的数量   |
| $T$                   | 时期的数量   |
| $L_t$                 | 时期 $P_t$ 中节点的数量   |

### 3.2.1 NS 的定义

正如前面分析的,用户的信息转发行为表明其易感性于该信息,并且转发的频率越高,易感程度越大.为量化这种随机转发行为,我们引入一种新颖且通用的量化指标——节点易感性 NS,NS 的定义及计算主要基于第 2 节中介绍的 DSSI 模型及 UGF 方法.

在 DSSI 模型中,主要考虑两个随机变量:施加于一个部件(或系统)上的压力及该部件(或系统)应对于压力而不失效的应力.在这里,用户  $U_n$  等价于 DSSI 模型中一个系统,在时期  $P_t$ ,用户  $U_n$  接收的信息  $I_m$  的数量  $X_{nmt}$  等价于压力,用户  $U_n$  转发的信息  $I_m$  的数量  $Y_{nmt}$  等价于应力,下面给出 NS 的具体定义.

**定义 6.** 用户易感性 NS.

假设在时期  $P_t$ , $U_n$  接收的信息  $I_m$  的数量  $X_{nmt}$  和转发的信息  $I_m$  的数量  $Y_{nmt}$  是两个随机变量,则  $U_n$  在时期  $P_t$  关于信息  $I_m$  的易感性  $NS_{nmt}$  定义为  $Y_{nmt}$  大于  $X_{nmt}$  的概率,即

$$NS_{nmt} = \Pr(Y_{nmt} > X_{nmt}) \quad (10)$$

关于定义 6,有以下几点需要说明:

(1)  $X_{nmt}$  的统计特征是基于一组观测值,即观测参数  $A_{nmtj}$  的.同样, $Y_{nmt}$  的统计特征是基于另一组观测值,即观测参数  $B_{nmtj}$  的.这就意味着  $X_{nmt}$  的概率分布和  $Y_{nmt}$  的概率分布可分别由其观测参数得到.

(2) 由于  $A_{nmtj}$  和  $B_{nmtj}$  与时间区间  $[Q_{t,j-1}, Q_{tj}]$  及时间节点  $Q_{tj}$  有关,因此时间区间的长短会对 NS 值大小有一定影响.在一个固定时期内,当时间节点个数增多(即时间间隔缩小)时,观测参数  $A_{nmtj}$  和  $B_{nmtj}$  的集合规模也会增大(因为观测次数增多),因此,  $X_{nmt}$  和  $Y_{nmt}$  的概率分布就会更准确,进而使得

NS 值的计算更精确.

(3) 由于观测数据是客观的,所以  $NS_{nmt}$  不依赖于决策者的主观判断,故基于  $NS_{nmt}$  的决策结果也是客观的.另一方面,观测数据越多,其决策结果也越准确.

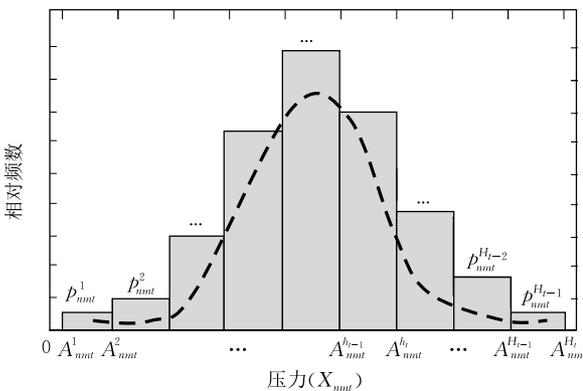
(4) 用户转发行为的随机性导致了信息传播的不确定性,而转发行为与信息接收数量  $X_{nmt}$  及信息转发数量  $Y_{nmt}$  有关.式(10)表明易感性 NS 是一个与  $X_{nmt}$  和  $Y_{nmt}$  相关的概率值,其值的大小反映出在相应时期,用户对于相应信息的易感程度.用户在不同时期易感程度的变化反映出其在转发行为上的差异,这也正是信息传播不确定性的体现.

### 3.2.2 NS 的计算方法

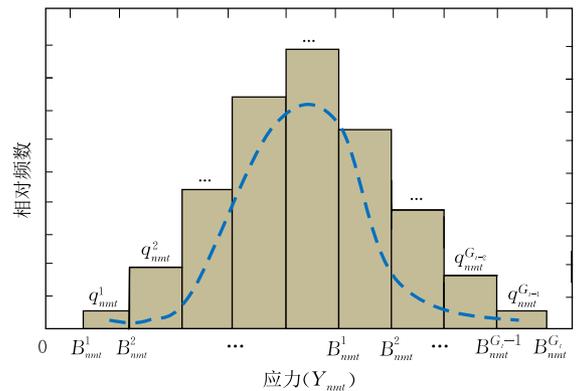
根据前面介绍的 UGF 方法和 DSSI 模型,下面给出 NS 的计算步骤.

(1) 推导  $X_{nmt}$  的概率分布和  $Y_{nmt}$  的概率分布.

由于  $X_{nmt}$  和  $Y_{nmt}$  是两个离散的随机变量,我们需要从它们的观测参数来获取其概率分布.假设  $X_{nmt}$  的观测参数为  $A_{nmt1}, A_{nmt2}, \dots, A_{nmtL_t}$ ,  $Y_{nmt}$  的观测参数为  $B_{nmt1}, B_{nmt2}, \dots, B_{nmtL_t}$ .将这两组参数用直方图表示,如图 3 所示.由此可获得观测值的组距区间以及相应的相对频数.例如在图 3(a)中,横坐标  $A_{nmt}^{h_t-1}$  表示组距区间的端点值,纵坐标  $p_{nmt}^{h_t-1}$  表示组距区间  $[A_{nmt}^{h_t-1}, A_{nmt}^{h_t}]$ ,  $h_t = 2, 3, \dots, H_t$  的相对频数,其中  $H_t$  表示组距区间端点的个数,因此  $X_{nmt}$  的观测参数组距区间的个数为  $H_t - 1$ .类似地,在图 3(b)中,  $B_{nmt}^{g_t-1}$  和  $q_{nmt}^{g_t-1}$ ,  $g_t = 2, 3, \dots, G_t$  分别表示组距区间  $[B_{nmt}^{g_t-1}, B_{nmt}^{g_t}]$  的端点值和相对频数,组距区间端点的个数为  $G_t$ ,因此  $Y_{nmt}$  的观测参数组距区间的个数为  $G_t - 1$ .



(a) 压力观测值的直方图



(b) 应力观测值的直方图

图 3 压力及应力观测值的直方图

$X_{nmt}$  的概率分布由两个要素确定, 一是  $X_{nmt}$  的可能取值, 二是每个可能取值的相应概率. 为此, 我们将每个组距区间  $[A_{nmt}^{h_t-1}, A_{nmt}^{h_t}]$ ,  $h_t = 2, 3, \dots, H_t$  的中点值作为  $X_{nmt}$  的可能取值, 每个组距  $[A_{nmt}^{h_t-1}, A_{nmt}^{h_t}]$ ,  $h_t = 2, 3, \dots, H_t$  的相对频数作为相应概率值. 因此, 可以获得如下  $X_{nmt}$  的概率分布:

$$X_{nmt} = \left( \frac{A_{nmt}^2 - A_{nmt}^1}{2}, \frac{A_{nmt}^3 - A_{nmt}^2}{2}, \dots, \frac{A_{nmt}^{H_t} - A_{nmt}^{H_t-1}}{2} \right) \\ \triangleq (X_{nmt}^1, X_{nmt}^2, \dots, X_{nmt}^{H_t-1}) \quad (11)$$

$$p_{nmt} = (p_{nmt}^1, p_{nmt}^2, \dots, p_{nmt}^{H_t-1}) \quad (12)$$

其中  $X_{nmt}^{h_t} = \frac{A_{nmt}^{h_t} - A_{nmt}^{h_t-1}}{2}$ ,  $h_t = 1, 2, \dots, H_t$ .

类似地, 将每个组距  $[B_{nmt}^{g_t-1}, B_{nmt}^{g_t}]$ ,  $g_t = 2, 3, \dots, G_t$  的中点值作为  $Y_{nmt}$  的可能取值, 每个组距  $[B_{nmt}^{g_t-1}, B_{nmt}^{g_t}]$ ,  $g_t = 2, 3, \dots, G_t$  的相对频数作为相应概率值. 因此, 可以获得如下  $Y_{nmt}$  的概率分布:

$$Y_{nmt} = \left( \frac{B_{nmt}^2 - B_{nmt}^1}{2}, \frac{B_{nmt}^3 - B_{nmt}^2}{2}, \dots, \frac{B_{nmt}^{G_t} - B_{nmt}^{G_t-1}}{2} \right) \\ \triangleq (Y_{nmt}^1, Y_{nmt}^2, \dots, Y_{nmt}^{G_t-1}) \quad (13)$$

$$q_{nmt} = (q_{nmt}^1, q_{nmt}^2, \dots, q_{nmt}^{G_t-1}) \quad (14)$$

其中  $Y_{nmt}^{g_t} = \frac{B_{nmt}^{g_t} - B_{nmt}^{g_t-1}}{2}$ ,  $g_t = 1, 2, \dots, G_t$ .

(2) 推导  $X_{nmt}$  的 UGF,  $Y_{nmt}$  的 UGF 和  $f(X_{nmt}, Y_{nmt})$  的 UGF.

根据定义 1, 离散型随机变量的 UGF 与它的概率分布一一对应, 而在步骤 1 中, 我们已经分别推导出  $X_{nmt}$  和  $Y_{nmt}$  的概率分布, 因此, 由定义 1,  $X_{nmt}$  的 UGF 和  $Y_{nmt}$  的 UGF 分别如下:

$$u_{X_{nmt}}(z) = p_{nmt}^1 z^{X_{nmt}^1} + p_{nmt}^2 z^{X_{nmt}^2} + \dots + p_{nmt}^{H_t-1} z^{X_{nmt}^{H_t-1}} \\ = \sum_{h_t=1}^{H_t-1} p_{nmt}^{h_t} z^{X_{nmt}^{h_t}} \quad (15)$$

$$u_{Y_{nmt}}(z) = q_{nmt}^1 z^{Y_{nmt}^1} + q_{nmt}^2 z^{Y_{nmt}^2} + \dots + q_{nmt}^{G_t-1} z^{Y_{nmt}^{G_t-1}} \\ = \sum_{g_t=1}^{G_t-1} q_{nmt}^{g_t} z^{Y_{nmt}^{g_t}} \quad (16)$$

由于  $f(X_{nmt}, Y_{nmt})$  是  $X_{nmt}$  和  $Y_{nmt}$  的一个函数, 根据定义 2 和定义 3, 可获得  $f(X_{nmt}, Y_{nmt})$  的 UGF 如下:

$$u_f(z) = \otimes (u_{X_{nmt}}(z), u_{Y_{nmt}}(z)) \\ = \otimes \left( \sum_{h_t=1}^{H_t-1} p_{nmt}^{h_t} z^{X_{nmt}^{h_t}}, \sum_{g_t=1}^{G_t-1} q_{nmt}^{g_t} z^{Y_{nmt}^{g_t}} \right) \\ = \sum_{h_t=1}^{H_t-1} \sum_{g_t=1}^{G_t-1} (p_{nmt}^{h_t} z^{f(X_{nmt}^{h_t}, Y_{nmt}^{g_t})} q_{nmt}^{g_t} z^{f(X_{nmt}^{h_t}, Y_{nmt}^{g_t})}) \\ = \sum_{r=1}^R \lambda_r z^{f_r} \quad (17)$$

其中  $f_r$  和  $\lambda_r$  ( $r = 1, 2, \dots, R$ ) 分别是函数  $f(X_{nmt}, Y_{nmt})$  的可能取值和相应概率, 且有  $R \leq H_t \times G_t$ .

(3) 根据 DSSI 模型计算  $NS_{nmt}$ .

假设  $f(X_{nmt}, Y_{nmt}) = Y_{nmt} - X_{nmt}$ , 根据定义 5,  $NS_{nmt}$  可根据式(18)计算

$$NS_{nmt} = \Pr(f(X_{nmt}, Y_{nmt}) > 0) = \sum_{r=1}^R \lambda_r \sigma(f_r) \quad (18)$$

其中  $\sigma(f_r)$  是一个定义在函数  $Y_{nmt} - X_{nmt}$  可能取值集合上的二元值函数, 即  $\sigma(f_r) = \begin{cases} 1, & f_r > 0 \\ 0, & f_r \leq 0 \end{cases}$ . 可以看出, 式(18)将 NS 定义中的概率转化为一个与  $\lambda_r$  和  $\sigma(f_r)$  相乘相加的函数表达式, 这样会更有利于计算机处理计算和决策过程.

### 3.3 动态提取信息易感用户

不失一般性, 我们给出关于信息  $I_m$  的易感用户的提取过程. 根据 3.2 节中 NS 的计算步骤, 可计算得到不同时期每个用户关于信息  $I_m$  的易感性, 如下所示:

$$NS_{1m1}, NS_{2m1}, \dots, NS_{Nm1}, \\ NS_{1m2}, NS_{2m2}, \dots, NS_{Nm2}, \\ \vdots \\ NS_{1mT}, NS_{2mT}, \dots, NS_{NmT},$$

而且

$$NS_{n_1m1} = \max_{1 \leq n_1 \leq N} \{NS_{n_1m1}\}, 1 \leq n_1 \leq N \quad (19a1)$$

$$NS_{n_2m2} = \max_{1 \leq n_2 \leq N} \{NS_{n_2m2}\}, 1 \leq n_2 \leq N \quad (19a2)$$

$\vdots$

$$NS_{n_TmT} = \max_{1 \leq n_T \leq N} \{NS_{n_TmT}\}, 1 \leq n_T \leq N \quad (19aN)$$

从式(19)可以提取出在不同时期关于信息  $I_m$  的易感用户. 例如, 在时期  $P_1$ , 用户  $U_{n_1}$  是信息  $I_m$  的易感用户; 在时期  $P_T$ , 用户  $U_{n_T}$  是信息  $I_m$  的易感用户. 此外, 这一决策结果也可对用户后续(时期  $P_T$  之后)的转发行为进行简单预测. 例如, 假设在时期  $P_T$ , 用户  $U_1$  和  $U_2$  分别是信息  $I_{m_1}$  和  $I_{m_2}$  的易感用户, 其中  $m_1, m_2 \in \{1, 2, \dots, M\}$ , 那么, 在下一个时期,  $U_1$  转发信息  $I_{m_1}$  的概率会高于其他用户, 类似地,  $U_2$  转发信息  $I_{m_2}$  的概率也会高于其他用户. 由于本文的决策目标是信息易感用户的提取, 关于用户随机转发行为预测的研究我们将在后续工作中继续展开.

从上述的提取过程可以看出, 影响决策结果的不是每个用户 NS 值的大小, 而是所有用户 NS 值的排序, 即 NS 值排序靠前的就是相关信息的易感用户. 因此, 在不同规模的观测参数集合下, 虽然同

一用户的 NS 值大小会有所不同(因为精确度不同),但所有用户的 NS 值排序差异不大,因为一旦确定了观测参数集合,所有用户 NS 值的计算精度都是一样的。

另一方面,在现实的决策过程中,如果需要根据实际情况提取更多的信息易感用户,决策者(例如网络商家)只需要将  $NS_{nmt}, t=1, 2, \dots, T; n=1, 2, \dots, N$  进行排序,然后设定阈值提取排序靠前的  $NS_{nmt}, t=1, 2, \dots, T; n=1, 2, \dots, N$ ,便可提取多个信息易感用户。

## 4 案例分析

在这一节中,我们将利用一个案例来分析和说明所提模型的可行性及实用性。

### 4.1 问题描述

不失一般性,我们选取现实生活中一种较为常见的在线社交网络——在线团购网站作为案例的研究对象。假设  $\Omega$  是某一在线团购网站,它可以汇集本地消费产品和零售服务,从商家那里获取代金券,并以折扣的方式销售给其会员用户。在运营的过程中, $\Omega$  会向用户(会员社区)推送不同种类(如美食,休闲娱乐,购物)的商业信息,这些信息通过会员间的转发实现网络推广,为商家找到最合适的消费者、为消费者推送最需要的服务是  $\Omega$  的目标。

在案例中,假设  $\Omega$  的会员用户人数为 120(即  $N=120$ ), $\Omega$  向用户推送 3 种类型(即  $M=3$ )的商业信息,分别是美食  $I_1$ ,休闲娱乐  $I_2$  和购物  $I_3$ 。分析四个月的信息转发数据,每个月对应于一个时期(即  $T=4$ ),假设每个月为 30 天,每一天代表一个时间节点(即  $L_t=30, t=1, 2, 3, 4$ )。在每个时间节点上,我们随机生成了用户在上一时间区间内接收和转发信息数量的观测参数(即  $A_{nmtj}$  和  $B_{nmtj}$ ),分别列于附件中的观测参数(一)和(二)。本案例的决策目标是,根据用户易感性的量化指标—NS 值,运用信息易感用户动态提取模型,在不同时期,对于不同种类的服务信息,提取 20 名该类信息的易感用户,为相关决策者制定有效的网络推广提供理论依据。

需要说明的是,利用随机生成的数据演示模型的应用并不影响模型有效性的验证。这是因为所提模型从建立到求解,其过程中并不涉及具体的变量数值,换句话说,所提模型可应用于观测参数  $A_{nmtj}$

和  $B_{nmtj}$  取任何数值的情形。另一方面,在实际的社交网络运营中,决策者(有权限的商家或网络管理人员)可以从用户日志(或用户数据)中获取这两类观测参数的真实数据。

### 4.2 结果与分析

利用统计软件 SPSS(具体计算过程可参见附录 A)计算用户信息接收量和转发量的概率分布,基于概率分布及式(18),可得到用户在不同时期对于相应信息的易感性。因篇幅所限,这里我们仅列出第一个月( $t=1$ )用户对于三类信息(美食  $I_1$ ,休闲娱乐  $I_2$  和购物  $I_3$ )的易感性,如表 2 所示。对于每类信息,将用户的 NS 值按大小排序,再将 NS 值位于前 20 的用户提取出来,他们即为第一个月易感于相应信息的用户。为清晰起见,我们将提取的结果列于表 3。

由表 3 可以看出,一般地,不同的用户易感于不同类型的信息,或者说感兴趣于不同类型的信息。这一点可由表中三类信息易感用户的重复率看出,这也表明每个用户都有自己最感兴趣的那一类网络服务。当然,也有个别用户同时易感于两类信息。例如,用户  $U_{43}$  同时易感于美食信息  $I_1$  和娱乐信息  $I_2$ ,而用户  $U_{118}$  则同时易感于休闲娱乐信息  $I_2$  和购物信息  $I_3$ 。另外,我们也发现有一个用户  $U_{82}$  易感于所有服务信息,这种现象可以解释为该用户对社交网络  $\Omega$  递送的每种服务信息都很关注,并乐于向他人传达这些信息,而且该用户的会员级别也比较高,可能就是社交网络中的“高级会员”或“金牌会员”。

类似地,我们可以提取第二个月至第四个月各类信息的易感用户,列于表 4。由表 4 可以看出,一般而言,同一用户在不同时期会易感于不同类型的服务信息。例如,用户  $U_8$  在第一个月易感于美食信息  $I_1$ ,而在第二个月则易感于休闲娱乐信息  $I_2$ 。这种现象可以解释为用户的信息转发行为不仅被自己的兴趣影响,也会受其他因素的影响,例如商家的信息推广策略,当商家加大了宣传力度,或增加了奖励(如消费优惠),则会吸引更多用户的转发。

另外,基于用户的 NS 数据,还可以进一步分析哪种类型的信息在社交网络中对于用户的吸引力更大,或者说在用户间传播得更快。为此,我们计算了每类信息易感用户的 NS 均值,并用图 4 进行了直观的比较。可以看出,在不同时期,购物信息  $I_3$  的 NS 均值都高于另外两类信息,这也暗示购物类信息对于网络用户来说更具吸引力、更受欢迎。但是,

表 2 第一个月用户对于三类信息的易感性

| NS    | $U_1$     | $U_2$     | $U_3$     | $U_4$     | $U_5$     | $U_6$     | $U_7$     | $U_8$     | $U_9$     | $U_{10}$  | $U_{11}$  | $U_{12}$  | $U_{13}$  | $U_{14}$  | $U_{15}$  |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $I_1$ | 0.6122    | 0.6322    | 0.5822    | 0.6300    | 0.6056    | 0.6378    | 0.5493    | 0.6437    | 0.5349    | 0.6527    | 0.5441    | 0.6480    | 0.5539    | 0.5882    | 0.5972    |
| $I_2$ | 0.6433    | 0.6700    | 0.6156    | 0.6211    | 0.6356    | 0.6200    | 0.6182    | 0.6210    | 0.5902    | 0.5694    | 0.6012    | 0.6321    | 0.5742    | 0.5318    | 0.5849    |
| $I_3$ | 0.6756    | 0.5956    | 0.6222    | 0.6400    | 0.5844    | 0.6022    | 0.6032    | 0.5816    | 0.5625    | 0.6137    | 0.5752    | 0.5945    | 0.6068    | 0.6119    | 0.5367    |
| NS    | $U_{16}$  | $U_{17}$  | $U_{18}$  | $U_{19}$  | $U_{20}$  | $U_{21}$  | $U_{22}$  | $U_{23}$  | $U_{24}$  | $U_{25}$  | $U_{26}$  | $U_{27}$  | $U_{28}$  | $U_{29}$  | $U_{30}$  |
| $I_1$ | 0.5790    | 0.6525    | 0.5528    | 0.5297    | 0.5809    | 0.6641    | 0.5703    | 0.6415    | 0.5844    | 0.6330    | 0.5944    | 0.6091    | 0.6377    | 0.5474    | 0.6583    |
| $I_2$ | 0.6325    | 0.6014    | 0.5934    | 0.5453    | 0.6470    | 0.5939    | 0.6149    | 0.6012    | 0.5205    | 0.5993    | 0.6316    | 0.6350    | 0.5936    | 0.5605    | 0.6520    |
| $I_3$ | 0.6424    | 0.5737    | 0.6160    | 0.5834    | 0.6332    | 0.6240    | 0.6229    | 0.5890    | 0.5788    | 0.6213    | 0.5840    | 0.6289    | 0.5584    | 0.5967    | 0.6341    |
| NS    | $U_{31}$  | $U_{32}$  | $U_{33}$  | $U_{34}$  | $U_{35}$  | $U_{36}$  | $U_{37}$  | $U_{38}$  | $U_{39}$  | $U_{40}$  | $U_{41}$  | $U_{42}$  | $U_{43}$  | $U_{44}$  | $U_{45}$  |
| $I_1$ | 0.6114    | 0.6037    | 0.6380    | 0.6235    | 0.5993    | 0.6408    | 0.6253    | 0.5837    | 0.6356    | 0.5835    | 0.6064    | 0.5493    | 0.6687    | 0.5907    | 0.6120    |
| $I_2$ | 0.5842    | 0.6219    | 0.6034    | 0.6472    | 0.6187    | 0.5936    | 0.6008    | 0.6128    | 0.6092    | 0.6437    | 0.6312    | 0.5880    | 0.6432    | 0.5858    | 0.6351    |
| $I_3$ | 0.5731    | 0.6794    | 0.5983    | 0.6116    | 0.6345    | 0.6211    | 0.5739    | 0.6351    | 0.5930    | 0.6622    | 0.5936    | 0.6135    | 0.6059    | 0.6261    | 0.5736    |
| NS    | $U_{46}$  | $U_{47}$  | $U_{48}$  | $U_{49}$  | $U_{50}$  | $U_{51}$  | $U_{52}$  | $U_{53}$  | $U_{54}$  | $U_{55}$  | $U_{56}$  | $U_{57}$  | $U_{58}$  | $U_{59}$  | $U_{60}$  |
| $I_1$ | 0.5993    | 0.5900    | 0.6113    | 0.6273    | 0.5786    | 0.5772    | 0.6372    | 0.5833    | 0.6077    | 0.6112    | 0.6278    | 0.5995    | 0.6310    | 0.5975    | 0.5843    |
| $I_2$ | 0.5732    | 0.6303    | 0.6189    | 0.5981    | 0.6239    | 0.5828    | 0.6453    | 0.6154    | 0.5836    | 0.6254    | 0.6039    | 0.6213    | 0.5829    | 0.6146    | 0.5635    |
| $I_3$ | 0.6017    | 0.6465    | 0.6032    | 0.5735    | 0.6146    | 0.5542    | 0.6255    | 0.6379    | 0.5974    | 0.5843    | 0.5735    | 0.5777    | 0.6111    | 0.5579    | 0.5994    |
| NS    | $U_{61}$  | $U_{62}$  | $U_{63}$  | $U_{64}$  | $U_{65}$  | $U_{66}$  | $U_{67}$  | $U_{68}$  | $U_{69}$  | $U_{70}$  | $U_{71}$  | $U_{72}$  | $U_{73}$  | $U_{74}$  | $U_{75}$  |
| $I_1$ | 0.6592    | 0.5998    | 0.6258    | 0.6227    | 0.6443    | 0.6185    | 0.5739    | 0.5536    | 0.6318    | 0.6433    | 0.5644    | 0.5873    | 0.6439    | 0.5970    | 0.6152    |
| $I_2$ | 0.6301    | 0.6246    | 0.5937    | 0.6019    | 0.5972    | 0.6417    | 0.6521    | 0.5905    | 0.6082    | 0.5972    | 0.5973    | 0.5911    | 0.6008    | 0.6222    | 0.6336    |
| $I_3$ | 0.6348    | 0.6579    | 0.5621    | 0.6483    | 0.6204    | 0.6023    | 0.6118    | 0.5892    | 0.5834    | 0.6139    | 0.5871    | 0.6353    | 0.6237    | 0.6073    | 0.5739    |
| NS    | $U_{76}$  | $U_{77}$  | $U_{78}$  | $U_{79}$  | $U_{80}$  | $U_{81}$  | $U_{82}$  | $U_{83}$  | $U_{84}$  | $U_{85}$  | $U_{86}$  | $U_{87}$  | $U_{88}$  | $U_{89}$  | $U_{90}$  |
| $I_1$ | 0.5835    | 0.6378    | 0.5738    | 0.5983    | 0.6159    | 0.5997    | 0.6481    | 0.6046    | 0.5458    | 0.6441    | 0.5991    | 0.5684    | 0.6314    | 0.6458    | 0.5542    |
| $I_2$ | 0.6149    | 0.5748    | 0.5905    | 0.6091    | 0.6473    | 0.6018    | 0.6352    | 0.6185    | 0.5802    | 0.6310    | 0.6134    | 0.5427    | 0.5982    | 0.5910    | 0.5867    |
| $I_3$ | 0.5777    | 0.6041    | 0.6382    | 0.6257    | 0.5877    | 0.6393    | 0.6530    | 0.6427    | 0.5773    | 0.6583    | 0.5882    | 0.6204    | 0.6553    | 0.6223    | 0.5693    |
| NS    | $U_{91}$  | $U_{92}$  | $U_{93}$  | $U_{94}$  | $U_{95}$  | $U_{96}$  | $U_{97}$  | $U_{98}$  | $U_{99}$  | $U_{100}$ | $U_{101}$ | $U_{102}$ | $U_{103}$ | $U_{104}$ | $U_{105}$ |
| $I_1$ | 0.5458    | 0.5875    | 0.6540    | 0.5939    | 0.6306    | 0.6114    | 0.5579    | 0.6037    | 0.6264    | 0.5982    | 0.6500    | 0.6312    | 0.5403    | 0.5898    | 0.6172    |
| $I_2$ | 0.5639    | 0.6347    | 0.6332    | 0.6058    | 0.6182    | 0.6312    | 0.5832    | 0.5920    | 0.5817    | 0.6330    | 0.5748    | 0.6055    | 0.5726    | 0.6426    | 0.5908    |
| $I_3$ | 0.5995    | 0.6139    | 0.5978    | 0.6237    | 0.5841    | 0.6550    | 0.6071    | 0.5548    | 0.6153    | 0.6235    | 0.6137    | 0.6642    | 0.5931    | 0.6089    | 0.6335    |
| NS    | $U_{106}$ | $U_{107}$ | $U_{108}$ | $U_{109}$ | $U_{110}$ | $U_{111}$ | $U_{112}$ | $U_{113}$ | $U_{114}$ | $U_{115}$ | $U_{116}$ | $U_{117}$ | $U_{118}$ | $U_{119}$ | $U_{120}$ |
| $I_1$ | 0.6430    | 0.6274    | 0.5946    | 0.6450    | 0.5985    | 0.6421    | 0.5428    | 0.5794    | 0.5946    | 0.6092    | 0.5649    | 0.5791    | 0.6392    | 0.5624    | 0.6275    |
| $I_2$ | 0.6192    | 0.6083    | 0.5693    | 0.5831    | 0.5673    | 0.5841    | 0.5669    | 0.6278    | 0.5527    | 0.6377    | 0.5935    | 0.5925    | 0.6467    | 0.5743    | 0.6048    |
| $I_3$ | 0.5990    | 0.5725    | 0.6214    | 0.6692    | 0.5852    | 0.6003    | 0.5737    | 0.6159    | 0.6961    | 0.6429    | 0.6223    | 0.5440    | 0.6418    | 0.5995    | 0.5836    |

表 3 第一个月各类信息的易感用户

| 信息 $I_1$ 的易感用户 |           |           |           |           | 信息 $I_2$ 的易感用户 |          |           |           |           | 信息 $I_3$ 的易感用户 |           |           |           |           |
|----------------|-----------|-----------|-----------|-----------|----------------|----------|-----------|-----------|-----------|----------------|-----------|-----------|-----------|-----------|
| $U_8$          | $U_{10}$  | $U_{12}$  | $U_{17}$  | $U_{21}$  | $U_1$          | $U_2$    | $U_5$     | $U_{20}$  | $U_{27}$  | $U_3$          | $U_4$     | $U_{16}$  | $U_{32}$  | $U_{40}$  |
| 0.6437         | 0.6527    | 0.6480    | 0.6525    | 0.6641    | 0.6433         | 0.6700   | 0.6356    | 0.6470    | 0.6350    | 0.6756         | 0.6400    | 0.6424    | 0.6794    | 0.6622    |
| $U_{23}$       | $U_{30}$  | $U_{43}$  | $U_{61}$  | $U_{65}$  | $U_{30}$       | $U_{34}$ | $U_{40}$  | $U_{43}$  | $U_{45}$  | $U_{47}$       | $U_{53}$  | $U_{62}$  | $U_{64}$  | $U_{78}$  |
| 0.6415         | 0.6583    | 0.6687    | 0.6592    | 0.6443    | 0.6520         | 0.6472   | 0.6437    | 0.6432    | 0.6351    | 0.6465         | 0.6379    | 0.6579    | 0.6483    | 0.6382    |
| $U_{70}$       | $U_{73}$  | $U_{82}$  | $U_{85}$  | $U_{89}$  | $U_{52}$       | $U_{66}$ | $U_{67}$  | $U_{75}$  | $U_{80}$  | $U_{81}$       | $U_{82}$  | $U_{83}$  | $U_{85}$  | $U_{88}$  |
| 0.6433         | 0.6439    | 0.6481    | 0.6441    | 0.6458    | 0.6453         | 0.6417   | 0.6521    | 0.6336    | 0.6473    | 0.6393         | 0.6530    | 0.6427    | 0.6583    | 0.6553    |
| $U_{93}$       | $U_{101}$ | $U_{106}$ | $U_{109}$ | $U_{111}$ | $U_{82}$       | $U_{92}$ | $U_{104}$ | $U_{115}$ | $U_{118}$ | $U_{96}$       | $U_{102}$ | $U_{109}$ | $U_{115}$ | $U_{118}$ |
| 0.6540         | 0.6500    | 0.6430    | 0.6450    | 0.6421    | 0.6352         | 0.6347   | 0.6426    | 0.6377    | 0.6467    | 0.6550         | 0.6642    | 0.6692    | 0.6429    | 0.6418    |

表 4 第二个月至第四个月各类信息的易感用户

|      | 信息 $I_1$ 的易感用户 |           |           |           |           | 信息 $I_2$ 的易感用户 |           |           |           |           | 信息 $I_3$ 的易感用户 |           |           |           |           |
|------|----------------|-----------|-----------|-----------|-----------|----------------|-----------|-----------|-----------|-----------|----------------|-----------|-----------|-----------|-----------|
|      | $U_7$          | $U_9$     | $U_{14}$  | $U_{19}$  | $U_{25}$  | $U_4$          | $U_8$     | $U_{15}$  | $U_{28}$  | $U_{32}$  | $U_3$          | $U_5$     | $U_{16}$  | $U_{27}$  | $U_{38}$  |
| 第二个月 | $U_{29}$       | $U_{33}$  | $U_{46}$  | $U_{55}$  | $U_{62}$  | $U_{34}$       | $U_{39}$  | $U_{44}$  | $U_{48}$  | $U_{51}$  | $U_{45}$       | $U_{49}$  | $U_{56}$  | $U_{58}$  | $U_{64}$  |
|      | $U_{71}$       | $U_{76}$  | $U_{84}$  | $U_{88}$  | $U_{93}$  | $U_{59}$       | $U_{68}$  | $U_{77}$  | $U_{82}$  | $U_{90}$  | $U_{69}$       | $U_{70}$  | $U_{79}$  | $U_{85}$  | $U_{95}$  |
|      | $U_{94}$       | $U_{103}$ | $U_{107}$ | $U_{114}$ | $U_{119}$ | $U_{99}$       | $U_{109}$ | $U_{110}$ | $U_{116}$ | $U_{117}$ | $U_{101}$      | $U_{106}$ | $U_{111}$ | $U_{118}$ | $U_{120}$ |
|      | $U_6$          | $U_8$     | $U_{18}$  | $U_{29}$  | $U_{32}$  | $U_{11}$       | $U_{17}$  | $U_{24}$  | $U_{25}$  | $U_{33}$  | $U_9$          | $U_{12}$  | $U_{22}$  | $U_{26}$  | $U_{33}$  |
| 第三个月 | $U_{40}$       | $U_{43}$  | $U_{51}$  | $U_{54}$  | $U_{63}$  | $U_{39}$       | $U_{54}$  | $U_{55}$  | $U_{61}$  | $U_{67}$  | $U_{39}$       | $U_{40}$  | $U_{51}$  | $U_{60}$  | $U_{61}$  |
|      | $U_{72}$       | $U_{76}$  | $U_{81}$  | $U_{87}$  | $U_{94}$  | $U_{75}$       | $U_{84}$  | $U_{89}$  | $U_{92}$  | $U_{96}$  | $U_{64}$       | $U_{73}$  | $U_{82}$  | $U_{88}$  | $U_{91}$  |
|      | $U_{99}$       | $U_{104}$ | $U_{115}$ | $U_{116}$ | $U_{118}$ | $U_{102}$      | $U_{108}$ | $U_{109}$ | $U_{115}$ | $U_{117}$ | $U_{103}$      | $U_{105}$ | $U_{114}$ | $U_{116}$ | $U_{119}$ |
| 第四个月 | $U_{10}$       | $U_{14}$  | $U_{23}$  | $U_{35}$  | $U_{41}$  | $U_1$          | $U_3$     | $U_{13}$  | $U_{17}$  | $U_{26}$  | $U_2$          | $U_9$     | $U_{13}$  | $U_{20}$  | $U_{30}$  |
|      | $U_{44}$       | $U_{48}$  | $U_{50}$  | $U_{57}$  | $U_{61}$  | $U_{31}$       | $U_{33}$  | $U_{37}$  | $U_{47}$  | $U_{52}$  | $U_{31}$       | $U_{38}$  | $U_{43}$  | $U_{49}$  | $U_{54}$  |
|      | $U_{64}$       | $U_{73}$  | $U_{79}$  | $U_{87}$  | $U_{92}$  | $U_{62}$       | $U_{66}$  | $U_{70}$  | $U_{74}$  | $U_{83}$  | $U_{58}$       | $U_{67}$  | $U_{75}$  | $U_{79}$  | $U_{86}$  |
|      | $U_{100}$      | $U_{104}$ | $U_{111}$ | $U_{112}$ | $U_{120}$ | $U_{88}$       | $U_{94}$  | $U_{101}$ | $U_{107}$ | $U_{115}$ | $U_{93}$       | $U_{99}$  | $U_{100}$ | $U_{107}$ | $U_{118}$ |

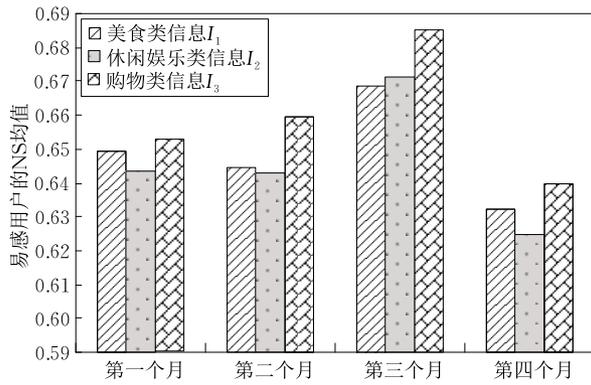


图 4 每类信息易感用户的 NS 均值

美食信息  $I_1$  和休闲娱乐信息  $I_2$  在不同时期的 NS 值则体现出一定的波动性, 这反映出用户对于这两类信息易感性的不确定性和动态性. 此外, 我们明显地看到在第三个月, 三类信息的 NS 值普遍较高, 说明在这一时期所有信息的转发频率都很高, 这种现象可以解释为社交网络  $\Omega$  在这期间进行了全网促销活动, 各类促销信息在会员用户间频繁转发, 例如现实生活中的“周年庆”或“双十一”.

#### 4.3 模型对比

为进一步说明本文模型在动态提取易感用户方面的作用与优势, 下面我们将基于同一输入参数集

合, 利用文献[28]的 Influence-Susceptibility-Cynical (ISC) 模型再次计算用户易感性, 并提取相关信息的易感用户.

在 ISC 模型中, 用户易感性  $S(v)$  指的是用户对于带有某种特定情感(正面、负面)信息的易感程度, 并将其量化为该用户转发带有某种情感信息的数量与接收带有同类情感信息的数量之比. 因此, 在模型对比中, 我们可将 ISC 模型中的带有某种特定情感的信息与本文模型中的某类特定信息对应起来, 即可基于 ISC 模型将用户对于某类信息的易感性量化为用户转发此类信息的数量与接收此类信息的数量之比.

根据上述量化方式, 基于附件中的输入参数集合, 我们分别利用本文模型及 ISC 模型计算用户对于不同种类信息的易感性. 由于 ISC 模型不涉及提取周期(即易感用户的提取是静态的), 为公平起见, 在对比中, 利用本文模型也仅分析一个月(第一个月)的用户易感性. 在本文模型与 ISC 模型中, 分别根据用户的 NS 值及  $S(v)$  值的排序提取 10 名相关信息的易感用户, 将结果列于表 5, 其中用方框标注的用户为两种不同模型下提取的不同易感用户.

表 5 本文模型与 ISC 模型的易感用户提取

| NS/ $S(v)$ | 信息 $I_1$ 的易感用户 |          |          |           |           | 信息 $I_2$ 的易感用户 |          |          |          |           | 信息 $I_3$ 的易感用户 |          |          |           |           |
|------------|----------------|----------|----------|-----------|-----------|----------------|----------|----------|----------|-----------|----------------|----------|----------|-----------|-----------|
|            | $U_{10}$       | $U_{17}$ | $U_{21}$ | $U_{30}$  | $U_{43}$  | $U_1$          | $U_2$    | $U_{20}$ | $U_{30}$ | $U_{34}$  | $U_1$          | $U_{32}$ | $U_{40}$ | $U_{62}$  | $U_{82}$  |
| 本文模型       | 0.6527         | 0.6525   | 0.6641   | 0.6583    | 0.6687    | 0.6433         | 0.6700   | 0.6470   | 0.6520   | 0.6472    | 0.6756         | 0.6794   | 0.6622   | 0.6579    | 0.6530    |
|            | $U_{61}$       | $U_{82}$ | $U_{89}$ | $U_{93}$  | $U_{101}$ | $U_{40}$       | $U_{52}$ | $U_{67}$ | $U_{80}$ | $U_{118}$ | $U_{85}$       | $U_{88}$ | $U_{96}$ | $U_{102}$ | $U_{109}$ |
|            | 0.6592         | 0.6481   | 0.6458   | 0.6540    | 0.6500    | 0.6437         | 0.6453   | 0.6521   | 0.6473   | 0.6467    | 0.6583         | 0.6553   | 0.6550   | 0.6642    | 0.6692    |
| ISC 模型     | $U_{10}$       | $U_{17}$ | $U_{21}$ | $U_{43}$  | $U_{61}$  | $U_2$          | $U_8$    | $U_{20}$ | $U_{30}$ | $U_{34}$  | $U_{27}$       | $U_{29}$ | $U_{32}$ | $U_{40}$  | $U_{62}$  |
|            | 1.1133         | 1.0801   | 1.1208   | 1.0611    | 1.0239    | 1.0838         | 1.0725   | 1.0661   | 1.0521   | 1.0994    | 1.0792         | 1.0825   | 1.0724   | 1.0695    | 1.0344    |
|            | $U_{82}$       | $U_{84}$ | $U_{93}$ | $U_{101}$ | $U_{120}$ | $U_{40}$       | $U_{52}$ | $U_{67}$ | $U_{86}$ | $U_{118}$ | $U_{82}$       | $U_{83}$ | $U_{96}$ | $U_{102}$ | $U_{109}$ |
|            | 1.0937         | 1.0311   | 1.0448   | 1.0655    | 1.1267    | 1.0902         | 1.0784   | 1.0364   | 1.0876   | 1.0042    | 1.0695         | 1.0712   | 1.0594   | 1.0148    | 1.0783    |

从表 5 可以看出, 在两种不同模型下所提取的 10 名信息易感用户基本一致(一致性达到 70% 以上), 这也从侧面说明了本文模型的有效性. 但是, 我们也看到利用不同模型提取信息易感用户的差异性. 以信息  $I_1$  的易感用户提取为例, 在本文模型中, 用户  $U_{30}$  和  $U_{89}$  的 NS 值均高于  $U_{84}$  和  $U_{120}$ , 而在 ISC 模型中, 用户  $U_{84}$  和  $U_{120}$  的  $S(v)$  值则高于  $U_{30}$  和  $U_{89}$ , 因此导致了提取结果的不同.

从这四个用户转发信息量的数据可以看出, 在某个时间节点上,  $U_{84}$  和  $U_{120}$  均出现了一次极高的转发量(57), 但是只出现了一次;  $U_{30}$  和  $U_{89}$  均没有出现极高转发量的时间节点, 但是他们在每个时间节

点的转发量相对“集中”, 这里的“集中”是指每个时间节点的转发数据偏离均值的程度不大. 虽然在个别时间节点上,  $U_{84}$  和  $U_{120}$  都有过极高的信息转发量, 导致其  $S(v)$  值偏高, 但是, 从统计学的角度讲, 这种偏离均值程度很大的异常值属于奇异点, 一般不将其作为变量统计规律分析的基本数据. 而且, 从现实的角度来看, 用户在某个时间节点上偶尔出现一次对某一信息的大量转发, 并不能直接说明该用户是此信息的易感用户. 由于本文模型中定义的用户 NS 值(定义 6)主要基于信息转发数据的统计特征, 而不是数据本身的数值大小, 因此, 从理论角度和实际角度来看, 以用户的 NS 值来度量其信息易

感程度更为科学、合理.

## 5 结束语

本文针对在线社交网络中的商业信息扩散问题,利用通用发生函数 UGF 方法及离散压力-应力 DSSI 理论,提出了一种新颖的信息易感用户动态提取模型.模型首先将用户的随机转发行为量化为节点易感性 NS,然后基于 UGF 方法及 DSSI 理论计算不同用户不同时期关于不同信息的 NS 值,最后通过 NS 值的动态排序提取信息易感用户.

本文模型的主要创新点有:(1)易感性被量化为一个与用户随机转发行为有关的概率值,在最大程度上模拟了现实社交网络中信息传播的不确定性;(2)模型的决策结果是客观且动态的,一方面不依赖于决策者的主观判断,另一方面也会随着观测数据的更新而更新;(3)易感用户的提取指标是唯一且通用的,一方面简化了决策过程,有助于实践者在现实中实施,另一方面该指标也可推广应用于其他具有高不确定性服务领域的易感用户提取.

本文模型也存在着几点不足.首先,虽然基于用户的 NS 值,模型能提供客观、动态的决策,但是却不能提供 NS 值在不同时期波动的原因,换句话说,决策者可能不知道为什么某些用户对服务信息不感兴趣,为什么有些服务信息不受用户欢迎.其次,用户的随机转发行为有可能会受到网络结构或用户间拓扑关系的影响,而本文模型并没有考虑到这一点.基于以上不足,下一步我们将从以下两个方面进行深入研究:(1)在信息易感用户动态提取模型中引入用户评价机制,并量化用户对相关信息的评价;(2)将网络结构或用户间的拓扑关系也作为影响信息扩散的因素之一考虑在模型之中,此时可能需要用到点集拓扑理论及图论来处理新的模型.

## 参 考 文 献

[1] Wang Yuan-Zhuo, Yu Jian-Ye, Qiu Wen, et al. Evolutionary game model and analysis methods for network group behavior. *Chinese Journal of Computers*, 2015, 38(2): 282-300 (in Chinese)  
(王元卓, 于建业, 邱雯等. 网络群体行为的演化博弈模型与分析方法. *计算机学报*, 2015, 38(2): 282-300)

[2] Bakshy E, Rosenn I, Marlow C, Adamic L. The role of social networks in information diffusion//*Proceedings of the 21st ACM Conference on World Wide Web*. Lyon, France, 2012, 519-528

[3] Traud A L, Mucha P J, Porter M A. Social structure of

facebook networks. *Physica A*, 2012, 391(16): 4165-4180

[4] Li Dong, Xu Zhi-Ming, Li Sheng, et al. A survey on information diffusion in online social networks. *Chinese Journal of Computers*, 2014, 37(1): 189-206(in Chinese)  
(李栋, 徐志明, 李生等. 在线社交网络中信息扩散. *计算机学报*, 2014, 37(1): 189-206)

[5] Wu Xin-Dong, Li Yi, Li Lei. Influence analysis of online social networks. *Chinese Journal of Computers*, 2014, 37(4): 735-752(in Chinese)  
(吴信东, 李毅, 李磊. 在线社交网络影响力分析. *计算机学报*, 2014, 37(4): 735-752)

[6] Liu Yan-Heng, Li Fei-Peng, Sun Xin, Zhu Jian-Qi. Social network model based on the transmission of information. *Journal on Communications*, 2013, 34(4): 1-9(in Chinese)  
(刘衍珩, 李飞鹏, 孙鑫, 朱建启. 基于信息传播的社交网络拓扑模型. *通信学报*, 2013, 34(4): 1-9)

[7] Jiang C, Chen Y, Ray Liu K J. Evolutionary dynamics of information diffusion over social networks. *IEEE Transactions on Signal Processing*, 2014, 62(17): 4573-4586

[8] Yagan O, Qian D, Zhang J, Cochran D. Conjoining speeds up information diffusion in overlaying social-physical networks. *IEEE Journal on Selected Areas in Communications*, 2013, 31(6): 1038-1048

[9] Miritello G, Moro E, Lara R. Dynamical strength of social ties in information spreading. *Physical Review E*, 2011, 83(4): 045102

[10] Jiang C, Chen Y, Ray Liu K J. Graphical evolutionary game for information diffusion over social networks. *IEEE Journal of Selected Topics in Signal Processing*, 2014, 8(4): 524-536

[11] Jiang J, Wen S, Yu S, et al. K-Center: An approach on the multi-source identification of information diffusion. *IEEE Transactions on Information Forensics and Security*, 2015, 10(12): 2626-2626

[12] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Physical Review Letters*, 2001, 86(14): 3200-3203

[13] Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blogspace//*Proceedings of the 13rd International Conference on World Wide Web (WWW)*. New York, USA, 2004: 491-501

[14] Donetti L, Hurtado P I, Munoz M A. Network synchronization: Optimal and pessimal scale-free topologies. *Journal of Physics A-Mathematical and Theoretical*, 2008, 41(22): 224008

[15] Wu Hong-Run, Qin Jun, Yi Yun-Fei, et al. A model of community-structure and scale-free network based on optimization theory. *Chinese Journal of Computers*, 2015, 38(2): 337-348(in Chinese)  
(吴泓润, 覃俊, 易云飞等. 基于优化理论的社区无标度网络模型. *计算机学报*, 2015, 38(2): 337-348)

[16] Lin M, Li N. Scale-free network provides an optimal pattern for knowledge transfer. *Physica A*, 2010, 389(3): 473-480

[17] Li Y, Qian M, Jin D, et al. Revealing the efficiency of information diffusion in online social networks of microblog. *Information Sciences*, 2015, 293: 383-389

[18] Park S J, Lim Y S, Park H W. Comparing Twitter and YouTube networks in information diffusion: The case of the “Occupy Wall Street” movement. *Technological Forecasting and Social Change*, 2015, 95: 208-217

[19] Kimura M, Saito K, Nakano R, Motoda H. Extracting influential nodes on a social network for information diffusion. *Data Mining and Knowledge Discovery*, 2010, 20: 70-97

[20] Wang Y, Cong G, Song G, Xie K. Community-based greedy algorithm for mining top- $k$  influential nodes in mobile social networks//*Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington, USA, 2010: 1039-1048

[21] Cao Jiu-Xin, Dong Dan, Xu Shun, et al. A  $k$ -core based algorithm for influence maximization in social networks. *Chinese Journal of Computers*, 2015, 38(2): 238-248 (in Chinese)  
(曹玖新, 董丹, 徐顺等. 一种基于  $k$ -核的社会网络影响最大化算法. *计算机学报*, 2015, 38(2): 238-248)

[22] Ilyas M U, Shafiq M Z, Liu A X, Radha H. A distributed and privacy preserving algorithm for identifying information hubs in social networks//*Proceedings of the IEEE INFOCOM*. Shanghai, China, 2011: 561-565

[23] Saito K, Kimura M, Ohara K, Motoda H. Super mediator—A new centrality measure of node importance for information diffusion over social network. *Information Sciences*, 2016, 329: 985-1000

[24] Zhang Yan-Chao, Liu Yun, Zhang Hai-Feng, et al. The research of information dissemination model on online social network. *Acta Physica Sinica*, 2011, 60(5): 050501(in Chinese)  
(张彦超, 刘云, 张海峰等. 基于在线社交网络的信息传播模型. *物理学报*, 2011, 60(5): 050501)

[25] Ma J, Li D, Tian Z. Rumor spreading in online social networks by considering the bipolar social reinforcement. *Physica A*, 2016, 447: 108-115

[26] Gu Yi-Ran, Xia Ling-Ling. The propagation and inhibition of rumors in online social network. *Acta Physica Sinica*, 2012, 61(23): 238701(in Chinese)  
(顾亦然, 夏玲玲. 在线社交网络中谣言的传播与抑制. *物理学报*, 2012, 61(23): 238701)

[27] Aral S, Walker D. Identifying influential and susceptible members of social networks. *Science*, 2012, 337: 337-341

[28] Lee R K, Lim E. Measuring user influence, susceptibility, and cynicism in sentiment diffusion//*Proceedings of the 37th European Conference on IR Research (ECIR 2015)*. Vienna, Austria, 2015: 411-422

[29] Hoang T, Lim E. Virality and susceptibility in information diffusions//*Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*. Dublin, Ireland, 2012: 146-153

[30] Chen J, Teng L, Yu Y, Yu X. The effect of online information sources on purchase intentions between consumers with high and low susceptibility to informational influence. *Journal of Business Research*, 2016, 69: 467-475

[31] Ushakov I. A universal generating function. *Soviet Journal of Computer and Systems Sciences*, 1986, 24(5): 118-129

[32] Lisnianski A, Levitin G. *Multi-State System Reliability: Assessment, Optimization and Applications*. Singapore: World Scientific, 2003

[33] Levitin G. *The Universal Generating Function in Reliability Analysis and Optimization*. London, UK: Springer, 2005

[34] Ebeling C. *An Introduction to Reliability and Maintainability Engineering*. Long Grove, USA: Waveland Press, Inc., 2005

附录 A. 利用 SPSS 计算  $X_{nmt}$  的概率分布和  $Y_{nmt}$  的概率分布.

(a) 生成  $X_{nmt}$  观测值的数据文件和  $Y_{nmt}$  观测值的数据文件

在“Variable View”中,将“ $X_{nmt}$ ”设置为变量名(如图附图 a1),然后将  $W(X_{nmt}$  观测值的个数)个值导入“Data View”中(如图附图 a2),即可生成  $W$  行一列的数据文件. 类似地,在

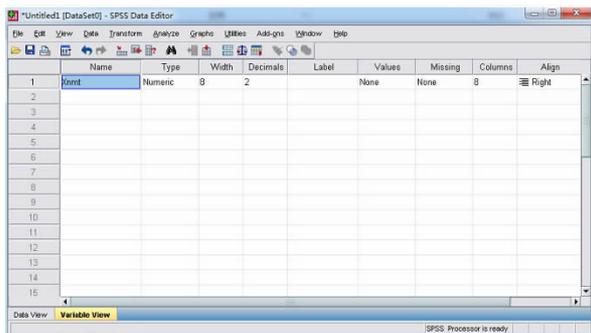


图 a1 “Variable View”窗口



图 a2 “Data View”窗口

“Variable View”中,以“ $Y_{nmt}$ ”为变量名导入  $Z(Y_{nmt}$  观测值的个数)个值到“Data View”中,生成  $Z$  行一列的数据文件.

(b) 获取以频率为纵坐标的直方图

在“Data Editor”对话框中依次选择“Graphs”,“Legacy Dialogs”,“Interactive”和“Histogram”(如图附图 a3),则会出

现“Create Histogram”对话框. 在“Assign Variables”列表中, 选择“ $X_{nmt}$ ”并将其拖入到横坐标的空白框内, 然后选择“Percent”并将其拖入到纵坐标的“Count”内(如附图 a4).

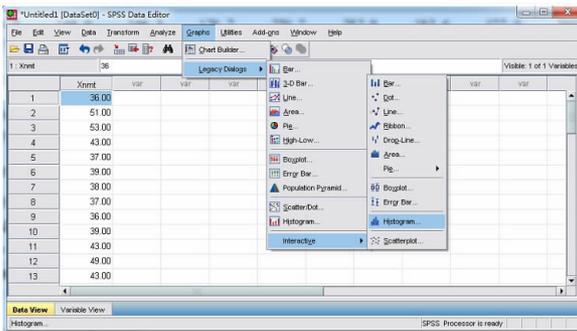


图 a3 “Data Editor”对话框

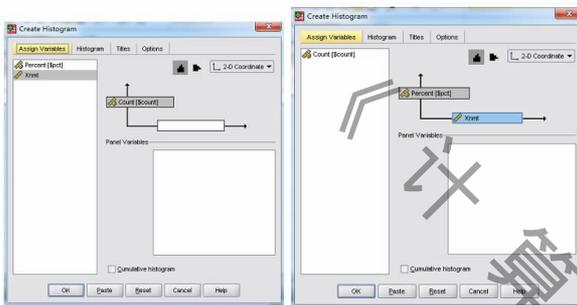


图 a4 “Create Histogram”对话框

在“Histogram”选项中, 勾选“Normal curve”的复选框, 表示正态曲线会出现在输出的直方图上. 另外, 将“Number of interval”设置为 7(决策者可根据需要设置其他数值), 表示直方图的组距区间个数为 7. 点击确定后, 即可得到  $X_{nmt}$  观测参数的直方图(如附图 a5). 类似地, 基于以上步骤也可获得  $Y_{nmt}$  观测参数的直方图. 两组观测数据直方图的一般形式如文中 3.2.2 小节的图 3.

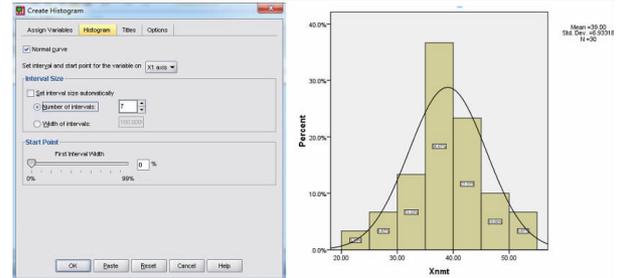


图 a5

(c) 基于(b)中获取的直方图推导  $X_{nmt}$  的概率分布和  $Y_{nmt}$  的概率分布

将两个直方图中每个组距区间的中点分别作为  $X_{nmt}$  和  $Y_{nmt}$  的可能取值, 将组距区间的频率作为相应的概率值, 因此, 基于 3.2.2 中的式(11)~式(14)即可得到  $X_{nmt}$  的概率分布和  $Y_{nmt}$  的概率分布.



**LI Ling**, born in 1981, Ph.D. candidate, lecturer. Her current research interests include service oriented computing and Web services composition.

**LIU Min**, born in 1970, professor, Ph.D. supervisor. His research interests include services computing and intelligent maintenance.

**CHENG Guo-Qing**, born in 1982, Ph.D. candidate, associate professor. His current research interests include reliability calculation and intelligent maintenance.

## Background

With the development of Internet and mobile information technologies, social networks have incrementally become the most popular applications. A social network connects organizations or individuals using smartphones with sharing of information through social networking applications such as Wechat, MySpace, Facebook, and scientific collaboration networks. Everyday users are sharing and exchanging the information by means of “word-of-mouth” communications in such large-scale social networks. Under such a circumstance, in-depth analyzing and quantifying random actions (to forward the information or not) of users for information

diffusion over social networks have become more and more important.

The rapid developments of communication and information technologies enable us to access, collect, and store the real-world big data on information diffusion, making the related research meaningful and versatile, meanwhile more challenging. Although the existing related works provide systematic approaches and useful tools for analyzing the effect of network users on information diffusion, the majority of them neglect the dynamic and random characteristics of this problem. The study presented in this paper is motivated

by the challenges of quantifying nodes' random forwarding actions and finding the most susceptible nodes, at the same time emphasizing the dynamics characteristics of this problem. The study aims to address three key questions: (1) which nodes are most susceptible? (2) which types of information they are most susceptible to? and (3) when they are most susceptible. The proposed model can help decision-makers to dynamically identify which users are most susceptible to the corresponding information at different periods of time.

This work is supported by the National Natural Science Foundation of China under Grant Nos. 71690234, 61573257, 71661016. These projects aim to provide better principles for the design and management of service systems. Many research papers have been published on respectable journals. The work in this paper contributes that it presents a new decision model for information diffusion over online social networks, which will open our mind in the research of information diffusion over service systems.

《计算机学报》