

在线社交媒体信息冗余现象建模与实证研究

秦烁¹, 吕欣^{1,2,3}, 孟凡辉⁴, 胡延庆⁴

1. 国防科技大学系统工程学院, 湖南 长沙 410073; 2. 中南大学商学院, 湖南 长沙 410083;
3. 卡罗林斯卡研究所公共卫生科学系, 瑞典 斯德哥尔摩 17177; 4. 中山大学数据科学与计算机学院, 广东 广州 510006

摘要

为深入了解社交媒体中的信息冗余现象, 定义了一种以个体好友信息发布的重复情况度量信息冗余程度的指标。推导个体在传播过程中出现冗余现象的概率, 分析了信息传播率、网络密度、集群系数对信息冗余的影响; 通过新浪微博的实际数据观察了冗余信息的扩散特点, 并从营销的角度探讨了信息冗余在广告推广和产品营销方面的意义。研究结果为进一步了解在线社交媒体的信息扩散特征提供了新的视角。

关键词

社交网络; 在线社交媒体; 信息传播; 信息冗余

中图分类号: N941.3

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2018050

Modeling and empirical research of information redundancy on online social media

QIN Shuo¹, LU Xin^{1,2,3*}, MENG Fanhui⁴, HU Yanqing⁴

1. College of Systems Engineering, National University of Defense Technology, Changsha 410073, China
2. School of Business, Central South University, Changsha 410083, China
3. Department of Public Health Sciences, Karolinska Institute, Stockholm 17177, Sweden
4. School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

Abstract

To learn more about redundancy phenomenon, a measurement to quantify this redundancy was proposed and the probability of individual redundancy in the process of information diffusion was deduced. Using simulated networks, how information propagation rate, network density, clustering coefficient affect information redundancy was investigated. The diffusion characteristics of redundant information through the actual data of Sina Weibo was observed and the perspective of information redundancy in advertising promotion and product marketing was discussed from marketing. The findings provide a new perspective for further understanding of the information diffusion characteristics of online social media.

Key words

social network, online social media, information diffusion, information redundancy

* 作者在国内外发表的论文均采用LU

1 引言

在线社交媒体是一类获取信息的重要渠道,更是一种新的信息传播方式。Facebook、Twitter、新浪微博、微信等作为一类基于朋友的关系型社交平台,通过整合用户的社交圈、生活圈、工作圈,满足用户在信息获取、交流沟通、娱乐消遣等方面的需求。但随着用户之间信息交互频繁,信息量日益增多,信息冗余现象也成为在线社交媒体发展过程中一个不可忽视的问题。当信息量超出人们的吸收能力时,就会使人们感到“数据窒息(datacide)”。直观上,当人们接收到过多重复信息或大量广告推送时,会影响用户获取信息的效率和体验感。对信息自身而言,信息冗余过于严重会导致信息传播效率低,信息丧失原有价值。除此之外,社交网络平台的活跃度及未来发展趋势也都与信息冗余问题息息相关。

Jin X^[1]等人在如何维持社交平台活跃度问题上,从网络增长模式和节点连接类型的角度研究了网络活跃性演化的机制。其实验结果也在一定程度上还原了一些真实社交网络平台的发展情况。人人网在早期顶着“中国Facebook”的光环,比它模仿的对象Facebook发展得更蒸蒸日上、充满朝气,最终,作为当时国内最大的社交平台和承载无数年轻人回忆的互联网产品却走向低谷。人人网除了在产品定位转型以及平台运营等方面存在问题之外,更重要的是人人网中总是分享一些“老梗”、过时或“老掉牙”的消息,原创力差,输出少,用户渐渐失去新鲜感和兴趣,从而淡出人人网,平台也因此逐渐走向衰落。再如Reuters机构通过对1 300位企业经理人所做的调查显示:信息过载会降低工

作满意度,甚至工作中的重要的决策也会因信息过载而被延迟,并希望能够通过技术手段解决信息过载问题^[2]。

除此之外,虽然信息冗余会造成一定的负面影响,但一定量的冗余在广告推广或产品营销方面却具有一定的意义。社交媒体的快速发展成为一股不可忽视的力量,为商业界带来了巨大的变革,很多企业都将在线社交媒体作为一种新的营销渠道,并制定符合新媒体特点的营销策略。在线社交媒体的广泛普及让营销者逐步开始淡化营销信息主要传播者的角色,转而以口碑为载体,激励用户进行信息传播。“口碑效应”的本质其实是社会学和心理学中提到的“从众行为”,指人们会很大程度地受到周围大多数人的影响。已有研究表明,用户之间的相互推荐会影响到用户的喜爱和偏好,从而促进产生购买和消费行为。因此,若广告营销类信息在社交媒体中得到大范围的传播和扩散,那必定会为相应产品带来很好的宣传和销量。

目前,复杂网络理论已比较成熟,将复杂网络理论用于社交网络分析的方法已有一定的基础和先例^[3-7]。如关于在线社交平台传播模型的构建、信息传播特性等的研究也都不在少数。Zhang Z K等人^[8]对近年来有关信息传播的研究进行了回顾,对不同理论和方法进行了统一,比较和评估不同模型与算法之间的性能差异,强调有关信息传播的研究将在各种领域都具有很强的科学深度。

但是,目前从网络和节点特性、传播机制以及广告推广或产品营销的角度出发,探究信息冗余问题的相关研究屈指可数。因此,运用复杂网络理论分析社交网络,从网络和节点特性、信息传播机制和营销的角度探讨冗余问题,不仅为进一步了解信息传播特性提供了新的视角,在对信息获取、管理、优化等方面也具有一定

的理论意义和实践价值。

2 问题描述

在大部分社交平台(如Facebook、Twitter、Instagram、BBM、LINE等^[8])上,信息冗余常常表现为新闻或热门头条的重复转发及推送,信息重复度高,零碎性强。刷屏现象就是信息冗余的典型表现。

现实社交平台中,热门信息常常以社会性新闻或娱乐新闻为主,如2017年11月12日,北京市“红黄蓝幼儿园”的“虐童事件”在各大社交平台上相继出现刷屏现象,并引起强烈的社会反响。仅仅3h内,新浪微博中某用户的30条推送中就有19条与该新闻相关。

基于以上现象,将信息传播过程中的节点类型分为两类:覆盖节点,即在该节点周围的邻居节点中,有一人或一人以上发布同一信息;冗余节点,即在该节点周围的邻居节点中,有两人或两人以上发布同一信息。并以网络中冗余节点所占比例作为度量网络信息冗余度的指标。若将社交网络问题提炼成由节点和边构成的社交关系图上的问题,那么用户就是节点,由一条边相连的两个节点具有“好友关系”,每个节点对于某条信息在某一时刻的状态只表现为“发布”或“未发布”。将朋友圈转换为网络关系拓扑,如图1所示。

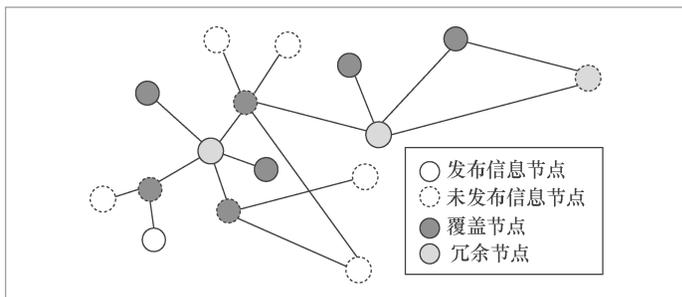


图1 “朋友圈”网络关系拓扑

3 信息冗余度模型

3.1 信息传播模型

为了深入研究社交媒体中的信息冗余现象,在社交平台用户关系网络中建立SI(susceptible-infected)信息传播模型,并根据实际信息的传播特点对传播过程加以限制。将用户对信息的“发布”状态设置为1,“未发布”状态设置为0。将关系网络中的所有节点状态分为两类:S(susceptible)为未发布信息的个体,I(infected)为已经发布信息的个体。其中,S在接触到传播者I后,会以一定概率 α (传播率)变为I,即该节点将分享信息到自己的朋友圈,此时呈现对该信息的状态为1。但是任何信息的热度都会随着时间以及传播而逐渐下降,因此,在传播模型中,每个节点在受到邻居节点发布信息的影响后,会做出“发布”或者“不发布”的决定,该决定只能产生一次,在初次决定后,节点的状态在之后的传播中不再受影响,也不再改变。

下面对冗余指标中的几个元素进行定义。 N 为网络规模, a_{ij} 为社交网络邻接矩阵中的元素,若 $a_{ij}=1$,则表明顶点 i 与 j 互为邻居节点,否则相反; $k(i)$ 表示任意节点的状态, $\xi(i)$ 表示任意节点周围是否出现信息冗余现象,可得到:

$$k(i) = \begin{cases} 0, \text{状态 S} \\ 1, \text{状态 I} \end{cases} \quad (1)$$

$$\xi(i) = \begin{cases} 0, \sum_{j=1}^N a_{ij} k(j) < 2 \\ 1, \sum_{j=1}^N a_{ij} k(j) \geq 2 \end{cases} \quad (2)$$

网络信息覆盖率 exp 为覆盖节点所占的比例,计算式为:

$$\text{exp} = \frac{\sum_{i=1}^N k(i)}{N} \quad (3)$$

同理,网络信息冗余率Re为冗余节点所占的比例,计算式为:

$$\text{Re} = \frac{\sum_{i=1}^N \xi(i)}{N} \quad (4)$$

3.2 网络结构模型及数据来源

由Watts D J 等人^[9]提出的WS小世界模型以及Barabasi A L 等人^[10]提出的BA无标度模型是复杂网络研究中的两个基本模型。现实生活中很多网络具备小世界和无标度网特性,如电力网、电影演员合作网等,社交网络也同样如此^[11-13]。研究中除了使用两个基本模型及可调集聚系数的无标度模型^[14]之外,在部分仿真实验中也选取了3个真实社交网络数据,包括美国社交平台Facebook网络、法国在线音乐社交网络、维基百科投票网络。网络的主要拓扑以及统计特征见表1。

4 研究结果

4.1 节点冗余概率

2016年,在微信平台上转发量达10万次以上的文章中,《罗一笑,你给我站住》

就是信息“病毒式”传播的典型代表。“罗一笑事件”的相关新闻在各大社交平台都出现了刷屏现象,事件相关转发量达到近1 000万次,传播链条最长达到74人,信息覆盖及二次曝光率位居当年前列。其传播路径可视化结果显示,信息传播呈现与随机游走方式相似的链条式特征,使得信息二次转发性强,扩散范围大。因此,首先假定社交网络中造成大范围冗余的信息传播模式为随机游走,探究节点在信息传播中发生冗余现象的概率 P_{O_i} 。

在随机游走过程中,在节点个数为 N 的网络中,任意节点 i 被访问的概率 Pr_i 与节点的度 d_i 成正比^[15-18]:

$$\text{Pr}_i = \frac{d_i}{\sum_{j=1}^N d_j} \quad (5)$$

对于节点 i ,设相邻节点的序号分别为 k_1, k_2, \dots, k_{d_i} ,当随机游走步长 $n=1$ 时,定义任意一个节点 i 的邻居节点中,有一个及以上节点发布同一信息的概率为:

$$p_i = 1 - p \left[\prod_{l=1}^{d_i} (1 - \text{Pr}_{k_l}) \right] \quad (6)$$

其中,令 i 节点周围任意节点都未被访问的概率为 z_i , $z_i^{(n)}$ 为 n 步下的概率,则:

$$z_i = \prod_{l=1}^{d_i} (1 - \text{Pr}_{k_l}) \quad (7)$$

n 步后,节点 i 是覆盖节点的概率为:

$$\text{Pe}_i = 1 - (1 - p_i)^n - \text{Pr}_i^2 \left[\sum_{l=0}^{n-2} (1 - \text{Pr}_i)^l \right] \quad (8)$$

表1 3种社交网络的主要拓扑及统计特征

网络名称	节点数 N /个	边数 M /条	平均度 k	最大度 k_{\max}	平均集聚系数 C
Facebook网络	4 039	88 234	43	1 045	0.605 5
在线音乐社交网络	41 773	125 826	6	112	0.091 2
维基百科投票网络	7 115	103 689	26	1 065	0.140 9

同理, $n=1$ 时, 定义任意一个节点*i*的邻居节点中有两个及以上节点被访问的概率为:

$$q_i = 1 - \left[\prod_{l=1}^{d_i} (1 - \text{Pr}_{k_l}) \right] - \sum_{l=1}^{d_i} \left[\text{Pr}_{k_l} \prod_{m=1, m \neq l}^{d_i} (1 - \text{Pr}_{k_m}) \right] = p_i - \sum_{l=1}^{d_i} \left[\text{Pr}_{k_l} \prod_{m=1, m \neq l}^{d_i} (1 - \text{Pr}_{k_m}) \right] \quad (9)$$

其中, 设节点*i*周围恰好有一个节点被访问的概率为 w_i , $w_i^{(n)}$ 为随机游走*n*步下的概率:

$$w_i = \sum_{l=1}^{d_i} \left[\text{Pr}_{k_l} \prod_{m=1, m \neq l}^{d_i} (1 - \text{Pr}_{k_m}) \right] \quad (10)$$

当 $n=2$ 时, 任意一个节点邻居节点中有两个及以上节点被访问的概率为:

$$q_i^{(2)} = p_i^{(2)} - w_i^{(2)} = p_i^{(2)} - (w_i z_i + z_i w_i) \quad (11)$$

以此类推, 随机游走*n*步后, 该节点邻居节点中有两个及以上节点被访问的概率为:

$$\text{Po}_i = \text{Pe}_i - (w_i z_i^{n-1} + z_i w_i z_i^{n-2} + \dots + z_i^{n-1} w_i) - t_i = \text{Pe}_i - C_n^1 w_i z_i^{n-1} - t_i \quad (12)$$

排除节点*i*周围任意节点在恰好只有一个邻居节点被访问的情况下, 被重复多次访问的概率:

$$t_i = \sum_{l=1}^{d_i} \left\{ \sum_{t=2}^n \left[C_n^t \text{Pr}_{k_l}^t (1 - \text{Pr}_{k_l})^{n-t} \prod_{m=1, m \neq l}^{d_i} (1 - \text{Pr}_{k_m})^n \right] \right\} \quad (13)$$

因此, 根据网络中的节点度, 就可以大致得到该节点在传播过程中出现冗余现象的概率。为了验证数学解析得到的节点冗余概率与传播过程中实际发生冗余现象节点比例的吻合程度, 建立由*N*个节点组成的无标度网络, 信息以随机游走方式进行传播。图2显示了对于特定 Po_i 值而言, 在多

次实验后, 实际有多少比例的节点出现了信息冗余现象。仿真结果显示, Po_i 与实际结果吻合较好, 散点大部分落在标准线附近。

4.2 不同因素对信息冗余的影响

本文主要从以下几个方面探索不同影响因素对信息冗余的影响。

(1) 传播率 α

在信息传播过程中, 定义传播率 α 为任意用户发布事件的可能性。研究中利用信息的传播率来代表微信、微博等现实社交平台中事件的热度。高频率出现或多人发布的内容往往是热门程度较高的消息, 如社会突发事件、娱乐八卦、时事政治等类型的事件, 往往表现出强大的传播和扩散能力。

(2) 节点度*d*

节点度指标给出了整个网络中每个节点拥有的好友数量, 是网络节点重要程度的初步体现。一般而言, 节点度值越大, 该节点与其他节点交互越多, 对网络具有更大的影响, 拥有更多获取信息的机会和途径。网络平均度*D*为网络中所有节点的度的平均值。

(3) 集群系数*C*

集群系数的高低代表网络中“三角形”结构数量的多少。

4.2.1 传播率对信息冗余的影响

首先, 建立网络规模为*N*的WS小世界网络模型, 分析不同传播率条件下, 信息冗余程度的变化趋势。从图3可以看出, 信息传播过程符合网络从一个传播节点开始, 直到所有个体都获取到信息的模型假设。并且值得注意的是, 若对传播中的某一时刻(如 $t=50$)进行分析, 可以发现, 同一时刻下, 信息重复率与传播率

正相关。这说明,传播率高的信息往往具有较为强大的扩散和传播能力,信息冗余现象受当下舆论热点话题、头条新闻等热门事件影响很大。

4.2.2 网络密度对信息冗余的影响

本文将网络平均度 D 作为衡量网络密度的指标。图4显示在同一时刻下(如 $t=70$),网络密度越大,信息重复率越高;除此之外,网络平均度达到某个值时会导致网络信息重复率突然升高。比如 $D=3$ 时,信息还不能得到有效的扩散,但当网络中每个节点平均只多一个好友时,即 $D=4$,在同样的传播时刻下,网络信息重复率出现大幅度升高,大大增加了信息的冗余。因此,网络信息重复率的高低、信息冗余程度不仅与节点度这一属性相关,还对节点度的变化非常敏感。

4.2.3 集聚系数对信息冗余的影响

社交网络中,集聚系数的大小代表网络中三角形关系结构的多少。一个人的朋友之间往往也可能是朋友,这种节点好友之间也互为朋友的结构会影响节点间获取信息的概率和机会。如在B、C分别是A的好友,但B、C之间互不相识的情况下(如图5(a)所示),A发布信息的状态对B是否能获取到该信息起决定性作用;但在B、C分别是A的好友且B、C之间也互为好友的情况下(图5(b)),即使A不发布信息,B也同样能从C处获取到信息,这大大增加了节点B获取信息的机会和途径。因此,在可调集群系数的无标度网络上分析集群系数与信息冗余之间的关系。如图6所示,在多次实验后,可以明显看出随着网络集聚系数的增加,网络中三角形结构增多,网络信息重复率得到了明显的抑制和降低。

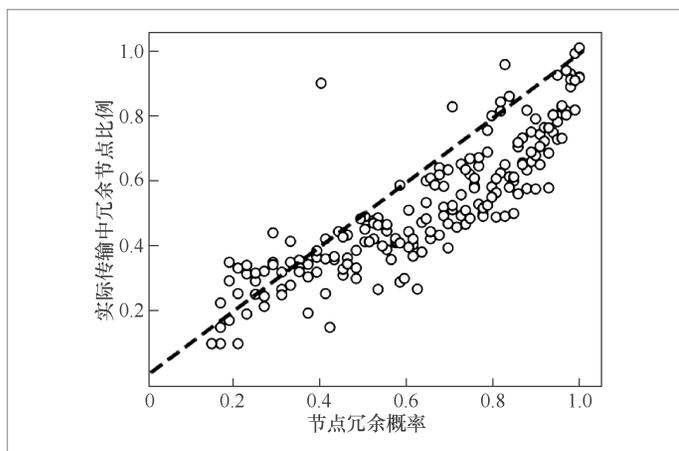


图2 节点冗余概率与传播过程中实际冗余节点比例吻合度

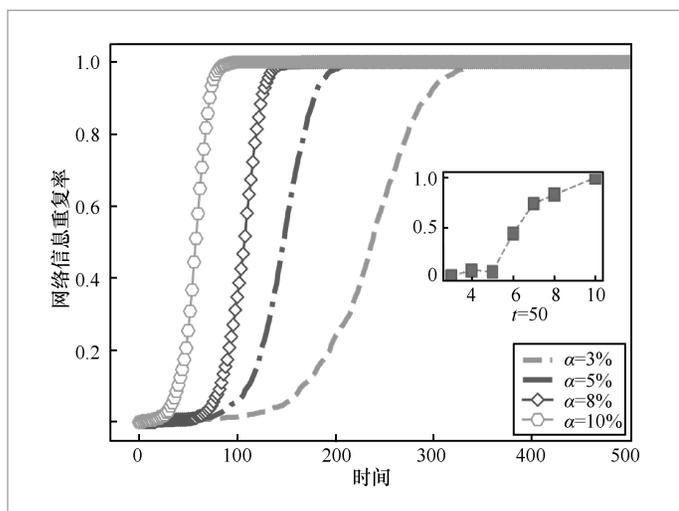


图3 传播率对信息重复率的影响(网络参数 $N=2\ 000$, $D=4$)

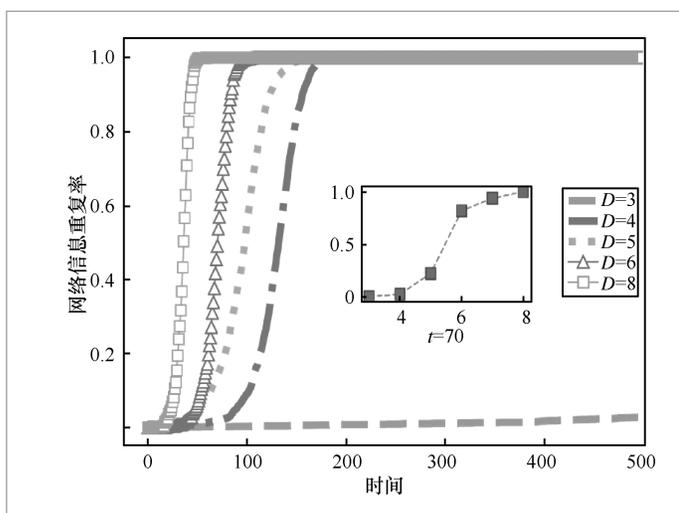


图4 网络复杂度对信息重复率的影响(网络参数 $N=2\ 000$, $\alpha=3\%$)

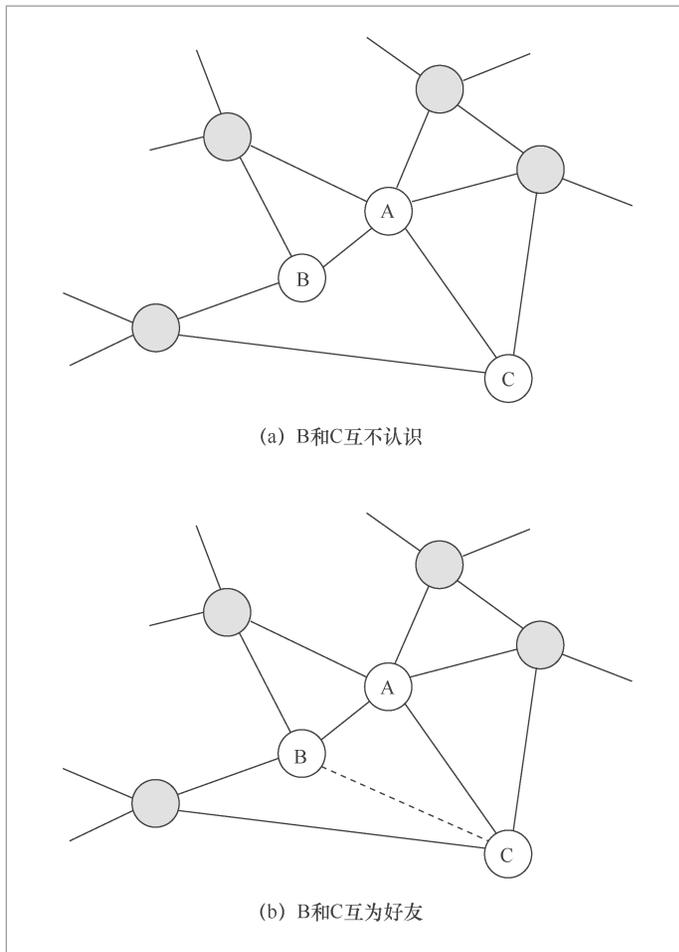
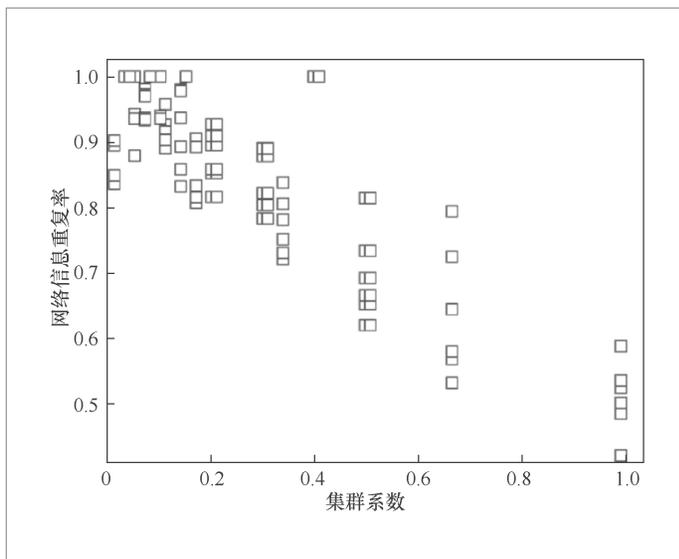


图5 三角形人际结构关系

图6 可调集群系数的无标度网络下, 信息重复率随集群系数变化趋势 (网络参数 $N=2\ 000$, $M=3$, $t=5$)

网络可视化有助于直观地了解网络特征、网络任意时刻的节点状态。因此, 为了探究集群系数增大导致网络信息重复冗余率被抑制的原因, 笔者选取了两个集群系数有所不同的网络模型, 分别对两种网络结构下的信息传播过程进行了可视化, 如图7所示。黑色和灰色分别代表已发布信息 and 未发布信息的节点。可视化结果显示, 在网络中信息覆盖节点数相同的情况下, 集群系数较大, 即网络中具有较多的“三角形”结构的B网络模型, 其信息扩散范围明显没有A模型大, 且B网络中已发布信息的节点S在网络中的位置非常集中, 而A模型中S类节点覆盖范围广, 从网络中心一直到网络的边缘, 传播链长。这说明三角形人际关系结构虽然增加了信息扩散的机会和途径, 但这种结构的“小圈子”封闭性会导致信息的传播受到抑制, 使信息传播存在“盲区”^[19], 信息传播限制在一个圈子中, 降低了信息扩散的范围。因此, 集群系数的增大对信息冗余有明显的抑制作用, 社交网络中这类三角形结构有助于抑制传播中的信息冗余现象。

4.3 冗余信息特征分析

为观察普通信息与二次转发性较强信息之间的差异, 分析信息冗余率随信息传播的变化趋势。分别在3种社交网络上进行了仿真, 结果如图8所示。随着信息转推层数的增长, 信息覆盖率和冗余率不断升高, 但随着时间以及信息转推层数的逐渐增多, 信息热度渐渐下降, 传播人数逐渐趋于稳定。同时, 在新浪微博约4.3亿个节点的网络上, 选取了近3个月内热门的10条信息进行了实际数据验证, 如图9所示, 信息覆盖率、冗余率增长趋势与仿真一致。

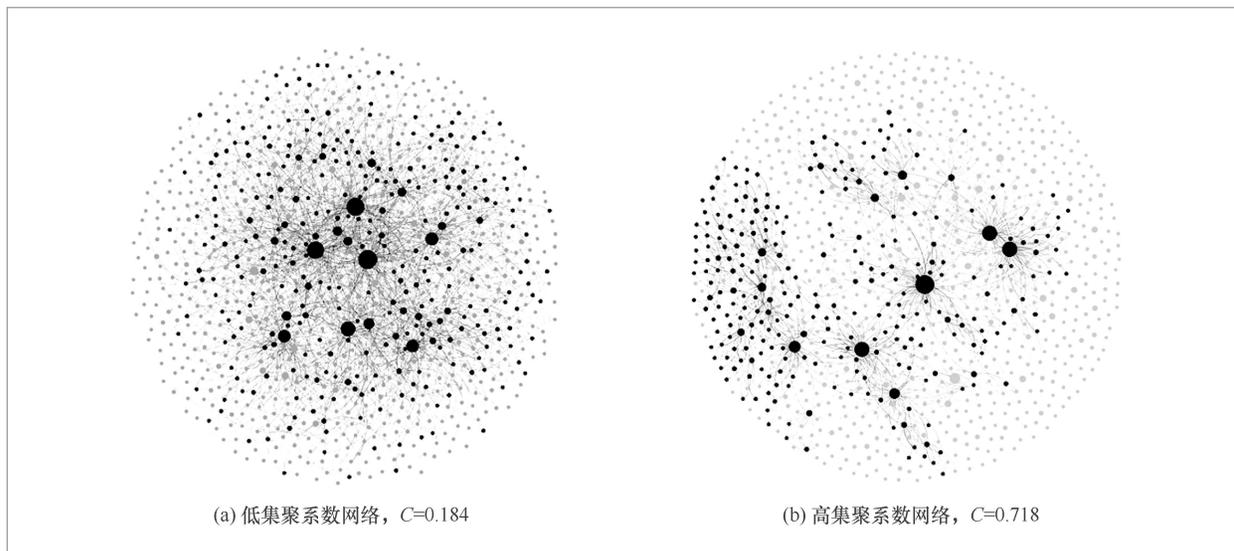


图7 不同聚类系数网络模型的信息传播可视化（网络参数 $N=2\,000$, $M=3$ ）

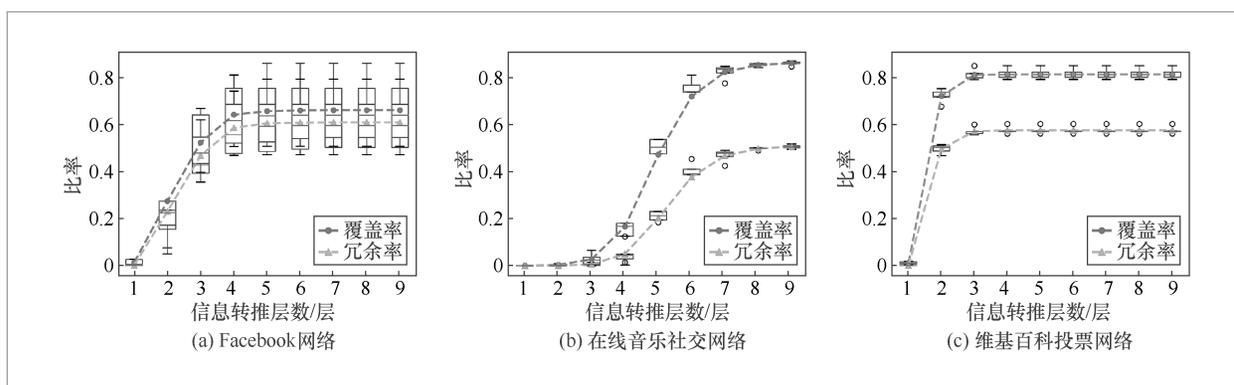


图8 3种真实社交网络上信息覆盖率及冗余率信息转推层数的变化趋势

值得注意的是，从10条信息的传播过程可以看出，热门信息及容易刷屏的信息主要分为两类：第一类信息如图9(a)、图9(b)所示，这类信息由大V博主发布，如娱乐明星或企业机构的官方微博，内容以广告推广或明星的个人宣传为主。在传播特点上，这类信息在第二层或第三层的转发量就可以达到全部转发量的80%左右，也就是说，博主仅依靠自身拥有的粉丝而产生的转发量就可以达到非常可观的覆盖效果。而在第二类信息传播过程中，信息覆盖率和冗余率变化有明显拐点，说明这类信息在传播过程中经过了一

些大度节点，并引起了新一轮小范围的爆发。这类信息的发布者可能只是一个粉丝量较少的普通用户，但在信息内容上极具价值，受众面较广，不针对某一类用户的喜好，如社会性新闻等。虽然在传播初期的覆盖面不广，但由于信息内容具有价值优势性，容易引起社会共鸣，虽由普通用户发布，但在后期却能经过大V传播，最终得以爆发。

除此之外，两类信息在传播达到稳定后，信息冗余率也相差较大。相比之下，第二类信息在传播稳定后信息冗余率更高，换言之，更容易出现刷屏现

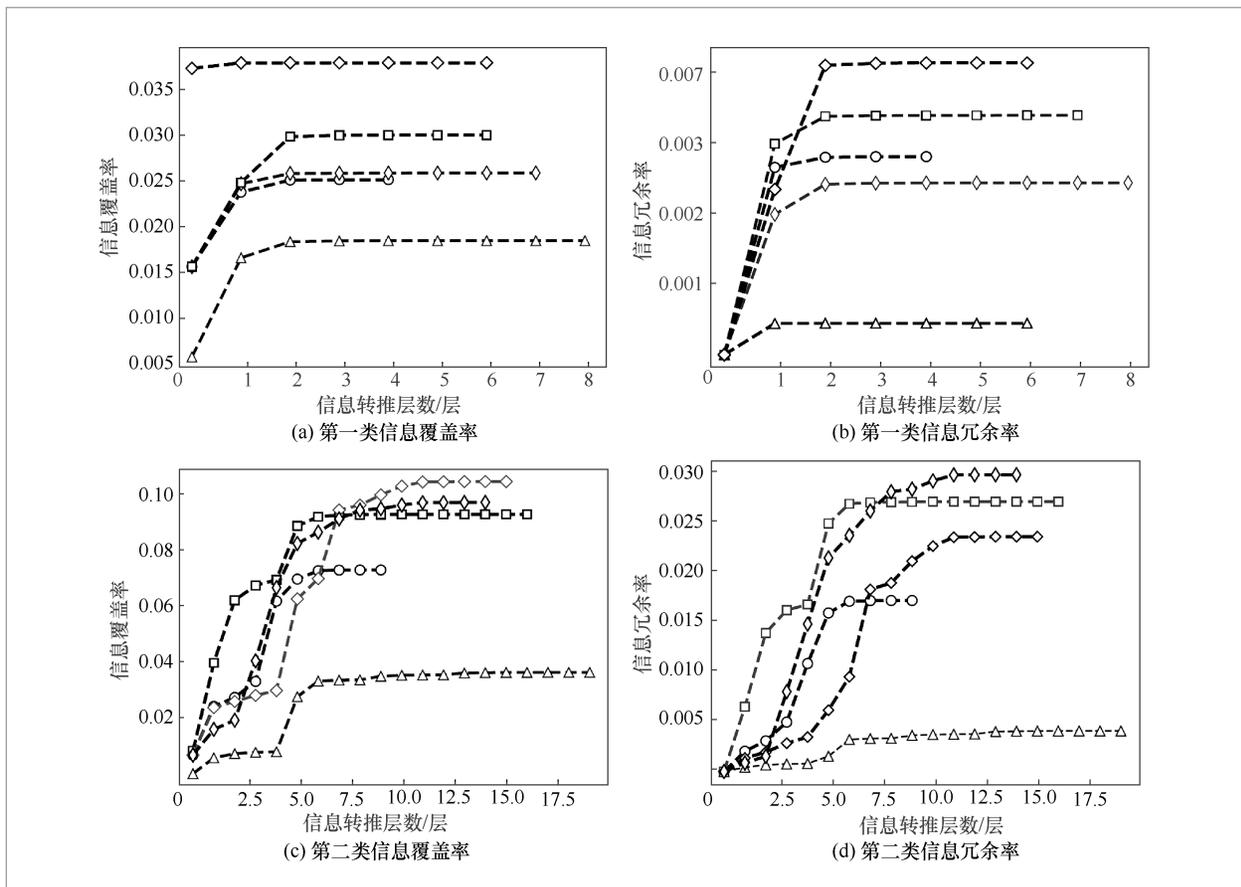


图9 新浪微博近3个月内10条热门信息覆盖率及冗余率随传播层数的变化趋势，每条曲线代表一条信息

象，维持热度久。这是由于进入大众视野的新闻虽然可能只是由一个小人物爆出的，但在内容上能引起社会的共鸣，而非通过“买热搜”“买水军”等方式产生的。同时，这一现象也说明，小度节点同样也可以带来很大的信息冗余。

4.4 基于冗余度传播策略研究

这一部分考察冗余率指标在广告推广和产品营销方面的意义。在进行广告推广时，营销者的首要目标是在广告信息投放后，尽可能使信息的覆盖率最大。但从信息冗余的角度出发，更希望在保证信息具有一定覆盖率的同时，也能

尽量降低信息冗余率，以达到在广告投放时，既能有良好的宣传作用，也不至于产生大量广告信息刷屏现象，从而引起用户体验感的降低以及对冗余信息的反感。

首先以不同的策略选择节点源，例如按随机选择、按节点度、 k -shell值、集聚系数等方式选择传播源，并在实际的社交网络上进行仿真。结果显示，在社交网络中，信息一旦传播开，信息的冗余率是不易控制的，无论选择节点源的标准如何，对信息传播后期冗余率的增长都没有实质性的影响，信息的冗余率都会随着信息覆盖率的增长而快速增长。由此，提出一种“定点投放”策略。同样是根据不同的标准在网络中选择节点源，但在该种策

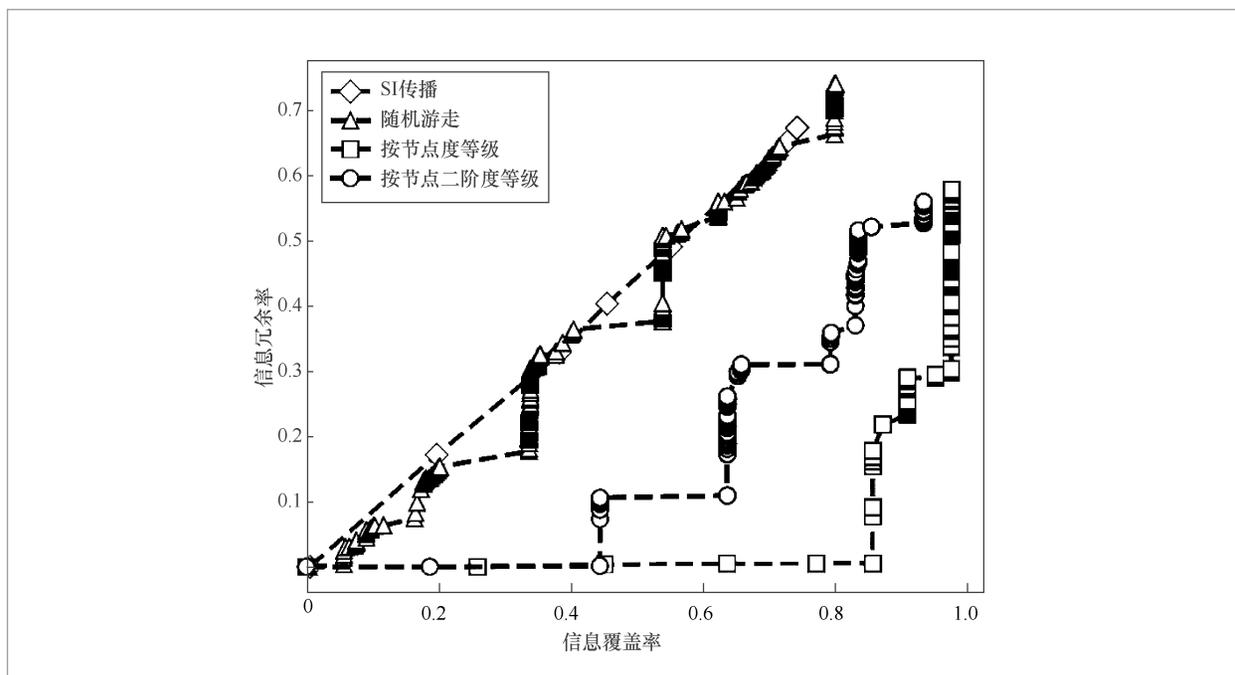


图10 Facebook网络中定点投放策略下节点信息冗余率随覆盖率的变化趋势

略中,信息不依靠转发的方式一层一层地向外扩散,而是让相应的节点进行信息展示和发布。如图10所示,通过在Facebook网络上进行仿真,可以得出,选择网络度较大的节点,如娱乐明星、大V以及官方微博等,可以很好地达到信息的大面积覆盖,并且当节点数目在一定范围内时,只会造成很小的信息冗余。但这一策略也存在缺陷,虽然实现了高覆盖、低冗余的目标,但就广告的收益而言,过低的冗余可能会造成大部分用户对广告的忽略,因此,合理的信息冗余才能实现广告宣传与用户体验的双赢。

5 结束语

本文基于复杂网络理论,定量地分析了社交网络中的信息冗余问题,定义了一种以个体好友信息发布的重复情况度量信息冗余程度的指标。首先,推导了个体

在传播过程中出现冗余现象的概率,分析了信息传播率、网络密度、集群系数对信息冗余的影响。仿真结果显示,信息传播率、网络密度与信息重复率正相关,而集群系数的增长对信息重复率有明显的抑制作用。其次,探究了信息传播过程中覆盖率与冗余率的变化趋势及冗余信息的扩散特点,为增加模型的可靠性和说服力,选取了新浪微博近3个月内的10条实际热门信息,与仿真结果进行比对。结果显示,在现实社交平台中,二次曝光及转发性较强的信息更容易进入大众视野并产生刷屏现象,小度节点也可以造成很大的信息冗余。最后,还从营销的角度探讨了信息冗余在广告推广和产品营销方面的意义。按照节点度等级对信息进行定点投放的策略,可以在传播初期很好地实现信息大范围覆盖,同时只造成很小的冗余。因此,本文的研究结果可以有效地评估节点发生信息冗余现象的概率,为进一步了解在线社交媒体

的信息扩散特征提供了新的视角,对信息传播在营销领域的应用也有一定的帮助,有助于深入了解网络属性对信息传播的影响,对掌握信息传播规律、冗余信息特点、社交平台舆情都有一定的理论和实践意义。

参考文献:

- [1] JIN X, JIN C, HUANG J, et al. Coupling effect of nodes popularity and similarity on social network persistence[J]. *Scientific Reports*, 2017(7): 42956.
- [2] ANGELA E, ANNE M. The problem of information overload in business organizations: a review of the literature[J]. *International Journal of Information Management*, 2000, 20(1): 17-28.
- [3] BAKSHY E, ROSENN I, MARLOW C, et al. The role of social networks in information diffusion[C]//The 21st International Conference on World Wide Web, April 16-20, 2012, Lyon, France. New York: ACM Press, 2012: 519-528.
- [4] GUILLE A. Information diffusion in online social networks[C]//The 2013 SIGMOD/PODS Ph.D. Symposium, June 23, 2013, New York, USA. New York: ACM Press, 2013: 31-36.
- [5] LI P, LI W, WANG H, et al. Modeling of information diffusion in Twitter-like social networks under information overload[J]. *Scientific World Journal*, 2014, 2014(5721): 914907.
- [6] HALBERSTAM Y, KNIGHT B. Homophily, group size, and the diffusion of political information in social networks: evidence from Twitter[J]. *Journal of Public Economics*, 2016(143): 73-88.
- [7] GRUHL D, GUHA R, LIBEN-NOWELL D, et al. Information diffusion through blogspace[C]// International Conference on World Wide Web, May 17-20, 2004, New York, USA. New York: ACM Press, 2004: 491-501.
- [8] ZHANG Z K, LIU C, ZHAN X X, et al. Dynamics of information diffusion and its applications on complex networks[J]. *Physics Reports*, 2016, 651: 1-34.
- [9] WATTS D J, STROGATZ S H. Collective dynamics of "small-world" networks[J]. *Nature*, 1998, 393(6684): 440.
- [10] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509.
- [11] NEWMAN M E. The structure of scientific collaboration networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98(2): 404-409.
- [12] BARABASI A L, JEONG H, NEDA Z, et al. Evolution of the social network of scientific collaborations[J]. *Physica A Statistical Mechanics & Its Applications*, 2002, 311(3): 590-614.
- [13] MA Y H, LI H J, ZHANG X D. Strength distribution of novel local-world networks[J]. *Physica A Statistical Mechanics & Its Applications*, 2009, 388(21): 4669-4677.
- [14] HOLME P, KIM B J. Growing scale-free networks with tunable clustering[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2002, 65(2): 026107.
- [15] LU X. Improving sample estimate reliability and validity with linked ego networks[J]. *Methodology*, 2012: arXiv:1205.1971v1.
- [16] SALGANIK M J, HECKATHORN D D. Sampling and estimation in hidden populations using respondent-driven sampling[J]. *Sociological Methodology*, 2004, 34(1): 193-240.
- [17] HECKATHORN D D. Extensions of respondent-driven sampling: analyzing continuous variables and controlling for differential recruitment[J]. *Sociological Methodology*, 2007, 37(1): 151-207.

[18] VOLZ E, HECKATHORN D D. Probability based estimation theory for respondent driven sampling[J]. Journal of Official Statistics, 2008, 24(1): 79-97.

[19] ZHANG Z K, ZHANG C X, HAN X P, et al. Emergence of blind areas in information spreading[J]. Plos One, 2014, 9(4): e95785.

作者简介



秦烁(1995-),女,国防科技大学系统工程学院硕士生,主要研究方向为复杂网络传播动力学。



吕欣(1984-),男,国防科技大学系统工程学院副教授,主要研究方向为大数据挖掘、人类行为动力学分析。



孟凡辉(1993-),男,中山大学数据科学与计算机学院硕士生,主要研究方向为计算传播学。



胡延庆(1980-),男,博士,中山大学数据科学与计算机学院副教授,主要研究方向为复杂系统理论。

收稿日期: 2018-08-15

通信作者: 吕欣, xin.lu@flowminder.org

基金项目: 国家自然科学基金资助项目(No.71522014, No.71771213, No.61773412, No.U1711265)

Foundation Items: The National Natural Science Foundation of China(No.71522014, No.71771213, No.61773412, No.U1711265)