

樊鹏翼, 李伟, 王晖

(国防科学技术大学 信息系统与管理学院, 湖南 长沙 410073)

: 近年来, 微博网络得到迅猛发展, 与之相应的网络群体意见传播建模研究得到人们的普遍关注。本文以 SIR 传染病模型为基础, 针对微博网络正负关系属性, 提出了正负关系网络意见传播模型。随后, 利用微博意见传播实际数据, 对该模型在微博网络中的传播过程进行对比分析, 验证了模型的有效性, 从理论与实证的角度推动了意见动力学的研究。

微博; 意见传播; 建模; 实证研究

0

近年来, 随着移动通信网络和 Web 2.0 技术的不断发展, 微博逐渐成为人们日常交流、通信、娱乐的基本工具。同时, 由于微博网络自身存在的交互性、实时性、广泛性、开放性等特点, 使其成为舆论传播中的一把双刃剑, 因此, 为了实现微博网络舆论的有效管理与引导, 舆论传播建模研究得到了人们的广泛关注。现有关于舆论(或意见)动力学建模方法研究, 大都是从社会学或者物理学的角度对社会网络节点意见的传播机理与交互机制进行数学抽象, 将其类比为物理学中粒子交互过程, 提出了基于粒子交互模型的意见动力学模型, 如传染病模型^[1]、Voter 模型^[2]、Majority Rule 模型^[3]、Sznajd 模型^[4]、Deffuant 模型^[5]、Hegselmann-Krause 模型^[6]等, 用于解释与分析舆论的形成与演化。然而, 以上模型缺乏实际意见传播系统的验证, 与实际的传播分析与控制应用仍存在一定的差距。

正负关系是社会网络的一个重要属性, 表示网络节点之间不仅存在正向(朋友、信任、支持等)关系, 而且也存在负向(敌对、不信任、反对等)关系, 这些网络称为正负关系网络或符号网络(Signed network)。现有的研究发现^{[7],[8]}, 许多在线社会网络均存在正负关系属性, 即节点之间存在“朋友”与“异己”两种不同的关系, 例如, Slashdot、Epinions 与 Wikipedia 等。同样, 微博网络中存在大量的“朋友”与“异己”关系, 即正负关系, 从而对意见的传播机理与传播过程带来较大的影响。一般而言, 正关系能够促成意见的一致, 而负关系将使得意见产生对立。因此, 考虑到关系的正负属性, 意见传播动力学模型仍需要进一步的改进与优化, 以适应微博网络中意见传播实际过程。

本文将传统的“朋友”关系网络扩展至更为泛化的正负关系网络, 提出了意见传播动力学模型。随后, 以微博网络“方韩大战”话题数据为依据, 将模型的演化过程与实际的意见传播过程进行对比分析, 发现本文提出的模型具有较好的扩展性, 能够较好的解释与分析微博网络中群体意见的传播过程。

1

网络传染病模型不仅仅用于传统的疾病、细菌、病毒等传播过程的建模、分析与控制, 还可以用来解释与分析信息、谣言、意见等在网络中的传播过程。基本的 SIR 模型中, 网络中的个体只能处于如下三种状态之一: 未感染状态 S, 表示个体不携带病毒, 但容易被病毒感染; 感染状态 I, 表示个体已经被病毒感染, 而且具有传染性; 免疫状态 R, 表示个体已被感染, 然后恢复健康并具有免疫性。本文主要采用 SIR 模型来描述个体意见在社会网络中的传播过程, 因此, 在具体的建模过程中需从以下三个方面进行考虑:

(1) 不同于病毒或者细菌, 个体意见具有更高维的属性, 很难用简单的二元状态(感染或未感染)进行描述。例如, 对于一个争议话题或者一场选举活动, 个体的意见往往呈现出支持、反对以及中立三种状态。因此, 在建模时应当从意见的基本属性出发, 对基本的 S-I-R 状态进行相应的扩展;

(2) 传统的网络流行病模型中, 节点之间的关系均为同构, 即节点之间仅存有朋友关系。然而, 实际的网络测量结果显示, 社会网络中不仅存有大量的朋友(或信任)关系, 而且还有许多异己(或不信任)关系, 即社会网络中的正负关系。因此, 在网络意见传播建模时, 必须考虑社会网络中节点关系的正负属

性特征,使其更符合实际的网络结构。

(3)一般而言,社会网络中节点关系的异构特性将直接影响个体意见在网络中的交互机理与传播过程。例如,相同或相似意见更倾向于在朋友之间进行传播,而异己关系节点之间更容易产生相异意见。因此,本文提出的模型,必须考虑异构关系对个体意见交互的影响与作用,使得模型更能揭示网络中意见的传播规律。

给定社会网络 $G = \langle V, E, A \rangle$, 其中 V 表示社会网络中的节点集合且 $|V| = N$, $E \subseteq V \times V$ 表示网络中节点相连形成的边集, 而 $A \in R^{N \times N}$ 表示网络中节点之间的正负关系: 如果节点 i 与节点 j 为朋友关系, 则 $a_{ij} = 1$; 如果节点 i 与节点 j 为异己关系, 则 $a_{ij} = -1$; 如果节点 i 与节点 j 之间互不关心或互不相识, 则 $a_{ij} = 0$ 。对于特定话题, 假设社会网络中的所有个体具有如下五种状态之一: (1) 中立未感染状态 S , 表示此时该节点持有中立观点, 同时也可以被其他节点的观点所影响而改变; (2) 正向感染状态 I_1 , 表示此时该节点持有正面观点, 同时也可以影响其他未感染的节点使其发生观点的变化; (3) 负向感染状态 I_2 , 表示此时该节点持有反面观点, 同时也可以影响未感染节点的观点; (4) 正向移除状态 R_1 , 表示此时该节点持有正面观点, 但由于感染时间过长导致话题兴趣的丢失, 使其不再向其他节点传播自己的观点, 即不再具有传染性; (5) 负向移除状态 R_2 , 与 R_1 定义相似, 此时该节点持有反面观点且不具有传染性。在市场营销学中, 状态为 I_1 和 I_2 的节点被称为助长者 (promoter) 与诋毁者 (detractor) [9]。

与基本的 SIR 模型相似, 在 t 时刻, 如果一个未感染节点 S 和网络中其他一个感染节点 I_1 或 I_2 进行意见交互, 则在 $t+1$ 时刻, 该节点将会以一定的概率 β 改变自身的观点, 节点状态由未感染状态转变为正向 (或负向) 感染状态; 同时, 感染节点也会在 $t+1$ 时刻以概率 γ 转变成为移除状态 R_1 或 R_2 。由于此处的概率刻画了单位时间内节点状态的转变速率, 因此, 如果没有特殊说明, 本文提到的传播速率、移除速率等同于此处的感染概率与移除概率。

与基本的 SIR 动力学过程不同之处是: 在意见传播过程中, 未感染节点的观点转移状态与转移概率是由交互节点之间的正负关系所决定, 并且正关系利于传播相同意见, 而负关系利于相异意见的传播。因此, 本论文提出的意见传播演化动力学可以描述为以下模式:

事件	传播速率	节点关系 a_{ij}
$S + I_1 \rightarrow I_1 +$	β_1	+1
$S + I_2 \rightarrow I_2 +$	β_1	+1
$S + I_1 \rightarrow I_2 +$	β_2	-1
$S + I_2 \rightarrow I_1 +$	β_2	-1

如上式所示, 当状态为 S 的节点 i 与感染节点 j 在 t 时刻相遇, 如果 $a_{ij} = 1$, 则在 $t+1$ 时刻, i 节点的状态以概率 β_1 转变成为与节点 j 相同的感染状态; 如果 $a_{ij} = -1$, 则 i 节点的状态以概率 β_2 转变成为与节点 j 相异的感染状态; 如果 $a_{ij} = 0$, 则两者互不影响, 将保持原有的状态。同时, 如果持相异观点的节点相遇 (I_1 与 I_2), 或者交互节点之一的状态为移除态 (R_1 或 R_2) 时, 交互双方也将保持原有的观点, 无法形成意见的传播。

2

微博意见传播是一个动态的演化过程, 主要体现在节点动态性与意见动态性两个方面, 其中节点动态

性是指网络中的节点随时间不断增加与消亡（不再发表微博），而意见动态性则是指参与话题的节点意见不断发生变化。此外，由于节点的意见并不是显式存在，需要借助情感分析方法对文本进行更深入的分析与挖掘，实现意见的量化描述，这些因素均对微博意见传播的实证分析研究带来了很大的困难。

本文选取 2012 年微博热点话题“方韩大战”展开数据采集，采用“监控模式”对“方舟子”与“韩寒”的相关微博及网友对其回复数据进行采集，获得方舟子相关博文（原创与网友回复博文）为 2910256 条，韩寒相关的博文为 280926 条。随后，我们从已有的话题数据集中选取“方舟子质疑韩寒身高造假”这一子话题数据，其始于 2012 年 5 月 5 日，方舟子公开发表微博质疑韩寒身高存在“造假”问题，由此引发网友对该子话题的不断争论，形成较为独立的热门话题。由于该子话题持续时间长，数据相对集中，便于意见传播的实证分析。以该子话题的起始时间 t_0 为基准，以日为时间单位，从数据集中提取十天内的 ($t = t_1, t_2, \dots, t_9, t_{10}$) 关于该子话题的用户行为数据与文本数据。为了克服节点动态性对最终结果的影响，此处的节点集 V 选取在 t_0 前“方韩大战”数据集中出现的节点，并利用情感分析方法，对这些节点发表的微博文本进行情感分析，得到该节点在一定时间段内对该子话题的意见值，例如，如果节点 A 对方舟子的情感值为正，则表示该节点支持“方舟子质疑韩寒身高造假”，反之，则表示该节点反对该子话题；相似的，如果节点 A 对韩寒的情感值为正，则表示该节点反对“方舟子质疑韩寒身高造假”，反之亦然。

因此，我们可以得到 t 时刻节点集 V 对该子话题的意见集合 $O_t = \{o_i^t \mid o_i^t = opinion(i) \in \{-1, 0, 1\}, i \in V\}$ ，结合离散时间 t ，建立群体意见时间序列 $O_{t_0}, O_{t_1}, O_{t_2}, \dots, O_{t_{10}}$ 。

为了与传播模型节点状态保持一致，令 $s(t) = \frac{|S(t)|}{|V|}$ ，其中 $S(t) = \{i \mid o_i^t = 0, i \in V\}$ ； $r_1(t) = \frac{|R_1(t)|}{|V|}$ ，其中

$R_1(t) = \{i \mid o_i^t = 1, i \in V\}$ ； $r_2(t) = \frac{|R_2(t)|}{|V|}$ ，其中 $R_2(t) = \{i \mid o_i^t = -1, i \in V\}$ 。

3

基于“方韩大战”话题数据集，我们首先对 t_0 之前的节点正负关系进行计算，构建子话题传播所需的正负关系网络。该网络是一个有向、权重网络，其中，边的方向表示一个节点对另一个节点的单向情感关系，其隐含着信息的单向传播过程。例如，节点 A 对 B （例如，方舟子）存在一个有向边 e ，表示节点 A 能够接收 B 的信息或意见，同时他对 B 的关系值为 $a = sign(w_e)$ ， w_e 为该有向边的权值。因此，该网络中的有向边既能表示信息的传播方向（与边的方向相反），也能反映节点的正负关系属性（与边的方向一致）。

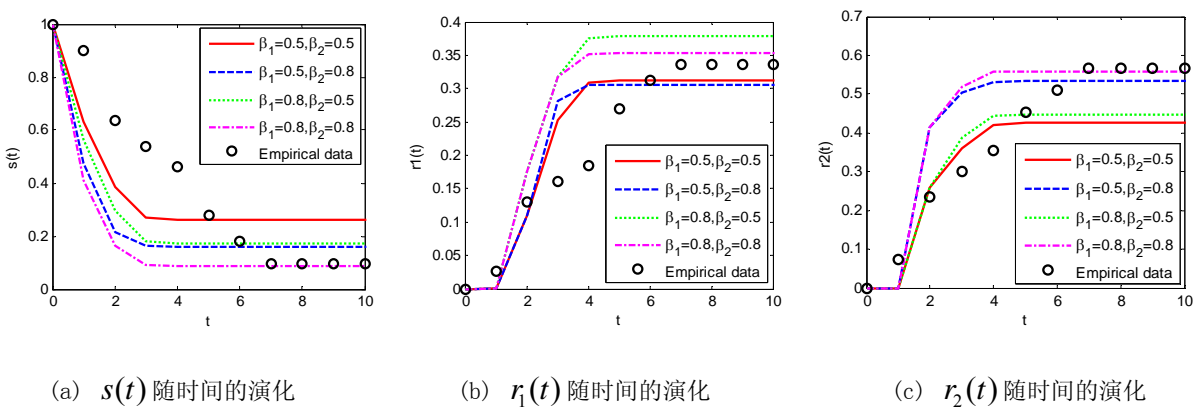


图 1 系统演化与实证传播结果对比

假设 $t = t_0 = 0$ 时，网络中节点意见的初始条件为 $o_{\text{方舟子}} = 1$ ， $o_i = 0, i \in (V - v_{\text{方舟子}})$ ，分别设定正关系传播速率 $\beta_1 = \{0.5, 0.8\}$ ，负关系传播速率 $\beta_2 = \{0.5, 0.8\}$ ，移除速率 $\gamma = 1$ 。在进行了 30 次独立实验后，意见在该网络中传播与演化的平均结果如图 1 所示，其中，实际意见传播结果用圆圈表示。我们可以发现：

β_1 与 β_2 的值越小，则最终未感染的节点密度越大，即未参与该话题的节点数目越少（如图 1(a)所示）。

β_1 与 β_2 的取值不同, 会使得最终意见演化的群体密度有一定变化 (如图 1(b)、图 1(c)所示)。

群体意见最终的传播结果与 $\beta_1 = 0.8, \beta_2 = 0.8$ 时模型演化结果一致, 但在随时间演化的过程中, 实际数据与模型的演化曲线仍存在一定的差异, 这主要是因为 SIR 模型假设节点进行意见传播 (感染) 的时间间隔相同, 均为单位时间, 而在实际的意见传播过程中, 用户行为往往不均匀, 具有幂率性与重尾特性^[1]未找到引用源。 , 从而造成意见实际传播过程与模型的演化过程不同步。

从以上的实证分析可以发现, 本文提出的传播模型能够较好的解释与分析微博网络中群体意见的传播过程, 而且参数 β_1, β_2 对系统的演化过程均产生重要影响。同时, 为了探寻参数 γ 对系统演化过程的影响, 本文假设以下场景: 方舟子在 $t = 0$ 时刻提出一个正向观点, 其他节点均为中立观点, 分别设定正关系传播速率 $\beta_1 = 0.8$, 负关系传播速率 $\beta_2 = 0.8$, 移除速率 $\gamma = \{1, 0.1, 0.01\}$ 。在进行 30 次独立实验后, 该话题在微博实证网络中传播与演化的平均结果如图 2 所示。

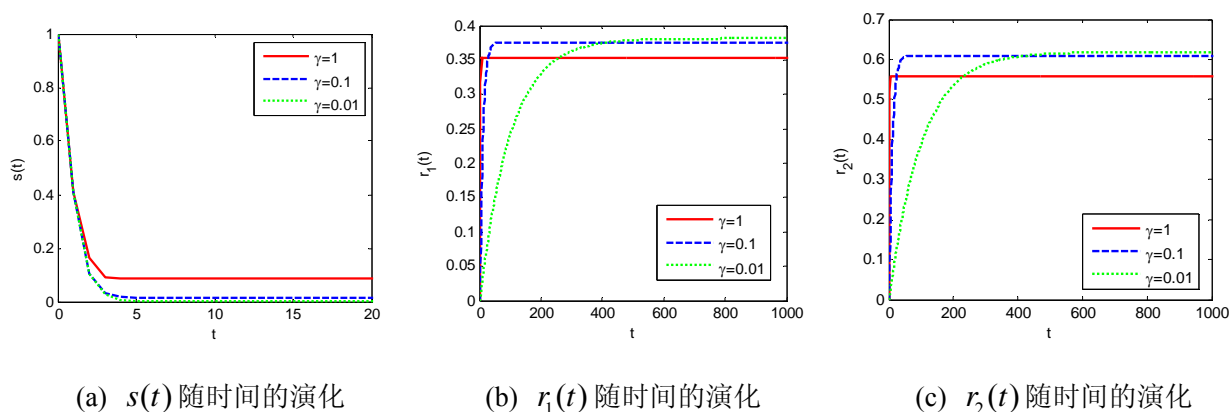


图 2 不同 γ 下系统演化结果

从图 2 可以看出, 系统的意见传播演化结果与参数 γ 的取值密切相关: 当 γ 较大时, 系统中未感染的节点密度较大, 而随着 γ 的减小, 系统稳定时间、感染节点密度都会不断增大。在微博网络中, 每个节点发表的微博都具有一定的传播时限, 因此, 如果能够延长微博的传播时限 (减小移除速率 γ), 将会使得更多的用户加入到话题的讨论, 从而形成意见更为广泛的传播。

从以上的分析可以发现, 当系统参数 β_1, β_2, γ 取值不同时, 最终的意见演化结果均产生一定的变化, 这说明该模型具有较好的扩展性, 便于真实意见传播数据的学习与训练, 适用于微博网络群体意见传播分析与预测。

4

本文将传统的“朋友”关系网络扩展至更为泛化的正负关系网络, 相应的提出了该类网络下的意见传播动力学模型。随后, 以微博网络“方韩大战”话题数据为依据, 将模型的演化过程与实际的意见传播过程进行对比分析, 发现本文提出的模型具有较好的扩展性, 并且能够较好的解释与分析微博网络中群体意见的传播过程, 从实际数据验证了动力学模型的有效性。同时, 本文的建模研究主要考虑了节点关系的正负属性, 并没有考虑关系的强弱属性, 其对意见的传播速率同样产生一定的影响, 因此, 本文提出的模型仍可以进行改进与优化。而且, 如何利用实证意见传播数据对本文提出的模型进行参数估计, 实现真实微博网络舆情传播预测应用, 也是下一步研究的主要方向。

- [1] Newman M E J. Spread of epidemic disease on networks[J]. Physical Review E, 2002, 66(1):016128.
- [2] Holley R A, Liggett T M. Ergodic theorems for weakly interacting infinite systems and the voter model[J]. The Annals of Probability. 1975, 3(4-6):643.
- [3] Galam S. Minority opinion spreading in random geometry[J]. The European Physical Journal B – Condensed Matter and

- Complex Systems. 2002, 25(4):403-406.
- [4] Sznajd-Weron K, Sznajd J. Opinion evolution in closed community[J]. International Journal of Modern Physics C. 2000, 11(6):1157-1165.
- [5] Deffuant G, Neau D, Amblard F, Weisbuch G. Mixing beliefs among interacting agents[J]. Advances in Complex Systems. 2001, 3:87-98.
- [6] Hegselmann R, Krause U. Opinion dynamics and bounded confidence models, analysis, and simulation[J]. Journal of Artificial Societies and Social Simulation. 2002, 5(3).
- [7] Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media[C]. // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10). Atlanta, USA, 2010:1361-1370.
- [8] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks[C]. // Proceedings of the 19th International Conference on World Wide Web (WWW'10). Raleigh, USA, 2010:641-650.
- [9] Chen W, Collins A, Cummings R, Ke T, Liu Z, Rincon D. Influence maximization in social networks when negative opinions may emerge and propagate[C]. // Proceedings of SIAM International Conference on Data Mining. 2011: 379-390.