**分类号**  **学号 M201676063**

**学校代码10487 密级**

****

**硕士学位论文**

**一种基于意见领袖博弈与记忆方式的舆情演化模型研究**

|  |  |  |
| --- | --- | --- |
| **学位申请人** | **：** | **叶玉腾** |
| **学科专业** | **：** | **软件工程** |
| **指导教师** | **：** | **沈 刚 教授** |
| **答辩日期** | **：** |  |

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**for the Degree for the Master of Engineering**

**Research of a Public Opinion Evolution Model Based on Opinion Leaders Through Game and Memory**

**Candidate : Ye YuTeng**

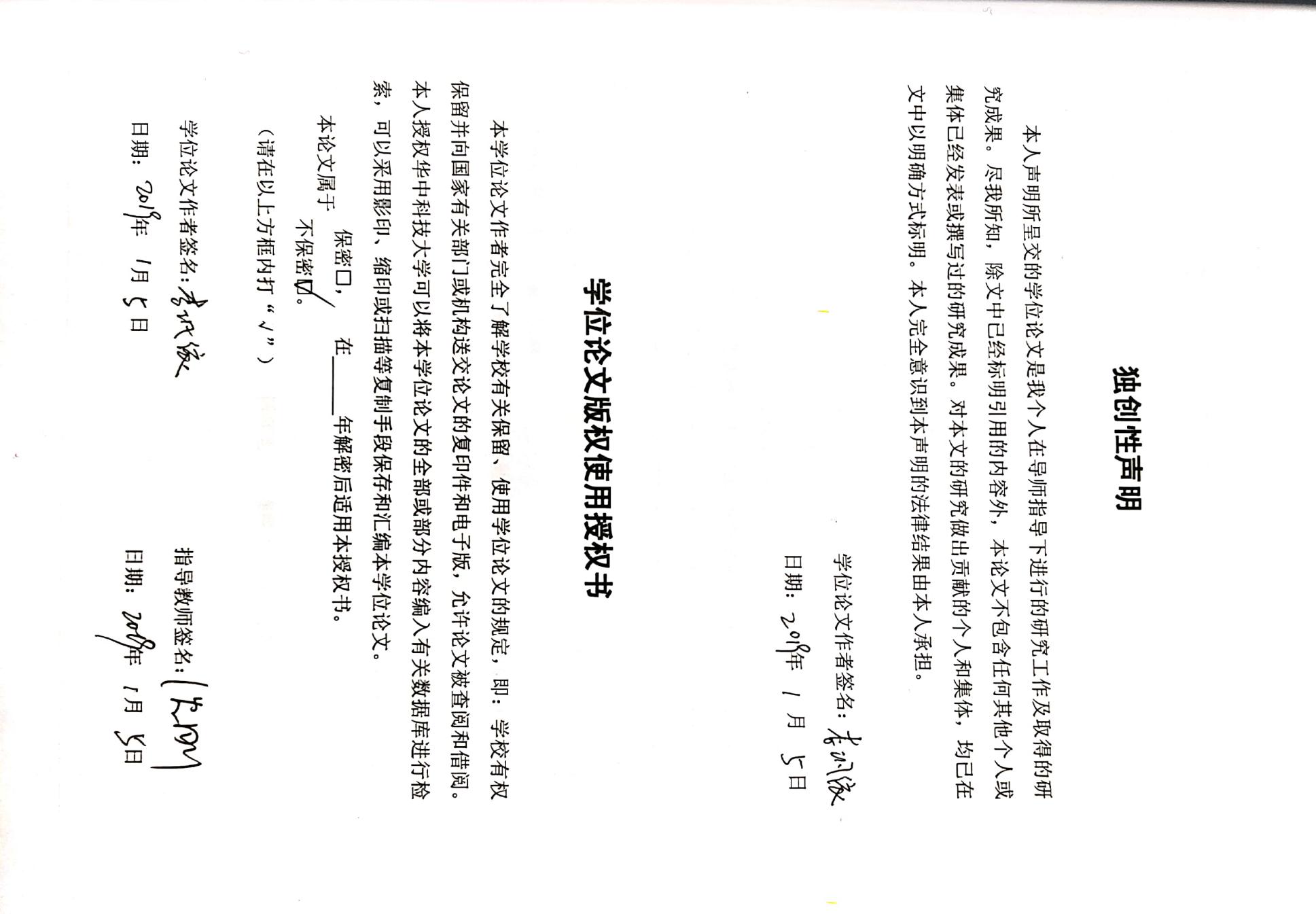
**Major : Software Engineering**

**Supervisor : Prof. Shen Gang**

**Huazhong University of Science and Technology**

**Wuhan 430074, P. R. China**

**April, 2019**



# 摘 要

舆情演化方式是舆情学的重要研究方向，意见领袖是社会舆情网络中影响力大的节点，在舆情演化过程中作用重要。但是，能较快达成舆情观点收敛的经典舆情动力学模型Hegselmann-Krause模型，其改进模型中融合意见领袖的方式是简单的加权意见值组合方式，忽略了意见领袖舆情驱动原因，也没有充分考虑舆情载体的网络拓扑结构。这导致现有Hegselmann-Krause模型及其改进模型的演化结果与真实舆情演化结果相差大。

基于上述现象，结合复杂网络，扩展Hegselmann-Krause模型，提出一种基于意见领袖的博弈与记忆方式的舆情动力学模型。在改进的复杂网络上，将意见领袖基于“囚徒博弈”原理做演化，普通网民基于HK模型做演化，再用记忆方式和节点亲密度方式将意见领袖观点引入普通网民意见变化中。博弈结果决定意见领袖演化方向与步长。博弈过程中，有界信任范围内的节点数直接影响博弈结果。针对经典BA无标度网络不可聚类特性，以不同概率引入了次领域与近领域连接，使改进网络同时具有小世界特性和无标度特性。并通过仿真，得到类似Twitter网络特性的参数。这个网络是舆论动力学模型的载体。

以微博关于“05.30头条腾讯大战”事件数据为训练集，使用遗传算法调节参数得出四个模型的合适参数。使用通过MATLAB仿真，得出舆情引导结论。研究表明：改变意见领袖发布负面意见的风险大小可以控制舆论走向。最后，通过微博关于“10.28重庆万州公交车坠江事件”事件数据为验证集，证明了改进模型更好拟合现实舆情数据。

**关键词：**复杂网络 意见领袖 博弈论 连续型舆情演化模型

# Abstract

The public opinion evolution is an important research direction of public opinion. Opinion leaders are the most influential nodes in the social opinion network, and play an important role in the evolution of public opinion. However, the classic opinion dynamics model Hegselmann- Krause, which can quickly reach the convergence of opinion views, the way to integrate opinion leaders in the improved model is a simple combination of weighted opinion values, ignoring the reasons for opinion leaders driving opinion, and no fully consider the network topology of the public opinion carrier. This leads to the difference between the evolution of the existing Hegselmann- Krause model and its improved model and the actual public opinion evolution.

Based on the above phenomena, combined with the complex network, the Hegselmann -Krause model is extended, and a lyric dynamic model based on the game and memory of opinion leaders is proposed. On the improved complex network, the opinion leaders are evolved based on the principle of “Prisoner Game”. The ordinary netizens make evolution based on the HK model, and then use the memory method and node intimacy to introduce the opinion leaders' opinions into the changes of ordinary netizens' opinions. The outcome of the game determines the direction and step size of the opinion leader. During the game, the number of nodes in the bounded trust directly affects the game outcome. Aiming at the non-clustering characteristics of the classical BA scale-free network, the sub-domain and the near-field connection are introduced with different probabilities, so that the improved network has both small world characteristics and scale-free characteristics. And through simulation, get the parameters of the network characteristics similar to Twitter. This network is the carrier of the paradox dynamics model.

Using Weibo's "05.30 Toutiao Tencent war" event data as the training set, using the genetic algorithm to adjust the parameters to obtain the appropriate parameters of the four models using MATLAB simulation, the results of the public opinion guidance show that: Manipulating the risk parameters of the opinions expressed by opinion leaders can control the direction of public opinion. Finally, through the microblogging modeling evolution of “10.28 Chongqing Wanzhou Gong Crashing River crash”, it is proved that the improved model better fits the actual public opinion data.

**Key words**：Complex network Opinion leader Game theory

Continuous evolution model of public opinion

# 目 录

[摘 要 I](#_Toc534792642)

[Abstract II](#_Toc534792643)

1 绪论

[1.1 研究背景与意义 (1](#_Toc534792646))

[1.2 国内外研究概况 (2](#_Toc534792647))

[1.3 本文研究内容 (5](#_Toc534792648))

2 相关工作

[2.1 传统谱聚类算法 (7](#_Toc534792650))

[2.2 大尺寸图像分割 (10](#_Toc534792651))

[2.3 本章小结 (14](#_Toc534792653))

[3 改进谱聚类](#_Toc534792654)算法

[3.1 递减距离影响的可聚类无标度网络 (15](#_Toc534792655))

[3.2 基于有界信任的意见领袖内部博弈演化模型 (19](#_Toc534792656))

[3.3 基于记忆的普通网民与意见领袖交互改进模型 (23](#_Toc534792657))

[3.4 本章小结 (27](#_Toc534792658))

[4 实验结果分析](#_Toc534792659)

[4.1 仿真数据实验结果 (28](#_Toc534792660))

[4.2 Berkeley dataset实验结果 (31](#_Toc534792661))

[4.3 参数敏感性分析 (36](#_Toc534792662))

[4.4 （其他结果得补充） (40](#_Toc534792663))

[4.5 本章小结 (46](#_Toc534792664))

[5 总结与展望](#_Toc534792665)

[5.1 全文总结 (47](#_Toc534792666))

[5.2 展望 (48](#_Toc534792667))

[致 谢 (49](#_Toc534792668))

[参考文献 (50](#_Toc534792669))

# 1 绪论

## 1.1 研究背景与意义

舆论是指民众在社会生活中对具体事件、人物等所持的观点态度的集合。早在1922年，WalterLippma[1]就撰写书籍《Public Opinion》，它被认为是舆论学的奠基之作。在我国，20 世纪 80 年代后，学者们才开始对舆论学的研究。而在一段时间里，在一定的社会空间内，围绕某社会事件的发展过程，民众对其中涉及的人物、机构等观点集合的整体状况被称为“舆论的情况”，简称为舆情。

截止2018年6月，中国网民数量达到8.02亿人，上半年新增网民数量为2968万人，互联网普及率为57.7%[2]。网络成了人们表达意见主要渠道之一。社交媒体对于突发事件、社会民生、环境保护等热点事件进行讨论，瞬间扩散能力强，许多热点事件传播以“亿”为单位，无论是数量还是传播范围、传播频率，均超过传统媒体[3]。在《2011中国危机管理年度报告》中，指出我国主要危机事件由网络首发的比例上升67%[4]。2018年5月，“空姐深夜滴滴打车遇害事件”经被害者好友微博传播引起广泛社会传播，最终演化成民众对顺风车的大声讨。2018年7月，“毒疫苗事件”由自媒体“紫竹张先生”发文披露，经微信朋友圈转发迅速发酵最终成为全民关注事件。2018年9月，央视《开学第一课》广告过多问题，经家长老师通过微信群、qq群、微博等渠道传播，得到社会广泛关注最终央视道歉。上述统计数据表明，互联网舆情已成为社会舆情的重要组成部分。

对于社交网络中舆情信息传播与观点演化还有助于了解具体社交网络中舆论的形成和发展及消亡，把握影响舆情发展的关键参数与节点有助于舆情学发展，有利于理解真实社交网络的拓扑结构，对复杂网络理论有一定意义。除此以外，舆情研究需要用到计算机仿真模拟舆情变化、自然语言处理获取基础数据等，理论研究可以加深研究工具的进步。可以说舆情学涉及到的方法论以及对复杂系统的深入研究将极大提高人们认知世界的能力，推动包括经济、信息和人工智能等诸多领域的发展[5]。

舆情学的研究范围广，包括舆论传播过程、舆情演化过程、热点舆情事件发现和意见领袖发现等方面。舆情演化过程主要指：在一个网络结构载体下，基于演化模型，在时间序列上对用户行为做出预测，从收敛时间、收敛速度、收敛簇数目、最终情感正负比等角度考察舆论的最终走向。舆情演化的模型又包括离散型舆情动力学模型、连续型舆情动力学模型、基于复杂网络的建模、基于博弈论思想的演化模型等。它们分别从不同方面考虑了舆情演化的过程。

基于有界信任思想的连续型舆情动力学模型体现了网民间“道不同不相为谋”的思想，表现出的网民交互方式与现实中人们意见变化方式相符。有界信任是指：个体i在下一时刻的意见值由网络中与i的观点差值在信任阈值内的用户的观点均值决定。连续型舆情动力学模型包括Deffaunt模型和Hegselmann-Krause模型。这两个舆情动力学模型认为网络中其他用户观念在一定阈值范围以内才会影响在下一时间序列时刻目标用户，我们可以认为这其实是一种类别划分。聚类方式就是观点阈值在观点相差度在一定阈值范围内。这种方式没有考虑用户群体的不同。虽然常常虽然Deffaunt模型存在用其它方式引入意见领袖观念的影响，但是Deffaunt模型个体没有充分考虑节点周围邻域，本身收敛较慢。而在收敛较快的Hegselmann-Krause模型中，普遍以加权值方式引入其他影响因素（如意见领袖、极端意见者、节点亲密关系）且其它影响因素值常常没有考虑网络拓扑结构特征。

这种方式忽略了意见领袖与普通网民舆情驱动原因的不同，意见领袖的演化常常受利益驱动，与普通网民自发受周边人影响不同。事实上，意见领袖作为舆情传播中的利益相关者和重复影响普通用户者，显然不能和普通用户使用同一模型。同时，目前舆情动力学模型与复杂网络常常分开研究，Hegselmann-Krause演化中并未用到网络拓扑结构的特征。为了解决上述问题，在特定网络拓扑结构下，需要改进Hegselmann-Krause模型，使其更好符合网络舆情演化现象，解释网络舆情演化规律。

## 1.2 国内外研究概况

舆情的研究涉及到多个方面，包含舆情载体-复杂网络、网络舆情信息挖掘和处理、和舆情传播演化等方面。其中网络拓扑结构是舆情传播的基础，舆情必须通过具体网络结构才能传播，复杂网络比简单的网络结构更符合现实网络特点。网络舆情信息挖掘是研究舆情传播的基础，将具体的舆论信息抽象成数字化可处理数据。网络舆情信息处理则是处理些如何确定不同群体，如何确定影响力较大的中心用户，不同族群用户交流等，涉及到文本聚类等技术手段。舆情传播演化则是指信息从少量用户扩散到其他用户的传播过程和其中舆情发生的变化。

对舆论传播演化的研究始于1948年，Lasswell在其著作中提出了的5W理论[6]。之后学者们对舆论动力学的建模方式主要分为自宏观到微观的建模方式和自微观到宏观的建模方式。传染病模型就是经典自宏观到微观的舆论传播模型[7]，它模拟了疾病传播的过程。1985年，Sudbury等人最先提出用传染病模型来进行谣言传播规律的研究[8]。Sudbury将传染病模型与网络拓扑结构结合，在均匀随机网络中进行试验，采用SIR模型，得出谣言传播的人群范围最多可达 80%的结论[9]。2012年，Xiong[10]在SIR模型基础上构建了 SCIR (C表示用户已接触到信息但未确定是否参与传播信息)，用于研究用户转发信息而导致信息扩散的过程。自微观到宏观的建模方法主要通过舆论参与者之间相互作用来考察舆情整体变化。舆情动力学模型以用户观点值能否为有限个分为两类。离散型舆情动力学模型用户仅包含有限个意见值，二项离散舆情动力学模型中用户意见值仅有正面意见观念和负面意见观念两个值，一般选为0或1。这类模型包括元细胞自动机模型、投票者模型、多数决定者模型等[11]。离散型演化模型研究较早，各方面改进模型比较成熟。离散型模型擅长模拟二值问题，但是忽略了现实中得到最终结果的过程。人们的观点值会随着时间变化在赞成与反对之间徘徊。连续型模型解决了这个问题，观点值可以是[0,1]区间内(0表示对观点持彻底否定态度，1表示对观点持肯定态度)任一值，这很好模拟了个体对一些敏感问题模糊不清的态度。连续型模型中，Deffuant模型基础模型简单，早期学者们对其研究更为充分。2013年后，对于Hegselmann-Krause模型的优化研究逐渐增多。

Guillaume Deffuant[12]于 2000 年提出在有界信任范围内建立Deffuant 连续型观点交互模型。有界信任的意思是指，仅有个体i与个体j的观点相差值在一定阈值范围内，个体j才能影响下一时刻个体i的观点。比如在二维网络内，有N个个体，随机选取其中两个个体i和j,他们对于某事件的观点值分别为和,设为个体受邻居影响的大小。若，则j在下一时刻影响用户i的观点。研究表明，收敛参数和系统规模只会影响最终态达到时间，而不会影响最终收敛结果。而信任阈值会影响最终收敛结果，当时，整个网络最终会收敛，否则会形成多个观点簇。

Deffuant模型提出时间更早，且更为简单，因此早期对Deffuant连续型舆情演化模型研究较多，成果比较丰富。2013年，张伟则提出认为收敛参数是不定的[13]，不同个体受邻居的影响程度不定，实验结果下，正态分布下的会比随机分布的更容易达到收敛。2014年，Shang等人[14]考察了非固定情况下的Deffuant模型，每次选取的阈值为[0,1]之间的随机数，仿真结果表明当阈值期望大于0.5，最终观点会在0.5处收敛。

Hegselmann-Krause模型是Hegselmann和Krause在2000年提出的演化模型[15]。在2013年后，学者们对其改进模型研究增多，目前仍旧有充分的改进空间。其与Deffuant模型主要区别是，更新时考虑所有邻居和个体观点在有限信任阈值内的网络节点的观点。HK模型不带记忆功能，用户i在t时刻的观点值x(t)无法影响x(t+1)。同时，由于HK模型是标准的时间序列模型，每次更新必须是同步的。

Jan Lorenz[16]在2015年时将基于交互式马尔科夫链的模型与传统HK模型做对比。在其马尔科夫模型中，将连续型的观点变成离散的观点，即n种观点，代表t时刻的观点分布。建模仿真后，得到交互式马尔科夫模型同时也符合传统HK模型的研究结果。2013年，Miguel Pineda[17]研究了HK模型中的噪声问题,每个个体都以一定的概率在全局范围或[]范围内选一个值为其下一个观点值，若不选才以hk模型为演化规则演化。实验结果表明，噪声模型在每个个体很多的簇和每个簇内个体数很低的极端情况下发生。簇的数量随着噪声增加而增加。在2018年，Su W,Chen G[18]在Miguel Pineda[17]的基础上引入噪声方案来改进hk模型。张亚楠[19]则在2017年时引入在hk模型中引入节点亲密度关系。

随着舆情动力学模型的发展，各种各样的影响因子（意见领袖、极端意见者等）被以各种方式引入上述经典舆情动力学模型中。2010年，Kitsak[20]提出用k-shell分解算法确定意见领袖比用网络中影响力最大节点确定意见领袖更准确。2014年，黄庆花[21]在Deffaunt模型中引入节点记忆模型。2016年，胡珑瑛[22]在研究政府与网民博弈中，引入博弈行为判断下一步舆情走向，但是没有引入经典舆情动力学模型。2017年，彭小兵[23]则基于SIR模型引入网络媒体、意见领袖等变量来探究邻壁效应。2018年，曾子明则引入BP神经网络处理感染节点的传染影响度[24]，区别了中心节点和普通节点在传染病模型中传染的区别。Zhao Y[25]等人在2016年通过区分意见领袖并将其以非普通民众权值方式引入hk迭代模型中。

随着复杂网络理论的发展，将复杂网络与舆情动力学模型相结合成了一种新的研究趋势。2009年，Albert[26]提出将复杂网网络与系统动力学结合，其成果可以应用于社交舆情、交通治理等问题上。2012年，Xiong[10]从 BA 无标度网络与规则网络两个网络结构上模拟仿真出SCIR模型。2018年，马永军[27]等基于Deffuant模型，构建加权网络拓扑结构,，研究群体中节点平均度对演化结果影响。

随着机器学习的发展，BP神经网络和常见优化参数等也被用于舆情领域做参数调节。游丹丹[28]则在2016年引入改进粒子群和BP神经网络方式用于网络舆情预测。黄亚驹和陈福集[29]等则在2018年将混合算法、BP神经网络和遗传算法用于舆情领域。

## 1.3 本文研究内容

基于网络舆情在公众中影响日益增加这一事实，针对舆情发生时政府、公司等应采取合适策略来引导舆情。意见领袖是社会舆情网络中影响力大的节点，在社交网络中作用重要，但是充分考虑节点邻域特征Hegselmann-Krause的舆情演化模型中融合意见领袖的方式是简单加权。忽略了意见领袖与普通网民舆情驱动原因的区别，也没有充分考虑网络的本身结构。针对上述问题，论文研究工作如下：

（1）通过阅读文献等方式对舆情演化过程做详细了解，对复杂网络相关模型、舆情动力学相关模型、博弈论相关理论等优缺点进行归纳总结，得出研究方向。

（2）舆情的演化离不开传播的节点网络，优化舆情承载结构是本文的研究重点之一。分析在线社交网络的特征结构，融合其他学科思想，将近领域概念引入BA无标度网络中并将次领域与近领域以一定比例融合来构造可聚类的无标度在线网络。

（3）选取经典连续舆情动力学HK模型，充分考虑网络拓扑结构，将网络人群分为意见领袖和普通网民两种身份。以经典“囚徒”博弈方式引入意见领袖对HK模型演化思想，并考虑到舆情演化过程中的记忆效应，提出改进普通网民与意见领袖交互方式的HK演化模型。

（4）采用python脚本语言和Matlab建模工具对社交网络的舆论形成、演化过程做了建模仿真，进而提出与现实相符的舆论动力学模型。通过MATLAB仿真，修改模型参数得到不同的收敛结果，得出舆情演化中关键节点和建议措施。以关于“05.30头条腾讯大战”事件的微博数据为原始数据，得到模型参数。然后以“10.28重庆公交坠江” 事故的微博数据做验证集，证明改进模型能比较好的拟合实际情况。

# 2 基本理论

本章主要介绍后文改进模型与调节参数时候所用的理论技术和方法。先介绍了网络评价指标和现有网络拓扑结构，基于此得出后文改进网络的评价方向与方式。最后介绍了基础的博弈理论，重点介绍囚徒博弈的方法。

## 2.1 网络特征指标

网络是表示诸多对象及其相互关系的模型，它一般由节点和节点之间的连线构成。在现实生活中，它可以表示不同个体相互连接和作用，因此大量复杂系统都可以通过各种网络结构加以描述。其中，复杂网络是随着图论兴起而兴起的一门学科。图论可以追溯到1736的经典“七桥”问题，欧拉解决该问题的同时提出了图论这一模型来拟合现实中的节点和节点的关系。1950年Erdos和Renyi提出随机图论后，图论开始快速发展[30]。通过大量实验，学者发现真实社会网络往往不是规则网络也不是随机网络，因此提出了一种新概念为复杂网络。**近几年，学者们对**复杂网络的理论及其应用研究不断深入。复杂网络是具有自组织性、小世界性、无标度性等特性的网络，它被用于研究复杂系统[31]。研究复杂网络的主要工具仍旧是欧拉所创立的图论，学者们将现实中的复杂系统的节点抽象成图中节点，将现实节点的关系抽象成图中的边。人们使用复杂网络相关理论来探寻现实世界中各个复杂系统的规律，例如社交网络、交通网络等。

复杂网络关注的点包括网络节点本身、节点与节点之间的相互关系、网络的整体结构方面等[31]。对于任意网络，可以用图G=(V,E)表示，其中V为网络点集和与E为网络边的集合。E中的每条边都对应着V中的两个节点。本小节主要从复杂网络的统计特性和复杂网络拓扑结构来介绍复杂网络。

评价网络性能要通过具体的评价指标，本小节主要介绍当前热门的网络评价指标。

（1）路径长度

对于图G中任意两个节点i和j,可以得到许多边集，使i可以通过这些路径连通到j，我们将这些边集称为i到j的路径，对应边集的大小称为路径长度。其中最小的路径长度称为节点i到节点j的距离，定义这个值为。

点集V中的任意两点之间距离的平均值是网络平均路径长度，记为L，如式2-1所示。

(2-1)

网络直径是指网络最大路径长度，也是衡量一个网络规模的因素，它代表着网络中任意用户能联系到另一用户的最大值，记为D，如式2-2所示。

(2-2)

经过学者研究发现，真实网络普遍网络直径和网络平均路径长度都比较小。这种现象符合心理学家Stanley Milgram[32]提出的“六度分离”理论：任意两个陌生人可以通过多个自己的熟人建立联系，而熟人的个数不超过五个。

（2）介数

介数指经过计算指标的最短路径与网络中所有最短路径的比例，分为边介数和节点介数。计算指标指需要计算的点或边。介数可以体现出计算指标在网络中的作用。

假设网络中一共有m条最短路径，对于任意节点i和边e,有条经过节点i，有条经过边e。则节点i的介数为，节点e的介数为。如果一个成员i位于其他成员的多条最短路径上，那么节点i在网络中的作用就比较大，也称为节点i具有较大的介数中心性。若节点或边的介性中心度高，表明节点或边在网络中地位重要，一旦去除对网络中其它节点或边造成的影响大，很多其它点或边之间的最短路径会被改变。

（3）度分布

在复杂网络中，度是一个非常重要的统计概念，它反映了网络的结构特征。与节点i之间直接有边相连的节点称为i的邻居，节点i的度是指与它连接的边的总数，它的度数与它的邻居数相等，体现了该节点i的重要程度。对于有向网络中，又分为入度和出度，比如社交网络中用户关注数称为出度数，其粉丝数称为入度数。本文主要考虑无向网络，因此主要介绍无向网络度分布。

任意非负整数k，度数是k的节点在所有节点中的比例为，如式2-3所示。定义du为节点k的度数，n为网络中所有节点的总数。也是从网络中随机选取一个节点，其恰好度数为k的概率。

(2-3)

假设网络中节点度数最大值为,则显然有。计网络中节点的平均度为<K>,则其表达式如式2-4所示。

(2-4)

度分布是指对于不同，的分布函数。大量研究表明，真实的社交网络的度分布大多满足幂率分布[33]，即满足。

（4）聚类系数

聚类系数代表了一个网络中节点之间的聚集程度，即其邻居的邻居也互相是邻居的比例。对于网络中任意节点i，其度数为-1，即其有个邻居。那么这个节点之间实际的边与个节点是全连接时的边的比例为节点i的聚类系数，反应了i的紧密程度。定义这k个节点之间实际边数为，则i聚类系数如式2-5所示。

(2-5)

聚类系数高代表着节点之间连接紧密，是小世界网络的显著特性之一。大量研究结果表明，现实的社交网络都具有高聚类系数。任意选取两个节点，它们与i之间两两连接，则称此三节点组成了一个与i连接的三角形；任意选取两个节点，其分别于i相邻，则称此三节点组成了一个与i连接的三元组。从拓扑结构来看，式2-5等价于：

记网络聚类系数为C,它是所有节点的聚类系数的平均值，则其表达式如式2-6所示。

(2-6)

## 2.2 网络拓扑结构

（1）规则网络

规则网络是指节点按照固定规则连接的一种网络结构，它是一种简单网络，其节点的度相同。从网络拓扑结构来看，规则网络的度分布是δ 分布。由于规则网络比较简单，早期学者多在规则网络上运行Ising 模型等简单舆情演化模型。规则网络每个节点与其它节点连接的方式相同，包括全局耦合网络、最近邻耦合网络和星形网络等[31]。有20个网络节点的全局耦合网络的图像如图2.1所示。

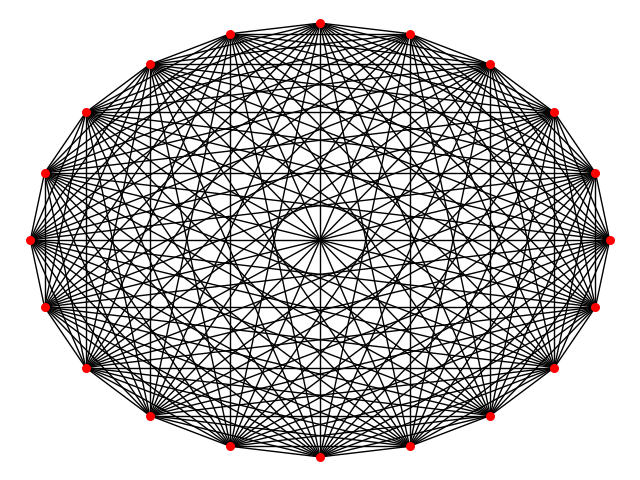


图2-1 规则网络拓扑图

（2）随机网络

随机网络是图论与概率论的交叉学科，是指通过各个节点之间生成边的过程带有随机性的网络。由N个顶点构成的无向图中,最多可能有N\*(N-1)条边，从这N\*(N-1)条边中随机选取任意M条边做连接，由这N个顶点和M条边所构成的网络就叫随机网络，也称为随机图。

Erdös P和Rényi出的ER网络[30]是一种经典的随机网络。假设网络中存在n个节点，任意两节点之间都以常数的可能性相连，网络的度分布满足Poission分布，平均度 ，它的聚类系数C=p，。全连接网络在现实中比较少，因此用随机网络模拟实际网络时，概率p不大，则其聚类系数就比较小。现实中的社会网络的聚类系数大多高于同规模的随机网络，因此随机网络不能很好模拟现实网络。用python库networkx生成有20个节点，连接概率为0.2的随机网络如图2-2所示。

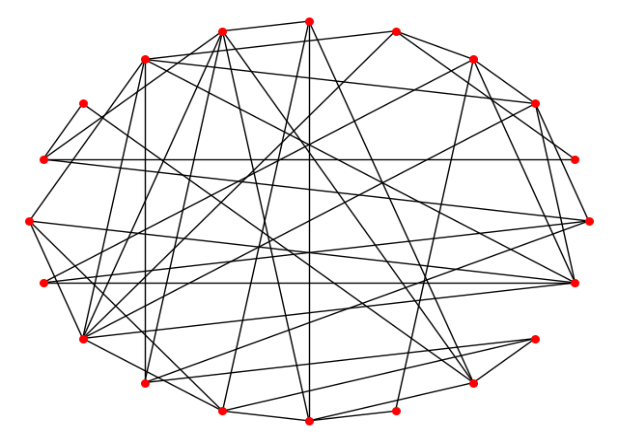


图2-2 N=20，p=0.2的随机网络

（3）小世界网络

在图论中，小世界网络是网络中任意两点的距离远远小于网络规模的网络，它具有很小的平均路径长度和很高的聚类系数，这种网络的特性称为小世界特性。“六度分割”现象就是小世界理论的体现，它是复杂网络的一种。它的生成边方式介于规则网络和随机网络之间。由Watts和Strogatz[34] 提出的WS 小世界网络是经典小世界网络模型，它在规则网络基础上做了部分边的随机重连，引入了随机性。WS 小世界网络具有较明显的高聚类特性和较小的路径长度。WS小世界网络构造步骤是将有N个节点的最近邻耦合网络的边以概率p进行断边重连其他任意节点。用python库networkx生成有20个节点，每个节点有4个邻居，重连概率为0.3的WS小世界网络如图2-3所示。

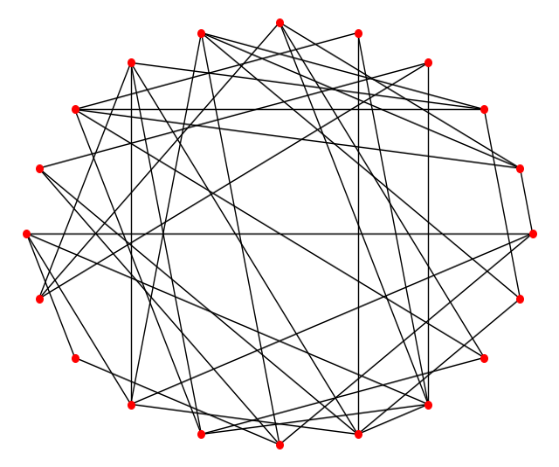


图2-3 ws小世界网络

（4）无标度网络

20世纪末，Albert-László与Réka Albert等人[35]研究发现：绝大多数网页中超链接的个数少于4个，但少数的网页具有的超链接个数超过1000。这表明真实网络中大部分的节点的度并不是相差无几，而是少部分节点有很大的度，大部分节点的度很小。在社交网络中，新加入网络的节点总是优先偏向于与影响力大的节点相连，这个现象符合“马太效应”。因此，规则网络、随机网络和小世界网络由于度是均匀分布的，并不能很好符合这个特点。这种网络中节点的度分布不均衡，少数节点度数很大，大部分节点度数很小的特性被称为无标度特性。

其中BA无标度网络是无标度网络中的重要模型。Barabási 和 Albret [36]指出：规则网络、随机网络和小世界网络都忽略了实际网络中少量节点与网络中其它大量节点相连的特性。基于上述特性，他们将每次新加入节点i与老节点j之间以p=的概率建立连接。这样加入了N个节点后，BA无标度网络的度是数k的概率如公式2-7所示，显然该分布符合幂率分布。

(2-7)

BA无标度网络的聚类系数较大。因此如何得到聚类系数小且有无标度特性的类似真实网络的网络是本文改进重点之一。

则上述四种网络的网络特性评价指标如表2-1所示。

表2-1 经典网络拓扑结构指标

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 规则网络 | 随机网络 | WS小世界网络 | BA无标度网络 |
| 度分布 | δ 分布 | Poisson分布 | 均匀分布 | 幂率分布 |
| 聚类程度 | 不定 | 不定 | 高 | 低 |
| 平均路径长度 | 不定 | 不定 | 低 | 低 |

## 2.3 博弈相关理论

博弈论中，多人在对局中权衡利益得失采取不同策略，达到取胜的目的[37]。早在古代，《孙子兵法》等著作中就体现了博弈论的思想。博弈论中的均衡是指博弈的结果处于一种稳定的状态[38]，参与博弈的主体都对博弈结果表示满意不再更改自身博弈策略。博弈论的基本概念包括博弈主体局中人、博弈策略、局中人行为和博弈过程中各方收益等。博弈过程中主要要素有博弈主体、博弈策略和收益。局中人通过衡量自身收益做出理性的选择策略以实现利益均衡。

由Flood和Dresher提出的囚徒困境博弈是博弈论中的经典模型之一[39]。从行为的时间序列性上看，囚徒博弈属于静态博弈：在博弈中，多个参与人是否同时选择下一步的行为对博弈结果无影响，但稍后做出行动的参与者不知道先行动者的行为。同时，囚徒博弈还是一种经典非零和博弈，即所有参与者收益之和不一定为零，所得方所得不一定都是所失方所失。这表明个人选择与团体最佳选择无关。

囚徒困境故事指：两名罪犯在犯罪后被警察抓获，他们被关在不同地方，无法交流。现有证据无法使警察确认两人的罪责。博弈主体是两个囚徒，博弈主体能选择的博弈策略有坦白或抵赖，博弈中的收益是判刑年数。不同策略会影响博弈主体的利益：如果两人选择抵赖，则会判刑一年；如果两人都坦白，则各判刑八年；如果两人中一人坦白而另一人抵赖，坦白的放出去，抵赖的判十年。显然，任意一个博弈主体的最佳博弈策略都是坦白：同伙抵赖的情况下，自己抵赖会判一年，坦白则被释放，坦白比不坦白的收益更高；同伙坦白的情况下，自己抵赖会判十年，坦白则判八年，坦白行为带来的收益仍旧优于抵赖行为带来的收益。但是使群体收益最高的行为是双方都采取抵赖的策略。如果罪犯都选择坦白，各判刑八年。如果两人都抵赖，各判一年，显然都抵赖结果(合作结果)是群体最优解。对于囚徒博弈模型，我们可以设参与集合I中元素为A和B，策略集合S中元素为合作与背叛。双方参与博弈，当博弈主体选择合作时，二者得到收益为R；一方选择合作另一方选择背叛时，合作方收益为S，背叛方收益为T；双方都背叛时收益为P。则囚徒博弈矩阵如表2-2所示。

表2-2 囚徒博弈策略及收益

|  |  |  |
| --- | --- | --- |
|  | 合作 | 背叛 |
| 合作 | （R，R） | （S，T） |
| 背叛 | （T，S） | （P，P） |

根据上面的囚徒博弈故事，显示A与B都合作时收益是群体最大；但是一人合作一人背叛时，背叛者可以达到个人收益最大，但是合作者会付出较大代价；当A与B都选择背叛时，群体利益是最低的，但是个人损失会小于一人合作一人背叛时合作者代价。因此参数满足下式。

(2-8)

但是囚徒博弈是非零和博弈，因此参数的关系不是固定的，参数大小应根据实际情况调整。

演化博弈是1973年由J.Maynard.Smith和G.R.Price在生物进化中基因演化的基础上提出的演化稳定策略[40]。它与传统博弈模型区别是，参与者不是完全理性并且也不完全了解信息。这个条件的加入突破了传统博弈的一些限制，外在因素也可以影响最终结果。它认为，人类通常是不断试错来调整自己的博弈策略，最终的达到博弈均衡[41]。因而历史、制度因素以及均衡过程的某些细节均会对博弈的多重均衡的选择产生影响。

## 2.4 本章小结

本章主要对复杂网络理论、博弈理论做了介绍。先介绍了当前网络评价指标，衡量一个网络结构应当从网络度分布、聚类系数、平均路径长度等入手，衡量一个节点或一条边应从其介数等入手。然后介绍了网络拓扑结构的发展过程，阐明了规则网络、随机网络、小世界网络和无标度网络等特点。最后，介绍了博弈论相关知识，介绍了囚徒博弈和演化博弈两种博弈思路。为下文改进基于网络拓扑结构和基于意见领袖的改进连续型舆情演化模型提供了理论基础。

# 3 改进演化模型

本章尝试在特点网络拓扑结构上改进经典舆情演化HK模型。考虑到意见领袖是社会舆情网络中影响力大的节点，在社交网络中起重要作用。但是，目前收敛时间较快的Hegselmann-Krause的舆情演化模型中融合意见领袖的方式是简单加权。忽略了意见领袖与普通网民舆情驱动原因的区别，也没有考虑复杂网络本身的结构。考虑意见领袖发布意见经常带有经济效应和利益关系，意见领袖观点演化采用囚徒博弈思路。这样虽改进了意见领袖观点，但是没有使意见领袖影响所有网民，无法体验意见领袖的作用。因此用记忆方式和基于拓扑结构的权重值将意见领袖意见影响于普通网民。选取特点网络拓扑结构为BA无标度网络，引入次领域与近领域加权方式使网络具有可聚类特性。本文基于传统BA无标度网络，改进生成边算法，引入部分小世界网络的高聚类特性。

## 3.1 递减距离影响的可聚类无标度网络

章节2.2中所示的四种经典网络拓扑结构都不能很好拟合真实社交网络。基于2.1中网络衡量指标，由斯坦福大学公开的社交网络数据集可以得到主流社交网络(圈子)指标如表3-1所示。

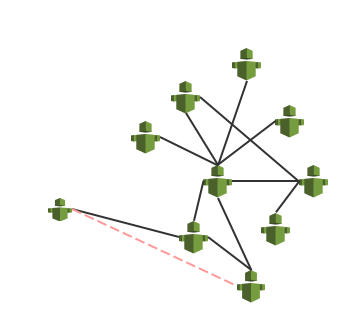
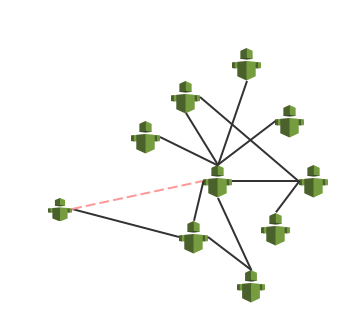
表3-1 社交网络指标

|  |  |  |  |
| --- | --- | --- | --- |
| 指标 | Twitter | Facebook | Google+ |
| 节点数 | 81306 | 4039 | 107614 |
| 边数 | 1768149 | 88234 | 13673453 |
| 网络直径 | 7 | 8 | 6 |
| 网络平均路径长度 | 4.5 | 4.7 | 3 |
| 网络聚类系数 | 0.5653 | 0.6055 | 0.4901 |

真实世界的社交网络大多同时具有小世界特性和无标度特性。如第二章复杂网络中所介绍，BA无标度网络和WS小世界网络是当前研究复杂网络的重要模型。但是它们与真实世界复杂网络相比都有一定区别，BA无标度网络和WS小世界网络分别倾向于考虑了复杂网络的度分布特性和小世界特性。

为了在类似真实社交网络的网络结构上演化模型，我们需要选取一个同时具有无标度和小聚类特性的网络。本文选取的BA无标度网络本身满足幂等分布，满足无标度特性，因此主要任务是改进新加入网络节点的生成边算法，引入高聚类特性。

一般而言，生成边的连接主要有优先连接和邻域连接两种。优先连接指节点生成边时，会优先与网络中影响力大的节点建立连接，与网络中任一节点连接的概率正比这个节点的度占网络中总边数的比例。领域连接指节点生成边时，会优先与其距离近的节点生成边。使用优先连接方式构建的网络的度分布具有幂律分布的特性。使用邻域连接方式构建的网络具有小世界特性。则二种方式生成边如图3-1所示，图（a）为以优先连接方式生成边，图（b）为以邻域连接方式生成边。



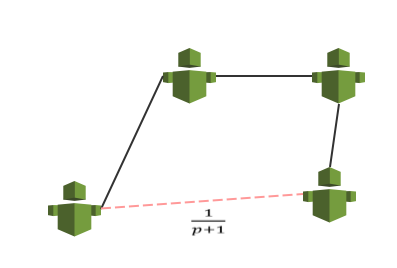
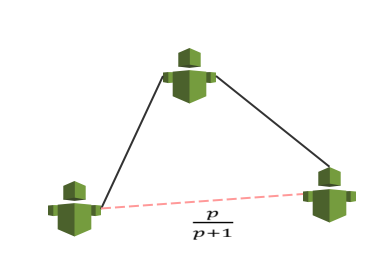
（a）优先连接 （b）邻域连接

图3-1 领域连接与优先连接

每次新加入的网络节点时，按照BA无标度网络生成算法生成新的边后可以再按照节点连接性以一定概率加入新边。如果A同时关注B和C，那么B和C也互相的概率是很大的。此时B与C以一定概率生成新边，则可使网络结构具备高聚类特点。Holeme&Kim提出的三元无标度网络构建算法[42]就是基于上述思路，引入三角连接策略来提高无标度网络的聚类系数。

现实社交网络中仅仅考虑三角连接策略是不够的，人们会以一定的概率关注朋友的朋友的朋友，一般而言超出三层关系的继续关注的概率会小很多。Nicholas A.Christakis曾提出“三度影响力”理论，该理论指出：影响力能够穿透三度人脉，你的朋友、你的朋友的朋友、你的朋友的朋友的朋友[43]。在社交网络上，由于具有某个朋友，社交网络推荐给用户这个朋友的朋友是正常现象，用户关注了这个朋友后，自然会显示这个朋友的朋友的粉丝数，而用户可能以较低概率关注其朋友的朋友的朋友。我们受三度内朋友的影响，也影响着三度内的朋友，且影响概率和程度虽度数递增而递减。超出三度范围内的人对用户的影响程度基本是由于传递信息本身所带来，而不是因为用户之间联系所带来。

基于上述分析，我们可以将“三度影响力”理论邻域连接方式中。若A与B相邻，则A优先与B的邻域C建立连接称为次领域连接。若A与C建立连接后又与C的领域D进行了连接，称为近领域连接。因此，设近领域与次领域连接的比值为p。三度内，影响力递减；采用次领域连接的生成边必然多余近领域连接的生成边。因此,p小于1。则按照“三度影响力”理论进行领域连接如图3-2所示，其中图(a)表示按次邻域进行连接，图(b)表示按照近邻域进行连接。



（a）次领域连接 （b）近领域连接

图3-2 邻域连接方式

为了使网络同时具有小世界特性和无标度特性，构建网络的过程中，生成边算法融合优先连接和无标度连接。改进的网络模型，按照pt的概率按原始BA无标度生成边方式对新加入节点采用优先连接方式生成边，按照1-pt的概率按照邻域连接的方式生成边。按照上节分析，不同度的对象影响力肯定是不同的。在领域连接中，设置次领域与近领域连接的生成边比例为p。

改进后的无标度网络构造算法如表3-2所示：

表3-2 递减距离影响的可聚类无标度网络

|  |  |
| --- | --- |
| Step1 | 初始化：假设该网络具有n个节点,生成一个包括m0个节点的全连接图; |
| Step2 | 从n个节点中选择未被加入网络的节点i,将i加入网络中,新加入的i将生成m条边（m<m0）； |
| Step3 | While （当前网络中有节点未遍历）：{  选取当前网络未遍历节点j;  考虑以pt 概率进行优先连接：节点i与节点j以概率建立连接关系。其中代表j节点的度，为当前所有网络中节点度之和;  考虑以1-pt 概率进行三角连接：节点i以(1-pt)的概率与j的领域做三角连接，以(1-pt)的概率对j的领域的领域做次近邻连接。  } |
| Step4 | 转Step2直至网络中有n个节点。 |
| Step5 | 结束网络构建。 |

选取m=5,即新加入节点的平均度为5。这主要是因为snap数据集公布的Twitter圈子数据中90%的节点有效直径为4.5。则，采用python库NetworkX生成N=200，每个新加入节点的加入边m=5进行作图。原始BA无标度网络的网络结构如图3-3所示。

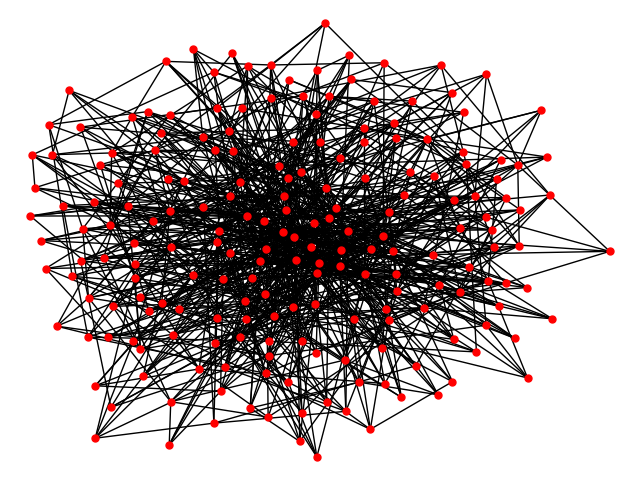


图3-3 BA无标度网络

图3-1中，原始BA无标度网络生成的N=200，m=5生成的网络的平均聚类系数为0.124。选取pt =0.7进行优先连接，选取p=0.35进行领域连接。改进后的无标度算法得到的平均聚类系数为0.4468。而从斯坦福大学公开snap数据源上，Twitter 社交圈的平均聚类系数为0.5653。可以看出改进结果更符合实际情况。改进三度可聚类无标度网络结构如图3-4所示。

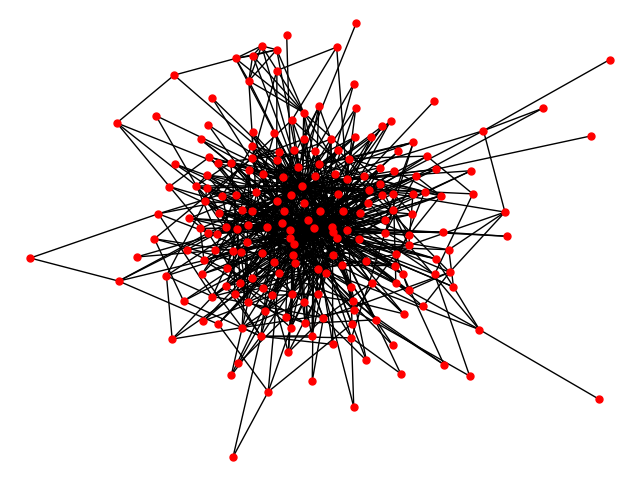


图3-4 递减距离影响的可聚类无标度网络

在使用Python实现过程中，新算法的实现改进了networkX库中优先连接的算法，库中使用将每个节点按度数大小重复添加数组中，然后从数组中随机抽取该网络期望节点度数个节点作为下一个加入网络节点的邻居。这一方法虽然考虑了优先连接性，但是每个节点的邻居都是固定的，忽略了节点度数的随机性。

## 3.2 基于有界信任的意见领袖内部博弈演化模型

本节在3.2的改进递减距离影响的可聚类无标度网络上进行舆情的演化改进。离散型舆情演化模型过于简单，不适合应用于观点多样的社会问题上。采用基于有界信任的连续型舆情演化模型是当前研究的主流方向。

Hegselmann-Krause的演化模型是经典连续型舆情演化模型，个体i在下一时刻的意见值由网络中与i的观点差值在信任阈值内的用户的观点均值决定。传统的HK模型主要是考虑取有限信任阈值内的网络个体的均值，类似投票者模型的思想。网络中每个节点高度同质化，没有充分考虑网络拓扑结构特点。传统HK模型中任意个体i在下一时刻的演化公式如式3-1所示。

(3-1)

目前各个改进模型大多数是在上述基础上将这些因素（如意见领袖、观点距离值）以加权值方式引入计算模型，如式3-2。

(3-2)

公式3-1中，和代表某一改进影响因子对网络中不同用户的影响参数的向量，除和以外还可以加入其它影响因素。

Hegselmann-Krause的改进模型普遍以加权值方式引入其他影响因素（如意见领袖、极端意见者）。这种方式忽略了意见领袖与普通网民舆情驱动原因的不同，意见领袖的演化常常受利益驱动，与普通网民自发受周边人影响不同。意见领袖跟随者众多，且网络发声会给他们带来直接的利益效应。普通群众对热点事件普遍持“吃瓜”态度，会按自身意见发表观点。但是意见领袖发布的观点往往涉及他们的网络关注度（更多用户会关注他们，从而可能带来经济收益）、社会地位和利益群体。因此，按照普通群众的演化规则来演化意见领袖是不合适的。意见领袖的观点变化涉及到各方利益群体的博弈，会按照他们的综合利益最大化来进行演化。

现实情况下，对于热点事件，很多大v（关注度很高的用户）往往会对事件发表看法，他们的看法会被个体用户大量转发评论，拥有较大影响力。因此改进模型主要从共同好友也会影响用户看法和意见领袖对群体的影响两方面入手。除此以外，共同好友较多个体可能会认识存在线下交流，即使观点差别较大，仍会影响个体行为。

传播学认为，信息的传播方式是从意见领袖传播向受众。可以将网络中所有个体分为意见领袖和普通网民两种群体。其中意见领袖指社交网络（如微博）获得关注度高的用户，诸如大v、网红用户、热点事件爆出用户。普通群众是指网络中普通用户。

意见领袖对应着网络中影响力大的节点，是能左右多数人态度倾向的少数人。目前对于意见领袖的选取有多种方式。下面列举三种常见选取方法。

（1）直接选取网络中度较大的节点。

（2）特征向量中心性算法。即与高度数相邻的节点的中心性也很大。

（3）k-shell分解算法得到影响力最大的节点。先找出所有度为1的节点，将其删除。然后在剩下的节点中继续查找度为1的节点，并删除。直到网络中没有度为1的节点，将之前删除的所有度为1的节点shell值赋值为1。再用同样方法处理度为2的节点，设置shell值为2。重复上述步骤，直到所有节点都有shell值。选出shell值最大点为影响力大的节点。2010年，Kitsak[20]提出用k-shell分解算法确定意见领袖比用网络中影响力最大节点确定意见领袖更准确。

本文的研究重点不在于如何得到关注度高的节点，因此简化意见领袖的发现这一问题如定义1。

定义1：意见领袖指网络中度最多节点，表示被网络中更多节点关注。因此，网络中度数前5%的节点为意见领袖。

引入“囚徒博弈”思想来进行意见领袖的演化，这是网民内部的双方博弈。针对热点事件，意见领袖可以采取积极或者消极的意见去评论事件；普通网民可以采取被影响和不被影响两方面来应对。虽然HK模型的基础是连续型舆情演化模型，但是现实情况下意见领袖的初始意见表达往往具有极端性，意见领袖观点普遍具有脸谱性。当意见领袖意见大于0.5时，我们认为他的意见是支持；否则持反对意见。

当意见领袖发表肯定观点时会给他带来网民关注度，此时定义收益为U；当意见领袖发表否定观点时也会带来的网民关注度，定义这个收益为V。由于网民在社交网络上表达的观点消极性，负面观点更容易被接受，因此发表负面观点会为用户带来额外的效应V1()。当意见领袖i发表观点与相应网民观点不同时，会带来掉粉被喷现象，对意见领袖带来一定的损失，定义这个损失为M1。当意见领袖发表不符当局认知的消极观点时，可能会给意见领袖带来线上封号或者线下被控制的风险损失，定义这个损失为V2。因此当意见领袖发表观点普通民众采纳时，意见领袖收益为U，不采纳时收益为采纳时收益减去“掉粉”风险值。意见领袖发表负面观点被采纳时，他的收益为关注收益U与“喷子”文化带来的收益加成与政府可能管控的风险加成。意见领袖发表负面观点且不被普通网民接纳时，其受益仅仅有因为发布观点带来的关注收益与政府可能管控的风险。与普通网民的观点差值在信任阈值范围内时，普通网民会采纳意见领袖的观点，否则会不采纳。详细博弈矩阵见表3-3。

表3-3 意见领袖与普通民众博弈矩阵

|  |  |  |
| --- | --- | --- |
| 普通网民  意见领袖 | 采纳 | 不采纳 |
| 积极肯定 |  |  |
| 消极否定 |  |  |

在这段博弈中，设置下一时刻意见领袖选择发布积极肯定的帖子概率为，选择发布消极否定的帖子概率为。网民选择是否采纳意见领袖观点仍由二者观点差值决定。初始时，参与事件的意见领袖一般会对热点事件有观点鲜明的表现。不妨设普通网民中会心信任意见领袖的概率为y，本模型基于有界信任这一概念，因此确定t时刻意见领袖i的观点值和t时刻其他网民的观点值下，y是不变值，用N表示网络总规模，如式3-3所示。

(3-3)

用表示意见领袖发布积极肯定的观点时的收益期望，表示意见领袖发布消极否定观点时的收益期望，为意见领袖的总收益期望。由于在时刻t不变，因此下面公式中用y代替t时刻的。

(3-4)

(3-5)

(3-6)

令表示意见领袖选择策略的函数，其自变量为。公式如3-7所示。

(3-7)

则分析式3-7，当时：

（1）当 时，无论取何值，。因此，此时无论意见领袖意见如何演化，都不会影响其下一时刻的收益，此时选择意见领袖意见不变。

（2）当 时，令,此时分两种情况：当y <时，， 恒成立， =1是平衡点，即下一时刻意见领袖应向正面观点演化；反正，当y > 时，， 恒成立， = 0是平衡点，即下一时刻意见领袖应向负面观点演化。

在博弈型意见领袖观点演化中，y是一个固定值不是变量。这也可以为政府或者社交网络提供平台（如新浪微博公司）提供管控意见。当管控方希望舆情向正面观点方向演化时，普通网民基数大并不好直接控制普通网民，因此少量的意见领袖可以成为管控方向。提高意见领袖提负面观点的风险值，即将值增大可以有效促进意见向正面方向演化。

由于本模型是连续型舆情演化模型，因此上述结论仅仅得到意见领袖演化方向。计v=，则意见领袖下一时刻最终演化意见如式3-8所示。

(3-8)

## 3.3 基于记忆的普通网民与意见领袖交互改进模型

在网络舆情传播中，普通网民一般倾向于查看与自己类似的观点，以期获得认同感，HK模型正是基于这个基础。但是，意见领袖由于是大V用户，普通用户会通过各种渠道（网络平台主动推送或线下与人讨论）获得大V用户观点。当观点与普通网民观点相差过大时，普通网民不会信任这个观点，但是观点会在普通用户的记忆中发酵。当意见领袖影响力非常大，即使用户不同意该观点，仍旧会被动反复看到意见领袖观点，记忆中观点沉淀值达到一定程度，则该观点可能对普通用户产生影响。这种记忆效应对于普通网民内部舆情演化影响比较小，因为大部分网民不会主动去获取与自己观点相差较大观点。

针对上述现象，本模型针对意见领袖对普通网民的影响方式进行了改进。在普通网民j的观点演化中，若意见领袖i观点值与其观点值在信任阈值内，则意见领袖与其他普通网民一样按照HK模型规则对其影响。若意见领袖i观点值与网民j观点值相差较大，则会引入记忆值，迭代次数与记忆值累计达到一定程度后，意见领袖观点参与网民意见演化。现实中，观点偏激用户受观点明确的意见领袖的影响是比较小的。即使他们反复阅读意见领袖观点，他们仍旧对之嗤之以鼻。因此当时，j不会记忆意见领袖观点。

记Sj\_i为普通网民j的对意见领袖i的记忆值，其初始值为0。当 时，则将意见领袖i的观点引入j记忆值中。对于普通用户j与意见领袖i,j在一时刻产生的记忆值为。同时设置代表用户j浏览意见领袖i的观点次数。事实上，网民选择相信意见领袖观点往往是多次重复浏览后，设必须多次浏览次数为C。计算公式如式3-9所示。

(3-9)

设记忆阈值为M，则有以下两种情况：

新引入的记忆值与用户j现有记忆值叠加超过了记忆阈值，此时将记忆值的观点引入主模型，并将Sj重置为初始化状态。每次Sj的更新公式如式3-10所示。

(3-10)

若,即用户j的记忆值超过记忆阈值并且用户j已多次浏览意见领袖i观点时，扩展用户j的领域。更新公式如式3-11所示。

(3-11)

若,即用户j的记忆值超过记忆阈值但是用户还没有多次浏览这个信息，常见于意见领袖观点与j相差较大时容易出现情况。按照有界信任原则，这时普通用户是不会受与其观点相距甚远的观点影响。因此，此时意见领袖i不会影响普通用户j，其迭代公式如式3-12所示。

(3-12)

若,即用户j已经浏览足够多次意见领袖的观点，但是意见领袖对普通用户j的影响还是比较小。此时需要记忆继续积累，等到超过阈值意见领袖才会影响普通用户。因此，按照式3-10迭代后不需要进行其它操作。

若,即用户j对意见领袖i的记忆值比较小重复记忆次数也较少，说明此时意见领袖对用户影响还不够，记忆值仍需继续迭代。因此，按照式3-10迭代后不需要进行其它操作。

当 时，此时意见领袖的观点已经和普通网民观点按照公式3-1进行交互碰撞。因此，已有的记忆值应该归零重新计算。此时演化迭代如上式3-12所示。综合上述过程，则普通网民i与意见领袖基于记忆的交互过程如图3-4所示。

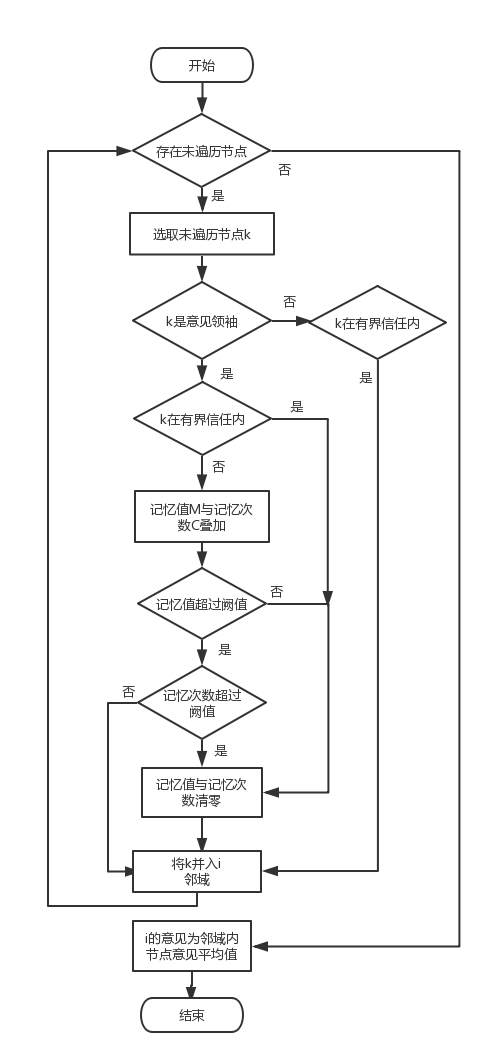


图3-4 普通网民i演化流程图

除此以外，为了充分考虑网络拓扑结构影响，引入文献[30]中的基于节点亲密度带来不同权重的简化方法。若节点i与节点j无边，则亲密度矩阵中X将亲密度定义；如果有边，亲密值 。

综合3.3.1中提出的意见领袖内部博弈演化模型和3.3.2中提出的普通网民对意见领袖基于记忆观念的改进模型，得到基于Hegselmann-Krause模型的改进模型，命名为基于意见领袖的连续型演化模型，演化算法如表3-4所示。

表3-4 基于意见领袖的博弈与记忆方式的改进hk演化模型

|  |  |
| --- | --- |
| Step1 | 初始化：假设该网络具有N个节点,它们的观点值记为O。取网络中度数前5%的节点为意见领袖，其余节点为普通网民。初始化大小为N的数组visit, 每个数组元素初始值都为0。 |
| Step2 | 在t时刻，选取任一网络节点q，计w=0表示本次迭代是否有变化；  if(visit[q]):  返回step2初始，重新选择节点；  else:  visit[q] = 1  if(q是意见领袖节点)：  转到step3，进行意见领袖节点演化；  if(q是普通网民节点)：  转到step4，进行普通网民节点演化. |
| Step3 | 对于意见领袖j=q,令y=;  if(y = )：  ，  else :  ，w=1  令t = t+1，转到step2; |
| Step4 | 对于普通网民i=q,令k=0.  while(k < N):{  if():{  if(k是意见领袖节点)：{  if():{  =    }else if(){  =  }else{  ;  +=  }  }else if() { |

续表3-4 基于意见领袖的博弈与记忆方式的改进hk演化模型

|  |  |
| --- | --- |
| Step4 | if(k是意见领袖节点)：{  =  }  }  N++；  }  转到step2; |
| Step5 | if(w == 0)  表示所有节点的观点O不再变化，则转至step6;  else:  令t=t+1,转至step2。 |
| Step6 | 结束演化 |

## 3.4 本章小结

由于收敛时间较快的Hegselmann-Krause的舆情演化模型中融合意见领袖的方式是简单加权。忽略了意见领袖与普通网民舆情驱动原因的区别，也没有考虑网络载体本身结构对于舆情演化过程影响。本章对舆情演化的网络拓扑结构和演化动力学模型做了改进。

针对ba无标度网络不可聚类这个缺点，以一定权值引入邻域连接方式，使网络具有无标度特性和小世界特性。引入近领域概念，并以一定概率区别了近领域与次领域节点与新加入节点建立连接的可能性。仿真结果表明，新得到的可聚类无标度网络的幂率分布满足无标度网络的幂率分布且其聚类系数高，较原始无标度网络更加接近于Twitter、微博等真实社交网络的聚类系数。

然后，本章在HK模型的基础上构建基于意见领袖的连续型模型。主要改进点有两点，第一是意见领袖观点演化方法中引入博弈的方法。第二是改进意见领袖对普通网民的影响模型。意见领袖与普通网民的交互模型中，普通网民对于有界信任范围的意见领袖引入记忆思想；同时根据递减距离影响的可聚类无标度网络的网络结果做节点亲密度分析，不同亲密度的节点的影响权重不同。最后给出了改进的基于意见领袖的连续型模型演化步骤。

# 4 实验设计与结果分析

本节首先介绍实验进行的环境、实验设计思路和数据预处理过程。使用网格搜索方法通过Python确定改进递减距离影响的可聚类无标度网络参数。通过爬取 “05.30头条腾讯大战”微博数据为测试数据集，使用遗传算法通过MATLAB确定改进HK模型的参数。然后使用对第三章所提改进模型进行仿真，控制变量得到不同情况下的演化结果。从结果中可以看出控制舆情演化的方法。然后通过爬取 “10.28重庆公交坠江事件”微博数据为原始数据，运用模型得到演化结果，与实际结果相比，证明模型的可行性。

## 4.1 实验设计与实现

本文实验主要通过MATLAB和Python完成，实验环境如表4-1所示。

表4-1 实验环境

|  |  |
| --- | --- |
| 硬件环境 | Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz  8.00GB RAM |
| 软件 | win10旗舰版(x64)  八爪鱼爬虫  MATLAB R2017b，Pycharm |
| 编程语言 | Python3.0 |
| 数据来源 | 微博话题 |

验证改进模型能反应真实社交网络舆情变化的过程主要通过爬取舆情初始阶段数据和舆情结束阶段数据，提取微博评论句进行情感分析，然后将舆情初始阶段情感分析结果做为原始数据导入模型中，将舆情结束阶段的情感分析数据作为最终衡量结果。上述步骤主要通过Python实现，用到的情感分析库主要有wordCloud、jieba、snowNlp，三个库都是github开源库。然后将原始数据存为MATLAB可读取的格式，在MATLAB中读取数据做为原始舆情分布数据，演化结果再与舆情结束阶段的情感分析数据比较。案例分析过程如图4-1所示。

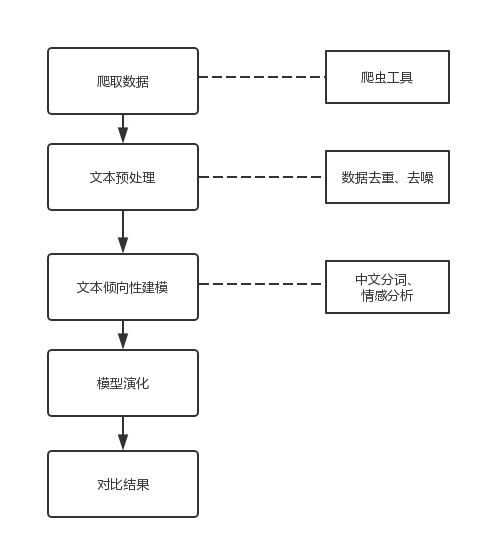


图4-1 案例分析过程

通过比较常见的爬取工具，选择“八爪鱼爬取器”对微博数据进行爬取。八爪鱼采集器使用简单，其内置爬取微博数据的模板，输入爬取微博的关键词和时间就可以一键完成爬取指定时间内微博的操作。采用关键词“重庆公交”或“公交车坠江”进行爬取数据。使用python读入数据，遍历csv文件，一行代表一条微博数据，逐行筛选需要的舆情数据。当微博内日期数据在需求日期集合内时，则将微博评论字符串放入单条微博处理程序中。

然后使用Jieba对中文文本做情感分析，然后使用wordCloud做词云图。Jieba主要做对微博评论做分词操作，每句话的情感值由其分词情感值组合决定。对于在Jieba分词训练库中存在的词，Jieba生成句子中汉字所有可能生成词情况所构成的有向无环图(DAG), 再采用了动态规划查找最大概率路径，找出基于词频的最大切分组合；对于未在Jieba分词训练库注册过的词，采用了基于汉字成词能力的HMM模型，使用了Viterbi算法。词云图能展示文本中出现频率较高的“关键词”，过滤掉大量的低频低质的文本信息，使得浏览者只要一眼扫过文本就可领略文本的主旨。WordCloud是python的词云图生成库，功能强大。但是对中文支持不够友好，无法识别中文字体，需要自行导入中文字体。SnowNlp是github上一款python写的类库，可以方便进行中文文本处理，主要处理unicord编码文本。它自带了中文正负情感的训练集，主要是评论的语料库。使用snowNlp对初始数据进行情感分析，可以得到舆情事件开始与结束时候的情感值。处理单条微博话题的情感值的过程如图4-2所示。

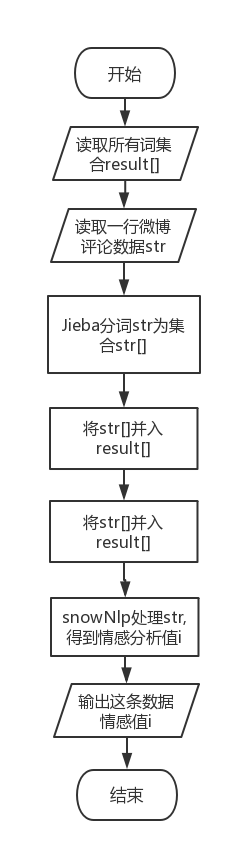


图4-2 单条微博处理过程

数据类型位list的result变量主要用于后续计算所有微博数据的词频生成词频图。舆情评价指标主要选取网络中所有个体情感值的均值和持正面意见者与负面意见者比值。sum代表所有个体意见总值，num代表导入的个体数，则网络中所有个体情感值的均值为。positive\_num代表持正面意见者的个数，negative\_num代表持负面意见者的个数。当情感分析值i>0.5时，这条微博对应的网民为持正面意见者；否则，这条微博对应的网民为持负面意见者。则网络中持正面意见者与负面意见者比值为。

将处理好的数据导入到MATLAB中。使用3.1节中递减距离影响的可聚类无标度网络生成的网络中，意见领袖应是较早加入网络中节点。则意见领袖内部舆情演化代码图如图4-3所示。

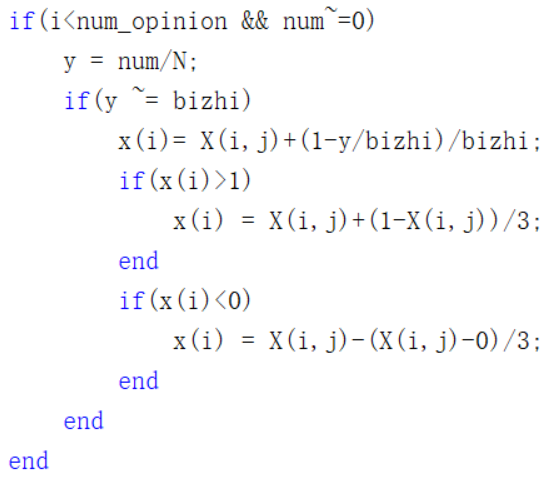


图4-3 意见领袖内部舆情演化代码图

普通网民节点与其它普通网民节点按照原HK模型的方式做交互，二者意见差值小于记忆阈值则并入邻域。普通网民节点与意见领袖节点引入了记忆方式，若二者意见差值大于记忆阈值，则普通网民节点会对意见领袖节点的观点进行记忆，实现代码如图4-4所示。

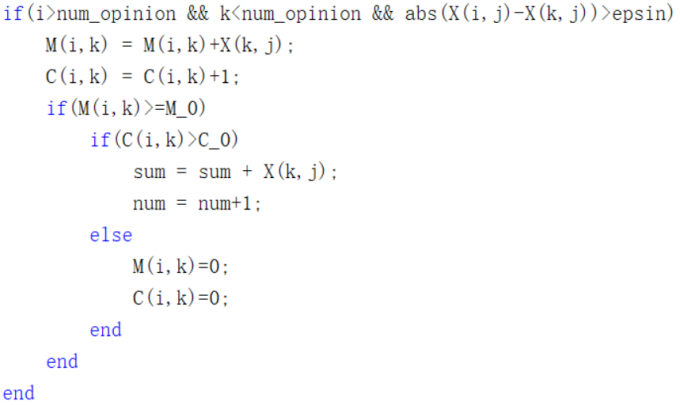


图4-4 基于记忆的普通网民与意见领袖交互代码图

## 4.2 模型参数获取

为了在改进后距离递减影响的可聚类无标度网络上运行基于意见领袖的改进HK模型，首先需要确定改进网络拓扑结构的参数，采用网格搜索方式建立调整参数。使用脚本语言python构建网络，优先连接与邻域连接比值pt从0.1开始，按照步长0.1进行搜索；在领域连接中，近领域与次领域生成边p也从0.1开始，按照步长0.1进行搜索。则生成的三维图像如图4-1所示。其中x轴表示pt值，y轴表示p值，z轴表示聚类系数。按照表3-1算法，生成规模为1000的网络。

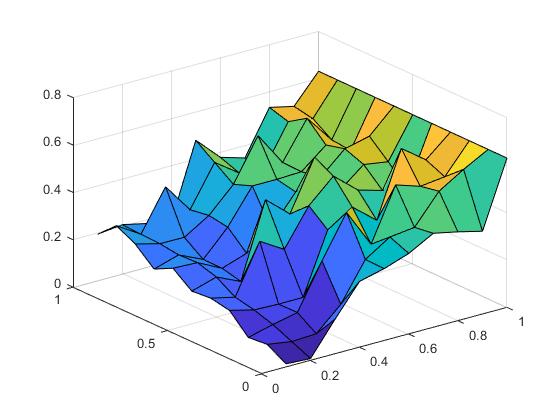


图4-5 距离递减影响的可聚类无标度网络参数图

微博的性质较表3-1的三种网络中最接近于Twitter。将Twitter的网络聚类系数值作为参考值为0.5653。参考网格搜索结果，当pt = 0.8,p=0.25时，网络聚类系数为0.5612，接近Twitter网络聚类数据。且由于pt较大，网络的无标度特性比较好。则生成的拟合曲线如图4-2所示。Barabasi与Albert R[36]的研究结果得出，BA无标度网络的幂率分布指数近似2.3，以0.1的范围波动。图4-6中，度分布近似服从幂率分布。

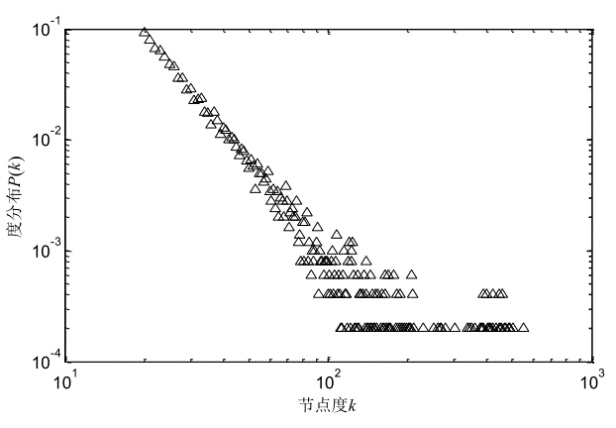


图4-6 度分布图

在改进的网络基础上获取新的基于意见领袖的博弈与记忆方式的改进hk演化模型参数，由于参数较多且不同组合对演化结果影响不定，为了优化参数选择过程，采用遗传算法来选取近似全局最优值。选取收敛较好的参数，采用遗传算法求得参数。

训练集为微博 “05.30头条腾讯大战”事件数据。“头腾大战”是指在2018年5月30日到6月初期间，今日头条公司与腾讯公司之间通过网络互骂和起诉等方式起矛盾冲突。事件起因是由于2018年5月30日，新华网发布稿件《多少道文件才能管住网游对少年儿童的戕害》，经过新浪社会专栏、百度新闻进行标题变更后，今日头条与百度新闻弹窗同样的内容《新华社：要多少文件腾讯才肯收手》推送给关注互联网资讯的活跃用户。随后双方进行了大规模网络骂战，最终演化为互为被告。“头腾大战”微博数据量如图4-7所示。

图4-7 “头腾大战”微博数据量

通过爬虫爬取5月30日到6月6日的微博数据。经分析，6月2日微博数据量达到最高峰。将5月30日到5月31日的数据作为初始舆情值带入演化模型内，将6月5日到6月6日数据作为结果值。由于舆情热度消散，且实际舆情演化过程中参与人员发生了一些随机变化，取比例结果来判断模型正确性。由6月5日到6月6日的数据可以得到，舆情均值为0.599，正面意见与负面意见比值为1.518。

采用二进制编码方式处理参数其中意见阈值,意见领袖发表负面意见带来的风险与带来的额外关注度收益比值, 记忆值阈值M。记忆次数阈值C。Matlab工具中内置遗传算法相关工具，使用MATLAB进行求参。则遗传算法参数设置如下：

（1）编码方式：采用实数转二进制编码方式。其中精度为小数点后2位，范围为1，则分别有100种可能，共有种组合。C精度为整数个位数，范围为4，共有4种可能。因此，一共有种组合，而，需要用22位二进制数做编码。先将的组合转化成实数，然后将实数转化为2进制编码。正实数与四参数之间相互转换关系式如式4-1与4-2所示。

(4-1)

=b

(4-2)

（2）适应度函数：设置适应度函数如公式4-3所示。代表“头腾大战”6月5日至6月6日实际数据值的均值，S=1.518代表实际数据值中正面意见与负面意见的比值。e代表演化数据的均值，s代表演化数据的正面意见与负面意见的比值。

(4-3)

（3）选择操作：采用轮盘赌法做选择操作。每个个体被选中概率与这个个体对应的适应度占种群中所有个体适应度之和的比值成正比，如式4-4所示。

(4-4)

（4）变异操作：本文随机产生一个[0,11]之间随机正整数i，将对应位发生变异。变异后的值对应的若大于4000000，则重新选择i值。

（5）控制参数设置：群体规模设定为30，迭代次数为50，群体的变异率为9%，群体的遗传交叉率为80%。

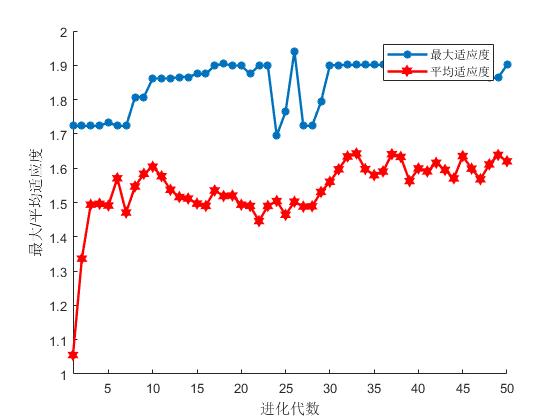


图4-8 遗传算法适应度变化趋势

则根据上述遗传算法过程，得出参数局部最优解。最大适应度的值为1.9412，对应的。根据式4-8可以求得局部最优参数值。而遗传算法迭代过程中，参数值转化为实数值与适应度y的关系散点图如图4-9所示。

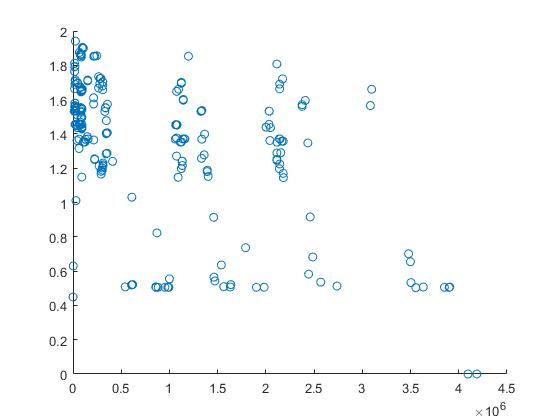


图4-9 实数x与适应度关系值

从图4-9中可以看出，当x整除后余160000左右拟合最好，解码为参数，说明在0.16时对数据集拟合较好。综合递减距离影响的可聚类网络的参数，参数值如表4-2所示。

表4-2 改进模型参数结果

|  |  |  |
| --- | --- | --- |
|  | 意义 | 值 |
| Pt | 优先连接与邻域连接比值 | 0.7 |
| P | 次领域与近领域比值 | 0.35 |
|  | 有界信任阈值 | 0.2 |
|  |  | 0.51 |
| M | 记忆值阈值 | 0.94 |
| C | 记忆次数阈值 | 2 |

## 4.3 改进模型仿真与分析

首先在相同条件下对经典HK模型做仿真。取节点数N=1000，初始加入节点=5。按照3.2节改进网络的研究结果，取做优先连接，在邻域连接中，取次领域与近领域连接比值为0.35。选取表4-2中参数，将定义为0.2.经典hk模型演化过程如图4-10所示。

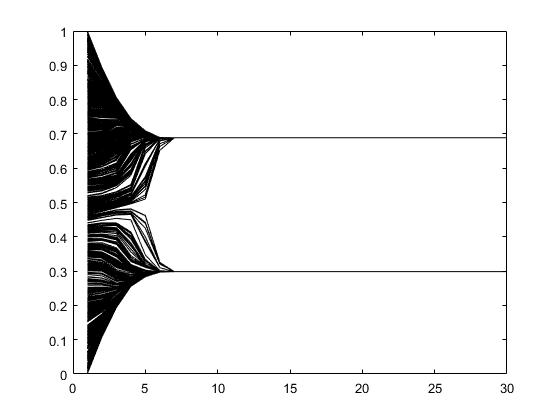
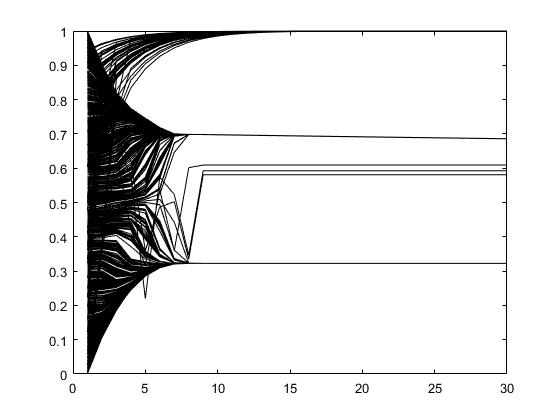
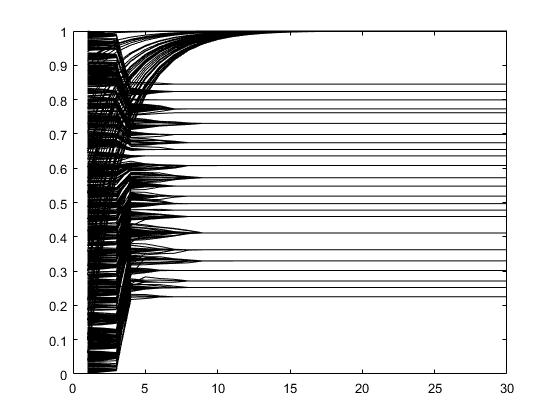
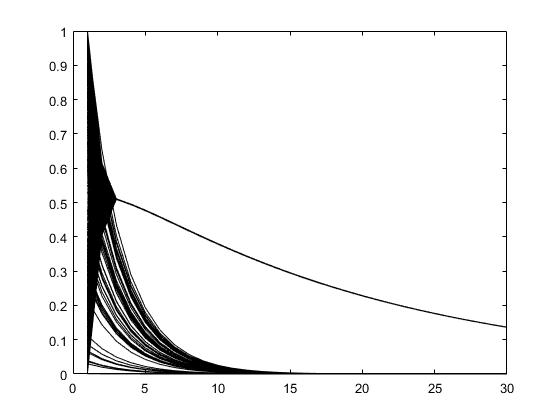
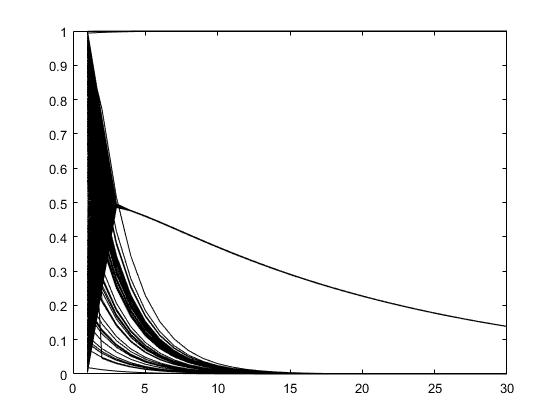


图4-10 原始hk模型

然后初始化改进模型。固定三参数v ,C的值为表4-2中所调参数。若记忆阈值小于信任阈值，则意见领袖将不会以记忆方式对普通民众观点起作用。信任阈值分别为0.01、0.2、0.5和0.8，则基于意见领袖的博弈与记忆方式的改进hk演化模型的演化过程分别如图4-11所示。



(a) (b)



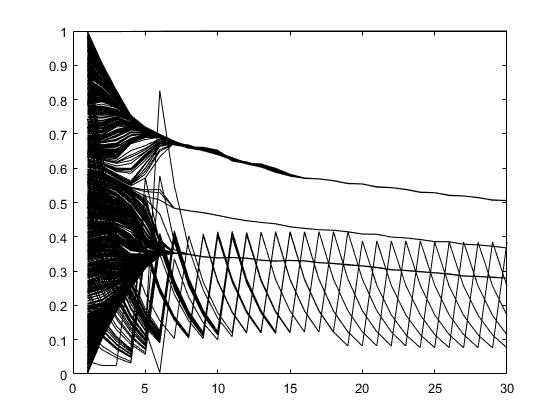
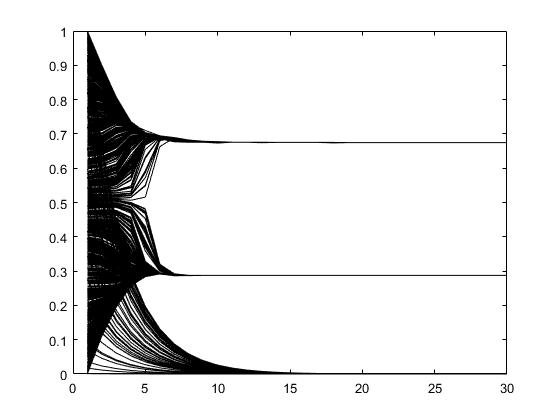
(c) (d)

图4-11 经典模型不同信任阈值下收敛情况

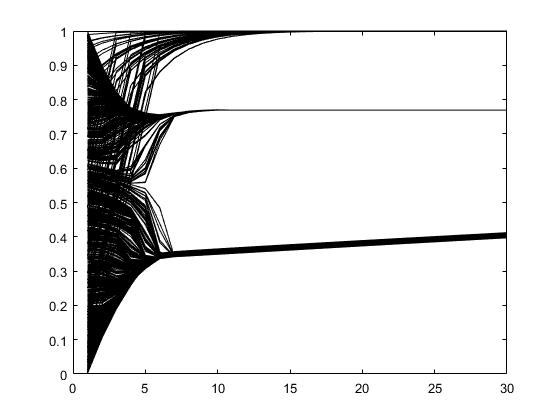
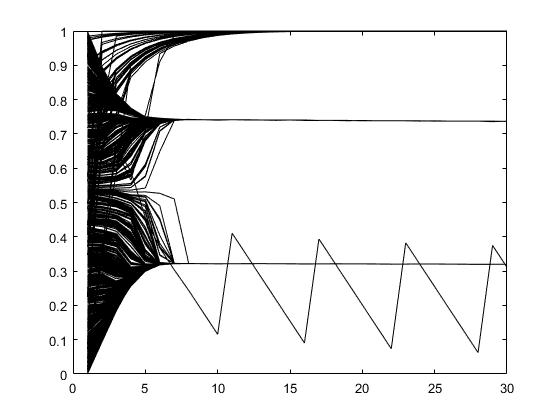
这里按照第三章定义1，将意见领袖简单的理解为度数前5%的节点。比较图4-10与4-11（b），按照基于意见领袖的博弈与记忆方式的改进hk演化模型改进后，群体意见值不再直接收敛成两个簇，部分在原始hk演化模型中最终观点收敛到0.4左右的人群会像正面意见演化。这是因为意见领袖带来的作用，由于意见领袖不按照观念邻域内节点均值为下一步演化值，其受普通用户影响方式不同，会持续引领舆论变化。这说明，新模式通过意见领袖的方式对舆情结果产生影响，意见领袖的引入可以让信任阈值较小的情况下，意见值朝一个方向演化，虽然不能最终收敛，仍能改善舆情比例结构。这个情况与现实相符，大V持续不断的发声会带来持续不断的舆情热度引导人群持续不断的进行观点演化。原始hk模型不能体现这个现象。

比较图4-11中四个图，可以发现在发布负面观念损失风险与发布观点收益的比值v=0.51，记忆阈值M=0.94，记忆次数阈值为2的情况下，当信任阈值较小，针对初始意见随机分布的情况，新模型会让意见偏正面演化；当信任阈值值较大，新模型意见偏负面演化。当信任阈值大于0.5后，新模型中信任阈值的增加不会对演化最终结果产生影响，仅有个别点会因此有更加平滑的舆情变化过程。而当模型走向不变化后，收敛时长也没发生变化。这说明，在基于意见领袖的博弈与记忆方式的改进hk演化模型中，不同信任阈值是会影响舆情最终走向。

为了观察发表负面言论风险和发表言论的收益的比值v对舆论演化模型的影响，固定三参数，v，C的值为表4-2中所调参数。比值分别为0.01、0.2、0.6和0.8，则基于意见领袖的博弈与记忆方式的改进hk演化模型的演化过程分别如图4-12所示。



(a) (b) v



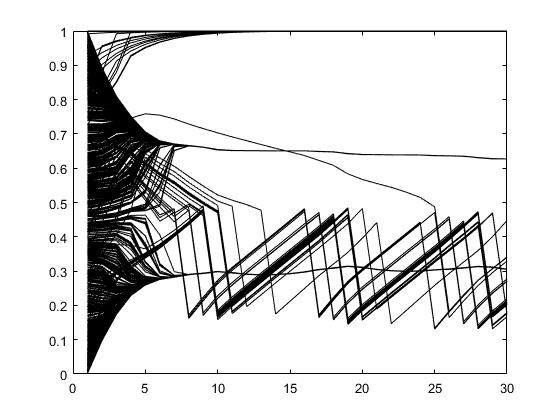
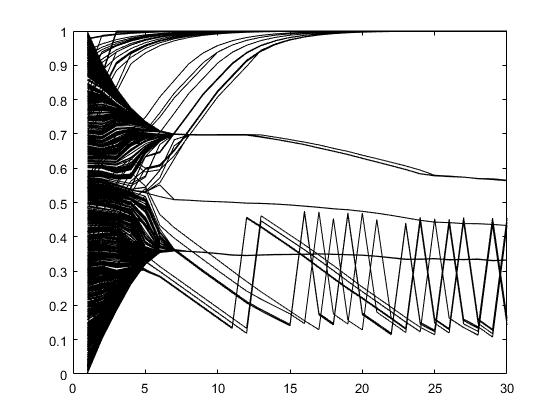
(c) (d)

图4-12 改进模型不同比值v下收敛情况

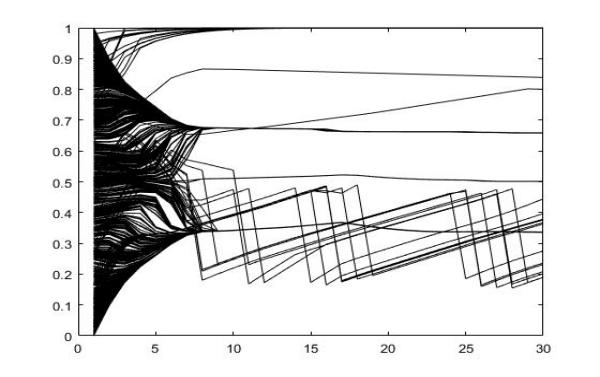
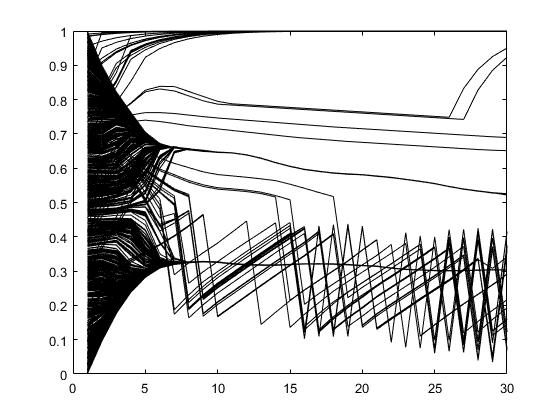
分析图4-10与4-12，当发布负面观念损失风险与发布观点收益的比值v非常小时，由于发布负面观点会给意见领袖带来额外关注，而均匀分布意见下，意见领袖的意见邻域内用户数与所有用户数比值大多小于这个比值v，此时意见领袖会迅速向负面意见靠拢以换取额外关注度，但是由于此时与普通民众信任阈值较大，普通民众几乎不受其影响。发布负面观念损失风险与发布观点收益的比值v比较小时，由于发布负面意见会给意见领袖带来额外关注度，他们仍旧会以较慢速度朝着发布反面意见的方向演化，此时由于意见领袖演化与普通网民的信任差值不大，基于有界信任观点，会影响普通网民朝着负面方向发展。但是当意见领袖发布负面观念损失风险与发布观点收益的比值v足够大时，会引导整体舆情向正向发展。这说明，若要引导舆情发展，可以通过增大意见领袖发布负面意见的惩罚措施来影响意见领袖的观念，进而影响整体舆情的趋势。

根据以上分析，我们可以得出舆情治理思路。在网络人群中，由于管控大众比较困难，而直接关闭话题容易引起民众累计不满。因此在舆情事件中，政府可以通过多种方式改变意见领袖观点意见（如交谈教育等），通过意见领袖去影响群众观点。

记忆阈值M与记忆次数C对意见领袖与普通网民的交互起作用。它们之间具有关系，当M较大时，需要足够大的记忆值才能影响普通网民，可能导致多次累计才会影响普通网民；当C较大时，也会导致多次累计观念才会影响普通网民。当信任阈值，发布负面观念损失风险与发布观点收益的比值v=0.51时，取四组不同的M与C值比较。此时舆情演化图如图4-13所示。



(a)M=0.3,C=2 (b)M=0.3,C=5



(c)M=0.8,C=2 (d)M=0.8,C=5

图4-13 改进模型M与C时演化情况

从图4-13的四个子图可以看出，改变M与C的值对整体演化趋势影响较小，整体观点值都是一部分值演化到1成为极端意见，一部分观点值在(0.5,1)之间，一部分观点值在(0.1,0.5)之间震荡。这说明，记忆模型作为将意见领袖观点引入整体模型的方式，其值的不同只能小范围影响具体意见值的变化，无法影响舆情整体趋势的改变。比较图4-8(a)(b)和(c)(d)，M值较小时会更易形成观点稳定趋势，观点大多集中于1或者震荡；而M值较大时，观点分散在(0.5,1)之间的簇数显著增多。这说明M值的增大有利于形成局部发散观点，此时要意见领袖与普通网民记忆累积到足够大才能影响普通网民。M值与C值作为记忆阈值和记忆次数阈值，主要取决于网民自身及网民对意见领袖的看法。当普通网民受教育程度较高时，他们的信任阈值和累积次数阈值会更大。

综合上述分析，可以得出舆情治理思路。改变参数v值，即政府将意见领袖发表负面观点所付出的代价提高到一个程度，最终意见会向正面方向发展。

## 4.4 重庆公交坠车事件案例分析

选取“10.28重庆万州公交坠江事件”微博数据为模型的验证集。在2018年10月，因司机与乘客争吵导致的公交车坠江事件[44]在微博、知乎等国内网络社交场合上引起了公众的广泛的关注。在本文分析案例上，“10.28重庆万州公交坠江事件”的主要发送场合为微博空间，传播形式包括短文本、短视频、图片等。但是短视频和图片等等往往起到传播事件作用，表达用户情感的主要形式仍是文本形式。不同于微信朋友圈和Facebook以表达用户私人生活为主，微博是一个带有强烈社会网络特性的社交网络。其中，微博的舆情汇聚是指发生于微博空间中的人们对某一舆情事件持有的态度的聚合过程[45]。因此选用样本案例“10.28重庆万州公交坠江事件”的微博数据来验证上述舆情演化模型。

采用现实数据做舆情分析首先要爬取真实数据，并对文本进行分析得到初始舆论情感值，对这个值做归一化处理，建立初始舆情数据值。然后将数据值导入到matlab的模型中，对模型进行演化。同时，将真实舆情结束时的微博文本用同样的文本处理手段做情感值分析，将将演化结果与真实舆情结束时的情感值作对比。经过对比，可以得出基于意见领袖的改进HK演化博弈模型更加符合真实情况。

通过比较现有爬取工具，选择“八爪鱼爬取器”对微博数据进行爬取。采用关键词“重庆公交”和“公交车坠江”进行爬取数据。

“10.28重庆万州公交坠江事件”发生于2018年10月28日10时，重庆市万州区22路公交车在[万州长江二桥](https://baike.baidu.com/item/%E4%B8%87%E5%B7%9E%E9%95%BF%E6%B1%9F%E4%BA%8C%E6%A1%A5/3836773" \t "_blank)桥面与小轿车发生碰撞后，坠入江中。从10月28日中午开始，该事件在微博上持续发酵。2018年10月29日，公安机关初步核实失联人员15人（含公交车驾驶员1人）[46]。11月02日上午10点，人民日报发文披露公交黑匣子内解密结果：一名女性乘客因下错站与男性司机激烈争执互殴致车辆失控。至11月04日，该事件热度逐渐消退。因此，爬取数据集中于10月28日12点到11月04日12点，可以得到各个时间微博数如图4-14所示。

图4-14 “10.28重庆万州公交坠江事件”微博数据量

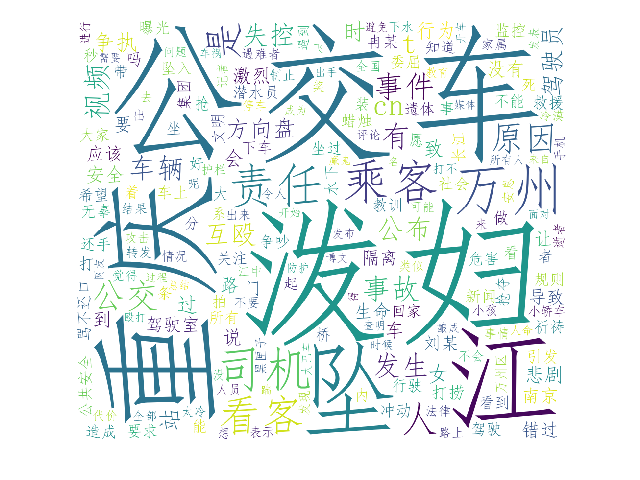
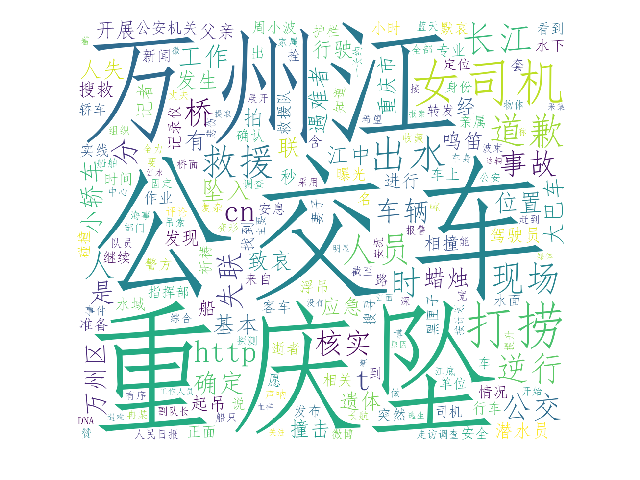
从图中可以看出，“10.28重庆万州公交坠江事件”从10月28日中午12起在微博发酵，10月29日微博数量达到一个高潮。随着救援进度开展，不断有新报道传出，但是事件热度逐渐衰退。到11月2日，随着黑匣子解密后发布事故原因，舆情达到一个最高点。因可控因素导致此悲剧事件让公众对该事件讨论程度达到一个新高度，舆情在人民日报发布黑匣子解密后登顶。随后，没有新的爆料下，舆情正常演化，话题热度逐渐消退。从图4-14中可以看到，到11月4日，相关微博数已经较少。节选代表性微博如表4-3所示。

表4-3 节选微博数据

|  |  |  |  |
| --- | --- | --- | --- |
| 微博时间 | 微博内容 | 点赞数 | 评论数 |
| 10/28 12:00 | 【#人民网快讯#【重庆万州公交车坠江 与小车相撞后冲出护栏掉入江中】… | 366 | 453 |
| 10/28 13:30 | #重庆坠江 22路公交车#真的觉得揪心，希望平安 | 0 | 1 |
| 10/28 15:00 | #重庆坠江 22路公交车#穿高跟鞋的女司机，你到底心里会不会痛？ | 10 | 9 |
| 10/29 12:00 | 事故现场视频曝光，或许公众欠那位女司机一声道歉。长沙·伍家岭 | 3 | 1 |
| 10/30 12:00 | 重庆万州公交车坠江：第二批潜水员出水，捞出两具遇难者遗体。 | 18 | 1 |
| 11/02 10:28 | #重庆万州公交车坠江#重庆公交坠江原因公布：乘客与司机激烈争执互殴致车辆失控… | 908 | 563 |
| 11/02 10:28 | 说重庆公交这件事，我认为主要责任在于公交司机… | 22 | 40 |
| 11/03 10:28 | 公交车坠江的总结：泼妇：本来只错过一站，结果错过了一生… | 30 | 26 |
| 11/04 10:28 | 万州公交车坠江总结 :总结你若不为正义站岗，你就得为邪恶陪葬！ ​ | 29 | 16 |

经过表4-1分析，10月28日到11月4日之间，公众意见在11月2日有了极大的转折。在10月28日到11月2日上午10点之前，微博用户对事件的关注主要在于被撞小轿车女司机是否有责任、救援进展等方面。到11月02日，公布事故原因后用户主要争论点在于悲剧产生责任在于打闹的女乘客、无责任的司机还是麻木的乘客。明显发现，在11月02日前后，公众对于这件事情的评论焦点不同。因此，将这个事件的舆情分为两部分：10月28日中午12点到11月02日上午10点的微博舆情和11月02日上午10点到11月04日晚上23点的微博舆情。

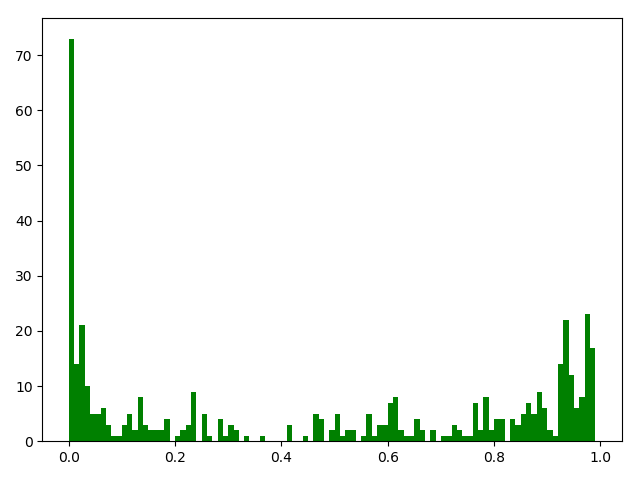
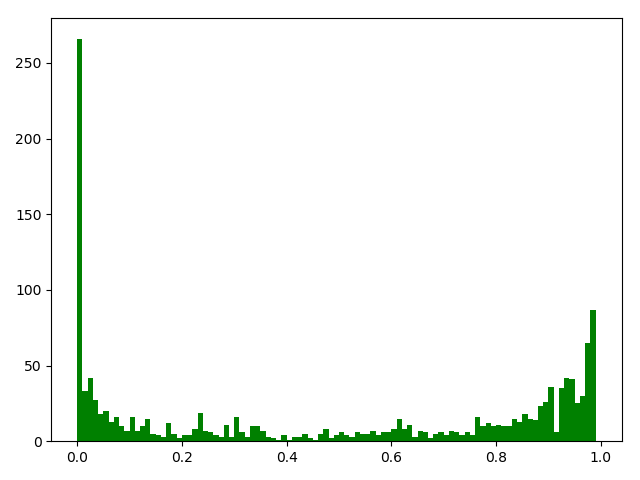
首先使用python对文本预处理，将重复微博与微博中不含“重庆万州公交坠江”等关键词的微博去除。中文文本处理第一步是做中文分词，分词是自然语言交互的基础模块。先对10月28日到11月02日黑匣子视频公布前的微博数据做分词处理，得到的分词图如图4-15(a)所示。再对11月02日到11月04日黑匣子视频公布前的微博数据做分词处理，得到的分词图如图4-15(b)所示。



（a）通报前分词 （b）通报后分词

图4-15 舆情分词图

如上分词处理是用jieba库做的分词处理，用wordCloud库做的分词展示。使用snowNlp对初始数据进行情感分析，可以得到舆情事件开始与结束时候的情感。图4-16为黑夹子公布前的原始情感分布图，其中图(a)代表10月28日到10月29日的情感数据分布图，图(b)代表11月01日情感数据分布图。



（a）初始情感值 （b）最终情感值

图4-16 黑夹子公布前情感分布值

这里我们选取意见领袖为评论及转发高的用户，将评论和转发权重各设置为0.5。则初始情感分析时，评论与转发之和排前5%的用户为意见领袖。按照4.2中得出的结论，我们取信任阈值。取发布不好消息的惩罚与发布不好消息带来的额外关注的收益比。取记忆阈值M为0.19，记忆次数阈值C为2。

将初始情感分布值导入MATLAB中的基于意见领袖的连续型hk改进模型，按照表3-3的步骤进行演化。则在黑夹子前改进模型的舆情演化过程如图4-17所示。

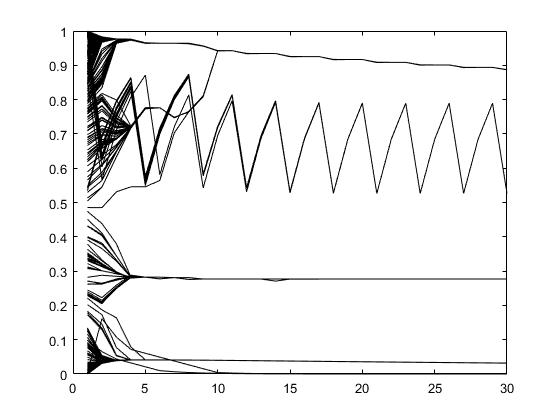
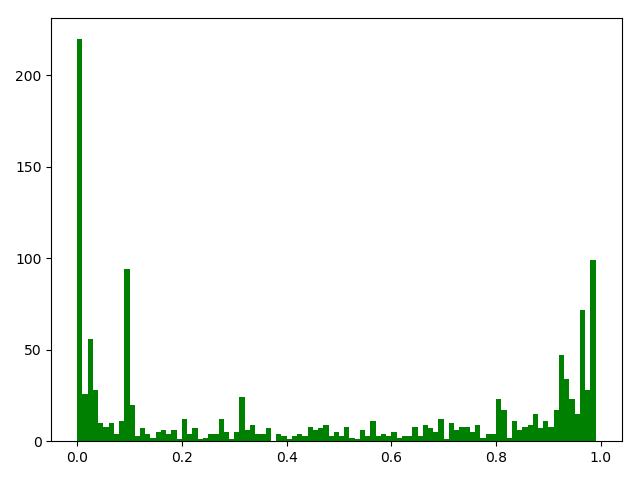
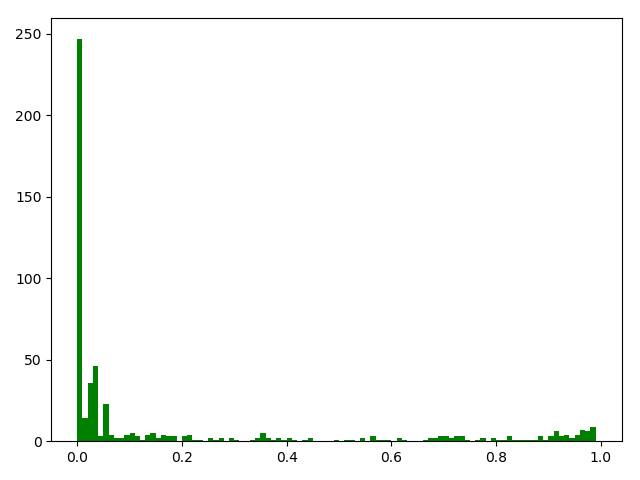


图4-17 黑夹子曝光前改进模型演化过程

采用公式4-3作为评价指标的计算公式，则将最终演化结果与计算结果为1.81，结果靠近于2。

由于黑夹子曝光后，事实真相带来了舆情内容巨大的改变。用户关注点从私家车女司机与救援过程转化为讨论公交司机与乘客大站。将黑夹子曝光后舆情演化过程另外分析，初始数据是11月02日情感分布数据，最终比较结果数据是11月08日到11月10日情感数据。则情感分析图如图4-18所示，其中图(a)代表11月02日的情感数据分布图，图(b)代表11月08日到11月10日情感数据分布图。



（a）初始情感值 （b）最终情感值

图4-18 黑夹子公布后情感分布值

同样选取意见领袖为评论及转发高的用户，将评论和转发权重各设置为0.5。则初始情感分析时，评论与转发之和排前5%的用户为意见领袖。按照4.1中得出的结论，我们取信任阈值。取发布不好消息的惩罚与发布不好消息带来的额外关注的收益比。取记忆阈值M为0.94，记忆次数阈值C为2。

将图4-18中的初始情感分布值导入MATLAB中的基于意见领袖的连续型hk改进模型，按照表3-3的步骤进行演化。则在黑夹子曝光后改进模型的舆情演化过程如图4-19所示。

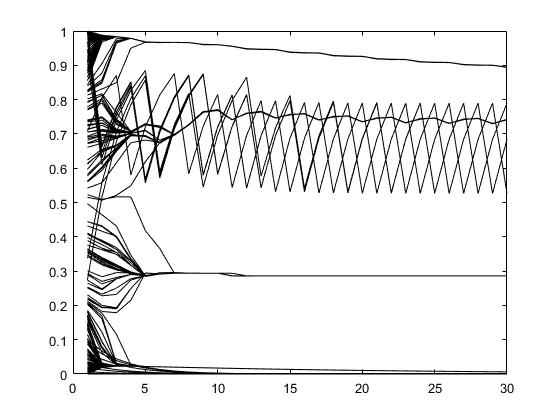


图4-19 黑夹子曝光后改进模型演化过程

此时正面意见领袖与负面意见领袖比值为0.3991，采用公式4-3作为评价指标的计算公式，则将最终演化结果与计算结果为1.85，结果靠近于2。

针对黑夹子曝光后的舆情演化博弈模型改进点如表4-4所示。

表4-4 改进效果表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 收敛簇 | 收敛时长 | 正面/反面 | 意见均值 |
| 真实数据 | 不完全收敛 | 长 | 0.333 | 0.3197 |
| 原始hk模型[15] | 2 | 短 | 0. 72 | 0.43 |
| 记忆Deffaunt模型[21] | 3 | 长 | 0.63 | 0.671 |
| hk噪声模型[18] | 不完全收敛 | 长 | 0. 657 | 0. 4021 |
| 改进模型 | 4 | 较长 | 0.399 | 0.37 |

## 4.5 本章小结

本章根据第三章中提出的基意见领袖的博弈与记忆方式的改进HK模型的步骤，先用微博数据集“头腾大战”作为训练数据集使用遗传算法得到参数值。然后固定三个参数值改变一个参数值，进行模型仿真并得出结论：改变意见领袖发布负面意见的风险大小可以控制舆论走向。最后，使用热点事件“10.28重庆公交坠江事件”微博数据集进行试验，给出实际结果证明了改进模型较原始HK模型更能拟合真实情况。

在本章中，首先使用MATLAB建模仿真，在3.1提出的改进聚类无标度网络上做了初始值服从均匀分布的舆情演化仿真。通过仿真，得出模型的合适参数。再将合适网络参数带入到模型中，使用遗传算法对基于意见领袖的博弈与记忆方式的改进hk演化模型参数做编码和迭代。将上述得到的参数带入模型中建模仿真，得出舆情演化意见，可以通过提高意见领袖发布负面意见的代价使舆论朝证明发展。最后，通过将原始HK模型、基于记忆的Deffaunt模型、引入噪声的hk模型和加权引入意见领袖方式的hk模型和改进后的基于意见领袖的改进HK演化博弈模型与“10.28重庆公交坠江事件”中较后日期的微博数据比较，得出改进模型有效。

# 5 总结与展望

## 5.1 全文总结

随着移动化网络的发展，社交网络在人们生活中占有越来越大的比重。网络舆情在舆情统计与演化中作用越来越大。目前对于舆情研究的切入点和方法都很多，结合物理方法的舆情动力学模型是研究热点之一。目前经典连续型舆情动力学主要集中于Deffuant模型和Hegselmann-Krause,而Deffuant模型做为早期连续型舆情模型研究比较成熟。目前关于HK模型的研究尚不充分，演化中考虑实际情况引入意见领袖、亲密关系等因素大多通过引入特定权重的观点加成完成。这忽略了意见领袖与普通民众观点变化不同的驱动原因，也没有充分考虑舆情载体的网络拓扑结构。

本文提出了基于意见领袖的博弈与记忆方式的改进Hegselmann-Krause模型。针对目前的HK模型没有充分考虑节点身份不同带来演化方式区别和网络拓扑结构的问题，完成了以下两方面的研究：

（1）提出了一种递减距离影响的可聚类无标度网络，它是舆情演化过程的网络载体。首先分析了在线网络的拓扑特点，然后在Holeme&Kim提出的三元无标度网络基础上，引入近领域概念，与次领域形成一定的比例使无标度网络具有可聚类特性。该方法引入的“三度影响力”理论能更好模拟现实社交网络的邻域连接情况，经过调参可以使聚类系数靠近Twitter社交网络的聚类系数。

（2）提出了一种基于意见领袖的博弈与记忆方式的改进Hegselmann-Krause模型*。*它充分考虑意见领袖作为舆情中的利益者，对意见领袖的演化引入基于有界信任的“囚徒博弈”思路。然后，通过引入记忆值方法，来加强意见领袖对普通网民的影响，使意见领袖的影响渗透于网络中其它用户。这个新模型进一步体现出了舆论演化过程中意见领袖的作用，反应了意见领袖与普通网民的异质性，更加符合真实的社交网络舆论演化情况。

最后，使用微博关于“05.30头条腾讯大战”事件数据为训练集，通过 Matlab构建遗传算法调参得到合适参数。爬取“10.28重庆万州公交坠江事件”的微博数据，对其做情感建模，将改进模型应用于此数据集上，结合仿真所得参数，得出改进模型能很好的演化结果接近真实网络舆情演化结果的结论。

## 5.2 展望

随着网络舆情的发展，未来舆情动力学模型的优化和与复杂网络的融合仍是研究重点。在递减距离影响的可聚类无标度网络上，使用基于意见领袖的博弈与记忆方式的改进Hegselmann-Krause模型，选取“10.28重庆万州公交坠江事件”作为数据集，证明了改进模型较原始hk模型更符合实际过程。然而，在后期研究中还是有问题有待解决。

首先，演化模型时仅仅用到了网络中节点最大度的方式来确定意见领袖，这个方法度量网络中意见领袖不够准确。网络的拓扑结构对舆情演化时的领域影响不仅在于最大度数上，如何进一步融合网络拓扑结构特点来改进领域计算方式是下一步研究方向。

其次，演化模型的参数是固定的，这与现实情况不符。整个舆情演化过程中，发布负面意见的收益与发布负面意见的风险比值是动态变化的。舆情潜伏期和爆发期，意见领袖发布负面意见的收益会较高;而随着舆情发展，意见领袖发布负面意见的风险会逐渐增高。除此以外，不同地区和不同受教育程度的网民对同一热点事件的意见领袖意见的记忆阈值和记忆次数阈值是不同的。如何动态改变参数是下一步研究方向。

最后，演化模型与复杂网络的结合还比较少。随着复杂网络的研究深入，将复杂网络和舆情演化模型结合是一个趋势。本文对网络载体的应用仅有确定意见领袖节点和用两节点是否有边确定节点亲密度，后续需要进一步在舆情动力学模型中引入网络拓扑结构特点。

# 致 谢

伴随着毕业论文的完结，我的研究生生涯阶段也要结束了，以前总觉得时间很长，我还是个在学校无忧无虑的学生，一转眼学生时代都将过去。在这里首先要感谢我的导师沈刚教授！感谢他对我在硕士期间学习生活上给予的各种教导，通过自己的行为感染我们做一个正直有责任感的人。沈老师总能规划好各种事情，积极阅读业内最新资料，尝试不同的新方法，对待科研非常勤恳，经常工作到很晚。研一时，沈老师还在百忙之中给我们定期召开组合，讲授机器学习相关课程，亲切询问我们的困难解答我们的疑惑。除此以外，沈老师还会积极给我们分享一些新知识，比如在参加完一些会议和比赛评审工作之后，他会将这些总结之言都告诉我们。这些知识和内容是我们平时接触不到的，让我们的思路和视野更宽一些这些。是沈老师的渊博知识和勤恳负责工作态度让我学会了科研的方式。我相信，这种态度也会深深感染我以后的工作方式。今后我也要严格的要求自己，向老师学习。

除此以外还要感谢我的父母，每当我遇到挫折时，是他们鼓励我坚持住努力前行。三年时光我曾碰到很多挫折，正是父母的鼓励让我能顺利毕业。

最后，还有感谢陪伴我三年的同学们，感谢我的室友，是她们的鼓励包容让我在硕士三年能够安心学习生活。

# 参考文献

1. 刘建明. 舆论研究的建树与未来-舆论学发展的百年回顾与展望. 新闻知识, 2000(5): 6-7
2. 中国互联网络信息中心. 第 42 次中国互联网络发展状况统计报告. [http: //www. cac. gov. cn/2018-08/20/c\_1123296882. htm](http://www.cac.gov.cn/2018-08/20/c_1123296882.htm)
3. 人民网舆情监测室. 2017年中国互联网舆情研究报告. [http: //zjnews. china. com. cn/yuanchuan/2018-01-04/138792. html](http://zjnews.china.com.cn/yuanchuan/2018-01-04/138792.html)
4. 文学国, 范正青. 中国危机管理报告(2011). 北京: 社会科学文献出版社, 2011
5. Conradt L, List C. Group decisions in humans and animals: a survey. Philosophical Transactions of the Royal Society B Biological Sciences, 2009, 364(1518): 719-742
6. Lasswell H. D. The structure and function of communication in society. 1948: 86-86
7. 张亚楠. 社交网络上舆论形成模型研究. 河南科技大学, 2017
8. Frs J. D. M. Mathematical Biology. 北京: 世界图书出版社, 2013: 88
9. Sudbury A. The Proportion of the Population Never Hearing a Rumour. Journal of Applied Probability, 1985, 22(2): 443-446
10. Xiong F., Liu Y., Zhang Z., et al. An information diffusion model based on retweeting mechanism for online social media. Physics Letters A, 2012, 376(30): 2103-2108
11. Zhou J., Liu Z., Li B. Influence of network structure on rumor propagation. Physics Letters A, 2007, 368(6): 458-463
12. Deffuant G., Neau D., Amblard F., et al. Mixing beliefs among interacting agents. Advances in Complex Systems, 2014, 03(01n04): 0000007
13. 张伟, 何明升, 白淑英等. 基于 Weisbuch-Deffuant 模型的网络舆论演化模式研究. 情报杂志, 2013(7): 43-48
14. Shang Y. L. An agent based model for opinion dynamics with random confidence threshold. Communications in Nonlinear Science and Numerical Simulation, 2014, 19(10): 3766-3777
15. Hegselmann R., Krause U. Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation. Journal of Artificial Societies & Social Simulation, 2002, 5(3): 2
16. Lorenz, Jan. Consensus Strikes Back in the Hegselmann-Krause Model of Continuous Opinion Dynamics Under Bounded Confidence. Journal of Artificial Societies and Social Simulation, 2006, 9(1): 8
17. Pineda M., Toral R., Hernández-García E. The noisy Hegselmann-Krause model for opinion dynamics. European Physical Journal B, 2013, 86(12): 490
18. Su W., Chen G., Yu Y. Finite-time elimination of disagreement of opinion dynamics via covert noise. Iet Control Theory & Applications, 2018, 12(4): 563-570
19. 张亚楠, 孙士保, 张京山等. 基于节点亲密度和影响力的社交网络舆论形成模型. 计算机应用, 2017, 37(4): 1083-1088
20. Kitsak M., Gallos L. K., Havlin S., et al. Identification of influential spreaders in complex networks. Nature Physics, 2010, 6(11): 888-893
21. 黄庆花, 宋玉蓉. 基于优先选择和记忆效应的观点动力学研究. 计算机工程, 2014, 40(11): 36-41
22. 胡珑瑛, 董靖巍. 网络舆情演进过程参与主体策略行为仿真和政府引导. 中国软科学, 2016(10): 50-61
23. 彭小兵, 邹晓韵. 邻避效应向环境群体性事件演化的网络舆情传播机制——基于宁波镇海反PX事件的研究. 情报杂志, 2017, 36(4): 150-155
24. 曾子明, 黄城莺. 基于BP神经网络的突发传染病舆情热度趋势预测模型研究. 现代情报, 2018(5)
25. Zhao Y., Kou G. 2016 6th International Conference on Computers Communications and Control (ICCCC) - Opinion evolution of a social group with extreme opinion leaders, 2016: 70-74
26. Greenberg, Albert, hamilton, et al. VL2: A Scalable and Flexible Data Center Network. Communications of the Acm, 2009, 54(4): 95-104
27. 马永军, 杜禹阳. 基于复杂网络Deffuant模型的舆情演化规律研究. 情报杂志, 2018(6)
28. 游丹丹, 陈福集. 基于改进粒子群和BP神经网络的网络舆情预测研究. 情报杂志, 2016, 35(08): 156-161
29. 黄亚驹, 陈福集, 游丹丹. 基于混合算法和BP神经网络的网络舆情预测研究. 情报科学, 2018, 36(02): 24-29s
30. Erdös P., Rényi A. On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci., 1960, 5: 17-60
31. 郭世泽, 陆哲明. 复杂网络基础理论. 北京: 科学出版社, 2012: 122-148
32. Milgram S. The small world problem. Psychology today, 1967, 2(1): 60-67
33. 周涛, 柏文洁. 复杂网络概述. 物理, 2016, 34(1): 10-15
34. Watts D. J., Strogate S. H. Collective dynamics of ‘small-world’ networks, Nature, 1998, 393(6684): 440-442
35. 汪小帆, 李翔, 陈关荣. 网络科学导论. 北京: 高等教育出版社, 2012: 56-60
36. Barabasi A. L., Albert R. Emergence of Scaling in Random Networks. Science, 1999, 286(5439): 509-512
37. 罗伯特·吉本斯著. 博弈论基础. 高峰译. 北京: 中国社会科学出版社, 1999
38. 宣飞, 查章林, 华德志等. “囚徒困境”博弈的理论模型及现实思考. 现代商贸工业, 2011, 23(17): 30-31
39. 王增. 基于博弈论的蜂窝网络功率控制欲多维效率优化研究. 北京: 北京邮电大学, 2018
40. Smith J. M., Price G R. The Logic of Animal Conflict. 1973
41. 李燕凌, 丁莹. 网络舆情公共危机治理中社会信任修复研究——基于动物疫情危机演化博弈的实证分析. 公共管理学报, 2017(4): 91-101
42. Holme P., Kim B. J. Growing scale-free networks with tunable clustering. Physical Review E, 2002, 65(2): 95-129
43. Christakis N. A., Fowler J. H. Social Network Sensors for Early Detection of Contagious Outbreaks. Plos One, 2010, 5(9): e12948
44. 重庆坠江公交车沉入江底, 已打捞出两具遗体. 四川日报电子版. [https: //cbgc. scol. com. cn/news/100541](https://cbgc.scol.com.cn/news/100541)
45. 高承实, 荣星, 陈越. 微博舆情监测指标体系研究. 情报杂志, 2011, 30(9): 66-70
46. 新浪新闻. 重庆坠江公交初步核实15人失联, 车辆位置基本确定. [http: //news. sina. com. cn/c/2018-10-29/doc-ihnaivxq2861526. shtml](http://news.sina.com.cn/c/2018-10-29/doc-ihnaivxq2861526.shtml)