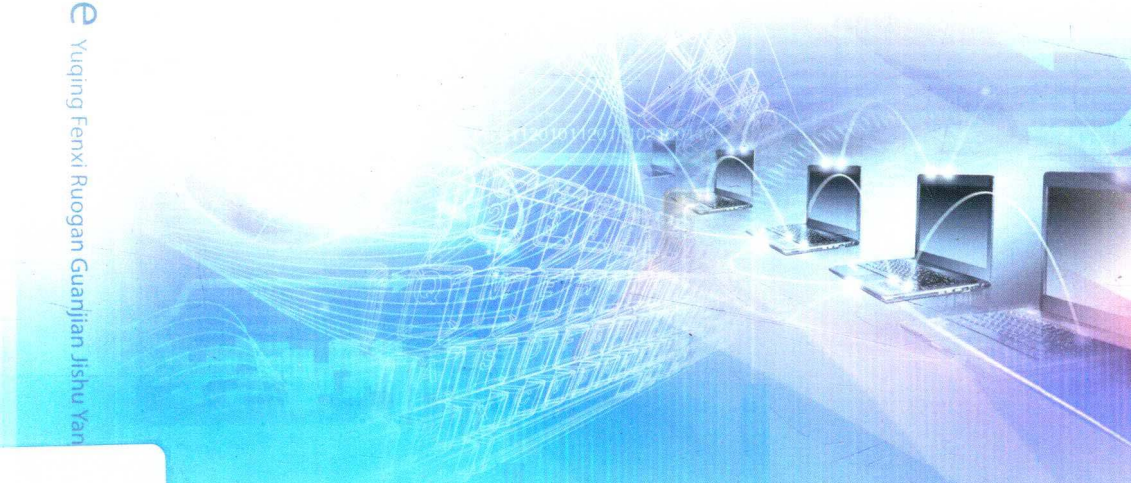
河南省科技计划重点科技攻关项目
（172102210174、172102210105、182402210025、192102210258）资助
河南省高等学校重点科研项目（18A520020）资助
河南城建学院科研能力提升工程研究项目资助

Mianxiang Weibo Tufa Huati De Yuqing Fenxi Ruogan Guanjian Jishu Yan

面向微博突发话题的舆情分析若干关键技术研究

董国忠　著



中国矿业大学出版社

河南省科技计划重点科技攻关项目(172102210174、172102210105、182402210025、192102210258)资助
河南省高等学校重点科研项目(18A520020)资助
河南城建学院科研能力提升工程研究项目资助

面向微博突发话题的舆情分析若干关键技术研究

董国忠 著

中国矿业大学出版社

内 容 简 介

舆情挖掘是计算机科学领域的重要内容之一，是多种技术交叉融合的产物。本书面向微博突发话题针对舆情挖掘的若干关键技术进行分析与研究，主要内容包括面向微博舆情感知的微博数据预处理、面向实时微博消息流的在线突发事件检测、面向微博突发话题的社区关键用户挖掘和面向微博突发话题的突发模式挖掘。

本书适用于对面向微博突发话题的舆情挖掘若干关键技术感兴趣的本科生、研究生与科研人员，以及从事微博突发话题舆情挖掘的工程技术人员阅读参考。

图书在版编目(CIP)数据

面向微博突发话题的舆情分析若干关键技术研究 /
董国忠著. —徐州 ：中国矿业大学出版社，2018.10
ISBN 978-7-5646-4198-6

Ⅰ. ①面… Ⅱ. ①董… Ⅲ. ①互联网络－舆论－研究
Ⅳ. ①G206.2

中国版本图书馆 CIP 数据核字(2018)第 237483 号

书　　名　面向微博突发话题的舆情分析若干关键技术研究
著　　者　董国忠
责任编辑　褚建萍
出版发行　中国矿业大学出版社有限责任公司
（江苏省徐州市解放南路　邮编 221008）
营销热线　(0516)83884103　83885105
出版服务　(0516)83995789　83884920
网　　址　http://www.cumtp.com　**E-mail**:cumtpvip@cumtp.com
印　　刷　徐州中矿大印发科技有限公司
开　　本　787×960　1/16　**印张** 6.5　**字数** 150 千字
版次印次　2018 年 10 月第 1 版　2018 年 10 月第 1 次印刷
定　　价　26.00 元

前　言

随着国内外主流社会媒体的快速发展，社会媒体已经逐渐取代传统媒体，成为人们发布、分享信息的主要平台。社会媒体给信息传播提供便利的同时也成为突发话题产生与传播的重要平台。与传统媒体不同，微博产生的突发话题可以不受时间、空间的限制，大大增加了面向微博突发话题的检测与挖掘分析的难度。当微博中涉及敏感信息的消息大规模爆发形成突发话题时，如果不能及时有效地检测以及挖掘分析突发话题，突发话题产生的负面舆情将不断发展，最终会成为影响广泛的社会事件，危及整个社会的安全。因此，面向微博突发话题的舆情挖掘分析研究已经得到学界和业界的重点关注。

由于微博平台具有数据量大、信息碎片化严重、用户质量良莠不齐、信息传播快等特性，通过人工方式进行实时监测并不能实时有效地检测和挖掘微博突发话题。因此，面向微博等社会媒体舆情产生的主要媒介，如何面向微博消息流实现突发话题检测以及突发话题挖掘分析，从而有效阻止微博舆情危机爆发并正确引导微博舆论是社会媒体舆情领域亟待解决的重要问题。本书以最具代表性的国内外主流微博平台作为研究对象，旨在面向微博突发话题研究突发话题检测、突发话题挖掘分析方法与技术。主要针对如下关键问题展开研究：

(1) 面向微博舆情感知的微博数据预处理方法

为了有效地检测突发话题，本书提出了适用于微博舆情感知的关键词和微博用户预处理方法。在关键词预处理方面，为了避免发现伪突发关键词，提出一种基于社会信任和动力学模型的突发关键词检测方法，该方法基于物理学中动力学的基本概念，将微博中的关键词突发现象抽象为关键词动量的变化，然后采用 MACD 指标计算每个关键词的突发权值并根据突发阈值判断该词在特定的时间窗口是否为突发关键词。在大规模新浪微博数据集上的实验结果表明此方法能够检测到微博中的突发关键词并且最大限度地避免发现伪突发关键词。在微博用户预处理方面，针对微博平台中存在大量低质量的僵尸粉丝用户，为了有效过滤僵尸粉丝等营销用户对突发话题检测准确率的影响，提出了一个基于交互图模型的僵尸粉丝检测方法。此方法利用用户交互关系构建用户交互图模

型，根据交互图模型提出了高鲁棒性的基于交互的僵尸粉丝发现特征，并利用不同的机器学习分类器对提出的特征的有效性进行实验验证。实验结果表明本书提出的基于交互特征的方法能够更加有效地发现僵尸粉丝。本书从突发关键词及僵尸粉丝用户两个角度提出的预处理方法为后续突发话题检测及挖掘分析奠定了基础。

(2) 面向实时微博消息流的在线突发事件检测方法

针对面向实时微博消息流的突发话题检测问题，考虑涉及微博负面舆情产生的突发话题通常是社会事件类突发话题，本书提出了一种面向微博消息流的突发事件检测方法，此方法首先基于滑动时间窗口构建高效的二层哈希表存储及更新模型，然后提出一个自适应调整阈值的候选突发消息检测算法提取突发消息，并从候选的突发消息中去除僵尸粉丝用户发布的突发消息，最后融合突发关键词及事件特征对突发消息进行增量聚类从而形成突发事件。实验结果表明本方法能够更加准确地检测实时微博消息流中的突发事件。

(3) 面向微博突发话题的社区关键用户挖掘方法

针对微博突发话题关键用户挖掘问题，考虑促使突发话题形成的关键用户对舆情事件传播的影响，提出了一种面向突发话题的社区关键用户发现方法。此方法基于突发话题用户关系对突发话题建立突发话题用户图模型，并利用基于随机游走的社区发现方法挖掘突发话题用户关系图中的用户社区。针对大规模的用户社区，利用基于排序的方法检测关键用户。该方法与其他关键用户检测方法相比能够更加有效地挖掘出促使突发话题早期传播与扩散的关键用户。

(4) 面向微博突发话题的突发模式挖掘方法

针对突发话题的突发模式挖掘问题，本书提出了一种面向突发话题的突发模式挖掘方法。此方法基于突发话题用户关系对突发话题建立突发话题用户图模型，结合突发话题用户图模型提出了宏观及微观突发模式挖掘方法，在宏观突发模式挖掘方面，提出了面向突发话题特征的层次聚类挖掘方法，该方法能够挖掘出不同类别的突发话题；在微观突发模式挖掘方面，提出了面向不同类别突发话题的频繁子图挖掘方法，该方法能够挖掘出不同类别突发话题中的频繁信息流模式。

本书的撰写参考了大量的国内外研究成果，他们的研究成果和贡献是本书的基础和思想源泉，在此对涉及的研究人员表示衷心的感谢！哈尔滨工程大学博士生导师杨武教授在百忙中认真、细致地审阅了全部书稿，并提出了建设性的指导意见和建议，在此向杨武教授表示衷心的感谢！河南城建学院计算机与数

据科学学院何宗耀教授等为本书的撰写提供了许多有益的指导与帮助，陈东莹、史春雷对书稿进行了校对工作，本书的出版得到了河南省科技计划重点科技攻关项目(172102210174、172102210105、182402210025、192102210258)、河南省高等学校重点科研项目(18A520020)、河南城建学院科研能力提升工程研究项目的资助，在此一并表示感谢。

舆情挖掘技术发展迅速，涉及的技术多，其理论和应用均有大量的问题亟待进一步深入研究。由于笔者才疏学浅，仅略知一二，书中不妥和错谬之处在所难免，敬请同行专家和读者批评指正，将不胜感激。

著　者

2018.5

目　录

第1章 绪 论

1.1 研究背景及意义

随着2006年Twitter正式上线以及Web 2.0技术的快速发展，以互联网为载体的各种社交网络平台成为Web 2.0时代最具代表性的应用[1]，其中微博客（以下简称微博）作为其中主要的平台得到了广大网民的关注。国内包括新浪、腾讯、搜狐和网易等主要网络媒体平台自2009年开始分别推出各自的微博服务，微博正式进入中文上网主流人群视野[2]。如图1.1所示，根据中国互联网信息中心2017年1月的《中国互联网发展状况统计报告》显示，截至2016年有37.1%的网民使用微博，较2015年有明显的增长，微博是现在网民分享信息、传递信息的主要社会媒体平台。

与论坛、博客等传统内容分享平台不同，微博是一种通过用户关注、消息转发等机制分享、传播实时简短信息的社交网络平台[3]。具有如下特点：

① 传播速度的即时性：微博用户可以通过固定互联网和移动互联网使用微博平台，实现了固定终端与移动终端的融合，使得微博与其他媒体相比实时性更高，传播速度更便捷、更迅速[4]。

② 传播内容的自主性：微博不仅仅表现为一个传播平台，同时又是一个通过用户发布等方式实现内容自创的平台，让微博用户成为内容的制造者、传播者以及评论者[5]。微博平台通过文字、超链接、图片和视频等多种信息载体，为微博用户提供了多元、多层次和多角度的扩展性能，增加了传播内容的自主性。

③ 传播方式的互动性：微博实现了真正意义上的双向互动传播，微博用户可以通过一对多、多对一、多对多等多种形式进行信息传递。

微博的上述特点满足了人们在工作、娱乐等碎片化时间里的发布信息、分享交流等需求，吸引了大批用户活跃在微博平台中[6]。由于移动互联网的发展及移动终端设备的普及，使得网络信息在微博平台的产生更加方便，在微博用户间的信息传播也更加快捷迅速，微博已经成为互联网用户发布、分享信息的重要途径，逐渐演变成为大众化的互联网舆论平台，越来越多新闻媒体及明星等影响力用户都通过微博来发布或传播信息[7]。由于微博的即时性、自主性以及互动性，许多非常规突发事件发生后，微博作为人们信息发布的主要载体为突发事件提

	2016年		2015年		
应用	用户规模(万)	网民使用率	用户规模(万)	网民使用率	全年增长率
即时通信	66 628	91.1%	62408	90.4%	6.8%
搜索引擎	60 238	82.4%	56623	82.3%	6.4%
网络新闻	61 390	84.0%	56440	82.0%	8.8%
网络视频	54 455	74.5%	50391	73.2%	8.1%
网络音乐	50 313	68.8%	50137	72.8%	0.4%
网上支付	47 450	64.9%	41618	60.5%	14.0%
网络购物	46 670	63.8%	41325	60.0%	12.9%
网络游戏	41 704	57.0%	39148	56.9%	6.5%
网上银行	36 552	50.0%	33639	48.9%	8.7%
网络文学	33 319	45.6%	29674	43.1%	12.3%
旅行预定[17]	29 922	40.9%	25955	37.7%	15.3%
电子邮件	24 815	33.9%	25847	37.6%	-4.0%
论坛/bbs	12 079	16.5%	11901	17.3%	1.5%
互联网理财	9 890	13.5%	9026	13.1%	9.6%
网上炒股或炒基金	6 276	8.6%	5892	8.6%	6.5%
微博	27 143	37.1%	23045	33.5%	17.8%
地图查询	46 166	63.1%	37997	55.2%	21.5%
网上订外卖	20 856	28.5%	11356	16.5%	83.7%
在线教育	13 764	18.8%	11014	16.0%	25.0%
互联网医疗	19 476	26.6%	15211	22.1%	28.0%
互联网政务	23 897	32.7%	—	—	—

图 1.1　中国互联网用户各类应用使用率

供了第一传播平台，包括“2013 年吉林省松原地震”以及“青岛石油管线爆炸”等突发事件，微博都是最早的信息来源。

微博中对社会突发事件的报道和讨论对于危机应对和态势感知是有积极意义的，但是涉及负面舆情的突发事件一旦通过微博在短时间内不断被转发扩散，将会产生极为不良的影响，对于这类由于社会事件形成的突发话题应尽早发现、控制和疏导以及挖掘形成机理，将其不利影响降至最低，以保证微博网络舆情乃至互联网舆情的健康发展。因此，以微博平台为研究对象，建立统一的面向微博突发话题的舆情分析研究框架对于微博舆情监管具有重要意义。

准确地感知微博突发话题、挖掘与分析突发话题是微博舆情监管中亟待解决的重要问题，具体包括以下四个研究问题：

(1) 研究适用于面向微博舆情感知的数据预处理方法。

由于微博平台中存在大量与突发话题无关的碎片化数据，需要从关键词的角度提出一种突发关键词检测方法，进而可以更好地构建与描述突发话题。针对微博平台中存在大量低质量的僵尸粉丝用户，为了有效过滤僵尸粉丝等营销用户对微博舆情感知准确率的影响，需要研究适用于微博的僵尸粉丝检测方法。上述两个方面为后续的舆情感知及挖掘分析奠定了基础。

(2) 研究微博舆情事件相关的突发话题检测方法。

由于微博平台的特点，微博中的舆情事件通常是在大规模传播后才被相关部门发现，导致微博舆情难以控制。因此，在事件未形成负面舆情前研究尽早地检测微博消息流中舆情事件相关的突发话题的方法，对于控制负面舆情的扩散具有重要意义。

(3) 研究促使突发话题形成的关键用户挖掘方法。

由于微博舆情在微博中的传播扩散速度非常快，通过实时突发话题检测可达到监测微博舆情的目的。但如何有效地控制负面舆情的扩散以及对舆情进行正确地引导，需要进一步研究挖掘微博突发话题关键用户的方法。因此，挖掘促使突发话题形成的关键用户对于微博舆情的全面控制具有重要意义。

(4) 研究突发话题的突发模式挖掘方法。

对于已经形成相当规模的突发话题，由于信息已经大规模的传播，需要研究突发话题传播扩散的突发模式，进而有助于深入理解突发话题的形成机理以及优化突发话题的检测方法。

综上所述，作为新型社交网络媒体，微博已成为我国网络舆情的重灾区。微博中的突发话题则是影响微博舆情的重要因素，本书所研究的面向微博舆情感知的微博预处理方法、微博舆情事件相关的突发话题检测方法、促使突发话题形成的关键用户挖掘方法以及面向微博突发话题的突发模式挖掘方法具有重要的理论意义和实际应用价值。

本课题在微博舆情感知及挖掘分析的应用背景下，以为微博用户提供安全、可靠的社会媒体平台为目的，对面向微博舆情感知的数据预处理方法、微博舆情事件相关的突发话题检测方法、促使突发话题形成的关键用户挖掘方法以及突发话题的突发模式挖掘方法展开研究。本课题将积极探索社会计算、数据挖掘以及网络舆情管理相结合的新思路，并提供相关基础理论及关键技术。综上所述，本课题的选题具有较强的理论研究和实际应用意义。

1.2 国内外研究现状

本课题主要涉及社会计算、社会网络舆情感知、社会网络挖掘与分析等相关

理论及方法。Schuler[8]在1994年首次提出了社会计算的概念，其认为社会计算可以是任何一种类型的计算应用，以软件作为媒介进行社交关系的应用，此后不同领域的学者对社会计算给予了不同的定义。中国学者主要从交叉学科的角度对社会计算进行定义。中国人民大学的孟小峰等认为，社会计算是使用系统科学、人工智能、数据挖掘等科学计算理论作为研究方法，将社会科学理论与计算理论相结合，为人类更深入地认识社会、改造社会，解决政治、经济、文化等领域复杂性社会问题的一种理论和方法论体系[9]。

微博作为主流的社会媒体，是网络舆情产生与传播扩散的重灾区。正是由于微博的特性使得许多非常规突发事件发生后，微博作为人们信息发布和传播的主要载体为突发事件提供了平台。对于这类由于社会事件形成的突发话题应尽早感知并挖掘分析其形成机理及突发模式，以保证微博网络舆情乃至互联网舆情的健康发展。

为了深入理解微博舆情感知与挖掘分析的相关研究，本书将从社会网络相关概念、微博舆情感知、微博挖掘与分析的国内外相关研究等方面进行具体的介绍。

1.2.1 社会网络相关概念

社会网络又称社交网络(Social Network)，社交网络由代表不同个人或团体的节点构成，呈现出实体之间的关系网络，其作为一种虚拟社交媒介，为用户提供了保持联系、分享信息的平台[10]。社交网络的理论基础与经典的“六度分隔理论”(Six Degrees of Separation)[11]和“150法则”(Rule of 150)[12]密切相关。2001年哥伦比亚大学的P. S. Dodds等通过邮件系统中连锁信实验证实了该理论在社交媒体中的真实存在[13]。近年来随着社交媒体的飞速发展，人与人之间的距离不断被缩短，Backstrom等通过对Facebook的分析发现，Facebook上的用户间的距离已经减小到4，并根据这个现象提出了“四度分隔理论”(Four Degrees of Separation)[14]。

社会网络呈现出复杂网络中的“弱关系”[15]和“无标度”[16]特征。社会网络与人类现实生活紧密相关，现实生活中的同质性与偏见性在社交网络中也得到了体现。同质性主要是指社会网络中的用户与他们的邻居节点或“朋友”具有一定的相似性。相关学者的研究[17,18]证实了在社会网络中同质性的真实存在。偏见性[19]主要是指社会网络中的用户间存在大量的社交关系，但是仅有有限的部分用户与该用户存在更加频繁的交互关系。

进入21世纪后，随着互联网新兴技术的兴起，大量优秀的社会媒体平台涌现，例如国外的Facebook、Twitter以及国内的新浪微博、腾讯微博等，这些社交网络平台为社会计算的研究提供了更大的发展空间。微博作为典型的社交网络

平台，用户群不断增加，导致微博逐渐代替传统的新闻媒体成为人们获取信息的主要来源。但由于微博信息传播碎片化严重以及存在大量低质量用户，导致了舆情事件在微博中更加容易产生。相关学者举例描述了社会计算在舆情分析和Web 2.0安全方面的应用，表明社会计算为信息安全工作者提供了新的视野、建模方法和处理数据的工具[20]。

为了能够更好地描述社会媒体用户形成的社交网络，关于社交网络用户与用户之间的网络结构、用户关系的重要统计概念被相关学者提出，下面对经常使用的基本概念进行简单介绍。

节点度(Degree)：节点度主要是指与节点相连接的边的个数，是衡量网络中节点影响力的重要指标之一。

社交网络又分无向社交网络和有向社交网络，典型的无向社交网络有Facebook、人人网等，有向社交网络最有代表性的有Twitter、新浪微博等。在无向社交网络中节点度的计算如式(1-1)所示：

$$Degree(v_i) = \sum_{j \in L_i} e_{i,j} \tag{1-1}$$

式中 $e_{i,j}$——表示 i 用户和 j 用户间的边；

L_i——表示 i 用户的邻居节点集合。

在有向社交网络中节点度又分为节点入度 $Degree_{in}(v_i)$ 和节点出度 $Degree_{out}(v_i)$。节点入度指的是节点的连入边，节点出度指的是节点的连出边，具体公式如式(1-2)和式(1-3)所示。

$$Degree_{in}(v_i) = \sum_{j \in IN_i} e_{i,j} \tag{1-2}$$

式中 IN_i——表示所有连入 i 用户的邻居节点集合，在微博网络中等同于用户的粉丝集合。

$$Degree_{out}(v_i) = \sum_{j \in OUT_i} e_{i,j} \tag{1-3}$$

式中 OUT_i——表示所有连出 i 用户的邻居节点集合，在微博网络中等同于用户的关注集合。

在对网络中的节点进行度的计算的基础上，可以求解网络中节点的累积度分布函数(Cumulative Degree Distribution Function)[21,22]，$CDDF(k)$ 表示节点度不小于 k 的节点的概率分布，如式(1-4)所示。

$$CDDF(k) = \sum_{\lambda=k}^{\infty} P(\lambda) \tag{1-4}$$

式中 $P(\lambda)$——表示一个随机选定的节点度为 λ 的概率，如果 $P(\lambda) \propto \lambda^{-\alpha}$ 则符合幂指数为 $\alpha-1$ 的幂律。

结构洞(Structural Holes)[23]：结构洞是表示节点重要性的重要特点之一，主要是指链接多个区域的桥接节点，即当删除该节点后，网络将会出现空洞。

中心性(Centrality)：中心性是衡量节点在网络中重要程度的重要指标之一，主要的中心度包括：度中心性、介数中心性等。

(1) 度中心性(Degree Centrality)：度中心性主要是指社会网络中的节点，其邻居节点的个数越多其影响力也越大[24]。度中心性可以表示为式(1-5)：

$$DC_i = \frac{k_i}{N-1} \tag{1-5}$$

式中 k_i——表示 i 的度。

(2) 介数中心性(Betweenness Centrality)[25]：节点的介数中心性主要是指网络中的全部最短路径中，经过该节点的最短路径所占的比例。节点 i 的中心性可以表示为式(1-6)：

$$BC_i = \sum_{j,k} \frac{d_{jik}}{d_{jk}} \tag{1-6}$$

式中 d_{jk}——表示从节点 j 到节点 k 的最短路径个数；

d_{jik}——表示从节点 j 到节点 k 并且经过 i 的最短路径个数。

结强度(Tie Strength)[15]：结强度是衡量两个用户关系紧密度的指标之一，主要是指两个用户间的共同邻居用户的数量，其公式如式(1-7)所示：

$$Ts(a,b) = \frac{|L_a \cap L_b|}{|L_a \cup L_b|} \tag{1-7}$$

最短路径长度(Shortest Path Length)：最短路径是衡量节点间信息传播难易程度的一个重要指标，主要是指从网络中的一个节点到另一个节点所经过的最短距离。最短路径越短，表示两个用户间更加容易传递信息。

平均路径长度(Average Path Length)[22]：平均路径长度可以用来衡量网络中信息传播的难易程度，表示网络中任意两个节点间的最短路径长度的平均值。其计算公式如式(1-8)所示：

$$APL = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geqslant j} dl(i,j) \tag{1-8}$$

式中 $dl(i,j)$——表示 i 用户到 j 用户的最短路径的长度。

网络密度(Network Density)：网络密度是衡量网络中实际存在的边数和理论上最大存在的边数的比值。用以衡量网络中节点间的互联程度，互联程度越高信息在网络中的传播也就更为容易。其计算公式如式(1-9)所示：

$$ND = \frac{|E|}{\frac{1}{2}N(N+1)} \tag{1-9}$$

同配系数(Assortativity Coefficient)[22]：同配系数主要用于计算社会网络中的节点是否倾向于连接和自己度值相近的节点。当同配系数大于0时表示网络是同配的，当同配系数小于0时表示网络是异配的。其计算公式如式(1-10)所示：

$$r=\frac{M^{-1}\sum_i j_i k_i-\left[M^{-1}\sum_i \frac{1}{2}(j_i+k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2+k_i^2)-\left[M^{-1}\sum_i \frac{1}{2}(j_i+k_i)\right]} \tag{1-10}$$

式中 j_i,k_i——表示第 i 条边两边端点的度数。

聚类系数(Clustering Coefficient)：节点 i 的聚类系数主要是刻画节点 i 的邻居节点间的连接程度，如式(1-11)所示：

$$CC_i=\frac{E_i}{\frac{1}{2}l_i(l_i-1)} \tag{1-11}$$

式中 E_i——表示 i 邻居节点间的实际边数；

l_i——表示 i 节点邻居节点的个数。

此外，还可以从全网的角度量化聚类系数，即网络中节点聚类系数的平均值，其计算公式如式(1-12)所示：

$$\overline{CC}=\frac{1}{n}\sum_{i=1}^{n}CC_i \tag{1-12}$$

$\overline{CC}=0$ 表示该社会网络中不存在任何边，$\overline{CC}=1$ 表示该社会网络为全连通图。

1.2.2 微博舆情感知研究现状

微博舆情是指广大网民在微博社交平台上呈现出来的对社会上的人或事的看法以及引起的讨论。微博舆情的传播则是由微博用户推动的微博舆情在微博网络内的传播过程，微博用户推动的微博舆情传播结果形成了微博舆情传播网络。微博网络中的舆情事件从产生、传播扩散到大规模爆发的时间非常短，使得微博网络中的舆情感知的复杂性陡增。

代表不同主题的微博舆情事件会引起不同微博用户群体的关注，进而引发不同规模的讨论，越能引起更多用户群体关注的微博舆情事件在微博中讨论越为激烈。用户对微博舆情事件的讨论热度也会随着事件的发展不断变化，在微博舆情产生与扩散过程中，由于微博用户可以通过转发、评论、@等用户行为广泛参与，导致微博舆情形成爆发、裂变式的扩散等特征。具体来看，微博舆情呈现出以下几个特征：

(1) 来源广泛

人们在社会生活中往往会考虑现实生活中的各种原因而不能把自己意见或情绪在社会生活中自由地表述出来。而微博等社会媒体的出现则克服了现实社会生活中时间与空间的限制，微博等社会媒体平台上的每个微博用户都可以通过社会媒体自由地表达对社会生活中的各类事件的态度或意见。正是微博等社会媒体的开放与包容性，使得每个微博用户都可以成为微博舆情事件信息的发布者，并且微博面向用户粉丝的自动推送机制，在一定程度上促使了微博舆情信息的广泛传播。从微博舆情所涉及的事件类型来看，微博用户对社会生活中出现的各种现象和各类热点事件都有所关注，微博舆情事件涉及政治、经济、文化、娱乐、科技、体育等各个领域。

(2) 即时性

微博平台的出现，强化了舆情事件产生的即时性。随着移动互联网技术的快速发展，微博手机客户端的普及，微博用户可以通过各种移动终端随时随地发布自己感兴趣的舆情事件关联的微博消息，并且时刻关注事件的进展情况。每当现实生活中舆情事件有所发展时，关注该事件的微博用户便会参与到该事件动态的最新讨论中。微博平台低门槛的内容发布模式也使得微博用户在发布微博时无须考虑过多语言润色以及复杂的格式编排，仅需发布用户自身感兴趣的舆情事件相关的消息即可，更加深化了微博舆情产生的即时性。

(3) 碎片化

微博平台传统的文本微博消息的字数要求限制在140字以内。大部分普通用户发布的微博内容比较简短，仅仅通过一条微博消息很难了解整个微博舆情事件的起因、发展等动态信息，微博舆情的消息呈现出了碎片化的特性。微博平台的长微博等功能虽然在一定程度上能够降低微博平台的碎片化，但是由于使用长微博功能的用户所占比例较小，改变不了微博舆情的碎片化特性。正是由于微博舆情的来源广泛和微博的即时性等特征，当某个用户想要在微博舆情事件发展的过程中了解某舆情事件的始末，需要浏览大量的微博信息，给微博用户了解舆情事件的整体发展造成了困难。

(4) 去中心化

微博平台的社交和媒体属性，使微博平台不同于传统的新闻媒体，微博用户在微博平台中扮演着微博舆情信息的制造者和微博舆情的传播者两种角色。双重身份的用户体验调动了众多微博用户的发布和传播信息的积极性，使得普通的微博用户从传统媒体传播情境中的旁观客体转变为微博信息传播中的参与主体。当舆情事件发生后，微博中的任意用户都可以成为舆情事件的参与主体，通过自由发布有关该事件的信息促使微博舆情的扩散。因此，微博打破了传统新闻媒体的中心化传播模式，为普通微博用户提供了一个自由发布信息、平等交流