## **Information propagation model based on hybrid social factors of opportunity, trust and motivation（6.2 李昊展）**

基于机会、信任和动机的混合社会因素的信息传播模型

### **基本信息**

期刊：《Neurocomputing》

发表时间：2018

作者：Jihong Wan, Xiaoliang Chen, Yajun Du, Mengmeng Jia

作者信息：

School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

Department of Computer Science and Operations Research, University of Montreal, Montreal, QC H3C3J7, Canada



### **文章主要研究内容**

信息在网络中的传播是一个复杂的动态过程。围绕共同的社会因素建立准确的信息传播模型从其识别有价值的信息和监管公众舆论的进化模拟中受益匪浅。本文基于三元封闭原则，在社会网络工作中提出了机会、社会信任和博弈选择动机，并构建了结合这三种社会因素的新信息传播模型。首先，两个用户之间的兴趣相似性便于衡量接收某些信息的机会。其次，通过将用户的网络影响力和内容贡献耦合来计算社会信任的门槛。第三，带有计算最佳契合度规则的游戏选择已被公认为用户传播信息的动机。最后，提出了基于页面排名算法（GCIP-页面排名）的游戏选择信息传播模型。实验结果表明，由用户主导的信息传播经历了信息接触、信息信任和信息传播等阶段。与此同时，用户的社交信任可以通过同时考虑网络结构和信息内容来加速微博社交网络中的信息传播。

### **文章内容**

**1.问题提出**

传播模型已经被用来解释和模拟信息如何在社会网络上扩散。大规模的在线社区，如Facebook、Twitter和微博，在人们的信息交流和社会交往中越来越受欢迎。开发一个高度准确的信息传播模型对于研究社交媒体中潜在的信息传播过程有很大的好处。在公认的信息流分析中可以找到完美的例子，如社会网络中的舆论预测和引导。这些研究提供了许多合理的技术手段来避免威胁社会安全和造成社会动荡的舆论传播。

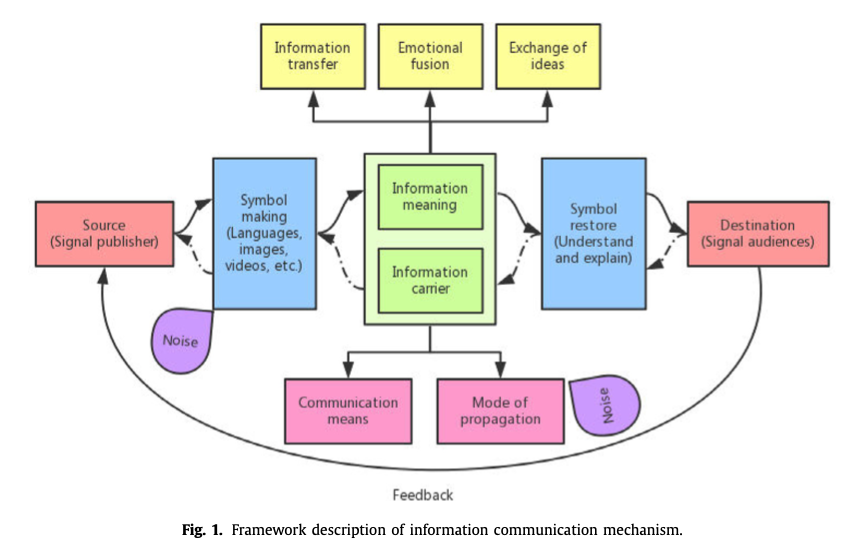
社会网络中的信息传播模式已被广泛关注。然而，在研究信息传播问题的过程中，现有的工作存在着过度的滥用，忽略了一些重要的社会、动机和有效的因素。在传统的模型中可以找到这样的例子，这些模型只用无权的无向图来表达。这样的模型将作为降低复杂性的有效途径。然而，从长远来看，构建社会网络的简化模型可能无法正确反映真实情况。一些与社会因素有关的最重要的特征会被丢失。因此，很可能得到一个与真实情况相反的模型。

本文依靠社会网络中的三段式封闭原理，定义了三类社会因素，即机会、社会信任和游戏选择动机。在信息传播的过程中，引入了传播者的社会和情感等传播因素。从网络拓扑结构和内容贡献两个角度来看，社会信任可以加速微博社交网络的信息传播。此外，以进化论博弈中个人收入最大化的特点作为用户选择信息传播的动机。 实验结果表明，以进化博弈为桥梁，将机会、信任和动机这三个社会因素整合到信息传播模型中，可以得到一个更加真实的信息传播模型。

**2.信息传播模型相关研究现状**

**（1）信息传播框架**

社会网络分析是社会网络理论研究的核心。它依赖于根据实际的传播机制构建信息传播模型，其中，社会网络中的传播机制和分享关系机制相辅相成。信息传播机制可以准确表达某种信息从发布者到接受者的框架。图1显示了信息传播的形式、方法和过程，包括传播者、传播渠道、媒介和接收者等基本信息。在信息载体的基础上，符号信息（如语言、图片、视频等）的编码和解码使信息能够被传送到接收者。而信息具有信息传递、情感融合和思想交流的意义。



**（2）基于过程的信息传播模型**

基于传播过程的传播模型描述了用户接受信息的状态和信息传播状态的变化过程。常见的模型源于SI、SIS、SIR、SIRS等传染病模型。这种模型通常只构建了信息传播规则，但没有考虑网络拓扑结构对传播行为的影响。鉴于信息在博客网络上的传播与博客网络本身的特点密切相关，Leskovec等人基于SIS模型构建了博客网络中的信息级联传播模型。此外，Xiao等人考虑到影响热门话题信息扩散的外部和内部驱动因素，保留了基于用户多分类属性和进化博弈的热点传播模型，并结合传统的SIR流行病模型。

这些信息传播模型主要关注传播过程的动态和个体的再分配。但它们没有考虑到用户本身在信息传播过程中的一些关键社会因素（如机会、信任和动机），没有显示出信息传播过程中的关键传播特征。

**（3）基于意见领袖的传播模型**

舆论领袖的评论或评论在特定领域具有一定的影响。舆论领袖在信息传播中发挥着重要作用，他们是针对紧急情况而出现的，可以影响和指导公众的决策选择。目前，大多数研究旨在准确识别网络中的观点领导者，进一步的工作主要考虑意见领袖在信息传播中的作用和目的。例如，为了分析舆论领袖在在线社会网络工作（OSN）中的作用和新浪微博信息传播的生命周期，提出了公共应急信息传播模型和一种新方法意见领导模型。Yong等人通过用户之间的信任指标和不信任指标评估了意见领袖的影响，其中指标使用知识分数（KS）、匹配系数（MC）和Jaccard系数（JC）通过利用正负链接分数来反映用户之间信任关系的强度。

虽然这些研究人员采用用户信任机制来确定舆论领袖在信息传播过程中的关键作用，但他们没有关注舆论领袖和用户信息传播之间信任机制的联系和重要性。

**（4）基于博弈论的传播模型**

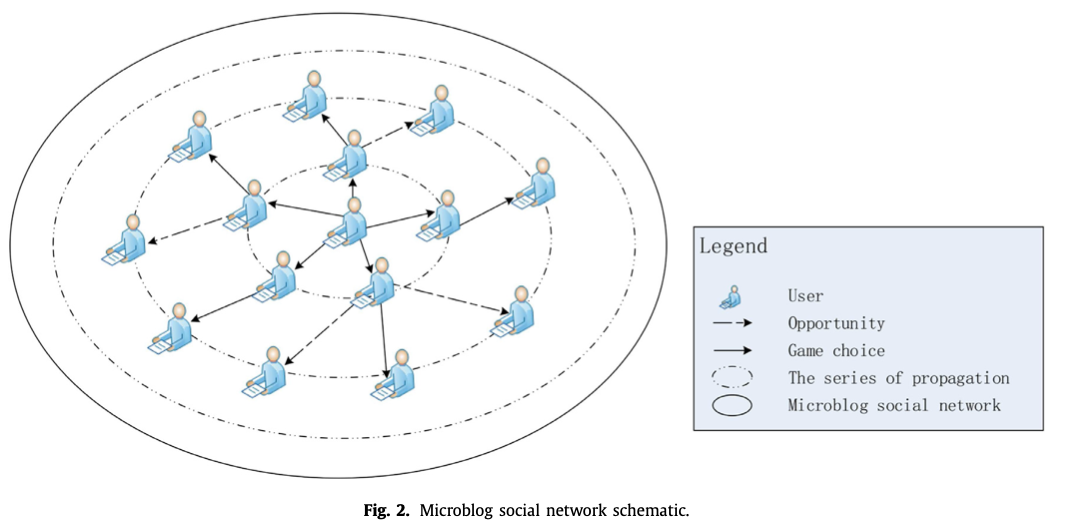
一些研究从博弈论的角度关注信息转发行为。特别是Jafari和Navid提出了一种博弈论的方法来模拟社会网络上的竞争性扩散，这表明信息内容、图的拓扑结构和个人的初始倾向对扩散过程有重大影响。Zinoviev和Duong利用行为人的知识水平、信任度和影响力来形成发布和评论的策略。并提出了星型社会网络的单向信息转发和反馈机制的博弈论模型，该模型考虑了参与者的个性。Qiu等人利用效用函数将用户的自身特征（如知识、信念、说服力、记忆和声誉等）与用户的行为特征（特别是传播和接收）相结合，在进化传播中获得用户的利益。因此，一个潜在的信息传播模型以评论和转发的形式被构建出来。

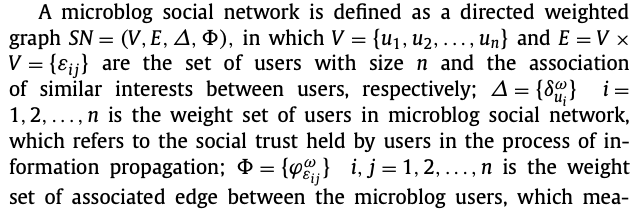
这些研究只是片面地将信任作为用户行为的衡量标准，而没有考虑到它在整个网络拓扑结构的信息传播过程中所起到的关键桥梁作用。

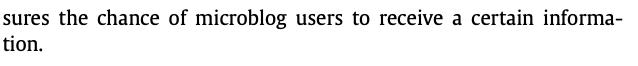
**3.三个社会因素及其算法**

**（1）微博信息传播网络**

图2是一个微博社会网络，表示微博用户发布信息后，兴趣相近的邻居节点将有机会收到信息，这一点用虚线表示。在衡量用户的游戏收益后，选择利益最大化的邻居节点进行传播。虚线椭圆表示信息传播的系列。

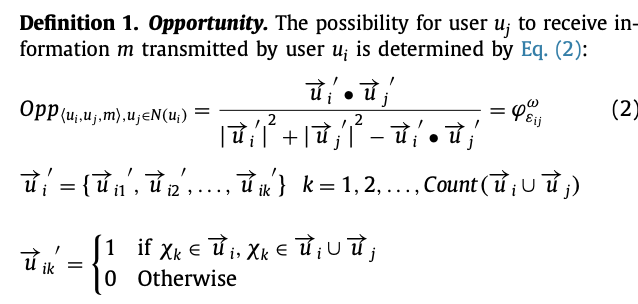
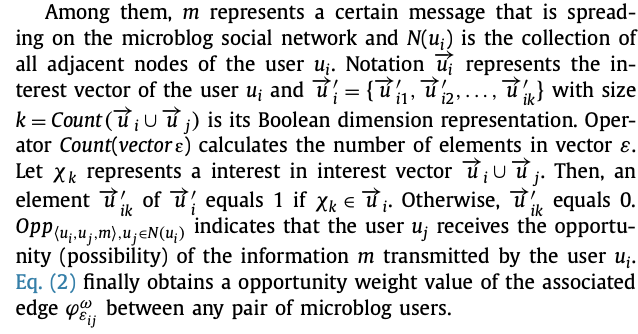




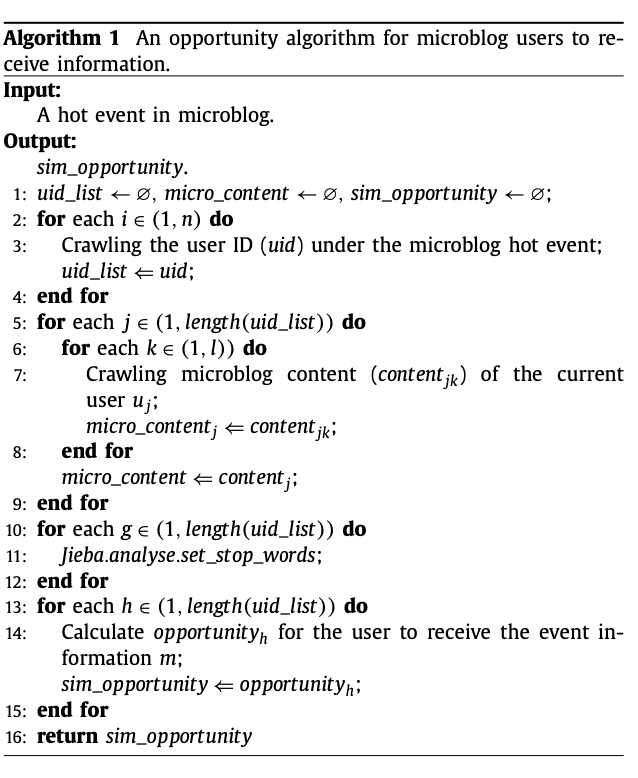
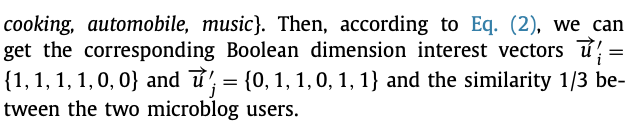
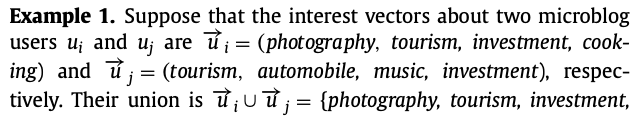


**（2）机会**

与用户ui有相似兴趣的邻居用户可能会关注同一个话题，并有机会收到用户ui在微博上发布的信息m。因此，两个相邻用户之间的兴趣相似度被作为接收某些信息的机会指标。根据微博用户的兴趣特征，一个用户的重点兴趣通常以布尔向量的形式呈现。改进的Tanimoto相似度系数可以用来计算布尔度量中个体之间的相似度（衡量共同兴趣）。

例子：

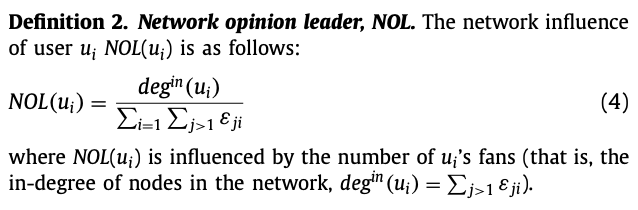


**（3）社交信任**

当微博用户有可能收到来自其邻居的信息时，确定信息的可信度对他/她来说是必要的，这通常取决于发布信息的微博用户的社会信任。用户由此产生了阅读、分享、评论和赞美等反馈行为。本文将网络结构和微博内容这两个方面全面结合起来衡量用户的社会信任，即网络结构和内容贡献。

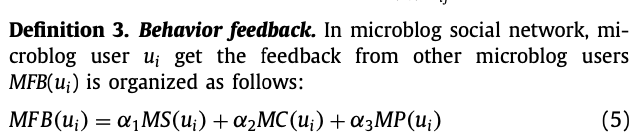
**基于网络结构的社交信任（意见领袖地位）**

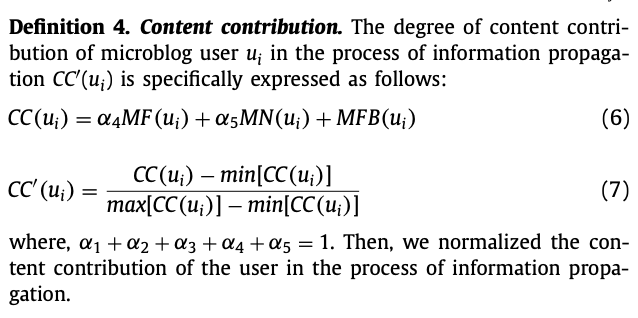
由于信息内容的影响力无法根据网络结构来衡量，网络用户的影响力越大，社会信任就越大。因此，本研究将发布信息的用户的网络影响力作为衡量用户社会信任的标准之一。

****

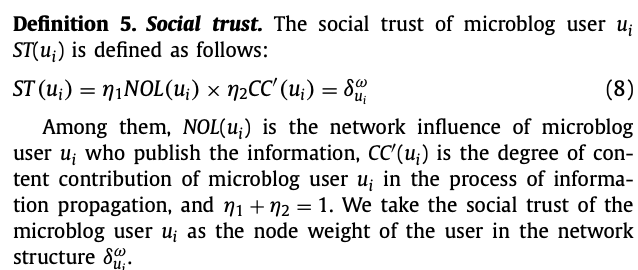
基于内容贡献的社交信任

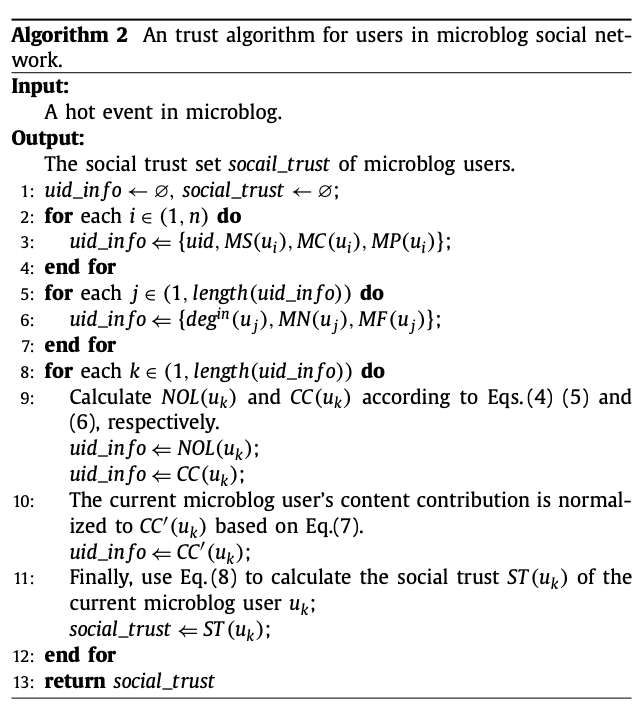
微博用户ui 的CC(ui)在信息传播过程中的内容贡献主要由三部分决定：在规定时间内发布微博的频率MF(u )、微博帖子的总数MN(u )和来自其他微博用户的反馈数量MFB(ui )(分享MS(ui )，合作MC(ui)，和赞美MP(ui))。





**社交信任**

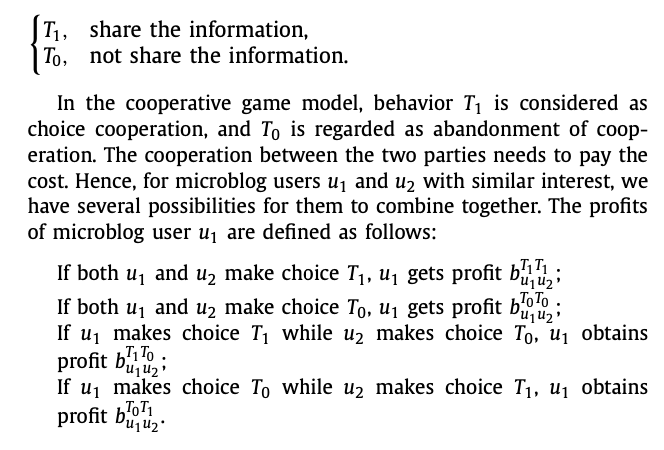
****

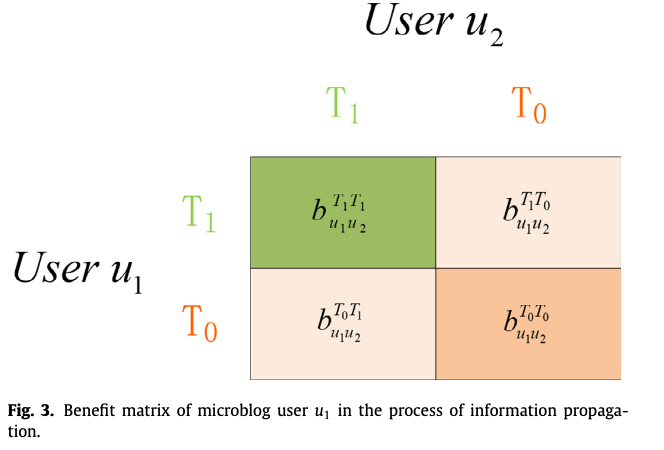
****

**（4）****博弈选择动机**

社会网络中的每个用户总是选择对自己利益最有利的信息进行传播和融合。因此，我们用博弈论作为用户选择传播信息的动机。在信息传播过程中，参与者的动机是保证信息传播规则的关键因素。为了更好地说明微博社交网络中信息传播的方向，我们采用信息传播特征的分享机制（发布、分享、评论、点赞等）来描述信息传播的过程。用博弈论的核心--纳什均衡定理来描述参与者在信息传播过程活动中的动机。

纳什均衡定理假定用户选择的策略是对彼此的最佳回应，也就是说，用户在选择策略时具有一致性。用户在互动过程中实现了平衡状态，没有任何个体可以通过单方面改变最佳策略来增加其收入。因此，纳什均衡定理可以用来判断用户在稳定状态下的决策分布。社会网络中的用户可以被视为进化游戏中的玩家。而每个用户的两种可能行动，即分享或不分享，都对应着两种策略。

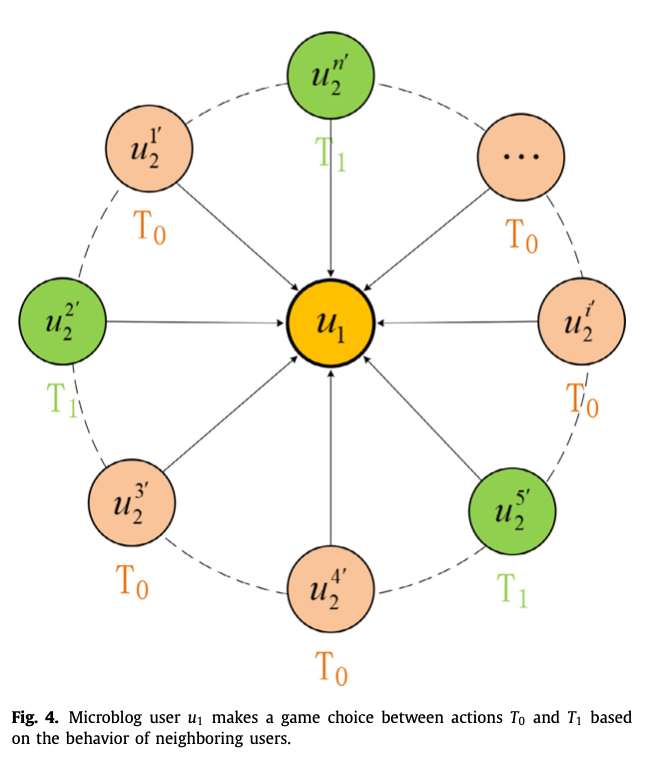




本文将微博社交网络中的信息传播过程视为用户通过博弈选择最佳决策、实现自身利益最大化的过程。而我们将用户与多个邻居节点之间的博弈视为多个双重博弈的组合。微博社交网络中的用户决策具有高度的自主性和独立性。同时，用户的行为决定会受到邻居的影响。因此，我们引入了邻居传播模式和候选传播节点这两个概念。

**邻居传播模式**

给定微博社交网络SN，SN中的顶点u（信息发布者？）和半径r，节点u的邻居传播模式p显示了距离u的r跳内的顶点的邻居，表示为PSrN（u）。SN的邻居传播路径是指距离节点u有r跳的路径。SN的邻居传播树由所有从u开始的、距离小于r的最长路径组成，因此，SN的邻居传播子图由距离u有r跳的所有顶点构建。



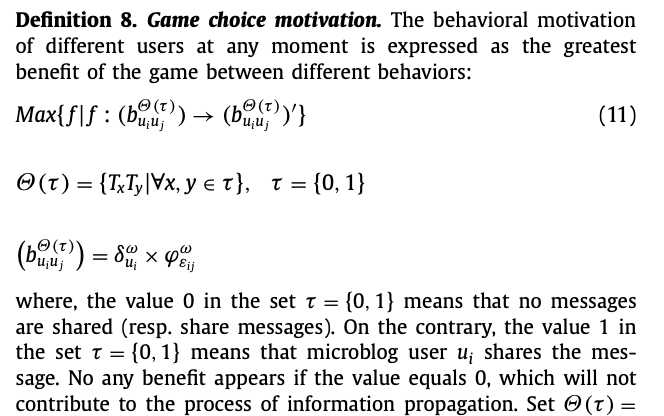
**候选传播者**

图4展现了ui的基于邻居传播模式的候选传播者



**博弈选择动机**

不同用户在任何时刻的行为动机都表现为不同行为之间的最大利益博弈。



其中，集合τ={0，1}中的值0意味着没有信息被分享。相反，集合τ={0, 1}中的值1意味着微博用户ui分享了信息。

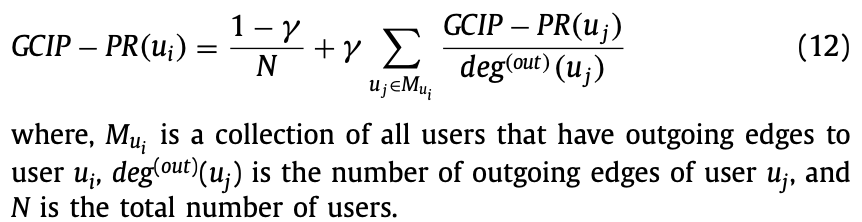
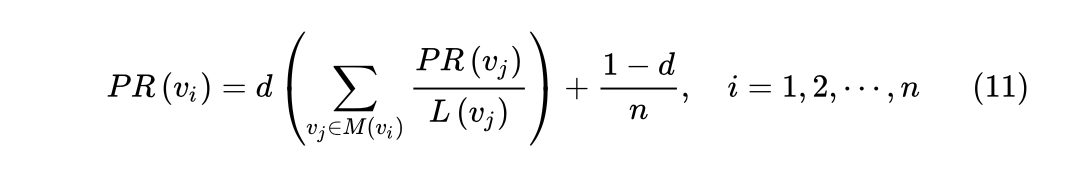
**4. 基于机会、信任和动机的混合社会因素的信息传播模型**

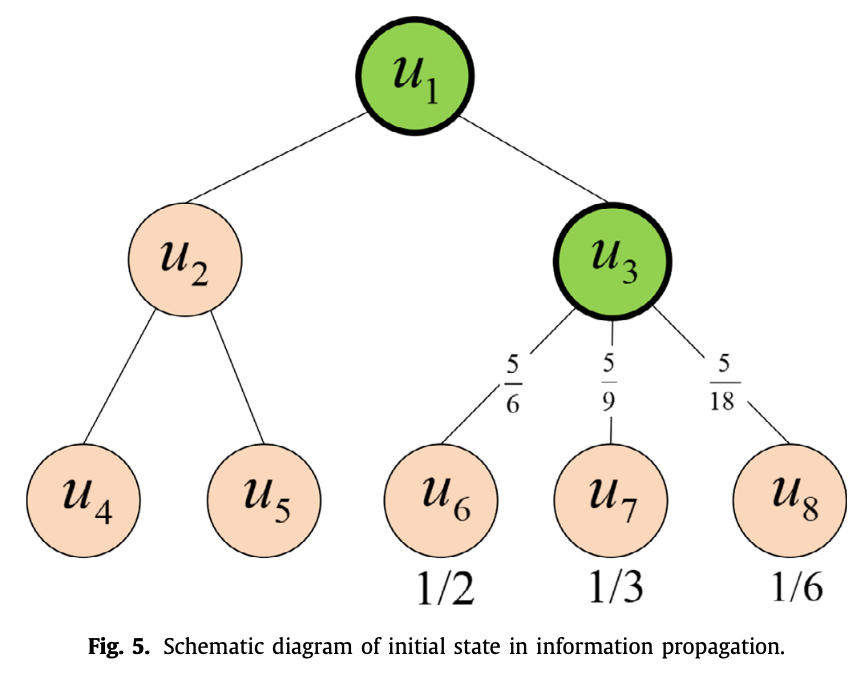
GCIP-PageRank

我们使用GCIP-PageRank来计算稳态概率分布向量值作为每个节点的信息收益。对比不同用户的收益，我们选择收益最大的节点作为上层微博用户的博弈选择结果，也就是信息传播过程中的下一个信息接收者。利用GCIP-Page Rank算法，不断迭代用户的选择过程，建立信息传播模型。

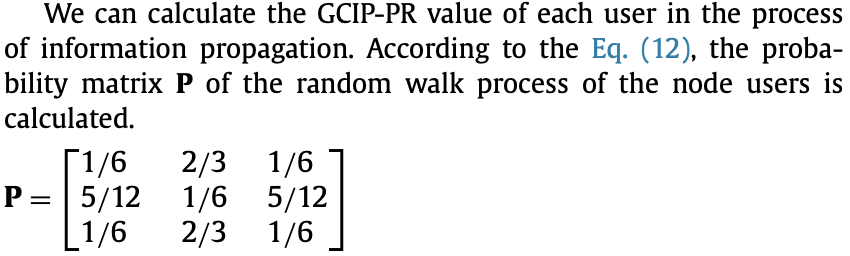
与网页的Page Rank算法不同，GCIP-Page Rank算法的初始状态的概率分布是机会和社会信任的耦合向量值。它反映了用户是否选择传播信息，这是由他接受上层微博用户信息的机会（或可能性）和他自己的社会信任决定的。

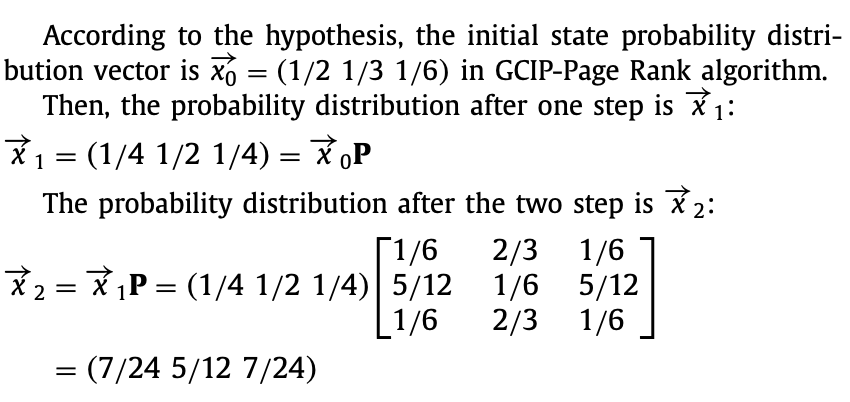
我们将机会衡量为两个相邻节点之间的边的权重，社会信任度为节点的权重。假设节点的初始状态为1，相应的初始状态概率分布向量就是机会和社会信任的耦合值。如图5所示，节点u1是一个信息的发布者，根据我们的规则，它向u3传播信息。u3收到信息后，分别计算u3与其邻居节点u6、u7和u8之间的兴趣相似度。我们把这些值（5/6,5/9,5/18）分别作为u6、u7和u8接收信息的机会量。在GCIP-Page Rank算法中，u3的社会信任度（3/5）和与其邻居节点u , u7和u8的机会度量的耦合值被用作初始状态概率分布向量-→x =（bΘ（τ））=δω ×φω =（1/2 1/3 1/6）。我们假设u3和u6、u7、u8之间的机会关系是双向的，并且γ=0.5。

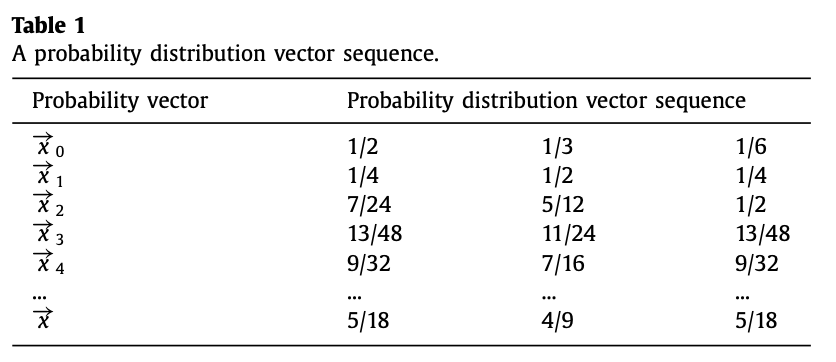


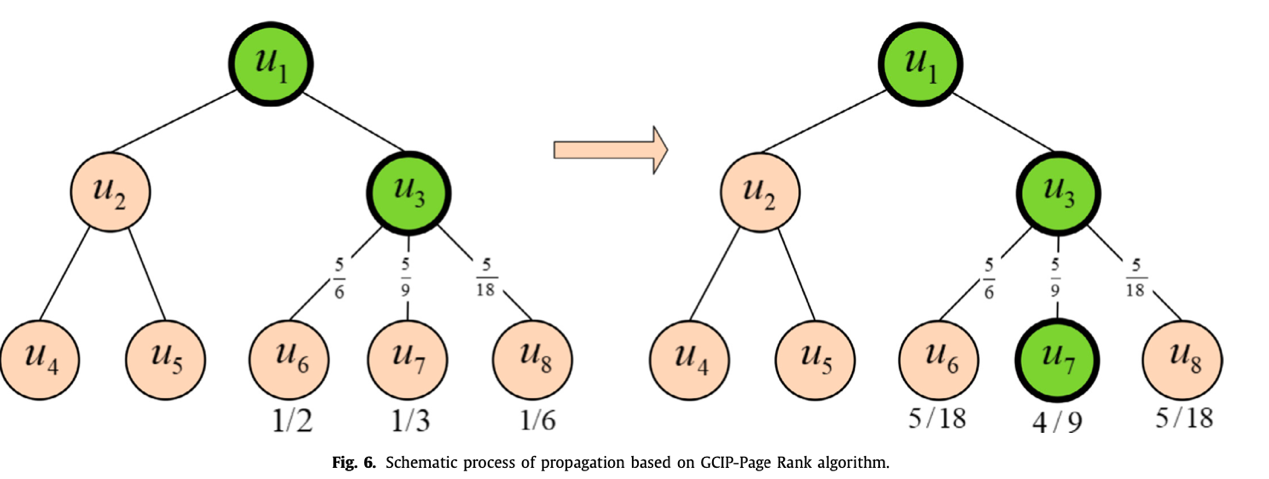


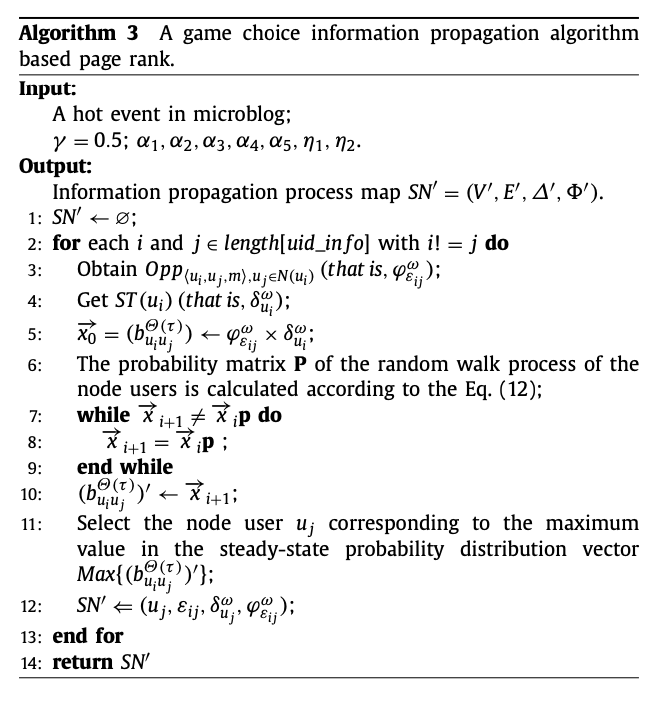
例子：











**5. 实验**

**（1）数据集**

公开数据集+微博爬虫数据

KDD Cup 2012, Track 1公开数据集作为数据集1。此外，从中国最大的在线社交网站新浪微博（http://m.sinna.com.cn）中抓取了三个微博热点事件的数据作为真实数据的样本，分别是数据集2（社会事件1：成都女司机被打）、数据集3（娱乐事件2：白百何出轨）和数据集4（国际新闻事件3：THAAD事件）。通过设置特定的时间间隔，我们得到了三个热点事件在新浪微博平台上的相关微博帖子数量。然后，抓取该微博热点事件的用户ID、发布时间、发布内容、分享次数、评论次数、点赞次数等细节。最后，对搜索返回的结果进行评估，有效文章占返回结果数量的比例分别为95%、96%和98%。

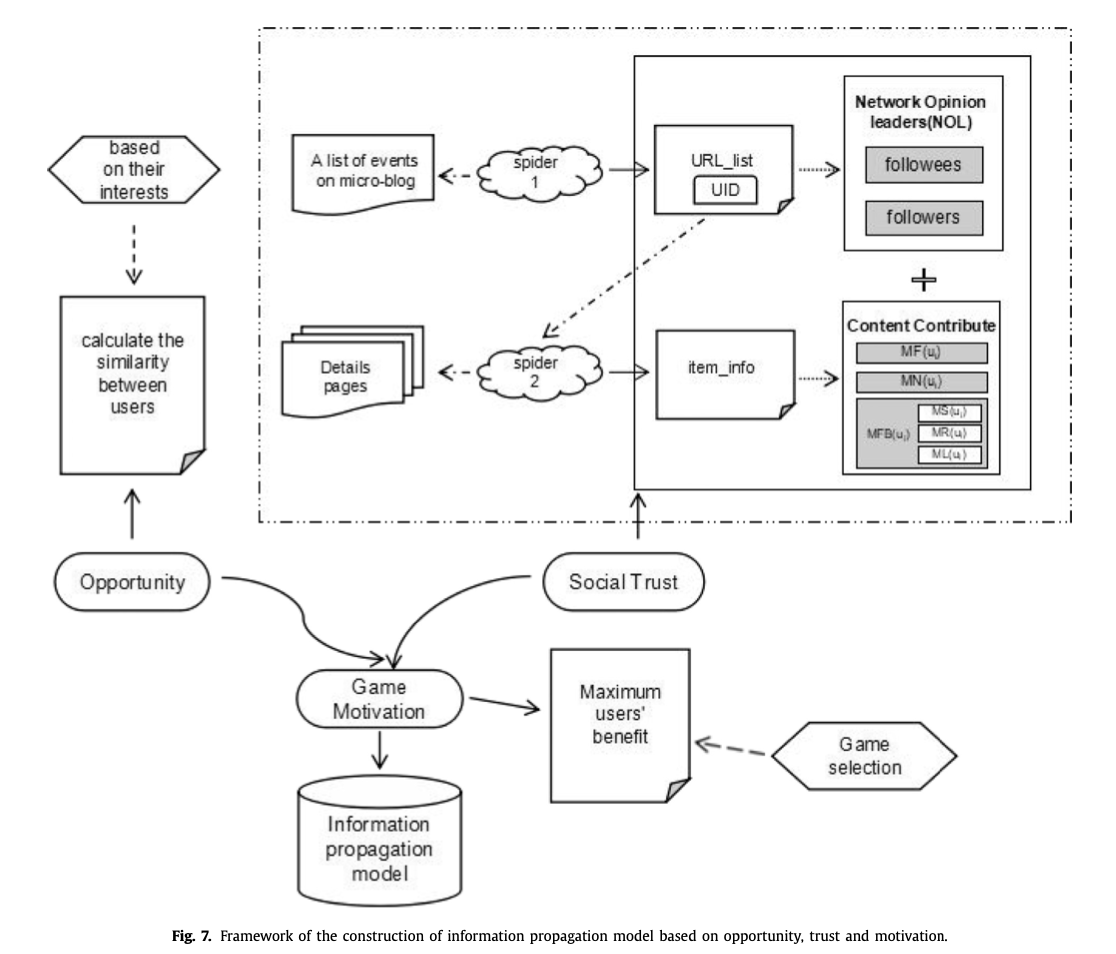
**（2）实验设计**

根据本文提出的基于机会、信任和动机的信息传播模型，研究了微博网络热点事件中信息传播的深度和广度。如图7所示，从以下步骤中具体考虑了信息传播模型的构建框架：

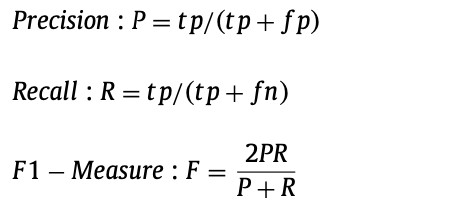
我们从微博社交平台获取数据，获得相关用户对某一热点事件的个人信息和微博帖子。利用自然语言处理方法分析文本数据，提取用户的兴趣特征，并形成兴趣特征集。根据用户的兴趣特征集，使用改进的Tanimoto系数计算相邻两个用户之间的兴趣相似度，该相似度被定义为用户接受某一信息的机会变量。

通过抓取微博的热点事件，找到信息传播中的用户ID。首先，根据用户ID抓取每个节点的关注者和被关注者的数量，从而计算出用户的个人影响力NOL(ui)。然后，根据用户ID抓取用户的基本信息，包括用户的主页信息、微博发帖量MN(ui)、指定时间内的微博发帖量MF(ui)、分享量MS(ui)、评论量MC(ui)和点赞量MP(ui)。因此，根据公式（5）和（6）计算出用户CC（ui）的内容贡献。最后，通过公式（8）得到微博用户ST（ui）的社会信任度。

使用本文提出的GCIP-Page Rank算法，计算每个节点的稳定状态的概率。选择最大概率作为博弈选择的结果，该结果被用作下一步信息传播的对象。其中，将机会测量与社会信任相结合的方法被用作GCIP-Page Rank算法的初始状态概率分布向量。信息传播过程的模型是成熟的，使用博弈进化理论作为用户选择传播信息的动机。

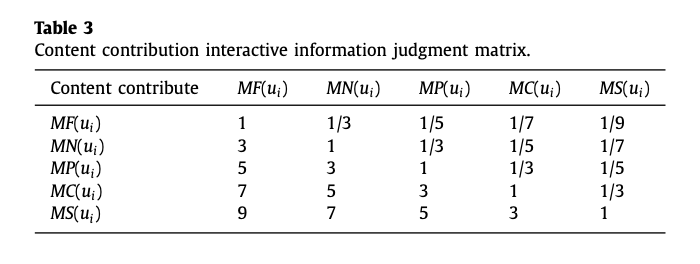


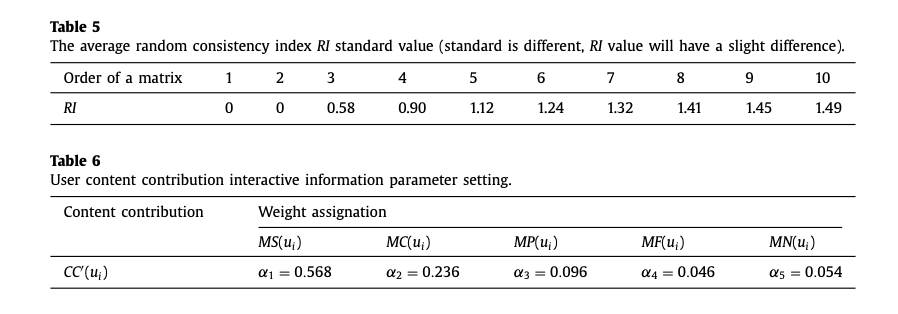
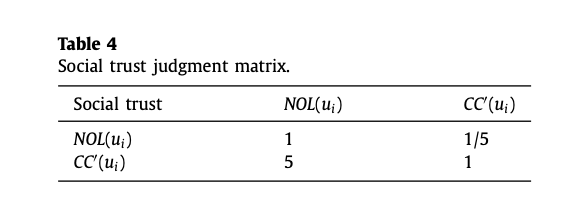
**（3）性能衡量（传播路径预测？）**

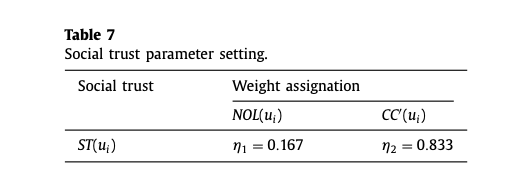


**（4）参数设置**

AHP确定权重







**（5）比较方法**

M1: CT+一般阈值模型，结合CT模型和一般阈值模型的基线方法。

M2: GT（页面排名），GT模型使用页面排名来计算用户的影响力，并使用社会报酬来扩散。

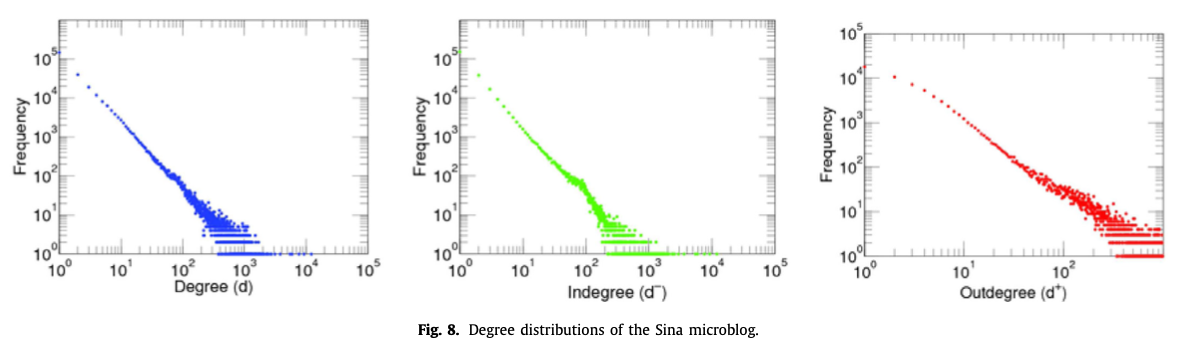
M3: GT (传播级联), GT模型使用传播级联来计算用户的影响，并使用社会报酬来扩散。

M4：GCIP-PageRank，基于Page Rank的游戏选择信息传播。

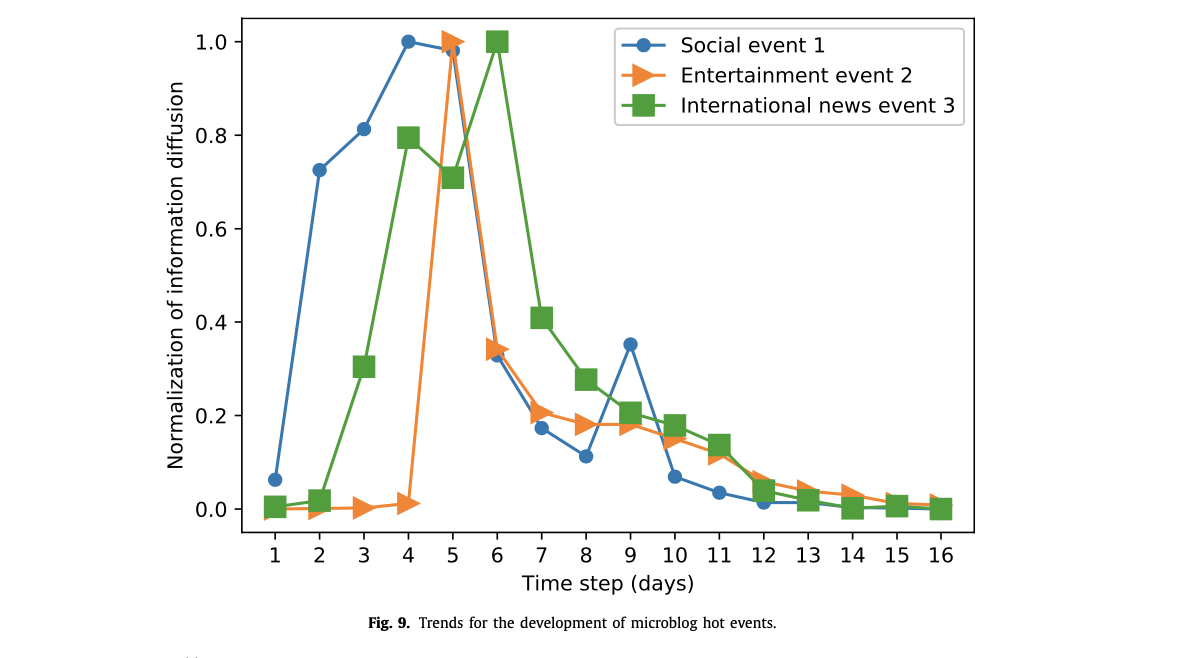
**（6）实验结果分析**

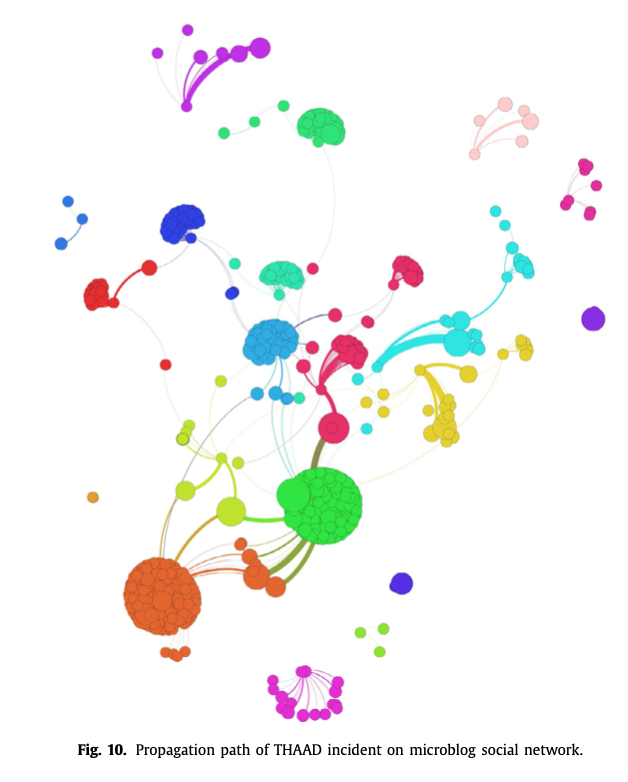
**基本网络描述性分析**

评估了新浪微博社交网络的信息传播过程。新浪微博数据集包含293,437个用户和1,629,665条边，其中的边意味着两个用户之间的联系。在图8中绘制了新浪微博的度分布，包括网络用户的度、不度和外度。从新浪微博的度分布中，我们可以看到，用户的度从1到10000不等，90%的用户的度在1000以下。从度数分布上看，新浪微博网络体现了用户的聚类现象。这反映了微博社会网络中信息的爆炸性传播特征。



我们以2015年至2017年的三个微博热点事件为数据集，对某些信息在微博社交网络中的传播情况进行分析，包括数据集2：成都女司机被打，数据集3：白某出轨，数据集4：THAAD事件。微博热点事件的发展趋势如图9所示。特别是，我们将数据集4（THAAD事件）作为一个重点研究对象。其在微博社交网络上的传播路径如图10所示。圆圈之间的连接厚度（边权重）表示微博用户之间兴趣的相似度。兴趣的相似性越大，连接就越粗。

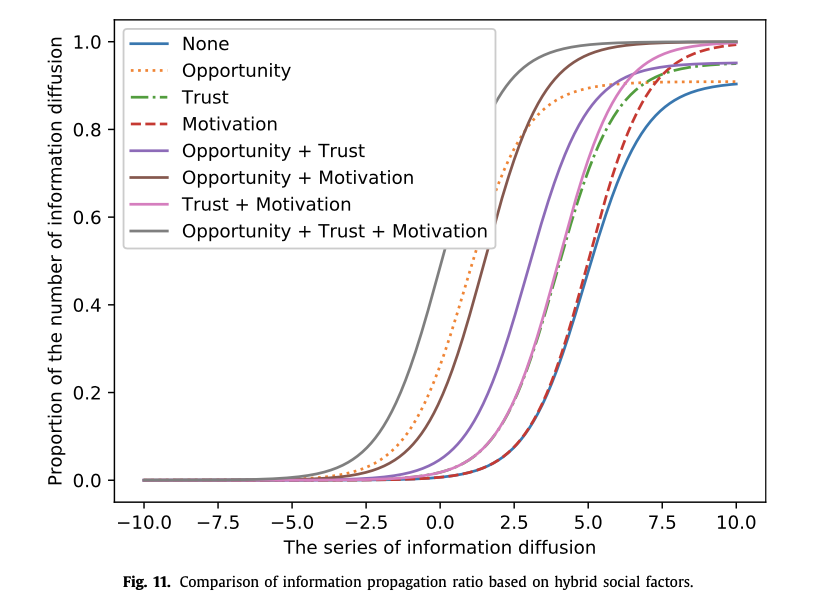




**传播效果分析**

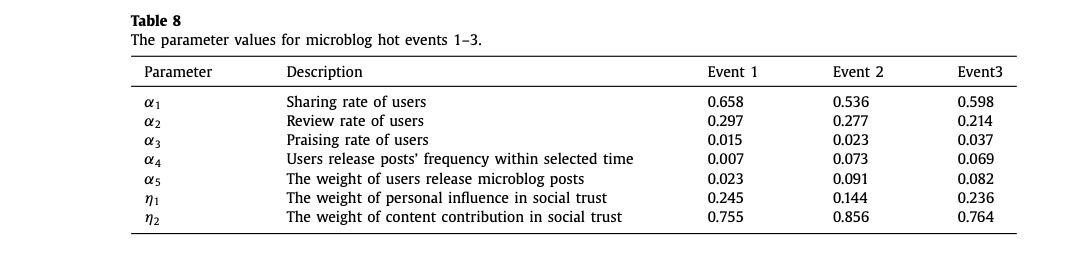
接下来，分析社会因素（机会、信任和积极性）对信息传播的影响。图11是一个显示不同社会因素在微博社交网络中的事件传播比例的图。

根据图11，可以发现，微博社交网络中的信息传播数量与S型曲线相似。同时，事件传播经历了三个阶段：传播接触（0-1系列）、信息信任（1-3系列）和信息传播（3-6系列）。此外，我们可以看到，不同的社会因素对信息传播有着不同的影响角度和程度。进一步说，我们可以观察到三个明显的现象：机会可以促进传播行为，缩短传播接触的时间；社会信任可以加速指定内的传播，缩短信息传播的生命周期；动机影响传播的深度和广度，使重要的信息尽可能广泛地传播给微博社交网络中的每个人。



**算法性能分析**

调整参数：通过分析层次过程（AHP）确定每个参数的估计值，然后通过不断调整参数获得最佳参数，由F1分数来衡量。此外，由于用户的游戏收入是目标函数，我们使用选定的样本值来进行无监督学习，这受到目标函数极值的约束。最后，我们计算出参数的最优值。



从表8中我们可以看出，对于一条信息来说，信息的传播受到两个方面的影响：一是传播信息的人，二是信息内容本身。其中，在某一特定领域，价值，有时也称为信息的意义（即信息在传播过程中的内容贡献）在传播中起着非常重要的作用。这体现在社交网络中用户对信息的行为（如分享、评论和赞扬）。此外，用户分享在社会网络中发挥着重要作用（相当于现实世界中的口碑传播）。

