## **Misinformation influence minimization problem based on group disbanded in social networks（5.26 李昊展）**

基于社交网络中解散的群体的错误信息影响最小化问题



### **基本信息**

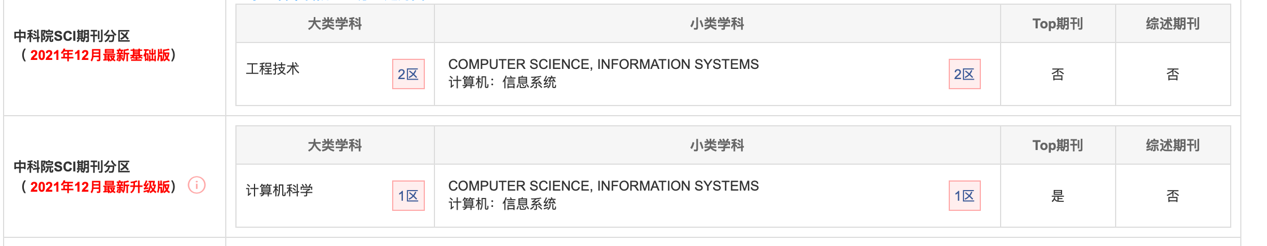
期刊：《Information Science》

发表时间：2021

作者：Jianming Zhu, Peikun Ni, Guoqing Wang, Yuan Li

作者信息：

School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China

School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

### **文章主要研究内容**

在线社交媒体的蓬勃发展改变了人们发布和访问信息的方式。内容的真实性被削弱，社交媒体上的各种错误信息迅速蔓延。在在线社交网络（OSN）中，用户任意组建私人团体/社区，这大大增加了错误信息的曝光率。考虑到群体的回声室效应非常存在，本文研究了OSN中私人群体的解散策略，以尽量减少回声室效应（MSME）作用下的错误信息的传播。给定有向无环OSN G（V;E;C）， C表示一组私人群体，MSME的问题在于从C中选择K组，以便通过解散这些组来最大限度地减少错误信息的传播。针对这个问题，构建了一种贪婪算法，并提出了几种启发式算法来求解目标函数。我们对四个现实世界数据集的实验模拟验证了我们构建算法的有效性。

### **文章内容**

1.问题提出

如今，在线社交网络（OSNs）上一些醒目的错误信息足以演变成网络舆论，影响人们的日常生活，甚至引发社会动荡。例如，在2017年10月，当加州发生毁灭性的丛林火灾时，政府不仅需要救援居民，还需要处理OSNs中关于林地火灾的谣言。2020年1月初，新型冠状病毒肺炎外显子（COVID-19）在中国武汉爆发后，中国一些地区的群众感到恐慌。然而，恐慌主要不是由冠状病毒引起的，而是由各种社交平台上大量的假新闻引起的。 错误信息在OSNs上的广泛传播，严重影响了正常的社会秩序，也对中国政府抗击疫情计划的正确实施造成干扰等等。

在OSNs中，用户不仅可能从有直接联系的朋友那里得到错误信息，还可能从他们加入的团体中得到错误信息。换句话说，用户可能会受到与他们之前建立了联系的朋友的影响，但也可能会受到与他们没有直接联系的团体成员的影响，但他们是私人团体的一部分。用户在OSN上任意组建私人团体/社区，大大增加了错误信息的暴露率，提高了用户之间错误信息的互动频率，扩大了错误信息的传播范围。考虑到这一点，有必要提出一种有效的策略来削弱群体的整体影响力，从而降低错误信息的暴露率和错误信息的传播速度。然而，目前的错误信息控制策略只考虑了用户受其朋友的影响，而忽略了其群体中其他用户的影响。在此基础上，本文考虑了群体对其用户的整体影响，并研究了OSNs中私人群体/社区的解散策略，以尽量减少错误信息的传播。

2.概念定义

给定有向无环社交网络G（V;E;C），其中V是节点集（用户），E表示边集（用户之间的关系），C是一些用户自发形成的私有组/社区。不同的群体在群体活动、群体程度和群体封闭等方面具有不同的特征。接下来，我们给出了群体特征的具体定义。

（1）团体度

团体度分为出度和入度。群体ci的出度ODi被定义为初始点位于Ni，终端点为V\Ni的有向边数。其中N i为群体ci中的节点



群体ci的入度IDi被定义为初始点位于V\Ni，终端点为Ni的有向边数。

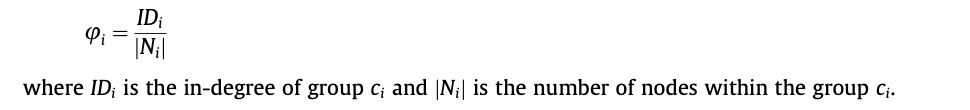


团体度即为



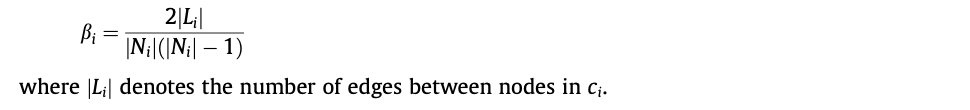
（2）组闭包

OSN中的组闭包主要是指组内节点和组外节点之间信息的平均交互。因此，群ci的群闭包被定义为



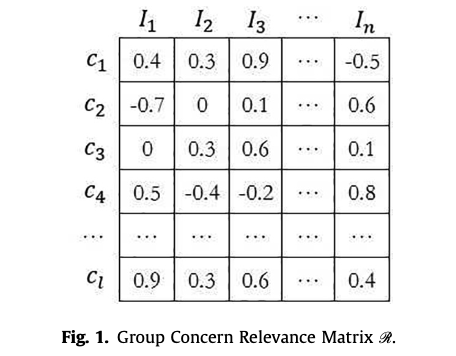
（3）组密度

本文使用组中节点的聚类程度来描述组密度。组越密集，节点之间的信任度越高，节点之间的信任度越高，组中私密的信息节点交换越多，不同节点的同质性就越高。ci的群密度定义为



（4）团体关注相关性

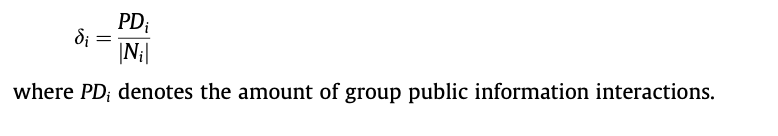
在OSN中，用户可以接收不同类型的信息，但这些不同类型的信息之间存在相关性。相关性的强度取决于信息的类型。ci组中的节点关注一些或几种类型的信息，他们关注的信息也与收到的信息Ij有关。以简化的方式，ci组和收到的信息Ij之间存在关注相关性。不同组和不同类型信息之间的关注相关性矩阵R如图所示。



矩阵R中的元素正值和负值分别表示正和负相关，零表示无关。每行都表示ci组所涉及的信息与不同类型的信息之间的相关性，每列表示Ij类型信息与不同组关注信息之间的相关性。本文为了方便地计算其影响概率，当相关性< 0时，我们取相关性=0。

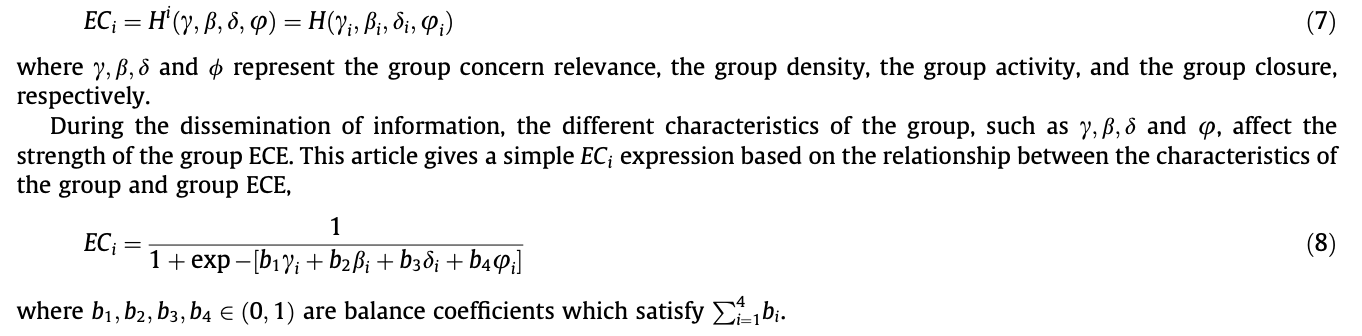
（5）团体活跃度

群组公共信息互动的数量是指群组成员在给定时间段内发送给群组的消息总数。当小组成员高度参与主题讨论时，小组中的公共信息互动总量非常高；相反，当群组中的用户对特定信息不感兴趣，并且几乎没有用户参与讨论时，信息互动的总量非常低。Ci的组活跃度是指在给定时间间隔内ci组中公共信息的相互作用的平均次数。



（6）回声室效应

“回声室效应”的概念起源于1956年，主要用于政治领域的研究。回声室效应是指在相对封闭的环境中（群体、社区等）补充类似的想法或意见，使群体/社区中的大多数人认为这些观点或意见是真实的。不同的结构特征和用户属性使不同的群体表现出不同的回声室效应。



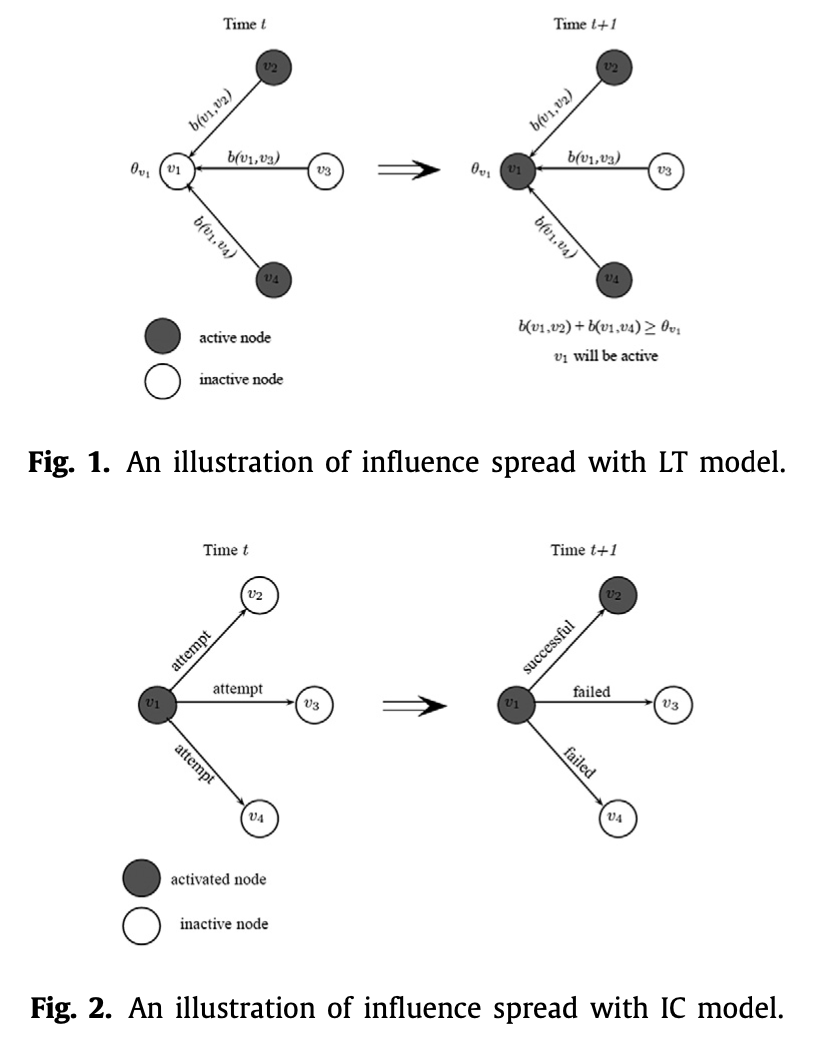
（7）团体解散

团体解散是指破坏其节点自发形成的相对封闭的环境，使组中的节点无法在封闭环境中发送和接收信息，并且只能通过节点之间的私有连接（边缘）传播信息。实际上，相对封闭的环境被一些方法摧毁，例如将组中的所有节点静音或与组所有者进行干预以自发解散组。

3.扩散模型

独立级联模型（ICM）和线性阈值模型是信息扩散领域的两个最经典的模型。在社交网络中，用户经常转发或共享他人发布的信息，但这些信息通常来自用户的近邻（朋友），因此信息在用户之间的传播会产生信息级联。此外，尽管人们的行为或决定总是受到增量激活用户数量的强烈影响，但ECE集团的影响也被视为一个独立事件。因此，本文基于ICM研究了MSME（最小化错误信息传播）的问题。

独立的级联模型，其中扩散过程在离散时间戳中触发级联激活。在模型中，每个边缘e =（u，v）都与节点v被节点u成功激活的概率有关。在时间戳t时，激活节点u尝试使用以概率激活节点v。如果激活成功，节点v将在时间戳t+1时变得活跃，否则它将保持不活跃。无论激活成功还是失败，节点u都没有机会在节点v上再次尝试，直到过程结束。这个过程将结束，直到不再激活节点。



4.问题表述

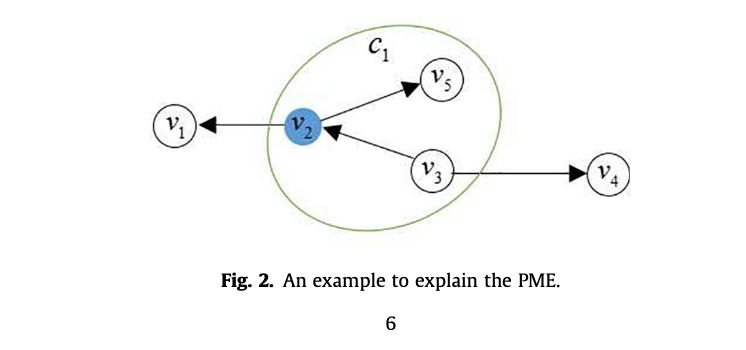
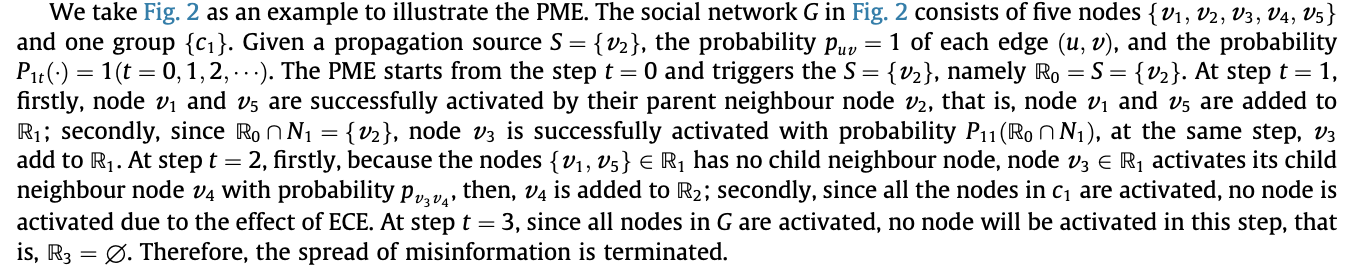
（1）PME问题（回声室效应下的错误信息传播模式）

ci组中的非活动节点v不仅受到其已激活的父邻居节点的直接影响，还受到组ECE的影响。

激活规则：

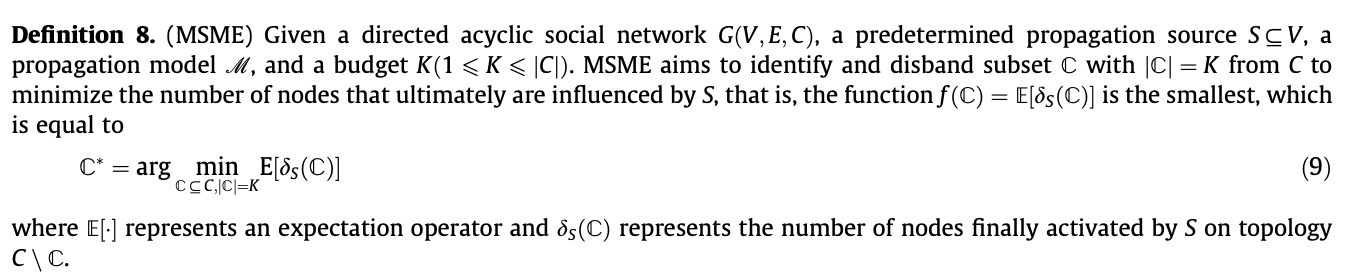
节点u只有一次机会以概率puv激活其非活动子邻居v；

由于回声室效应，如果Rt1 \ Ni不是空的，它只有一次机会以概率为Pit激活组内的非活动节点。



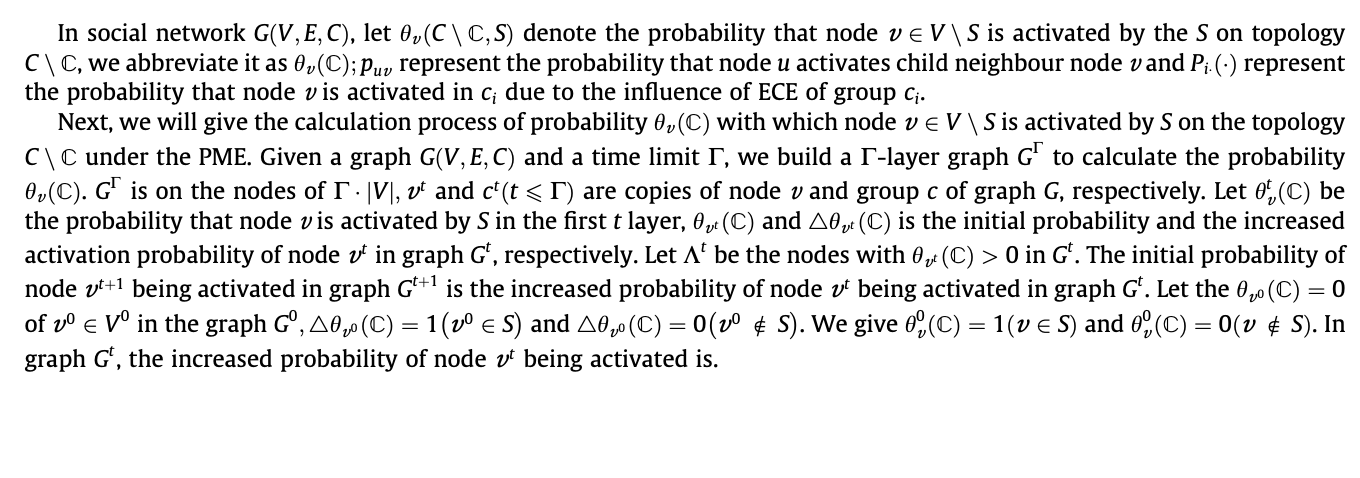
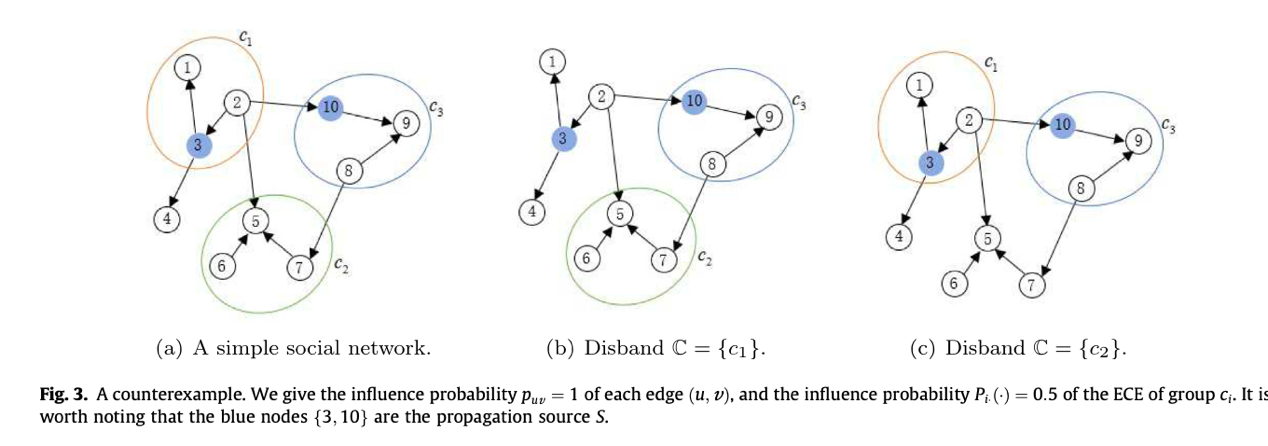
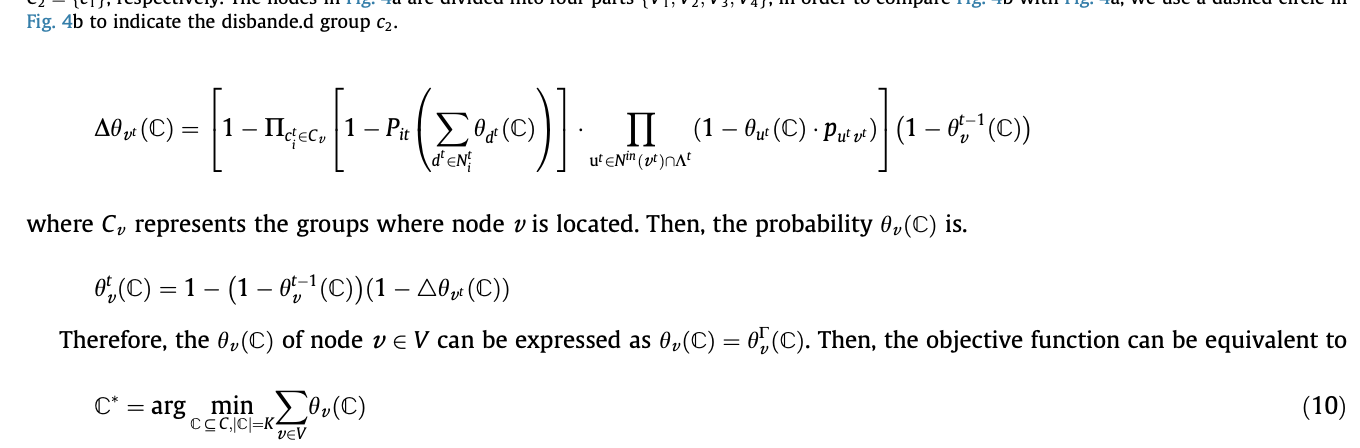
（2）MSME问题（最小化错误信息传播）

本文考虑了ECE在OSN中的影响，并提出了MSME的问题。本文的目标是识别和解散一些组，以尽量减少最终受错误信息影响的节点数量，即减少受组ECE影响的节点数量，以便大多数节点可能只从其激活的父邻居节点接收错误信息。MSME的问题被定义为：



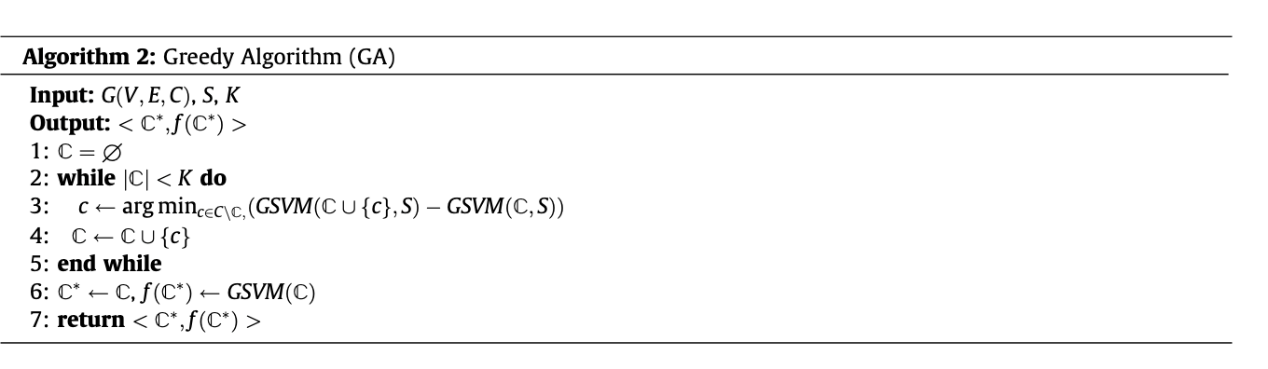
5.解决方法

计算整体激活概率

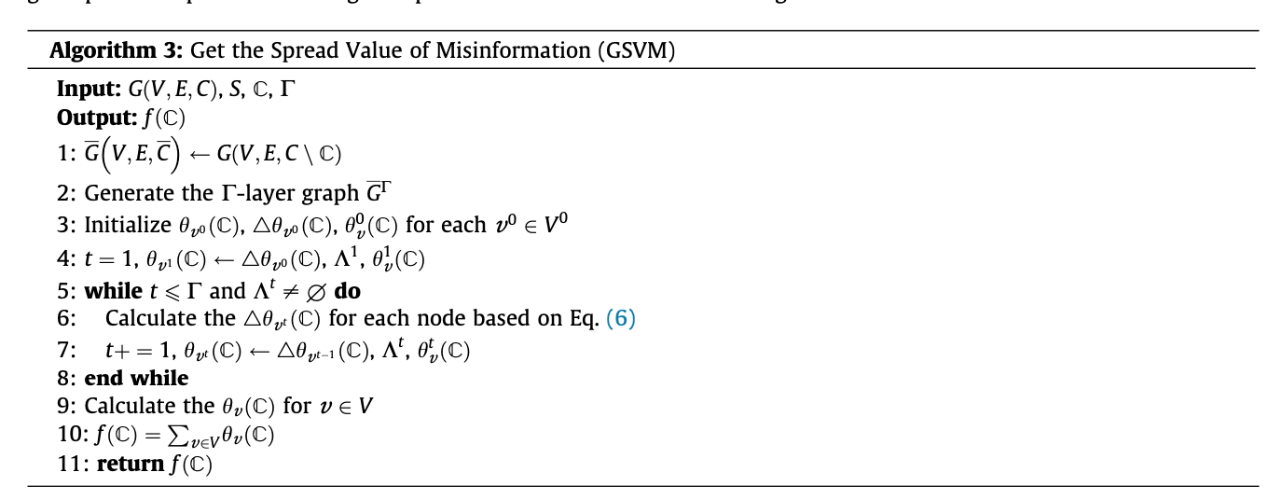
** **

（1）贪婪算法

遍历所有解散组的结果，每次选择解散后效果最好的一组，直至达到K组



错误信息传播值计算

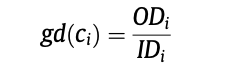


（2）启发式算法（目的：降低回声室效应）

在现实世界中，OSN非常大，包含许多组，因此使用GA解决目标函数需要大量的内存和时间。本文提出了几种启发式算法，用于基于群度、群密度和群闭包等群属性求解目标函数。与GA相比，这些启发式算法具有运行时间更短、内存使用更少的优点。

团体度评分算法（SGDA）

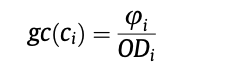
组度表示组ci从组外部接收并向外部发送错误信息的错误信息通道的总和。IDi越大，它收到的错误信息范围就越广，群体成员思想的两极分化就越小，即ECE的影响就越弱。ODi越大，该群体向外部世界传播错误信息的能力就越强，也就是说，将ECE传播到群体外节点的能力就越强。



SGDA每次都选择gd最高的组来屏蔽或解散，以尽量减少最终被S激活的用户数量。

组闭包评分算法（SGCA）

群闭包表示Ni和V\Ni之间的平均错误信息相互作用。节点Ni与组外节点的错误信息交互越少，从组外获得的错误信息就越少，就越容易受到组中错误信息的影响。当组内外的错误信息交互更多时，组内的节点获得的错误信息就越多，然后组中的用户将组中收到的错误信息与从组外获得的错误信息进行比较，然后决定是否接受。



SGCA每次都选择gc低的组来屏蔽或解散，以尽量减少最终被S激活的用户数量。

组密度评分算法（SGEA）

群密度表示节点之间的聚集程度。在ci组内。群密度越大，组中节点之间的私密错误信息互动就越多，ECE的影响就越弱。当组更密集时，组中用户之间的私有交换越多，节点之间的信任度就越高，不同节点的同质性就越高。



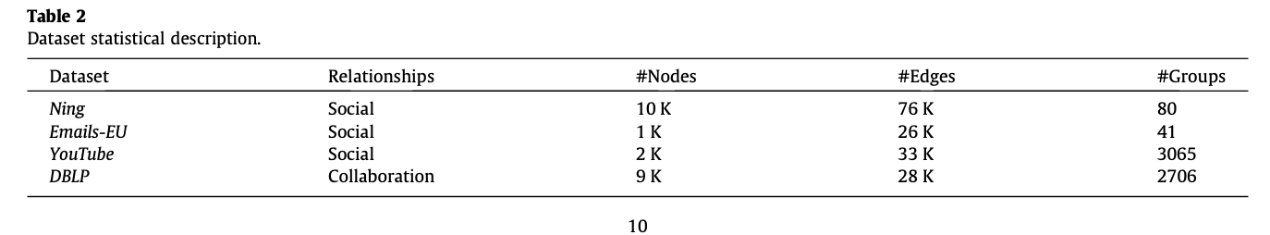
我们迭代选择K组解散或屏蔽的最低de，以尽量减少最终接受错误信息的用户数量

随机算法（RDA）

随机选择k组

6.实验

（1）数据集

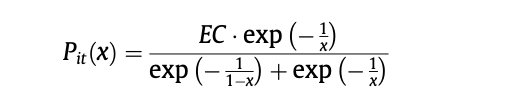


（2）设定影响概率

一种按0.1固定值设定、另一种按照{0.2、0.05、0.01}随机分配

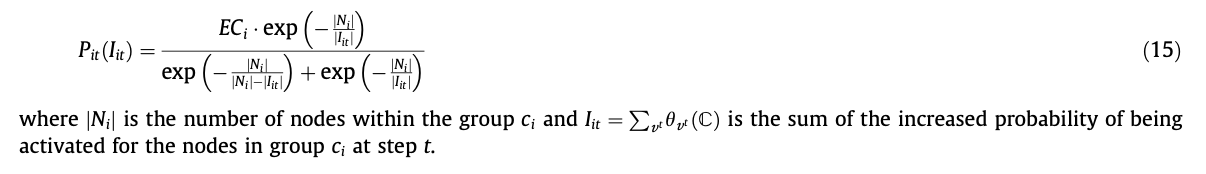
（3）激活概率

由于群体Ci回声室效应作用下的影响概率

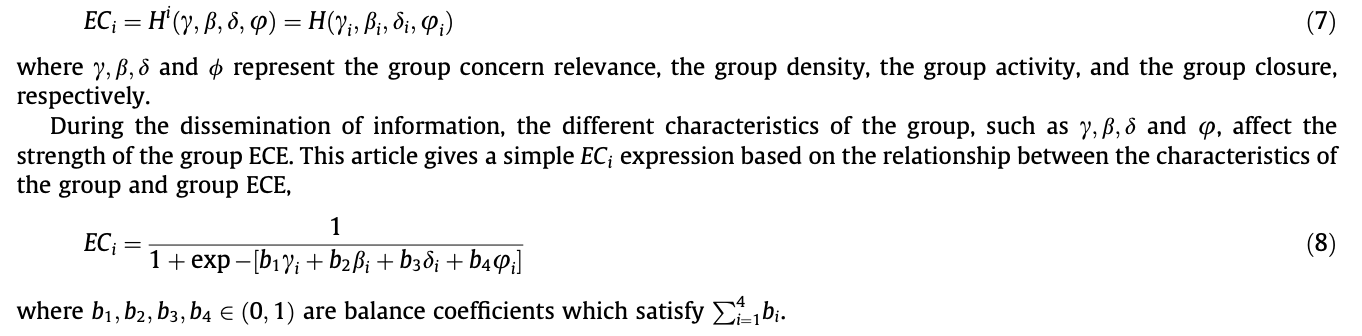
****

x是组内在时间t被新激活的节点

由于每个群体都有不同的特征和属性，不同的群体对错误传播有不同的影响，也就是说，不同的群体有不同的EC值。因此，影响概率Pit可以表示为

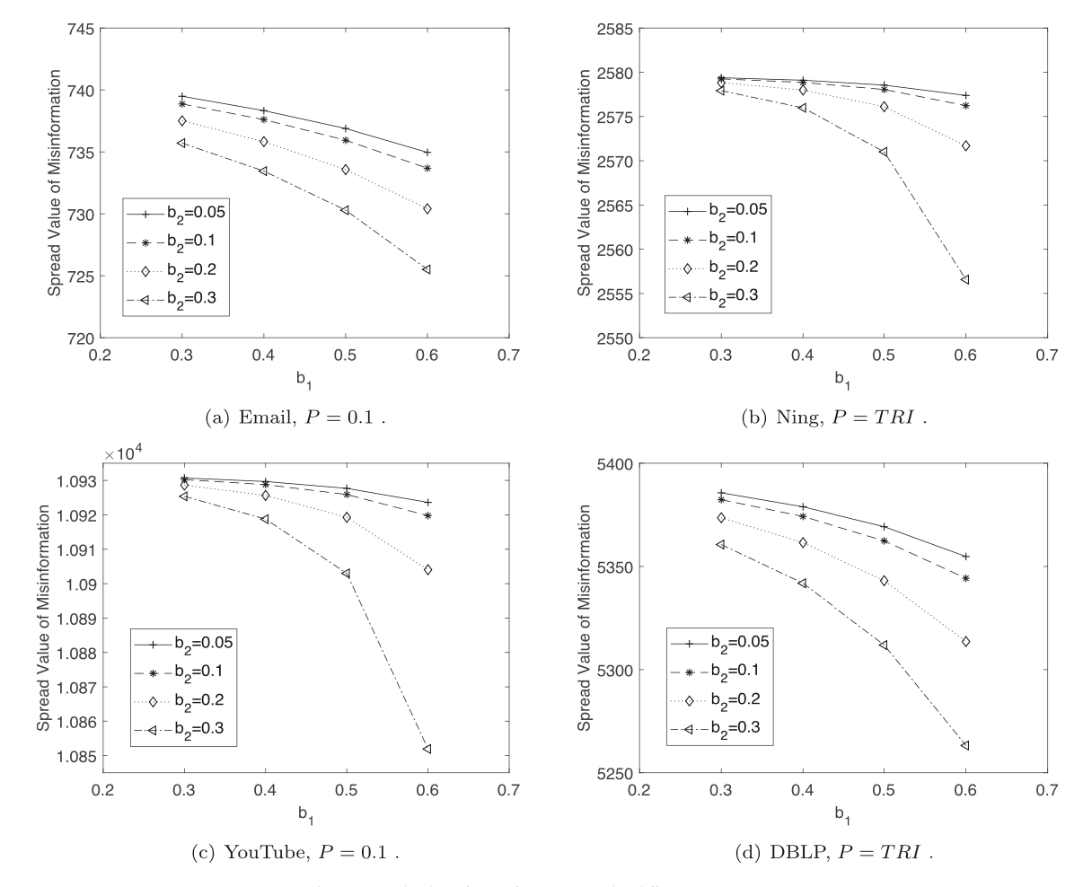
****

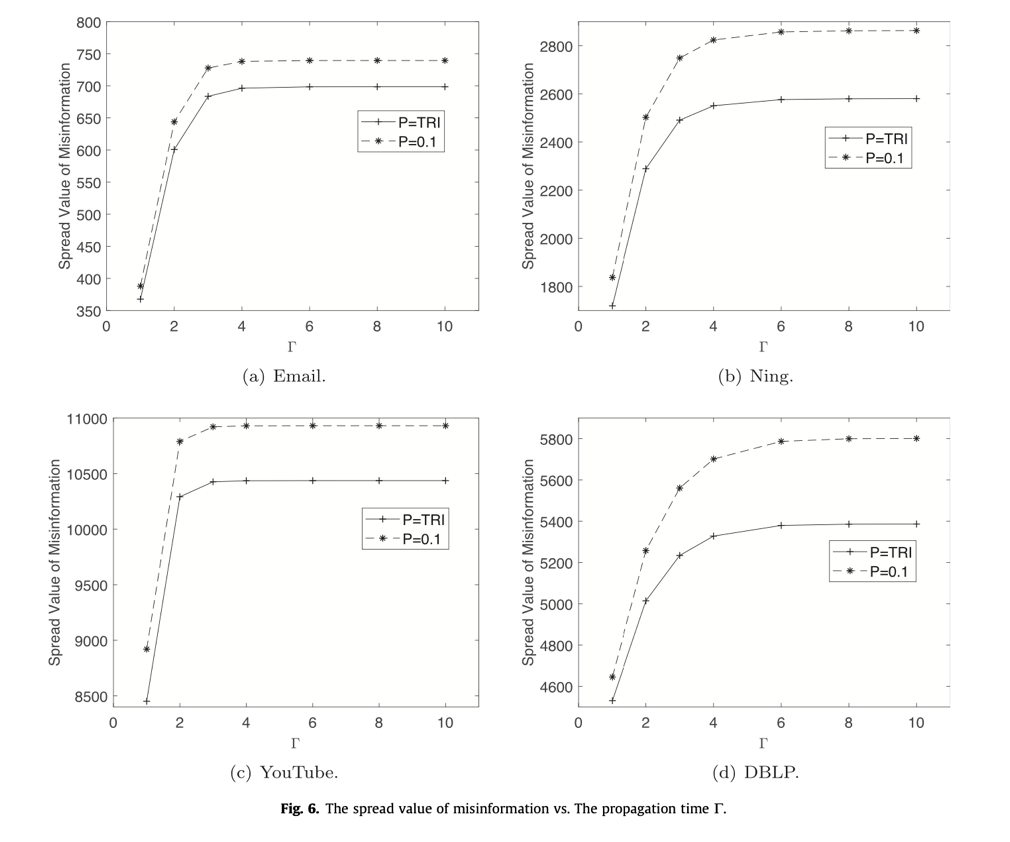
（4）参数设定



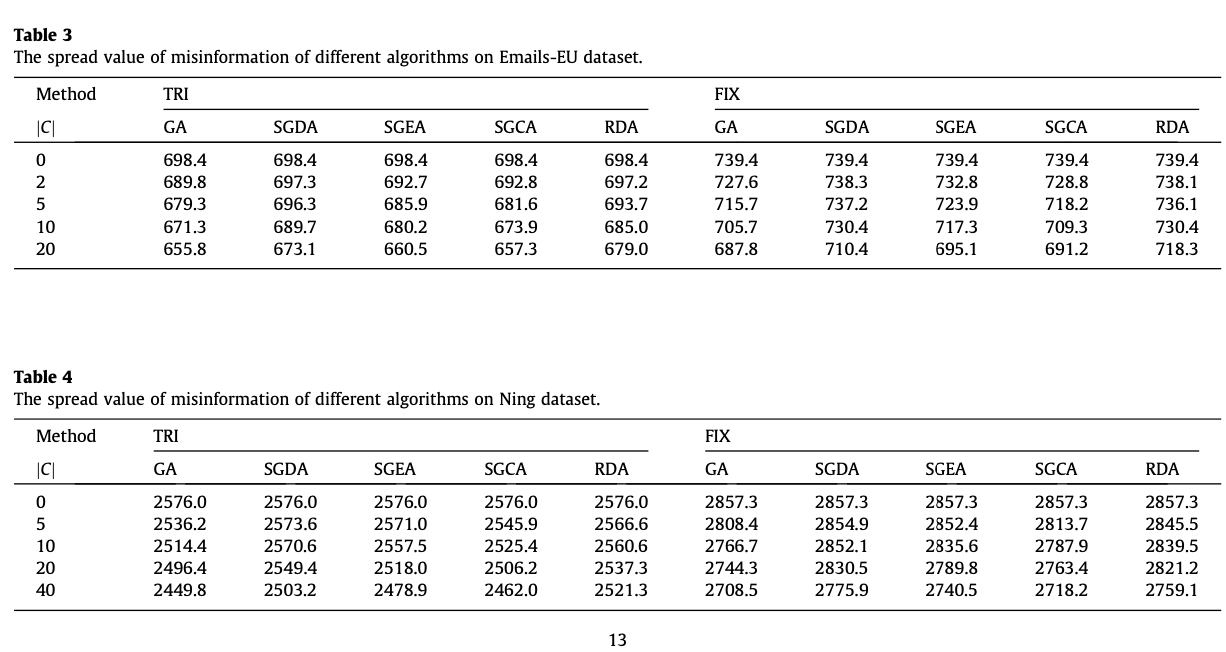
在OSN中，也很难获得给定时间C中组用户的公共错误信息交互量，因此本文不考虑组中公共错误信息交互的数量，即系数b3等于0。在以下实验中，我们设置了b1 =0.3；b2 =0.05；b4=0.65和时间步为5。

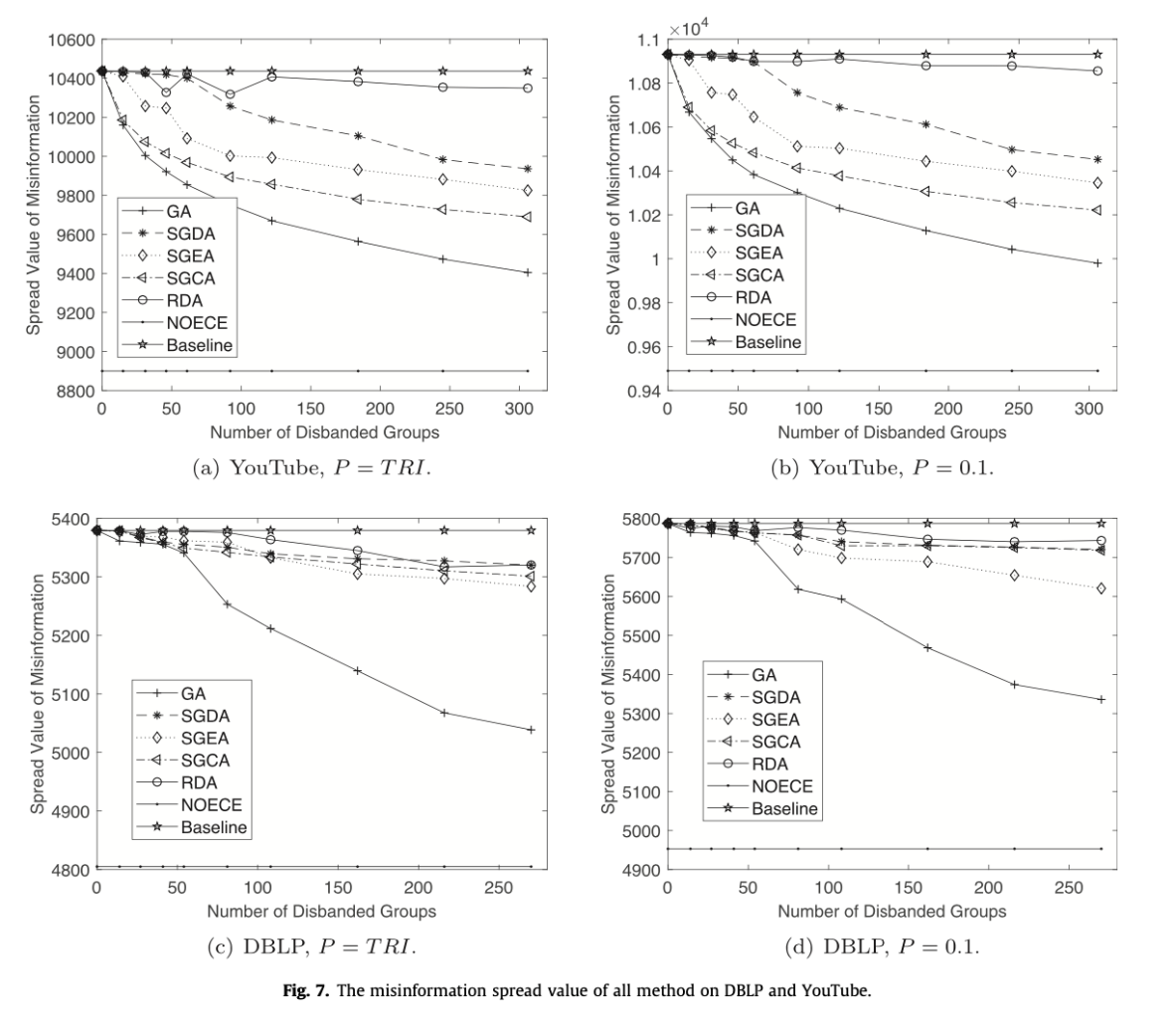
（5）参数设定实验





（6）算法对比





baseline表示当没有团体解散时，错误信息的传播价值，NOECE在不考虑ECE的情况下表示错误信息的传播价值。我们将该组的总体影响定义为基线和NOECE之间的差异。