



计算机科学

Computer Science

ISSN 1002-137X, CN 50-1075/TP

《计算机科学》网络首发论文

题目：网络结构影响传播效果的解耦分析
作者：崔允松，吴晔，许小可
收稿日期：2022-09-13
网络首发日期：2023-04-10
引用格式：崔允松，吴晔，许小可. 网络结构影响传播效果的解耦分析[J/OL]. 计算机科学. <https://kns.cnki.net/kcms/detail/50.1075.tp.20230407.1930.034.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

网络结构影响传播效果的解耦分析

崔允松¹ 吴晔^{2,3} 许小可¹

1 大连民族大学信息与通信工程学院 辽宁 大连 116000

2 北京师范大学计算传播学研究中心 广东 珠海 519085

3 北京师范大学新闻传播学院 北京 100875

(337579873@qq.com)

摘要 如今越来越多的人通过社交网络传播信息，在线社交网络逐渐改变了人们交换信息的方式，因此基于在线社交网络研究信息传播效果的影响因素受到众多研究人员的关注，尤其是网络结构对信息传播效果的影响。以往研究中大都强调某种网络统计量对信息传播效果的影响，但各个网络统计量之间相互耦合的现象是客观存在的，某种网络统计量的改变可能导致其他网络统计量同步改变，从而可能影响最终的传播效果。提出一个解耦的零模型框架，通过零模型解耦不同网络统计量间的耦合作用，然后使用 SIR 传播模型仿真实验来分析网络统计量在没有耦合作用下对信息传播效果的影响，最后使用线性阈值模型仿真实验验证 SIR 模型的实验结论在社会强化中的适用性。通过 Facebook 和 Twitter 实证网络的传播模型仿真实验表明：网络的平均最短路径是影响信息传播速度和信息传播范围的主要因素，聚类系数是影响信息传播范围的次要因素。

关键词：传播效果；零模型；平均最短路径；聚类系数

中图法分类号:TP301

Decoupling Analysis of Network Structure Affecting Propagation Effect

CUI Yunsong¹, WU Ye^{2,3} and XU Xiaoke¹

1 School of Information and Communication Engineering, Dalian Minzu University, Dalian, Liaoning 116000, China

2 Research Center of Computational Communication, Beijing Normal University, Zhuhai, Guangdong 519085, China

3 School of Journalism and Communication, Beijing Normal University, Beijing 100875, China

到稿日期：2022-09-13

返修日期：2022-11-22

基金项目：国家自然科学基金(62173065)；辽宁省自然科学基金(2020-MZLH-22)

This work was supported by the National Natural Science Foundation of China (62173065) and the Natural Science Foundation of Liaoning Province, China(2020-MZLH-22).

通信作者：许小可 (xuxiaoke@foxmail.com)

Abstract: As more and more people spread information through social networks, online social networks have gradually changed the way people exchange information, so the influencing factors of information dissemination effect based on online social networks have attracted the attention of many researchers, especially the influence of network structure on the effect of information dissemination. In previous studies, most of the research has emphasized the influence of a certain network statistic on the effect of information dissemination, but the phenomenon of coupling between various network statistics is objective, and the change of a certain network statistic may lead to synchronous changes of other network statistics, which may affect the final propagation effect. In this study, a decoupled zero model framework is proposed, which decouples the coupling between different network statistics through the zero model, and then uses the SIR propagation model simulation experiment to analyze the influence of network statistics on information propagation effect without coupling, and finally uses linear threshold model simulation experiments to verify the applicability of the experimental conclusions of the SIR model in social reinforcement. Simulation experiments of the propagation model of Facebook and Twitter empirical networks show that the average shortest path of the network is the main factor affecting the speed and scope of information dissemination, and the clustering coefficient is the secondary factor affecting the scope of information dissemination.

Keywords Propagation effect, Null model, Average shortest path length, Clustering coefficient

1 引言

美国学者 Lasswell 在 1948 年提出了 5W 模式,即构成传播过程的 5 种基本要素——传播者、信息、渠道、接受者、传播效果。其中,传播效果是 5W 模式里最重要的一部分。传播效果分析主要集中在大众传播在改变受众固有立场、观点上有多大的威力上,但也涉及了大众传播对社会及文化所造成的影响,因此传播效果分析一直都是传播研究领域中最长、最有现实意义的研究内容。

在传播的过程中,对于正能量的信息,说服力越高,代表传播效果越好。我们一般认为拥有权威的传播者比缺乏权威的传播者更有说服力^[1],但有研究发现说服力信息成功通常取决于传播者和受众力量之间的互动,高权威传播者会产生可以说服比较权威的受众的信息,低权威传播者会产生说服权威低的受众的信息^[2],这说明了信息导向与受众力量的匹配可以提高信息的说服力^[3]。文献[4]在复杂网络上基于网络指标识别有影响力的节点进行传播动力学研究,研究发现基于节点介数中心性的目标免疫策略最适用于标准无标度网络,而基

于节点度的目标免疫策略最适用于其他非标准无标度网络。文献[5]表明传播渠道类型确实会对传播结果产生影响,但这种影响取决于信息(是否与受众一致)和传播者(是否值得信任)。由此可以了解到现有研究对传播者、内容、受众等对信息传播效果的影响分析得比较透彻,但是对传播渠道对信息传播效果影响的理解并不像其他因素那样深刻。

随着互联网的普及,在线社交网络逐渐成为最受欢迎的传播渠道,因此在线社交网络对信息传播效果的影响越来越引起研究者的重视,尤其是网络结构对信息传播的影响。文献[6-8]基于在线社交网络进行传播动力学研究,发现高聚类网络可抑制信息传播范围,之后又有研究发现行为在高聚类网络中传播更快^[9]。Xing 等发现网络聚类系数的降低和平均度的升高可以增加信息传播范围,平均最短路径增长会降低信息传播速度^[10]。Xu 等发现网络度分布熵越小,谣言越容易传播^[11]。Zhou 等研究表明网络的平均最短路径长度是影响传播速度的关键性因素^[12]。这些研究表明社交网络的传播效果受多种网络统计量影响。在现实社交网络中,多种

网络统计量之间往往存在着很强的耦合作用, 一种网络统计量变化的同时会导致其他网络统计量的同步变化。现有研究中主要强调某种网络统计量的连续变化对信息传播效果的影响, 并没有说明某种网络统计量的变化会导致其他网络统计量变化从而可能使信息传播效果呈现出不同的结果。本研究的目的是分析一种网络统计量在没有与其他网络统计量耦合作用下对信息传播效果的影响。

零模型是故意排除兴趣机制模式的生成模型, 作为一种流行的分析工具, 由于其具有出色的探索图的非凡特性的能力, 因此被广泛应用于复杂网络传播动力学研究中^[13]。在网络科学中, 零模型主要用于比较量化复杂的网络属性, 如社团结构^[14]、路由效率^[15]、模式检测^[16]等, 它可以揭示一些在某种情况下无法直接量化的重要网络属性指标。本文提出一种解耦网络统计量之间相互作用的零模型框架, 其可以在保持一种网络统计量不变的同时使得另一种网络统计量连续变化, 从而应用零模型定性、定量去分析结果。

本文将从网络结构的微观尺度特性(度分布、联合度分布、聚类系数等)和宏观尺度特性(平均最短路径)两类网络统计量展开研究。首先, 基于 Facebook 和 Twitter 实证网络分别构造解耦零模型和不考虑耦合作用的零模型; 然后, 在 SIR 模型上做零模型的仿真传播实验, 通过比较解耦前后零模型的传播曲线与原始网络传播曲线之间的差异辨别影响传播的网络统计量; 接下来, 根据解耦零模型的传播实验观察某个统计量单独连续改变对信息传播效果造成的影响, 从而得出实验结论; 最后, 使用线性阈值模型仿真实验验证 SIR 模型得到的结论在社会强化中的适用性。实验结果表明: 网络的平均最短路径是影响信息传播速度和信息传播范围的主要因素, 聚类系数是影响信息传播范围的次要因素。

2 研究方法

2.1 解耦不同尺度网络统计量的零模型构造

在静态无权无向网络中, 常用的零模型置乱方法是随机断边重连^[17]。该方法主要是在原始网络上随机断开原有连边后重连此前不存在的边, 使原始网络尽可能地随机化。因此, 随机断边重连构造零模型是将原始网络随机化的过程。本研究中应用了随机断边重连的方法来构建零模型^[18]。

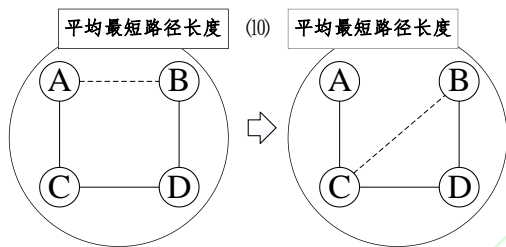
2.1.1 保持平均最短路径基本不变的微观尺度特性零模型

网络的微观尺度特性主要包括平均度、度分布、匹配系数和聚类系数等涉及单个或少量几个节点之间的网络统计量。平均度是指所有节点度的总和的平均数。度分布是指网络中节点度的概率分布。匹配系数是衡量度相近的节点之间互相连接的倾向性, 若网络的匹配系数大于零, 则该网络为同配网络, 反之该网络则为异配网络。聚类系数是衡量一个网络中节点聚集程度的系数, 也即衡量一个节点的两个邻居节点相连的概率。0-3 阶零模型是保持微观尺度特性的一种零模型框架, 随着零模型阶数的增加, 保持的微观尺度特性种类会增多。在各阶零模型保持微观尺度特性的过程中, 宏观尺度特性也会随之发生变化, 那么就不能明确地判断网络的某些非平凡行为到底是某种微观尺度特性的变化所导致还是宏观尺度特性的变化所导致。为解决此问题, 本文提出保持平均最短路径基本不变的微观尺度特性零模型, 其中平均最短路径是一种重要的宏观尺度特性网络统计量。由于该统计量是对全局网络的描述, 即使几条边的变化也会导致整个网络的平均最短路径发生变化, 很难保证在大规模的断边重连后网络的平均最短路径完全不变, 因此本文采用近似的方式来保持平均最短路径基本不变。

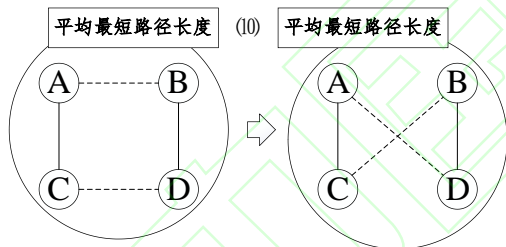
(1) 保持平均最短路径基本不变的 0-3 阶零模型

首先介绍不考虑平均最短路径变化的 0-3 阶零模型^[19]的定义: 0 阶零模型(与原始网络具有相

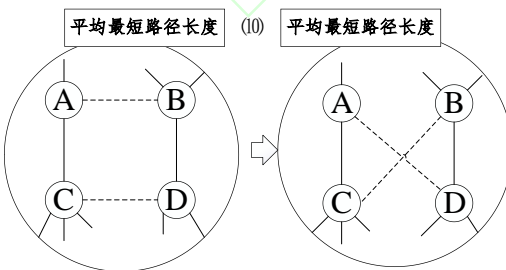
同的平均度); 1 阶零模型 (与原始网络具有相同的度分布); 2 阶零模型 (与原始网络具有相同的联合度分布, 联合度分布是指每条边两端连接节点的度值数目或概率); 3 阶零模型 (与原始网络具有相同的联合边度分布, 具有相同的联合边度分布就是要求零模型网络和原始网络具有相同的开三角形和闭三角形分布)。保持平均最短路径基本不变的 0-3 阶零模型在上述定义的基础上, 多了保持平均最短路径基本不变的条件。图 1 给出了保持平均最短路径基本不变的 0-3 阶零模型的构造过程。



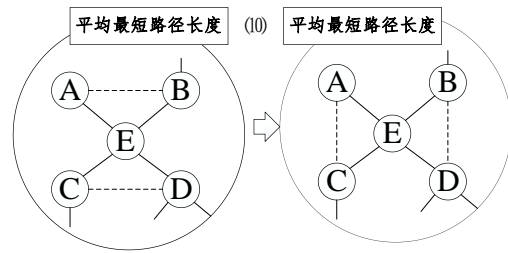
(a) 保持平均路径基本不变的 0 阶零模型



(b) 保持平均路径基本不变的 1 阶零模型



(c) 保持平均路径基本不变的 2 阶零模型



(d) 保持平均路径基本不变的 3 阶零模型

图 1 保持平均最短路径基本不变的 0-3 阶零模型的构造过程图

Fig. 1 Construction process diagram of keeping the average shortest path basically unchanged 0-3 order null model

保持平均最短路径基本不变的 0 阶零模型是 0-3 阶零模型中随机化程度最高的零模型, 只需要与原始网络具有相同的节点数和平均度。0 阶零模型的构造过程如图 1(a)所示, 首先随机选择一条边 AB , 然后随机选择两个不相连的节点 A 和 D , 在断边重连之前计算网络的平均最短路径。断开边 AB 的同时, 在节点 A 和 D 之间添加一条边, 然后计算新网络的平均最短路径。若断边重连前后网络的平均最短路径基本保持不变, 那么说明本次断边重连成功; 否则需要重新选边。为了使网络充分随机化, 一般应根据网络的规模将上述断边重连过程重复多次, 直到原始网络足够随机化后才认为生成了对应的零模型。

保持平均最短路径基本不变的 1 阶零模型的生成条件要高于 0 阶零模型, 需要在度分布不变的前提下进行随机断边重连。1 阶零模型的构造过程如图 1(b)所示, 随机选择两条边 AB 和 CD , 且保证 AD 和 BC 之间没有边, 断边重连之前计算网络的平均最短路径, 断开边 AB 和 CD , 连接 AD 和 BC 后计算新网络的平均最短路径。若断边重连前后网络的平均最短路径保持基本不变, 那么说明本次断边重连成功; 否则需要重新选边。根据网络规模和实际需要进行迭代, 直到网络足够随机化。

保持平均最短路径基本不变的 2 阶零模型的

随机断边重连算法是保持联合度分布不变的连边置乱, 具体步骤与上述 1 阶零模型一样, 只是多了一个限制条件, 即要求交换前后两条边连接节点的度值是不变的。2 阶零模型的构造过程如图 1(c)所示, 随机选择两条边 AC 和 BD , 且保证 AD 和 BC 之间没有边以及节点 B 和节点 D (或节点 A 和节点 C) 的度值相同, 断边重连之前计算网络的平均最短路径, 断开边 AB 和 CD , 连接 AD 和 BC 后计算新网络的平均最短路径。若断边重连前后网络的平均最短路径保持基本不变, 那么说明本次断边重连成功; 否则需要重新选边, 然后重复这个过程, 直到网络达到足够随机化。

保持平均最短路径基本不变的 3 阶零模型是在 2 阶零模型的基础上, 保持与原始网络有相同度相关的聚类系数、相同开闭三角形数量和相似的平均最短路径。3 阶零模型的构造过程如图 1(d)所示, 首先随机选择两条边 AC 和 BD , 且保证 AD 和 BC 之间没有边以及节点 B 和节点 D (或节点 A 和节点 C) 的度值相同, 并计算断边重连之前两条边两端节点度相关的聚类系数、选边节点及其邻居节点组成的子图中开闭三角形的数量和网络的平均最短路径, 断边重连之后计算新生成的连边连接节点度相关的聚类系数、新边节点及其邻居节点组成的子图中开闭三角形的数量和新网络的平均最短路径。若断边重连前后节点度相关的聚类系数数值不变、开闭三角形的数量不变和网络的平均最短路径基本不变, 那么说明本次断边重连成功。若不满足上述 3 个条件之一, 则需要重新选边, 然后重复这个过程, 直至网络足够随机化。

本文对零模型算法复杂度的评估指标选用断边重连的成功率, 即成功断边重连次数与尝试断边重连次数的比值。在相同的随机化程度下, 零模型的算法复杂度越低, 断边重连的成功率越高。保持平均最短路径基本不变的 1-3 阶零模型随着零模型阶数的增加, 限制条件越来越多, 因此算法的复杂度越来越大。

选取一个高聚类 Facebook 社交网络和一个低

聚类 Twitter 社交网络 (数据集详细信息会在 2.1 节中介绍), 在生成真实网络零模型后, 得到了各阶零模型的断边重连成功率。在 Facebook 社交网络中, 保持平均最短路径基本不变的 0 阶零模型的断边重连成功率为 28.6%, 保持平均最短路径基本不变的 1 阶零模型的断边重连成功率为 37%, 保持平均最短路径基本不变的 2 阶零模型的断边重连成功率为 0.4%, 保持平均最短路径基本不变的 3 阶零模型的断边重连成功率为 0.1%。在 Twitter 社交网络中, 保持平均最短路径基本不变的 0 阶零模型的断边重连成功率为 17.2%, 保持平均最短路径基本不变的 1 阶零模型的断边重连成功率为 37%, 保持平均最短路径基本不变的 2 阶零模型的断边重连成功率为 7.5%, 保持平均最短路径基本不变的 3 阶零模型的断边重连成功率为 5.3%。基于真实社交网络零模型的断边重连成功率, 证实了我们对零模型复杂度分析的正确性。

(2) 保持平均最短路径基本不变而改变聚类系数的零模型

在微观尺度特性中, 聚类系数属于高阶特性。为了实现聚类系数单独改变的同时不改变其他网络统计量, 本文提出保持平均最短路径基本不变而改变聚类系数的零模型。该零模型在 2 阶零模型的基础上, 不仅保持了平均最短路径基本不变, 还保持了 0-2 阶特性不变, 从而保证了聚类系数单独连续变化。

保持平均最短路径基本不变而增大聚类系数的零模型的构造过程为: 首先, 随机从原始网络中选取两个度相同的节点及其连边; 其次, 断边重连之前, 计算两条边及其邻居的平均聚类系数和网络的平均最短路径, 在断边重连之后计算新生成的连边及其邻居的平均聚类系数和新网络的平均最短路径。若断边重连后所选边上节点的平均聚类系数数值大于断边重连前的平均聚类系数, 且网络的平均最短路径基本保持不变, 则说明本次断边重连成功。若断边重连后所选边上的节点的平均聚类系数小于或等于断边重连前的平均聚类系数或平均最

短路径数值相差较大, 则说明所选的边不符合条件, 需要重新选边。

保持平均最短路径基本不变而减小聚类系数的零模型的构造过程为: 首先, 随机从原始网络中选取两个度相同的节点及其连边; 其次, 断边重连之前, 计算两条边及其邻居的平均聚类系数和网络的平均最短路径, 在断边重连之后计算新生成的连边及其邻居的平均聚类系数和新网络的平均最短路径。若断边重连后所选边上的节点的平均聚类系数数值小于断边重连前的平均聚类系数, 且网络的平均最短路径基本保持不变, 则说明本次断边重连成功。若断边重连后所选边上的节点的平均聚类系数大于或等于断边重连前的平均聚类系数或平均最短路径相差较大, 则说明所选的边不符合条件, 需要重新选边。

2.1.2 保持微观特性不变而改变平均最短路径的零模型

平均最短路径是复杂网络重要的宏观尺度特性网络统计量, 定义为任意两个节点之间距离的平均值。保持微观特性不变而改变平均最短路径的零模型, 在保持微观尺度特性不变的同时, 可以使平均最短路径增大或减小, 从而得到只有平均最短路径连续变化的网络集合。

保持微观特性不变而增大平均最短路径的零模型的构造过程为: 首先, 随机从原始网络中选取两个度相同的节点及其连边; 其次, 断边重连之前, 计算两条边及其邻居的平均聚类系数和网络的平均最短路径, 在断边重连之后计算新生成的连边及其邻居的平均聚类系数和新网络的平均最短路径。若断边重连前后所选边上的节点的平均聚类系数不变且平均最短路径增大, 则说明本次断边重连成功。若断边重连前后所选边上的节点的平均聚类系数不一致, 或断边重连前的平均最短路径小于或等于断边重连后的平均最短路径, 则说明所选的边不符合条件, 需要重新选边。

保持微观特性不变而减小平均最短路径的零模型的构造过程为: 首先, 随机从原始网络中选取两个度相同的节点及其连边; 其次, 断边重连之前, 计算两条边及其邻居的平均聚类系数和网络的平均最短路径, 在断边重连之后计算新生成的连边及其邻居的平均聚类系数和新网络的平均最短路径。若断边重连前后所选边上的节点的平均聚类系数不变且平均最短路径减小, 则说明本次断边重连成功。若断边重连前后所选边上的节点的平均聚类系数不一致, 或断边重连前的平均最短路径大于或等于断边重连后的平均最短路径, 则说明所选的边不符合条件, 需要重新选边。

2.2 传播模型

经典的信息传播模型包括独立级联模型、线性阈值模型、传染病模型和博弈论模型等。本文的研究框架具有通用性, 因此可以在不同的传播模型上进行仿真实验。以往信息传播研究中广泛应用了传染病模型, 考虑到传染病模型中的 SIR 模型机制较为简单且符合真实场景的传播过程, 本文主要使用 SIR 模型来进行仿真实验。然而, 社会强化是信息传播中极为重要的一部分, 传染病模型没有考虑到这种效应。线性阈值模型是一种价值积累模型, 基于邻居节点的影响力来进行传播, 满足社会强化的要求, 因此我们选用线性阈值模型进行对比实验。

2.2.1 传染病模型——SIR 模型

SIR 模型最早在 Kermack 与 McKendrick 于 1927 年发表的论文中提出, 逐渐发展为最成功的传染病模型之一。该模型把网络中的节点分为 3 种状态: 易感态 (S 态), 即节点处于未得病的状态, 但缺乏免疫能力, 与感染态节点接触后有一定概率受到感染, 其数量记为 $s(t)$, 表示 t 时刻未染病但有可能被该疾病传染的节点数; 感染态 (I 态), 即感染疾病者, 它可以传播给 S 态的节点, 其数量记为 $i(t)$, 表示 t 时刻已被感染成为病人而且具有传染力的节点数; 免疫态 (R 态), 即被隔离或因病愈而具有免疫力的节点, 其数量记为 $r(t)$, 表示 t 时刻已经从

感染态中免疫的节点数^[20]。设总人口为 $N(t)$, 则有 $N(t)=s(t)+i(t)+r(t)$ 。

SIR 模型的构建基于 3 个假设: 1) 人口总量保持不变, 不考虑出生率、死亡率等种群动力因素; 2) 感染态节点与易感态节点接触就有一定概率被感染, 假设 t 时刻单位时间内, 一个感染态节点能感染的易感态节点数量与此环境中易感态节点总数 $s(t)$ 成正比, 比例系数为 β , 从而在 t 时刻单位时间内被所有感染态节点传染的节点数为 $\beta s(t)i(t)$; 3) 单位时间内从感染态转换成免疫态的节点数与感染态节点数成正比, 比例系数为 γ , 单位时间内免疫者的数量为 $\gamma i(t)$ 。基于这 3 个假设条件, 给出感染机制:

$$\begin{cases} S(i) + I(j) \xrightarrow{\beta} I(i) + I(j) \\ I(i) \xrightarrow{\gamma} R(i) \end{cases} \quad (1)$$

当易感态节点与感染态节点都存在的时候, 感染态节点的增长率为 $\beta s(t)i(t) - \gamma i(t)$, 易感态节点的下降率为 $\beta s(t) i(t)$, 免疫态节点的增长率为 $\gamma i(t)$ 。节点从易感态到免疫态的过程可用微分方程表示如下:

$$\begin{cases} \frac{ds(t)}{dt} = -\beta i(t)s(t) \\ \frac{di(t)}{dt} = \beta i(t)s(t) - \gamma i(t) \\ \frac{dr(t)}{dt} = \gamma i(t) \end{cases} \quad (2)$$

由此可将 β 作为感染率, γ 作为免疫率。

在本文的 SIR 模型仿真实验中, 感染率 β 设置为 0.6, 免疫率 γ 设置为 0.5, 传播的总时间为 T , 初始时刻 $t=0$, 最初网络总节点数为 N , 选取网络中度中心性最高的 5 个节点作为感染态节点 ($I=5$), 剩余节点均为易感态 ($S=N-1$), 免疫态节点数为 0 ($R=0$)。接下来的每一时间步内, 每个感染态节点都会以感染率 β 对其易感态邻居节点进行传染, 并且感染态节点以免疫率 γ 变成免疫态, 直至网络中没有感染态节点时传播结束。

2.2.2 影响力模型——线性阈值模型

影响力模型属于应用传播学线性视角的范畴。线性传播从信息发出者的角度出发, 认为信息的传播是一个说服的过程, 信息的接收者是被动的。本文用到的线性阈值模型^[21] (LT 模型) 就属于影响力模型的一种。

首先将社交网络中的节点分为激活态和待激活态两种状态, 每个节点由系统随机分配一个阈值 θ , 且 $\theta \in [0,1]$ 。用 θ_v 表示节点 v 的激活阈值, 用 b_{uv} 表示节点 u 对其邻居节点 v 的影响力, 用 $N_a(v)$ 表示所有处于激活态的节点 v 的邻居节点的集合。当且仅当处于待激活态的节点 v 所接收到的来自其所有处于激活态的邻居节点所传递的影响力的总和超过了点 v 自身所对应的阈值 θ_v 时, 节点 v 被激活。因此, 节点 v 从待激活态变为激活态的条件可表示为:

$$\sum_{u \in N_a(v)} b_{uv} \geq \theta_v \quad (3)$$

线性阈值模型的具体传播过程为: 最初, 所有节点都处在待激活态; 在时间步长为 0 时, 将预先设定的 ρ_0 个种子节点转换为激活态; 在随后的时间步骤中, 如果一个待激活态节点的激活邻居的比例超过阈值, 该节点就会成为激活态节点, 并且一旦被激活, 该节点将永远保持激活态。按照这些规则, 当网络中没有节点可以变成激活态节点时, 网络达到稳定状态。

在本文的线性阈值模型仿真实验中, 阈值 θ 设置为 0.3, 传播的总时间为 T , 初始时刻 $t=0$, 随机选取网络中 20% 的节点作为种子节点, 将种子节点转换为激活态, 剩下的节点设置为待激活态。接下来的每一个时间步内, 若待激活态节点的阈值小于其激活态邻居节点在所有邻居节点中的比例, 则该节点会转换成激活态, 当网络中没有可以变成激活态的节点时传播结束。

2.3 传播效果评价指标

在现实生活中, 文字信息、音频信息、视频信

息等多媒体信息被丰富的信息传播渠道所传递,例如电视、广播、互联网等。本文研究中,不考虑传播的信息内容、传播的平台等现实问题,只针对实证网络中节点对节点之间的传播行为,用到的传播效果评价指标为信息传播速度和信息传播范围。

2.3.1 信息传播速度

本文的信息传播速度指在信息传播模型里,由部分传播到整体的平均速率。在 SIR 模型中,我们将感染的总节点数与总时间步数的比值作为传播速度。在初始时刻 $t=0$, 由于选择度中心性最高的 5 个节点作为初始感染节点, 因此最初感染节点数 $N_0=5$; 在 T 时刻传播结束, 此时感染节点的总数为 N_T , 传播速度 $v_{SIR} = \frac{N_T - N_0}{T}$ 。在线性阈值模型中, 传播速度的计算方式与 SIR 模型相似, 计算方式为在 T 时刻传播完成后的激活态节点总数 N_T 与初始时刻的种子节点数 N_0 的差值除以总传播时间 T , 即 $v_{LT} = \frac{N_T - N_0}{T}$ 。

2.3.2 信息传播范围

本文的信息传播范围指信息传播过程中所覆盖的总节点数。在 SIR 传染病模型中, 由于最终感染态节点会清零, 因此在传播结束 T 时刻的免疫态节点总数 R_T 可以代表信息传播范围 F_{SIR} , 即 $F_{SIR} = R_T$; 在线性阈值模型中, 在 T 时刻传播结束的最终的激活态节点数 N_T 可以代表信息传播范围 F_{LT} , 即 $F_{LT} = N_T$ 。

3 仿真实验

3.1 网络数据集

本文选取了两个无权无向的真实在线社交网络, 分别为 Facebook 相互喜欢网络^[22] (以下简称 Facebook 网络) 和 Twitter 转发提及网络^[23] (以下简称 Twitter 网络), 两个实证网络的数据均来自于网络仓库网站。Facebook 网络的节点表示页面, 边表示页面之间互相喜欢, 该网络是 2017 年 11 月收集的关于不同类别的蓝色验证

Facebook 页面的数据; Twitter 网络的节点表示用户, 边表示转发或者提及, 该网络是 2010 年在哥本哈根举行的联合国会议的转发和提及网络, 数据是在会议的两周内收集的。网络数据集参数如表 1 所列, 从表中可以看出 Facebook 网络是一个高聚类的网络, 仅仅 620 个节点就有 2102 条边和 8800 个三角形, 网络平均度比较大, 最大度达到了 132; 相比 Facebook 网络, Twitter 网络中三角形数量要少得多, 761 个节点仅含有 1029 条边, 说明 Twitter 是一个低聚类的网络。

表 1 真实社交网络具体统计参数

网络	节点数	边数	平均度	最大度	三角形数量
Facebook	620	2102	6.781	132	8800
Twitter	761	1029	2.704	37	447

3.2 基于 SIR 模型的信息传播效果影响因素分析

3.2.1 网络微观尺度特性对信息传播效果的影响分析

为了探究在减少不同尺度特性间耦合作用后网络微观尺度特性对信息传播效果的影响, 首先构建 Facebook 网络和 Twitter 网络不考虑平均最短路径变化的 0-3 阶零模型和保持平均最短路径基本不变的 0-3 阶零模型, 然后比较原始网络和减少耦合作用前后的 0-3 阶零模型传播曲线之间的差异性。

表 2、表 3 列出了原始网络、不考虑平均最短路径变化的 0-3 阶零模型和保持平均最短路径基本不变的 0-3 阶零模型的网络统计量。从表中可以看出 0 阶零模型的各个网络统计量均与原始网络有较大的差异, 这是由于 0 阶零模型的随机化程度高。1-3 阶零模型的各个网络统计量随着阶数的增大, 越来越接近原始网络; 保持平均最短路径基本不变的 0-3 阶零模型的平均最短路径均与原始网络近似相同。

表 2 Facebook 网络 0-3 阶零模型的网络统计量

Table 2 0-3 order null model network statistics of Facebook network

Facebook 网络	聚类系数	平均最短路径	匹配系数
原始网络	0.331	5.089	-0.028
0 阶/保持路径 0 阶	0.009/0.151	3.572/5.080	0.022/0.044
1 阶/保持路径 1 阶	0.070/0.243	3.321/5.035	-0.055/-0.032
2 阶/保持路径 2 阶	0.058/0.265	3.463/5.087	-0.028/-0.028
3 阶/保持路径 3 阶	0.331/0.331	4.355/5.088	-0.028/-0.028

表 3 Twitter 网络 0-3 阶零模型的网络统计量

Table 3 0-3 order null model network statistics of Twitter network

Twitter 网络	聚类系数	平均最短路径	匹配系数
原始网络	0.076	5.353	-0.099
0 阶/保持路径 0 阶	0.004/0.052	6.702/5.353	0.037/-0.089
1 阶/保持路径 1 阶	0.008/0.021	4.878/5.335	-0.063/-0.119
2 阶/保持路径 2 阶	0.010/0.035	4.993/5.348	-0.099/-0.099
3 阶/保持路径 3 阶	0.076/0.076	5.096/5.347	-0.099/-0.099

图 2 给出了两个实证网络不考虑路径变化的 0-3 阶零模型和保持路径基本不变的 0-3 阶零模型在 SIR 传染病模型上的传播效果, 传播曲线的数据是实验 1000 次取均值的结果。其中, 图 2(a)和图 2(b)是传播速度图, 图 2(c)和图 2(d)是传播范围图, 为了便于比较, 原始网络的传播指标用虚线标出。从图中可以看出, 与不考虑路径变化的 0-3 阶零模型相比, 保持路径基本不变的 0-3 阶零模型在两种传播指标上均更加接近于原始网络, 说明平均最短路径对信息传播效果的影响大于 0-3 阶特性。从图中还可以看出解耦前后的 3 阶零模型的传播指标均比 0-2 阶零模型更加接近原始网络。为探究这一现象的原因, 图 3 给出了保持平均最短路径基本不变而改变聚类系数的零模型的传播效果, 横轴是聚类

系数, 纵轴是传播效果指标, 传播曲线的数据是实验 1000 次取均值的结果。其中, 图 3(a)和图 3(b)是传播速度图, 图 3(c)和图 3(d)是传播范围图。从图中可以看出, 两组网络聚类系数零模型的传播速度数值总是在原始网络的传播速度数值附近浮动, 说明保持平均最短路径基本不变时, 聚类系数的变化对信息传播速度的影响很小。两组网络零模型随着聚类系数的增大, 传播范围越来越小, 说明即使保持平均最短路径基本不变, 聚类系数对信息传播范围仍有影响, 且聚类系数越大信息传播范围越小。

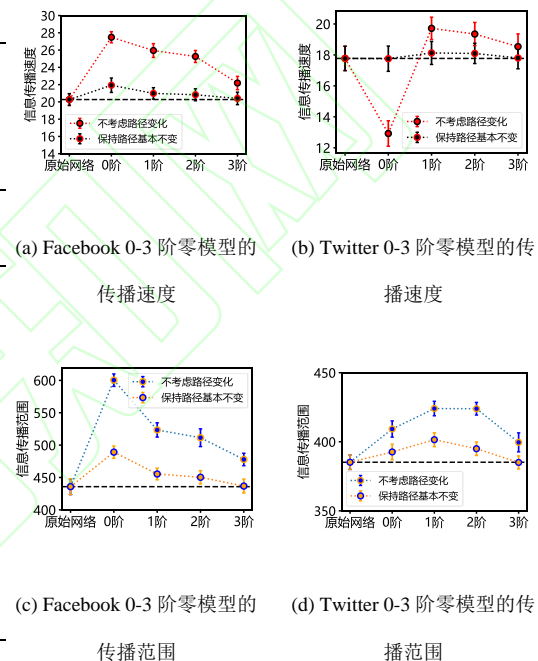
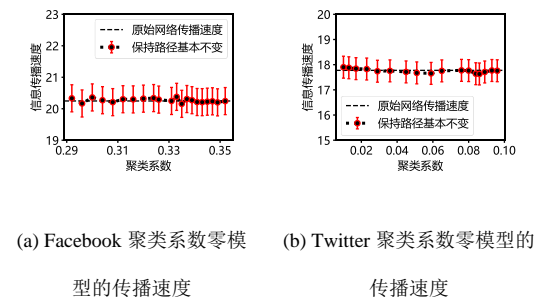


图 2 真实网络 0-3 阶零模型 SIR 传播效果图

Fig.2 SIR propagation effect of real network

0-3 order null model



(a) Facebook 聚类系数零模型的传播速度 (b) Twitter 聚类系数零模型的传播速度

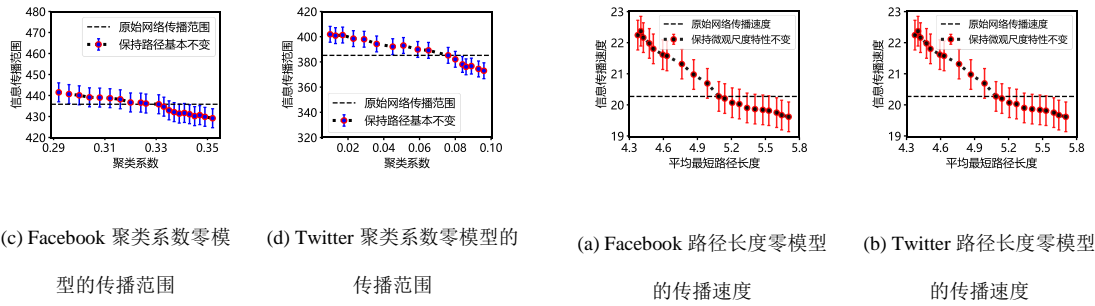


图3 真实网络聚类系数零模型的 SIR 传播效果曲线

Fig.3 SIR propagation effect curve of real network with clustering coefficient null model

3.2.2 网络宏观特性对信息传播效果的影响分析

3.2.1 节实验的结论中已表明网络的平均最短路径是影响信息传播的重要因素，为探究平均最短路径影响信息传播效果的规律，我们在保持微观特性不变而改变平均最短路径的零模型上进行 SIR 传播实验。图 4 为平均最短路径零模型的 SIR 传播效果图，横轴是平均最短路径，纵轴是传播效果指标，传播曲线的数据是实验 1000 次取均值的结果。图 4(a)和图 4(b)是传播速度图，图 4(c)和图 4(d)是传播范围图。从图中可以看出，两组网络的平均最短路径零模型的传播速度随着平均最短路径的增大而减小，说明平均最短路径的变化对传播速度起着决定性的作用。两组网络的平均最短路径零模型的传播范围随着平均最短路径的增大而减小。与图 3(c)和图 3(d)相比，图 4(c)和图 4(d)中传播范围变化的幅度更大，说明平均最短路径对信息传播范围的影响大于聚类系数的影响。综上所述，平均最短路径是影响信息传播速度的主要因素，平均最短路径越长，信息传播速度越小；平均最短路径是影响信息传播范围的主要因素，聚类系数是影响信息传播范围的次要因素，平均最短路径越长，信息传播范围越小。

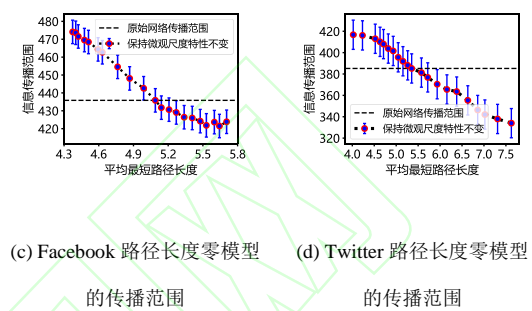


图4 真实网络平均最短路径零模型的 SIR 传播效果图

Fig.4 SIR propagation effect diagram of real network average shortest path null model

3.3 基于线性阈值模型的对比实验

在传播动力学研究中，对社交网络信息传播的研究主要集中在网络结构上。近年来，社会强化对信息传播的影响逐渐引起了一些学者的关注^[24]。本文为了讨论在 SIR 传播模型中得到的结论在社会强化中的适用性，使用线性阈值模型进行对比实验。对比实验基于保持平均最短路径基本不变而改变聚类系数的零模型和保持微观特性不变而改变平均最短路径的零模型，通过两种信息传播指标的传播曲线来进行分析比较。图 5 给出了两种零模型在线性阈值模型的信息传播速度图。从图 5(a)和图 5(b)中可以看出，随着聚类系数的变化，零模型的信息传播速度在原始网络传播速度附近浮动且十分接近原始网络，说明聚类系数对信息传播速度的影响较小。从图 5(c)和图 5(d)中可以看出，平均最短路径越大，信息传播速度越慢。

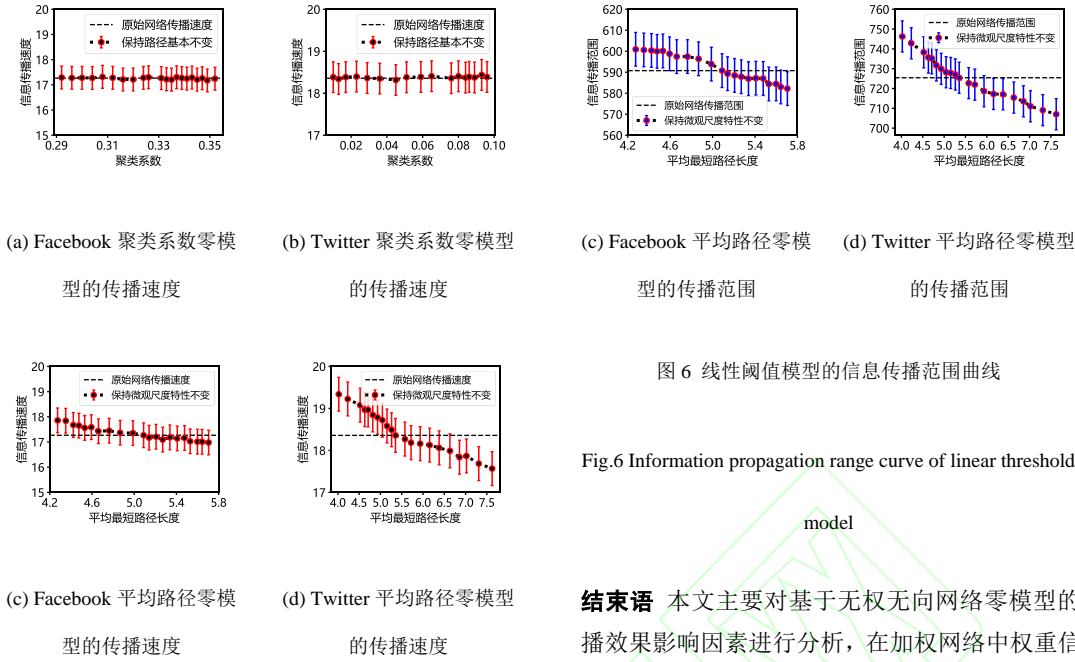


图 5 线性阈值模型的信息传播速度曲线

Fig.5 Information propagation speed curve of linear threshold model

图 6 是两种零模型在线性阈值模型的信息传播速度图。在图 6(a)中, 随着聚类系数的增大, 信息扩散范围有下降的趋势, 但是下降的幅度比较小; 在图 6(b)中, 可以明显地看出聚类系数越大, 信息传播范围越小。从图 6(c)和图 6(d)中可以看出, 平均最短路径越长, 信息传播范围越小, 且平均最短路径的改变使信息传播范围缩小的速度大于聚类系数的改变。线性阈值传播模型的结论与 SIR 传播模型的结论相同, 由此说明在 SIR 传播模型中得到的结论也在社会强化中适用。

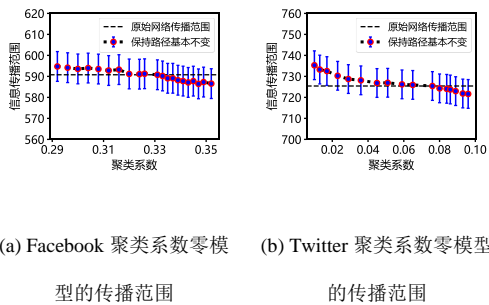


图 6 线性阈值模型的信息传播范围曲线

Fig.6 Information propagation range curve of linear threshold model

结束语 本文主要对基于无权无向网络零模型的传播效果影响因素进行分析, 在加权网络中权重信息对网络的信息传播有非常重要的影响。在今后的研究工作中, 我们会综合考虑权重信息与拓扑结构对网络信息传播的影响。

与之前的相关研究相比, 本文的亮点是研究方法的不同。我们提出了一种零模型框架, 通过保证某种特性不变的同时改变另一种特性的方法, 构造了关于不同尺度特性的零模型网络, 从而减少了各个网络统计量间的耦合作用; 然后使用零模型与信息传播模型相结合的方式, 探讨了原始网络和零模型之间拓扑结构的差异对信息传播的影响。通过分析与总结信息传播仿真实验的结论可知: 信息传播速度的变化是由网络中平均最短路径所决定的, 平均最短路径越长, 信息传播速度越慢, 微观特性网络统计量对信息传播速度的影响微乎其微。信息传播范围的变化是由聚类系数和平均最短路径共同影响的, 网络的聚类系数越大, 信息传播范围越小, 网络的平均最短路径越长, 信息传播范围越小, 其中平均最短路径对信息传播范围的影响要大于聚类系数。

依据实验的结论, 面对现实传播问题时, 可以根据不同的信息传播类型采取不同的措施。若是疫情传播、谣言传播等传播行为需要抑制, 应采取尽可能切断传染源头与被感染者的传播路径, 增加传播网络聚类等措施; 若是广告营销等传播行为需要

扩散,应选择传播路径短且聚类系数较小的社交网络或圈子进行传播。

参考文献

- [1] LAMMERS J, DUBOIS D, RUCKER D D, et al. Power gets the job: Priming power improves interview outcomes[J]. *Journal of Experimental Social Psychology*, 2013, 49(4): 776-779.
- [2] DUBOIS D, RUCKER D D, GALINSKY A D. Dynamics of Communicator and Audience Power: The Persuasiveness of Competence versus Warmth[J]. *Journal of Consumer Research*, 2016, 43(1): 68-85.
- [3] YUAN M, LIU N. Power and persuasion: The value of message-audience matching and fluency[J/OL]. <https://doi.org/10.1007/s12144-022-02915-4>.
- [4] WEI X, ZHAO J, LIU S, et al. Identifying influential spreaders in complex networks for disease spread and control[J]. *Scientific Reports*, 2022, 12(1): 1-11.
- [5] WORCHEL S, ANDREOLI V, EASON J. Is the Medium the Message? A Study of the Effects of Media, Communicator, and Message Characteristics on Attitude Change[J]. *Journal of Applied Social Psychology*, 1975, 5(2): 157-172.
- [6] NEWMAN M E J. Properties of highly clustered networks[J]. *Physical Review E*, 2003, 68(2): 026121.
- [7] NEWMAN M. Random graphs with clustering[J]. *Physical review letters*, 2009, 103(5): 058701.
- [8] GLEESON J P, MELNIK S, HACKETT A. How clustering affects the bond percolation threshold in complex networks[J]. *Physical Review E*, 2010, 81(6): 066114.
- [9] CENTOLA D. The Spread of Behavior in an Online Social Network Experiment[J]. *Science*, 329(5996): 1194-1197.
- [10] XING X Y. Research on the evolution of online social network structure and its influence on information dissemination[D]. Hefei:Hefei University of Technology,2012.
- [11] XU X D, XIAO Y T, ZHU S R. Research on rumor spreading simulation in micro blog community. *Computer Engineering*, 2011, 37(10): 272-274.
- [12] ZHOU J Y, LIU Z Z, XU X K. Empirical analysis of influential factors of network communication with reference to zero model[J]. *Complex Systems and Complexity Science*, 2019, 16(3): 40-47.
- [13] XUAN Q, WANG Z, WANG J, et al. Data Augmentation Based on Null Model for Graph Classification[J]. arXiv: 2112.00476, 2021.
- [14] CAZABET R, BORGNAT P, JENSEN P. Enhancing space-aware community detection using degree constrained spatial null model[C]//International Workshop on Complex Networks. Cham:Springer, 2017: 47-55.
- [15] NIAN X, FU H. Efficient routing on two layer degree-coupled networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2014, 410: 421-427.
- [16] ULRICH W, GOTELLI N J. Pattern detection in null model analysis[J]. *Oikos*, 2013, 122(1): 2-18.
- [17] SHANG K K. Research on zero model construction and behavior prediction of online social network[D]. Qingdao:Qingdao Technological University, 2013.

[18] SHANG K K, XU X K. Construction and application of complex network null model based on Scrambling Algorithm[J]. Journal of University of Electronic Science and Technology of China, 2014, 43(1): 7-20.

[19] XU X K, CUI W K, CUI L Y, et al. Construction and application of zero model of non weighted network[J]. Journal of University of Electronic Science and Technology of China, 2019, 48(1): 122-141.

[20] ZHOU H Y, LI S L, LIU H K. Study on the epidemic situation of novel coronavirus pneumonia based on Sir / IR model —— Based on the data from January to may 2020 in Hubei Province[J]. Journal of Hunan University of Technology, 2021,35(6): 89-94.

[21] PENG H, NEMATZADEH A, ROMERO D M, et al. Network modularity controls the speed of information diffusion[J]. Physical Review E, 2020, 102(5): 052316.

[22] ROZEMBERCZKI B, DAVIES R, SARKAR R, et al. GEMSEC: Graph Embedding with Self Clustering[C]//Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2019: 65-72.

[23] AHMED N K, BERCHMANS F, NEVILLE J, et al. Time-Based Sampling of Social Network Activity Graphs[C]//Proceedings of the Eighth Workshop on Mining and Learning with Graphs. 2010: 1-9.

[24] TUR E M, ZEPPINI P, FRENKEN K. Diffusion with social reinforcement: The role of individual preferences[J]. Physical Review E, 2018, 97(2): 022302.

postgraduate.His main research interests include null model of complex social network .



XU Xiaoke, born in 1979,, ph.D, professor.His main research interests include network science and social network analysis.

(责编: 柯颖)



CUI Yunsong, born in 1997, ,