

第一部分

社交网络与建模方法

本部分给出本书社交网络及其建模方法的基础理论方法,包括社交网络及其演化的基本知识、博弈论的基础理论、演化博弈的基础理论及当前的研究进展,以及社交演化博弈模型方法。其中,

第1章介绍社交网络及其演化的基本概念、形式化表示和研究要素等基本概念和知识。

第2章介绍博弈论的基本概念和性质。首先介绍了博弈论的基本框架、理性行为、有限理性等的基础知识,以及常见的博弈模型。随后介绍了博弈论中的重要概念纳什均衡、纳什均衡的存在性以及混合策略纳什均衡。最后介绍了合作博弈的相关知识。

第3章介绍演化博弈论及网络演化博弈的研究进展。首先介绍了演化博弈的基本结构、种群博弈、复制者动态以及演化博弈中的重要概念——演化稳定策略。随后在静态网络、动态网络和相互依赖网络这3个层面上回顾了近年来网络演化博弈的研究进展。

进一步地,第4章给出面向社交网络演化研究的社交演化博弈模型,首先给出了社交演化博弈模型的建模方法,进而根据不同类型的社交网络个体信息交互行为给出了模型设置方法和仿真结果,探讨了个体信息交互行为与社交网络演化的相互影响。

第 1 章

社交网络

1.1 概述

互联网正逐步演变为无处不在的计算平台和信息传播平台。在线社交网站、微博、博客、论坛、维基等社交网络的出现和迅猛发展,使得人类使用互联网的方式产生了深刻变革——由简单信息搜索和网页浏览转向网上社会关系的构建与维护,以及基于社会关系的信息创造、交流和共享。

基于互联网的社交网络正在成为人类社会关系维系和信息传播的重要渠道和载体,对国家安全和社会发展都会产生深远的影响:①社会个体通过各种连接关系在社交网络上构成“关系结构”,包括以各种复杂关系关联而成的虚拟社区;②基于社交网络的关系结构,大量网络个体围绕某个事件而聚合,并相互影响、作用、依赖,从而形成具有共同行为特征的“网络群体”;③基于社交网络关系结构和网络群体,各类“网络信息”得以快速发布并传播扩散形成社会化媒体,并反馈到现实社会,使得社交网络与现实社会之间形成互动,对现实世界产生影响。

针对网络结构、群体和信息之间的互动关系,研究社交网络中各种关系结构(例如,虚拟社区等)的形成机制,发现社交网络关系结构的稳定特性;研究网络群体建模方法,揭示网络群体演化规律,发现网络群体的生成机理;研究网络群体行为规律,分析群体内部的相互作用及群体之间的相互影响,发现社交网络中的信息传播模式;研究多种社交网络之间信息传播的相互影响以及与传统媒体之间的互动规律。这些问题的研究涉及计算机科学、社会学、管理学、心理学等多个学科领域,可以揭示社交网络中的关系结构、网络群体、网络信息之间的复杂交互关系和互动规律,为社交网络分析与网络信息传播研究提供重要的理论支撑。

1.2 社交网络的研究要素

在线社交网络是一种由社会个体集合及个体之间在信息网络上的连接关系构成的社会性结构,包含关系结构、网络群体与网络信息 3 个

要素。其中,社交网络的关系结构是社会个体成员之间通过社会关系结成的网络系统。个体也称为节点,可以是组织、个人、网络 ID 等不同含义的实体或虚拟个体;而个体间的相互关系可以是亲友、动作行为、收发消息等多种关系。其中,社交网络中的虚拟社区为该社交网络的一个子集,且虚拟社区内节点之间关联密切,不同虚拟社区的节点间关联稀疏。网络群体行为是指网络个体就某个事件在某个虚拟空间聚合或集中,相互影响、作用、依赖,有目的地以类似方式进行的行为。基于社交网络的信息传播是指社交网络中的个体与个体之间、个体与群体之间、群体与群体之间的信息传递。这种传递作用可以迭代进行。

与传统的 Web 应用及信息媒体应用相比,社交网络应用具有以下主要的新特点:①信息的发布和接收异常简便、迅速。用户可以通过手机和浏览器随时随地发布和接收信息。②“核裂变”式的信息传播。消息一经发布即刻被系统推送到所有关注者,一旦被其转发,又立即传播到下一批关注者,呈现“核裂变”式的几何级数扩散态势,为普通网民创造了意见扩散的渠道。③人人都有机会成为意见领袖。与传统媒体相比,社交网络服务中的广大网民都有机会通过社交网络形成意见领袖,并在突发事件的产生、发酵、传播、爆炒等环节中具有重要地位。④呈现自媒体形态,能够快速形成虚拟社区。由于社交网络中的个体都具有提供、发布信息的手段和渠道,呈现自媒体形态,并依靠社交网络快速传播信息,快速形成网上的虚拟社区。

1.2.1 网络结构

在社交网络中存在着一定的网络结构,可以用图的形式进行描述和记录。针对社交网络的静态结构特性的分析,可以从宏观、中观和微观 3 个层面展开。

在社交网络的宏观静态结构特性分析方面,大量已有研究验证了在真实世界中各种不同的社交网络具有许多复杂网络所共有的结构特性。例如:六度分隔、小世界现象、无标度、幂律分布和结构鲁棒性等。2003 年,美国哥伦比亚大学的 Watts 等人在 6 万个节点规模的电邮网络上验证了六度分割和小世界模型。2005 年,学者们又通过研究 Yahoo 在线社区的路径长度,发现了网络规模最大时的平均路径长度和有效路径长度分别为 8 和 10,比通常认为的“六度分隔”要大。

在社交网络的中观静态结构特性分析方面,社区结构是研究热点之一。根据社区结构的定义,社区结构可分为不可重叠的社区和可重叠的社区。在不可重叠的社区结构方面,目前最受关注的是 2004 年由美国密歇根大学的 Newman 等人^[1]提出的通过寻找使得社区的模块度最大的网络划分来发现网络社区的算法。2006 年,Newman 等人^[2]又从矩阵谱分析的角度阐述了模块度优化和传统网络划分的关系,提出了基于模块度矩阵谱分析的社区发现方法。随后,学者们发现,模块度优化方法存在分辨率限制问题,使得基于模块度优化的方法无法识别出一些较小的社区。在可重叠的社区结构方面,2005 年,匈牙利科学院的 Palla 等人^[3]提出了一种基于 K -完全子图渗流的重叠社区发现方法,该方法的优点在于能够揭示网络社区间的重叠现象,不足之处是其参数选择缺乏有效的理论指导。2008 年,美国华盛顿大学的 Rosvall 等人^[4]把网络社区发现问题看作网络压缩问题,社区发现在于寻找网络的一种高效压缩表示,该压缩表示是保留网络拓扑结构特征和网络压缩比率之间的折中。此后,学者们开始研究网络层次化社区的发现问题。2009 年,中国科学院计算技术研究所的程学旗等人^[5]进一步提出了网络层次化重叠社区的发现方法,能够同时揭示网络

的层次化和重叠社区结构,并通过分析信息通路和社区结构的关系,给出了一种基于网络传导率的社区发现方法。

在社交网络的微观静态结构特性分析方面,从微观层面分析节点间的链接关系,发现有向图中节点的出度可以被模型化为一个节点对其他节点的影响程度,即该节点的活跃度;类似地,将有向图中节点的入度考虑为节点受欢迎程度的一个因素,并发现,在社交网络中,流行度同样也服从幂率分布。部分学者还提出了利用频繁模式挖掘算法发现社交网络中微观结构特征。

1.2.2 群体行为

网络用户通过各种信息行为间的作用关系有机地关联在一起,形成具有一定拓扑结构关系的网络群体。一般地,网络群体是指互联网上有相似特质、共同目标或价值观以及团体意识的、由两个(或两个以上的)以一定方式进行线上、线下交互或活动的用户而构成群体。网络群体不但具有一般群体的特点,还具有其自身的特点:依托于社会关系网络,网络群体不受地理等因素的影响,成员分布范围极广,网络群体的时效性比一般群体强等。网络群体形成的方式多种多样,大致归类为以下几种:①由某种固定关系形成的群体,通常是现实社会中的实际关系在网络中的具体映射,例如由同学及校友关系形成的网络校友群体,这类群体的结构比较稳定,形成之后随时间发生的变化不大,成员的活跃程度有限;②由某种不确定的共同需求形成的群体,例如,微信群、微博中的微群、豆瓣的群组等,这类群组中的成员多是由共同的需求或目的聚集在一起,随着时间的变化,群体成员会发生一定的流动,但是流动主要发生在度数相对较低的节点;③由某种突发事件驱动而形成的群体,这类群体具有很强的时效性,通常由一个焦点事件引发而产生,其结构变化与该事件的关注度涨落有关,在事件平息后的短时间内群体就会解散。

在社交网络中,宏观层面的网络群体行为所表现出来的特征比微观层面的个体信息行为更加复杂。网络群体的行为受到社会性和随机性等特性的影响,使得群体的行为往往表现出极大的不确定性,而更加值得关注的现象是群体行为的涌现。由用户行为导致的信息的传播一般都具有从逐渐积累到一夕爆发的特点,这样的积累并不是线性的,也不是用户行为产生效应的简单叠加,而是一系列小的局部变化,每一个小的局部变化都不足以对整体系统造成影响,但当这些变化的积累达到某个临界状态时,整个系统出现临界相变。整个系统表现出来的现象是所有局部行为相互作用的结果。社会关系网络中的大量涌现现象包括信息扩散数量上的涌现、网络结构特征的涌现、特定网络群体的涌现以及特定网络群体行为的涌现等。

为了理论简化和直观表达,对网络群体行为进行分析和建模是社交网络研究的重要内容。目前,对于网络群体行为的模型的研究集中在对网络群体行为定性描述的模型上,主要作为表现个体信息行为复杂过程的一种方法。社会关系网络是一个多维度的复杂系统,网络内部节点之间的相互影响与相互作用频繁。然而,网络群体行为的社会性、随机性和复杂性使得传统的统计研究方法不能对网络群体行为产生的影响进行预测,因此迫切需要新的基础理论和研究方法。目前,网络群体行为模型的相关研究工作主要包括信息传播及观点交互模型、群体行为演化模型、群体结构模型和突发群体事件模型等,但还没有一个公认的

成体系的研究思路。近年来,对社交网络的建模方面进行了大量工作,比如对于社会网络静态社区结构特性的研究、在演化过程中社会网络动态社区结构变化规律的发现等。但是,这些研究工作的关注点主要集中在网络拓扑结构上,对于网络群体行为的预测调控工作还没有有效的研究方法。

1.2.3 网络信息

与传统媒体相比,基于在线社交网络的网络信息传播在传播媒介、传播渠道、信息受众、反馈机制等方面迥然不同。具体来说,在线社交网络体现出以下几个突出的特点。

(1) 平民化、个性化: 在在线社交网络中,每个人都从旁观者转变为当事人,每个网民都可以利用网络来表达自己的观点,传递他们的生活的阴晴圆缺,构建自己的社交网络。在线社交网络因而成为网民张扬个性、表现自我的最佳场所。

(2) 低门槛、易运作: 在线社交网络中,用户可以发布文字、音乐、图片、视频等信息,创建属于自己的媒体。这不需要投入多大的成本,也不要求有太多的专业技术知识。其进入门槛低,操作运作简单,让在线社交网络大受欢迎,迅速发展。

(3) 传播迅速、交互性强: 在线社交网络大大降低了时空限制,任何时间、任何地点,只要能够上网,用户就可以经营自己的媒体,再加上媒体服务网站所提供的信息推送功能,使得信息能够迅速地传播,时效性大为增强。而且,由于作为一种新的社会媒体形式在线社交网络与受众是零距离的,不但能够迅速地将信息传播到受众中,受众也可以迅速地对信息传播的效果进行反馈,因此其交互性的强大是任何传统媒体望尘莫及的。

因为上述特点,在线社交网络以 Facebook、Twitter、新浪微博和微信等的诞生为标志,在短短十多年的时间里得到了飞速的发展。网络空间给广大网民提供了平等表达自己意见的“新公共领域”。调查显示,有超过一半的网民表示喜欢在互联网上发表评论。网络空间“新公共领域”的特征有助于成为社会冲突的“安全阀”,良好、通畅的对话空间有助于缓解社会矛盾,促进社会和谐。近年来,我国政府积极倡导倡议并引导网络参政议政,广大网民通过互联网渠道评论时事、反映民生、建言献策,网络已经成为推进社会主义民主政治建设的重要力量。

1.3 社交网络的演化

1.3.1 网络信息的传播

信息传播的过程其实也是信息与信息传播者的状态进行演化的过程。目前,得到最广泛且深入研究的信息传播模型是流行病模型。几乎所有的流行病模型都可以用 MSEIR 框架模型来描述,其中, M (passive immune) 表示被动免疫态, S (Susceptible) 表示易感态, E (Exposed) 表示潜伏态, I (Infective) 表示感染态, R (Recovered) 表示恢复态。在流行病模型中,研究最为彻底、应用最为广泛的是 SIS 模型和 SIR 模型^[6]。在 SIS 模型中,个体不能对病毒产生免疫,而在 SIR 模型中被感染的个体能够从感染态恢复过来并获得免疫,从而进入恢复态。由于所研究细节的不同, SIS 和 SIR 模型产生了很多的变体,如 SIRS^[7]、SIDR^[8]和 SAIR^[9]。然而,这些模型都无法反映病毒在使宿主产生传染能力之前有一个潜伏期这

一事实。为解决这一问题,潜伏态被引入了 SIR 模型,从而有了 SEIR 模型^[10]。近来,对 SIS 与 SIR 系列模型又有了不少新的改进用以建模信息传播行为。譬如,为了刻画信息传播中广泛存在的点到群的传播模式,清华大学的学者 Yuan 和 Chen 对 SEIR 模型进行了改良,提出了 e-SEIR 模型^[11]。在文献[12]中,Hill 等人提出了 SISa 模型来研究行为作为一种特殊的信息通过社交网络在人与人之间的扩散。他们在 SIS 模型中引入了个体自动或自发接受某种行为的可能性。

在信息演化传播模型中,也采用类似于 SI 模型的级联模型和阈值模型。在级联模型中,每当一个个体获得信息,都有一定的概率将信息传播给它的相邻个体。而且个体一旦被感染,将一直处于感染态。研究表明,在均匀网络中,存在有限的传播临界值。当有效传播率大于该临界值时,感染个体能够将病毒传播开,并使得整个网络感染个体总数最终稳定于某一平衡状态;如果有效传播率低于此临界值,则感染个体数呈指数衰减,无法大范围传播。阈值模型则假设每个个体对信息有一个接受阈值,个体之间的关系带有权重,当节点在其所有邻居对其影响之和大于接受阈值时,则接受信息。除了上面几种最流行的信息传播模型之外,学者们还提出了其他一些模型。譬如,在文献[13]中,Young 把社会革新的动力性建模成一个在社交网络上进行的协调博弈游戏。在文献[14]中,Iribarren 和 Moro 发现,信息传播展示出非马尔可夫的分支动力学特性,能够用一个两步 Bellman-Harris 分支过程来建模。

目前对网络信息传播的建模工作通常都基于流行病或类似于流行病模型的级联模型和阈值模型。然而,病毒传播与信息传播有着很显著的差异^[15]。例如,在信息传播中,一个个体往往根据多个不同的因素(如信息的内容、自身的喜好等)来决定是否把接收到的信息传播给他的朋友。如果决定要传播的话,则会进一步根据信息的特征、朋友的喜好等因素决定传给哪些朋友,而且往往不是一次就传播给所有要传的朋友。也就是说,现有的信息传播模型往往忽略了传播信息个体之间的差异性,以及他们可以综合多种因素做出决策的能力。因此,现有信息传播模型并不能准确地刻画信息在社会媒体中的传播模式。

1.3.2 网络结构的演化

在社交网络演化分析方面,学者们从社交网络演化中的统计规律展开研究,并提出了面向不同类型社交网络的演化模型。

在社交网络演化规律方面,2005 年美国斯坦福大学的 Leskovec 等人^[16]通过演化网络的快照分析,发现了增长的网络顶点与网络直径反关联的现象。2006 年,美国加州大学伯克利分校的 Chakrabarti 等人^[17]提出了综合时间点社区质量和时间轴社区变化率的进化聚类算法,建立了社交网络的社区演化模型。2007 年,Chi 等人^[18]扩展了相似性计算方法,用图分割(graph cut)作为测度社区结构和社区进化的指标,首次提出进化谱聚类算法。同年,Tantipathananandh 等人^[19]根据社会经验列举了社交网络上中观层面子结构变化的各种情况,并构造出相应变化的损耗 cost 评价体系,建立了以个体消耗、组消耗、颜色变化消耗这 3 种因素合成的最优化模型。Leskovec 等人^[20]还在 2008 年研究了社交网络的微观演化过程,发现边的生成频度与节点间的已有跳数成反比。

在社交网络演化过程建模方面,为了揭示社交网络的形成和演化机制,学者们提出了众多的演化模型来分析社交网络的形成过程。1999 年,Barabasi 和 Albert^[21]提出的 BA 模型则同时考虑了点和边的演化:新加入的节点会引入新边连接到已有的老节点。因此,BA 网

络模型有两个重要特性：一是网络节点不断增加，网络处于动态变化中；二是择优连接，即网络中新增节点连边并不是随机连接到其他节点的，而是与节点的度有关。这两个特点使得 BA 模型显著区别于 ER 模型和 WS 小世界模型。针对权重网络，Barrat 等人^[22]在 2004 年提出加权演化网络模型 BBV 模型，该模型认为节点间的权重与网络的拓扑结构是一种共生演化关系，相互驱动。Kumar 等人^[23]通过对真实网络数据分析，提出了一种简单的择优连接机制以模拟社交网络的生成过程，发现不同类型用户会对网络生长过程产生不同的影响。Bu 等人^[24]提出了一个基于最新更新时间的局部世界模型，用以描述社交网络中的一些结构特征。Yuan 等人^[25]针对微博中用户双向好友关系，提出了一种混合连边规则的网络演化模型，该模型生成的网络与真实网络在众多拓扑特征方面能够保持一致。Sun 等人^[26]则提出了一种基于社交活动的网络演化模型。

1.3.3 群体行为的演化

社会关系网络中的群体是一个复杂系统，群体由各种具有合作关系的个体有机关联形成，个体间的相互影响和作用较为频繁。对于网络群体行为的建模和预测，已逐渐成为众多学者关注的问题。相关的研究工作包括对网络个体行为的建模与分析、网络群体模型与行为分析、群体行为建模与演化性分析以及群体行为预测等等。但社会关系网络群体中每个个体的行为都体现了对自身效用最大化的追求，而这种追求又往往随着内部因素和外部条件的改变发生着演化，这使得网络群体所表现的行为不能用个体的行为进行简单叠加，对网络群体行为的研究既要考虑群体中的结构关系和信息传播的影响，同时也要考虑人的因素。传统的模型和分析方法无法研究这样的复杂系统，而必须从本质上探索系统化的理论与方法。本项目将演化博弈模型的思路引入网络群体行为的描述中，虽然在复杂网络的研究中已有一些探索，但是目前还没有一个公认的成体系的研究方法和研究思路。

网络个体行为与传统实体行为不同，不仅取决于个人的心理、性格和社会经济地位等因素，还会受到网络信息、虚拟平台互动等元素的影响。因此，对于网络个体行为的研究突破了传统行为学的理论框架，而将视角转向网络系统和信息用户之间的互动模式及其影响因素。

网络个体行为模型 网络个体行为建模主要是通过对人们信息行为动机和结果的刻画，来描述微观层面个体行为的模式和特征。其中，Schramm 等人^[27]从传播学的角度，对个体的行为与动机进行了分析，建立了个体行为特征模式。Pei 等人^[28]提出利用个体属性中多维性抽取和个体行为特征密切关联的属性取值，从而达到自动学习个体喜好和行为规律的目标。黄志荣等人^[29]提出了一种基于关联分析的属性独特性度量方法，用于挖掘群体中的个体行为特征。

个体间观点交互模型 观点互动模型主要研究社会关系系统中由于个体之间决策的影响与外界公共信息的影响，群体中对某些特定事件或事物所持的不同观点的形成 (formation) 和演化 (evolution) 等现象，并包括观点的一致性 (consensus) 与多样性 (diversity) 保持等问题。Sznajd 模型重点描述的是观点从内向外扩散的现象。该模型最初使用的场景为链状结构^[30]，之后，研究者们对其在二维网格以及各种复杂网络结构下的演化进行了研究。Deffuant 模型^[31]和 Krause-Hegselmann 模型^[32]是有限信任模型的代表，主要描述了具有一定交互门限的个体中的观点演进现象。在 Krause-Hegselmann 模型中，个体选择更新自己观点的折中值不仅受到双方观点值的影响，还受到节点间的权重值的影

响。Di Mare 和 Larota^[33]在2007年首先提出了观点交互的博弈模型,描述了现实中个体在交互中都倾向于说服对方而避免被对方说服的现象,称为SO(Stubbornness-Orator)模型。

个体间博弈模型 社会关系网络中,个体的行为通常受自身利益的驱使,即个体倾向于自身效用最大化的方向来选择行为,这就构成了个体间的博弈关系。其中,Santos 和 Pacheco^[34]研究了无标度网络上的博弈行为,发现无标度网络结构可以同时使囚徒博弈和雪堆博弈中出现几乎完全合作的状态。Gomez-Gardeñes 等人^[35]通过研究网络中个体策略的动态组织结构,直观地解释了无标度网络促进合作的原因。Tang 等人^[36]研究了网络平均度对博弈中合作行为的影响。Ren 等人^[37]发现网络结构的随机性和动力学随机性能够同时促进网络博弈中的合作行为。

根据群体分析的目标和方法,网络群体行为演化分析研究,从网络群体形成的因素、群体的结构、群体的演化等角度进行。Du 等人^[38]利用博弈模型研究了不同的伙伴选择方式对个体策略和群体结构的影响,结果发现,个体对声誉的偏好会促使群体形成,并使群体结构呈现星型。Chen 等人^[39]提出了群体形成博弈模型,研究了群体发现和群体重叠问题。Palla 等人^[3]提出了一种基于完全子图渗流的重叠社区发现方法。目前,完全子图渗流算法仍然是重叠社区发现方法中应用最为广泛的方法之一,被应用到生物网络、信息网络、社会网络等多种类型的网络中。网络群体的演化性是社会关系网络的一个基本特性。近几年,在前述群体发现研究的基础上,群体演化得到了越来越多的关注。Palla 等人^[40]基于他们提出的 k-clique 群体发现方法研究群体演化,得到一个有趣的结论,小群体的稳定性是保证其存在的前提,大群体的动态性是其存在的基础。群体内合作的驱动力(driving force)可用演化博弈论(evolutionary game theory)进行建模分析。

合作行为的涌现是指以自我利益最大化为目的的自私人最终常常会得到合作态占多数的演化结果的现象。单纯针对这一现象的研究主要集中在复杂网络上的演化博弈讨论。Wu 等人^[41]研究了动态群体的结构对于合作水平的影响,他们发现,合作个体与非合作个体之间的脆弱连接越多,合作涌现越有可能发生。Poncela 等人^[42]在一个不断增长的呈现网络结构的群体中讨论了演化博弈,结果显示,即使是适度的收益优先选择策略也会促使高度数节点的出现;随着网络范围的扩大,出现了高层次的合作,但是作为静态网络相同的网络结构却并没有促进合作。文献[43]提出了一种模型,能够提高在复杂网络中的演化博弈情境下的合作涌现水平。

1.4 社交网络中的分析

1. 网络结构特性与群体行为演化的相互影响

在建立社交网络的演化分析模型时,有一些关键因素会对演化结果产生巨大的影响,这些关键因素也是研究者关注的热点。网络结构特性和个体策略更新机制是对群体行为演化结果有重要影响的因素。

网络结构对群体行为演化的影响 Gomez-Gardenes 等人^[44]将群体结构的影响引入了演化博弈模型,得出的结论是网络局部结构提高了合作行为出现的概率,每个群体人数的增加会降低合作水平。Taylor 等人^[45]讨论了同质网络结构对合作行为的演化的影响。F. C.

Santos 等人^[46]将网络结构的异质特征引入了演化博弈动力学,发现增加异质性有利于合作态的出现。同时,他们阐述了合作行为对无标度网络中的复杂联系的依赖性。

群体行为演化对网络结构的影响 Brian Skyrms^[47]的研究指出:随着演化的进行,生成的网络结构倾向于将整个结构分割成一个个小的群体的结论,这激发了很多学者在社区发现方面的研究,比如常见的复杂网络方向^[2],随机游走在近几年也被用来做相关工作^[48,49]。

2. 个体行为策略更新机制对群体行为演化的影响

Arne Traulsen^[50]研究了允许任意数量的策略在进化过程中突变,发现,当基因突变率高于临界值时,群体由混合策略组成;当低于临界值时,群体趋向同质。他在文献^[51]中讨论了个体以一定概率随机选择策略对演化产生的影响,结果证明,当突变率足够小时,策略选择只与收益的排名有关,其他参数的影响可以忽略。William H. Sandholm^[52]讨论了个体能够以独立泊松过程更改自身策略的情况,也考虑了噪声对选择的影响。

社交网络表象上的无序性导致了信息生成与传播呈现出一种无序的状态,但是其中群体行为的聚集性和可预测性却使得群体行为具有规律和模式可循。Barabási 等人^[53]对匿名手机用户的活动模式的研究发现,人们的活动遵循有规律的模式,93%的人类活动行为是可以预测的,并且人类个体行为的规律性和可预测性在人口统计类别方面没有明显不同。目前,社会关系网络上的用户行为预测的研究主要集中在网络群体中的关系预测和网络群体行为涌现性预测。

3. 社交网络中的个体关系预测分析

Liben-Nowell 和 Kleinberg^[54]首次提出关系预测的定义,预测关系在未来时间内是否产生。基于这种关系预测的算法,按方法分为 3 种。一是基于有监督的学习算法,这种方法将预测问题视为分类问题。比较有代表性的工作包括文献^[55,56]。尽管有监督的学习算法是目前比较流行的算法,但它在分类特征选择方面经常会面临不平衡性(imbalance),必须依赖于实验数据集的先验分布知识,因而预测结果准确率不高。为了解决这一问题,有学者提出了基于非监督的学习算法的关系预测方法。这类方法试图在两个用户之间定义一个合理的相似度指标,越相似的用户越有可能产生关系。非监督的方法不用考虑关系预测的数据集的先验分布知识,因而准确度较高。经典的相似度指标包括 Adamic-Adar 指标、优先连接(preferential attachment)、Katz 指标。最近人们提出了一些新的相似度指标,如基于地理位置的指标^[57]、基于时差的指标^[58]、基于用户观点的相似性指标^[59]等。第 3 种方法是将关系预测问题视为关系推荐问题^[60],推荐系统中的方法就可以应用到关系预测领域。周涛等人则利用关系预测研究了关系产生的演化机制。

4. 社交网络中的群体行为涌现性分析

涌现性(emergence)是指系统在微观层次的个体相互作用在宏观层次造成的并不直接蕴含在微观相互作用中的现象。涌现性被有的研究者称为是复杂系统研究的圣盘(holy grail)。在网络关系社会中,我们认为,没有全局控制的、各自追求自身利益最大化的各个个体,以相对简单的交互规则相互作用,最终形成宏观上的群体效应(collective behavior)。