

社会化标签系统中基于社会网络的知识推送网络演化研究^{*}

易 明 毛 进 邓卫华 曹高辉

摘 要 利用网络分析方法,本文将社会化标签系统中基于社会网络的知识推送网络构造方法描述为六个步骤。借鉴演化理论,将基于社会网络的知识推送网络演化动力归纳为内生动力和外生动力。移植生物进化论的相关理论,提出多重动力机制作用下基于社会网络的知识推送网络演化周期性模型;每个周期的起点就是新一轮社会网络的构建,每个周期内部的演化过程分为选择阶段、遗传阶段、变异阶段。依托原型系统搜集不同时间点的相关数据,利用网络结构计量方法对基于潜在社会网络的知识传播网络演化过程进行实证分析,印证上述理论研究成果。图8。表9。参考文献13。

关键词 社会化标签 社会网络 知识推送 网络分析 网络演化

分类号 G250

Evolution of Knowledge Push Network Based on Social Network in Social Tagging System

Yi Ming, Mao Jin, Deng Weihua & Cao Gaohui

ABSTRACT According to the network evolution theory, the evolution dynamics of knowledge push network can be concluded as the endogenous dynamics and the exogenous dynamics. By transplanting biological evolution theory, this paper puts up a periodic evolution model for knowledge push network based on social network under the action of multiple dynamic mechanisms. The periodic evolution model is iterative: at the start of each period, a brand new social network is constructed, while during each period, the selection stage, the entailing stage and the mutation stage are involved. In the experiments, firstly a prototype system is used to collect relevant data at different times. Then, the whole evolution process of knowledge push network based on latent social network is observed and analyzed empirically by using network structure metrics. As a result, the theoretical significance of the research is confirmed. 8 figs. 9 tabs. 13 refs.

KEY WORDS Social tags. Social network. Knowledge push. Network analysis. Network evolution.

1 引言

当前,用网络的观点描述客观世界的相关学

科有数学领域的图论、物理学领域的复杂网络、社会学领域的社会网络分析、图书情报学领域的信息计量^[1]。虽然这些学科形成了各自的研究范式,但它们都是在实践着网络分析的思想,即:从网络的

^{*} 本文系国家自然科学基金重大招标项目“基于多维度聚合的网络资源知识发现研究”(13&ZD183)和国家社会科学基金青年项目“基于信任的网络社区口碑信息传播模式及其演化研究”(12CTQ044)的研究成果之一。

通讯作者:易明, Email: yiming0415@sina.com

视角观察客观世界,用点代表实体,用连接两点的线表示两个实体之间的关系,用网络结构这种比较独特的方式来描述各种实体之间的特定关系,进而通过形成的关系网络来发现规律,解决问题。

作为一种新的分析方法,网络分析关注的不仅是单个实体的个体特征,更为重要的是各种实体之间的网络结构关系;其直接目的是从网络结构与网络功能的交互作用入手,揭示网络结构对网络功能的影响。综合国内外学者的已有研究,网络分析的主要内容包括:①网络结构计量及其应用,其中网络结构计量指标是核心。利用网络结构计量指标,针对特定时间点形成的网络,可以从网络结构与网络功能的交互作用入手,揭示网络结构对网络功能的影响,进而产生实际的应用。②网络动态演化分析及其应用这是网络分析的动态观,强调设计网络生成机制以建立网络动态演化模型。通过网络动态演化模型,能够呈现网络演化的全过程。搜集不同时间点的网络数据,利用网络结构计量指标进行对比分析,可以发现网络结构如何随时间演化。由此,结合网络结构与网络功能的交互作用,网络动态演化分析不仅能够预测网络功能的发展趋势,而且可以用于网络结构动态优化,使网络功能在网络结构的影响下向有利于主体目标的方向发展。

学者在应用网络分析方法研究基于社会化标签的知识推送问题时,关注的重点是第一个层次,即对社会化标签系统中的社会网络进行网络结构计量分析,进而与知识推送过程有机融合^[2-5]。显然,从网络分析的第二层次来研究基于社会化标签的知识推送问题,能够探索知识推送网络的动态演化规律,进而应用于动态监测并实时优化知识推送网络演化过程,有利于从宏观层面提高知识推送的实际效果,具有重要的学术价值与实践意义。本文依托社会化标签系统中的社会网络,研究基于社会网络的知识推送网络演化问题,以拓展相关研究。

2 社会化标签系统中社会网络的类型

Web2.0产生的一个重要背景就是Web用户

的社会性需求日益强烈,这一需求也是未来发展的主要趋势。社会化标签系统作为Web2.0的典型应用技术之一,可以通过构建如下两种类型的社会网络以满足Web用户的社会性需求。

(1) 显在社会网络

现在通常所讨论的社会网络实质上就是显在社会网络,是指因各种类型的“关系”而形成的网络连接,包括血缘及姻缘的亲属关系、市场上的交易关系、同事之间的合作关系、朋友之间的友谊关系等。这些关系的集合,形成了实际存在的各种关系网络。利用社会化标签系统的社会性功能,Web用户能够在使用社会化标签系统过程中有效地交流与协作,从而可以自主发展显在社会网络。在这种显在社会网络中,Web用户之间有着共同的兴趣,他们很容易彼此了解,能够明确意识到彼此之间的人际关系,从而可以满足Web用户的社会性需求。由此,利用社会化标签系统,可以收集到相关的社会网络资料,进而获得社会网络中的节点信息,构建显在社会网络。

(2) 潜在社会网络

社会化标签体现的是Web用户的思维认知,它是Web用户自由发挥主体能动性的产物,体现了Web用户的认知结构特征^[6],基于社会化标签的Web用户聚类便能够将那些具有相似认知结构的Web用户聚集。在认知心理学中,认知结构就是储存于个人长时记忆系统内知识的实质性内容和它们彼此之间的联系,也就是通常所理解的知识结构,所以可以借用布鲁克斯知识方程式来阐述基于社会化标签的Web用户聚类的实际价值:在布鲁克斯知识方程式 $K(S) + \Delta I = K(S + \Delta S)$ 中, $K(S)$ 为原有的知识结构, ΔI 为情报的改变量, ΔS 为改进的效果, $K(S + \Delta S)$ 为新的知识结构^[7],情报 ΔI 只有与原知识结构 $K(S)$ 相吻合才能相加,从而导致 $K(S + \Delta S)$ 的出现,否则既无情报吸收,又无知识结构的变化。利用社会化标签对Web用户进行聚类后,同属一个聚类的各个Web用户的主观知识结构存在相似的可能性,彼此之间产生知识互动的概率相对较大。根据这种思路,便可以通过基于社会化标签的Web用户

聚类来构建潜在社会网络。在这种潜在社会网络中,Web 用户之间可能根本不认识,彼此之间只是一种潜在的知识互动关系,而且这种关系也难以被 Web 用户所察觉。然而,通过社会化标签系统中的智能代理系统,不仅能够发现这种潜在的知识互动关系,还可以利用这种关系实现 Web 用户之间的有效互动。

3 社会化标签系统中基于社会网络的知识推送网络构造

在社会化标签系统构建显在社会网络或潜在社会网络之后,便可以利用网络结构计量指标中的关系强度、社团等指标建立各种知识推送模型,也就是网络分析方法的第一层次:网络结构计量及其应用。事实上,在基于社会网络的知识推送过程中,同一 Web 用户可以同时属于多条知识推送路径,而且在 Web 用户之间、社团之间、Web 用户与社团之间都有不同层次知识推送路径的交叉和融合,这些众多的知识推送路径便形成了知识推送网络。而且,在不同的时间点,这种知识推送网络将会表现出不同的结构特征,具有明显的动态性。具体而言,社会化标签系统中基于社会网络的知识推送网络构造方法可以描述为六个步骤。

(1)假设前提:已构建社会网络,如显在社会网络或潜在社会网络。

(2)初始状态:将社会网络映射为知识推送网络。起初,社会化标签系统依据社会网络中节点之间的关系强度,利用协同过滤方法进行知识推送,第一步构建完成的显在社会网络或潜在社会网络可以视为知识推送网络的初始状态。理论上,社会化标签系统可以采取综合利用显在社会网络和潜在社会网络的组合推送方法进行知识推送,但最终所形成的知识推送网络与具体的组合推送算法密切相关,具有很大的不确定性,因此本文只考虑单独利用其中一种社会网络进行知识推送而形成知识推送网络的情形。

(3)加边过程:利用社会网络发现新的知识推送路径。一般而论,新的知识推送模型可以通

过两种路径来构建:节点之间知识资源的相似性和节点之间知识资源的异质性。例如,在社会网络中属于同一社团的节点之间的知识资源相似性相对较大,第一种路径便可以通过社会网络中同一社团内部节点之间关系的传递来实现;属于社会网络中不同社团的节点之间的知识资源异质性相对较大,第二种路径便可以通过在不同社团之间建立“桥”来构建(例如不同社团之间的重叠节点),利用“桥”便可以实现异质性知识资源的跨社团流动。

(4)减边过程:在动态演化过程中,如果节点之间的关系强度低于最小阈值,则让它们之间的连边断裂。需要指出的是,无论是显在社会网络还是潜在社会网络,确定最小阈值的方法原理都与该社会网络的构建过程密切相关。以构建潜在社会网络为例,完成了基于社会化标签的 Web 用户聚类之后,一个重要的问题就是判断属于同一聚类的两个 Web 用户之间是否应该建立连接关系,判断的依据就是这两个 Web 用户所拥有知识资源的相似性。具体方法为:首先计算每一个聚类中每一对 Web 用户之间知识资源的相似性;然后依据相似性值的分布情况,通过专家主观判断或者机器自主学习等方式确定一个合适的相似性值作为临界点。如果属于同一聚类的两个 Web 用户对应的相似性值不低于临界点,则他们之间就可以建立连接关系,这个临界点就是进行减边操作时的最小阈值。当然,这个最小阈值还会结合网络演化过程的实际情形动态调整。

(5)边的衰减:当节点度达到一定数值时,该节点将不再与其他节点连边,并且需要选取与该节点关系强度相对较低的邻节点强制进行边的断裂,即使它们之间的关系强度大于步骤四中的最小阈值。每个节点能够容纳的度不能无限增加,这是因为社会化标签系统的负载能力有限,并且每个节点的知识接收能力也有限。同时,这也有利于每个节点都有机会建立新的连接,使基于社会网络的知识推送网络能够可持续演化。

(6)循环:新 Web 用户和新知识资源不断增加,进入第一步重构社会网络。社会化标签系统

中新 Web 用户的增加,意味着新的知识需求在产生;新知识资源的不断增加,不仅意味着知识内容的丰富,同时也意味着 Web 用户知识需求的变化。当这些变化达到一定程度之后,社会化标签系统需要重构社会网络,进而形成新的知识推送网络。当然,在重构社会网络之前,如果技术手段允许,也可以利用已有的社会网络将这些新增的知识资源推送给相应节点,从而改变节点之间的关系强度。

4 基于社会网络的知识推送网络演化动力分析

演化是指事物渐进的变化和长时间的发展过程,它以遗传、变异、选择为基本的存在形式^[8]。动力作为推动事物运动和发展的作用力,是引导事物主动变化的原因,能够使事物产生目的性的变化^[9]。在社会化标签系统中,基于社会网络的知识推送网络演化动力主要分为内生动力和外生动力。其中,内生动力是基于社会网络的知识推送网络自身所具备的持续发展力量,外生动力则是影响它的外部环境因素。

4.1 内生动力

动力具有可控制性,主体能够根据实现目标的需要决定动力的有无及大小。促使事物发生变化的因素有很多种,但它们大多都无法控制,在事物变化的可控制因素中,只有一部分能使事物产生合乎目的的变化。因此,主体需要选取符合目的要求的影响因素作为动力。在基于社会网络的知识推送网络演化过程中,内生动力是推动网络演化的重要力量,主要包括加边机制、边的衰减机制和知识反馈机制。

(1) 加边机制

加边机制是利用社会网络发现新的知识推送路径,从而建立新的边。如前所述,加边机制主要有两种路径。理论上分析,第一种路径是依据节点之间知识资源的相似性,所发现的节点之间的关系属于“强关系”,据此构建的知识推送路径相对可

靠;第二种路径是依据节点之间知识资源的异质性,所发现的节点之间的关系属于“弱关系”,据此构建的知识推送路径相对新颖。

(2) 边的衰减机制

边的衰减机制是基于社会网络的知识推送网络持续演化的必要条件,具体包括本文第3节提到的步骤四(减边过程)和步骤五(边的衰减)。减边过程意味着如果节点之间的关系强度低于最小阈值,则它们之间的连边断裂。同时,当节点度达到一定值时,该节点将不能再与其他节点连边,并且需要选取与该节点关系强度相对较低的邻居节点强制进行边的断裂,以保证该节点能够有机会接纳新的知识源。以上两个方面构成了边的衰减机制,它们均可以促使基于社会网络的知识推送网络持续演化。

(3) 知识反馈机制

在社会化标签系统中,Web 用户之间的交互十分频繁,使得知识反馈可以成为基于社会网络的知识推送模型的重要组成部分。知识推送效果直接影响节点之间的关系强度,因此可以用于社会网络优化,进而实现“社会网络→知识推送→知识反馈→社会网络优化→知识推送……”的良性循环。由此,知识反馈机制也就成为了基于社会网络的知识推送网络演化的重要内生动力之一。引入知识反馈机制之后,即使在不考虑其他动力因素的情形下,基于社会网络的知识推送网络也会自我优化,从而呈现出持续动态演化的发展趋势。

4.2 外生动力

根据耗散结构理论,系统要保持持续发展就需要不断与外部进行能量、信息的交换,即保持系统的开放性^[10]。基于社会网络的知识推送网络不是一个封闭的系统,外部环境的变化不但会直接影响其知识推送过程,而且还会改变社会网络,进而推动基于社会网络的知识推送网络可持续演化。

基于社会网络的知识推送网络与外部环境的交互主要是通过社会化标签系统来传导的。通过与外部环境的交互,社会化标签系统中的 Web 用户、知识资源和社会化标签等元素不断发生变化,

从而促使外生动力作用下基于社会网络的知识推送网络演化。

(1) 知识推送内容的变化

知识推送内容是知识推送模型的基本要素之一,知识推送内容的变化将直接影响知识推送效果。在社会化标签系统中,Web 用户会不断上传新的知识资源,使得社会网络中能够被用于知识推送的知识资源变得更加丰富,从而促发更多知识推送行为的产生。知识推送行为的增加,必然会改变 Web 用户之间的关系强度,然后透过知识反馈机制反映到社会网络之中,使社会网络进一步优化,进而推动基于社会网络的知识推送网络持续演化。但是,在实际应用过程中,如果社会化标签系统无法将新增的知识资源导入已构建的社会网络之中,这种外生动力将不复存在,也就不能推动基于社会网络的知识推送网络演化。

(2) 知识接收者的变化

在知识推送的一般模型中,知识接收者是影响知识推送效果的关键要素,因为知识推送的最终目的就是为了满足知识接收者的个性化需求。在社会化标签系统,Web 用户会不断上传新的知识资源、标注新的社会化标签。这些新增的知识资源和社会化标签一方面丰富了社会化标签系统,另一方面也会在一定程度上暗示 Web 用户的兴趣正慢慢发生变化,而一旦知识接收者的兴趣发生了重要变化,必然会影响知识反馈环节,进而反映到社会网络之中,从而推动基于社会网络的知识推送网络不断演化。

(3) 社会网络的重构

随着时间的推移,社会化标签系统的元素会不断发生变化,社会化标签系统会经历初始的量变到后期的质变过程。例如,社会化标签系统中会新增很多 Web 用户,他们上传了大量的知识资源,并使用了各种社会化标签;同样,在原有的 Web 用户群体之中,一部分会停止使用社会化标签,另外一部分则持续上传新的知识资源,并使用新的社会化标签。这些变化使得先前构建的社会网络的局限性更加突出。第一,大量新增 Web 用户没有融入社会网络,他们

无法获得有效的知识推送服务;第二,已经停止使用社会化标签系统的 Web 用户占据了社会网络的节点,浪费了系统资源,增加了知识推送的计算复杂性;第三,持续活跃的 Web 用户的知识需求会发生重要变化,已有的社会网络已经无法准确体现其所包含的 Web 用户之间的关系特征。由此,社会化标签系统需要重构社会网络,并据此建立知识推送模型,推动其持续演化。

5 基于社会网络的知识推送网络演化总体模型

图 1 描述了社会化标签系统中基于社会网络的知识推送网络演化总体模型。从总体上看,这一演化过程具有明显的周期性特征,每个周期的起点就是新一轮社会网络的构建,由此循环往复。在单个周期内部,可以将知识推送网络演化划分为三个阶段:选择阶段、遗传阶段和变异阶段^[11,12]。其中,选择阶段是起点,遗传阶段和变异阶段可以在选择阶段之后随时出现,彼此之间没有绝对的先后关系,它们只是各种动力作用的不同结果。而且,从前一个周期向后一个周期的转换也会呈现出不同的演化模式特征。

(1) 选择阶段

选择阶段就是在单个周期的起始阶段确定哪些 Web 用户参与知识推送网络,哪些知识资源作为知识内容在网络中进行个性化推送。

基于社会网络的知识推送网络是依托社会网络而构建,而社会化标签系统主要是依据社会网络中节点之间的关系特征进行个性化知识推送。因为在构建完成的社会网络中,已经包含了相关的 Web 用户和知识资源,由此构建社会网络即是完成了选择阶段的核心任务。换言之,构建完成的社会网络便可以视为基于社会网络的知识推送网络在特定周期的初始状态。然后,在相关动力的作用下不断演化。

此外,在外生动力的作用下,如果社会化标签系统需要重构社会网络,基于社会网络的知识推送网络便会进入新一轮的选择阶段,即实现了周期之

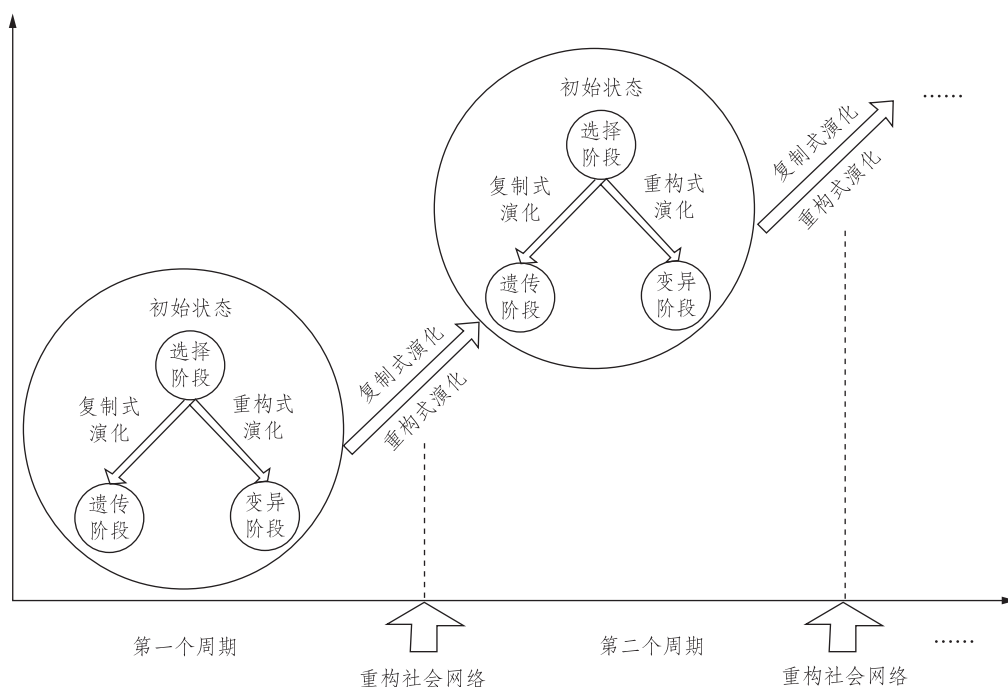


图1 基于社会网络的知识推送网络演化总体模型

间的转换。新构建的社会网络便可以视为图1所示模型在新周期的初始状态。

(2) 遗传阶段

遗传阶段对应着复制式演化模式,即:知识推送网络按照“节点之间知识资源相似性”原则不断复制增长新的连接以实现可持续演化。

在单个周期内部,在加边机制作用下,社会化标签系统将按照 Web 用户之间知识资源的相似性,为 Web 用户进一步挖掘其他可能的知识节点以建立连接关系,从而实现复制式演化。需要指出的是,这种复制式演化模式不仅可以出现在单个周期内部,同时也会体现在前一周期向后一周期的转换阶段。在外生动力的作用下,社会化标签系统需要重构社会网络,以实现跨周期的转换。

社会化标签体现了 Web 用户的认知结构特征,认知结构特征具有明显的继承性与稳定性。由于社会化标签系统中社会网络的构建主要依据 Web 用户的协同标注行为,所以新构建的社会网络与前一个周期对应的社会网络必然有着较强的关联性,并且还可能新增一些拥有类似知识资源的

其他 Web 用户,从而使得前一个周期向后一周期的转换体现了复制式演化模式。

(3) 变异阶段

变异阶段对应着重构式演化模式,即:知识推送网络按照“节点之间知识资源异质性”原则新增具有异质性关系的知识联结以实现可持续演化。

在单个周期内部,在加边机制和边的衰减机制作用下,社会化标签系统可以为特定 Web 用户寻找拥有异质性知识资源的其他 Web 用户以建立连接,逐步实现重构式演化。同样,这种重构式演化模式既可以出现在单个周期内部,同时也会体现在前一周期向后一周期的转换阶段。在外生动力的作用下,社会化标签系统会增加很多新的 Web 用户和异质性知识资源,同时原有 Web 用户的兴趣也会发生重要的改变,因此社会化标签系统需要重构社会网络,重构之后社会网络中的知识资源必然与前一个周期对应的社会网络中的知识资源存在着一定的异质性,从而使得前一个周期向后一周期的转换体现了重构式演

化模式。

6 实证分析

要深入理解社会化标签系统中基于社会网络的知识推送网络演化问题的本质,需要采用规范研究与实证研究相结合的方法。本文依托原型系统构建潜在社会网络,并通过搜集不同时间点的相关数据,利用网络结构计量方法对基于潜在社会网络的知识推送网络演化过程进行实证分析。

6.1 原型系统

原型系统基于“Linux+Apache+MySQL+PHP”程序架构,在“网上二手书 C2C 平台”的基础上,以图书资源作为知识资源实例,并嵌入社会化标签系统的相应功能,进而开发社会化标签资源导航、潜在社会网络构建、知识推送等知识服务功能(见图2)。原型系统利用 Zend Studio 编程工具进行程序开发,借助 PhpStudy 集成开发软件,在 Apache 服务器、PHP 运行环境和 MySQL 数据库中进行系统测试、实施和运行。



图2 原型系统首页

原型系统按功能结构可分为依托平台和实证平台两部分,总体功能结构如图3所示。其中,底层基础数据层为两个平台共同使用的数据库,业务处理与挖掘层及用户交互层在逻辑上可以分别对待。业务数据管理和资源组织体系功能属于依托平台的业务处理层,知识子类生成、知识主题形成、潜在社会网络构建等属于原型系统实证平台的核

心。在用户交互层中,包括了后台管理系统、前台展示系统、知识推送、知识反馈等功能。原型系统中包含在线网站和离线计算两种运作模式。在线网站以 Web 网站形式呈现给用户,进行在线操作。离线计算通过程序开发工具调用、程序运行环境调用等方式直接调用,为了可视化管理和操作,在实证过程中以程序开发工具调用方式进行。

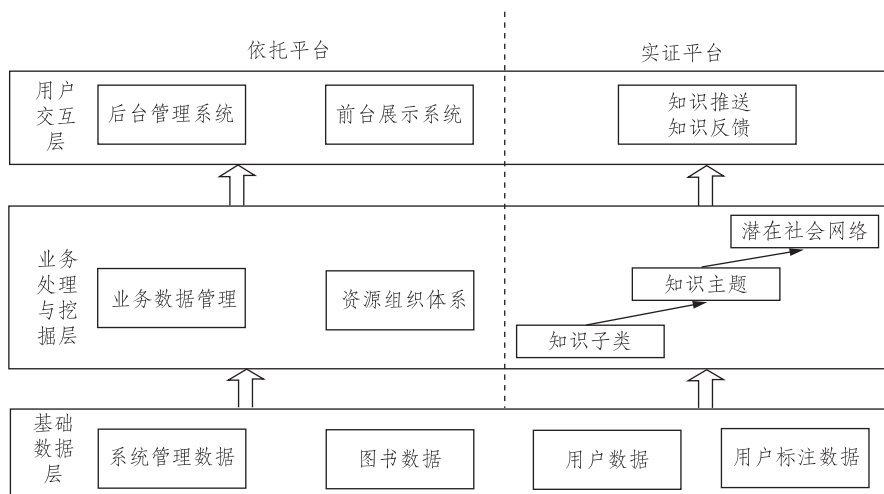


图3 原型系统总体功能结构

6.2 数据准备

原型系统主要需要三类基础数据:用户、图书和社会化标签。在实证过程中,共征集43名本科生和研究生参与调研。图书和社会化标签数据的采集步骤如下:

(1)由15名学生根据学院学生的一般兴趣爱好共选择400本图书,相对均匀地分布在经济与管理、英语与其他外语、考试、小说、计算机与互联网、文学、传记、信息传播等类目下。

(2)每位学生继续添加自己感兴趣而原型系统中还没有的图书,每人至少添加10本图书,且每本图书至少标注3个社会化标签。

(3)每位学生结合自己的兴趣领域,标注自己感兴趣的图书20本以上,且每本图书至少标注3个社会化标签。

经过以上步骤和预处理环节,原型系统的初始数据统计如下:32位有效用户、738本图书、1,610个社会化标签、5,235条社会化标签记录。原型系统中还包含11条无效用户记录,这部分用户只在系统中进行了注册,但并没有任何系统使用记录。

通过离线计算方式,经过知识子类生成、知识主题形成、潜在社会网络构建等步骤(核心方法参见课题组已发表的学术成果^[13]),构建了原型系统中的潜在社会网络,包含30个用户,与之相关联了

533本图书和4,707条社会化标签记录,部分边及权重的示例见表1。利用社会网络分析软件UCINET将其转变为网络拓扑图,并利用NetDraw对其可视化(见图4)。

6.3 IKTN1网络→IKTN2网络的演化分析

依据本文第3节的内容,最初构建的潜在社会网络可以视为知识推送网络的初始状态。在此实例中,数据准备阶段构建的潜在社会网络即是基于潜在社会网络的知识推送网络的初始状态(记为IKTN1网络),原型系统以此为基础进行知识推送,30位用户对接收的图书资源进行知识反馈。在本阶段,并未引入新的用户和新的图书资源,但是30位用户可以对原型系统中自己感兴趣的图书资源继续添加社会化标签。由于原型系统功能的限制,这些新增社会化标签及关联的图书资源并不能自动导入IKTN1网络关联的潜在社会网络。1个月之后,搜集数据,形成了新的基于潜在社会网络的知识推送网络(记为IKTN2网络)。利用NetDraw将IKTN2网络进行可视化显示(见图5)。

6.3.1 IKTN1网络与IKTN2网络的对比分析

基于实证分析的需要,最终选择网络结构计量指标中的网络规模、密度、直径、平均距离、网络群聚系数、线的中介中心度、社团、度中心度、接近中

心度、中介中心度、特征向量中心度、流介数中心度 等 12 个指标进行对比分析。

表 1 潜在社会网络边及权重示例

节点 A	节点 B	权重	节点 A	节点 B	权重
11	19	2	50	54	9
11	26	3	50	55	5
11	27	1	50	58	10
11	28	18	50	59	12
11	29	8	50	60	20
11	33	2	51	59	1
11	34	3	52	54	3
11	39	13	52	55	1
11	41	4	52	58	2
11	43	3	52	59	1

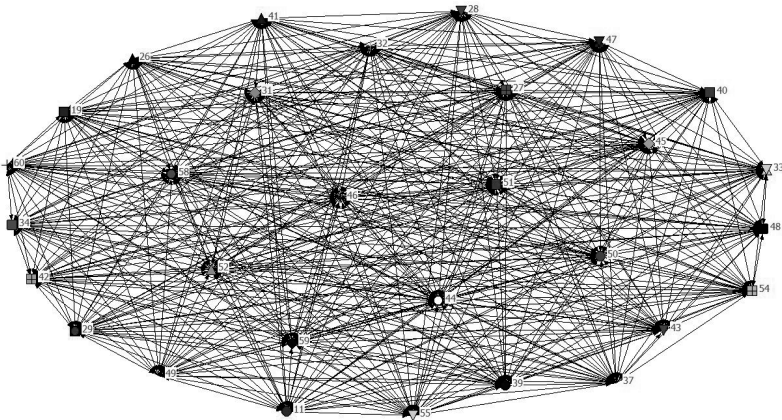


图 4 潜在社会网络拓扑结构图

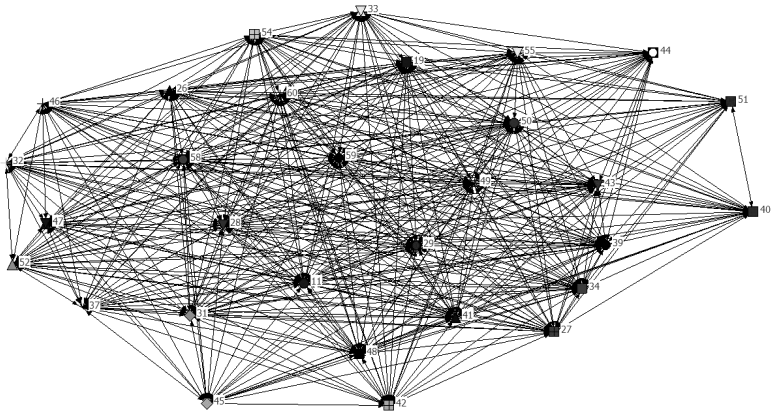


图 5 IKTN2 网络拓扑结构图

表 2 给出了 IKTN1 网络和 IKTN2 网络之间部分指标的对比。可以发现,IKTN2 网络的边由原来的 346 条变为 404 条,网络密度由 0.7954 变为 0.9287。而且,IKTN2 网络的平均距离也相应降低,网络群聚系数也得以提升,达到 0.944。总体来看,IKTN2 网络中节点之间的关系变得更加紧密。

表 2 IKTN1 网络和 IKTN2 网络部分指标对比

指标	IKTN1 网络	IKTN2 网络
网络规模	30	30
密度	0.7954	0.9287
直径	2.000	2.000
平均距离	1.205	1.071
网络群聚系数	0.878	0.944

通过对 IKTN1 网络进行线的中介中心度分析,发现边的最大中介中心度值为 4.495,对应的边为(40,51)和(49,51)(40,49,51 为节点编号),说明这两条边在 IKTN1 网络中处于比较重要的中介地位,但是所有边的中介中心度值的变化幅度并不大。在 IKTN2 网络中,边(48,51)(48,51 为节点编号)在该指标的取值最大,为 1.521。这表明,在加边机制的作用下,IKTN2 网络增加了一些边,从而

削弱了部分边的中介功能,降低了线的中介中心度。

针对 IKTN1 网络,运用 N-Clique 进行社团分析,N 值取为 2 时,得到 1 个社团(见图 6)。同样,针对 IKTN2 网络利用 N-Clique 分析也只得到 1 个社团,与 IKTN1 网络相同。

表 3 和表 4 描述了 IKTN1 网络与 IKTN2 网络在度中心度、接近中心度、中介中心度、特征向量中心度、流介数中心度等五种中心度指标排名前 10 的节点对比。可以发现:第一,五种中心度指标中测度值最大的节点均发生了变化,即在网络中处于各种重要性位置的节点发生了改变;第二,IKTN2 网络中度中心度、接近中心度、中介中心度的前 10 名节点均相同,而在 IKTN1 网络中并未出现这种一致性,这应该是演化动力作用的结果;第三,IKTN2 网络中节点的度中心度、接近中心度、特征向量中心度的测度值较 IKTN1 网络有所提升,而节点的流介数中心度和中介中心度的测度值较 IKTN1 网络有所下降,其主要原因还是加边机制的影响;第四,IKTN2 网络出现了在五种中心度指标均排名第一的节点“11”,而 IKTN1 网络中节点“59”只在三种中心度指标中排名第 1,表明在演化动力的作用下,IKTN2 网络中明星节点的影响力越发突出。

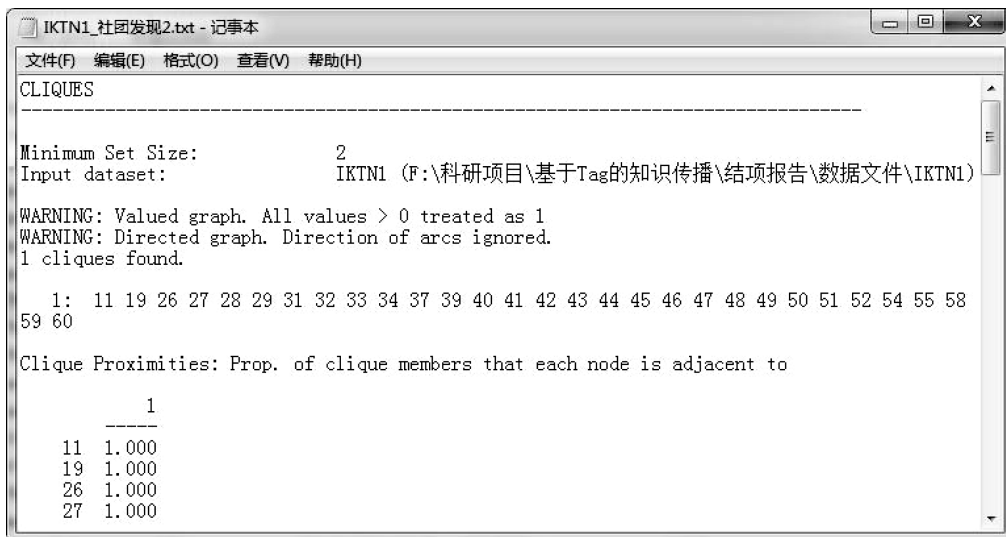


图 6 IKTN1 网络的社团分析结果

表3 IKTN1网络与IKTN2网络的中心度指标排名前10的节点对比(1)

度中心度				接近中心度				中介中心度			
IKTN1 节点	值	IKTN2 节点	值	IKTN1 节点	值	IKTN2 节点	值	IKTN1 节点	值	IKTN2 节点	值
59	100	11	100	59	100	11	100	59	2.159	11	0.375
41	96.552	43	100	41	96.667	43	100	49	1.974	43	0.375
43	96.552	26	100	43	96.667	26	100	41	1.974	26	0.375
49	96.552	27	100	49	96.667	27	100	43	1.573	27	0.375
29	93.103	28	100	29	93.548	28	100	29	1.413	28	0.375
50	93.103	29	100	50	93.548	29	100	33	1.236	29	0.375
27	93.103	49	100	27	93.548	49	100	58	1.151	49	0.375
58	93.103	50	100	58	93.548	50	100	50	1.033	50	0.375
33	89.655	39	100	33	90.625	39	100	54	0.928	39	0.375
34	89.655	34	100	34	90.625	34	100	60	0.91	34	0.375

表4 IKTN1网络与IKTN2网络的中心度指标排名前10的节点对比(2)

特征向量中心度				流介数中心度			
IKTN1 节点	值	IKTN2 节点	值	IKTN1 节点	值	IKTN2 节点	值
44	31.869	11	48.617	39	15.756	11	10.506
54	31.769	39	44.009	42	15.098	39	8.072
37	31.588	28	43.866	58	14.804	28	7.82
47	31.515	50	41.117	33	14.522	29	6.311
50	31.385	29	39.408	29	14.089	50	4.83
59	31.169	60	32.621	45	13.89	60	4.664
19	30.783	34	32.53	51	13.792	43	4.648
41	30.654	58	30.271	55	13.619	37	4.244
28	30.118	43	28.799	60	13.442	41	4.18
32	29.567	19	26.775	47	7.751	11	10.506

6.3.2 IKTN1网络→IKTN2网络的演化动力分析

IKTN1网络→IKTN2网络的演化过程并未涉及新用户和新增的图书资源,而且30位用户新标注的图书资源和社会化标签也不会导入IKTN1网络关联的潜在社会网络之中,所以不能直接分析外生动力。但是,30位用户在继续标注其他图书资源的过程中,如果他们的兴趣发生了较大变化,也就意味着知识接收者的变化(外生动力二),这样仍然会通过知识反馈机制传导到潜在社会网络之中。由此,在IKTN1网络→IKTN2网络的演化过程中,外生动力应该还是存在的,只是作用有限。由

此,此部分重点关注内生动力。

(1) 加边机制

由于IKTN1网络中节点之间已经基本建立了点对点的互动关系,而且也不存在两个以上的社团结构,所以在实证过程中主要利用节点之间图书资源的相似性进行加边操作,从而新增58条边,使得IKTN2网络共有404条边。

表5描述了节点“29”在IKTN1网络和IKTN2网络中的边权重以及权重增量值。需要说明的是IKTN2网络中边的权重计算方法:若用户通过知识反馈确认某一图书资源为其感兴趣的知识资源,则

表明该用户接受了这一知识资源,该图书资源关联的相关用户就会与该用户建立 1 次知识互动关系;若两个用户通过多次积极的知识反馈建立了频繁的知识互动关系,则将积极知识反馈的次数计为边的增量权重,并与 IKTN1 网络中的初始权重相加,得到 IKTN2 网络中边的权重。从最终的变化来看,节点“29”共有 27 条边权重增加,新增边 2 条(分别

是与节点“44”和节点“52”新增边)。

(2) 边的衰减机制

表 6 给出了在此次演化过程中权重未变的边,即这些边的增量权重为 0。若只考虑增量过程,相对于其他边而言,这些边实际上就是衰减了。如果需要减边或者断掉关系强度相对较低的边,那么它们就是首选。

表 5 节点“29”在 IKTN1 网络和 IKTN2 网络的边权重对比

源节点	目标节点	IKTN1 原权重	IKTN2 新权重	权重增量
29	11	8	31	23
29	19	14	21	7
29	26	8	15	7
29	27	7	12	5
29	28	18	32	14
29	31	7	11	4
29	32	2	5	3
29	33	5	11	6
29	34	23	31	8
29	37	2	8	6
29	39	20	31	11
29	40	15	19	4
29	41	8	15	7
29	42	1	6	5
29	43	14	22	8
29	45	3	6	3
29	46	1	5	4
29	47	3	10	7
29	48	2	7	5
29	49	5	10	5
29	50	19	31	12
29	51	2	7	5
29	54	9	13	4
29	55	6	14	8
29	58	13	22	9
29	59	14	17	3
29	60	8	17	9
29	44	-	5(新增边)	5(新增边)
29	52	-	3(新增边)	3(新增边)

表 6 IKTN2 网络中权重未变边

IKTN2 网络中权重未变边(共 56 条)	
(19,37)(19,40)(19,42)(19,46)(26,40)(26,46)(27,31)(27,40)(27,45)(27,46)	
(31,32)(31,37)(31,40)(31,42)(31,44)(31,45)(31,46)(31,47)(31,52)(32,37)	
(32,45)(32,46)(32,47)(32,52)(33,46)(33,52)(34,40)(37,42)(37,45)(37,46)	
(37,47)(37,52)(37,54)(37,55)(40,45)(40,49)(40,51)(40,54)(40,55)(42,45)	
(42,47)(42,52)(42,54)(42,55)(44,46)(44,59)(45,46)(45,47)(46,47)(46,52)	
(46,54)(46,59)(47,52)(47,59)(52,54)(52,55)	

(3) 知识反馈机制

在此阶段的演化过程中,共得到 564 条知识反馈数据。以节点“29”为例,原型系统向其推送了 35 本图书资源,节点“29”在知识反馈过程中确认了其中 20 本为其感兴趣的图书资源。以此为基础,可以调整节点之间的关系强度,进而优化潜在社会网络,以推动 IKTN1 网络→IKTN2 网络的演化。

6.4 IKTN2 网络→IKTN3 网络的演化

在 IKTN1 网络→IKTN2 网络的演化完成之后,实证分析进入第二阶段,即:原型系统环境发生

重要变化,推动潜在社会网络的重构,从而促使网络演化的周期转换。具体而言,前期已注册但无系统使用记录的 11 位用户和原先的 32 位用户继续使用原型系统,他们可以继续添加新的图书资源,并标注自己感兴趣的图书资源。一个月之后,在原型系统中形成了 42 条有效用户记录,845 本图书、1,814 个社会化标签、6,972 条社会化标签使用记录。按照类似的方法,重构的潜在社会网络包含 41 位用户,672 本图书,5,846 条社会化标签记录,它便是新一轮演化周期的初始状态(记为 IKTN3 网络)。利用 NetDraw 将 IKTN3 网络可视化显示结果见图 7。

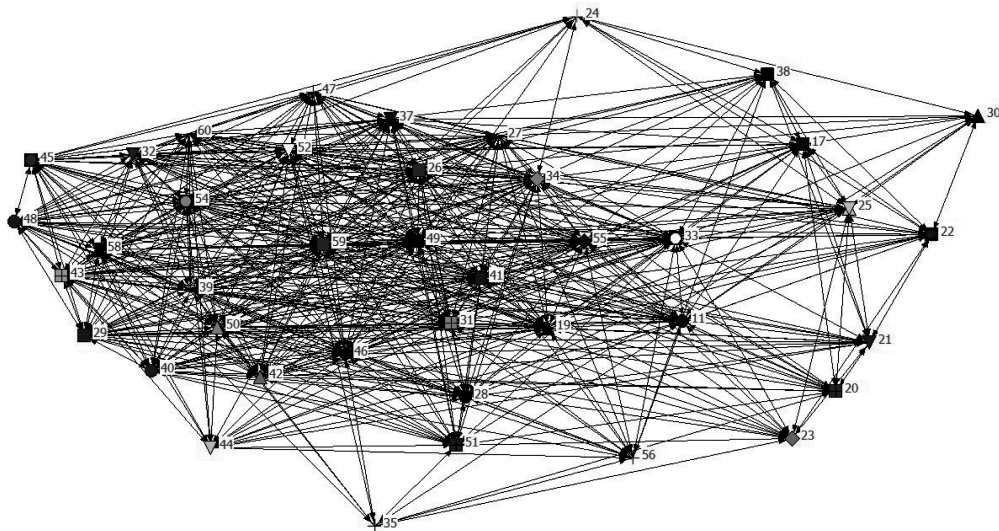


图 7 IKTN3 网络拓扑结构图

6.4.1 IKTN2 网络与 IKTN3 网络的对比分析

表 7 给出了 IKTN2 网络和 IKTN3 网络之间部

分指标的对比。可以发现,IKTN3 网络结构相对松散,网络密度仅为 0.6463,比 IKTN2 网络密度 0.9287

降低很多。虽然 IKTN3 网络的边增加至 530 条,但是直径和平均距离都在增大,网络群聚系数也降低到 0.732。造成这一现象的主要原因就是新用户的引入。随着用户数量的增多,用户之间建立关系的机会在增多,但不会随着用户数量的增加而线性增加。这一点可以通过比较边与节点的比值得到印证:在 IKTN2 网络中边与节点的比值为 13.47 (404/30),而 IKTN3 网络中边与节点的比值为 12.93 (530/41)。

表 7 IKTN2 网络与 IKTN3 网络部分指标对比

指标	IKTN2 网络	IKTN3 网络
网络规模	30	41
密度	0.9287	0.6463
直径	2.000	3.000
平均距离	1.071	1.355
网络群聚系数	0.944	0.732

通过对 IKTN3 网络进行线的中介中心度分析,发现线的中介中心度最大的边为 (26,24),值为

3.925,而 IKTN2 网络中线的中介中心度最大值仅为 1.521。可以认为,IKTN3 网络结构相对松散,因此核心边的中介中心度变大。

图 8 是针对 IKTN3 网络利用 N-Clique 分析得到的社团分析结果,N 取值为 2。结果发现,IKTN3 网络出现了两个社团,表明外生动力正推动 IKTN3 网络出现较为重要的结构性变化。

表 8 和表 9 描述了 IKTN2 网络与 IKTN3 网络在五种中心度指标排名前 10 的节点对比。可以发现:第一,五种中心度指标排名均发生明显变化,即在 IKTN3 网络中各重要性位置的节点均发生了改变;第二,在 IKTN2 网络中度中心度、接近中心度、中介中心度的前 10 名节点均相同,而在 IKTN3 网络中已完全发生了改变,回到了与 IKTN1 网络类似的状态;第三,IKTN3 网络中节点的度中心度、接近中心度、特征向量中心度和流介数中心度的测度值较 IKTN2 网络有所下降,只有节点的中介中心度的测度值较 IKTN2 网络有所提升;第四,IKTN2 网络中出现了在五种中心度指标值均排名第一的明星节点,而 IKTN3 网络中还未出现相似的明星节点,与 IKTN1 网络类似。

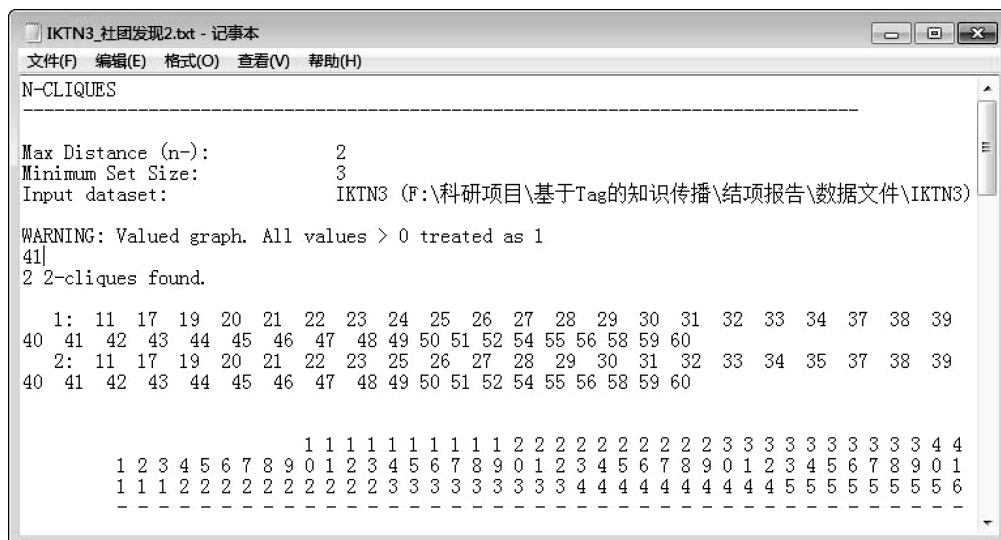


图 8 IKTN3 网络的社团分析结果

表8 IKTN2网络与IKTN3网络的中心度指标排名前10的节点对比(1)

度中心度				接近中心度				中介中心度			
IKTN2 节点	值	IKTN3 节点	值	IKTN2 节点	值	IKTN3 节点	值	IKTN2 节点	值	IKTN3 节点	值
11	100	59	85	11	100	59	86.957	11	0.375	11	1.957
43	100	41	85	43	100	41	86.957	43	0.375	33	1.813
26	100	34	82.5	26	100	34	85.106	26	0.375	46	1.469
27	100	49	82.5	27	100	49	85.106	27	0.375	59	1.449
28	100	33	82.5	28	100	33	85.106	28	0.375	41	1.401
29	100	19	80	29	100	19	83.333	29	0.375	26	1.364
49	100	50	80	49	100	50	83.333	49	0.375	34	1.358
50	100	26	80	50	100	26	83.333	50	0.375	49	1.31
39	100	55	80	39	100	55	83.333	39	0.375	25	1.246
34	100	27	77.5	34	100	27	81.633	34	0.375	39	1.24

表9 IKTN2网络与IKTN3网络的中心度指标排名前10的节点对比(2)

特征向量中心度				流介数中心度			
IKTN2 节点	值	IKTN3 节点	值	IKTN2 节点	值	IKTN3 节点	值
11	48.617	50	37.755	11	10.506	11	3.789
39	44.009	34	36.716	39	8.072	17	2.151
28	43.866	29	36.386	28	7.82	19	3.621
50	41.117	60	32.385	29	6.311	20	2.576
29	39.408	39	31.887	50	4.83	21	2.14
60	32.621	37	31.555	60	4.664	22	2.411
34	32.53	59	30.979	43	4.648	23	1.492
58	30.271	41	30.849	37	4.244	24	1.45
43	28.799	27	30.12	41	4.18	25	2.596
19	26.775	43	29.525	11	10.506	26	2.999

6.4.2 IKTN2网络→IKTN3网络的演化动力分析

IKTN2网络→IKTN3网络的演化体现了外生动力作用下的周期转换。由此,此部分重点分析本文4.2所述的外生动力三,即潜在社会网络的重构。在具体分析过程中,不再重复阐述潜在社会网络的构建方法,而是通过数据说明重构潜在社会网络的必要性。

第一,新用户引入之后,原型系统中的有效用户数量达到42位,原有的潜在社会网络只包含了其中的30位用户,出现了大量用户无法通过已有

潜在社会网络获得知识推送服务的情形。

第二,新的图书资源引入之后,原型系统中的图书资源达到845本,远远超出了已有潜在社会网络所涉及的533本,出现了大量图书资源无法通过已有潜在社会网络推送给目标用户的情形。

第三,新的社会化标签引入之后,原型系统中的社会化标签达到1,814个,涉及的社会化标签记录达到6,972条,远远超过已有潜在社会网络所涉及的1,610个社会化标签和4,707条社会化标签记录,使得原先基于1,610个社会化标签和4,707条

社会化标签记录而构建的潜在社会网络面临严峻挑战。

第四,已有潜在社会网络中的30位用户在继续使用原型系统过程中,其知识兴趣可能会发生重要改变。经过对比分析发现,原有的30位用户中已有7位用户所标注的图书发生了明显变化,暗示着他们知识兴趣的重要改变,使得已有潜在社会网络不能准确反映这30位用户之间的知识互动关系,其局限性进一步显现。

基于以上重要变化,原型系统需要重构潜在社会网络,进而依据新的潜在社会网络建立知识推送模型,从而推动IKTN2网络→IKTN3网络的演化,实现网络演化的周期转换。

6.5 相关结论

在各种动力机制的作用下,通过实证分析所构建的基于潜在社会网络的知识推送网络实现了跨周期的演化,不仅进一步印证了前文的相关研究,而且利用网络结构计量方法进行对比分析之后,还可以得到如下结论。

(1)在网络演化的单个周期内部,在以内生动力为主导的动力机制作用下,网络密度在增大,平均距离在减小,网络群聚系数在增大,线的中介中心度在减小,这些都表明网络结构变得更为紧凑。一方面说明内生动力能够有效促进用户之间的知识共享,同时也能表明网络演化有利于知识推送效果的优化。

(2)在网络演化的单个周期内部,重要节点更能保持自身在网络中的地位,在一定程度上体现了“马太效应”。通过对比节点在IKTN1网络和IKTN2网络中的度数,可以发现绝大部分度数较高的重要节点在演化过程中仍然保持了较高的度数。此外,IKTN2网络出现了在五种中心度指标均排名第一的节点,也是“马太效应”的重要体现。

(3)在网络演化的周期转换阶段,重构的潜在社会网络便是基于潜在社会网络的知识推送网络在新周期的初始状态,其网络结构相对松散,明星节点的地位相对弱化,在很多方面都与上一周期的初始状态相类似,体现了网络演化的周期性特征。

(4)社团指标是监测基于潜在社会网络的知识推送网络演化过程的重要指标,也是复制式演化模式和重构式演化模式产生的重要标志之一。复制式演化模式是按照“节点之间知识资源相似性”原则不断复制增长新的连接以实现可持续演化,一般不会引发社团结构的变化;而重构式演化模式是按照“节点之间知识资源异质性”原则新增具有异质性关系的知识联结,往往会慢慢引发社团结构的变化。IKTN1网络→IKTN2网络的演化主要是复制式演化,结果是IKTN1网络与IKTN2网络都只包括一个社团且成员相同;而IKTN2网络→IKTN3网络的演化蕴含了重构式演化模式,结果是IKTN3网络出现了两个社团。

7 结束语

本文应用网络分析的基本原理,依托社会化标签系统中的社会网络,分析了基于社会网络的知识推送网络演化动力,构建了基于社会网络的知识推送网络演化总体模型。虽然本文的相关理论研究成果得到了实证研究的进一步印证,但是原型系统采集的数据量有限,如果能够进一步扩大实验数据规模,或许能够发现更为深入的规律与结论。同时,本文所构建的知识推送网络主要是基于无向图,是对实际网络一般特征的简化,具有一定的局限性。此外,如何设计有向网络的演化模型,产生与实际网络一般特征类似的拓扑结构,捕捉知识推送有向网络的生成机制与演化规律,也是值得深入研究的一个难点问题。

参考文献

- [1] 易明,毛进,曹高辉,等.互联网知识传播网络结构计量研究[J].情报学报,2013,32(1):44-57.(Yi Ming,Mao Jin,Cao Gaohui,et al. Measuring the structure of web knowledge dissemination network[J]. Journal of the China Society for

- Scientific and Technical Information, 2013,32(1): 44-57.)
- [2] 吴江.豆瓣图书、标签、读者推荐构成的网络分布特征研究[J].计算机与网络,2009(3-4):145-148.(Wu Jiang. Research on the books,tags and recommendation composing network of Douban scattering character[J].China Computer & Network, 2009(3-4): 145-148.)
- [3] Guan Ziyu,Wang Can,Yang Kun,et al. Document recommendation in social tagging services[C]// Proc. of the 19th International Conference on World Wide Web. Raleigh, 2010: 391-400.
- [4] Liu Kaipeng,Fang Binxing. Integrating social relations into personalized tag recommendation[C]//Proc. of the 2nd International Conference on Intelligent Human-Machine Systems and Cybernetics. Nanjing, 2010: 292-295.
- [5] Feng Wei,Wang Jianyong.Incorporating heterogeneous information for personalized tag recommendation in social tagging systems[C]// Proc. of the 18th ACM SIGKDD International conference on knowledge discovery and data mining. New York, 2012: 1276-1284.
- [6] Bielenberg K, Zacher M. Groups in social software: Utilizing tagging to integrate individual contexts for social navigation [D]. Dissertation of Master Degree,Universität Bremen, 2005.
- [7] 贺颖,孟鹏,宋文胜.情报用户知识结构的认知视角分析——布鲁克斯方程式的进一步探讨[J].情报杂志,2003,22(7): 6-8.(He Ying,Meng Peng,Song Wensheng. The analysis study of knowledge structure from the visual angle of cognition-The further discussion for brooks equation [J].Journal of Information, 2003,22(7): 6-8.)
- [8] 库尔特·多普菲.演化经济学:纲领与范围[M].北京:高等教育出版社,2004.(Dopfer K. Evolutionary economics: Program and scope[M].Beijing: Higher Education Press, 2004.)
- [9] 王福涛,钟书华.创新集群的演化动力及其生成机制研究[J].科学学与科学技术管理,2009(8):72-77.(Wang Futao,Zhong Shuhua. Research on the evolutionary dynamic and emergence mechanism of innovation clusters [J]. Science of Science and Management of S. & T., 2009(8): 72-77.)
- [10] 李平.创新集群的演化动力与影响因素研究[J].特区经济,2011(12):280-283.(Li Ping. Research on evolution powers and factors of innovation clusters [J].Special Zone Economy, 2011(12): 280-283.)
- [11] Zollo M,Winter S G. From organizational routines to dynamic capabilities [M].INSEAD,1999.
- [12] 董俊武,黄江圳,陈震红.基于知识的动态能力演化模型研究[J].中国工业经济,2004(2): 77-85.(Dong Junwu, Huang Jiangzhen, Chen Zhenhong. Evolution of dynamic capabilities based on knowledge [J]. China Industrial Economy, 2004(2): 77-85.)
- [13] 易明,曹高辉,毛进等.基于 Tag 的知识主题网络构建与 Web 知识推送研究[J].中国图书馆学报,2011(5):4-9.(Yi Ming,Cao Gaozhui,Mao Jin,et al. Knowledge topic network construction and Web knowledge push based on tag[J].Journal of Library Science in China, 2011(5): 4-9.)

易明 华中师范大学信息管理学院副教授。

通讯地址:湖北省武汉市华中师范大学信息管理学院。邮编:430079。

毛进 武汉大学信息资源研究中心博士研究生。

通讯地址:湖北省武汉市武汉大学信息管理学院。邮编:430072。

邓卫华 华中农业大学公共管理学院信息管理系副教授。

通讯地址:湖北省武汉市华中农业大学公共管理学院信息管理系。邮编:430070。

曹高辉 华中师范大学信息管理学院副教授。

通讯地址:湖北省武汉市华中师范大学信息管理学院。邮编:430079。

(收稿日期:2013-07-22;修回日期:2013-09-15)