



知识标签网络生成机制研究

韩 仪¹, 冯 鑫^{2*}, 周金连³, 吴 晔⁴, 肖井华¹

(1. 北京邮电大学理学院 北京 海淀区 100876; 2. 燕山大学经济管理学院 河北 秦皇岛 066004; 3. 中山大学传播与设计学院 广州 510275;
4. 北京师范大学计算传播学研究中心 北京 海淀区 100875)

【摘要】在知识生成社区中,现实中的网络关系由于具有一些人为实际演化特征,而与传统经典模型并不相符。该文将网络的批量增长和交叉连接特征引入到知乎社区标签网络中,对于网络演化过程进行动态分析,提出了一种标签网络优先连接模型,用来模拟知识标签网络中参数时间序列的复杂性。结果发现,该模型的仿真效果与实际情况下的知识标签网络吻合良好,且用时变方法分析网络的动态趋势较静态分析方法更为合理,该模型有助于后期深入了解现实世界中的复杂网络。

关键词 可调指数; 知识生成; 标签网络; 网络动力学; 幂律分布

中图分类号 O157.5; TP393.0 文献标志码 A doi:10.12178/1001-0548.2020084

The Generation Mechanism of Label Network

HAN Yi¹, FENG Xin^{2*}, ZHOU Jin-lian³, WU Ye⁴, and XIAO Jing-hua¹

(1. School of Science, Beijing University of Posts and Telecommunications Haidian Beijing 100876; 2. School of Economics and Management, Yanshan University Qinhuangdao Hebei 066004; 3. School of Communication and Design, Sun Yat-sen University Guangzhou 510275;
4. Center for Computational Communication Research, Beijing Normal University Haidian Beijing 100875)

Abstract In the knowledge generation community, the network relationship in the real world does not match the traditional classic model because of some artificial evolution characteristics. In this paper, the parameter distribution and time-varying characteristics of the network are introduced into the network, and the evolution process of the network is analyzed dynamically, which provides an objective reference for the construction of the network. Based on the quantitative analysis of label network parameters, this paper proposes a knowledge label connection model based on Barbési-Albert (BA) model, named batch cross-linking BA model, to simulate the complexity of the time-varying parameter series in generation process of the knowledge label network. The simulation results of the model are in good agreement with the actual knowledge label network. This model is helpful for us to understand the complex network in the real world. At the same time, it is more reasonable to use time-varying analysis method to analyze the dynamic trend of the network than static analysis method.

Key words adjustable index; knowledge generation; label network; network dynamics; power law distribution

知识是在社会化合作进程中的产物之一。在线问答社区是知识生成和交流的一种众包平台,用户可以在其中自主互动。以在线问答平台知乎为例,用户可以根据个人认知对问题进行标注。通过出现在同一个问题上的标签之间建立连接,构建出一个标签的共现网络。标签信息具有可使用性,用其作为导航和推荐能有效挖掘出用户的行为和偏好^[1],提高个性化系统推荐的准确性^[2]。标签网络是一种

人工知识网络^[3],它反映了现实世界中知识单元间的复杂关系,通过运用网络分析的方法,找出社会知识系统中的一些特征。例如:度比较大的节点代表知识元素的核心,标签网络中的度分布代表社会知识体系中的宏观主题结构等。此外,通过探索标签网络的动态演化机制,也可以进一步发现知识网络生成和演进的规律。

复杂网络的无标度特性,即幂律分布主导着复

收稿日期: 2020-02-27; 修回日期: 2020-07-17

基金项目: 国家自然科学基金(11905042, 61672108); 教育部人文社会科学研究青年基金(16YJC630022); 河北省“三三三人才工程”(A202001015)

作者简介: 韩仪(1994-),男,主要从事系统科学、复杂网络方面的研究。

通信作者: 冯鑫, E-mail: 149987543@qq.com

杂网络,而非度分布的特定平均指标。以往的研究大多使用抽样数据来验证静态标签网络的无标度特性,证明了高频标签相对较少,而低频标签则相反。文献[4]从2004年-2007年间用户在Flickr平台发布的照片中随机抽取了5200万张包含了至少一个标签的照片^[4],证明标签频率分布和每张照片的标签数量分布都遵循幂律分布。另外一项研究随机选择了20万个问题,发现大约50%的问题包含3个或数量更多的标签^[4]。文献[5]对《美国国家科学院院刊》关键词的统计特性和进化性质进行了频率分布、时间缩放行为和衰减因子分析等经验研究,研究显示1991-2006年间所有关键词出现的频率近似遵循Zipf定律,指数为0.86。文献[6]从知乎和Quora社区随机抽取了100万个问题,观察到问题标签遵循重尾Zipf分布,其中98%以上的标签(Quora=99.02%, Zhihu=98.43%)出现次数不超过500次。同时,许多研究采用复杂网络动态建模的方法研究社会标签规模的增长规律。其中,文献[7]在del.icio.us(美味书签)上研究了用户对图书的标注行为,发现在平台的早期,用户的标注行为在一定程度上呈现出随机性。然而,随着时间的推移,标签规模的分布服从Zipf定律,基于用户标注行为相互模仿机制的模型能很好地解释标签的分布特征。文献[8]以del.icio.us和BibSonomy作为数据源,进一步发现标签增长符合Heaps定律。即随着新标签的增加,系统中不同标签的数量呈指数增长。文献[9]发现已有用户的背景知识可以解释标签增长的Heaps定律。事实上,Zipf定律和Heaps定律并不独立,这两个定律经常同时出现。文献[10]对Zipf定律和Heaps定律之间的关系进行了清晰描述,即Heaps定律是Zipf定律的衍生现象。这类研究探索了网络动态演化过程,但主要关注的是标签网络的增长规律。

综上所述,现有的标签网络研究还存在两个不足之处:一是大部分的研究使用了抽样数据,而在实际的社会系统和网络中,社会标签的分类具有复杂、随机、数据量巨大等特点,采样数据可能存在样本代表性的问题;二是静态网络只能表示数据采集时间节点时的网络的最终状态,不能反映标签网络的动态演化过程。而文献[11]提出无标度网络的幂律特性可以由以下两种机制产生:其一为增长,即实际的网络并不是静态的、不变的,而是通过不断地加入新的节点进行增长;其二为优先连接,即在添加新节点时,与旧节点的连接不是随机

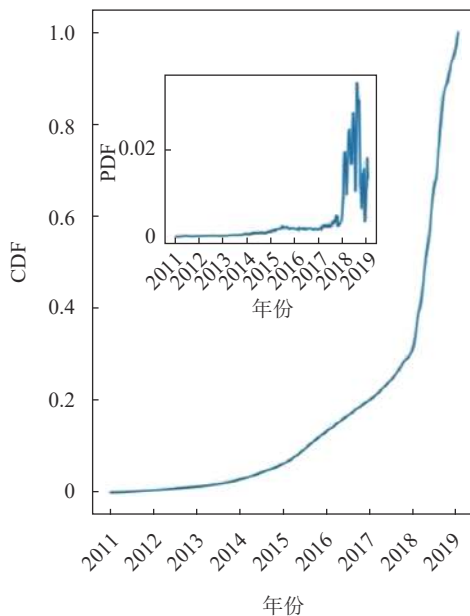
的,而是一个存在优先选择的过程;从而提出了Barbési-Albert模型(BA模型)。基于这两种机制,BA模型可以模拟无标度网络的动态演化,因此被广泛应用于互联网、科学引文网络、短信通信网络、演员合作网络等。自其被提出以来,研究者们陆续在BA模型基础上提出了许多变种模型,主要是通过增加新影响因素来拟合复杂网络的生成结构和统计特征,其中文献[12]提出可以调整优先链接中节点被选中的概率来控制网络结构;文献[13]则从增加重连功能入手讨论网络变化;文献[14]提出可增加节点老化特性来模拟节点随着“年龄”增长而竞争力下降的情形,此外还有学者增加了复制特性来描述引文网络中多重引用的情形^[15]。

上述基于BA模型的众多衍生研究都表明,BA网络作为一个基础普适性模型意义重大且影响深远,但在描述现实生活中一些具体网络时却不能较好地适配,知乎中的知识标签网络就是一个典型例子,它是一种用户生成的社会化知识标签网络,统计分析结果表明它存在无标度特性,且其在动态演化过程中也有增长和优先连接机制。因此,本文利用BA模型模拟了标签网络的形成过程。然而,基础BA模型只反映了复杂网络演化的最基本特征。本文根据知乎平台标签网络演化的特点对模型进行了以下改进:首先,新标签是以用户提出问题的形式生成的,一个问题可以由多个新标签标记,即一次可以在网络中涉及多个标记节点,其中新标记的数量并不一定是一个。其次,在生成连边的过程中,多个标签之间都会产生连边,而不是像传统的BA模型,只在新节点和旧节点之间产生连边。即知乎上问题的产生将改变网络中旧标签的连接状态。因此,本文研究主要涉及两个方面:首先,以知乎平台的数据为基础,验证标签数量的分布特征。然后,通过2011-2018年共计9年间的数据,探索标签网络的动态演化机制,由此探究社会化知识网络的演化规律。

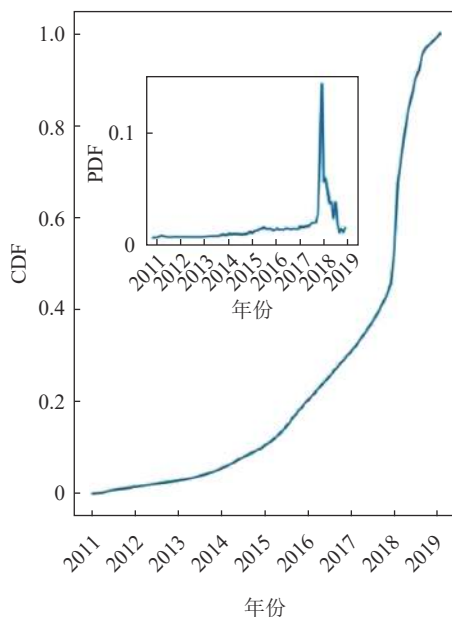
1 数据来源和网络统计描述

在知乎平台上,每个问题都可用多个标签进行标记。由问题标签数据定义的标签网络反映了标签间连接关系,节点代表标签,节点间边代表标签共现关联,即两个标签同时标记某一问题其间会有连边。为了获得标签网络的拓扑特征和时间演化规律,本文使用了完整的知乎数据集,包含问题创建时间和标签,涵盖了2011-2018年的问题。经过清

洗,一共有来自 2034404 个问题的 76379 个不同标签。图 1 绘制了 2011–2018 年问题的累计数量和不同标签累计数量的曲线,研究了数据集的拓扑特征和动态演化趋势,找出了数据的分布特征和演化特征,为建立合理模型提供统计支持。对知乎标签被使用次数与排名之间关系进行分析,未呈现幂律关系,如图 2 所示。再探索不同标签的数量与累积标签使用数间的关系,基本呈现幂律关系,符合 Heaps 定律如图 3 所示。



a. 问题数量的累计分布



b. 标签数量的累计分布

图 1 问题数量和标签数量的概率分布密度曲线图(小图分别代表问题数量和标签数量的概率分布密度曲线)

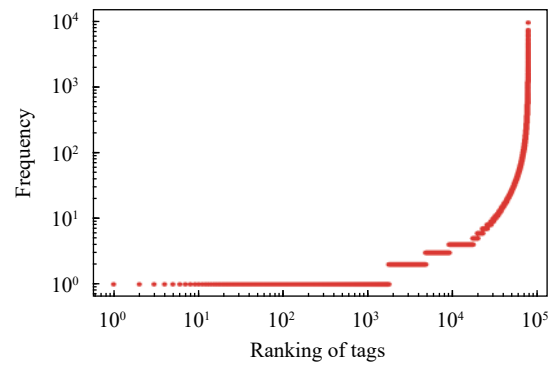


图 2 知乎标签被使用次数与排名的概率分布密度曲线图

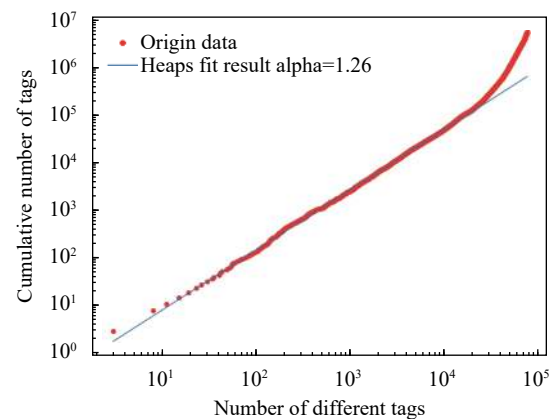


图 3 知乎不同标签的数量与累积标签使用数的概率分布密度曲线图

1.1 度分布

度分布是图理论和网络科学中的一个重要概念^[16],反映了网络的结构特点。网络度分布 $p(k)$ 一般定义为网络中度为 k 的节点占比。在随机网络中,因其节点之间随机连边,度分布呈现泊松分布。无标度网络是复杂网络中另一种重要类型,其典型特点是网络中大多数节点仅有极少边,而少数节点具有大量边,度分布为幂律分布,许多真实世界网络属于此类,如互联网、金融系统网络及社交网络等。经典的 BA 模型用来解释复杂网络的无标度特性。文献 [17] 在对引文网络的度分布进行研究时指出 PR 的引文网络的度分布呈现对数正态分布。而在知乎标签网络中出现了度分布从幂律分布到对数正态分布的演化,幂律分布是具有如下形式的概率分布,

$$P(x) \propto x^{-a} \quad (1)$$

尽管幂律分布很常见,但其检测和描述仍是一个复杂问题,因其分布的尾部通常是波动的。常用幂律分布数据分析方法,如最小二乘拟合,

可能会产生明显的估计参数不准确的问题。本文使用 python 工具包 powerlaw 来解决此问题, 它基于文献 [18-19] 为了对复杂网络的度分布进行分析和拟合而提出的幂律分布分析方法构建。在 powerlaw 中, 可以使用 distribution_compare 工具比较不同分布拟合数据的效果, 返回值代表了不同的候选分布之间的对数似然比, 如果返回值是整数, 代表更加符合第一种分布, 负数代表更加符合第二种分布。当使用这种方法对比对数正态分布和 stretched exponential 的拟合效果时, 返回值是 7.4359187185161595, 代表对数正态分布的拟合效果更好。构建一个全体数据的标签网络如图 4 所示。蓝色星形标记代表度分布数据, 不同颜色实线代表多种拟合方法结果, 标签网络的度分布 (蓝色星形曲线) 明显偏离了幂律分布而更加接近对数正态分布^[20]。

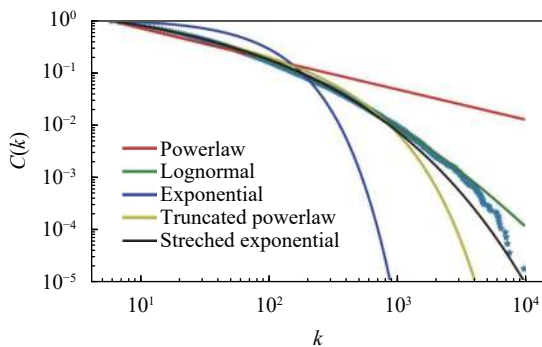


图 4 2011-2018 年数据构建的标签网络的度分布和拟合结果图

为了探究此现象成因, 将数据按照时间以年为单位切片获得 8 个子网络, 计算其网络参数, 可得子网切片的度分布特征。比较不同切片间度分布的差别, 观察标签网络演化和发展的动态过程, 为后续分析标签网络的度分布与幂律分布的偏差提供线索。

对 2011-2018 年各年标签网络度分布进行分析和拟合, 如图 5 所示, 黑色点代表度分布数据, 黑色实线代表幂律分布拟合结果, 红色实线代表对数正态分布拟合结果, 前 5 年更接近幂律分布, 后 3 年逐渐偏离更接近对数正态分布。综上, 在用真实知乎数据构建的标签网络中, 总体网络的度分布服从对数正态分布, 对子网而言网络度分布从幂律分布到对数正态分布的演变趋势。

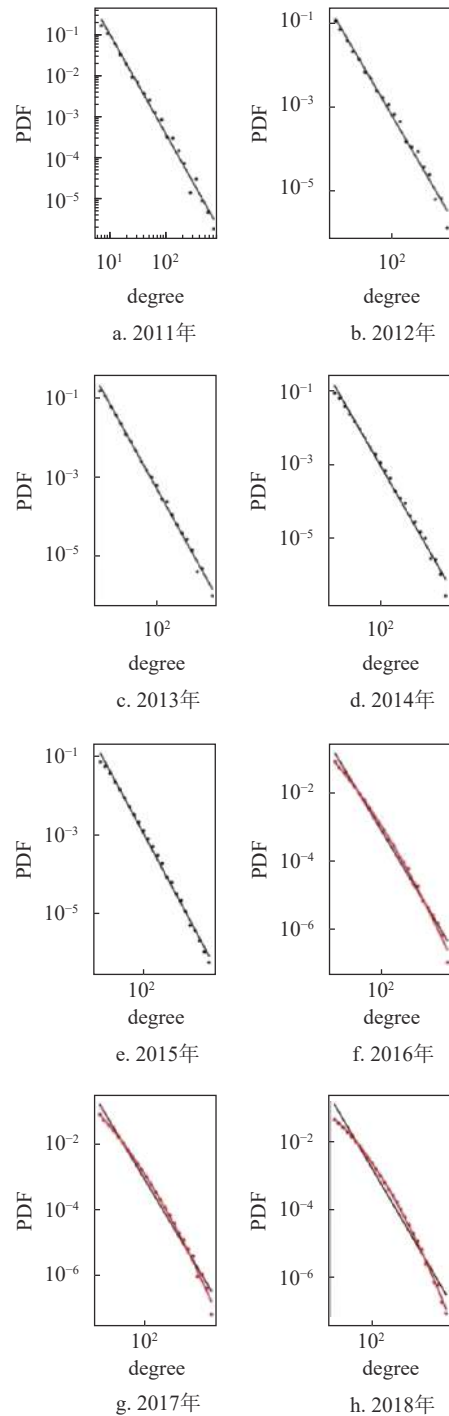


图 5 2011-2018 年年度标签网络的度分布与拟合结果图

1.2 幂律分布指数的改变

用幂律分布对 8 个年份的子网进行拟合, 可以观察到幂律分布的指数变化, 如图 6a 所示。可以发现知乎标签网络的度分布的幂指数明显小于 BA 模型所给出的幂指数 3, 同时可以发现每年的指数总体呈下降趋势, 在 powerlaw 工具中, 当使用幂律分布对结果进行拟合时, 会同时给出幂律分布的拟合指数及标准差 sigma, 当对 8 年的度分布数据

使用幂律分布拟合时,标准差结果如表1所示。

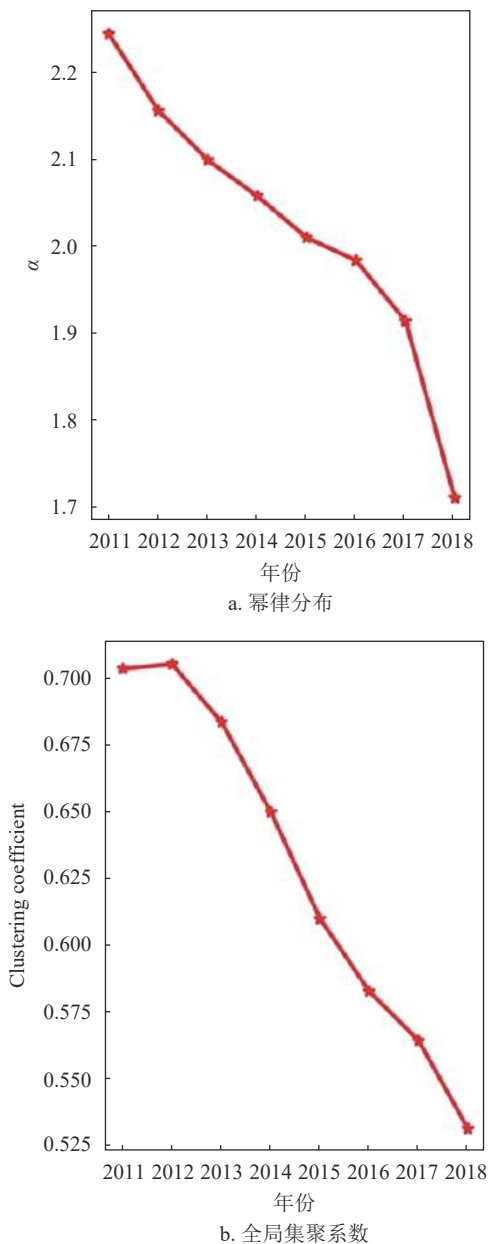


图6 2011–2018年8个子网络的幂律分布的指数和全局集聚系数的变化图

表1 2011–2018年幂律分布拟合指数的标准差

年份	标准差	年份	标准差
2011	0.020 9	2015	0.008 91
2012	0.020 1	2016	0.007 72
2013	0.013 8	2017	0.006 347
2014	0.012 3	2018	0.003 72

这说明度值极大的节点的占比逐渐增加,知乎标签网络中度分布的异化情况逐渐减弱,度值的分布相较以往逐渐呈现一定的平均化趋势。

1.3 网络直径的改变

网络中,节点*i*和*j*的相互通信能力取决于其最短路径,所有节点之间最短路径的最大值定义为网络直径,在一定程度上反映网络连通性。由计算得2011–2018年的8个子标签网络直径,如表2所示知乎标签网络中标签间距离相对较近,尽管随网络规模逐年扩增,其直径依旧基本稳定在8左右,说明标签间距离并未随网络规模增加而显著增加,也从侧面说明了网络节点间连边较为密集。

表2 2011–2018年的8个子标签网络直径

年份	2011	2012	2013	2014	2015	2016	2017	2018
网络直径	10	8	8	8	8	9	8	8

表2给出的每年的网络直径,对每年计算了一个最短路径等于直径的通路,结果表3所示。2016年开始,因为网络规模极度增大,计算两两节点对之间的路径算法复杂度很高,计算一个实例需要数十天,因此只计算了2011–2015年的实例。

表3 tag链条示例表

年份	通路实例
2011	[经济全球化, 经济形势, 欧洲联盟(EU), 经济, 生活, Microsoft Windows, 软件推荐, 网络日记本, 日记, 雷锋]
2012	[ZARD, 日本流行乐(J-Pop), 动漫, 互联网, 世纪佳缘, 艾尚派, 交友网站, 恋爱交友网站]
2013	[康复治疗, 康复医学, 医疗, 美国, 流行音乐, 韩国流行音乐团体(GROUP), 韩国流乐歌(SOLO), 希澈]
2014	[弱者, 强大, 咨询, 职业规划, 私募证券投资基金, 黑石集团(Blackstone), 投资机构职场, Bridgewater Associates]
2015	[交规(待合并话题), 驾驶, 学习, 数据库, 集群, dubbo, 微服务架构, 面向服务的架构(SOA)]

1.4 全局集聚系数

集聚系数常用来描述节点间聚集程度^[21],也可以用于描述节点与邻居的互连程度,可分为全局集聚系数和局部集聚系数两种,前者可评价全网聚集程度,后者可度量每个节点局部聚集程度,如图6b所示,集聚系数较大代表了网络中节点较为密集,其下降代表表明随时间推移和网络规模渐增,网络逐渐变得稀疏。知乎标签网络全局集聚系数呈逐年稳步下降趋势。

2 模型与仿真

BA 模型基于增长过程和优先连接机制两个假设, 其构建过程为 1) 增长: 从一个初始网络 G 开始, 一次增加一个新节点; 2) 连接: 每个新节点都会连接到 m 个旧节点上; 3) 优先连接: m 个旧节点的选择标准是度大优先, 即一个已有节点 i 的度是 d_i , 则新节点选择它的概率是:

$$p_i \propto d_i \quad (2)$$

BA 模型给出了无标度网络生成机制简洁且合理的解释, 已被成功应用于许多种实际网络, 但却只能生成度分布幂指数为 3 的网络, 故对知乎标签网络而言并不适用。这是因为二者生成机制不相同, 且 BA 模型难以解释知乎标签网络度分布指数取值以及从幂律分布到对数正态分布的演化。观察知乎标签网络生成过程及参数分布发现: 1) 每个问题都被一组标签标记; 2) 在一组标签中, 新节点数量并不总是 1; 3) 标记了同一个问题的一组标签之间会产生连边。

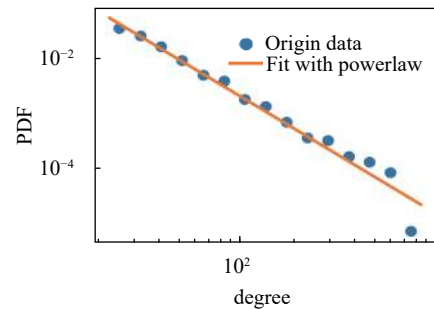
因此, 基于 BA 模型结合知乎标签网络特点, 本文提出了知乎标签网络生成模型。相比前者, 新模型主要做出了以下调整: 1) 批量增长 (batch growing): 网络生成过程依赖新节点加入, 但每次并非单一新节点加入而是有一组节点参与连接, 总数为 m , 其中新节点占比为 p , 故调整后新节点增量为 mp , 旧节点数量是 $m(1-p)$ 。2) 交叉连接 (cross linking): 所有 m 个节点之间均可能产生连接, 即新节点和新节点间、新节点和旧节点间、旧节点和旧节点间均可能产生连接, 已有节点间连接关系可能发生改变。对比 BA 模型新生边必在新和旧节点间, 即边增长只发生于新和旧节点间。3) 优先连接: 旧节点的选择准则依旧和 BA 模型保持一致, 度越大的旧节点具有越高优先度。基于上述改进, 将新模型命名为“标签网络优先连接模型”。对比知乎标签网络生成过程, 从定性角度看, 不难发现标签网络有限连接模型与标签批量标记和节点批量新增的机制更加吻合, 也更能贴近真实地反映知乎标签网络的生成过程。

2.1 模型仿真

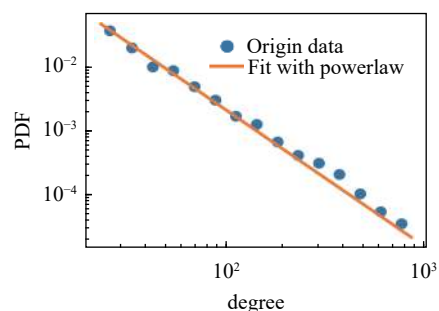
如前所述, BA 模型生成的无标度网络与知乎标签网络的度分布有所差异: 1) 幂指数不同, 前者恒定为 3, 后者明显小于 3 且出现了减小趋势; 2) 度分布的分布类型发生了变化, 后者在后期逐渐呈现了偏离幂律分布的对数正态分布。故为保证标签网

络有限连接模式的有效性, 它必须能生成度分布满足幂律分布且幂指数可以发生变化的无标度网络, 也必须能生成度分布符合对数正态分布的复杂网络。

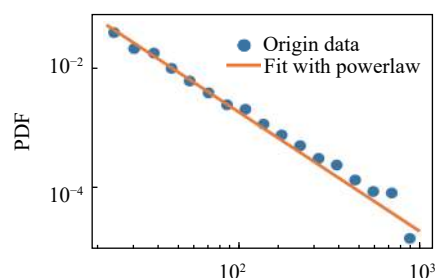
为了验证模型有效性, 通过仿真来研究模型的特点。对于新模型, 如果固定 m 和 p , 可以生成一个符合幂律分布的网络, 如图 7, 使用不同的 p 值产生的网络的度分布和使用幂律分布拟合的结果。蓝色圆形数据点代表度分布数据, 橙色实线代表使用幂律分布拟合的结果。而且幂律分布的指数可以通过 p 进行调整, 如图 8a 所示。此外, 如果在网络生成过程中, 新节点的比例 p 逐渐减小, 可以得到一个符合对数正态分布的网络, 如图 8b 所示。这两个仿真结果表明, 本文提出的模型确实可以满足知乎标签网络的度分布特点。图 8a 为图 7 中幂律分布的指数的变化, 图 8b 通过在仿真过程中逐渐减小 p 值得到的网络的度分布以及使用对数正态分布拟合的结果, 其中的蓝色圆形数据点代表度分布数据, 橙色实线代表使用对数正态分布拟合的结果。



a. $P=0.25$



b. $P=0.20$



c. $P=0.17$

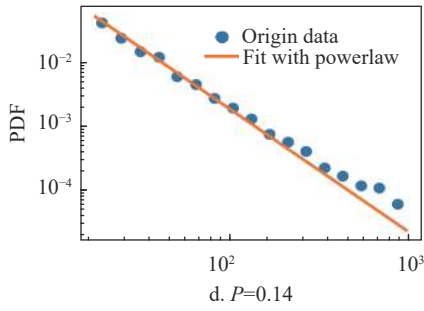


图7 度分布数据和使用幂律分布拟合

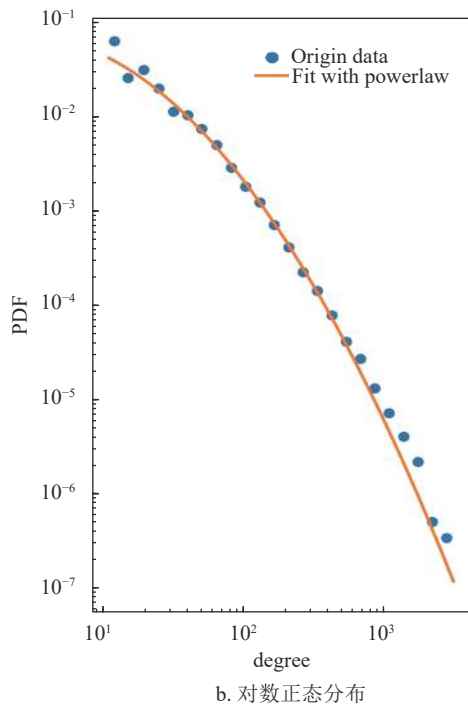
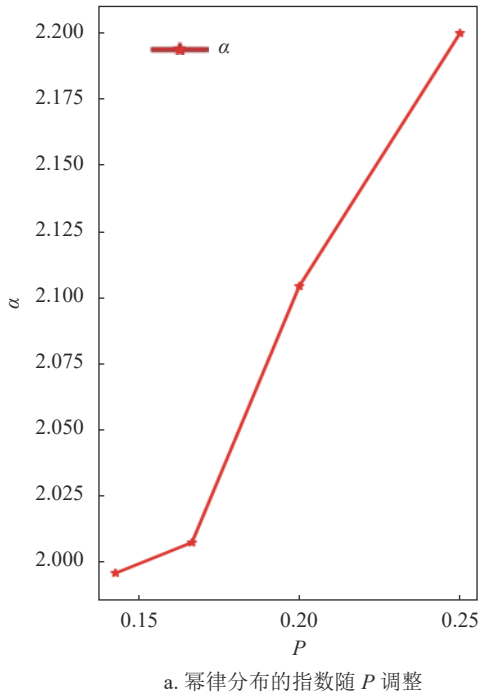


图8 对数正态分布的网络图

2.2 模型验证

为了进一步验证提出的新生成模型适用于知乎标签网络，需要新模型能够根据知乎标签网络的参数，生成仿真的复杂网络。如果仿真的复杂网络的度分布与知乎标签网络的度分布一致，说明本文提出的模型符合知乎标签网络的特点，是有效的。

因此，本文使用从真实数据中提取的 m 和 p 值进行仿真，对比真实网络的度分布和仿真网络的度分布，结果如图9，可以发现仿真网络的度分布与真实网络的度分布吻合较好。

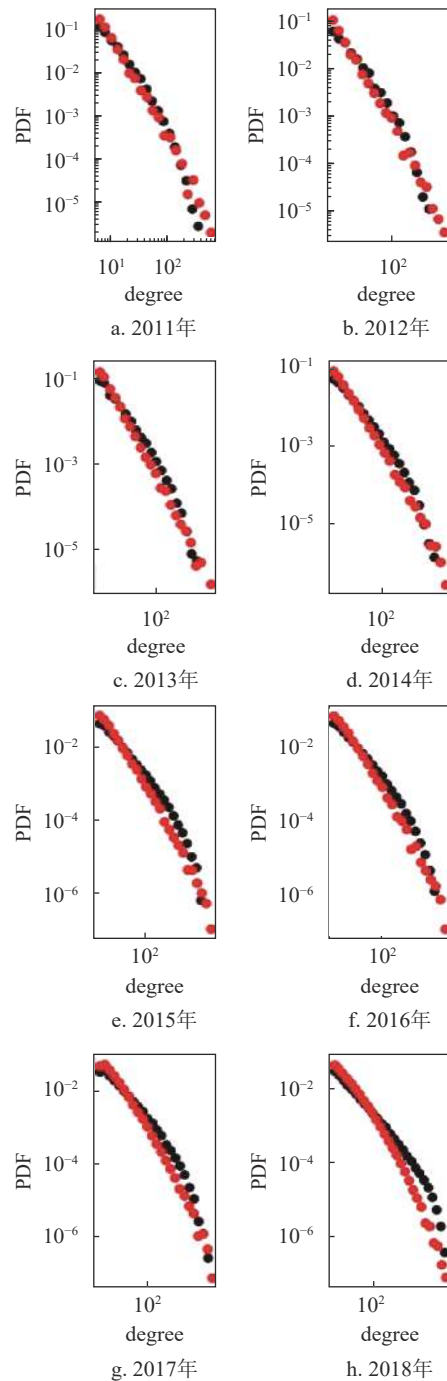


图9 度分布数据图

使用从真实网络中提取的 p 值和 m 值, 对比通过仿真生成的网络的度分布和真实网络的度分布如图 9a~图 9h, 图中的红色数据点代表真实网络的度分布数据, 黑色数据点代表仿真网络的度分布数据。知乎标签网络的度分布从幂律分布逐渐向对数正态分布进行演化, 在此过程中幂律分布的幂指数也逐步下降。度分布呈现幂律分布代表了网络的度分布呈现一种异化的分布类型, 即大多数节点的度都很小, 而极少量的节点具有很大的值。幂指数越大, 度值较大的节点的占比越小, 度分布的异化越明显。知乎标签网络的幂指数的下降以及从幂律分布到对数正态分布的演化都表明, 知乎标签网络的度分布在晚期相较早期有一定的均匀化趋势, 度分布的异化程度逐渐降低。从知识演化的角度来看, 这种变化也在一定程度上反映了社会群体对于知识点的关注从早期的集中化开始逐渐呈现了一定的平均化, 也反映了随着时间的发展社会群体的兴趣逐渐呈现一定的多样化趋势。

通过图中真实网络的度分布与仿真网络的度分布的对比, 也可以看出根据模型仿真得到的结果十分贴近真实结果, 说明了模型中提出的全连接和可变化的新节点比例的假设是有效的。综上所述, 该模型从定性和定量的角度都更加接近真实数据, 可以解释知乎标签网络的度分布以及演化规律。

3 结束语

本文基于 2011-2018 年知乎 2034404 条问答记录的实证数据, 深入分析了其标签网络生成特点和统计规律, 总结出了知乎标签网络“批量增长”和“交叉连接”特性, 并对经典 BA 模型假设进行了符合上述特点的修改适配, 提出了一种新的标签网络优先连接模型, 能够有效地解释知乎标签网络的度分布规律和动态生成机制。

通过仿真模型和实证数据对比发现, 知乎标签网络度分布从早期幂律分布逐渐趋于对数正态分布, 因此经典 BA 模型难以解释此演化过程, 且其度分布幂指数也明显小于恒为 3 的 BA 模型幂指数, 但本文提出的标签网络优先连接模型却能很好地描述知乎标签网络的动态生成过程。该模型可以充分利用现有网络节点和新旧节点间相互关系, 实现批量交叉连接, 能够很好地拟合知乎标签网络的静态结构参数及动态演化机制, 这也表明其生成机制确实受到“批量增长”及“交叉连接”这两个特性的控制, 且其知识标签的连接确实是不同类型节

点间交叉组合连接两两互相连接, 而非只有新旧节点间简单随机连接。该模型揭示了标签网络生成机制, 还原了标签网络生成过程, 实现了调节参数来生成其演化过程中不同幂指数的幂律分布以及对数正态分布, 它也适用于其他多种生成过程与标签网络类似的共现网络, 如: 引文网络和科学家合作网络^[22]等。

知乎标签网络作为一种知识网络, 它可以代表社会群体对于知识的关注, 知乎标签网络的度分布幂指数的减小以及从幂律分布到对数正态分布的演化在一定程度上可以代表人们对于知识的关注点从原本的集中化逐渐变得分散和多样化。同时较小的网络直径和相对较大的集聚系数也代表着不同知识点和学科之间联系比较紧密, 在一定程度上体现了现在的学科融合和知识跨界的现象。对于这类知识网络的研究可以帮助我们理解知识网络的形成, 为进一步研究社会知识的生成, 知识结构的构成、演化与变迁^[23], 社会关注点的变化, 乃至舆情监控^[24]奠定基础。

参 考 文 献

- [1] ZHANG Z K, ZHOU T, ZHANG Y C, et al. Tag-aware recommender systems: A state-of-the-art survey[J]. *Journal of Computer Science and Technology*, 2012, 26(5): 767-777.
- [2] ZHANG Z K, ZHOU T, ZHANG Y C. Personalized recommendation via integrated diffusion on user-Item-tag tripartite graphs[J]. *Physica A: Statistical Mechanics and Its Applications*, 2010, 389(1): 179-186.
- [3] SEUFERT A, BACK A, VON KROGH G. Knowledge management in networked environments: leveraging intellectual capital in virtual business communities[M]. New York: Amacom, 2003.
- [4] SIGURBJÖRNSSON B, VAN ZWOL R. Flickr tag recommendation based on collective knowledge[C]// *Proceedings of the 17th international conference on World Wide Web*. [S. l.]: ACM, 2008: 327-336.
- [5] NIE L, ZHAO Y L, WANG X, et al. Learning to recommend descriptive tags for questions in social forums[J]. *ACM Transactions on Information Systems (TOIS)*, 2014, 32(1): 5.
- [6] WU Y, WU W, LI Z, et al. Improving recommendation of tail tags for questions in community question answering[C]// *The 30th AAAI Conference on Artificial Intelligence*. Phoenix, Arizona, USA: [s.n.], 2016: 3066-3072.
- [7] HALPIN H, ROBU V, SHEPHERD H. The complex dynamics of collaborative tagging[C]// *Proceedings of the 16th International Conference on World Wide Web*. [S. l.]: ACM, 2007: 211-220.
- [8] CATTUTO C, BARRAT A, BALDASSARRI A, et al.

- Collective dynamics of social annotation[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(26): 10511-10515.
- [9] DELLSCHAFT K, STAAB S. An epistemic dynamic model for tagging systems[C]//Proceedings of the 19th ACM conference on Hypertext and Hypermedia. Pittsburgh, PA, USA: ACM, 2008: 19-21.
- [10] LÜ Lin-yuan, ZHANG Zi-ke, ZHOU Tao, et al. Zipf's law leads to heaps' law: Analyzing their relation in finite-size systems[J]. *PLoS One*, 2010, 5(12): e14139.
- [11] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [12] KRAPIVSKY P L, REDNER S. Organization of growing random networks[J]. *Physical Review E*, 2001, 63(6): 066123.
- [13] DOROGOVTSEV S N, MENDES J F F. Evolution of networks[J]. *Advances in Physics*, 2002, 51(4): 1079-1187.
- [14] DOROGOVTSEV S N, MENDES J F F. Evolution of networks: From biological nets to the internet and WWW[M]. Oxford: OUP Oxford, 2013.
- [15] KRAPIVSKY P L, REDNER S. Network growth by copying[J]. *Physical Review E*, 2005, 71(3): 036118.
- [16] NEWMAN M E J. The structure and function of complex networks[J]. *SIAM Review*, 2003, 45(2): 167-256.
- [17] REDNER S. Citation characteristics from 110 years of physical review[J]. *Phys Today Online*, 2005, 58(6): 49-54.
- [18] CLAUSET A, SHALIZI C R, NEWMAN M E J. Power-law distributions in empirical data[J]. *SIAM Review*, 2009, 51(4): 661-703.
- [19] ALSTOTT J, BULLMORE D P. Powerlaw: A python package for analysis of heavy-tailed distributions[J]. *PLoS One*, 2014, 9(1): e85777.
- [20] SHERIDAN P, ONODERA T. A preferential attachment paradox: How preferential attachment combines with growth to produce networks with log-normal in-degree distributions[J]. *Scientific Reports*, 2018, 8(1): 1-11.
- [21] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440.
- [22] BARABÁSI A L, JEONG H, NÉDA Z, et al. Evolution of the social network of scientific collaborations[J]. *Physica A: Statistical Mechanics And Its Applications*, 2002, 311(3-4): 590-614.
- [23] PALLA G, BARABÁSI A L, VICSEK T. Quantifying social group evolution[J]. *Nature*, 2007, 446(7136): 664-667.
- [24] 顾明毅, 周忍伟. 网络舆情及社会性网络信息传播模式[J]. *新闻与传播研究*, 2009, 16(5): 67-73.
GU Ming-yi, ZHOU Ren-wei. A research on social network information distribution pattern with internet public opinion formation[J]. *Journalism & Communication*, 2009, 16(5): 67-73.

编辑 税红