

大数据驱动的开发社区中知识交流网络的分析

达一菲¹ 刘旭东² 孙海龙²

(北京航空航天大学中法工程师学院 北京 100083)¹ (北京航空航天大学计算机学院 北京 100083)²

摘 要 开发社区一般包括博客、问答和论坛等多个版块,这些版块共同构成了用户贡献和交流软件开发知识的平台。文中以 CSDN 为研究对象,通过分析 CSDN 平台上积累的大数据构建了开发者知识交流网络,并基于复杂网络理论对其进行了分析,发现多版块的知识交流网络具有小世界、无标度等复杂网络特性。基于知识交流网络进一步分析了知识贡献者的分布情况,发现多版块用户中有较多的知识贡献者,其在知识交流网中起着比较重要的作用。

关键词 开发者知识交流,复杂网络,大数据,数据分析

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.09.017

Big Data Driven Analysis of Knowledge Exchange Network in Developer Community

DA Yi-fei¹ LIU Xu-dong² SUN Hai-long²

(Sino-French Engineer School, Beihang University, Beijing 100083, China)¹

(School of Computer Science and Engineering, Beihang University, Beijing 100083, China)²

Abstract Developer community generally has various sections such as blog, QA, bbs, etc. These sections together constitute a platform where users contribute and communicate software development knowledge. This paper concentrated and analyzed the big data accumulated in CSDN and constructed knowledge exchange networks. Analysis of the networks based on complex network method indicates the small-world effect and scale-free property of multi-section knowledge exchange network. Further study shows that there exist relatively more knowledge distributors in multi-section users and they play a relative more important role in the knowledge exchange network.

Keywords Developer knowledge exchange, Complex network, Big data, Data analysis

1 引言

开发社区,如 GitHub¹⁾, Stack Overflow²⁾, CSDN³⁾, 已经成为软件开发人员的重要知识来源。一方面,这些社区积累了大量代码、技术问答、技术博客,即“知识”;另一方面,这些社区提供了技术交流的平台。GitHub 的 pull/request 的过程、问答平台一问一答的过程、博客平台互相评论的过程为用户提供了沟通和学习的渠道,这些渠道将用户们联系起来,使存储在用户大脑中的知识也可能成为平台来为其他用户提供帮助。因此,开发社区实际能为用户提供的帮助可以分为两部分:已经存储在平台服务器中的“显式”的“知识”,以及存储在用户们大脑中的“隐式”的知识。如果能够知道用户擅长什么以及擅长的程度,哪些用户愿意热心地回答别人的问题,并有针对性地去询问他们,那么这些“隐式”的知识就能够“显现”且得以利用。

在以往的研究中,对开发者知识交流特点的研究几乎都

只针对某一种开发者关系,如研究 GitHub 网站中软件合作开发关系,研究 Stack Overflow 网站中技术问答关系,研究博客网站的博文评论关系等。然而,一个优秀的博主是否也是一个优秀的问题回答者呢?目前,问答平台 Stack Overflow 已经在用户属性中增加了“GitHub 账号”一项,但大部分用户还没有使用这项功能把两个网站关联起来。本文在前人的研究基础上,将多版块开发社区作为研究目标,结合多个版块,研究知识交流特点,从而更好地揭示各个版块相互影响的作用,得到多版块用户的行为特点。

CSDN 是世界最大的中文技术论坛,可通过一个账号使用博客、论坛、问答等多版块。本文将 CSDN 的博客版块和问答版块联系起来,研究多版块的用户知识交流行为,更好地利用开发社区以辅助软件开发人员的工作。

2 相关工作

目前,对于单独的博客版块和单独的问答版块,已有大量

¹⁾ <https://github.com/>

²⁾ <https://stackoverflow.com/>

³⁾ <http://www.csdn.net/>

投稿日期:2017-10-03 返修日期:2017-12-20 本文受国家重点研发计划课题(2016YFB1000804)资助。

达一菲 女,硕士生,主要研究领域为软件开发和数据挖掘;刘旭东 男,主要研究领域为软件开发方法及工具、软件中间件技术等;孙海龙 男,博士,副教授,主要研究领域为软件开发、分布式计算等,E-mail:sunhl@act.buaa.edu.cn(通信作者)。

的文献对开发者知识交流特点进行了研究。文献[1]对博客平台的评论过程进行了研究,将博客看作由特定部分(specific part)组成的集合,每个特定部分描述了一个完整的事情或观点。将所有特定部分和博客评论结合在一起组成新的集合,该集合即可吸引其他用户来评论。文献[2-3]以 Stack Overflow 为研究对象。文献[2]综述了对问答平台的研究中采用的特征,将所有特征分成3类:关于问题结构的特征(number of tags, length of question 等),关于问题质量的特征(number of views, score 等),关于用户提问历史的特征(reputation 等),并通过 Weka 进行统计和计算。该研究认为关于问题质量的特征是决定一个问题能否得到回答的最重要的因素。文献[3]则利用卷积神经网络研究了问答网站现存问题之间的语义联系。上述文献分析了可成为知识的特征,但没有在知识与用户之间建立联系。

文献[4-6]构造了“知识交流网络”,研究了问答平台上用户知识交流的特点(文献[5]将 CSDN 论坛作为问答平台的一种形式)。知识交流网首先分析了用户以问答为中间介质的知识交流特点,之后去掉中间介质,构建用户之间的网络;然后将用于网页链接分析的 PageRank, HITS 等方法应用于开发者关系的研究中,刻画了知识交流网的整体性质。上述文献研究了问答平台知识交流网络的特点,但对于多版块网站,问答平台不能全面地刻画用户知识交流行为。

本文基于前人的工作,将单一版块知识交流网络的特点应用于多版块知识交流网的构建和预测中。

3 多版块开发者知识交流网络的构建

3.1 数据获取

本文数据来自 CSDN 的博客和问答两个版块,博客版块的用户有 27079 人,共 1037101 篇博客,博客评论 81573 条。问答版块用户主要的知识交流方式有两种:回答其他用户的问题;评论别人的问题或回答。本文共获取了 157437 条问题,259269 条回答记录,70481 条评论别人问题和回答的记录。其中,参与了问题回答活动的用户有 103590 人,参与评论问答活动的用户有 30277 人。

对于获取的数据,本文首先进行了数据清洗。

每个 CSDN 用户有其独一无二的、字符串形式的 id,显示在用户主页上。由于博客、问题、回答、评论上显示的是该用户的昵称,因此首先通过博客、问题、回答、评论获得用户主页链接,以此获得用户 id。本文将此 id 作为主要研究对象。

对于研究范围,CSDN 有大量不活跃用户(只阅读别人发布的内容,自己不发布任何内容),由于未在网站上留下任何痕迹,无法进行研究,因此不在本文研究范围内。另有大量只进行过 1-2 次内容发布的用户,数量众多,对网站做出了贡献,本文将这种现象纳入研究范围。其余数据均为复杂网络研究的必要数据,不做筛选。

对于数据缺失情况,有少量发布了内容之后注销账号的用户,通过发布内容无法链接到其主页。这种情况只出现过 2 次,相对研究而言不具有普遍性,因此删除响应内容并忽略该用户。

经过整理,得到互相之间有评论关系、问答关系的用户列表,将其作为以下研究的主要对象。

3.2 单一版块知识交流网络的构建

本文对 CSDN 博客和问答版块进行研究,构建多版块的开发者知识交流网络。

首先,分别对两个版块构建知识交流网络。

在博客版块中,开发者的交流沟通主要通过评论博主的方式进行,因此构建如图 1 所示的网络。

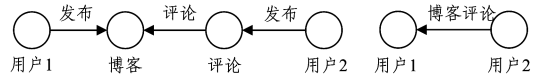


图 1 博客版块的知识交流网

Fig. 1 Knowledge exchange network in blog section

通过研究 CSDN 发现,对博客的评论主要包括感谢博主或进一步提问来请博主解答。因此,博客评论关系说明写博客的用户 1 在该领域的知识水平高于评论人用户 2。本文为了使用 PageRank 算法,用从 2 指向 1 的连线来代表这种关系。

问答版块中有两种关系,即回答问题和评论问答,分别构建如图 2 所示的网络。图中,用户连线方向与博客版块的原因相同。

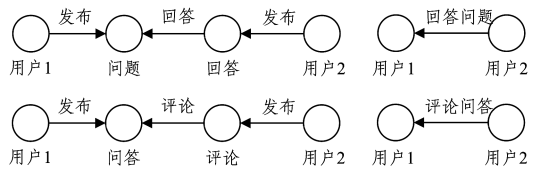


图 2 问答版块的知识交流网

Fig. 2 Knowledge exchange network in QA section

3.3 多版块知识交流网络的构建

首先,本文对 3.1 节中获取的数据进行筛选。选取博客问答网中至少使用过 1 次问答版块的用户;选取问题回答网和问答评论网的并集中,至少使用过 1 次博客版块的用户。将经过筛选的用户作为点,将博客评论关系、问题回答关系、问答评论关系均作为边,建立开发者网络。

值得注意的是,经过筛选,多版块开发者知识交流网络中,每个点至少与其他点有 2 次联系。这些联系以不同的属性形式存储在连线中。

4 知识交流网络的量化分析

4.1 研究问题与假设

4.1.1 研究问题

以往的研究利用复杂网络对知识交流进行研究,对网络的整体进行描述和刻画,以合理量化网络的性质。这些研究为本文提供了思路,即:

RQ1:多版块开发者知识交流网复杂网络特性——多版块开发者知识交流网络的网基本特征与构成该网络的单一版块知识交流网络是否一致?

复杂网络作为研究网络结构的重要工具,常用于对社交网络的分析中。

1) 出入度分布是网络重要的性质。CSDN 知识交流网度的分布可以刻画出 CSDN 用户中不同角色在网络中的分布规律。

2) 出入度相关性在本文的知识网络中可以刻画用户主动回应别人和用户得到别人回应的比例。

3) 网络聚集度可以刻画社交网络中人与人之间形成的明显的“群”或“簇”。对于 CSDN 知识交流网络,可以刻画 CSDN 用户是否会因为知识交流关系形成一定的社交群(cluster)。

4) 小世界性质和无标度性质。规则网络(regular networks)具有大的集聚系数和大的平均距离;随机网络(random networks)具有小的集聚系数和小的平均距离。1998 年,Watts 等^[7]构造出一种新网络,其具有大的集聚系数和小的平均距离,即小世界效应(small-world effect)。1999 年,Albert 等^[8]基于万维网提出了“无标度网络”的概念。真实世界的复杂系统多具备小世界和无标度的特征。

本文将多版块开发者知识交流网络的上述性质进行分析,并与单一版块的性质进行对比。

RQ2:多版块开发者网络的知识贡献者发现研究——多版块开发者网络中是否包含了更多的知识贡献者?

多版块开发者知识交流网中,开发者参与了至少两种活动,不难想到,这些用户在 CSDN 的活跃度有可能相对较高,为其他用户贡献的知识可能较多。本文希望对该网络进行知识贡献分析。

对于社区问答中的专家发现已有很多不同的技术和方法,主要方法有用户模型、基于语义主题的分析方法和链接分析算法。本文主要从用户本身和知识交流网的角度对开发者社区中的知识贡献者进行识别和研究。

基于用户模型的方法是基于用户以往的贡献来对用户能力进行打分和评级排名^[9]。本文采用出入度作为知识贡献者的标准,入度越高,说明其对别人的帮助越大。

基于链接分析的方法,Jureczyk 等^[4]将用户看作节点,并利用 HITS 算法寻找权威用户^[10]。此外,林鸿飞等^[6]验证了基于类别参与度的专家发现,PageRank 算法的效果相对更好。本文采用 PageRank 算法^[11],利用已经构建的知识交流网络进行分析,定义公式如下:

$$P(i) = \frac{(1-\sigma)}{n} + \sigma \sum_{k=1}^n \frac{P(Y_k)}{C_{out}(Y_k)}$$

其中, n 为用户总数, $P(i)$ 为用户 i 的 PageRank 值, $P(Y_i)$ 为知识交流网中连接到用户 i (即得到用户 i 的知识的用户)的 PageRank 值,初始值均为 $1/n$ 。 C_{out} 为 Y_i 的出度, σ 的取值设为 0.85。

4.1.2 假设

根据数据基础统计结果和以往研究结果,做出下列假设。

假设 1 对于出入度分布,多版块知识交流网与单一版块知识交流网具有相同的特征和趋势。

根据文献^[5],基于 CSDN 论坛的知识交流网络的出入度符合幂律分布。而通过对 CSDN 博客和问答版块的基础数据统计可知,在每个版块中,用户出入度的最大值较大(几千或几百),而出入度的标准差在 1 以上。因此可以假设:在博客版块、问答版块、多版块的知识交流网络中,点的出入度分布成反比关系,均呈现幂律分布。

假设 2 对于集聚系数,几种知识网络均具有比同等规

模随机网络更高的集聚系数。

通过查看样本发现,有的用户会经常回复同一个人的发布内容。因此本文假设,通过知识网可以基于兴趣形成一定的“群体”,群体内的成员虽然没有形成稳固的社交关系,但相对于随机网络,具有一定的明显的集聚性。

假设 3 多版块、博客版块、问答版块的知识交流网络具有小世界和无规则特性。

无标度性质是指网络中节点的度满足幂律分布,由假设 1 中的相同原因,可以假设几种知识交流网络具有无标度性质,节点度服从幂律分布。

$$P(k) \sim k^{-\gamma}$$

小世界网络具有较高的集聚系数和较小的平均距离。与假设 2 相同,本文预计了大的集聚系数。因此,预计知识交流网络是小世界网络。

假设 4 相比于单一版块使用者,多版块知识交流网的用户能为其他用户提供更多的知识。即,博客得到更多的评论,在问答版块更多地回答了别人的问题,在问答版块的问答得到了更多的评论。

多版块知识交流网的用户尝试多种版块,极有可能是网站中相对活跃的用户,从知识交流的角度,其可能更加愿意“帮助别人”,也可能确实具备比较多的知识,因此才有能力帮助别人。这类人即是本文要寻找的知识贡献者。

虽然 CSDN 博客评论的现象很多,但是知识交流的特性并不十分明显,大多数博客评论都是阅读者对作者的感谢(如“学习了,谢谢博主!”)。而有的用户虽然通过博客学习了许多知识,但并不进行评论。由博客评论关系联系起来的用户,虽然往往会因为相同的兴趣和能力继续关注对方,但难以由此定量地反映博客的水平,因此,预计通过博客评论关系进行专家挖掘效果相对较低,同时,预计多版块用户的博客评论量与整体相差不多。

与之相对,问答因为本身“一问一答”的形式容易在得到回答后进行追问、讨论,从而促进社交关系的建立。虽然问答的评论关系与博客版块一样,属于评论关系,但根据样本数据,在问答版块进行评论的人往往也能得到持续的追问和讨论。因此,预计多版块用户也往往是问答版块入度较高的用户。

假设 5 多版块知识交流网的用户,在每个单一版块的网络中处于更重要的位置。即,对博客、问答版块的知识交流网络分别使用 PAGERANK 算法对版块中的用户进行排名,多版块知识交流网络的用户排名较高。

与假设 4 相同,多版块用户中很有可能包含我们希望找到的知识贡献者。如果情况属实,这些知识贡献者在知识交流网络中应该起着比较重要的作用,在网络中占据比较重要的位置。本文使用判断网页重要程度的 PageRank 算法,得到的结果能呈现多版块用户在 PageRank 中的排名,有较多的用户在 PageRank 中的排名较高。

4.2 实验结果

对于 4.1 节中的研究问题和 5 个假设,对数据进行了分析和实验,实验结果如下。

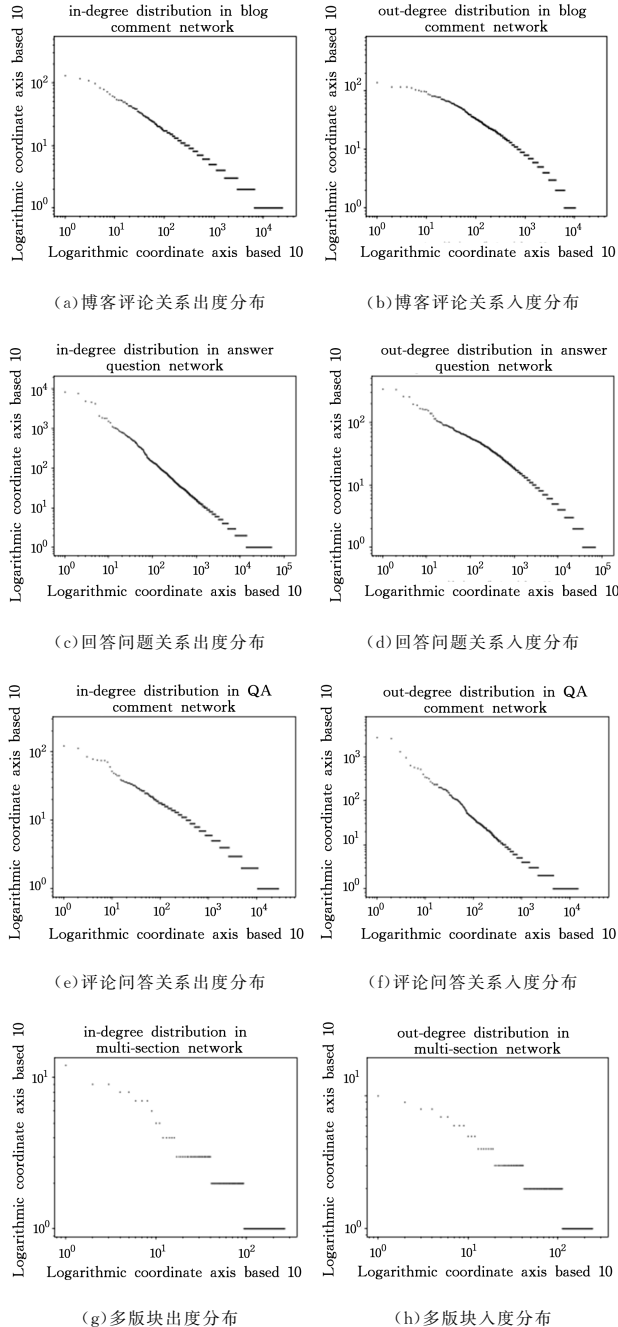
RQ1:多版块开发者知识交流网复杂网络特性

博客版块、问答版块、多版块知识交流网络的基本统计信息如表 1 所列。

表 1 基本统计信息
Table 1 Basic statistical information

	节点数(N)	边数(L)	节点平均度(k)
博客版块	27079	42053	0.55
问答版块回答问题	103590	208637	2.01
问答版块评论问答	37277	54023	1.44
多版块	275	466	1.69

上述 4 个知识交流网的出入度分布如图 3 所示。



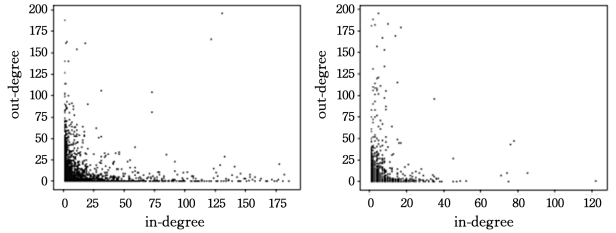
注:图中坐标是以 10 为底的双对数坐标

图 3 知识交流网出入度分布

Fig. 3 Distribution of out/in-degree of knowledge exchange network

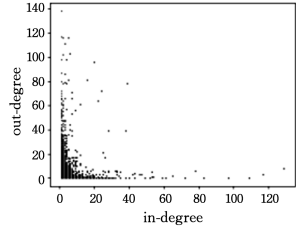
4 种知识交流网络均基本符合幂律分布,由此可以证明假设 1 正确;同时,也说明 4 种知识交流网络均具有无标度性质。

对出入度相关性的研究发现,单一版块的出入度分布具有很弱的相关性,如图 4 所示。



(a) 博客出入度相关性

(b) 问题回答出入度相关性



(c) 评论问答关系出入度相关性

图 4 单一版块出入度相关性

Fig. 4 Correlation of out/in-degree of one certain section

而多版块网络出入度相关性则相对明显,如图 5 所示。

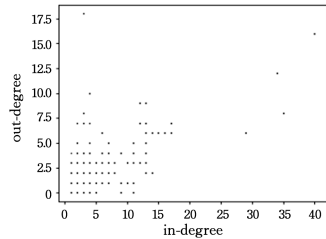


图 5 多版块出入度相关性

Fig. 5 Correlation of out/in-degree of multi-section

对上述网络的聚集系数和平均距离进行计算,并与同等规模网络进行比较,同等规模的 ER 随机网络按照与知识交流网节点数相同、度平均相等的方式构建网络。节点和度满足泊松分布^[12]。

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \approx \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$$

知识交流网络的聚集系数和平均距离如表 2 所列。

表 2 知识交流网络的聚集系数和平均距离

Table 2 Clustering coefficients and mean distance of knowledge exchange network

	聚集系数	同等聚集	平均距离	同等距离
博客评论	0.0118	0.00004	3.26	56
回答问题	0.0189	0.00001	11.39	49
评论问答	0.0431	0.00111	4.27	55
多版块	0.0211	0.011	1.86	9

由计算结果可知,单一版块和多版块上的知识交流网络在聚集系数上与同等规模的网络相比有非常明显的提升,平均距离与同等规模的网络相比,则有明显的减小,符合小世界性质。这说明假设 3 是正确的,同时也证明了假设 2 正确。

RQ2:多版块开发者网络中是否包含了更多的知识贡献者?

在获得了知识交流网后,可以很容易地对网络中的点根

据其出入度进行排名。由第 3.2 节中对网络的构建方法可知,本文所构建的知识交流网中,入度越大,代表为其他人提供的知识越多,即知识贡献者。因此,可以以入度为根据,对博客和问答版块的所有用户进行排名。本文对排名前 10、前 100、前 1000 的人进行分析,寻找其中多版块用户的个数,得到表 3。

表 3 根据入度排名的知识贡献者分布

Table 3 Distribution of knowledge distributors based on in-degree ranking

	前 10 人中 多版块用户 的个数	前 100 人中 多版块用户 的个数	前 1000 人中 多版块用户 的个数	多版块用户占 该版块用户的 比例
博客评论	3	7	13	0.010
回答问题	0	11	43	0.002
评论问答	1	12	42	0.007

从绝对数量来看,每个版块中,多版块用户数量不多,上述 3 个网络中前 1000 名中分别只有 13 人、43 人、42 人是多版块用户。但是从比例来看,多版块用户占整体的比例较小。以前 1000 人为研究对象,由知识提供者分布比例 $0.013 > 0.010, 0.043 > 0.002, 0.042 > 0.007$ 可知,多版块用户中,知识提供者的数量明显高于整个版块的平均水平。其中,博客版块相对较低,符合博客版块不能正确反映知识交流状况的假设。综合可得出,多版块用户对单个版块的贡献相对较高。

当计算为他人提供知识最多的前 1/2 的用户时,多版块用户在博客和问题回答网络分别有 223 人、200 人,即多版块用户几乎都分布在博客和问题回答的知识提供者中。而评论他人问答最多的前 1/2 的用户中,有 130 人是多版块用户,不足全部多版块用户的一半,说明多版块用户不常用评论问答的方式提供知识。因此可以认为假设 4 正确。

使用同样的方法,计算使用 PageRank 算法排序处理的用户列表,得到如表 4 所列的结果。

表 4 根据网络排名的知识贡献者分布

Table 4 Distribution of knowledge distributors based on ranking of network

	前 10 人中 多版块用户 的个数	前 100 人中 多版块用户 的个数	前 1000 人中 多版块用户 的个数	多版块用户占 该版块用户的 比例
博客评论	0	3	16	0.010
回答问题	0	10	43	0.002
评论问答	1	13	48	0.007

仍以 1000 人为研究对象,由 $0.016 > 0.010, 0.043 > 0.002, 0.048 > 0.007$ 同样可以得出结论:多版块用户在整个知识网络中的贡献度较大,处于比较重要的位置。因此假设 5 正确。

4.3 实验结果

上述实验对多版块知识交流网络进行了分析,发现此网络具有与单一版块相对一致的网络特性,且包含较多的活跃用户和有能力用户,说明该数据集值得进一步深入研究。对实验结果的反思主要有以下几点。

1) 针对多版块知识交流网的网络特征分析

在构建多版块知识交流网的过程中可以发现,介于博客和问答版块的多版块用户相对于整体用户量非常少,在单一

版块的数据量均为几十万甚至几百万的情况下,多版块用户人数仅有 275 人。不足 300 人的数据量对于计算结果可能有一定局限性。这种局限性比较小,因为这些数据在数据量较少的情况下体现出了非常明显的规律性。但是在下一步的研究中,将争取获得更多的数据进行实验。

2) 对于知识贡献者分布的研究

在进行知识贡献者发现的过程中,使用了简单的人度和 PageRank 排名作为判断知识贡献者贡献的标准。有可能出现以下几种情况:①一篇博客下虽然有多人评论,但评论人并非因为博客的技术而给予评论,即未进行知识交流;②一个人虽然回答了很多问题,但是很少或者从来没有回答被选为最佳答案;③一个人在问答版块提出的问题或回答引起了非常多人的评论,但实际上该回答本身并没有价值。

通过对网页的调查可知,出现上述情况的可能性非常小,但还是具有一定的可能性。因此在进行知识贡献者发现的过程中,本文重点关注单一版块排名前 1000 的用户。对于这些用户,上述情况出现的可能性较小(比如,排名第 1000 的博主,共获得评论 213 条,这种情况下非技术原因的评论较少,大多数还是技术评论,可以证明博主的知识贡献),对于知识贡献者发现的说服力较强。

下一步的研究中会使用文本分析的方法对知识贡献者的分布进行更准确的描述。

结束语 本文对 CSDN 网站的博客和问答版块进行了研究,利用用户博客评论关系、问题回答关系、问答评论关系建立了 3 个知识交流网络;对数据进行筛选,建立多版块用户知识交流网络。通过对比,主要有以下几点发现。

1) 与单一版块不同,多版块网络出入度呈现明显相关性,多版块用户对其他人的知识贡献越多,得到的知识就越多;2) 与同等规模的网络相比,多版块知识交流网络具有大的聚集系数和小的平均距离,呈现出小世界的性质;3) 对于出入度的分布,多版块知识交流网络出入度均呈现幂律分布和无标度性质,与单一版块相同;4) 相对于单一版块,多版块用户对他人知识贡献次数相对较多,并主要体现在问答版块;5) 相对于单一版块,多版块用户在 Pagerank 算法中的排名较高,并主要体现在问答版块。

本文发现了开发者平台隐含的复杂网络特征,揭示了知识贡献者的版块特征。下一步研究将充分利用网络的特征,参考多维关系的研究,对开发者复杂网络进行深入分析;同时,我们的研究涉及专家发现,将多版块作为重要特征,并将其应用于开发者社区的专家发现和知识热点采集的工作中。

参 考 文 献

- [1] BANSAL T, DAS M, BHATTACHARYYA C. Content driven user profiling for comment-worthy recommendations of news and blog articles[C]//Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 2015:195-202.
- [2] SAHA R K, SAHA A K, PERRY D E. Toward understanding the causes of unanswered questions in software information sites:a case study of stack overflow[C]//Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering. ACM, 2013:663-666.
- [3] XU B, YE D, XING Z, et al. Predicting semantically linkable

knowledge in developer online forums via convolutional neural network[C]//Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering. ACM, 2016: 51-62.

- [4] JURCZYK P, AGICHTSTEIN E. Discovering authorities in question answer communities by using link analysis[C]//Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management. ACM, 2007: 919-922.
- [5] PENG H B, WANG J. Topology of the Knowledge Communication Network in Virtual Communities; Based on CSDN[J]. New Technology of Library and Information Service, 2009, 25(4): 44-49. (in Chinese)
彭红彬, 王军. 虚拟社区中知识交流的特点分析——基于 CSDN 技术论坛的实证研究[J]. 现代图书情报技术, 2009, 25(4): 44-49.
- [6] LIN H F, WANG J, XIONG D P, et al. Category participation-based approach to find experts for community question answer services[J]. Computer Engineering & Design, 2014, 35(1): 333-338. (in Chinese)

林鸿飞, 王健, 熊大平, 等. 基于类别参与度的社区问答专家发现方法[J]. 计算机工程与设计, 2014, 35(1): 333-338.

- [7] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440.
- [8] ALBERT R, BARABÁSI A L. Statistical Mechanics of Complex Networks[J/OL]. Reviews of Modern Physics, 2002, 74: 47-97.
- [9] CHEN L, NAYAK R. Expertise analysis in a question answer portal for author ranking[C]//Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society, 2008: 134-140.
- [10] KAO W C, LIU D R, WANG S W. Expert finding in question-answering websites: a novel hybrid approach[C]//Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 2010: 867-871.
- [11] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [12] 何大韧, 刘宗华, 汪秉宏. 复杂系统与复杂网络[M]. 北京: 高等教育出版社, 2009: 111.

(上接第 103 页)

试状态, 从而影响了测试覆盖率; 在很多状态下, 测试依赖其他状态, 如查看“饿了么”历史订单需要订餐记录, 否则测试难以覆盖。本文的未来工作将会扩展规则的语言, 形成功能通用的 DSL(Domain Specific Language)语言, 以低耦合的方式驱动测试过程, 达到更好的测试覆盖率。

参 考 文 献

- [1] DU R Y, WANG C H, HE K. Location Privacy Protection Technology on Smart Mobile Devices [J]. ZTE Communications, 2015, 21(3): 23-29. (in Chinese)
杜瑞颖, 王持恒, 何琨. 智能移动终端的位置隐私保护技术[J]. 中兴通讯技术, 2015, 21(3): 23-29.
- [2] CHOUDHARY S R, GORLA A, ORSO A. Automated test input generation for android: Are we there yet?[C]//2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2015: 429-440.
- [3] MACHIRY A, TAHILIANI R, NAIK M. Dynodroid: An input generation system for android apps[C]//Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering. ACM, 2013: 224-234.
- [4] AMALFITANO D, FASOLINO A R, TRAMONTANA P, et al. Using GUI ripping for automated testing of Android applications[C]//Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering. ACM, 2012: 258-261.
- [5] AZIM T, NEAMTIU I. Targeted and depth-first exploration for systematic testing of android apps[C]//ACM SIGPLAN Notices. ACM, 2013: 641-660.
- [6] MEMON A M, BANERJEE I, NAGARAJAN A. GUI Ripping: Reverse Engineering of Graphical User Interfaces for Testing [C]//2003 10th Working Conference on Reverse Engineering. 2003: 260.
- [7] ANBUNATHAN R, BASU A. A recursive crawler algorithm to

detect crash in Android application[C]//2014 IEEE International Conference on Computational Intelligence and Computing Research (ICIC). IEEE, 2014: 1-4.

- [8] MAHMOOD R, MIRZAEI N, MALEK S. Evodroid: Segmented evolutionary testing of android apps[C]//Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering. ACM, 2014: 599-609.
- [9] MIRZAEI N, MALEK S, PŠŠSREANU C S, et al. Testing android apps through symbolic execution [J]. ACM SIGSOFT Software Engineering Notes, 2012, 37(6): 1-5.
- [10] AMALFITANO D, FASOLINO A R, TRAMONTANA P, et al. A toolset for GUI testing of Android applications[C]//2012 28th IEEE International Conference on Software Maintenance (ICSM). IEEE, 2012: 650-653.
- [11] AMALFITANO D, FASOLINO A R, TRAMONTANA P. A gui crawling-based technique for android mobile application testing[C]//2011 IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops (ICSTW). IEEE, 2011: 252-261.
- [12] LI X, JIANG Y, LIU Y, et al. User guided automation for testing mobile apps[C]//2014 21st Asia-Pacific Software Engineering Conference (APSEC). IEEE, 2014: 27-34.
- [13] ZHANG C, XUE Y Z, CHEN J C. Design and application of Android platform-based GUI capture-replay testing tool[J]. Computer Application and Software, 2012, 29(12): 6-9. (in Chinese)
张灿, 薛云志, 陈军成. 一种基于 Android 平台 GUI 录制回放工具的设计与实现[J]. 计算机应用与软件, 2012, 29(12): 6-9.
- [14] CHOI W, NECULA G, SEN K. Guided gui testing of android apps with minimal restart and approximate learning[C]//ACM SIGPLAN Notices. ACM, 2013: 623-640.
- [15] ZHANG D W, GUO X, HAN Z. Security and Trusted Intelligent Mobile Terminal [J]. ZTE Communications, 2015, 21(5): 39-44. (in Chinese)
张大伟, 郭旭, 韩臻. 安全可信智能移动终端研究[J]. 中兴通讯技术, 2015, 21(5): 39-44.