

文章编号:1001-9081(2015)09-2565-04

doi:10.11772/j.issn.1001-9081.2015.09.2565

面向大规模学术社交网络的社区发现模型

李春英^{1,2}, 汤庸², 汤志康^{3*}, 黄泳航², 袁成哲², 赵剑冬¹

(1. 广东技术师范学院 计算机网络中心, 广州 510665; 2. 华南师范大学 计算机学院, 广州 510631;

3. 广东技术师范学院 计算机科学学院, 广州 510665)

(*通信作者电子邮箱 fzutang@126.com)

摘要:针对基于标签传播的复杂网络重叠社区发现算法中预先输入参数在真实网络中的局限性以及标签冗余等问题,提出一种基于标签传播的面向大规模学术社交网络的社区发现模型。该模型通过寻找网络中互不相交的最大极大团(UMC)并对每个UMC中的节点赋予唯一标签来减少冗余标签,提高社区发现的效率以及稳定性。标签更新时以UMC作为核心单位采用亲密度的方式由中心向四周更新UMC邻接节点的标签及权重,以权重最大值的方式更新网络中非UMC邻接节点的权重。后期处理阶段采用自适应阈值方式去除节点标签中的噪声,有效克服了预先输入重叠社区个数在真实网络中的局限性。通过在学术社交网络平台——学者网数据集上的实验表明,该模型能够将具有一定共性的节点划分到同一个社区中,并为学术社交网络平台进一步的好友推荐、论文分享等精确的个性化服务提供了支持。

关键词:社交网络;社区发现;重叠社区;标签传播;最大极大团;自适应阈值

中图分类号: TP391.3 **文献标志码:**A

Community detection model in large scale academic social networks

LI Chunying^{1,2}, TANG Yong², TANG Zhikang^{3*}, HUANG Yonghang², YUAN Chengzhe², ZHAO Jiandong¹

(1. Computer Network Center, Guangdong Polytechnic Normal University, Guangzhou Guangdong 510665, China;

2. School of Computer, South China Normal University, Guangzhou Guangdong 510631, China;

3. School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou Guangdong 510665, China)

Abstract: Concerning the problem that community detection algorithm based on label propagation in complex networks has a pre-parameter limit in the real network and redundant labels, a community detection model in large scale academic social networks was proposed. The model detected Utmost Maximal Cliques (UMC) in the academic social network and arbitrary intersection between the UMC is the empty set, and then let nodes of each UMC share the unique label by reducing redundant labels and random factors, so the model increased the efficiency and stability of the algorithm. Meanwhile the model completed label propagation of the UMC adjacent nodes using closeness from core node groups (UMC) to spread around, Non-UMC adjacent nodes in the network were updated according to the maximum weight of its neighbor nodes. In the post-processing stage an adaptive threshold method removed useless labels, thereby effectively overcame the pre-parameter limitations in the real complex network. The experimental results on academic social networking platform—SCHOLAT data set prove that the model has an ability to assign nodes with certain generality to the same community, and it provides support of the academic social networks precise personalized service in the future, such as latent friend recommendation and paper sharing.

Key words: social network; community detection; overlapping community; label propagation; Utmost Maximal Clique (UMC); adaptive threshold

0 引言

随着 Web2.0 以及在线大规模社交网络的快速发展,从用户数据中发现社区结构成为网络分析中的一个热点问题。社交网络可以看成是由用户(节点)以及用户之间的关系(边)构成的无权重无向图,与现实生活中由各种关系构成的圈子类似,其可以划分成若干个社区。社区结构是社交网络中非常有意义的属性,同一个社区内部成员之间关系紧密,不

同社区的成员之间关系稀疏。利用该特点发现大规模学术社交网络中的社区结构以及每个用户所属的社区,发现的结果或者能够揭露一些隐藏的关系进而解释一些现象;或者代表了真实的社会团体,团体内部成员或属于相同的研究领域,或属于同一课程或科研团队,或具有相同的兴趣及其他相似的行为。因此,发现学术社交网络中的社区结构,有助于更好地理解和开发学术社交网络,提供精确的个性化服务。

近年来复杂网络社区发现的相关研究得到了学者的广泛

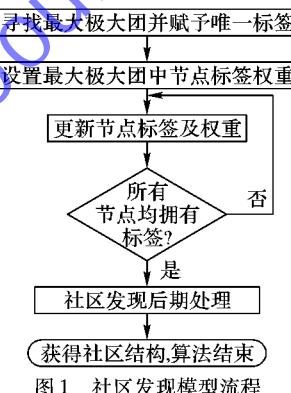
收稿日期:2015-04-20;修回日期:2015-06-18。 基金项目:国家863计划项目(2013AA01A212);国家自然科学基金资助项目(61272067, 61370229);广东省自然基金团队研究项目(S2012030006242);广东省自然科学基金—博士科研启动项目(2014A030310238);广东省教育厅特色创新项目(2014WTSCX078);广东技术师范学院校级项目(2014)。

作者简介:李春英(1978-),女,黑龙江齐齐哈尔人,副教授,博士研究生,CCF会员,主要研究方向:社交网络与大数据应用、服务计算、社区发现;汤庸(1964-),男,湖南张家界人,教授,博士生导师,博士,CCF会员,主要研究方向:信息搜索与数据挖掘、协同计算与移动互联网应用;汤志康(1978-),男,山东临沂人,讲师,硕士,主要研究方向:社交网络与大数据应用、社区发现。

关注,相关研究成果可以从多个角度进行阐述^[1~8]。本文则着重阐述基于标签传播的社区发现研究成果。2007 年,Raghavan 等首先提出简单快速的基于标签传播的社区发现算法(Label Propagation Algorithm, LPA)^[4],并且得到了广泛的应用。LPA 要求每个节点只能拥有一个标签并通过迭代更新节点的标签,当算法收敛时发现不相交的社区。基于该算法仅适合于非重叠社区发现以及存在健壮性和高随机性等问题,一些学者对其进行了改进。其中,重叠社区标签传播算法(Community Overlap Propagation Algorithm, COPRA)^[9]和 SLPA(Speaker-listener Label Propagation Algorithm)^[10]用于发现重叠社区,但 COPRA 和 SLPA 标签初始化时为每一个节点赋予唯一标签导致自身拥有过多冗余标签,以至于算法结果不稳定。基于 LPA 简单高效的特点,很多学者^[11~14]从不同方面对其进行了改进。平衡标签传播 BLP(Balanced Label Propagation)^[15]、并行 SLPA(Parallel Speaker-listener Label Propagation Algorithm)^[16]以及基于标签传播的可并行复杂网络重叠社区发现算法(Minimal Maximal Clique Label Propagation Algorithm, MMCLPA)^[17]等社区发现算法采用了超级计算的模式,将大型/超大型复杂网络的节点分成 N 个部分并分配给相应的 N 个处理器并行处理以此用于对大规模社交网络进行社区发现,解决了复杂网络大数据对单机系统带来的挑战问题,使得对现实中普遍存在的超大规模复杂网络进行社区发现成为可能。

1 社交网络社区发现模型

为了有效地开发大规模学术社交网络并为网络中的用户提供精确的个性化服务,本文以学术社交网络学者网(www.scholat.com)为数据平台,将改进的 MMCLPA 应用于学术社区发现模型。该模型主要由 3 部分组成:寻找最大极大团并初始化标签、标签更新及后期处理。在标签初始化阶段,模型首先在复杂网络中寻找最大极大团(如定义 1),并将每一个最大极大团中的节点赋予唯一一个标签,即每一个最大极大团中的节点获得相同标签。初始化阶段完成后,将获得标签的节点权重统一设置为 1,然后以最大极大团为核心单位,按照标签更新规则由内向外扩散进而更新其他节点的标签及权重。当所有节点都获得了标签及权重后按照社区发现后期处理规则对节点的标签及权重去除噪声,保留合理的标签及权重,进而获得复杂网络的社区结构。模型具体流程如图 1 所示。



1.1 标签初始化

定义 1 最大极大团(Utmost Maximal Clique, UMC)。学

术社交网络用户之间关系图由 $G = \{U, E\}$ 表示,其中 U 表示用户节点的集合, $E \subseteq U \times U$ 表示用户之间相邻关系(边)的集合。对 U 中的节点按照度数由大到小进行排序,以 U 中未赋予标签的度数最大的节点 u_i 及其相邻节点中未赋予标签的度数最大的节点 u_j 构成的边 $e(u_i, u_j)$ 为初始边寻找完全图 G_m ,则 G_m 称为团。若 $G_m \subseteq G$,且不存在任何完全图 $G_t \subseteq G$,使得 $G_m \subset G_t$,则称 G_m 为最大极大团(UMC)。

定义 2 重叠节点。 C_1 和 C_2 是两个不同的社区,若节点 u_i 同时属于 C_1 和 C_2 社区,则称节点 u_i 为重叠节点。即:

$$\text{overlap}(u_i) \Leftrightarrow \exists u_i (u_i \in C_1 \cap C_2 \wedge C_1 \neq C_2)$$

定义 3 社区合并。存在一些社区,是其他较大社区的子社区,则将这些社区合并为一个社区。即: $\exists C_i \exists C_j (C_i \neq C_j \wedge C_i \subseteq C_j) \Rightarrow C_i = (C_i \cup C_j) = C_j$ 。

定义 4 叶子节点所属社区。若叶子节点 u_i 的邻居节点 u_j 为重叠节点,则 u_i 归属于 u_j 所属的最大社区中。即:

$$\left\{ \begin{array}{l} \exists u_i \in U \wedge d(u_i) = 1 \wedge e(u_i, u_j) \in E \wedge \text{overlap}(u_j) \in \\ C_1 \cap C_2 \wedge |C_1| \neq |C_2| \Rightarrow u_i \in C_1, |C_1| > |C_2| \\ \exists u_i \in U \wedge d(u_i) = 1 \wedge e(u_i, u_j) \in E \wedge \text{overlap}(u_j) \in \\ C_1 \cap C_2 \wedge |C_1| \neq |C_2| \Rightarrow u_i \in C_2, |C_2| > |C_1| \end{array} \right.$$

通过分析学术社交网络拓扑结构得知,最大极大团是社区的核心单位,处在同一个最大极大团中的节点,必然处在同一个社区。因此对学者网用户组成的复杂网络进行标签初始化时,着重考虑寻找网络中的最大极大团。模型为寻找到了网络中的每一个最大极大团中的节点赋予一个相同的标签,后续则不再考虑已具有标签的节点,因此大量减少了冗余标签,提高了模型的稳定性。标签初始化规则如下。

- 1) 对网络中的所有节点按度数进行降序排序,令 $id = 1$;
- 2) 取没有赋予过标签的度数最大的节点 u_i 及其相邻节点中没有赋予过标签的度数最大的节点 u_j ;
- 3) 以 $e(u_i, u_j)$ 为初始边,依据定义 1 寻找网络中的最大极大团 UMC;
- 4) 将标签 id 赋予给 UMC 中的每一个节点;
- 5) 设置 $id = id + 1$;
- 6) 重复步骤 2) ~ 5),直到没有满足要求的节点,初始化过程终止。

根据标签初始化规则,以一个简单的网络拓扑结构为例。其中图 2 为原始网络拓扑结构,图 3 为经过标签初始化后得到的网络拓扑结构。在图 3 中共找出 3 个最大极大团 UMC,分别为节点群(4,11,8)、(3,5,6) 和(12,13,14),则按照标签初始化规则分别为每一个 UMC 中的节点赋予相同的标签:即节点 4,11,8 对应的标签均为 1;节点 3,5,6 对应的标签均为 2;节点 12,13,14 对应的标签均为 3。

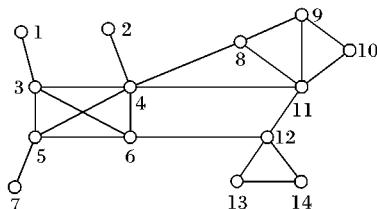


图 2 网络原始拓扑结构

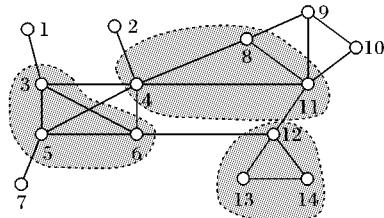


图3 网络初始化后拓扑结构

1.2 标签更新

通过对学术社交网络的拓扑结构分析得知,每一个社区至少有一个影响力比较大的核心朋友圈,社区的拓扑关系则由核心朋友圈向外围扩散,而此核心朋友圈即为本文定义的最大极大团(UMC)。标签在传播时,采用从核心朋友圈UMC向外围扩展,能提高算法的稳定性和效率。另外,节点是否属于某个社区以及是否为重叠节点则与节点在多个社区中的度数分布有关。标签更新规则具体操作步骤如下所示。

1) 将标签初始化后的UMC中的每一个用户节点的标签权重设置为1。

2) 统计其他用户节点与UMC相邻的亲密度 $F/d(u_i)$ 。其中: F 为用户节点 u_i 与UMC相邻的次数; $d(u_i)$ 为用户节点 u_i 的度数,该亲密度即为用户节点 u_i 在UMC标签下的权重。

3) 将非UMC及其相邻的节点随机排列,对每一个节点进行标签更新,节点拥有其所有邻居节点的标签号,每个标签号的权重为其在邻居节点中权重最大的一个。

对初始化后的结果(图3)按照标签更新规则更新网络中节点的标签及权重,结果如表1所示。表中UMC ID/Label表示UMC的标签,即该最大极大团中全部节点对应的标签,UMC Node表示UMC中的节点,UMC N-node表示与UMC相邻的节点,按照标签更新规则Weight为UMC N-node节点与UMC的亲密度,也为该节点对应的权重,Label则继承了UMC的标签号。

表1 标签更新过程

UMC ID/Label	UMC Node	UMC N-node	Weight	Label
1 4,11,8	2	1	1	1
	3	1/4	1	1
	5	1/4	1	1
	6	1/4	1	1
	9	2/3	1	1
	10	1/2	1	1
	12	1/4	1	1
2 3,5,6	1	1	2	2
	4	1/2	2	2
	7	1	2	2
	12	1/4	2	2
3 12,13,14	6	1/4	3	3
	11	1/5	3	3

1.3 社区发现后期处理

针对COPRA算法中预先输入重叠社区个数在未知复杂网络社区发现中存在的缺陷,本文深入分析网络节点标签权重与社区个数的关系得知,若节点 u_i 拥有 C 个标签,则该节点在每个标签下的有效权重均应大于 $1/(C+1)$,采用这种自适应阈值方式对节点的标签队列进行后期处理,有效克服了预先输入参数对未知网络的局限性。社区发现后期处理结果中

节点拥有的每一个标签代表一个社区,标签相同的节点组成一个社区。具体过程如下所示。

1) 当节点 u_i 拥有多个标签时,统计该节点的标签个数 C ,删除权重小于等于阈值 $1/(C+1)$ 的标签;若所有标签权重都小于 $1/(C+1)$,则保留权重最大的标签,删除其他所有标签,如果有多个权重最大,同时又都小于 $1/(C+1)$ 的标签时,随机选择一个标签保留,删除其他所有的标签。

2) 若存在一些社区,是其他较大社区的子社区则按照定义3将该子社区与较大的社区合并成为一个社区。

3) 叶子节点所属的社区按照定义4进行处理。

4) 只有一个标签的节点归为该标签所属的唯一社区中,拥有多个标签的节点则属于多个社区。

对表1的数据按照标签更新后期处理规则进行标签后期处理,过程如表2所示。表2中Node表示节点,Label(Weight)表示当前节点拥有的标签(权重),W-threshold($1/(C+1)$)表示该节点对应的自适应阈值,Result表示处理结果,Community表示节点所属的社区。从结果可以看出:社区 C_1 中节点为(2,4,8,9,10,11),社区 C_2 中节点为(1,3,4,5,6,7),社区 C_3 中节点为(12,13,14),该网络拓扑结构图中重叠节点为节点4。重叠节点4属于的2个社区长度均为6,因此与重叠节点4相邻的叶子节点2则按照后期处理过程后的结果属于 C_1 社区,与实际情况相符。如图4所示。

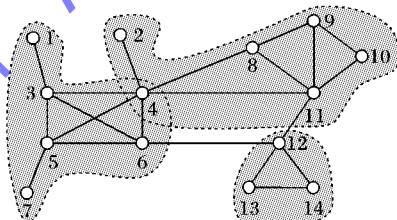


图4 网络社区划分结果

表2 标签更新后期处理过程

Node	Label(Weight)	W-threshold($\frac{1}{C+1}$)	Result	Community
1	2(1)	1/2	2(1)	C_2
2	1(1)	1/2	1(1)	C_1
3	2(1), 3(1/4)	1/3	2(1)	C_2
4	1(1), 2(1/2)	1/3	1(1), 2(1/2)	C_1, C_2
5	1(1/4), 2(1)	1/3	2(1)	C_2
6	1(1/4), 2(1)	1/3	2(1)	C_2
7	2(1)	1/2	2(1)	C_2
8	1(1)	1/2	1(1)	C_1
9	1(2/3)	1/2	1(2/3)	C_1
10	1(1/2)	1/2	1(1/2)	C_1
11	1(1), 3(1/5)	1/3	1(1)	C_1
12	1(1/4), 2(1/4), 3(1)	1/4	3(1)	C_3
13	3(1)	1/2	3(1)	C_3
14	3(1)	1/2	3(1)	C_3

1.4 模型的复杂度分析

假设社交网络中有 n 个节点且平均度数为 k ,算法第一层取没有赋予过标签的度数最大的节点 u_i 及其邻居节点中没有赋予过标签的度数最大的节点 u_j ,该过程产生的时间复杂度为 $O(kn)$ 。算法第二层以节点 u_i, u_j 构成的边 $e(u_i, u_j)$ 为初始边按照定义1寻找UMC,该过程产生的时间复杂度为 $O(n)$ 。因此寻找网络中的最大极大团在最坏情况下产生的时

间复杂度为 $O(kn^2)$ 。在标签更新阶段,假设标签传播的迭代次数 T ,每个节点接受一个标签分两个过程,一是接受每个邻居节点的新标签,再是对与邻居节点相同的标签进行比较,选择权重最大的标签;该过程在最坏情况下产生的时间复杂度为 $O(Tkn)$ 。后期处理是对所有节点进行遍历,按照自适应阈值清除节点标签中的噪声,并将属于同一标签的节点归为同一个社区中;该过程的时间复杂度为 $O(n)$ 。因此,该社区发现算法在最坏情况下的时间复杂度近似为 $O(kn^2)$ 。

2 算法实验

实验采用了学术社交网络平台(学者网)2015年3月30日的社交网络好友关系数据集,并对该数据集去除噪声,保留了学者网当日用户数量最多的一个连通分量,共计2106个公开信息的用户数据节点以及8867条好友关系构成的学术社交复杂网络。在该数据集中应用本文提出的社区发现模型进行社区发现,系统标签初始化阶段共获得116个最大极大团,亦即产生了116个标签。与COPRA和SLPA标签初始化时为每一个节点赋予唯一标签产生2106个标签相比,标签的数量减少了1990个,减少冗余标签比率高达94.49%,因此极大地提高了算法的稳定性以及社区发现的质量。经过标签传播及后期处理在该大规模社交网络中共发现91个社区,923个重叠节点,其中最大社区含637个用户节点,最小社区为3个用户节点。由实验可知,在学者网社区发现中应用该模型产生的重叠节点数量较多,这与目前学者网数据集的特点有关。学者网数据集中的大多数用户因相同工作单位而属于同

一个社区;也可能因不同的学术组织团体而聚集在一起;或因同一个科研团队/课程团队而属于另外一个社区;或因相同研究兴趣而与不同单位不同地域的研究者构成一个社区;因此成为社区发现中的重叠节点。以学者网创始人汤庸教授为例,在该连通分量中共属于24个不同的社区,在表1中列出了其因为不同角色而被归属6个不同社区的情况。

为了更好地阐述本社区发现模型在学者网中的有效性,对社区发现结果中的社区进行了详细的分析,发现每一个社区都具有一定的共性。现从中挑选10个最具有代表性并能够反映学者网目前提供服务的社区进行特点分析,如表3所示。从表3可以看出,社区内节点具有一定的共性。例如社区1则为以汤庸教授为核心的计算机相关研究方向的大社区;社区5和社区17则为计算机相关学术机构的社交网络平台;社区11和社区68为课程平台所产生的社区,且社区68中只包含了2014级广东技术师范学院“网络与新媒体”班级与授课教师李春英存在好友关系的小社区;社区15和社区66则为相应的科研团队社区;社区13为以汤庸教授为核心的师生关系构成的社区;社区19为具有相同研究兴趣的人员组成的社区,社区26为同事/师生关系组成的社区。因此,在学者网中的社区或者代表了一个工作单位、或者代表师生关系、或者具有相同的研究兴趣、或者属于同一课程/科研团队、或者是一些社会团体的社交网络平台等。通过分析可知,该社区发现模型在学者网数据中取得了良好的划分效果,为进一步的网络结构分析、功能演化预测以及类似于好友推荐、论文分享等精确的个性化服务提供了支持。

表3 学者网社区特点分析

序号	社区编号 (标签)	社区 节点数	代表节点	社区内节点的共性
1	1	637	汤庸、韩国强、姚剑等	以汤庸教授为核心的不同高校计算机相关专业朋友圈社区
2	5	106	汤庸、印鉴、傅秀芬等	广东省计算机学会数据库分会社区
3	11	102	曾碧卿、彭小飞、熊芳敏等	《C语言程序设计》课程社区
4	13	71	汤庸、潘炎、吴凌坤等	以汤庸教授为核心的师生关系社区
5	15	66	汤庸、蒋运城、赵淦森等	广东省服务计算工程技术研究中心社区
6	17	74	汤庸、赵淦森、蔡永铭等	中国计算机学会青年计算机科技论坛广州分论坛(CCF YOCSEF 广州)社区
7	19	63	汤庸、李建国、李春英等	学者网研究开发团队社区
8	26	39	胡天宝、贺凯山、陈云慧等	华南师范大学职业教育学院社区
9	66	15	李祥杰、熊建斌、李沁雪等	广东石油化工学院工业安全与无线传感器网络研究室社区
10	68	11	李春英、黄炜文、老贝华等	《MS Office 高级应用》课程“网络与新媒体”班级社区

3 结语

本文提出了一种面向大规模社交网络的社区发现模型,通过分析网络中的拓扑结构以及节点的度数分布与社区的关系,提出首先获取网络中的最大极大团,并将最大极大团作为社区的核心节点群,利用其他节点与最大极大团UMC的亲密度作为标签的权重并采用邻接节点标签权重最大值的方式进行标签传播,自适应阈值的后期处理方式有利于在真实网络中实现社区划分。相比于每个节点分配唯一一个标签的重叠社区发现算法,该社区发现模型在标签初始化阶段大量减少了冗余标签,减少了传播时不必要的判断开销,进而提高了算法的效率,并在一定程度上提高了算法的稳定性以及社区发现的质量。本文提出的模型对真实社交网络平台学者网的功能演化及预测提供了有力支持,下一步将对学者网的全部用户数据(包括多个连通分量)进行社区发现,为后续将推出的

精确个性化服务奠定基础。

参考文献:

- [1] PALLA G, DERÉNYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society [J]. Nature, 2005, 435(7043): 814–818.
- [2] AHN Y Y, BAGROW J P, LEHMANN S. Link communities reveal multiscale complexity in networks [J]. Nature, 2010, 466(7307): 761–764.
- [3] LANCICHINETTI A, FORTUNATO S, KERTÉSZ J. Detecting the overlapping and hierarchical community structure in complex networks [J]. New Journal of Physics, 2009, 11(3): 033015.
- [4] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks [EB/OL]. [2015-01-08]. <http://wenku.baidu.com/view/d6c2d36ba98271fe910ef9c9.html>.

(下转第2573页)

具有较高的精准度;算法运行时相对较低的时间复杂度,能够高效率地给用户合理地推荐一些潜在感兴趣的图书,是图书管理系统中的一种较好的个性化推荐方法。

参考文献:

- [1] LIU S. Research on the key issues for the recommender system [D]. Hefei: University of Science and Technology of China, 2014. (刘士琛.面向推荐系统的关键问题研究及应用[D].合肥:中国科学技术大学,2014.)
- [2] ZHANG F. Survey of online social network based on personalized recommendation [J]. Journal of Chinese Computer Systems, 2014, 35(7): 1470 – 1476. (张富国.基于社交网络的个性化推荐技术[J].小型微型计算机系统,2014,35(7):1470 – 1476.)
- [3] KONG Y. Recommendation algorithms in the big data era [D]. Xiamen: Xiamen University, 2014. (孔远帅.基于大数据的推荐算法研究[D].厦门:厦门大学,2014.)
- [4] WANG Z, HE M, DU Y. Text similarity computing based on topic model LDA [J]. Computer Science, 2013, 40(2): 229 – 232. (王振振,何明,杜永萍.基于LDA主题模型的文本相似度计算[J].计算机科学,2013,40(2):229 – 232.)
- [5] ZHU W. Research on user similarity function of recommendation system [D]. Chongqing: Chongqing University, 2014. (朱文奇.推荐系统用户相似度计算方法研究[D].重庆:重庆大学,2014.)
- [6] BLEI D M. Introduction to probabilistic topic models [EB/OL]. [2015-01-11]. http://www.cs.princeton.edu/~blei/papers/Blei2011.pdf?origin=publication_detail.
- [7] BOBADILLA J, ORTEGA F, HEMANDO A, et al. Improving collaborative filtering recommender system results and performance using genetic algorithms [J]. Knowledge-Based Systems, 2011, 24 (8): 1310 – 1316.
- [8] YANG Y, XIE K, ZHU Y, et al. Implementation of association rules recommendation model in recommendation system of e-commerce Web [J]. Computer Engineering, 2004, 30 (19): 57 – 59. (杨引霞,谢康林,朱扬勇,等.电子商务网站推荐系统中关联规则推荐模型的实现[J].计算机工程,2004,30(19):57 – 59.)
- [9] LI K, LIANG Z. Personalized book recommendation algorithm based on multi-feature [J]. Computer Engineering, 2012, 38(11): 34 – 37. (李克潮,梁正友.基于多特征的个性化图书推荐算法[J].计算机工程,2012,38(11):34 – 37.)
- [10] BLEI D M, ANDREW Y N, JORDAN M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(1): 993 – 1022.
- [11] ZHANG Z, MIAO D, GAO C. Short text classification using latent Dirichlet allocation [J]. Journal of Computer Applications, 2013, 33 (6): 1587 – 1590. (张志飞,苗夺谦,高灿.基于LDA主题模型的短文本分类方法[J].计算机应用,2013,33(6):1587 – 1590.)
- [12] YAGER R R, YAGER R L. Social networks: querying and sharing mined information [C]// Proceedings of the 2014 47th Hawaii International Conference on System Sciences. Washington, DC: IEEE Computer Society, 2013: 1435 – 1442.
- [13] MA Z. Bayesian estimation of the Dirichlet distribution with expectation propagation [C]// Proceedings of the 20th European Signal Processing Conference. Piscataway: IEEE, 2012: 689 – 693.
- [14] GRIFFITHS T, STEYYERS M. Probabilistic topic models [J]. Handbook of Latent Semantic Analysis, 2007, 427(7): 424 – 440.
- [15] TU D, SHU C, YU H. Using unified probabilistic matrix factorization for contextual advertisement recommendation [J]. Journal of Software, 2013, 24(3): 454 – 464. (涂丹丹,舒承椿,余海燕.基于联合概率矩阵分解的上下文广告推荐算法.软件学报,2013,24(3):454 – 464.)

(上接第2568页)

- [5] JIN H, WANG S, LI C. Community detection in complex networks by density-based clustering [J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(19): 4606 – 4618.
- [6] XIA Z, BU Z. Community detection based on a semantic network [J]. Knowledge-Based Systems, 2012, 26: 30 – 39.
- [7] BARBIERI N, BONCHI F, MANCO G. Cascade-based community detection [C]// Proceedings of the 6th ACM International Conference on Web Search and Data Mining. New York: ACM, 2013: 33 – 42.
- [8] DEV H, ALI M E, HASHEM T. User interaction based community detection in online social networks [M]// BHOWMICK S S, DYRESON C, JENSEN C S, et al. Database Systems for Advanced Applications, LNCS 8422. Berlin: Springer, 2014: 296 – 310.
- [9] GREGORY S. Finding overlapping communities in networks by label propagation [EB/OL]. [2015-01-08]. http://iopscience.iop.org/1367-2630/12/10/103018/pdf/1367-2630_12_10_103018.pdf.
- [10] XIE J, SZYMANSKI B K. Towards linear time overlapping community detection in social networks [M]// TAN P-N, CHAWLA S, HO C K, et al. Advances in Knowledge Discovery and Data Mining, LNCS 7302. Berlin: Springer, 2012: 25 – 36.
- [11] LIU X, MURATA T. Advanced modularity-specialized label propagation algorithm for detecting communities in networks [J]. Physica A: Statistical Mechanics and its Applications, 2010, 389(7): 1493 – 1500.
- [12] XIE J, SZYMANSKI B K. Community detection using a neighbor-
- hood strength driven label propagation algorithm [C]// Proceedings of the 2011 IEEE Network Science Workshop. Piscataway: IEEE, 2011: 188 – 195.
- [13] CORDASCO G, GARCANO L. Community detection via semi-synchronous label propagation algorithms [C]// Proceedings of the 2010 IEEE International Workshop on Business Applications of Social Network Analysis. Piscataway: IEEE, 2010, 1 – 8.
- [14] LOU H, LI S, ZHAO Y. Detecting community structure using label propagation with weighted coherent neighborhood propinquity [J]. Physica A, 2013, 392(14): 3095 – 3105.
- [15] UGANDER J, BACKSTROM L. Balanced label propagation for partitioning massive graph [C]// Proceedings of the 6th ACM International Conference on Web Search and Data Mining. New York: ACM, 2013: 507 – 516.
- [16] KUZMIN K, SHAH S Y, SZYMANSKI B K. Parallel overlapping community detection with SLPA [C]// Proceedings of the 2013 International Conference on Social Computing. Washington, DC: IEEE Computer Society, 2013: 204 – 212.
- [17] LI C, TANG Y, LIN H, et al. A parallelize overlapping community detection algorithm in complex networks based on label propagation [EB/OL]. [2014-12-28]. <http://info.scichina.com/sciF/CN/10.1360/N112014-00258>. (李春英,汤庸,林海,等.基于标签传播的可并行复杂网络重叠社区发现算法[EB/OL]. [2014-12-28]. <http://info.scichina.com/sciF/CN/10.1360/N112014-00258>.)