

# 基于社团结构的链接预测和属性推断 联合解决方法

王 锐, 吴玲玲, 石 川, 吴 斌

(北京邮电大学智能通信软件与多媒体北京市重点实验室, 北京 100876)

**摘 要:** 链接预测与属性推断是社交网络数据挖掘的两项重要任务. 之前的大部分研究工作将链接预测和属性推断视为不同的问题, 分别研究解决方法. 然而, 根据网络结构的同质性理论, 社交网络中的链接与属性之间具有内在关联. 本文提出了基于社团结构的链接预测和属性推断联合解决方法 (LAIC), 将社团结构作为链接预测与属性推断的关联因子, 利用用户属性和社团结构进行链接预测, 利用链接信息得到社团属性进而推断用户属性. LAIC 不仅同时解决了链接预测和属性推断问题, 而且通过迭代使链接预测和属性推断的准确率可以相互提升. 两个真实数据集上的实验证明 LAIC 方法是有效的.

**关键词:** 社交网络; 链接预测; 属性推断; 社团结构

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2016)09-2062-06

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2016.09.006

## Integrating Link Prediction and Attribute Inference Based on Community Structure

WANG Rui, WU Ling-ling, SHI Chuan, WU Bin

(Beijing Key Lab of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** Link prediction and attribute inference are two important tasks in social network mining. Most of the previous studies treated link prediction and attribute inference as different problems and sought for solutions separately. However, according to the theory of homophily, there are intrinsic relations between links and attributes in social network. We propose the link and attribute inference based on community (LAIC) solution which utilizes the community structure to connect link prediction and attribute inference. LAIC employs users' attribute and community structure for link prediction, and takes advantage of link information to get the attributes of communities for attribute inference of users. LAIC is not only able to predict attributes and links simultaneously, but also promotes the precision of link prediction and attribute inference mutually through iterations. Experiments on two real datasets verify the effectiveness of LAIC.

**Key words:** social network; link prediction; attribute inference; community structure

### 1 引言

在社交网络中, 链接预测可以帮助用户找到潜在好友, 改善用户体验. 属性推断可以完善用户信息, 为用户提供有针对性的服务. 因此, 链接预测和属性推断是社交网络数据挖掘的两项重要任务.

之前的研究将链接预测与属性推断视为两个不同的问题. 链接预测的研究主要基于节点相似性与拓扑

结构相似性. 例如, Chen 等人<sup>[1]</sup>在有向图中利用随机游走到达时间衡量用户相似性, 进行聚类, 在同一个聚簇中的节点被认为是朋友. 属性推断方法大致分为两类: 基于特征的方法<sup>[2,3]</sup>与基于网络结构的方法<sup>[4,5]</sup>. 基于特征的方法致力于寻找有效特征训练分类器. 基于网络结构的方法依据用户与好友的紧密性推断用户属性.

然而, 根据同质性理论<sup>[6]</sup>, 用户的相同属性越多, 用户间存在链接的概率越大. 反之, 用户间若存在链接,

收稿日期: 2015-02-05; 修回日期: 2015-05-27; 责任编辑: 覃怀银

基金项目: 国家重点基础研究发展计划 (No. 2013CB329602); 国家自然科学基金 (No. 61375058, No. 71231002); 北京市高等教育青年英才项目 (No. YETP0444)

则他们具有相同属性的概率越大. 因此, 链接预测和属性推断之间存在内在关联. 至今, 只有极少部分工作将链接预测和属性推断问题联合解决. Yin 等人<sup>[7]</sup>通过社会属性网络中的随机游走将链接预测与属性推断结合起来. 尹绪森<sup>[8]</sup>等人利用两层人工神经网络, 建立可同时解决链接预测与属性推断问题的综合框架.

为了同时解决链接预测和属性推断问题, 本文提出了基于社团结构的链接预测和属性推断联合解决方法 LAIC(Link and Attribute Inference Based on Community). LAIC 基于可重社团探测, 首先利用节点的社团信息和属性信息进行链接预测. 之后使用基于社团的随机游走获得社团属性, 通过用户所属社团的属性推断用户属性. 最后通过迭代使链接预测与属性推断相互提高.

## 2 基于社团结构的链接预测与属性推断联合解决方法

### 2.1 社会属性网络

本文使用社会属性网络解决链接预测和属性推断问题, 这里给出定义.

#### 定义 1 社会属性网络 SAN

给定网络  $G = (V, E)$ ,  $V$  是节点集,  $E$  是边集. 构造网络  $G' = (V', E')$ , 对  $G$  中的每个节点, 在  $G'$  中也相应构造一个节点, 称为用户节点  $V_p$ . 对  $G$  中的每条边, 在  $G'$  中也相应构造一条边. 对每个用户属性, 在  $G'$  中构造一个节点, 称为属性节点  $V_a$ ,  $V' = V_p \cup V_a$ . 若某个用户具有某个属性, 则在该用户节点与该属性节点间构造一条边.

图 1 为 SAN 的示意图. 图中矩形表示属性节点, 人物表示用户节点, 虚线表示用户属性, 实线表示用户间的好友关系.

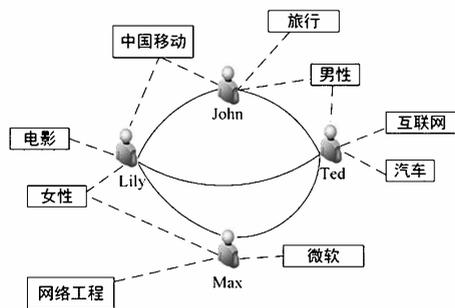


图1 社会属性网络示意图

### 2.2 基本思路

社团包含网络的结构信息, 两个用户的社团重叠次数越多, 他们之间越可能存在链接<sup>[9]</sup>. 两个用户的共同属性越多, 则两个用户越相似, 他们之间也越可能存在链接. 因此可以通过社团信息和用户属性求得缺失

的链接.

此外, 社团与属性不是独立的. 社团中用户的属性决定了社团的属性. 反之, 若已知社团属性, 则可推断社团中用户的属性. 每个用户可同时属于多个社团, 而每个社团具有各自的社团属性. 因此可根据用户所属社团的属性推断用户的属性.

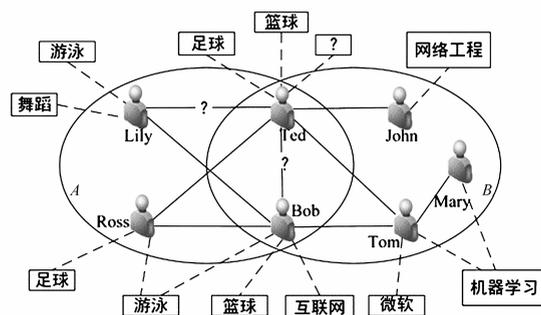


图2 基于社团结构的链接预测与属性推断举例

图 2 说明了 LAIC 的思路. 图中椭圆形表示社团, 其它图标含义与图 1 相同. 图 2 中, Ted 和 Bob 同属于这两个社团且拥有共同属性“篮球”, 而 Ted 和 Lily 只同属于一个社团且没有共同属性. 因此, 相对于 Lily, Ted 和 Bob 之间链接存在的概率更大. 此外, Ted 同时属于社团 A 和 B, 因此, 对社团 A 和 B 越重要的属性越可能是 Ted 的属性.

### 2.3 具体方法

#### 2.3.1 社团探测

本文使用 alpha-beta 社团探测方法<sup>[10]</sup>进行社团探测, 该方法适用于可重社团划分, 高效、简单且易于并行化. 其他能够准确发掘可重社团结构的方法亦可用于本文提出的框架.

假设通过可重社团划分, 将  $G$  划分为  $K$  个社团, 并且求得节点与社团的关系  $F$ . 用户-社团关系  $F$  为  $N * K$  的矩阵,  $F_{uc} = 1$  表示节点  $u$  属于社团  $c$ , 否则  $F_{uc} = 0$ .

#### 2.3.2 链接预测

为构建利用社团结构和用户属性预测链接的概率模型, 本文提出 2 点假设:

- (1) 社团结构和用户属性联合影响节点间链接存在的概率.
- (2) 不同社团对链接存在概率的影响是相互独立的.

仅考虑社团信息时, Leskovec 等人<sup>[9]</sup>提出同属于社团  $c$  的节点  $u$  和  $v$  间存在链接的概率为:

$$P_{uv}(c) = 1 - \exp(-F_{uc} \cdot F_{vc}) \quad (1)$$

当  $u$  和  $v$  中的任一点不属于社团  $c$  时,  $F_{uc} = 0$  或  $F_{vc} = 0$ , 则  $P_{uv}(c) = 0$ . 假设每个社团以概率  $1 - \exp(-F_{uc} \cdot F_{vc})$  连接  $u$  和  $v$ . 由于不同社团对链接存在概率的影响是相互

独立的,  $u, v$  间不存在链接的概率可写成  $u, v$  在所有社团下不存在链接的概率之积:

$$1 - P_{uv} = \prod_c (1 - P_{uv}(c)) = \exp\left(-\sum_c F_{uc} \cdot F_{vc}\right) \quad (2)$$

$u, v$  间存在链接的概率  $P_{uv}$  为:

$$P_{uv} = 1 - \exp\left(-\sum_c F_{uc} \cdot F_{vc}\right) \quad (3)$$

由假设知, 用户属性对链接存在概率有影响, 且用户具有的相同属性越多, 用户间存在链接的概率越大. 考虑到用户属性, 我们将式(3)改进为如下形式:

$$P_{uv} = 1 - \beta \cdot \exp\left(-\sum_c F_{uc} \cdot F_{vc}\right) - (1 - \beta) \cdot \exp\left(-\sum_a A_{ua}^V \cdot A_{va}^V\right) \quad (4)$$

$A^V$  是表示用户-属性关系的 0-1 矩阵.  $A_{ua}^V = 1$  表示用户  $u$  具有属性  $a$ ; 否则用户  $u$  不具有该属性或该属性缺失.  $A_u^V, A_v^V$  分别为用户  $u, v$  的属性, 当  $u$  和  $v$  没有共同属性时,  $A_u^V \cdot A_v^V = 0$ . 当  $u$  和  $v$  既不同属于任何社团又没有共同属性时,  $P_{uv} = 0$ . 参数  $\beta$  控制社团和属性信息对链接存在概率的影响.

### 2.3.3 属性推断

为构建利用社团结构和社团属性推断用户属性的模型, 本文提出 2 点假设:

(1) 社团中所有节点的属性均为该社团的属性, 但每个社团属性与社团的关联强度存在区别.

(2) 节点的属性由其所属社团的属性决定.

本文用 SAN 描述用户间以及用户和用户属性的关系. 不同于 Yin 等人<sup>[7]</sup> 利用从单个节点重启的随机游走进行链接预测和属性推断, 我们利用从社团重启的随机游走得到社团和各属性的关联强度. 对每个社团  $C_k$ , 设置随机游走从  $C_k$  中的每个节点以相同概率重启, 即若  $u \in C_k$ , 则随机游走子以概率  $\alpha \cdot \frac{1}{|C_k|}$  从节点  $u$  重启, 其中  $|C_k|$  表示社团  $C_k$  的节点数.

从社团  $C_k$  重启时随机游走的迭代公式为:

$$r_{C,p}^{(t)} = (1 - \alpha) \sum_{p' \in N_u(p)} w(p', p) r_{C,p'}^{(t-1)} + (1 - \alpha) \sum_{a' \in N_u(p)} w(a', p) r_{C,a'}^{(t-1)} + \alpha \cdot \frac{1}{|C_k|} r_p^{C_k} \quad (5)$$

$$r_{C,a}^{(t)} = (1 - \alpha) \sum_{p' \in N_u(a)} w(p', a) r_{C,p'}^{(t-1)} \quad (6)$$

$r_{C,p}, r_{C,a}$  分别为从社团  $C_k$  重启时, 用户和属性节点的到达概率,  $t$  为迭代次数,  $p, p'$  为用户节点,  $a'$  为属性节点.  $N_p(p)$  是与用户节点  $p$  相连的用户节点的集合,  $N_a(p)$  是与用户节点  $p$  相连的属性节点的集合,  $N_p(a)$  是与属性节点  $a$  相连的用户节点的集合.  $w(p', p)$  是从用户节

点  $p'$  到用户节点  $p$  的转移概率,  $w(a', p)$  是从属性节点  $a'$  到用户节点  $p$  的转移概率,  $w(p', a)$  是从用户节点  $p'$  到属性节点  $a$  的转移概率.  $\alpha$  是随机游走重启概率, 当  $p \in C_k$  时,  $r_p^{C_k} = 1$ , 否则等于 0.

因为属性节点只能与用户节点相连, 所以从属性节点到用户节点的转移概率设置如下:

$$w(a, p) = \frac{1}{N_p(a)} \quad (7)$$

用户节点可与属性或用户节点相连, 我们用参数  $\lambda$  权衡属性与用户节点权重. 一般 SAN 中的随机游走区分用户-用户与用户-属性两种边, 同类型边的转移概率相同. 考虑到节点间的影响力是有区别的, 我们提出利用链接存在概率表示边的影响力. 从用户到用户节点的转移概率设置如下:

$$w(p', p) = \begin{cases} \frac{(1 - \lambda) P_{p'p}}{\sum_{i \in N_p(p')} P_{p'i}}, & \text{若 } |N_p(p')| > 0 \text{ 且 } |N_a(p')| > 0 \\ \frac{P_{p'p}}{\sum_{i \in N_p(p')} P_{p'i}}, & \text{若 } |N_p(p')| > 0 \text{ 且 } |N_a(p')| = 0 \\ 0, & \text{其他} \end{cases} \quad (8)$$

当  $(p', p) \in E$  时,  $P_{p'p} = 1$ . 否则  $P_{p'p}$  等于边  $(p', p)$  的存在概率.

根据 Yin 等人<sup>[7]</sup> 提出的混合权重的方法, 我们用  $w_p(a)$  衡量对用户  $p$  而言, 属性  $a$  的重要性:

$$w_p(a) = g(a) \times l_p(a) \quad (9)$$

其中,  $g(a)$  和  $l_p(a)$  分别是属性  $a$  的全局重要性和本地重要性:

$$g(a) = \frac{\sum_{(u,v) \in E} e_{uv}^a}{\binom{n_a}{2}} \quad (10)$$

$$l_p(a) = \sum_{p' \in N_u(p)} I(p', a) \quad (11)$$

将  $w_p(a)$  归一化作为从用户节点到属性节点的转移概率:

$$w(p, a) = \begin{cases} \frac{\lambda w_p(a)}{\sum_{a' \in N_p(p)} w_p(a')}, & \text{若 } |N_a(p)| > 0 \text{ 且 } |N_p(p)| > 0 \\ \frac{w_p(a)}{\sum_{a' \in N_p(p)} w_p(a')}, & \text{若 } |N_a(p)| > 0 \text{ 且 } |N_p(p)| = 0 \\ 0, & \text{其他} \end{cases} \quad (12)$$

当从社团  $C_k$  重启的随机游走收敛时, 各属性节点

的到达概率  $r_{c,a}$  表示社团  $C_k$  与各属性的关联强度:

$$\mathbf{r}_{c,a} = [r_{c,a_1}, r_{c,a_2}, \dots, r_{c,a_v}] \quad (13)$$

当从  $K$  个社团重启的随机游走都收敛时,得到社团-属性矩阵  $\mathbf{A}^C$ :

$$\mathbf{A}^C = [\mathbf{r}_{c,a}, \mathbf{r}_{c,a}, \dots, \mathbf{r}_{c,a}]^T \quad (14)$$

用社团-属性矩阵  $\mathbf{A}^C$  和用户-社团关系  $\mathbf{F}$  的乘积填充用户-属性矩阵  $\mathbf{A}^V$  中的缺失项:

$$P(A_{ua}^V = 1) = \sum_{c_i \in \Omega} F_{uc_i} \cdot A_{c_i a}^C \quad (15)$$

$P(A_{ua}^V = 1)$  为用户  $u$  具有属性  $a$  的概率,  $\Omega$  为网络  $G$  中所有的社团.  $A_{c_i a}^C$  为社团  $C_i$  与属性  $a$  的关联强度. 更新  $\mathbf{A}^V$  矩阵:

$$(A_{ua}^V)^t = \begin{cases} P(A_{ua}^V = 1), & \text{若 } (A_{ua}^V)^{t-1} = 0 \\ (A_{ua}^V)^{t-1}, & \text{其他} \end{cases} \quad (16)$$

$(A_{ua}^V)^t$  是第  $t$  次迭代后用户  $u$  具有属性  $a$  的概率. 得到更新后的  $\mathbf{A}^V$  矩阵, 返回链接预测步骤重新计算节点间链接存在的概率. 链接预测完成后接着执行属性推断, 如此反复迭代, 直到达到指定的迭代次数. 迭代完成后, 对每个用户  $u$  返回  $P_{uw}$  最大的  $k$  个用户作为推荐的链接, 返回  $A_{ua}^V$  最大的  $k$  个属性作为推断的缺失属性.

### 2.3.4 时间复杂度分析

LAIC 算法的时间复杂度由链接预测和属性推断两部分构成. 链接预测的时间复杂度为  $O((m+c)n^2)$ , 属性推断的时间复杂度为  $O(t_1 cm(m+n))$ , 其中  $m$  为属性数,  $c$  为社团数,  $n$  为用户节点数,  $t_1$  为随机游走收敛所需的迭代次数. 因此 LAIC 算法的时间复杂度为  $O(t_2((m+c)n^2 + t_1 cm(m+n)))$ , 其中  $t_2$  为链接预测和属性推断过程相互迭代的最大次数.

## 3 实验

本节首先介绍实验数据集, 随后介绍对比方法, 最后展示实验结果及结果分析.

### 3.1 数据集

新浪微博数据集: 我们从新浪微博上爬取该数据集 (<http://weibo.com>), 包括 34,199 个用户和 691,522 个链接. 我们抽取用户的标签、学校及公司作为用户属性. 随后删去度小于 20 的点和属性少于 3 的点以及与这些点相连的边. 最后得到 2503 个用户, 56,768 个链接和 228 个属性.

表 1 数据集信息统计

数据集	节点数	链接数	属性数
新浪微博数据集	2,503	56,768	228
电信数据集	34,905	70,495	127

电信数据集: 该数据集来自第一届中国大数据竞赛 (<http://www.bigcloudsys.com/ccf2013/detail2>. ht-

ml), 包括用户间的电话、短信记录和用户信息. 我们抽取身份为学生的 77,577 个用户作为备选节点. 用户的学校作为用户属性. 若用户间相互发送超过 5 条短信, 则用户间存在一条链接. 删除处理后孤立的节点得到 34,905 个用户, 70,495 个链接和 127 个不同的学校 id.

两个数据集的统计信息如表 1 所示.

### 3.2 对比方法

共同邻居 (CN): 链接预测算法. 将与目标用户共同邻居最多的用户推荐给他.

共同属性 (CA): 属性推断算法. 将目标用户的邻居拥有最多的属性推荐给他.

带重启的随机游走 (RW)<sup>[11]</sup>: 链接预测算法. 设置重启概率从目标用户重启, 游走过程收敛后, 将到达概率高的节点推荐给他.

社会属性网络上的随机游走 (SAN\_RW)<sup>[7]</sup>: 链接预测和属性推断联合解决方法. 根据用户的好友关系和用户属性构造社会属性网络 SAN. 在 SAN 中进行带重启的随机游走, 游走过程收敛后, 分别将到达概率高的用户和属性节点作为链接预测与属性推断的返回结果.

### 3.3 有效性实验

每次实验随机选取 10% 的链接和用户属性作为测试集, 即  $E^d$  和  $(A^V)^d$ . 用剩下的网络  $G' = (V, E', (A^V)')$  训练模型, 使用训练出的模型在测试集上进行实验. 链接预测和属性推断的结果不断迭代更新直到收敛或达到实验设置的最大迭代次数 40 次. RW, SAN\_RW 和 LAIC 的重启概率  $\alpha$  设为 0.8. LAIC 和 SAN\_RW 的参数  $\lambda$  设为 0.3. 新浪微博数据集和电信数据集中 LAIC 的参数  $\beta$  分别设为 0.9 和 0.3.

实验重复进行 50 次, 链接预测平均结果如图 3 所示, 横坐标 TOP  $K$  表示为用户推荐链接的数量. 图 3 中, LAIC 在两个数据集上的准确性始终优于其他方法. 原因是我们同时使用了网络社团结构和用户属性信息且链接预测与属性推断之间相互迭代提高, 这是其他算法没有的.

属性推断平均结果如表 2 所示. 由表 2 知, 新浪微博数据集中, 所有方法的准确率都偏低, 这是因为我们

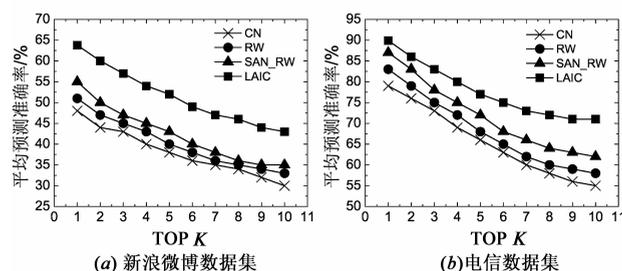


图 3 不同方法链接预测准确率对比

没有对属性做语义归并,如:我们的算法将“北邮”和“北京邮电大学”看作不同的属性.并且,新浪微博中除了社会关系还有一些兴趣关系,在这种社交网络中进行属性推断比较难.电信数据集中,用户的学校 id 不存在歧义,且电信网络是基于社会关系的,所以属性推断

准确率较高.表 2 中,LAIC 的准确率总是最高的,原因是同一社团的用户通常具有相同的属性,我们利用网络的结构信息弥补了属性信息的缺失.并且,LAIC 利用链接预测步骤得到的链接存在概率弥补遗失的链接信息,链接信息越充分时,推断用户属性越准确.

表 2 不同方法属性推断效果对比

新浪微博数据集				召回率			电信数据集		
准确率						准确率			
方法	Top1	Top2	Top3	方法	Top1	Top2	Top3	方法	Top1
CA	17.9%	12.7%	14.5%	CA	17.2%	23.3%	40.6%	CA	53.2%
SAN_RW	20.5%	17.7%	14.8%	SAN_RW	20.2%	33.9%	40.1%	SAN_RW	83.4%
LAIC	25.3%	22%	19.5%	LAIC	22.8%	37.5%	47.2%	LAIC	88.5%

### 3.4 不同缺失比例对比实验

我们在两个数据集上扩大删除比例来验证缺失信息更多的情况下 LAIC 的适用性.依次选取 10%,20%,30%,40%,50% 的链接和用户属性作为测试集,每次为用户推荐排名第一的链接或属性.相关参数设置如前所述,实验重复进行 50 次,平均结果如图 4 所示.图 4 中,随着测试集比例的增大,所有方法的准确率都在下降,但 LAIC 在两个任务中的准确率始终最高.原因是 LAIC 是基于社团结构的,网络中个别用户缺失的信息可由该用户所属社团中其他用户的信息部分弥补.此外,LAIC 中链接预测与属性推断相互迭代使缺失信息不断得到补充.

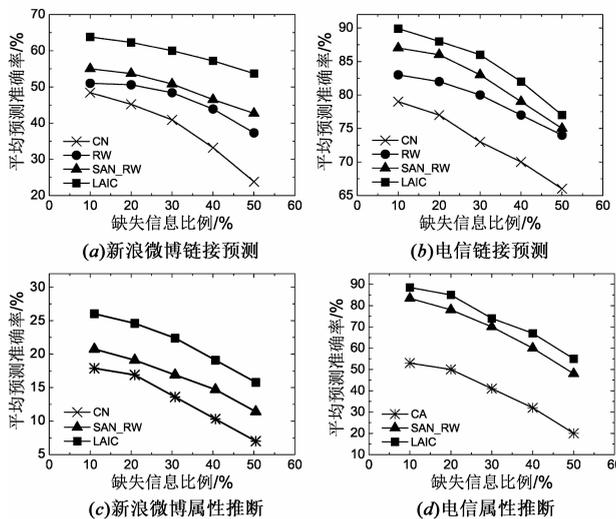


图 4 算法准确率随缺失信息变化对比图

### 3.5 算法迭代过程实验

算法迭代过程实验用于验证链接预测与属性推断相互迭代的效果.相关参数设置如前所述,迭代次数从 5 增加至 50,每次增加 5,对不同的迭代次数分别进行 50 次实验,实验平均结果如图 5.由图 5 可知迭代次数

大约在 25~30 次时,准确率的变化趋于平缓.在两个数据集中,随着迭代次数的增加,属性推断与链接预测的准确率都在增加,这证实了 LAIC 能使属性推断与链接预测相互提高.

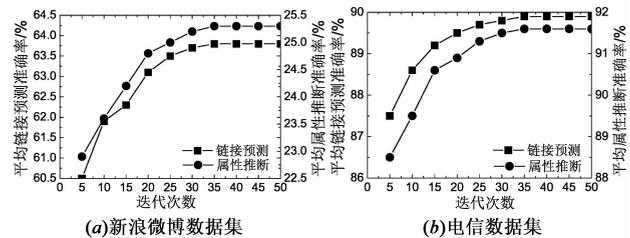


图 5 算法准确率随迭代次数变化对比图

## 4 结束语

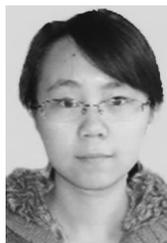
本文提出利用社团结构综合解决链接预测与属性推断问题的框架 LAIC.实验证明,LAIC 不仅能同时解决链接预测和属性推断问题,使其相互提升,且准确率高于现有的方法.该框架在一些方面还存在提升的空间.首先,可以通过算法优化解决算法时间复杂度较大的问题.其次,可以通过属性语义归并来提高属性推断的准确率.

### 参考文献

- [1] Chen M, Liu J, Tang X. Clustering via random walk hitting time on directed graphs [A]. Proceedings of the 23rd National Conference on Artificial Intelligence [C]. Chicago, Illinois, USA: AAAI, 2008. 616 - 621.
- [2] Rao D, Yarowsky D, Shreevats A, et al. Classifying latent user attributes in twitter [A]. Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents [C]. Toronto, Canada: ACM, 2010. 37 - 44.
- [3] Leuski A. Email is a stage: discovering people roles from email archives [A]. Proceedings of the 27th Annual Inter-

- national ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Sheffield, UK: ACM, 2004. 502 – 503.
- [4] Mislove A, Viswanath B, Gummadi K P, et al. You are who you know: inferring user profiles in online social networks [A]. Proceedings of the Third ACM International Conference on Web Search and Data Mining [C]. New York, USA: ACM, 2010. 251 – 260.
- [5] Coscia M, Rossetti G, Giannotti F, et al. DEMON: a local-first discovery method for overlapping communities [A]. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. Beijing, China: ACM, 2012. 615 – 623.
- [6] La Fond T, Neville J. Randomization tests for distinguishing social influence and homophily effects [A]. Proceedings of the International World Wide Web Conference [C]. Raleigh, NC, USA: Springer, 2010. 601 – 610.
- [7] Yin Z, Gupta M, Weninger T, et al. A unified framework for link recommendation using random walks [A]. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining [C]. Odense, Denmark: IEEE/ACM, 2010. 152 – 159.
- [8] Yin X, Wu B, Lin X. A unified framework for predicting attributes and links in social networks [A]. Proceedings of the 2013 IEEE International Conference on Big Data [C]. Santa Clara, CA, USA: IEEE, 2013. 153 – 160.
- [9] Yang J, McAuley J, Leskovec J. Community detection in networks with node attributes [A]. Proceedings of the 13th International Conference on Data Mining [C]. Dallas, TX, USA: IEEE, 2013. 1151 – 1156.
- [10] He J, Hopcroft J, Liang H, et al. Detecting the structure of social networks using  $(\alpha, \beta)$ -communities [A]. Proceedings of the 8th International Workshop on Algorithms and Models for the Web-Graph [C]. Atlanta, GA, USA: Springer, 2011. 26 – 37.
- [11] Tong H, Faloutsos C, Pan J Y. Fast random walk with restart and its applications [A]. Proceedings of the 6th IEEE International Conference on Data Mining [C]. Hong Kong, China: IEEE, 2006. 613 – 622.

#### 作者简介



王 锐 女, 1992 年 6 月生于河南洛阳. 北京邮电大学计算机学院硕士研究生. 研究方向为数据挖掘与机器学习.

E-mail: aboutstefanie@163.com



吴玲玲 女, 1989 年 4 月生于福建漳州. 2015 年毕业于北京邮电大学计算机学院. 研究方向为数据挖掘与机器学习.

E-mail: wulingling@bupt.edu.cn