

基于连续时间量子游走的链路 预测方法研究*

钱菁, 杨林涛, 余泽太, 刘守印

(华中师范大学物理科学与技术学院, 武汉 430079)

摘要: 链路预测是复杂网络研究的基础问题之一。目前研究者们已经提出了许多链路预测的方法, 其中大量的链路预测方法是基于经典随机游走。量子游走是经典随机游走的量子模拟。大量研究表明, 在诸如图匹配、搜索等很多领域, 基于量子游走的量子算法的性能远优于其对应的经典随机游走算法。但目前几乎没有关于量子游走的链路预测算法研究报道。提出了一种基于连续时间量子游走的链路预测方法。实验结果表明, 连续时间量子游走链路预测结果的 AUC 值和经典随机游走的结果非常接近。而在 Precision 和 Recall 指标上, 远优于经典随机游走的链路预测结果。

关键词: 复杂网络; 链路预测; 量子游走; 相似度

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2016)01-0104-04

doi:10.3969/j.issn.1001-3695.2016.01.024

Link prediction using continuous-time quantum walk

Qian Jing, Yang Lintao, Yu Zetai, Liu Shouyin

(College of Physical Science & Technology, Central China Normal University, Wuhan 430079, China)

Abstract: Link prediction is one of the key issues of complex networks. Many link prediction methods have been proposed so far. The classical random walk as an effective tool has been widely used to study the link prediction problems. Quantum walk is the quantum analogue of classical random walk. Numerous research results show that quantum algorithms using quantum walk outperform their classical counterparts in many applications, such as graph matching, searching etc. . But few of the research is about the link prediction based on quantum walk. This paper proposed a new link prediction method based on continuous-time quantum walk. Experiments show that the AUC result of continuous-time quantum walk is very close to that of the classical random walk, while the Precision and Recall results of continuous-time quantum walk are much higher than that of the classical random walk.

Key words: complex network; link prediction; quantum walk; similarity

0 引言

链路预测是复杂网络研究的基础问题之一。链路预测问题是指通过对已知网络结构的分析, 包括一些可能节点的其他信息, 来评估两个节点间未来产生链接的可能性, 从而实现对未来链路或对未知链接状态的预测^[1-3]。

链路预测对于研究网络的演化机制等问题具有重要的理论研究意义。又因其广泛的实际应用价值, 受到各领域研究者的广泛关注, 包括蛋白质相互作用网络、推荐系统、科学家合作网络、在线社交网络、移动社会网络、恐怖袭击网络等。

1 相关工作

目前为止, 研究者们已经提出了许多链路预测的方法^[3], 其中基于相似度的链路预测以其简单易行的优势而被广泛使用。应用节点间的相似性进行链路预测的一个重要前提, 就是

假设两个节点之间相似性越大, 它们之间存在链接的可能性就越大。

研究人员广泛关注的是基于网络的结构信息定义的结构相似性, 其中有相当数量的相似性指标是基于经典随机游走过程的。平均通勤时间 (ACT)^[4] 中, 两个节点间的相似性定义为随机游走者从其中任一节点出发到另一节点所需平均步数的倒数。重启的随机游走 (RWR)^[5] 是著名的网页排序算法 PageRank^[6] 的拓展。假定随机游走者从节点 v 开始, 他将以概率 C 移动到 v 的任意一个邻居节点, 以概率 $1-C$ 返回节点 v 。SimRank^[7] 指标考察两个随机游走者分别从节点 v 和 u 出发, 直至相遇总共经过多少步数。步数越少, 则节点越相似。以上指标都利用了整个网络的拓扑信息, 能够得到较准确的预测结果。但是这些指标的计算都必须通过多次迭代。由于计算过程中, 对规模巨大的矩阵求逆矩阵或者特征值几乎是不可能的, 所以它的计算十分耗时, 不适合于大规模网络和实时的应用。文献^[8] 提出两种新颖的相似性定义: 局部随机游走

收稿日期: 2014-09-02; 修回日期: 2015-10-31 基金项目: 湖北省自然科学基金资助项目(2013CFB210)

作者简介: 钱菁(1975-), 女, 湖北武汉人, 讲师, 硕士, 主要研究方向为复杂网络(appleqian@phy.ccnu.edu.cn); 杨林涛(1982-), 男, 湖北武汉人, 讲师, 博士, 主要研究方向为链路预测; 余泽太(1978-), 男, 湖北武汉人, 工程师, 硕士, 主要研究方向为复杂网络; 刘守印(1965-), 男, 湖北武汉人, 教授, 博士, 主要研究方向为无线通信。

(LRW)和叠加的随机游走(SRW)。这两种相似度定义只局限在局部游走,步数的选择一般都很小。实验表明,LRW指标和SRW指标,对比与其他的相似性指标,如CN指标、RA指标、HSM指标等,都显示出较优的性能。更重要的是,局部随机游走只需要较少次数的迭代,时间复杂度大大减小,尤其适合对稀疏网络的计算^[8,9]。

本文提出一种新的基于连续时间量子游走的链路预测算法。量子游走是经典随机游走的量子模拟,于1993年由Aharonov等人^[10]提出。量子游走与经典随机游走差别非常大。在经典随机游走中,游走者的运动是扩散的。而在量子游走中,游走者的运动状态是粒子态的量子叠加和这些叠加态的干涉。这些差别促使了基于量子游走的量子算法的研究^[11,12]。大量的研究表明,这些量子算法在很多应用,如搜索、元素甄别、图匹配等领域,远优于基于经典随机游走的算法。如文献^[13]中,定义了一种基于量子游走的量子导航方法。与基于经典随机游走的导航方法相比,该方法的收敛步数较小,并且可以解决基于经典游走方法中的退化问题。文献^[14,15]提出了基于量子游走的图匹配算法。研究表明,对随机产生的图,该算法可实现100%的正确匹配;对于复杂网络,该算法可确定两个图是否同构。相对于基于经典随机游走的图匹配算法,该算法的时间复杂度在量子计算机上,大大降低。虽然基于量子游走的量子算法在很多研究工作中,显示出其优越性,但目前关于基于量子游走的链路预测算法几乎没有研究报道。本文提出一种新的基于量子游走的链路预测算法。

类似于经典随机游走,量子游走有连续时间量子游走CTQW^[16,17]和离散时间量子游走DTQW^[10]两种基本类型。本文将连续量子游走应用于链路预测模型中,提出了基于连续时间量子游走的链路预测算法。实验结果表明,在连续时间量子游走的链路预测结果中,AUC值和经典随机游走的结果非常接近。而在Precision和Recall指标上,其预测结果远优于经典随机游走的链路预测结果。

2 连续时间量子游走

2.1 网络模型

$G=(V,E)$ 是一个无权无向无回路的复杂网络,其中 V 表示 G 的节点集合, E 表示 G 的边集合。 G 中的节点个数为 $|V|$,边的个数为 $|E|$ 。对任意的节点 v , $\Gamma_v=\{u \in V:(u,v) \in E\}$ 表示与 v 邻接的节点(邻居节点)。 A 是网络 G 的邻接矩阵,即

$$A_{uv} = \begin{cases} 1 & \text{if } (u,v) \in E \\ 0 & \text{else} \end{cases} \quad (1)$$

由于 G 是无权无向无回路的复杂网络,所以, $A_{uv}=A_{vu}$ 并且 $A_{vv}=0$ 。

2.2 基于连续时间量子游走的链路预测算法

网络 G 的节点集合 V 构成整个量子游走的状态空间,每一个节点 v 代表量子游走状态的一个基态,记为 $|v\rangle$ 。 $H \in C^N$ 是由基态 $|v\rangle$ 张成的希尔伯特空间。这些基态是相互正交的,表示如下:

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, |2\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, |N\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad (2)$$

在任意时刻 t ,随机游走者在 G 上的连续量子游走的状态 $|\varphi(t)\rangle$ 是所有基态的一个线性叠加态,即

$$|\varphi(t)\rangle = \sum_{v \in V} \alpha_v(t) |v\rangle \quad (3)$$

其中: $\alpha_v(t)$ 是对应基态 $|v\rangle$ 在 t 时刻的概率幅, $\alpha_v(t) \in C$ 且 $|\alpha_v(t)| \in [0,1]$ 。随机游走者在时刻 t 处于基态 $|v\rangle$ 的概率 $p(|v\rangle,t) = \alpha_v(t) \alpha_v^*(t)$,其中 $\alpha_v^*(t)$ 是 $\alpha_v(t)$ 的复共轭。在任意时刻 t ,满足 $\sum_{v \in V} p(|v\rangle,t) = 1$ 。

不同于经典随机游走,连续量子游走的游走过程不是一个马尔可夫链。其状态向量 $|\varphi(t)\rangle$ 随时间 t 的演变过程由如下的酉变换实现:

$$\frac{d}{dt} |\varphi(t)\rangle = iA|\varphi(t)\rangle \quad (4)$$

由式(4)可见,状态演变过程取决于邻接矩阵 A 。

假定游走者在初始时刻处于状态 $|v\rangle$ (对应于顶点 v),由式(4)可得:经过时间 t ,游走者所处的状态

$$|\varphi(t)\rangle = e^{-iAt} |v\rangle \quad (5)$$

$p_{vu}(t)$ 表示游走者初始时刻处于顶点 v ,经过时间 t 后处于顶点 u 的概率

$$p_{vu}(t) = | \langle u | \varphi(t) \rangle |^2 \quad (6)$$

类似文献^[8],本文定义了叠加的连续时间量子游走(SCTQW)相似度权重:

$$S_{uv}^{\text{SCTQW}}(T=n\delta t) = \sum_{l=1}^n S_{uv}^{\text{LCTQW}}(l\delta t) \quad (7)$$

其中:观测周期为 T , n 是大于1的正整数, δt 为单位观测时间。 $S_{uv}^{\text{LCTQW}}(l\delta t)$ 是基于局部连续时间量子游走(LCTQW)的相似度权重,定义如下:

$$S_{uv}^{\text{LCTQW}}(l\delta t) = \frac{|\Gamma(u)|}{2|E|} \cdot p_{uv}(l\delta t) + \frac{|\Gamma(v)|}{2|E|} \cdot p_{vu}(l\delta t) \quad (8)$$

其中: $\frac{|\Gamma(u)|}{2|E|}$ 和 $\frac{|\Gamma(v)|}{2|E|}$ 分别对应于节点 u 和 v 的初始资源分配。很明显, $S_{uv}^{\text{SCTQW}} = S_{vu}^{\text{SCTQW}}$ 。

综上,基于连续时间量子游走的链路预测算法如下:

- 对网络 G ,由式(1)得其对应的邻接矩阵 A ;
- 从 G 中任一节点 v 出发,依据(6)计算从节点 v 到节点 $u(u \in V \setminus v)$ 经过时间 T 的转移概率 $p_{vu}(t)$;
- 节点 v 和节点 u 的相似度由式(7)和(8)即得。
- 重复步骤b),可得 G 中所有节点对的相似度。

3 实验评价

3.1 实验评价模型

U 定义为所有可能的连边集合,其中有 $N \times (N-1)/2$ 个元素。为了测试算法的准确性,将已知的连边 E 分为训练集 E^T 和测试集 E^P 两部分,其中 $E = E^T \cup E^P$,且 $E^T \cap E^P = \emptyset$ 。将不属于 E 的边定义为不存在的边。本文中采用的衡量链路预测算法精确度的指标包括AUC^[18]、Precision^[19]和Recall^[19]指标。

AUC从整体上衡量算法的精确度。AUC中,每次随机从 E^P 中选择一条边,将其与 $(U-E)$ 中一条随机生成的边比较相似度值。在 n 次独立比较中,如果测试集边的分值高的次数为 n' 次,则

$$AUC = \frac{n' + 0.5(n - n')}{n} \quad (9)$$

如果算法产生的相似度是独立和均匀分布的,则 AUC 值为 0.5。AUC 值超过 0.5 的部分说明算法在多大程度上比随机选择更准确。

Precision 定义为在前 L 个预测边中被预测准确的比例。如果有 m 个预测准确,即排在前 L 的边中有 m 个在测试集中,则 Precision 定义为:

$$Precision = m/L \quad (10)$$

显然,Precision 越大,预测越准确。通过训练集 E^T 预测可能产生的连边,并将可能连边的相似度由大到小排列,即按照相似度的值进行排序,从中抽取前 L 条相似度最大连边。通过 Precision 指标确定算法的优劣程度。

Recall 与 Precision 定义类似,定义为预测准确的链路和所有存在链路的比率。Recall 越大,预测越准确。

3.2 实验数据集

为了评价算法的有效性,本文使用了六个典型的真实网络数据集开展实验。本章实验中使用到的数据集包括:a)悲惨世界人物合作网络:Lesmis 网络,包含 77 个节点,代表小说中的角色;127 条连边,代表两个角色同时出现在一幕或多幕中。b)美国橄榄球俱乐部网络:football 网络,每个节点代表了参加美国 2000 年橄榄球赛季的高校代表队,连边则表示相应的两支球队间至少有一场比赛。c)科学家合作网络(NS):选择了其中最大连通子图 379 个作者的关系网络,节点代表了作者,连边表示两个作者间的合作关系。d)爵士音乐人合作网络(jazz):这个网络包含了 198 个音乐人和 2 742 个音乐人合作关系。e)美国航空运输网络(USair):包含 332 个节点,代表航空站,2 126 条连边,代表航线。f)西班牙 Rovira i Virgili 大学邮件网络(email):其中节点代表邮件用户,连边表示用户之间的通信。表 1 总结了实验数据集的各项拓扑属性:

表 1 网络结构属性表

data sets	n	m	$\langle k \rangle$	$\langle d \rangle$	C
lesmis	77	254	6.6	2.64	0.5731
football	297	2148	14.46	2.46	0.2924
NS	379	914	4.82	6.04	0.7412
jazz	198	2742	27.7	2.24	0.6175
USAir	332	2126	12.8	2.74	0.6252
email	1133	5451	9.62	3.6	0.22

其中, N 代表节点数, m 代表连边数, $\langle k \rangle$ 代表平均度, $\langle d \rangle$ 表示平均距离, C 是集群系数。

3.3 基准预测器

为了考察本算法优劣,五种典型预测算法的结果被作为基础算法与本算法的结果相比较。其中两种算法是基于经典随机游走,local random walk 和 superposed random walk。另外,文献[20,21]指出,common neighbors(CN)指标,adamic adar(AA)指标^[22]和 resource allocation(RA)指标^[23]在很多情况下有更好的预测性能。本文也将这三种指标作为基础比较算法。这几种指标的简要介绍如下:

a) common neighbours (CN)

$$s_{xy}^{CN}(x, y) = |N(x) \cap N(y)| \quad (11)$$

其中: $N(x)$ 是节点 x 的邻居节点集合。CN 指标基于这样的观察:两个节点如果有更多的共同邻居,则它们更相似。

b) adamic-adar (AA)

$$S_{AA}(x, y) = \frac{1}{\sum_{z \in N(x) \cap N(y)} \log(k(z))} \quad (12)$$

其中, $k(z)$ 是节点 z 的度。AA 指标基于这样的观察:度小的共同邻居节点的贡献大于度大的共同邻居节点。

c) resource allocation index (RA)

$$S_{RA}(x, y) = \frac{1}{\sum_{z \in N(x) \cap N(y)} k(z)} \quad (13)$$

RA 指标是基于复杂网络中动态的资源分配提出的。

d) local random walk (LRW)

该指标是基于经典随机游走的,只考虑有限步数的随机游走过程。随机游走的系统演化方程为 $p_x(t+1) = P^T p_x(t)$ 。其中转移概率矩阵 P ,当节点 u 和 v 相连时, $P_{uv} = 1/k(u)$,否则, $P_{uv} = 0$ 。 P_{uv} 表示游走者当前时刻 t 处于节点 u , $t+1$ 时刻处于节点 v 的概率。设定各节点的初始资源分布为 $q_u = k_u / |E|$,则基于 t 步游走的相似性为

$$s_{uv}^{LRW}(t) = q_u \cdot P_{uv}(t) + q_v \cdot P_{vu}(t) \quad (14)$$

e) Superposed Random Walk (SRW)

在 LRW 基础上,将 t 步和其之前的结果求和,就得到

$$s_{uv}^{SRW}(n) = \sum_{t=1}^n s_{uv}^{LRW}(t) \quad (15)$$

该指标是是赋予邻近节点更多的权重,即加强短路径的贡献。

3.4 实验结果与分析

为了检测算法的正确性,在每个数据集中随机选择 90% 的边作为测试集 E^T ,余下的 10% 的边作为 E^p ,采用前述的三种检测方法的结果比较各预测算法的预测结果。在实验中,只考虑网络的最大连通子图,因为很多基于相似性的算法,如共同邻居 CN,基于随机游走的相似性指标等等,都是基于网络结构信息的,所有孤点(度数为 0 的节点)的相似性值都被赋予 0,而且对不同连通子图中的两个节点也会赋予 0。实际上,在划分训练集和测试集的过程中,必须保证训练集依然是连通的。表 2~4 分别为不同算法下,AUC、Precision 和 Recall 结果的比较。表格中,LRW 和 SRW 算法,括号中的数字是随机游走中的最优步数。SCTQW 算法中, $\delta t = 0.02$,括号中的数字对应连续时间量子游走中的最优周期 $T = n\delta t$ 中的 n 。

在表 2 中,在 6 个数据集上,基于经典随机游走和量子游走的 AUC 结果总体优于其他链路预测指标。基于量子游走的算法 AUC 结果均非常接近于最优算法结果。Email 数据集中,SCTQW 的结果比最优结果 SRW 低 0.93%,但明显优于 CN、AA 和 RA 指标的结果。在其他数据集中,SCTQW 的结果最多仅低于最优结果 0.37%。在数据集 NS 和 Usair 中,其差值仅为 0.04% 和 0.05%。

在 AUC 非常接近的情况下,本文比较 recall 和 precision 指标的结果。Recall 的比较结果如表 3 所示。基于量子游走的 Recall 结果总体优于其他链路预测指标。Lesmis 和 Jazz 数据集中,基于量子游走的指标预测结果优于基于随机游走指标的预测结果 13%~15%。在 Email 数据集中,基于量子游走的指标预测结果是基于随机游走指标预测结果的 1.3 倍左右。而在 Usair 数据集中,量子游走的预测结果更达到随机游走结果的 2.5 倍。

Precision 的比较结果如表 4 所示。Precision 指标表现出

和 Recall 指标类似的结果。基于量子随机游走的 Precision 结果总体优于其他链路预测指标。值得注意的是,在 jazz 数据集中,本文的算法预测结果是基于经典随机游走预测结果的 2 倍。在 Usair 数据集中,本文的算法预测结果是基于经典随机游走预测结果的 3 倍。在 Email 数据集中,则达到 5 倍。

表 2 AUC 预测精度比较

AUC	CN	AA	RA	LRW	SRW	SCTQW
lesmis	0.9547	0.9632	0.9634	0.9537(3)	0.9624(4)	0.9597(11)
football	0.8455	0.8455	0.8456	0.8717(3)	0.8705(4)	0.8702(5)
NS	0.9789	0.9826	0.9827	0.9900(3)	0.9924(2)	0.9921(19)
jazz	0.9563	0.9631	0.9721	0.9665(1)	0.9666(1)	0.9684(4)
usair	0.9544	0.9662	0.9723	0.9679(3)	0.9741(2)	0.9736(7)
email	0.8561	0.8581	0.8579	0.9326(5)	0.9349(8)	0.9256(9)

表 3 Recall 预测精度比较

Recall	CN	AA	RA	LRW	SRW	SCTQW
lesmis	0.5214	0.5579	0.5743	0.5065(2)	0.4517(2)	0.5847(16)
football	0.329	0.2855	0.2839	0.3409(1)	0.3409(1)	0.3824(34)
NS	0.4218	0.5605	0.5857	0.3693(1)	0.4052(2)	0.4398(17)
jazz	0.5218	0.5232	0.5399	0.4111(1)	0.4111(1)	0.5729(4)
usair	0.3858	0.3995	0.4643	0.1648(6)	0.1915(16)	0.4447(6)
email	0.1761	0.1585	0.1443	0.0625(6)	0.0833(40)	0.1592(4)

表 4 Precision 预测精度比较

Precision	CN	AA	RA	LRW	SRW	SCTQW
lesmis	0.2262	0.2168	0.2172	0.2019(4)	0.1954(10)	0.2062(1)
football	0.3191	0.287	0.2862	0.3110(1)	0.3110(1)	0.3270(27)
NS	0.3902	0.5303	0.5574	0.3634(1)	0.3920(2)	0.4190(18)
jazz	0.816	0.8366	0.8208	0.3518(8)	0.4075(19)	0.8845(1)
usair	0.6129	0.6205	0.6357	0.2536(40)	0.2052(40)	0.6601(6)
email	0.3093	0.3136	0.2568	0.0778(14)	0.0614(40)	0.3287(5)

5 结束语

本文把量子游走引入到链路预测问题中,以研究相对于经典随机游走,基于量子游走的链路预测方法是否可以提高预测精度。本文提出了基于量子游走的相似度指标,并将该指标的预测结果在 6 个真实数据集上与其他 5 种经典指标结果相比较。结果表明,连续时间量子游走的链路预测结果的 AUC 值和随机游走的结果非常接近。而在 Precision 和 Recall 指标上,远优于随机游走的链路预测结果。

本文是在几个不同数据集上比较各指标的预测结果。下一步将深入研究在不同结构特征的网络下,本指标的预测结果。另外,离散时间量子游走的特征与连续时间量子游走的特征有很大区别,基于离散时间量子游走的链路预测方法也是下一步的研究工作。

参考文献:

- [1] Getoor L, Diehl C P. Link mining: a survey[J]. SIGKDD Explorations, 2005,7(2):3.
- [2] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks[C]//Proc of the 12th international conference on Information and knowledge management, New Orleans, LA, USA, 2003, (ACM Press, New York) 2003:556-559.
- [3] Lü L, Zhou T. Link prediction in complex networks: a survey[J]. Physica A (Amsterdam), 2011,390(6): 1150.
- [4] Fouss F, Pirotte A, Renders J M, et al. Random-walk computation of

similarities between nodes of a graph with application to collaborative recommendation[J]. IEEE Trans. Knowl. Data. Eng,2007,19: 355.

- [5] Tong H, Faloutsos C, Pan J Y. Fast random walk with restart and its applications[C]//Proc of the 6th International Conference on Data Mining, Hong Kong, China, 2006, (IEEE Press, Washington) 2006: 613-622 .
- [6] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine, Comput. Netw. ISDN Syst, 1998,30:107.
- [7] Jeh G, Widom J. SimRank: a measure of structural-context similarity [C]//Proc of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, 2002, (ACM Press, New York) 2002: 271-279.
- [8] Liu W, Lü L. Link prediction based on local random walk[J]. Europhys. Lett. ,2010, 89: 58007.
- [9] Feng X, Zhao J C, Xu K. Link prediction in complex networks: a clustering perspective[J]. Eur. Phys. J. B. , 2012,85(1):3.
- [10] Aharonov Y, Davidovich L, Zagury N. Quantum random walks[J]. Phys. Rev. A, 1993,48(2):1687.
- [11] Childs A, Farhi E, Gutmann S. An example of the difference between quantum and classical random walks, Quan. Info. Proc. , 2002,1: 35.
- [12] Kempe J. Quantum random walks-an introductory overview, Contemp. Phys, 2003,44: 307.
- [13] S'anchez-Burillo E, Duch J, Gómez-Gardeñes J, et al. Quantum navigation and ranking in complex networks[J]. Nat. Sci. Rep, 2012,2: 605.
- [14] Emms D, Wilson R C, Hancock E R. Graph matching using the interference of continuous-time quantum walks[J]. Pattern Recognit, 2009,42: 985-1002.
- [15] Emms D, Wilson R C, Hancock E R. Graph matching using the interference of discrete-time quantum walks[J]. Image Vis. Comput, 2009,27: 934-949.
- [16] Farhi E, Gutmann S. Quantum computation and decision trees[J]. Phys. Rev. A, 1998,58:915.
- [17] Mülken O, Blumen A. Continuous-time quantum walks: models for coherent transport on complex networks[J]. Phys. Rep. , 2011, 37:502.
- [18] Hanely J A, McNeil B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. Radiology, 1982, 143: 29.
- [19] Herlocker J L, Konstann J A, Terveen K. Evaluating collaborative filtering recommender systems[J]. ACM Trans Inf Syst, 2004, 22(1): 5-53.
- [20] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks[J]. Amer. Soc. Inf. Sci. and Tech. , 2007,58(7): 1019.
- [21] Sachan M, Contractor D, Faruque T A, et al. Using content and interactions for discovering communities in social networks[C]//Proc of the 21st International Conference on World Wide Web, Lyon, France,2012, (ACM Press, New York) 2012: 331-340.
- [22] Adamic L A, Adar E. Friends and neighbors on the web, Soc. Networks, 2003,25: 211.
- [23] Zhou T, Lü L, Zhang Y C. Predicting missing links via local information[J]. Eur. Phys. J. B, 2009,71: 623.