

社会网络中基于主题的影响力最大化算法*

朱玉婷¹, 李雷¹, 施化吉¹, 周从华¹, 施磊磊¹, 徐慧²

(1. 江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013; 2. 大全集团, 江苏 扬中 212211)

摘要: 为了解决现有的影响力最大化研究没有充分考虑主题对影响力节点挖掘的影响而导致特定主题下节点集合的影响范围不大这一问题,提出了一种社会网络中基于主题的影响力最大化算法TIM。该算法首先根据主题敏感阈值对初始节点集进行预处理,剔除干扰节点,再在新的节点集合上分两个阶段进行节点挖掘。第一阶段挖掘主题权威性大的节点,第二阶段挖掘主题影响增量最大的节点,最后综合两个阶段的节点作为结果集并进行实验验证。实验结果表明,相比其他算法,TIM算法挖掘的节点集合在特定主题下的影响范围更大,时间复杂度更低。

关键词: 社会网络; 影响力最大化; 主题; 节点挖掘; 节点集

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2016)12-3611-04

doi:10.3969/j.issn.1001-3695.2016.12.021

Topic-based influence maximization algorithm on social networks

Zhu Yuting¹, Li Lei¹, Shi Huaji¹, Zhou Conghua¹, Shi Leilei¹, Xu Hui²

(1. School of Computer Science & Telecommunication Engineering, Jiangsu University, Zhenjiang Jiangsu 212013, China; 2. Daqo Group Company, Yangzhong Jiangsu 212211, China)

Abstract: To solve the problem that recent researches of influence maximization haven't fully considered that topic has an impact on influential nodes mining, which lead to low influence scope under specific topic, this paper proposed a topic-based influence maximization algorithm (TIM). This algorithm first pretreated the initial node set according to topic sensitive threshold, and removed the interference nodes, then mined nodes in two stages on the new node set. In the first stage, it mined the nodes with high topic authority; in the second stage, it mined the nodes with the biggest topic influence increment. At last, it combined the two stages of the node as a result set and made an experimental verification. The experimental results indicate that the nodes set mined by the proposed algorithm improve the influence scope under specific topic and the algorithm cost less time compare to other influence maximization algorithms.

Key words: social network; influence maximization; topic; nodes mining; node set

0 引言

随着社会网络的迅速普及,人们越来越依赖于通过社会网络进行交友、交流、产品推广等活动,由此不断产生了大量数据,其中所包含的丰富的语义内容和结构吸引了众多学者对其进行深入研究,如用户行为预测、话题检测、影响力最大化等,其中影响力最大化是当前的一个研究热点。影响力最大化是指从社会网络中找出一个高影响力的节点集合,通过该节点集合可以将信息最大限度地传播到更多的节点。该研究在朋友推荐^[1]、市场营销^[2]、谣言控制^[3]等领域具有广泛的应用价值。

影响力最大化问题由 Domingos 等人^[4]首次作为一个算法问题引入到社会网络领域,并且该问题已被证明是一个 NP-难问题,可以使用一种爬山贪心算法得到精确度为 $(1 - 1/e)$ 的近似最优解^[5]。另有一些学者在此工作的基础上又提出了一些算法^[6-8],有效地提高了挖掘影响力节点的时间效率。然

而,这些研究中一个节点对其他节点的影响力都是相同的,即一个节点激活其他节点的激活概率都是相同的。但事实上节点在社会网络中的影响力与节点之间的关系及主题有关,同一节点在不同主题下对其他节点的影响力不同^[9-11]。文献[12]通过节点的主题分布计算节点在主题层次上的激活概率,然后基于此概率提出一种 TopicRank 算法挖掘特定主题下最具影响力的节点集合,一定程度上提高了节点集合的影响范围。文献[13]结合节点的主题偏好,提出一种两阶段算法 GAUP 挖掘特定主题下最具影响力的节点集合,同样在一定程度上提高了节点集合的影响范围。但这些研究都只考虑了主题对节点的激活概率的影响,并没有充分考虑主题对挖掘影响力节点的影响,使得特定主题下节点集合的影响范围小且时间效率不高。为此,提出一个社会网络中基于主题的影响力最大化算法(topic-based influence maximization algorithm, TIM),以挖掘特定主题下最具影响力的节点集合。首先根据主题敏感阈值对社会网络中的所有节点进行预处理得到社会网络的一个节点子

收稿日期: 2015-09-22; **修回日期:** 2015-11-06 **基金项目:** 国家自然科学基金资助项目(71271117);江苏省科技支撑计划资助项目(BE2011156)

作者简介: 朱玉婷(1991-),女,江苏扬州人,硕士研究生,主要研究方向为社会网络分析、数据挖掘(zhu_yt_candy@163.com);李雷(1976-),男,河南南阳人,讲师,主要研究方向为智能信息处理、社会网络分析;施化吉(1964-),男,浙江台州人,教授,主要研究方向为智能信息处理、电子商务、社会网络分析;周从华(1978-),男,江苏盐城人,副教授,博士,主要研究方向为智能信息处理、数据挖掘;施磊磊(1989-),男,江苏海安人,博士研究生,主要研究方向为复杂系统、社会计算;徐慧(1980-),男,江苏扬中人,硕士,主要研究方向为智能制造、企业信息化。

集,然后在该节点子集上分两阶段进行影响力节点的挖掘。第一阶段从节点子集中静态地挖掘主题权威性大的节点;第二阶段考虑到社会网络的传播特性,将第一阶段挖掘出的节点利用基于主题的信息传播模型(topic-based information propagation model, TPM)进行信息传播,然后贪心地挖掘主题影响增量最大的节点;最后综合两阶段所挖掘的节点得到特定主题下的节点集合。

1 问题描述与相关定义

社会网络中基于主题的影响力最大化是在一个社会网络中对给定的主题找到一个指定大小的初始节点集合,该集合中的所有节点在初始时刻被激活成活跃状态,然后根据特定的信息传播模型,遵循传播机制进行信息传播,最终激活社会网络中最多的节点。下文首先给出基于主题的社会网络及其影响范围的定义。

定义1 社会网络 G 。 G 是一个基于主题的社会网络, V 是 G 中节点的集合, E 是 G 中边的集合, T 是 G 中主题的集合, 则 G 可表示成 $G(V, E, T)$, 其中 $V = \{v_1, v_2, \dots, v_n\}$, $|V| = n$; $E = \{\langle u, v \rangle \mid u, v \in V\}$, 当节点 u, v 之间有联系时, 它们之间则存在一条边, 反之则无; $T = \{1, 2, \dots, z\}$, $|T| = z$ 。

定义2 影响范围 $\delta(S|t)$ 。设 S 是一个大小为 K 的节点集合, 则 $\delta(S|t)$ 表示 S 在主题 t 下的影响范围, 即通过 S 进行主题信息传播后, 最终被激活的节点数目。

社会网络中基于主题的影响力最大化可以表述为: 给定一个参数 K 和主题 t , 在社会网络 $G(V, E, T)$ 中, 找到一个大小为 K 的初始节点集合 S , $|S| = K$, 使得 $\delta(S|t)$ 最大, 即求 $\max \{\delta(S|t) \mid |S| = K, S \subseteq V, t \in T\}$ 。

2 基于主题的信息传播模型

本章先介绍独立级联模型^[5], 然后给出基于主题的信息传播模型 TPM。

2.1 独立级联模型

独立级联模型(independent cascade model, IC)^[5] 是社会网络中应用最广泛的信息传播模型, 该模型用于模拟信息在社会网络中的传播。IC 模型中一个活跃节点通过与其邻居节点之间的激活概率去激活其不活跃的邻居节点, 激活概率越大, 激活成功的可能性越高。IC 模型的信息传播过程如图1所示。

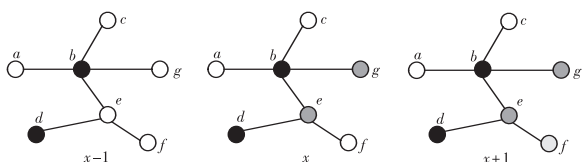


图1 独立级联模型信息传播过程

由图1可知, 在 $x-1$ 时刻有 b, d 两个活跃节点, 它们以一定的概率尝试去激活其邻居节点, b 可以去激活节点 a, c, e, g, d 可以去激活节点 e , 但是 b 和 d 都只有一次机会去激活其所有邻居节点, 若激活失败, 则不能再继续激活。又 b 和 d 有共同的邻居节点 e , 它们可以同时去激活节点 e , 但 e 只能被其中一个节点所激活。在 x 时刻, 将被成功激活的节点 e, g 加入到活跃节点集合中, 继续以上激活过程直至没有新的节点被激活为止。

2.2 基于主题的信息传播模型

IC 模型中激活概率是随机产生的, 而在真实的信息传播过

程中, 节点之间的激活概率与节点之间的社交关系以及主题相关, 且不同主题下节点之间的激活概率不同。基于主题的信息传播模型 TPM 通过计算特定主题 t 下节点 u, v 之间的激活概率 $P'_{u,v}$ 来更加真实地模拟社会网络中的信息传播。

$P'_{u,v}$ 不仅与节点之间的社交联系紧密程度相关且受到节点主题分布的影响。

节点之间的联系次数越多, 节点之间的关系越亲密, 激活概率越大, 节点之间的社交联系紧密程度可用用户亲密度来表示。

定义3 用户亲密度 $C_{u,v}$ 。 $C_{u,v}$ 是用户亲密度, 表示节点 u 和 v 之间联系的频率, 可通过式(1)计算获得。

$$C_{u,v} = \frac{R_{u,v}}{\sum_{i=1}^n R_{u,v_i} + \sum_{i=1}^n R_{v,v_i}} \quad u, v, v_i \in V \quad (1)$$

其中: R_{u,v_i} 表示社会网络中节点 u 与节点 v_i 联系的次数, $R_{u,v}$ 表示节点 u 与 v 的联系次数。

两个节点的主题分布越相似越容易发生信息传播行为, 节点之间主题分布的相似程度可用主题相似度来表示。

定义4 主题相似度 $\text{SIM}_{u,v}^t$ 。 $\text{SIM}_{u,v}^t$ 是主题相似度, 表示两个节点之间主题分布的相似程度。任意节点 u 的主题分布情况可由主题模型 LDA^[14] 计算得出, 记为 $\psi_u^t = (\varphi_u^1, \varphi_u^2, \dots, \varphi_u^z)$, ($u \in V$), 则节点 u 与 v 之间的主题相似度如式(2)所示。

$$\text{SIM}_{u,v}^t = \frac{\psi_u^t \times \psi_v^t}{|\psi_u^t| \times |\psi_v^t|} = \frac{\sum_{i=1}^z \varphi_u^i \times \varphi_v^i}{\sqrt{\sum_{i=1}^z \varphi_u^{i2} \sum_{i=1}^z \varphi_v^{i2}}} \quad u, v \in V, t \in T \quad (2)$$

$\text{SIM}_{u,v}^t$ 值越大, 表示两个节点的主题相似程度越大, 反之, 则越小。

此外, $P'_{u,v}$ 还与节点本身对该主题信息的接收情况有关, 若节点经常接收该主题的信息, 则当其邻居节点向它传播有关该主题的信息时, 激活它的可能性就更大。节点对主题信息的接收情况可用主题接收度来表示。

定义5 主题接收度 G_v^t 。 G_v^t 是主题接收度, 表示节点 v 对于某特定主题 t 的接收情况, 如式(3)所示。

$$G_v^t = \frac{m_t}{\sum_{j=1}^z m_j} \quad t, j \in T \quad (3)$$

其中: m_t 是节点 v 所接收的关于主题 t 的信息数目, m_j 是节点 v 所接收的关于主题 j 的信息数目。

因此, 特定主题 t 下节点 u, v 之间的激活概率 $P'_{u,v}$ 可通过式(4)计算。

$$P'_{u,v} = G_v^t \times (\alpha \text{SIM}_{u,v}^t + (1 - \alpha) C_{u,v}) \quad P'_{u,v} \in [0, 1] \quad (4)$$

其中: α 为调和因子, 用于调和不同情况下主题相似度 $\text{SIM}_{u,v}^t$ 与用户亲密度 $C_{u,v}$ 所占的比重。通常情况下 $\text{SIM}_{u,v}^t$ 和 $C_{u,v}$ 对整体激活概率的影响一样重要, 故 α 取值为 0.5。

TPM 模型的传播过程与 IC 模型一样, 每个活跃节点只有一次机会去激活其邻居节点, 且这一过程是相互独立的。不同的是, TPM 采用不同主题 t 下节点之间的激活概率 $P'_{u,v}$, 更加符合社会网络的信息传播情况。

3 基于主题的影响力最大化算法

基于主题的影响力最大化是在基于主题的社会网络 $G(V, E, T)$ 中, 找到特定主题 t 下一个大小为 K 的节点集合, 通过该集合可以将信息传播到最多的节点。为了挖掘出这些节点, 首先利用主题模型 LDA 对社会网络进行主题建模, 得到基于主题的社会网络 $G(V, E, T)$ 和节点的主题分布 $\psi_u^T = (\varphi_u^1, \varphi_u^2, \dots, \varphi_u^z)$ ($u \in V$)。然后用 TIM 算法进行节点挖掘。

3.1 节点数据预处理

在挖掘特定主题下影响力节点时,可能存在许多与该主题不相关或对该主题完全没有影响力的节点,这部分节点极大地影响了挖掘节点的效率。为此这里定义一个主题敏感阈值 θ_t 和节点主题敏感度 I'_u 。 I'_u 表示节点 u 对于主题 t 的敏感程度,可由节点 u 的主题分布中的元素 φ'_u 获得。 θ_t 可由所有节点的主题敏感度的平均值获得。 I'_u 和 θ_t 的计算如式(5)和(6)所示。

$$I'_u = \varphi'_u \quad u \in V, t \in T \quad (5)$$

$$\theta_t = \frac{\sum_{i=1}^n I'_{v_i}}{n} = \frac{\sum_{i=1}^n \varphi'_{v_i}}{n} \quad v_i \in V, t \in T \quad (6)$$

对 $\forall u \in V$, 当 $I'_u < \theta_t$ 时,可以认为节点 u 对于主题 t 的信息传播能力较差,难以成为主题 t 的影响力节点,可以先剔除这些节点再进行影响力节点的挖掘。

预处理后的节点子集 $U = \{u | I'_u \geq \theta_t, u \in V, t \in T\}$ 。由于 $|U|$ 明显低于整个社会网络的节点数目 $|V|$,故从 U 中进行影响力节点挖掘可以降低算法的时间复杂度且不影响结果的准确性。

3.2 主题权威节点挖掘

在影响力最大化中,初始传播节点的选择关系到之后的信息传播走向和影响范围,因此找到具有最高影响力的初始传播节点显得至关重要^[15]。而已有的一些影响力最大化算法中,初始传播节点都是从整个社会网络中挖掘的影响增量最大的节点,这些节点虽然在整个社会网络上具有较高的影响力,但是对于特定的主题并不一定具有高影响力,且计算每个节点的影响增量时间效率低下。

为此,这里选取特定主题下主题权威性 $\text{Authority}(u, t)$ 大的 $\lceil \beta K \rceil$ ($\beta \in [0, 1]$) 个节点作为初始传播节点,这样既节约了时间又保证了这些节点具有该主题下的高影响力,更加容易将主题信息传播到更多的节点^[16]。节点的主题权威性与节点传播关于该主题的信息被其他节点接收的能力有关,计算方法如式(7)所示。

$$\text{Authority}(u, t) = \frac{\sum N_{v_i, t}}{\sum W_{u, t}} \quad v_i, u \in U, v_i \neq u, t \in T \quad (7)$$

其中: $W_{u, t}$ 表示节点 u 所有有关主题 t 的信息数; $N_{v_i, t}$ 表示这些信息被其他节点(不包括 u)接收的数目。

3.3 主题影响力节点挖掘

第一阶段只是静态地选取 $\lceil \beta K \rceil$ 个主题权威性大的节点作为影响力节点,并没有考虑到社会网络的信息传播特性。因此,将第一阶段获取的节点作为初始传播节点,通过 TPM 模型模拟社会网络中的信息传播,然后再采用贪心策略迭代地挖掘主题影响增量最大的 $K - \lceil \beta K \rceil$ 个节点作为剩余的影响力节点。其中主题影响增量最大是指加入某个节点 u 之后,集合的影响范围与加入 u 之前影响范围的差值最大,如式(8)所示。

$$\Delta(u|t) = \max\{\delta(S \cup \{u\} | t) - \delta(S | t)\} \quad u \in U, S \subseteq U, t \in T \quad (8)$$

3.4 TIM 算法描述

首先进行数据预处理,然后静态选取主题权威性大的节点,最后贪心地选取剩余的主题影响力节点,具体的算法描述见算法 1。

算法 1 TIM 算法

输入: 基于主题的社会网络 $G(V, E, T)$ 、主题 t 、正整数 K 、参数 β 。

输出: 影响力节点集合 S 。

a) 初始化影响力节点集合 $S = \varphi$, 根据式(5)(6)计算主题 t 下节点的主题敏感度 I'_u 和主题敏感阈值 θ_t 。

b) 将 $I'_u < \theta_t$ 的节点剔除后剩余的节点集合作为节点挖掘集合 U , 即 $U \leftarrow \{u | I'_u \geq \theta_t, u \in V, t \in T\}$ 。

c) 根据式(7)计算 U 中所有节点的主题权威性 $\text{Authority}(u, t)$, 并对其进行排序, 选取其中权威性值排行前 $\lceil \beta K \rceil$ 的节点作为 S 中的节点, $S \leftarrow \text{Top}(\lceil \beta K \rceil)_{\text{sort}}(\text{Authority}(u, t))$ 。

d) 将上一步中得到的 S 中的节点设置为活跃状态, 剩下的 $U - S$ 集合中的节点为不活跃状态。将 S 中的节点作为初始传播节点通过 TPM 模型进行信息传播, 得到集合 S 的影响范围 $\delta(S | t)$ 。

e) 计算 $U - S$ 中所有节点的主题影响增量 $\Delta(u | t) = \delta(S \cup \{u\} | t) - \delta(S | t)$ 。

f) 选择主题影响增量最大的节点 u 加入到集合 S 中并记录下该最大增量 $\max \delta$, 即 $u = \{v | \max\{\Delta(v | t)\}, v \in U - S\}$, $\max \delta \leftarrow \Delta(u | t)$, $S \leftarrow S \cup \{u\}$ 。然后将新的 S 通过 TPM 继续进行信息传播并得到新的影响范围 $\delta(S | t)$ 。

g) 对于 $U - S$ 中的任意节点 v , 根据子模特性, 若 $\Delta(v | t) > \max \delta$, 则 $\Delta(v | t)$ 为最大的主题影响增量, 并且其他节点不需要再计算增量, 更新 $\max \delta \leftarrow \Delta(v | t)$, $S \leftarrow S \cup \{v\}$ 。

h) 重复 g) ($K - \lceil \beta K \rceil - 1$) 次, 循环结束。

i) 输出影响力节点集合 S 。

4 实验结果与分析

4.1 实验数据集

实验采用 Twitter2009 年 9 月份的部分数据集, 包括 52 863 个用户和 3 257 374 条 tweets。首先将每个用户所发表的 tweets 归入一个文档, 然后将所有用户文档作为文档集, 利用 LDA 主题模型进行主题分类, 得到 $Z = 60$ 个主题以及所有用户的主题分布。

根据得到的结果, 建立基于主题的社会网络 $G(V, E, T)$, 包含 52 863 个节点和 1 637 625 条有向边, 其中节点代表用户, 有向边代表用户之间存在的联系, T 表示 G 中的主题。

4.2 主题挖掘

利用主题模型 LDA 共挖掘出 60 个主题, 其中的三个主题及这些主题排行前 8 位的关键词如表 1 所示。

表 1 Twitter 数据主题挖掘

Topic 11		Topic 20		Topic 39	
WORD	PROB.	WORD	PROB.	WORD	PROB.
Music	0.0634	Game	0.0450	Tourism	0.0164
Songs	0.0349	Live	0.0364	Tourist	0.0150
Lyrics	0.0295	Team	0.0355	Scenery	0.0150
Band	0.0227	Club	0.0345	Mountain	0.0145
Concert	0.0205	Player	0.0259	Sight	0.0134
Video	0.0134	Basketball	0.0199	Food	0.0124
CD	0.0125	Backboard	0.0128	Guide	0.0108
Star	0.0104	Rockets	0.0117	Rest	0.0101

由表 1 可知, 这三个主题分别是 Music、Basketball、Tourism。下文以 Music 这一主题为例进行实验分析。

4.3 参数的确定

为了确定算法中的调节参数 β , 下面分别对不同 β 和 K 下节点集合的影响范围和算法的运行时间进行实验分析。为了减少实验的误差, 将进行 100 次重复实验计算平均值, 实验结果如图 2、3 所示。

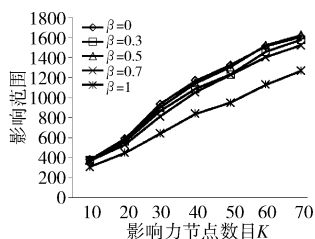


图2 节点影响范围

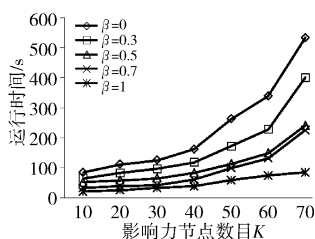


图3 算法运行时间

由图2、3可知,对于任意影响力节点数目 K , $\beta=1$ 的情况下影响范围最差,但运行时间最短,这是因为当 $\beta=1$ 时,算法完全是采用启发式思想,所以运行快,但没有考虑到社会网络的传播特性,故影响范围较小; $\beta=0$ 时影响范围较大,但用时相对较长,而 $\beta=0.5$ 时在一些 K 值下的影响范围几乎接近 $\beta=0$ 时的影响范围,且运行时间也相对较短,因此,综合考虑运行时间和影响范围,选取 $\beta=0.5$ 作为TIM算法的参数值。

4.4 与其他算法的比较

为了验证TIM算法的影响范围与运行效率,将其与TopicRank^[12]、GAUP^[13]以及不考虑主题的CELF算法^[7]进行比较。所有算法都在同配置的PC上运行,同样进行100次重复实验取平均值,实验结果如图4、5所示。

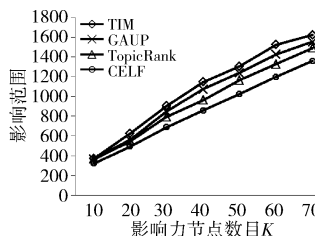


图4 节点影响范围

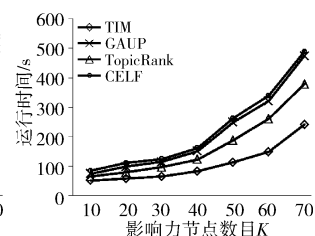


图5 算法运行时间

由图4可知,CELF算法的影响范围最低,这是因为其他三种算法都考虑了主题对影响力的影响,因而在特定主题下挖掘的节点的影响范围比CELF要大。同时,TIM算法明显优于考虑了主题的GAUP和TopicRank算法,这是因为TIM算法选取了主题权威性高的节点作为影响力节点,在特定主题下,这部分节点可以影响到更多的节点。而GAUP和TopicRank算法并没有考虑到节点的主题权威性,因此所选取的节点在特定主题下影响力不大。

由图5可知,TIM算法在任意影响力节点数目 K 下运行时间最短,这是因为TIM算法对数据进行了预处理操作,剔除了一些不影响结果的节点,并采用了静态的启发式思想,直接选取主题权威性大的影响力节点,相比其他仅采用贪心思想挖掘节点的算法,明显要快很多。实验结果表明,TIM算法在特定主题下,影响范围与时间效率都有了较大的提升。

5 结束语

本文主要研究了社会网络中基于主题的影响力最大化问题,提出了一个基于主题的信息传播模型TPM来更加真实地模拟社会网络中的信息传播,并提出了一个基于主题的影响力最大化算法TIM来挖掘特定主题下最具影响力的节点集合。实验表明,TIM算法在影响范围和时间效率上都比其他算法优越,适用于主题影响力节点的挖掘。当然,本文还有许多值得改进和进一步研究的地方。例如节点之间的激活概率有多种影响因素,如何综合这些因素更加精确地计算节

点间的激活概率;现有的研究没有考虑到社会网络中的时间概念,如何将时间因素融入到基于主题的影响力最大化算法中,等等。

参考文献:

- [1] 王珂,高琳.基于社交圈的在线社交网络朋友推荐算法[J].计算机学报,2014,37(4):801-808.
- [2] Zhang Yu, Wang Zhaoqing, Xia Chaolun. Identifying key users for targeted marketing by mining online social network [C]//Proc of IEEE International Conference on Advanced Information Networking and Applications Workshops. 2010: 644-649.
- [3] 谭娟.基于传染病模型的社交网络舆情话题传播[J].计算机工程与应用,2015,51(12):118-122.
- [4] Domingos P, Richardson M. Mining the network value of customers [C]//Proc of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2001: 57-66.
- [5] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network [C]//Proc of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2003: 137-146.
- [6] 曹玖新,董丹,徐顺,等.一种基于k-核的社会网络影响最大化算法[J].计算机学报,2015,38(2):238-248.
- [7] Leskovec J, Krause A, Guestrin C, et al. Cost-effective outbreak detection in networks [C]//Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2007: 420-429.
- [8] Tang Youze, Xiao Xiaokui, Shi Yanchen. Influence maximization: near-optimal time complexity meets practical efficiency [C]//Proc of ACM SIGMOD International Conference on Management of Data. 2014: 75-86.
- [9] 郭景峰,吕加国.基于信息偏好的影响最大化算法研究[J].计算机研究与发展,2015,52(2):533-541.
- [10] Tang Jie, Sun Jimeng, Wang Chi, et al. Social influence analysis in large-scale networks [C]//Proc of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009: 807-816.
- [11] Liu Lu, Tang Jie, Han Jiawei, et al. Mining topic-level influence in heterogeneous networks [C]//Proc of the 19th ACM International Conference on Information and Knowledge Management. 2010: 199-208.
- [12] Zhou Donghao, Han Wenbao, Wang Yongjun. Identifying topic-sensitive influential spreaders in social networks [J]. International Journal of Hybrid Information Technology, 2015, 8(2): 409-422.
- [13] Zhang Yunlong, Zhou Jingyu, Cheng Jia. Preference-based top-k influential nodes mining in social networks [C]//Proc of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. 2011: 1512-1518.
- [14] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [15] 田家堂,王铁彤,冯小军.一种新型的社会网络影响最大化算法[J].计算机学报,2011,34(10):1956-1965.
- [16] Barbieri N, Bonchi F, Manco G. Topic-aware social influence propagation models [J]. Knowledge and Information Systems, 2013, 37(3): 555-584.