



(12) 发明专利申请

(10) 申请公布号 CN 104484825 A

(43) 申请公布日 2015. 04. 01

(21) 申请号 201410740284. 8

(22) 申请日 2014. 12. 05

(71) 申请人 上海师范大学

地址 200234 上海市徐汇区桂林路 100 号

申请人 公安部第三研究所

(72) 发明人 张波 杨涛 宋倩倩 宋锋
胡斯卉

(74) 专利代理机构 上海申新律师事务所 31272
代理人 周云

(51) Int. Cl.
G06Q 50/00(2012. 01)

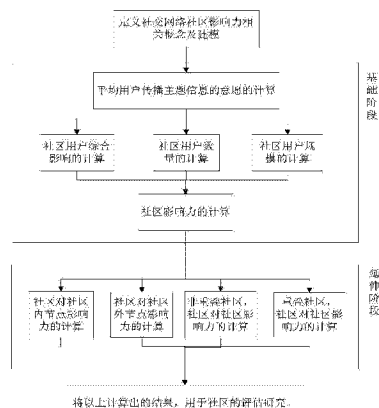
权利要求书6页 说明书13页 附图2页

(54) 发明名称

社交网络社区影响力评估算法

(57) 摘要

本发明社交网络社区影响力评估算法,是基于图论方法基础上的发展与应用,涉及社交网络技术领域。本发明第一、定义社交网络与社区影响力相关的概念及建模;第二、计算出用户传播主题信息的意愿、社区用户综合影响、社区用户数量、社区用户规模,再综合计算出社区影响力;第三、计算出社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力;第四、将计算出的社区影响力、社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,用于评估各社区间的影响力、节点对社区的重要性以及社区对节点作用。基于此,本发明对人人网,微博等社交网络影响进行评估,为更好传播信息以及网络安全等具有实际应用的前景。



1. 一种社交网络社区影响力评估算法,其特征为:基于图论方法,其基本步骤为:

第一步,定义社交网络与社区影响力相关的概念及建模;

第二步,基础阶段,计算出用户传播主题信息的意愿、社区用户综合影响、社区用户数量、社区用户规模,再综合计算出社区影响力;

第三步,延伸阶段,以基础阶段为基础,计算出社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力;

第四步,将计算出的社区影响力、社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,用于评估各社区间的影响力、节点对社区的重要性以及社区对节点作用。

2. 根据权利要求 1 所述的社交网络社区影响力评估算法,其特征为:所述第一步定义,包括步骤:

社交网络模型:社交网络可以形式化为一个二分图, $G = (V, E)$, V 代表社交网络中的用户集合, E 用来描述用户间的关系的集合;

社区模型:社区是由拥有相同兴趣爱好的用户组成,为社交网络的一个子图, $C = \langle CV, CE \rangle$, 当用户属于社区表示为: $CV \subseteq V$, 社区用户间的关系表示为: $CE \subseteq E$;

根据社区间是否拥有共同用户,将社区分为非重叠社区和重叠社区;

(1) 非重叠社区:如果社区中的所有用户均不属于另一社区,则该社区与另一社区为非重叠社区,非重叠社区表示为: $NC = \langle SV, CE \rangle$, $SV \subseteq V$, $CE \subseteq E$;

(2) 重叠社区:如果一个社区中的至少有一个用户同时也属于另一社区,则两个社区为重叠社区,重叠社区表示为: $OC = \langle MCV, SCV, CE \rangle$, $MCV \subseteq V$, $SCV \subseteq V$, $CE \subseteq E$;

根据上述的非重叠社区和重叠社区,得出下列属性:

- $\forall v_i, v_j \in C.CV \wedge \exists e(v_i, v_j) \rightarrow \forall e(v_i, v_j) \in C.CE$;
- $\forall e(v_i, v_j) \in C.CE \wedge \rightarrow \forall v_i, v_j \in C.CV$;
- $\exists OC \wedge OC.MCV \neq \emptyset \wedge OC.SCV \neq \emptyset \rightarrow OC.MCV \cap OC.SCV = \emptyset$;
- $\exists NC_i \neq \emptyset \wedge \exists NC_j \neq \emptyset \wedge NC_i \neq NC_j \rightarrow NC_i \cap NC_j = \emptyset$.

社区影响力评估算法:根据给出的社区模型以及与它相关的社交网络的信息来评估社区的影响力,社区影响力评估算法表示为: $CI(C_i) = f(G, C_i)$, $CI(C_i)$ 表示社区 C_i 的影响力, $f(G, C_i)$ 表示该模型是基于 G 和 C_i ;

用户影响力:每个用户在其所在的社区内存在一定的影响力,而用户的影响力是通过用户发布的信息影响的人数来衡量的,用户发布的信息影响的人数即转发或分享该信息的总人数,用户发布的信息,不仅会影响到社区内的用户,还会影响社区外的用户, UI 的取值在 0 到 1 之间, UI 的值越大,表示用户的影响力越大;

社区影响力:当兴趣爱好相同的人聚集形成社区时,社区就会形成一定的影响力,社区影响力是对社区内用户影响力的综合;

如述,社交网络是由多个重叠或非重叠社区组成,社区是由用户组成,每个用户是一个用户节点,每个用户节点可能属于一个或多个社区,为了便于分析,作出如下定义:

每一条主题信息 j 在社交网络中传播,都会形成一个图 g_j , 社交网络中,有 n 个主题,

且主题的数量随着时间的推移,在不断的增加,所以需要不停地对信息进行更新,主题 S 的集合表示为: $S = \{s_1, s_2, s_3, \dots, s_n\}$, 主题 S 传播过程中形成的图的集合表示为: $g = \{g_1, g_2, g_3, \dots, g_n\}$;

平均用户传播主题信息的意愿:每条主题信息由发布者开始传播,用户在面对该主题信息时,决定是否愿意传播下去,如果用户转发或分享,则说明用户愿意对该主题信息进行传播;反之,则不愿意,将用户对多条主题信息的反应综合,形成平均用户传播信息的意愿;

社区对社区内节点的影响:社区对社区内节点的影响是社区内的用户受到社区内其他用户的影响;

社区对社区外节点的影响力:社区内的用户发布的信息,在影响社区内的用户的同时,也可能对社区外的节点产生影响,将社区看作一个整体,社区对社区外节点的影响力即看作是节点对节点的影响力;

社区对社区的影响力:社交网络中的社区与社区之间由于用户之间的交互,使得社区与社区之间存在影响力。

3. 根据权利要求 1 所述的社交网络社区影响力评估算法,其特征在于,所述第二步基础阶段,计算出用户传播主题信息的意愿、社区用户综合影响、社区用户规模,统计出社区用户数量,再综合计算出社区影响力,包括步骤:

(1) 分别统计用户接收和传播主题信息的次数,计算出平均用户传播主题信息的意愿;

(2) 根据社区内平均用户传播主题信息的意愿及用户影响力,计算出社区内所有用户的综合影响;统计社区内的用户数量,得出社区用户数量;根据用户间的关系,计算出社区的用户规模;

(3) 结合上述计算出的社区内所有用户的综合影响、社区的用户规模,以及统计出的社区的用户数量,计算出社区影响力。

4. 根据权利要求 1 所述的社交网络社区影响力评估算法,其特征在于,所述第三步延伸阶段,计算出社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,包括步骤:

(1) 社区内节点受其他用户发布主题信息的影响情况计算出社区对社区内节点的影响力;

(2) 根据计算出的社区内用户的平均用户传播主题信息的意愿及社区外节点受社区内用户发布主题信息的影响情况计算出社区对社区外节点的影响力;

(3) 根据计算出的社区内的用户的平均用户传播主题信息的意愿及社区的重叠与否,分别计算出非重叠社区的社区对社区的影响力及重叠社区的社区对社区的影响力。

5. 根据权利要求 3 所述的社交网络社区影响力评估算法,其特征在于,所述 (1) 分别统计用户接收和传播主题信息的次数,计算出平均用户传播主题信息的意愿,还包括步骤:

用户在接收到主题信息时,可能传播下去,也可能不传播下去,同时,用户接收到该主题信息的途径可能有多个,可能从一个或多个用户那接收到该信息,用户 u 对主题信息 j 的标记为:

$$M_{us_j} = \begin{cases} 0, & \text{表明该用户从未接收过该信息} \\ 1, & \text{表明用户已接收过该信息} \end{cases} \quad (1)$$

用户 u 接收主题信息的次数, 初始值为 0, 若此时用户 u 对主题信息 j 的标记为 1, 则用户 u 接收主题信息的次数加 1, 虽然用户 u 可能从多个用户那接收到主题信息 j , 但是, 用户 u 对主题信息 j 的累加只进行一次, 用户 u 对主题信息 j 的传播, 是通过观察图 g_j ,

$$pro_{us_j} = \begin{cases} 0, & \text{if } outdegree(u)=0 \\ 1, & \text{if } outdegree(u) > 0 \end{cases} \quad (2)$$

其中, pro_{us_j} 为 0 时, 表明用户未对该主题信息进行传播, pro_{us_j} 为 1 时, 表明用户对该主题信息进行了传播; 当用户 u 的出度大于 0 时, 说明用户 u 对主题信息 j 进行了传播, 平均用户传播主题信息的意愿, 其计算方法为:

$$AP(u) = \frac{\theta \times \frac{Pro_u}{AcS_u} + (1-\theta) \times \sum_{s \in S(u), g \in g(u)} \left(\sum_{v \in A(u)} AP(v) \times w(v) + \sum_{b \in B(u)} AP(b) \times w(b) \right)}{num_s} \quad (3)$$

$$AcS_u = \sum_{s \in S} M_{us_j}, \quad Pro_u = \sum_{s \in S} pro_{us_j}, \quad w(v) = \frac{1}{outdegree(v)}, \quad w(b) = \alpha \times \frac{1}{outdegree(b)}$$

$$\alpha = \frac{t_{out \rightarrow in}}{t_{ac} \times num_{C_i}}$$

$AP(u)$ 为用户 u 平均传播主题信息的意愿; θ 为综合因子, 且 $\theta \in [0, 1]$; Pro_u 为用户传播主题信息的总次数; AcS_u 为用户接收主题信息的总次数; 集合 $A(u)$ 为社区内指向用户 u 的节点, $v \in A(u)$ 为节点 v 指向 u 被存放于集合 $A(u)$; 集合 $B(u)$ 为社区外指向用户 u 的节点, $b \in B(u)$ 为社区外节点 b 指向 u 被存放于集合; 每条主题信息的传播, 都会形成一个新的图, $S(u)$, $g(u)$ 分别为节点 u 发布的主题信息的集合, 发布主题信息后形成的主题信息传播图的集合, 对每个新图的指向 u 的节点和社区外节点对节点 u 产生的影响进行计算综合; $w(v)$, $w(b)$ 分别为节点 v 和节点 b 的权重, 而 $w(v)$ 和 $w(b)$ 是通过节点的出度来衡量的; α 为社区外节点对社区内节点的影响系数; $t_{out \rightarrow in}$ 表示社区外节点指向社区内节点的次数, t_{ac} 表示社区内节点指向社区外节点的次数, num_{C_i} 表示社区内用户的数量; 假设 $AP(u)$ 的初值为 1;

以上对平均用户传播主题信息的意愿的计算是通过将用户自身传播信息的意愿和社区内以及社区外节点对用户的影响进行综合, 得出客观的平均用户传播信息的意愿。

6. 根据权利要求 3 所述的社交网络社区影响力评估, 其特征在于, 所述 (2) 根据社区内平均用户传播主题信息的意愿及用户影响力, 计算出社区内所有用户的综合影响, 统计社区内的用户数量, 得出社区用户数量, 根据用户间的关系, 计算出社区的用户规模, 还包括步骤:

用户发布的主题信息, 将会对社区内和社区外的用户产生影响, 对用户影响力, 其计算方法为:

$$UI(u) = \xi \sum_{r \in S(u)} (sum_{in,r} + \beta \times sum_{out,r}) \quad (4)$$

$$\beta = \frac{t_{ac-c}}{t_{ac-c} \times num_{out}}, \quad \xi = \frac{1}{num_{C_i}^{sum_{S(u)}}}$$

其中, $UI(u)$ 为用户 u 的影响力; $S(u)$ 为用户 u 发布的主题信息的集合; $sum_{in,r}$ 为 r 条主题信息影响的社区内人数; $sum_{out,r}$ 为 r 条主题信息影响的社区外人数; β 为社区内节点对社区外节点的影响系数; t_{ac-c} 表示社区内用户指向社区外用户的次数, t_{ac} 表示社区外用户接收社区内主题信息的次数, num_{out} 表示社区外用户的数量, ξ 为用户影响力的影响因子, num_{C_i} 表示社区内用户的数量, $num_{S(u)}$ 表示用户 u 发布的主题信息的次数; 如果用户从未发布过信息, 就设用户的影响力为 0.01;

根据用户发布的所有主题信息, 对社区内和社区外产生影响的总人数的综合得出用户影响力;

相对于社区外的用户而言, 社区又被看作一个整体, 则整体就有其特有的影响力, 包括如下因素:

a. 社区内用户的综合影响, 其计算方法为:

$$UII(C_i) = \int_0^t \left(\sum_{u \in I(u)} AP(u) \times UI(u) \right) dt \quad (5)$$

$UII(C_i)$ 是社区 C_i 的用户综合影响; $I(u)$ 是社区内用户的集合; $AP(u)$ 是用户 u 的平均传播主题信息的意愿; $UI(u)$ 是用户 u 的用户影响力;

b. 社区用户的数量是社区内存在的用户的数量, 在计算社区影响力时起着至关重要的作用, 如果不考虑该因素, 则不同的社区可能有相同的社区影响力, 这是不合理的;

c. 社区用户规模是用户间的关系, 通过用户的出度来确定, 社区用户规模, 其计算方法为:

$$US(C_i) = \sum_{u \in I(u)} outdegree(u) \quad (6)$$

$US(C_i)$ 是社区 C_i 的用户规模, $outdegree(u)$ 是用户 u 的出度。

7. 根据权利要求 3 所述的社交网络社区影响力评估算法, 其特征在于, 所述 (3) 结合上述计算出的社区内所有用户的综合影响、社区的用户规模, 以及统计出的社区的用户数量, 计算出社区影响力, 还包括步骤:

综合上述三个因素, 得出社区影响力, 社区影响力, 其计算方法为:

$$CI(C_i) = M \times UII(C_i) + N \times NU + (1 - M - N) \times US(C_i) \quad (7)$$

其中, M 和 N 是影响因素, 因为在不同的社区影响力评估算法中各个影响因素的重要性不同, M 和 N 可根据实际情况进行取值。

8. 根据权利要求 4 所述的社交网络社区影响力评估算法, 其特征在于, 所述 (1) 社区内节点受其他用户发布主题信息的影响情况计算出社区对社区内节点的影响力, 还包括步骤:

社区对社区内的用户的影响力, 其计算方法为:

$$ICNI(h) = \sum_{u \in (I(u) - \{h\})} AP(u) \times \left(\sum_{s \in S(u), g \in g(u)} influ \right) \quad (8)$$

假设节点 h 在社区内, $I(u)$ 是社区内节点的集合, $I(u) - \{h\}$ 表示节点 h 不在集合内; $influ$ 用来标记节点 h 是否被用户 u 发布的影响影响, 如果节点 h 被主题信息 s 影响, $influ$

则是 1, 否则 influ 是 0;

社区对社区内节点的影响力的计算是将社区外节点受到社区内节点发布信息的影响和社区内用户影响力综合。

9. 根据权利要求 4 所述的社交网络社区影响力评估算法, 其特征在于, 所述 (2) 根据计算出的社区内用户的平均用户传播主题信息的意愿及社区外节点受社区内用户发布主题信息的影响情况计算出社区对社区外节点的影响力, 还包括步骤:

由于社区是由许多用户组成的群体, 这个群体势必会对社区外的节点产生影响力, 社区对社区外节点的影响力, 其计算方法如下:

假设节点 q 为社区外节点,

$$ICTN(q) = \beta \times \sum_{u \in I(u)} AP(u) \times \left(\sum_{s \in S(u), g \in g(u)} inf_s \right) \quad (9)$$

其中, $ICTN(q)$ 为社区对社区外节点 q 的影响力; β 为社区内节点对社区外节点的影响系数; $C(u)$ 为社区内用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(u), g(u)$ 分别为节点 u 发布的主题信息的集合, 发布主题信息后形成的主题信息传播图的集合; inf_s 为用户 u 发布的主题信息对节点 q 的影响, 若主题信息 s 对用户产生影响, 则 inf_s 为 1, 反之, 则 inf_s 为 0;

上述社区对社区外节点的影响力, 根据社区内所有节点发布的主题信息, 对社区外节点产生影响和社区内节点的平均传播主题信息的意愿综合得出。

10. 根据权利要求 4 所述的社交网络社区影响力评估算法, 其特征在于, 所述 (3) 根据计算出的社区内的用户的平均用户传播主题信息的意愿及社区的重叠与否, 分别计算出非重叠社区的社区对社区的影响力和重叠社区的社区对社区的影响力, 还包括步骤:

a. 非重叠社区

社区 G_A 和社区 G_B 是重叠社区, 对 G_A 对社区 G_B 的影响力, 其计算方法为:

$$ICTC(C_A \rightarrow C_B) = \beta \times \sum_{\substack{u \in D(C_A), v \in H(C_B) \\ s \in S(C_A), g \in g(C_A)}} AP(u) \times sum_{u \rightarrow v} \quad (10)$$

其中, $ICTC(G_A \rightarrow G_B)$ 为社区 G_A 对社区 G_B 的影响力; β 为社区内节点对社区外节点的影响系数; $D(G_A)$ 和 $H(G_B)$ 分别为社区 G_A 和社区 G_B 中用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(G_A)$ 和 $g(G_A)$ 分别为社区 G_A 中所有用户发布的主题信息的集合和所有用户发布的主题信息形成的主题信息传播图的集合; $sum_{u \rightarrow v}$ 为社区 G_A 中用户 u 发布的所有主题信息影响的社区 G_B 的人数;

上述社区对社区的影响力的计算是通过统计社区 G_A 发布的所有主题信息对社区 G_B 产生影响的人数, 综合得出了社区间的影响力;

b. 重叠社区, 其计算方法为:

$$ICTC(C_A \rightarrow C_B) = \beta \times \sum_{\substack{u \in (C_A - C_A \cap C_B), v \in (C_B - C_A \cap C_B) \\ s \in S(C_A - C_A \cap C_B), g \in g(C_A - C_A \cap C_B)}} AP(u) \times sum_{u \rightarrow v} + \sum_{\substack{u \in (C_A \cap C_B), v \in C_B \\ s \in S(C_A \cap C_B), g \in g(C_A \cap C_B)}} AP(u) \times sum_{u \rightarrow v} \quad (11)$$

其中, $ICTC(G_A \rightarrow G_B)$ 为社区 G_A 对社区 G_B 的影响力; $G_A - G_A \cap G_B$ 为社区 G_A 中减去与社区 G_B 重叠的用户所剩余的用户的集合; $G_B - G_A \cap G_B$ 为社区 G_B 中去掉与社区 G_A 重叠的用户所剩余的用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(G(A) - G(A) \cap G(B))$ 和 $g(G(A) - G(A) \cap G(B))$ 分别为社区 G_A 中去掉与社区 G_B 重合的用户, 所剩下的用户发布的

主题信息的集合和剩下的用户发布的主题信息形成的主题信息传播图的集合; $\text{sum}_{u \rightarrow v}$ 为社区 G_A 中去掉与社区 G_B 重合的用户, 所剩下的用户 u 发布的所有主题信息影响的社区 G_B 中去掉与社区 G_A 重合的用户的总人数; $G(A) \cap G(B)$ 为社区 G_A 与社区 G_B 重叠的用户的集合; $S(G(A) \cap G(B))$ 和 $g(G(A) \cap G(B))$ 分别为社区 G_A 与社区 G_B 重合的用户, 重合的用户发布的主题信息的集合和重合的用户发布的主题信息形成的主题信息传播图的集合; $\text{sum}_{in, u}$ 为节点 u 发布的主题信息影响的社区 G_B 内人数;

上述对重叠社区对社区的影响力的计算是通过将社区 G_A 中去掉与社区 G_B 重叠的用户, 剩下的用户对社区 G_B 中去掉与社区 G_A 重叠的用户产生的影响以及社区 G_A 和社区 G_B 重叠的用户, 对社区 G_B 产生的影响的综合, 得出社区对社区的影响力。

社交网络社区影响力评估算法

技术领域

[0001] 本发明涉及社交网络技术领域,更具体的说是社交网络社区影响力评估算法。

背景技术

[0002] 近年来,随着人们对获取信息和传递信息的需求越来越大,而社交网络又是满足人们这些需求的良好媒介,从而使得社交网络得到了迅速的发展,社交网络是由每个相对独立的用户组成的一个复杂的虚拟社会,它是当前互联网最热门的网络应用,深刻的改变了互联网信息传播的规则和方式,目前具有代表性的社交网络有:人人网,开心网,Facebook, Twitter 等。

[0003] 与真实的社会网络相同,社区是方便用户交流的重要元素,例如兴趣分享、主题讨论等。因此,社区将会对用户的观点、感受、决定产生影响,并且影响力越大的社区对社交网络的影响越大。因此,社区影响力的评估是对社交网络研究的关键挑战。社区影响力的研究有利于更好地管控网络带来的负面因素以及更好地利用网络带来的正面效应。如网络舆情监控中如何有效地防止各种有害信息在网络中迅速传播;商业领域的“口碑效应”和“病毒式营销”引出的影响力最大化问题等。

[0004] 目前,国内外许多学者都对社交网络的社区挖掘及用户影响力进行了研究,但很少对社区的影响力进行探讨。社区影响力是对社区内的用户以及用户之间的交互形成的相互影响的一种综合,也是对社区的质量等进行评估的一种方法。社区是社交网络中的一个重要的组成元素,它的作用是可以找到更多兴趣相同的朋友,而对社区的影响力进行评估,对稳定和发展社区有着至关重要的作用。近年来,对社区的研究已经开展了很多,潘磊等人基于边社区的思想,考虑每条边在网络中所归属的局部边社区,提出了基于局部信息的边社区挖掘算法;黄发良等人在给出的社交网络及社区挖掘的形成定义的基础上,提出了一个基于启发式策略的在线社交网络社区挖掘框架;杨长春等人研究微博社区中博主的交互行为,构建微博社区网络,建立评价指标,评价微博博主在网络中的影响力;肖宇等人研究特定区域的人群在社交网络中的信息传播行为,对用户的影响力进行评估,提出了社交网络中用户区域影响力评估算法;张波等人提出了基于信任的社会网络最大化影响节点的计算方法,该方法引入信任这一元素,考虑信任对节点的重要影响,将节点之间的信任度和影响值结合,得出综合影响值,再根据综合影响值得出社会网络影响最大化节点,对节点影响的计算是由节点的出度和节点间边的权重组成的。上述研究都没有对社区影响力进行评估,但是为社交网络中社区影响力的评估提供了一定的基础。

发明内容

[0005] 本发明的目的针对现有技术中存在的缺失和不足,提出社交网络社区影响力评估算法。

[0006] 本发明在现有的社区以及对区域内的用户影响力的评估的基础上,提出了一种社区影响力评估算法。计算出社区影响力后还作了相关方面的延伸,由三个因素组成:社区对

社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力。本发明,提出一种社区影响力评估算法,其基本流程(如图1所示):

[0007] 首先,通过分析用户以往的接收和传播主题信息的情况,得出平均用户传播主题信息的意愿,对用户传播主题信息及发布信息等的情况计算出用户影响力,根据社区内所有用户的综合影响、社区用户数量、社区用户规模计算出社区影响力。

[0008] 其次,根据计算出的社区内用户的平均用户传播主题信息的意愿及社区内节点受其他用户发布主题信息的影响情况计算出社区对社区内节点的影响力;根据计算出的社区内用户的平均用户传播主题信息的意愿及社区外节点受社区内用户发布主题信息的影响情况计算出社区对社区外节点的影响力;根据计算出的社区内的用户的平均用户传播主题信息的意愿及社区的重叠与否,分别计算出非重叠社区的社区对社区的影响力和重叠社区的社区对社区的影响力。

[0009] 最后,将计算出的社区影响力、社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,用于评估各社区间的影响力、节点对社区的重要性以及社区对节点作用。

[0010] 本发明社交网络社区影响力评估算法,具体步骤为:

[0011] A. 定义社交网络与社区影响力相关的概念及建模。

[0012] B. 基础阶段:(1)分别统计用户接收和传播主题信息的次数,计算出平均用户传播主题信息的意愿(2)根据社区内平均用户传播主题信息的意愿及用户影响力,计算出社区内所有用户的综合影响;统计社区内的用户数量,得出社区用户数量;根据用户间的关系,计算出社区的用户规模(3)结合上述计算出的社区内所有用户的综合影响、社区的用户数量,以及统计出的社区的用户规模,计算出社区影响力。

[0013] C. 延伸阶段:(1)社区内节点受其他用户发布主题信息的影响情况计算出社区对社区内节点的影响力(2)根据计算出的社区内用户的平均用户传播主题信息的意愿及社区外节点受社区内用户发布主题信息的影响情况计算出社区对社区外节点的影响力(3)根据计算出的社区内的用户的平均用户传播主题信息的意愿及社区的重叠与否,分别计算出非重叠社区的社区对社区的影响力和重叠社区的社区对社区的影响力。

[0014] D. 将计算出的社区影响力、社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,用于评估各社区间的影响力、节点对社区的重要性以及社区对节点作用。

[0015] 其中,所述步骤A的具体过程为:

[0016] 社交网络模型:社交网络形式化为一个二分图, $G = (V, E)$, V 代表社交网络中的用户集合, E 用来描述用户间的关系的集合。

[0017] 社区模型:社区是由拥有相同兴趣爱好的用户组成,并且是社交网络的一个子图, $C = \langle CV, CE \rangle$,当用户属于社区可以表示为: $CV \subseteq V$,社区用户间的关系可以表示为: $CE \subseteq E$ 。

[0018] 根据社区间是否拥有共同用户,将社区分为非重叠社区和重叠社区

[0019] (1) 非重叠社区:如果社区中的所有用户均不属于另一社区,则该社区与另一社区为非重叠社区,非重叠社区表示为: $NC = \langle SV, CE \rangle$, $SV \subseteq V$, $CE \subseteq E$ 。

[0020] (2) 重叠社区:如果一个社区中的至少有一个用户同时也属于另一社区,则两个

社区为重叠社区,重叠社区表示为: $OC = \langle MCV, SCV, CE \rangle$, $MCV \subseteq V$, $SCV \subseteq V$, $CE \subseteq E$ 。

[0021] 根据上述的非重叠社区和重叠社区,得出下列属性:

[0022] ● $\forall v_i, v_j \in C.CV \wedge \exists e(v_i, v_j) \rightarrow \forall e(v_i, v_j) \in C.CE$;

[0023] ● $\forall e(v_i, v_j) \in C.CE \wedge \rightarrow \forall v_i, v_j \in C.CV$;

[0024] ● $\exists OC \wedge OC.MCV \neq \emptyset \wedge OC.SCV \neq \emptyset \rightarrow OC.MCV \cap OC.SCV = \emptyset$;

[0025] ● $\exists NC_i \neq \emptyset \wedge \exists NC_j \neq \emptyset \wedge NC_i \neq NC_j \rightarrow NC_i \cap NC_j = \emptyset$ 。

[0026] 社区影响力评估算法:根据给出的社区模型以及与其相关的社交网络的信息来评估社区的影响力,社区影响力评估算法表示为: $CI(C_i) = f(G, C_i)$, $CI(C_i)$ 表示社区 C_i 的影响力, $f(G, C_i)$ 表示该模型是基于 G 和 C_i 。

[0027] 用户影响力:每个用户在其所在的社区内存在一定的影响力,而用户的影响力是通过用户发布的信息影响的人数来衡量的,用户发布的信息影响的人数即转发或分享该信息的总人数,用户发布的信息,不仅会影响到社区内的用户,还会影响社区外的用户, UI 的取值在 0 到 1 之间, UI 的值越大,表示用户的影响力越大。

[0028] 社区影响力:当兴趣爱好相同的人聚集形成社区时,社区就会形成一定的影响力,社区影响力是对社区内用户影响力的综合;

[0029] 如述,社交网络是由多个重叠或非重叠社区组成,社区是由用户组成,每个用户是一个用户节点,每个用户节点可能属于一个或多个社区,为了便于分析,作出如下定义:

[0030] 每一条主题信息 j 在社交网络中传播,都会形成一个图 g_j ,社交网络中,有 n 个主题,且主题的数量随着时间的推移,在不断的增加,所以需要不停地对信息进行更新,主题 S 的集合表示为: $S = \{s_1, s_2, s_3, \dots, s_n\}$, 主题 S 传播过程中形成的图的集合表示为: $g = \{g_1, g_2, g_3, \dots, g_n\}$ 。

[0031] 平均用户传播主题信息的意愿:每条主题信息由发布者开始传播,用户在面对该主题信息时,决定是否愿意传播下去,如果用户转发或分享,则说明用户愿意对该主题信息进行传播;反之,则不愿意,将用户对多条主题信息的反应综合,形成平均用户传播信息的意愿。

[0032] 社区对社区内节点的影响:社区对社区内节点的影响是社区内的用户受到社区内其他用户的影响。

[0033] 社区对社区外节点的影响力:社区内的用户发布的信息,在影响社区内的用户的同时,也可能对社区外的节点产生影响,将社区看作一个整体,社区对社区外节点的影响力即看作是节点对节点的影响力;

[0034] 社区对社区的影响力:社交网络中的社区与社区之间由于用户之间的交互,使得社区与社区之间存在影响力。

[0035] 其中,所述步骤 B 的具体过程为:

[0036] 用户在接收到主题信息时,可能传播下去,也可能不传播下去。同时,用户接收到该主题信息的途径可能有多个,可能从一个或多个用户那接收到该信息,用户 u 对主题信息 j 的标记为:

[0037]

$$M_{us_j} = \begin{cases} 0, & \text{表明该用户从未接收过该信息} \\ 1, & \text{表明用户已接收过该信息} \end{cases} \quad (1)$$

[0038] 用户 u 接收主题信息的次数, 初始值为 0, 若此时用户 u 对主题信息 j 的标记为 1, 则用户 u 接收主题信息的次数加 1。虽然用户 u 可能从多个用户那接收到主题信息 j , 但是, 用户 u 对主题信息 j 的累加只进行一次, 用户 u 对主题信息 j 的传播, 是通过观察图 g_j 。

$$[0039] \quad pro_{us_j} = \begin{cases} 0, & \text{if } outdegree(u)=0 \\ 1, & \text{if } outdegree(u) > 0 \end{cases} \quad (2)$$

[0040] 其中, pro_{us_j} 为 0 时, 表明用户未对该主题信息进行传播; pro_{us_j} 为 1 时, 表明用户对该主题信息进行传播。当用户 u 的出度大于 0 时, 说明用户 u 对主题信息 j 进行传播。平均用户传播主题信息的意愿, 作出了如下定义:

$$[0041] \quad AP(u) = \frac{\theta \times \frac{Pro_u}{AcS_u} + (1-\theta) \times \sum_{s \in S(u), g \in g(u)} \left(\sum_{v \in A(u)} AP(v) \times w(v) + \sum_{b \in B(u)} AP(b) \times w(b) \right)}{num_s} \quad (3)$$

$$[0042] \quad AcS_u = \sum_{s \in S} M_{us_j}, \quad Pro_u = \sum_{s \in S} pro_{us_j}, \quad w(v) = \frac{1}{outdegree(v)}, \quad w(b) = \alpha \times \frac{1}{outdegree(b)}$$

$$[0043] \quad \alpha = \frac{t_{out \rightarrow in}}{t_{ac} \times num_{C_i}}$$

[0044] $AP(u)$ 为用户 u 平均传播主题信息的意愿; θ 为综合因子, 且 $\theta \in [0, 1]$; Pro_u 为用户传播主题信息的总次数; AcS_u 为用户接收主题信息的总次数; 集合 $A(u)$ 为社区内指向用户 u 的节点, $v \in A(u)$ 为节点 v 指向 u 被存放于集合 $A(u)$; 集合 $B(u)$ 为社区外指向用户 u 的节点, $b \in B(u)$ 为社区外节点 b 指向 u 被存放于集合; 每条主题信息的传播, 都会形成一个新的图, $S(u)$, $g(u)$ 分别为为节点 u 发布的主题信息的集合, 发布主题信息后形成的主题信息传播图的集合, 对每个新图的指向 u 的节点和社区外节点对节点 u 产生的影响进行计算综合; $w(v)$, $w(b)$ 分别为节点 v 和节点 b 的权重, 而 $w(v)$ 和 $w(b)$ 是通过节点的出度来衡量的; α 为社区外节点对社区内节点的影响系数。 $t_{out \rightarrow in}$ 表示社区外节点指向社区内节点的次数, t_{ac} 表示社区内节点指向社区外节点的次数, num_{C_i} 表示社区内用户的数量。这里, 假设 $AP(u)$ 的初值为 1。

[0045] 以上对平均用户传播主题信息的意愿的计算是通过将用户自身传播信息的意愿和社区内以及社区外节点对用户的影响进行综合, 得出客观的平均用户传播信息的意愿。

[0046] (1) 用户影响力的计算

[0047] 用户发布的主题信息, 将会对社区内和社区外的用户产生影响, 对用户影响力, 作出了如下定义:

$$[0048] \quad UI(u) = \xi \sum_{r \in S(u)} (sum_{in,r} + \beta \times sum_{out,r}) \quad (4)$$

$$[0049] \quad \beta = \frac{t_{ac-e}}{t_{ac} \times num_{out}}, \quad \xi = \frac{1}{num_{C_i} \times num_{S(0)}}$$

[0050] 其中, $UI(u)$ 为用户 u 的影响力; $S(u)$ 为用户 u 发布的主题信息的集合; $sum_{in,r}$ 为

r 条主题信息影响的社区内人数 ; $\text{sum}_{\text{out}, r}$ 为 r 条主题信息影响的社区外人数 ; β 为社区内节点对社区外节点的影响系数 ; $t_{\text{ac-c}}$ 表示社区内用户指向社区外用户的次数, t_{ac} 表示社区外用户接收社区内主题信息的次数, num_{out} 表示社区外用户的数量, ξ 为用户影响力的影响因子, num_{C_i} 表示社区内用户的数量, $\text{num}_S(u)$ 表示用户 u 发布的主题信息的次数。如果用户从未发布过信息, 假设用户的影响力为 0.01。

[0051] 根据用户发布的所有主题信息, 对社区内和社区外产生影响的总人数的综合得出用户影响力。

[0052] (2) 社区影响力计算

[0053] 相对于社区外的用户而言, 社区又被看作一个整体, 则整体就有其特有的影响力, 包括如下因素 :

[0054] ①社区内用户的综合影响

[0055] 社区内用户的综合影响, 定义为如下 :

$$[0056] \quad UII(C_i) = \int_0^1 \left(\sum_{u \in I(u)} AP(u) \times UI(u) \right) dt \quad (5)$$

[0057] $UII(C_i)$ 是社区 C_i 的用户综合影响 ; $I(u)$ 是社区内用户的集合 ; $AP(u)$ 是用户 u 的平均传播主题信息的意愿 ; $UI(u)$ 是用户 u 的用户影响力。

[0058] ②社区用户的数量

[0059] 社区用户的数量是社区内存在的用户的数量, 在计算社区影响力时起着至关重要的作用, 如果不考虑该因素, 则不同的社区可能有相同的社区影响力, 这是不合理的。

[0060] ③社区用户规模

[0061] 社区用户规模是用户间的关系, 通过用户的出度来确定, 社区用户规模, 定义为如下 :

$$[0062] \quad US(C_i) = \sum_{u \in I(u)} \text{outdegree}(u) \quad (6)$$

[0063] $US(C_i)$ 是社区 C_i 的用户规模。 $\text{outdegree}(u)$ 是用户 u 的出度。

[0064] 综合上述三个因素, 得出社区影响力, 社区影响力定义如下 :

$$[0065] \quad CI(C_i) = M \times UII(C_i) + N \times NU + (1 - M - N) \times US(C_i) \quad (7)$$

[0066] 其中, M 和 N 是影响因素, 因为在不同的社区影响力评估算法中各个影响因素的重要性不同, M 和 N 可根据实际情况进行取值。

[0067] 其中, 所述步骤 C 的具体过程为 :

[0068] (1) 社区对社区内的用户的影响力计算

[0069] 社区对社区内的用户的影响力定义如下 :

$$[0070] \quad ICNI(h) = \sum_{u \in (I(u) - \{h\})} AP(u) \times \left(\sum_{s \in S(u), g \in g(u)} \text{influ} \right) \quad (8)$$

[0071] 假设节点 h 在社区内。 $I(u)$ 是社区内节点的集合, $I(u) - \{h\}$ 表示节点 h 不在集合内 ; influ 用来标记节点 h 是否被用户 u 发布的影响影响, 如果节点 h 被主题信息 s 影响, influ 则是 1, 否则 influ 是 0。

[0072] 社区对社区内节点的影响力的计算是将社区外节点受到社区内节点发布信息的影响和社区内用户影响力综合。

[0073] (2) 社区对社区外用户的影响力的计算

[0074] 由于社区是由许多用户组成的群体,这个群体势必会对社区外的节点产生影响,社区对社区外节点的影响力,作出了如下定义:

[0075] 假设节点 q 为社区外节点。

$$[0076] \quad ICTN(q) = \beta \times \sum_{u \in I(u)} AP(u) \times \left(\sum_{s \in S(u), g \in g(u)} inf_s \right) \quad (9)$$

[0077] 其中, $ICTN(q)$ 为社区对社区外节点 q 的影响力; β 为社区内节点对社区外节点的影响系数; $C(u)$ 为社区内用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(u), g(u)$ 分别为为节点 u 发布的主题信息的集合,发布主题信息后形成的主题信息传播图的集合; inf_s 为用户 u 发布的主题信息对节点 q 的影响,若主题信息 s 对用户产生影响,则 inf_s 为 1,反之,则 inf_s 为 0。

[0078] 上述社区对社区外节点的影响力,根据社区内所有节点发布的主题信息,对社区外节点产生影响和社区内节点的平均传播主题信息的意愿综合得出。

[0079] (3) 社区对社区影响力的计算

[0080] a. 非重叠社区

[0081] 社区 G_A 和社区 G_B 是重叠社区,对 G_A 对社区 G_B 的影响力,作出了如下定义:

$$[0082] \quad ICTC(C_A \rightarrow C_B) = \beta \times \sum_{\substack{u \in D(C_A), v \in H(C_B) \\ s \in S(C_A), g \in g(C_A)}} AP(u) \times sum_{u \rightarrow v} \quad (10)$$

[0083] 其中, $ICTC(G_A \rightarrow G_B)$ 为社区 G_A 对社区 G_B 的影响力; β 为社区内节点对社区外节点的影响系数; $D(G_A)$ 和 $H(G_B)$ 分别为社区 G_A 和社区 G_B 中用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(G_A)$ 和 $g(G_A)$ 分别为社区 G_A 中所有用户发布的主题信息的集合和所有用户发布的主题信息形成的主题信息传播图的集合; $sum_{u \rightarrow v}$ 为社区 G_A 中用户 u 发布的所有主题信息影响的社区 G_B 的人数。

[0084] 上述社区对社区的影响力的计算是通过统计社区 G_A 发布的所有主题信息对社区 G_B 产生影响的人数,综合得出了社区间的影响力。

[0085] b. 重叠社区

$$[0086] \quad ICTC(C_A \rightarrow C_B) = \beta \times \sum_{\substack{u \in (C_A - C_A \cap C_B), v \in (C_B - C_A \cap C_B) \\ s \in S(C_A - C_A \cap C_B), g \in g(C_A - C_A \cap C_B)}} AP(u) \times sum_{u \rightarrow v} + \sum_{\substack{u \in (C_A \cap C_B), v \in C_B \\ s \in S(C_A \cap C_B), g \in g(C_A \cap C_B)}} AP(u) \times sum_{in, u} \quad (11)$$

[0087] 其中, $ICTC(G_A \rightarrow G_B)$ 为社区 G_A 对社区 G_B 的影响力; $G_A - G_A \cap G_B$ 为社区 G_A 中减去与社区 G_B 重叠的用户所剩余的用户的集合; $G_B - G_A \cap G_B$ 为社区 G_B 中去掉与社区 G_A 重叠的用户所剩余的用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(G(A) - G(A) \cap G(B))$ 和 $g(G(A) - G(A) \cap G(B))$ 分别为社区 G_A 中去掉与社区 G_B 重合的用户,所剩下的用户发布的主题信息的集合和剩下的用户发布的主题信息形成的主题信息传播图的集合; $sum_{u \rightarrow v}$ 为社区 G_A 中去掉与社区 G_B 重合的用户,所剩下的用户 u 发布的所有主题信息影响的社区 G_B 中去掉与社区 G_A 重合的用户的总人数; $G(A) \cap G(B)$ 为社区 G_A 与社区 G_B 重叠的用户的集合; $S(G(A) \cap G(B))$ 和 $g(G(A) \cap G(B))$ 分别为社区 G_A 与社区 G_B 重合的用户,重合的用户发布的主题信息的集合和重合的用户发布的主题信息形成的主题信息传播图的集合; $sum_{in, u}$ 为节点 u 发布的主题信息影响的社区 G_B 内人数。

[0088] 上述对重叠社区对社区的影响力的计算是通过将社区 G_A 中去掉与社区 G_B 重叠的

用户,剩下的用户对社区 G_B 中去掉与社区 G_A 重叠的用户产生的影响以及社区 G_A 和社区 G_B 重叠的用户,对社区 G_B 产生的影响的综合,得出社区对社区的影响力。

[0089] 其中,所述步骤 D 的具体过程为:

[0090] 将计算出的社区影响力、社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,用于评估各社区间的影响力、节点对社区的重要性以及社区对节点作用。

[0091] 本发明的主要贡献有:(1) 与社区影响力相关的概念定义;(2) 社区影响力的计算方法包括:用户影响力计算和社区影响力计算。(3) 社区影响力的相关方面的延伸。

附图说明

[0092] 图 1 为本发明社交网络社区影响力评估算法流程框图;

[0093] 图 2 为本发明一个实例的网络节点图。

[0094] 其中:a 用户关系图,b 用户 u_1 的主题信息传播图,c 用户 u_1 的主题信息传播图,d 用户 u_2 的主题信息传播图。

具体实施方式

[0095] 下面结合附图和实施例对本发明进一步描述

[0096] 本发明首先对社交网络进行定义:

[0097] 定义 1 社交网络:社交网络形式化为一个二分图, $G = (V, E)$, V 代表社交网络中的用户集合, E 用来描述用户间的关系的集合。

[0098] 定义 2 社区:社区是由拥有相同兴趣爱好的用户组成,为社交网络的一个子图, $C = \langle CV, CE \rangle$,当用户属于社区表示为: $CV \subseteq V$,社区用户间的关系表示为: $CE \subseteq E$ 。

[0099] 定义 3 根据社区间是否拥有共同用户,将社区分为非重叠社区和重叠社区。

[0100] (1) 非重叠社区:如果社区中的所有用户均不属于另一社区,则该社区与另一社区为非重叠社区,非重叠社区表示为: $NC = \langle SV, CE \rangle$, $SV \subseteq V$, $CE \subseteq E$ 。

[0101] (2) 重叠社区:如果一个社区中的至少有一个用户同时也属于另一社区,则两个社区为重叠社区,重叠社区表示为: $OC = \langle MCV, SCV, CE \rangle$, $MCV \subseteq V$, $SCV \subseteq V$, $CE \subseteq E$ 。

[0102] 定义 4 社区影响力评估算法:根据给出的社区模型以及与其相关的社交网络的信息来评估社区的影响力,社区影响力评估算法表示为: $CI(C_i) = f(G, C_i)$, $CI(C_i)$ 表示社区 C_i 的影响力, $f(G, C_i)$ 表示该模型是基于 G 和 C_i 。

[0103] 定义 5 用户影响力:每个用户在其所在的社区内存在一定的影响力,而用户的影响力是通过用户发布的信息影响的人数来衡量的,用户发布的信息影响的人数即转发或分享该信息的总人数,用户发布的信息,不仅会影响到社区内的用户,还会影响社区外的用户,UI 的取值在 0 到 1 之间,UI 的值越大,表示用户的影响力越大。

[0104] 定义 6 社区影响力:当兴趣爱好相同的人聚集形成社区时,社区就会形成一定的影响力,社区影响力是对社区内用户影响力的综合。

[0105] 定义 7 平均用户传播主题信息的意愿:每条主题信息由发布者开始传播,用户在面对该主题信息时,决定是否愿意传播下去,如果用户转发或分享,则说明用户愿意对该主

题信息进行传播;反之,则不愿意,将用户对多条主题信息的反应综合,形成平均用户传播信息的意愿。

[0106] 定义 8 社区对社区内节点的影响:社区对社区内节点的影响是社区内的用户受到社区内其他用户的影响。

[0107] 定义 9 社区对社区外节点的影响力:社区内的用户发布的信息,在影响社区内的用户的同时,也可能对社区外的节点产生影响,将社区看作一个整体,社区对社区外节点的影响力即看作是节点对节点的影响力。

[0108] 定义 10 社区对社区的影响力:社交网络中的社区与社区之间由于用户之间的交互,使得社区与社区之间存在影响力。

[0109] 下面,给出本发明社交网络社区影响力评估算法

[0110] 社交网络社区影响力评估算法,具体步骤为:

[0111] A. 定义社交网络与社区影响力相关的概念及建模。

[0112] B. 基础阶段:(1) 分别统计用户接收和传播主题信息的次数,计算出平均用户传播主题信息的意愿(2) 根据社区内平均用户传播主题信息的意愿及用户影响力,计算出社区内所有用户的综合影响;统计社区内的用户数量,得出社区用户数量;根据用户间的关系,计算出社区的用户规模(3) 结合上述计算出的社区内所有用户的综合影响及社区的用户数量及社区的用户规模,计算出社区影响力。

[0113] C. 延伸阶段:(1) 社区内节点受其他用户发布主题信息的影响情况计算出社区对社区内节点的影响力(2) 根据计算出的社区内用户的平均用户传播主题信息的意愿及社区外节点受社区内用户发布主题信息的影响情况计算出社区对社区外节点的影响力(3) 根据计算出的社区内的用户的平均用户传播主题信息的意愿及社区的重叠与否,分别计算出非重叠社区的社区对社区的影响力及重叠社区的社区对社区的影响力。

[0114] D. 将计算出的社区影响力、社区对社区内节点的影响力、社区对社区外节点的影响力、社区对社区的影响力,用于评估各社区间的影响力、节点对社区的重要性以及社区对节点作用。

[0115] (1) 平均用户传播主题信息的意愿的计算

$$[0116] \quad AP(u) = \frac{\theta \times \frac{Pro_u}{AcS_u} + (1-\theta) \times \sum_{s \in S(u), g \in g(u)} (\sum_{v \in A(u)} AP(v) \times w(v) + \sum_{b \in B(u)} AP(b) \times w(b))}{num_s} \quad (3)$$

$$[0117] \quad AcS_u = \sum_{s \in S} M_{us_j}, \quad Pro_u = \sum_{s \in S} pro_{us_j}, \quad w(v) = \frac{1}{outdegree(v)}, \quad w(b) = a \times \frac{1}{outdegree(b)}$$

$$[0118] \quad \alpha = \frac{t_{out \rightarrow in}}{t_{ac} \times num_{C_i}}$$

[0119] $AP(u)$ 为用户 u 平均传播主题信息的意愿; θ 为综合因子,且 $\theta \in [0, 1]$; Pro_u 为用户传播主题信息的总次数; AcS_u 为用户接收主题信息的总次数;集合 $A(u)$ 为社区内指向用户 u 的节点, $v \in A(u)$ 为节点 v 指向 u 被存放于集合 $A(u)$;集合 $B(u)$ 为社区外指向用户 u 的节点, $b \in B(u)$ 为社区外节点 b 指向 u 被存放于集合;每条主题信息的传播,都会形成一个新的图, $S(u)$, $g(u)$ 分别为为节点 u 发布的主题信息的集合,发布主题信息后形成的主题信息传播图的集合,对每个新图的指向 u 的节点和社区外节点对节点 u 产生的影响进行

计算综合 ; $w(v)$, $w(b)$ 分别为节点 v 和节点 b 的权重, 而 $w(v)$ 和 $w(b)$ 是通过节点的出度来衡量的 ; α 为社区外节点对社区内节点的影响系数。 $t_{out \rightarrow in}$ 表示社区外节点指向社区内节点的次数, t_{ac} 表示社区内节点指向社区外节点的次数, num_{C_i} 表示社区内用户的数量。假设 $AP(u)$ 的初值为 1。

[0120] 以上对平均用户传播主题信息的意愿的计算是通过将用户自身传播信息的意愿和社区内以及社区外节点对用户的影响进行综合, 得出客观的平均用户传播信息的意愿。

[0121] (2) 用户影响力的计算

[0122] 用户发布的主题信息, 将会对社区内和社区外的用户产生影响, 对用户影响力, 作出了如下定义 :

$$[0123] \quad UI(u) = \xi \sum_{r \in S(u)} (sum_{in,r} + \beta \times sum_{out,r}) \quad (4)$$

$$[0124] \quad \beta = \frac{t_{ac-c}}{t_{ac} \times num_{out}}, \quad \xi = \frac{1}{num_{C_i} num_{S(u)}}$$

[0125] 其中, $UI(u)$ 为用户 u 的影响力 ; $S(u)$ 为用户 u 发布的主题信息的集合 ; $sum_{in,r}$ 为 r 条主题信息影响的社区内人数 ; $sum_{out,r}$ 为 r 条主题信息影响的社区外人数 ; β 为社区内节点对社区外节点的影响系数 ; t_{ac-c} 表示社区内用户指向社区外用户的次数, t_{ac} 表示社区外用户接收社区内主题信息的次数, num_{out} 表示社区外用户的数量, ξ 为用户影响力的影响因子, num_{C_i} 表示社区内用户的数量, $num_{S(u)}$ 表示用户 u 发布的主题信息的次数。如果用户从未发布过信息, 假设用户的影响力为 0.01。

[0126] 根据用户发布的所有主题信息, 对社区内和社区外产生影响的总人数的综合得出用户影响力。

[0127] (3) 社区影响力计算

[0128] 相对于社区外的用户而言, 社区又被看作一个整体, 则整体就有其特有的影响力, 包括如下因素 :

[0129] ①社区内用户的综合影响

[0130] 社区内用户的综合影响, 定义为如下 :

$$[0131] \quad UII(C_i) = \int_0^t \left(\sum_{u \in I(u)} AP(u) \times UI(u) \right) dt \quad (5)$$

[0132] $UII(C_i)$ 是社区 C_i 的用户综合影响 ; $I(u)$ 是社区内用户的集合 ; $AP(u)$ 是用户 u 的平均传播主题信息的意愿 ; $UI(u)$ 是用户 u 的用户影响力。

[0133] ②社区用户的数量

[0134] 社区用户的数量是社区内存在的用户的数量, 在计算社区影响力时起着至关重要的作用, 如果不考虑该因素, 则不同的社区可能有相同的社区影响力, 这是不合理的。

[0135] ③社区用户规模

[0136] 社区用户规模是用户间的关系, 通过用户的出度来确定, 社区用户规模, 定义为如下 :

$$[0137] \quad US(C_i) = \sum_{u \in I(u)} outdegree(u) \quad (6)$$

[0138] $US(C_i)$ 是社区 C_i 的用户规模。 $outdegree(u)$ 是用户 u 的出度。

[0139] 综合上述三个因素,得出社区影响力,社区影响力定义如下:

$$[0140] \quad CI(C_i) = M \times UII(C_i) + N \times NU + (1-M-N) \times US(C_i) \quad (7)$$

[0141] 其中, M 和 N 是影响因素,因为在不同的社区影响力评估算法中各个影响因素的重要性不同, M 和 N 可根据实际情况进行取值。

[0142] (4) 社区对社区内的用户的影响力计算

[0143] 社区对社区内的用户的影响力定义如下:

$$[0144] \quad ICNI(h) = \sum_{u \in (I(u) - \{h\})} AP(u) \times \left(\sum_{s \in S(u), g \in g(u)} influ \right) \quad (8)$$

[0145] 假设节点 h 在社区内。 $I(u)$ 是社区内节点的集合, $I(u) - \{h\}$ 表示节点 h 不在集合内; $influ$ 用来标记节点 h 是否被用户 u 发布的影响影响,如果节点 h 被主题信息 s 影响, $influ$ 则是 1, 否则 $influ$ 是 0。

[0146] 社区对社区内节点的影响力的计算是将社区外节点受到社区内节点发布信息的影响和社区内用户影响力综合。

[0147] (5) 社区对社区外用户的影响力的计算

[0148] 由于社区是由许多用户组成的群体,这个群体势必会对社区外的节点产生影响,社区对社区外节点的影响力,作出了如下定义:

[0149] 假设节点 q 为社区外节点。

$$[0150] \quad ICTN(q) = \beta \times \sum_{u \in I(u)} AP(u) \times \left(\sum_{s \in S(u), g \in g(u)} inf_s \right) \quad (9)$$

[0151] 其中, $ICTN(q)$ 为社区对社区外节点 q 的影响力; β 为社区内节点对社区外节点的影响系数; $C(u)$ 为社区内用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(u), g(u)$ 分别为为节点 u 发布的主题信息的集合,发布主题信息后形成的主题信息传播图的集合; inf_s 为用户 u 发布的主题信息对节点 q 的影响,若主题信息 s 对用户产生影响,则 inf_s 为 1, 反之,则 inf_s 为 0。

[0152] 上述社区对社区外节点的影响力,根据社区内所有节点发布的主题信息,对社区外节点产生影响和社区内节点的平均传播主题信息的意愿综合得出。

[0153] (6) 社区对社区影响力的计算

[0154] a. 非重叠社区

[0155] 社区 G_A 和社区 G_B 是重叠社区,对 G_A 对社区 G_B 的影响力,作出了如下定义:

$$[0156] \quad ICTC(C_A \rightarrow C_B) = \beta \times \sum_{\substack{u \in D(C_A), v \in H(C_B) \\ s \in S(C_A), g \in g(C_A)}} AP(u) \times sum_{u \rightarrow v} \quad (10)$$

[0157] 其中, $ICTC(G_A \rightarrow G_B)$ 为社区 G_A 对社区 G_B 的影响力; β 为社区内节点对社区外节点的影响系数; $D(G_A)$ 和 $H(G_B)$ 分别为社区 G_A 和社区 G_B 中用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(G_A)$ 和 $g(G_A)$ 分别为社区 G_A 中所有用户发布的主题信息的集合和所有用户发布的主题信息形成的主题信息传播图的集合; $sum_{u \rightarrow v}$ 为社区 G_A 中用户 u 发布的所有主题信息影响的社区 G_B 的人数。

[0158] 上述社区对社区的影响力的计算是通过统计社区 G_A 发布的所有主题信息对社区 G_B 产生影响的人数,综合得出了社区间的影响力。

[0159] b. 重叠社区

$$[0160] \quad ICTC(C_A \rightarrow C_B) = \beta \times \sum_{\substack{u \in (C_A - C_A \cap C_B), v \in (C_B - C_A \cap C_B) \\ s \in S(C_A - C_A \cap C_B), g \in g(C_A - C_A \cap C_B)}} AP(u) \times sum_{u \rightarrow v} + \sum_{\substack{u \in (C_A \cap C_B), v \in C_B \\ s \in S(C_A \cap C_B), g \in g(C_A \cap C_B)}} AP(u) \times sum_{u, u} \quad (11)$$

[0161] 其中, $ICTC(G_A \rightarrow G_B)$ 为社区 G_A 对社区 G_B 的影响力; $G_A - G_A \cap G_B$ 为社区 G_A 中减去与社区 G_B 重叠的用户所剩余的用户的集合; $G_B - G_A \cap G_B$ 为社区 G_B 中去掉与社区 G_A 重叠的用户所剩余的用户的集合; $AP(u)$ 为用户 u 的平均用户传播主题信息的意愿; $S(G(A) - G(A) \cap G(B))$ 和 $g(G(A) - G(A) \cap G(B))$ 分别为社区 G_A 中去掉与社区 G_B 重合的用户, 所剩下的用户发布的主题信息的集合和剩下的用户发布的主题信息形成的主题信息传播图的集合; $sum_{u \rightarrow v}$ 为社区 G_A 中去掉与社区 G_B 重合的用户, 所剩下的用户 u 发布的所有主题信息影响的社区 G_B 中去掉与社区 G_A 重合的用户的总人数; $G(A) \cap G(B)$ 为社区 G_A 与社区 G_B 重叠的用户的集合; $S(G(A) \cap G(B))$ 和 $g(G(A) \cap G(B))$ 分别为社区 G_A 与社区 G_B 重合的用户, 重合的用户发布的主题信息的集合和重合的用户发布的主题信息形成的主题信息传播图的集合; $sum_{in, u}$ 为节点 u 发布的主题信息影响的社区 G_B 内人数。

[0162] 上述对重叠社区对社区的影响力的计算是通过将社区 G_A 中去掉与社区 G_B 重叠的用户, 剩下的用户对社区 G_B 中去掉与社区 G_A 重叠的用户产生的影响以及社区 G_A 和社区 G_B 重叠的用户, 对社区 G_B 产生的影响的综合, 得出社区对社区的影响力。

[0163] 实施例:

[0164] 给出如下实施例说明节点的社会网络影响计算 (如附图 2 所示) 其中, a 用户关系图, b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图。

[0165] 1. 平均用户传播主题信息的意愿的计算

[0166] 在附图 2 之 a 用户关系图中假设每个用户的初始 AP 值为 1, 即 $AP(u_1) = AP(u_2) = \dots = AP(v_3) = 1$. 从 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图,

d 用户 u_2 的主题信息传播图得出 $a = \frac{3}{7 \times 3} = \frac{1}{7}$, $\theta = 0.6$ 。

[0167] (1) 计算用户 u_1 的 AP: 从 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图 $num_s = 3$, $Pro_{u_1} = 1 + 1 + 0 = 2$, $AcS_{u_1} = 1 + 1 + 1 = 3$, 因为 b 用户 u_1 的主题信息传播图和 c 用户 u_1 的主题信息传播图是由于用户 u_1 形成的, 所以 b 用户 u_1 的主题信息传播图和 c 用户 u_1 的主题信息传播图在集合 $S(u_1)$ 中。从 b 用户 u_1 的主题信息传播图得出 $A(u_1) = \{u_4\}$, $B(u_1) = \{v_1\}$; $w(u_4) = \frac{1}{3}$, $w(v_1) = \frac{1}{7}$, 从 c 用户 u_1 的主题信息传播图

得出: $A(u_1) = \{u_2, u_5\}$, $B(u_1) = \Phi$; $w(u_2) = \frac{1}{2}$, $w(u_5) = 1$

$$[0168] \quad AP(u_1) = \frac{0.6 \times \frac{2}{3} + 0.4 \times ((1 \times \frac{1}{3} + 1 \times \frac{1}{7}) + (1 \times \frac{1}{2} + 1 \times 1))}{3} = \frac{25}{63}$$

[0169] (2) 计算用户 u_2 的 AP: 从 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, $num_s = 3$, $Pro_{u_2} = 0 + 1 + 1 = 2$, $AcS_{u_2} = 1 + 1 + 1 = 3$, 因为 d 用户 u_2 的主题信息传播图是由于用户 u_2 形成的, 所以 d 用户 u_2 的主题信息传播图在集合 $S(u_2)$ 中。

从 d 用户 u_2 的主题信息传播图得出 : $A(u_2) = \{u_5\}, B(u_1) = \phi; w(u_5) = \frac{1}{4}$

$$[0170] \quad AP(u_2) = \frac{0.6 \times \frac{2}{3} + 0.4 \times (1 \times \frac{1}{4})}{3} = \frac{1}{6}$$

[0171] (3) 计算用户 u_3 的 AP: 从 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, $num_s = 3$, $Pro_{u_3} = 0 + 0 + 0 = 0$, $AcS_{u_3} = 0 + 0 + 1 = 1$, 用户 u_3 从未传播过主题信息。

$$[0172] \quad AP(u_3) = \frac{0.6 \times 0 + 0.4 \times 0}{3} = 0$$

[0173] (4) 计算用户 u_4 的 AP: 从 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, $num_s = 3$, $Pro_{u_4} = 1 + 1 + 1 = 3$, $AcS_{u_4} = 1 + 1 + 1 = 3$, 用户 u_4 从未传播过主题信息。

$$[0174] \quad AP(u_4) = \frac{(0.6 \times \frac{3}{3} + 0.4 \times 0)}{3} = \frac{1}{5}$$

[0175] (5) 计算用户 u_5 的 AP: 从 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, $num_s = 3$, $Pro_{u_5} = 0 + 1 + 1 = 2$, $AcS_{u_5} = 1 + 1 + 1 = 3$, 用户 u_5 从未传播过主题信息。

$$[0176] \quad AP(u_5) = \frac{(0.6 \times \frac{2}{3} + 0.4 \times 0)}{3} = \frac{2}{15}$$

[0177] 2. 用户影响力的计算

[0178] 从附图 2 之 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, 得出 $\beta = \frac{7}{7 \times 3} = \frac{1}{3}$

[0179] (1) 计算用户 u_1 的 UI : $num_{C_i} = 8$, $num_{S(u_1)} = 2$, $\xi = \frac{1}{8^2}$, 从 b 用户 u_1 的主题信息传播图, 得出 $sum_{in.r} = 4$, $sum_{out.r} = 2$, 从 c 用户 u_1 的主题信息传播图, 得出 $sum_{in.r} = 4$, $sum_{out.r} = 2$

$$[0180] \quad UI(u_1) = \frac{1}{8^2} \times ((4 + \frac{1}{3} \times 2) + (4 + \frac{1}{3} \times 2)) = \frac{7}{48}$$

[0181] (2) 计算用户 u_2 的 UI : $num_{C_i} = 6$, $num_{S(u_2)} = 2$, $\xi = \frac{1}{6^2}$, 从 c 用户 u_1 的主题信息传播图, 得出 $sum_{in.r} = 4$, $sum_{out.r} = 2$

$$[0182] \quad UI(u_2) = \frac{1}{6^2} (4 + \frac{1}{3} \times 2) = \frac{7}{54}$$

[0183] (3) 由于用户 u_3, u_4, u_5 从未传播过主题信息, 所以他们的值为 0.01. 即 $UI(u_3) = UI(u_4) = UI(u_5) = 0.01$

[0184] 3. 社区内用户综合影响的计算

[0185] 从附图 2 之 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, $t = 1$

$$[0186] \quad UH(C_i) = \int_0^1 (\frac{25}{63} \times \frac{7}{48} + \frac{1}{6} \times \frac{7}{54} + 0 \times 0.01 + \frac{1}{5} \times 0.01 + \frac{1}{15} \times 0.01) dt = 0.083$$

[0187] 4. 社区内用户数量的计算

$$[0188] \quad NU = 8$$

[0189] 5. 社区内用户规模的计算

$$[0190] \quad \text{从附图 2 之 a 用户关系图, } US(C_i) = 5+2+0+4+3 = 14$$

[0191] 6. 社区对社区内节点的影响力的计算

[0192] 从附图 2 之 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, 得出:

$$[0193] \quad ICNI(u_1) = AP(u_2) \times \sum \text{influ} = AP(u_2) \times 1$$

$$[0194] \quad ICNI(u_2) = AP(u_1) \times \sum \text{influ} = AP(u_1) \times (1+1)$$

$$[0195] \quad ICNI(u_3) = AP(u_1) \times \sum \text{influ} + AP(u_2) \times \sum \text{influ} = AP(u_1) \times (0+0) + AP(u_2) \times 1$$

$$[0196] \quad ICNI(u_4) = AP(u_1) \times \sum \text{influ} + AP(u_2) \times \sum \text{influ} = AP(u_1) \times (1+1) + AP(u_2) \times 1$$

$$[0197] \quad ICNI(u_5) = AP(u_1) \times \sum \text{influ} + AP(u_2) \times \sum \text{influ} = AP(u_1) \times (1+1) + AP(u_2) \times 1$$

[0198] 7. 社区对社区外节点的影响力的计算

[0199] 从附图 2 之 b 用户 u_1 的主题信息传播图, c 用户 u_1 的主题信息传播图, d 用户 u_2 的主题信息传播图, 得出:

$$[0200] \quad ICTN(v_1) = \beta \times (AP(u_1) \times \sum \text{inf} + AP(u_2) \times \sum \text{inf}) = \beta \times (AP(u_1) \times (1+1) + AP(u_2) \times 1)$$

$$[0201] \quad ICTN(v_2) = \beta \times (AP(u_1) \times \sum \text{inf} + AP(u_2) \times \sum \text{inf}) = \beta \times (AP(u_1) \times (1+1) + AP(u_2) \times 1)$$

$$[0202] \quad ICTN(v_3) = \beta \times (AP(u_1) \times \sum \text{inf} + AP(u_2) \times \sum \text{inf}) = \beta \times (AP(u_1) \times (0+0) + AP(u_2) \times 0)$$

[0203] 综上所述, 本发明所有的研究都是基于社区是已经挖掘到的社区, 充分利用了以往对用户影响力等方面的研究, 提出了社交网络社区影响力的评估算法。由于以往的对社区的研究, 很少对社区的影响力进行考虑, 针对以往的缺陷, 本发明对社区影响力、社区对社区内节点的影响力、社区对社区的影响力、社区对社区的影响力等问题进行了考虑。因此, 可以对人人网, 微博等社交网络进行影响评估, 对于更好的传播信息以及网络安全等具有实际应用的前景。

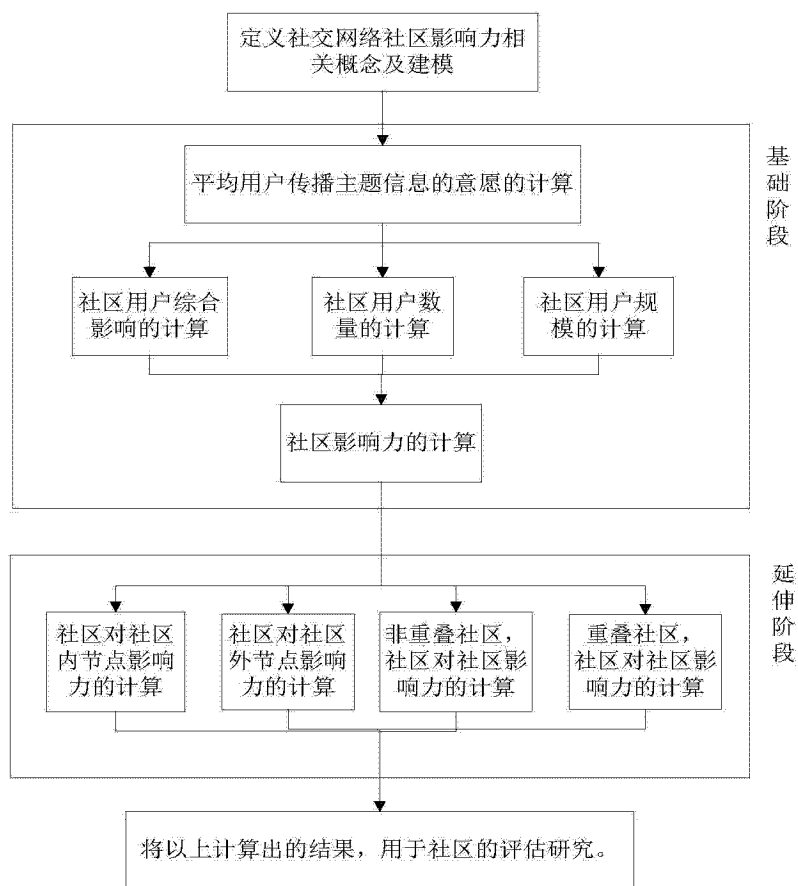


图 1

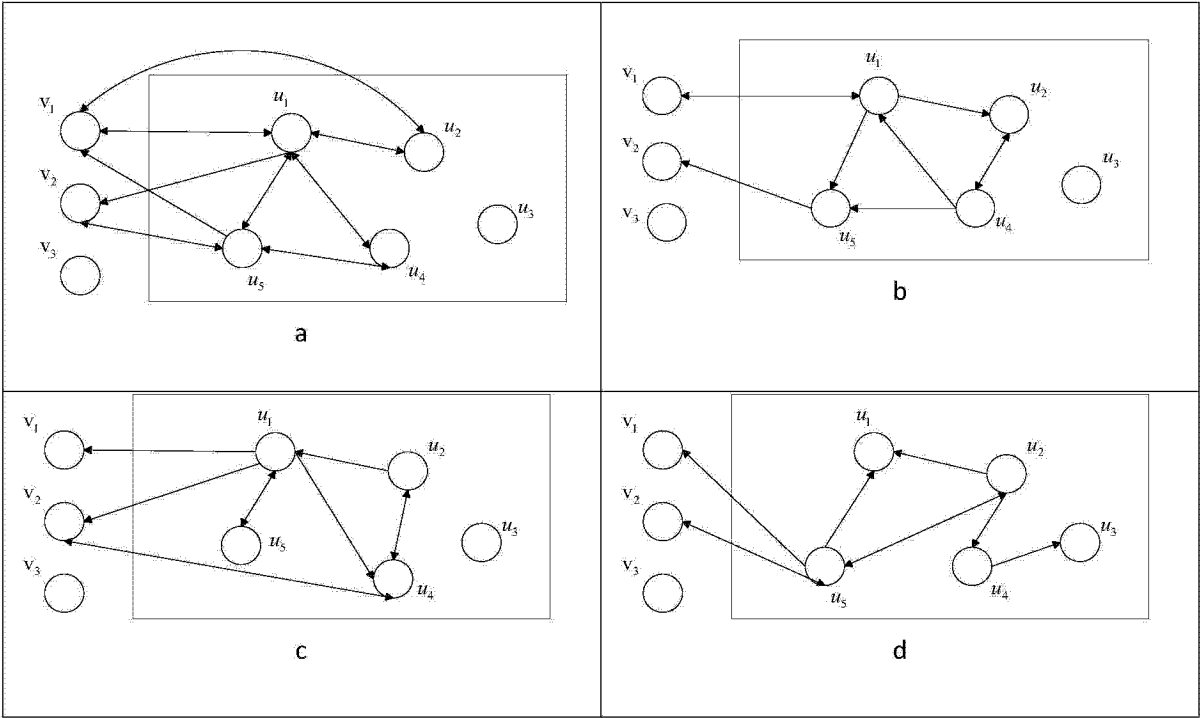


图 2