

网络演化中基于事件的节点影响力分析

熊 超 陈云芳 仓基云
(南京邮电大学计算机学院 南京 210023)

摘 要 社会影响力分析是社会网络研究领域的一个重要研究热点,目前大多数的影响力分析都是针对静态网络的。针对网络演化提出了一种基于个体事件的影响力分析方法,首先改进了传统扩散模型以适用于网络演化,并对在扩散中表现出的事件进行了定义,然后基于其中的个体事件测量了社交指数和影响力指数来衡量节点影响力,从而挖掘出动态网络影响力最大化问题中的关键节点。实验分别计算了节点的两个指标并对得出的结果分别进行了分析,找出了影响力较大的重要节点,然后对两个指标的性能进行了对比,结果表明利用社交指数发掘的节点在扩散前期传播的效率更高,而遇到扩散瓶颈时利用影响力指数发掘的节点则可以更快地突破瓶颈期。

关键词 网络演化,网络事件,影响力扩散,影响力最大化

中图法分类号 TP391 文献标识码 A

Event-based Node Influence Analysis in Social Network Evolution

XIONG Chao CHEN Yun-fang CANG Ji-yun

(School of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract Social influence analysis is an important research focus in the field of social network research, and most existing works on influence analysis focus on static network. This paper proposed a method of influence analysis based on individual events for network evolution. The traditional diffusion model was improved to adapt the evolution, and the events that exhibited in the diffusion was defined. Then two indicators based on the individual events which are social index and influence index were given to measure the influence of node. The value of the two indicators of nodes were analyzed separately in the experiment to find out the important influential node in the influence maximization problem. Then the performance of the two indicators was compared. The results show that the nodes of social index have higher efficiency than nodes of influence index at the initial diffusion stage, but when the diffusion meets the bottleneck, nodes of influence index can break the bottleneck faster so that they can infect more nodes in the network.

Keywords Network evolution, Network events, Influence diffusion, Influence maximization

1 引言

社会网络已经成为大众获取信息和交流信息的重要媒介,影响力传播是社会网络的重要特征之一。随着互联网的蓬勃发展,研究人员开始有机会在大量现实数据的基础上对社会影响力进行探讨。典型的社会网络是一个无标度网络,无标度网络的最显著特点就是“富人俱乐部现象”,即少数具有较高度数的节点统治着网络,这些极重要的点通常又被称为关键节点。关键节点在网络演化、社区划分、信息传播等应用领域均具有巨大影响,如何分析社会网络、找出这些影响力较大的关键节点成为众多领域的研究热点。

社会网络中的事物都是在不停地运动的,有的运动利于事物的发展,需要去促进,但有的运动会阻碍事物的发展,需要去抑制。比如,传染性疾病对人类健康造成巨大威胁,在传染病网络中,为了快速有效地控制病情,从而将损失尽可能降

低,可以通过社会网络分析的方法及评价节点影响力的方法等来找出传播的来源以及在传播过程中影响力较大的初始节点,然后将这些源头节点去除掉从而将其不良传播势头有效抑制住,进而阻止传染病毒的进一步传播扩散,将损失降到最低。反之,在“病毒式营销”中^[1],商家在推出一种新产品的时候,需要找出一些有影响力的个体,然后经由他们口口相传进行相互推荐,从而将效益最大化以及达到营销的目的,如何利用有限的资源找出影响力较大的体验者,通过他们的“口碑效应”^[2,3]使得产品的影响范围扩散到最大程度,就是影响力问题在病毒式营销中的运用。

影响力最大化问题首次提出后,大部分学者开始研究此问题,而目前大部分的影响力研究都是基于静态网络进行的,但对于大多数网络来说,它们的拓扑结构都是随着时间而变化的,而且现实世界中我们并不是总能获取完整的网络拓扑结构,所以研究在网络演化过程中影响力最大化问题是有意义的。

熊 超 硕士生,主要研究方向为社会计算;陈云芳 副教授,主要研究方向为社会计算、信息网络安全等;仓基云 硕士生,主要研究方向为社会计算。

2 相关工作

在社会网络中, Richardson 和 Domingos 等人^[4]首次引入了影响力最大化问题, 该问题的提出给后来研究者们指明了一个重要的研究方向。随后, Kempe 等人^[5]对影响力最大化问题进行详细研究, 总结了求解影响力最大化问题的两种一般传播模型, 即线性阈值模型和独立级联模型, 他们提出利用贪心算法来求该问题, 同时验证了在社会网络中求解该问题本身就是一个 NP 难题。目前有很多针对影响力最大化的研究使用贪心算法或者经过改进的贪心算法等来求解, 然而此类算法时间复杂度比较高, 并没有利用网络的拓扑性质来求解影响力最大化问题, 因而很难适用于规模较大的网络。

很多社会网络都存在着社区结构, 每个节点受其所在社区的其他节点影响的概率较大, 所以利用社区结构特性有助于解决影响力最大化问题。利用网络的社区性质来求解影响力最大化问题是 Galstyan 等人^[6]首次提出来的, 考虑到了社区性质, 但是他的实验是基于两个稀疏社区构成的随机网络, 网络过于简单, 所以算法存在着一定的局限性, 现实中很多社会网络至少存在两个社区。而 Cao 等人^[7]认为选取初始节点是一个求解最佳分配的过程, 他们提出一种算法将网络划分为多个社区, 然后使用动态规划算法将选取的最佳种子节点分配到每个社区中进行影响力最大化实验。

有很多关于社区的影响力最大化问题的研究已经被展开, 对于目前针对静态网络的研究工作而言, 考虑到现实生活中网络并不是一尘不变的, 而是在很长一段时期内不断变化的^[8], 网络演化过程中的影响力最大化问题开始成为研究热点。Berger-Wolf 等人^[9,10]对于动态网络提出了一种数学框架, 并设计算法来获取在此框架上动态网络的结构信息, 他们将动态网络看成是若干个静态网络的组成, 其中的每一个静态网络都作为一个快照, 相当于用图来刻画每一个时刻的快照, 基于这些图序列, 刻画网络的动态特性, 然而这种方法忽略了社会网络中个体行为的随机性和突发性。Zhuang 等人^[11]针对影响力最大化在动态网络环境下进行研究, 采取的方法是周期性地探测社会网络的一个节点集, 使得这个节点集合可以最大程度揭示真实网络影响力扩散的过程, 他们在此基础提出一种最大间隙探测算法来得到一个近似最优解, 最大可能地减小所观察网络和真实网络之间的误差。但是这种方法具有一定局限性, 即某个节点集合并不能够完全替代整个网络。

上述网络演化中的研究大多是基于一种“硬切分”的分析方法, 并且这种动态网络分析方法忽略了网络演化过程中的噪声和事件两个重要因素, 其中噪声由具有社会化特征的个体行为的不确定性和随机性所造成的, 而事件则是由个体或群体的异常行为所引起, 并且具有一定的持续时间, 同时也会具有扩散特性, 从而造成局部或整个网络的异常性变化^[12,13]。本文从事件的角度, 针对网络演化给出一种基于个体事件的影响力分析方法, 通过研究网络演化中的信息传播, 改进了传统影响力传播模型并描绘出了网络演化下的影响力扩散状态, 得出网络演化中影响力扩散的特点, 即不断出现的社区划分以及节点加入和离开等事件, 然后对网络中的事件

进行具体定义, 对其中的个体相关事件进行了分析, 通过测量节点的社交指数和影响力指数两个指标来衡量节点的影响力, 从而挖掘出在动态网络影响力最大化问题中的影响力关键节点。

3 基于事件的影响力分析

相比于网络的静态特征, 目前网络的动态演化特性也受到广泛的关注, 许多关于网络演化的分析方法被应用于社会网络的热点研究中^[8]。本文从事件的角度去研究网络演化中的影响力问题, 通过对网络演化中的影响力扩散模型的研究, 发现在扩散中不断会出现节点的加入、离开等事件, 而这些节点事件的发生也使得网络结构不断变化。给出网络演化中事件的具体定义, 并对其中的个体事件进行详细分析, 基于个体事件给出两个指标来衡量节点的影响力, 从而找出影响力关键节点。

3.1 网络演化中的影响力扩散

影响力传播模型定义了影响力在网络上的传播机制。在考察影响力的传播模型时, 通常会将网络 G 上的每一个节点都表示为有两种状态: 激活态和非激活态, 只有处于激活态的节点且与它相连的节点才能进行影响力传播, 当一个非激活态节点被其他节点成功影响时, 称此节点被激活, 且每个节点只能由非激活态转为激活态, 不能反转。目前在影响力传播研究中最基本的两种传播模型是独立级联模型和线性阈值模型^[5], 一般在研究影响力使用两种通用传播模型时, 是需要预先获取整个网络的拓扑结构的。对于线性阈值模型而言, 未激活态节点受到其周边激活态节点的影响, 一个个体有一次机会去感染另一个个体, 对于独立级联模型而言, 一个激活态节点尝试激活每一个非激活态邻居节点并且只有一次激活的机会。

静态网络中, 节点只有一次机会激活它的邻居节点, 不管这次是否被激活, 后续都不能再次激活此节点, 这与网络演化中的状况是不相符的, 例如对于节点 v , 我们希望它可以不断地去激活邻居节点, 所以区别于静态网络里的扩散, 为了描述网络演化中的扩散图, 对已被激活的节点设置感染特性(感染态)的概念, 即只要有邻居节点时这些激活态(即感染态)节点都可以去激活它的邻居节点。需要注意的是, 感染态与激活态并不冲突, 它是对传统激活态节点激活能力的一个补充。然后对于一些本身不能被激活、即使激活多次也无效的节点, 区别于非激活态, 将它们定义为孤立态。扩散过程的开始阶段, 网络中所有节点都是非激活态, 然后通过给予网络中少量节点资源, 使他们成为激活态节点以开始扩散过程, 这些激活态节点同时短时间内成为感染态节点, 在下一个时刻, 这些激活态节点尝试激活那些与它们交互的非激活态节点, 传递它们之前得到的信息, 然后激活态节点失去感染性, 被激活的节点具有感染性。接下来重复该过程, 新的感染态节点尝试激活它们的非激活态邻居节点, 直到网络中没有节点可以再激活。因为在网络演化中, 网络的拓扑结构是在不断变化的, 在接下来的时刻, 可能会有新的节点与之前已成为激活态的节点交互, 所以我们保留激活态节点的感染性, 在整个演化过程中, 每个激活态节点都可以尝试激活它的每个唯一邻居节点。

传统的影响力研究中,激活态节点只有一次机会去尝试激活它的邻居节点,然而这就忽略了现实网络拓扑不断改变形成社区划分、节点事件不断发生使得邻居节点不断变化的事实。将节点的扩散通过图1描述出来之后,可以看出网络演化中节点不断出现加入、离开的扩散特征和社区的划分。如 $T=3$ 时节点 5、6、7 合并形成一个新的社区,到 $T=4$ 时刻,原有社区的已被激活的感染态节点 4 加入该社区,开始进行激活,同时原社区新加入节点 8 与节点 2、3 交互,在 $T=6$ 时刻,离开的节点 4 同新社区一起与原有社区进行了合并。

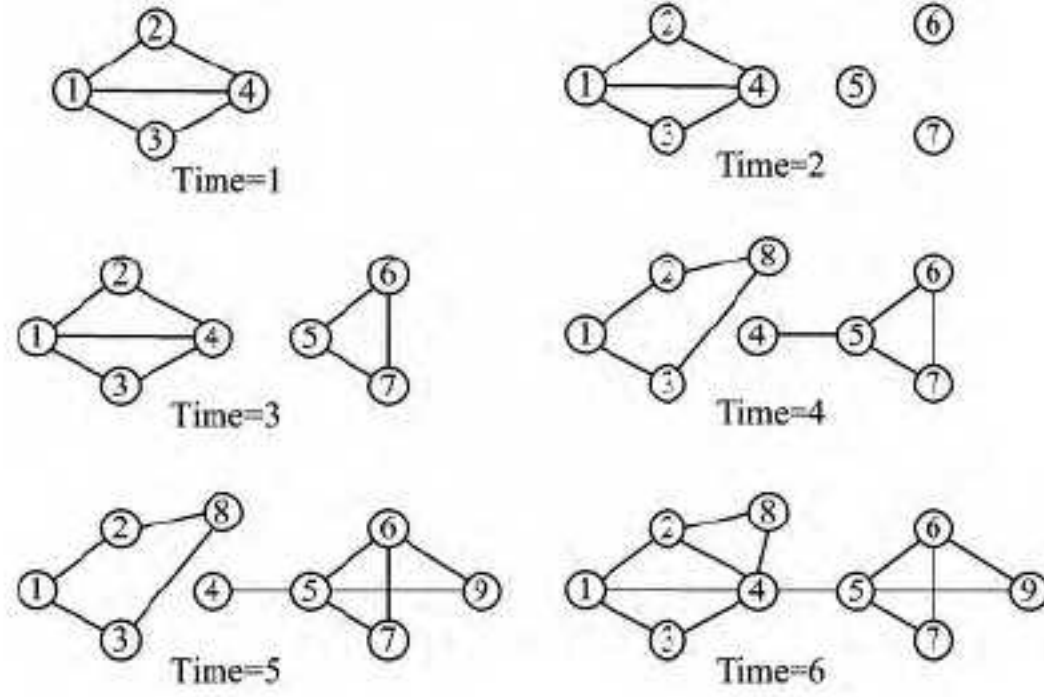


图1 网络演化中的扩散模型图

3.2 网络演化中的事件

根据 3.1 节的分析,可以看出事件是社会网络动态演化过程中的一个明显特点,为了充分利用事件进行影响力分析,下面首先从社区和个体两个方面给出网络演化中事件的定义。

关于社区的事件有:

(1) 延续。如果第 i 时刻的社区 j 里的点集 V_i^j 与第 $i+1$ 时刻的社区 k 里的点集 V_{i+1}^k 相同,就认为社区 C_i^j 在下一时刻得到了延续,为 C_{i+1}^k 。注意到,在延续事件中并没有限制边集,如图 2 中 $t=1$ 时的 C_1^1 和 $t=2$ 时的 C_2^1 ,虽然 $t=2$ 时,社区中有新边出现,但还是认为 C_1^1 得到了延续。延续事件表示为:

$$\text{Continue}(C_i^j, C_{i+1}^k) = 1 \text{ 且 } V_i^j = V_{i+1}^k \quad (1)$$

(2) k -合并。在时刻 i 对于两个不同的社区 C_i^k 和 C_i^l ,如果在下一个时刻 $i+1$ 中存在一个社区,该社区中包含至少 $k\%$ 的节点属于 C_i^k 和 C_i^l ,那么称 C_i^k 和 C_i^l 发生了 k -合并。 k -合并事件表示为:

$$\exists C_{i+1}^j, \text{ 使得 } \frac{|(V_i^k \cup V_i^l) \cap (V_{i+1}^j)|}{\max(|V_i^k \cup V_i^l|, |V_{i+1}^j|)} > k\% \quad (2)$$

并且 $|V_i^k \cap V_{i+1}^j| > \frac{|C_i^k|}{2}, |V_i^l \cap V_{i+1}^j| > \frac{|C_i^l|}{2}$ 。合并事件的发生只会在 V_i^k 和 V_i^l , 在 $i+1$ 时刻有新边的加入,如图 2 中 $t=3$,同时也说明了两个不同的社区在此时有了交互从而合并在了一起。

(3) k -分裂。社区 C_i^j 中如果在下一时刻有 $k\%$ 的节点出现在不同的两个社区里,那么称 C_i^j 发生了 k -分裂。 k -分裂事件表示为:

$$\exists C_{i+1}^k, C_{i+1}^l \text{ 使得 } \frac{|(V_{i+1}^k \cup V_{i+1}^l) \cap (V_i^j)|}{\max(|V_{i+1}^k \cup V_{i+1}^l|, |V_i^j|)} > k\% \quad (3)$$

并且 $|V_{i+1}^k \cap V_i^j| > \frac{|C_{i+1}^k|}{2}, |V_{i+1}^l \cap V_i^j| > \frac{|C_{i+1}^l|}{2}$ 。如图 2 中

$t=4$, 分裂事件说明原本同属于一个社区内的一些节点关系破裂导致不再交互,从而使得这些点在下一时刻分属于其它不同社区。

(4) 生成。如果在前一个时刻没有节点聚集在一起,在当前时刻这些点聚集在一起形成了社区,那么称该事件为社区的生成。生成事件表示为:

$$\exists C_i^j \text{ 使得 } V_{i+1}^k \cap V_i^j > 1 \quad (4)$$

如图 2 中 $t=5$, 生成事件说明新的合作关系的生成。

(5) 解体。如果前一个时刻的社区 C_i^j 在当前时刻再也没有一个节点属于该社区,那么称该事件为社区的解体。解体事件表示为:

$$\exists C_{i+1}^j \text{ 使得 } V_i^k \cap V_{i+1}^j > 1 \quad (5)$$

解体事件说明原有的合作关系破裂使得彼此间不再交互,同时解体的事件通常伴随着分裂事件。如图 2 中在 $t=5$ 中的 C_5^1 在 $t=6$ 时解体了,分裂为 3 个社区 C_6^1, C_6^2, C_6^3 。

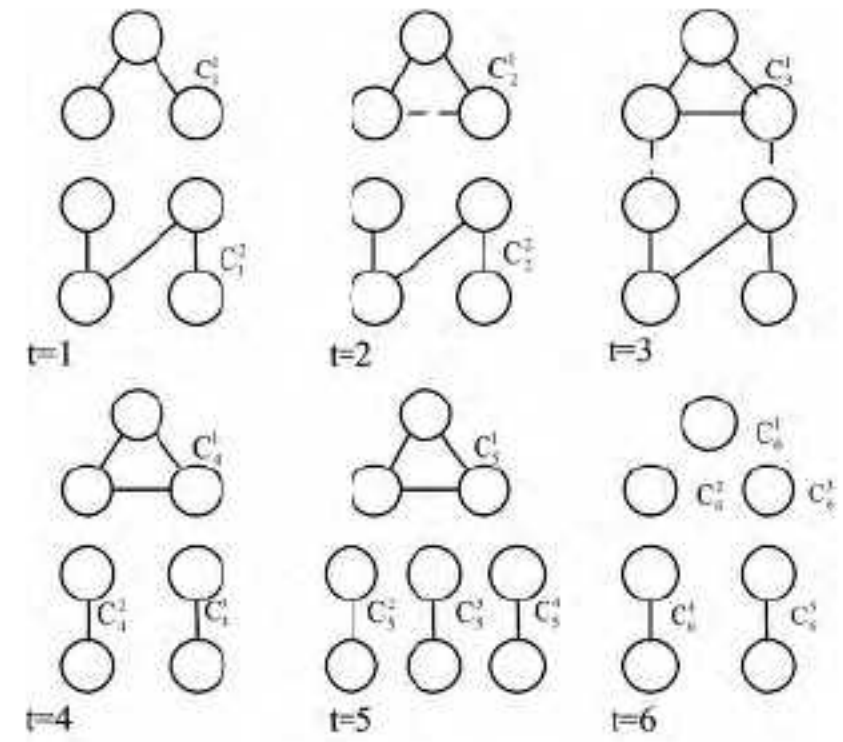


图2 时刻 $t=1$ 到 6 的网络快照图

关于个体的事件有:

(1) 出现。如果一个节点在当前网络快照社区中出现,而在之前的快照中都没有出现过,那么称该事件为节点的出现,节点的出现表示为:

$$\text{Appear}(v, i) = 1 \text{ 当且仅当 } v \notin V_{i-1}, v \in V_i \quad (6)$$

(2) 消失。如果一个节点在当前网络快照中的任一社区里都不再出现,而在之前的快照中出现过,那么称该事件为节点的消失,节点的消失表示为:

$$\text{Disappear}(v, i) = 1 \text{ 当且仅当 } v \in V_{i-1}, v \notin V_i \quad (7)$$

(3) 加入。如果一个节点在该时刻 i 存在于社区 C_i^j 中,但不存在于前一时刻的社区 C_{i-1}^k 中,且 $C_i^j \cap C_{i-1}^k > \frac{|C_{i-1}^k|}{2}$,那么称该事件为节点的加入,节点的加入表示为:

$$\text{Join}(v, C_i^j) = 1 \text{ 当且仅当 } \exists C_{i-1}^k, C_{i-1}^l \text{ 使得}$$

$$C_i^j \cap C_{i-1}^k > \frac{|C_{i-1}^k|}{2} \text{ 且 } v \notin V_{i-1}, v \in V_i \quad (8)$$

加入事件的发生一定也伴随着出现事件。在科学合作网络中,节点的加入可能是因为新生的入学。

(4) 离开。如果一个节点在前一时刻 $i-1$ 存在于社区 C_{i-1}^k 中,但不存在于当前时刻的社区 C_i^j 中,且 $C_i^j \cap C_{i-1}^k > \frac{|C_{i-1}^k|}{2}$,那么称该事件为节点的离开,节点的离开表示为:

$$\text{Leave}(v, C_i^j) = 1 \text{ 当且仅当 } \exists C_{i-1}^k, C_{i-1}^l \text{ 使得}$$

$$C_i^j \cap C_{i-1}^k > \frac{|C_{i-1}^k|}{2} \text{ 且 } v \in V_{i-1}, v \notin V_i \quad (9)$$

离开事件的发生一定也伴随着消失事件。在科学合作网络中,节点的离开可能是因为新生的毕业。

从上述关于社区和个体的事件定义,可以看出一般事件都是成对出现的。下一节将给出基于一对基本事件的分析方法。

3.3 节点事件分析

可以通过对个体行为进行分析来捕捉个体行为的特征等,从而帮助我们进行分析影响力等信息[8]。通过运用节点的两个基本事件,加入和离开来分析节点的行为信息。

社交指数是用来衡量一个节点在一段时间内与不同社区的交互次数,本文通过计算节点的加入和离开事件的次数来计算该点的社交指数,用 $cl_i(x)$ 表示时刻 i 节点 x 所属的社区。社交指数表示为:

$$SoI(x) = \frac{\sum_{i=1}^T (Join(x, cl_{i+1}(x)) + Leave(x, cl_i(x)))}{|Activity(x)|} \quad (10)$$

其中, $Activity(x) = \sum_{i=1}^T (x \in V_i)$, 该公式表明社交指数越高,节点与不同社区的交互次数就越多,在研究影响力最大化问题中,将社交指数高的节点选为初始节点有利于信息的扩散。而由于社区在一段时间里某些影响力较大的节点可能不会离开及加入到其他社区,即 $Moves(x) = 0$ 时我们以它们的度值 $Deg(x)$ 来衡量它们的影响力。

在网络演化中,我们也考虑到这样一种情况,如果存在着某一个节点离开或加入一个社区,而导致一大群节点跟着加入或离开该社区,这就表明那个节点对其他节点造成一定的影响。用 $Companions(x)$ 表示在指定时间内跟随节点 x 加入或离开社区的所有节点的数目, $Moves(x)$ 表示节点 x 加入或离开社区的次数,则节点 x 的影响力指数表示为:

$$Inf(x) = \frac{|Companions(x)|}{|Moves(x)|} \quad (11)$$

我们注意到一点,由于富人俱乐部现象,与高影响力交互的那些节点往往也有比较高的 Inf 值,如果要剔除这些追随者,需要进一步修正公式,用 $Max-Int-Node(x)$ 表示与节点 x 交互次数最多的节点, $Deg(x)$ 表示节点 x 的度值。如果下面两个式子中任意一个成立,那么节点 x 的影响力指数为 0,否则节点 x 的影响力指数为 $Inf(x)$ 的值。

$$\begin{cases} Inf(Max-Int-Node(x)) > Inf(x) \\ Deg(Max-Int-Node(x)) > Deg(x) \end{cases} \quad (12)$$

算法 1 社交指数和影响力指数计算

Input: node x , community C

Output: $SoI(x)$, $Inf(x)$

1. Initialize leave, join, activity $\leftarrow 0$
2. for $i=1$ to T do
3. leave = the times that x leave C_i + leave
4. join = the times that x join C_{i+1} + join
5. if $(x \in V_i)$ then
6. activity = activity + 1
7. end if
8. end for
9. $SoI(x) = (leave + join) / activity$
10. return $SoI(x)$
11. move $\leftarrow$ the times that x join and leave C_i
12. max-int-node(x) $\leftarrow$ the node interact with x for the most times
13. deg(x) $\leftarrow$ degree of x

14. for everytime x join or leave C_i do
15. companions = the numbers of nodes that followed x to join and leave C_i
16. end for
17. $Inf(x) = companions / move$
18. if $(Inf(max-int-node(x)) > Inf(x))$ or $deg(max-int-node(x)) > deg(x)$ then
19. $Inf(x) = 0$
20. return $Inf(x)$
21. else
22. return $Inf(x)$

4 实验分析

4.1 实验数据

本文研究计算机学科在 1994—2005 年之间的研究热点演化过程,使用社会网络演化研究的方法研究这 12 年来计算机科学的发展趋势,并用图形描绘出 12 年来的发展脉络,使得其发展历程更清晰。结合影响力最大化问题的研究,找出在这 12 年的演化时间里最具影响力的一些研究学科。

使用 1994—2005 年这 12 年的关于计算机方面的引文数据,将数据集分为 3 个时间段:1994—1997 年、1998—2001 年及 2002—2005 年,这主要考虑到大部分论文都是在校研究生或博士和其导师共同完成的,一般 4 年为一个周期。

4.2 实验结果

为了直观地展示网络演化过程,引入了冲击流图这一工具。当得到了每个时刻点的社区聚类之后,就想要揭示出随着时间变化,社区之间的变动趋势。冲击流图就能很好地揭示在一定时间段内网络社区结构的变化。冲击流图中横向每一栏对应于不同时刻网络具有的显著性的社团划分,纵向上不同色块对应于网络中划分的社区,块的高度代表社区的大小。我们绘制了冲击流图,并根据绘制出的演化流形进行分析。

实验一:网络演化的整体描绘和分析

图 3 为该计算机引文网络的完整冲击流图。从图中可以直观地看出在 1994—1997 年间,研究热点为以特征提取为主的相关话题,而在 1998—2001 年间,有两个比较受关注的热点,以知识获取为主的相关话题和以 HTML 为主的相关话题,在 2002—2005 年间,就形成了多个研究热点,主要有自然语言处理、密码学、工作流管理系统等。

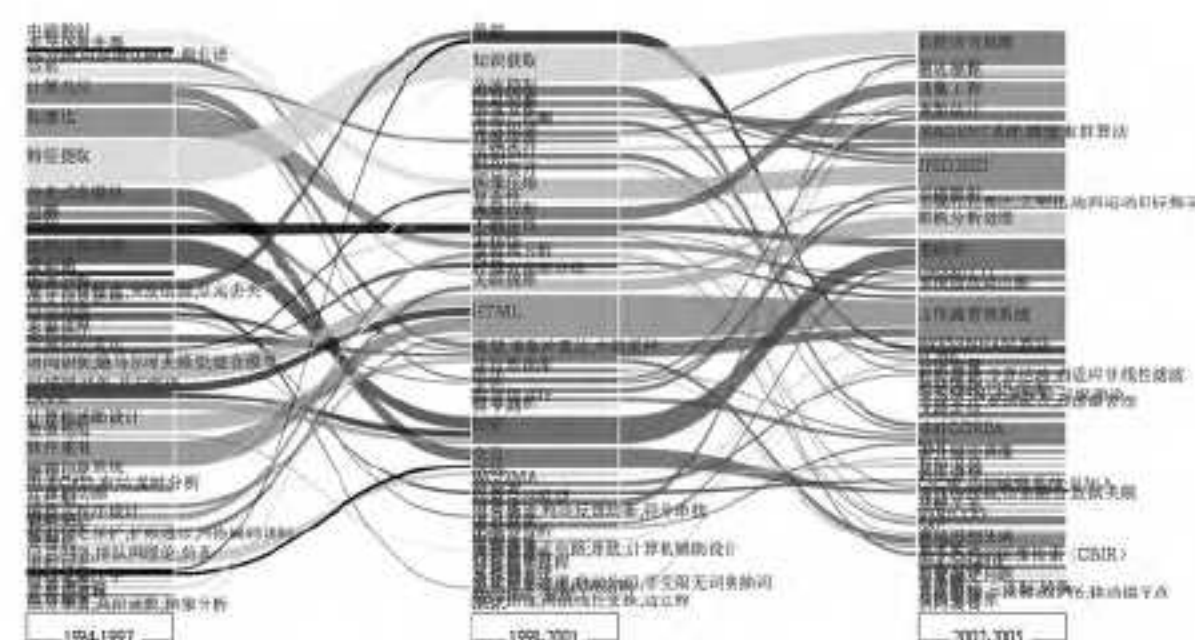


图 3 引文网络的冲击流图

以 2002—2005 年的经过演化形成的工作流管理系统社区为例,图 4 展示了该社区的演化过程。

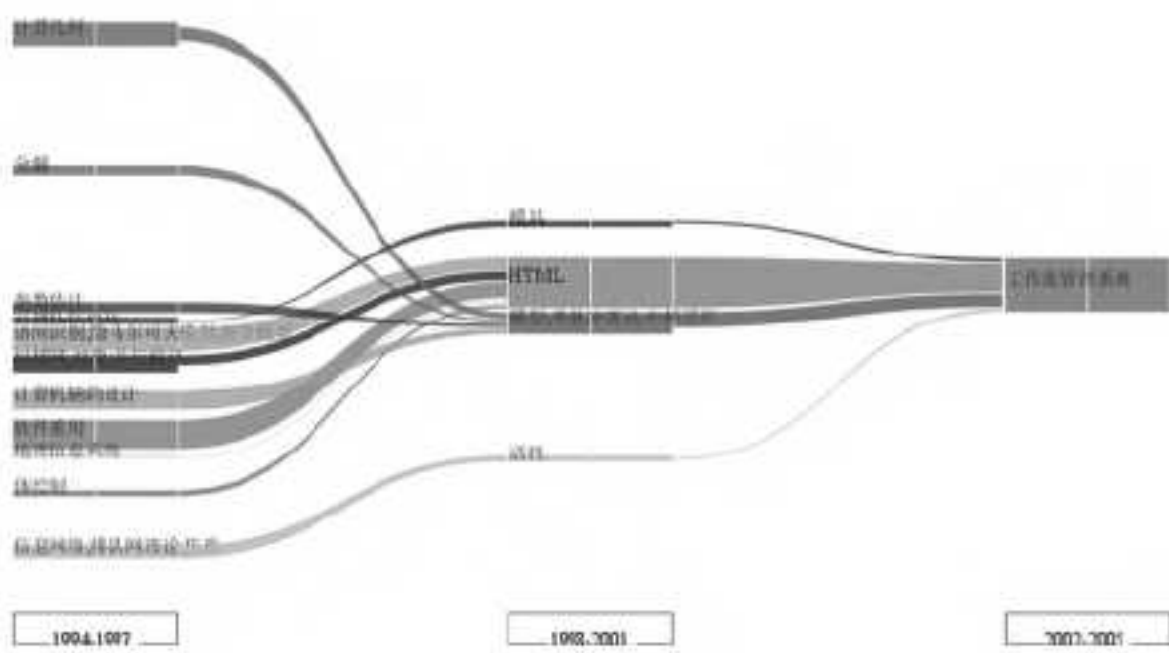


图4 社区“ workflow 管理系统”的演化流形

从图4以直观地看出,该社区的形成主要是通过很多中小社区的逐渐合并形成的,从最初1994—1997年间的11个社区,到1998—2001年间的4个社区,这表明很多学科的发展都是基于前面的研究展开的,如果能找出之前的研究热点方向,那么在以后的研究中就很容易抓住科研的主流方向。

实验二:社区内节点影响力的度量分析

由于该数据集较大,包含40400个节点,其中1994—1997年间有6609个节点,1998—2001年间有12462个节点,2002—2005年间有21329个节点,本文从中抽取典型社区来分析社区网络中的事件与节点的影响力等特征。

以初始1994—1997年间的初始社区特征提取为例,图5展示了该社区的演化。

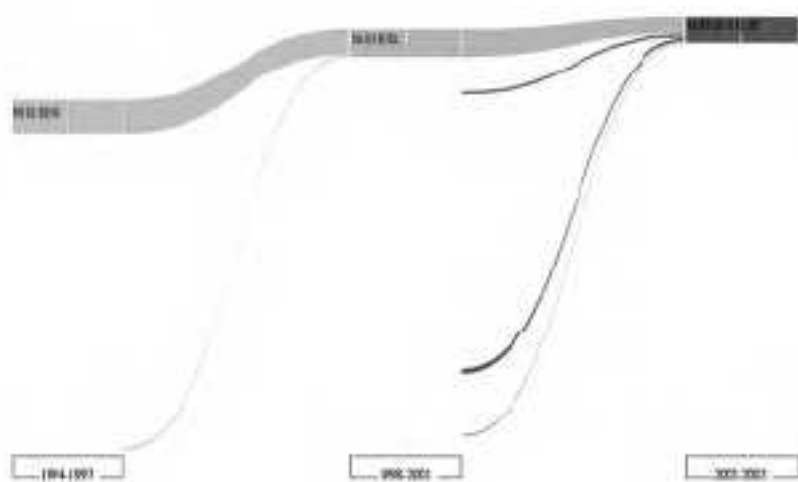


图5 社区“特征提取”的演化流形

从图5中可以直观地看出该社区与另外一个社区在1998—2001年间合并成为知识获取这个社区,而知识获取这个社区在2002—2005年间与另外3个社区最终合并成为自然语言处理社区。

先对这3个不同时间段社区内节点的度进行排序,如表1所列。

表1 各社区内节点度排序

时间段	1994—1997	1998—2001	2002—2005
神经网络	神经网络	神经网络	中文信息处理
专家系统	遗传算法	遗传算法	计算机应用
遗传算法	模式识别	模式识别	支持向量机
人工智能	专家系统	专家系统	人工智能
知识表示	人工智能	人工智能	模式识别
模式识别	机器学习	机器学习	信息检索
机器学习	知识库	知识库	人脸识别
知识获取	知识表示	知识表示	机器学习
人工神经网络	人脸识别	人脸识别	自然语言处理
特征提取	机器翻译	机器翻译	搜索引擎
聚类	自然语言处理	自然语言处理	语音识别
故障诊断	优化	优化	机器翻译
BP算法	信息融合	信息融合	相似度
学习算法	聚类分析	聚类分析	远程教育
矢量量化	语料库	语料库	人脸检测
自适应	时延	时延	知识库

从表1中可以看出在1994—1997年和1998—2001年这两个时间段内,排名前8的节点几乎一致,而在2002—2005年年间,只有人工智能和模式识别还保持在前8,而神经网络、遗传算法等都不再是2002—2005年间的研究重点了。

接下来对3个时间段社区内节点的社交指数和影响力指数进行了计算和分析,得出的结果如表2所列。

表2 Sol指数和Inf指数排序前15的节点

测量指数	Sol	Inf
神经网络	神经网络	神经网络
遗传算法	遗传算法	遗传算法
模式识别	中文信息处理	中文信息处理
人工智能	计算机应用	计算机应用
机器学习	人工智能	人工智能
自然语言处理	模式识别	模式识别
中文信息处理	信息检索	信息检索
语音识别	机器学习	机器学习
计算机应用	自然语言处理	自然语言处理
人脸识别	相似度	相似度
知识获取	支持向量机	支持向量机
聚类分析	机器翻译	机器翻译
机器翻译	模糊推理	模糊推理
特征抽取	人脸识别	人脸识别
组合优化	语音识别	语音识别

从表2中可以看出神经网络和遗传算法这两个研究点分别社交指数和影响力指数的前两名,但在社交指数里排在第3位的模式识别在影响力指数里却跌到了第6位,而在社交指数里排第7位的中文信息处理在影响力指数中排到了第3位,这主要是因为社交指数大的节点,其对应的度值也比较大,而在影响力排序中,如前文所述,由于存在富人俱乐部效应,高影响力节点会有一些与之交互的追随者,因此针对这一问题做出了改进,从而导致了部分社交指数大的节点的影响力指数有所降低。

实验三:社交指数和影响力指数挖掘关键节点性能对比

基于前面的节点事件分析,选取排序靠前的节点作为扩散的初始节点进行影响力扩散实验,在该实验中用下面的公式计算节点间的权重。

$$\omega = f/d \quad (13)$$

中, f 表示该节点的词频, d 表示该节点的度。

经过多次实验,将激活概率设置为0.4,初始节点数量与扩散度的关系如图6所示。

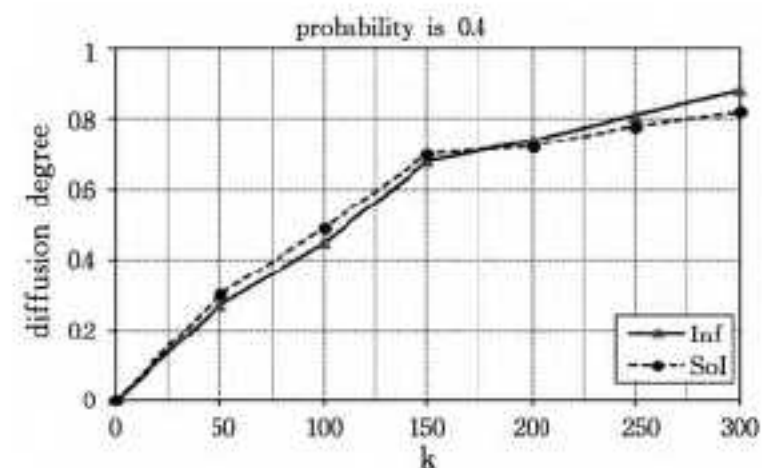


图6 Inf与Sol性能对比

从图6中可以看出,选取的初始节点少于150个时,利用社交指数发掘的节点在扩散过程中的效果要优于利用影响力指数发掘的节点;当节点数大于150个时,影响力指数的效果要优于社交指数。这主要是因为很多社交指数高的节点的度值也很高,在扩散初期,度数高的节点有利于信息扩散,而当扩散到一定程度时,由于富人俱乐部效应,那些影响力指数高于社交指数的节点往往能快速突破扩散瓶颈,从而更好地进行信息扩散。

结束语 目前绝大多数影响力问题的研究都是基于静态网络的,我们针对网络演化给出了一种基于事件的影响力分析方法,首先描绘出网络演化中的影响力扩散模型,总结了网络演化中扩散的特点,即不断出现的社区划分和节点的加入离开,接下来对网络演化中的个体事件和社区事件进行了定义,实验中对10年的计算机学科研究热点进行分析,基于个体事件给出了两个指标即社交指数和影响力指数来衡量学科的影响力,并找出影响力传播过程中的关键节点,最后对两个指标的性能进行了对比分析,结果表明用社交指数和影响力指数挖掘出的节点分别在扩散初期和扩散的瓶颈期更有利于信息的扩散。

参 考 文 献

- [1] Goldenberg J, Libai B, Muller E. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata[J]. *Academy of Marketing Science Review*, 2001, 9(3): 1-18
- [2] Goldenberg J, Libai B, Muller E. Talk of the network: A complex systems look at the underlying process of word-of-mouth[J]. *Marketing Letters*, 2001, 12(3): 211-223
- [3] Ma H, Yang H, Lyu M R, et al. Mining social networks using heat diffusion processes for marketing candidates selection[C]// *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. ACM, 2008: 233-242
- [4] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing[C]// *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2002: 61-70
- [5] Kempe D, Kleinberg J, Tardos é. Maximizing the spread of influ-

ence through a social network[C]// *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2003: 137-146

- [6] Galstyan A, Musoyan V, Cohen P. Maximizing influence propagation in networks with community structure[J]. *Physical Review E*, 2009, 79(5): 056102
- [7] Cao T, Wu X, Wang S, et al. OASNET: an optimal allocation approach to influence maximization in modular social networks[C]// *Proceedings of the 2010 ACM Symposium on Applied Computing*. ACM, 2010: 1088-1094
- [8] Asur S, Parthasarathy S, Ucar D. An event-based framework for characterizing the evolutionary behavior of interaction graphs[J]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2009, 3(4): 913-921
- [9] Berger-Wolf T Y, Saia J. A framework for analysis of dynamic social networks[C]// *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2006: 523-528
- [10] Habiba, Yu Y, Berger-Wolf T Y, et al. Finding spread blockers in dynamic networks[M]// *Advances in Social Network Mining and Analysis*. Springer Berlin Heidelberg, 2010: 55-76
- [11] Zhuang H, Sun Y, Tang J, et al. Influence maximization in dynamic social networks[C]// *2013 IEEE 13th International Conference on Data Mining (ICDM)*. IEEE, 2013: 1313-1318
- [12] Wu Bin, Wang Bai, Yang Sheng-qi. the evolution of the social network analytical framework based on events [J]. *Journal of Software*, 2011(7): 1488-1502
- [13] Ilhan N, Oguducu I G. Community Event Prediction in Dynamic Social Networks[C]// *2013 12th International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2013: 191-196

(上接第 373 页)

据操作方面的性能优于之前所提方案的,且计算开销较小。云存储环境下数据完整性验证问题的研究是一个非常活跃的方向,关于云存储数据完整性验证方面的研究还不成熟,无论是技术方面还是理论方面都有很大的不足,希望科技工作者们能够继续深入研究,争取有更大的研究成果。

参 考 文 献

- [1] 陈兰香. 一种基于同态 Hash 的数据持有性证明方法. *电子与信息学*[J]. 2011, 33(9): 2200-2204
- [2] Ateniese G, Berns R, Cutmola R, et al. Provable Data Possession at Untrusted Stores[C]// *Proc of the 14th ACM Conference on Computer and Communications Security*. New York: ACM, 2007: 598-609
- [3] Ateniese G, Pietro, R D, Mancini L V, et al. Scalable and Efficient Provable Data Possession[C]// *Proc of the 4th International Conference on Security and Privacy in Communication Networks Istanbul, Turkey*. ACM, 2008: 1-10
- [4] Erway C, Kupcu A, Papamanthou C, et al. Dynamic Provable Data Possession[C]// *Proc of the 16th ACM Conference on Computer and Communications Security*. Chicago, Illinois, USA: ACM, 2009: 213-222
- [5] Wang Q, Wang C, Li J, et al. Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing[C]//

Proc of the 14th European Conference on Research in Computer Security. Heidelberg, Berlin, 2009: 355-370

- [6] Hao Z, Zhong S, Yu N. A Privacy-preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(9): 1432
- [7] Curtmola R, Khan O, Burns R, et al. Multiple-Replica Provable Data Possession [C]// *28th IEEE ICDCS*. 2008: 411-420
- [8] Barsoum A F, Hasan M A. On Verifying Dynamic Multiple Data Copies over Cloud Servers [R]. *Cryptology ePrint Archive*, 2011: 447
- [9] Damgard I, Ren K, Lou W, et al. Toward Publicly Auditable Secure Cloud Data Storage Services[J]. *IEEE Network*, 2011, 24: 19-24
- [10] Wang C, Jurki M. A Simplification and some Application of Paillier's Probabilistic Public Key System [C]// *4th International Workshop on Practice and Theory in Public Key Cryptosystems*. 2001: 13-15
- [11] 于洋洋, 虞慧群, 范贵生. 一种云存储数据完整性验证方法[J]. *华东理工大学学报(自然科学版)*, 2013, 39(2): 211-216
- [12] 胡德敏, 余星. 一种基于同态标签的动态云存储数据完整性验证方法[J]. *计算机应用研究*, 2014, 31(5): 1362-1365
- [13] Chen L, Guo G. An Efficient Remote Data Possession Checking in Cloud Storage[J]. *International Journal of Digital Content Technology and its Applications*, 2011, 5(4): 43-50