

超图结构下的在线社交网络中隐性影响力评估

李泉林, 李超然, 马静宇*

(燕山大学经济管理学院, 河北 秦皇岛 066004)

摘要: 本文将超图结构、概率行为与信息理论三者进行合理综合, 并将它们应用到在线社交网络及其隐性影响力评估的研究中, 从而提出了一种新的有效数学分析方法. 本文首先利用超图理论并通过超路径建立了在线社交网络中用户之间的信息传播过程, 同时结合概率行为描述了信息传播的量化关系与波动强度. 然后, 本文通过平均互信息提出了一种新的在线社交网络隐性影响力评估的量化方法, 同时也使用数值算例验证了这种评估方法的有效性.

关键词: 在线社交网络; 超图; 超路径; 社团组织; 概率行为; 平均互信息; 隐性影响力

中图分类号: TP393

文献标识码: A

文章编号: 1000-5781(2020)01-0130-15

doi: 10.13383/j.cnki.jse.2020.01.011

Online social networks under hypergraph structure and their hidden influence evaluation

Li Quanlin, Li Chaoran, Ma Jingyu*

(School of Economics and Management, Yanshan University, Qinhuangdao 066004, China)

Abstract: This paper sets up a reasonable synthesis among hypergraph structure, probability behavior and information theory, and applies them to the study of online social networks. This motivates us in this paper to provide a new mathematical method for analyzing online social networks and their hidden influence. Therefore, this paper first applies hypergraph theory to set up the information transmission process by means of a hyperpath between any two users, and describes quantitative relation and fluctuation strength through combining hypergraph structure with probabilistic behavior. Then the average mutual information is established to provide a new mathematical evaluation method of hidden influence. Numerical examples verify the effectiveness of the evaluation method of hidden influence.

Key words: online social network; hypergraph; hyperpath; community; probabilistic behavior; average mutual information; hidden influence

1 引 言

目前, 在线社交网络已经成为人们日常生活与工作中不可或缺的重要组成部分. 根据eMarketer的预测: 2017年中国社交网络用户会达到6.62亿人, 并且在未来几年其规模将不断地快速增长. 这就推动了国内外许多学者对在线社交网络开展广泛的研究, 包括用户行为与属性分析^[1-6]、影响力评估^[7-10]、社团结构检

收稿日期: 2017-09-27; 修订日期: 2018-10-15.

基金项目: 国家自然科学基金资助重点项目(71932002); 国家自然科学基金资助项目(71471160; 71671158); 河北省自然科学基金资助项目(G2017203277); 北京市社会科学基金研究基地资助项目(19JDGLA004).

本文的初稿已经在中国运筹学会随机服务与运作管理分会 2017 学术年会上宣讲并被评选为优秀论文二等奖.

*通信作者

测^[11-15]与信息传播过程研究^[16-20]等多个方面。

在线社交网络通常描述了在线用户之间的各种复杂连接关系,而图论是多年来国内外学者用于构建网络拓扑结构的一种重要数学方法。自从1989年Berge^[21]提出超图理论以来,超图由于其自身顶点与超边的拓扑属性,相较于一般图能够更好的体现复杂网络中节点及其社团组织之间的拓扑结构关系,从而在数据挖掘^[22,23]、聚类分析算法^[24-26]、在线社交网络以及社会网络等方面获得了广泛的应用。

近年来,利用超图或者超网络研究在线社交网络以及社会网络中的实际问题是一个新的有较好发展前景的研究方向,其中已有的研究工作包括以下三个方面:(a)在信息传播与影响力扩散方面,Suo等^[27]基于社交网络中的不同关键属性与信息传播方式构建了一种超网络,由此对社交网络中的用户进行了等级评价,并为用户提供了个性化的推荐服务。Du^[28]提出了基于概率超图的一种新的信息传播模型,寻找了概率最大的传播路径,度量了源节点的领导能力并能够确定出意见领袖。Gangal等^[29]将超图中的节点代表行动者,超边代表组织群体,通过分析影响力扩散过程提出了一种新的影响力最大化方法。(b)在社区结构检测方面,Tao等^[30]针对社交网络使用超图构建了所有的局部链接结构,并采用hMETIS方法进行了超图分割,从而提出了一种基于超图的重叠社区提取方法。Yang等^[31]在文献[30]基础上,利用信息熵描述度分布并提出了一个名为EQHyperpart的社区检测工具,对无标度网络和一些经典的现实数据进行了评估测试。Zhang和Liu^[32]提出了一个进化的超图模型,并通过Folksonomy的底层结构研究了社交网络中的用户行为。(c)在话题关系与用户关系方面,肖玉芝和赵海兴^[33]利用超图理论建立了在线社交网络中用户行为的四层超网络模型,并通过兴趣话题的趋同性来促使用户之间可以建立朋友关系。Fang等^[34]在社交媒体网络中利用超图理论提出了一种有效的计算方法(TSIM),同时也推断了超图中每个节点在不同话题中的影响力。You等^[35]通过对腾讯QQ数据集的分析,得出群体及其成员之间三种不同类型的网络:群体超图、群体网络与用户网络,揭示了微观层面的社会关系互动。

在线社交网络中用户之间的影响力分析是多年来国内外社交网络领域中一个重要的研究课题。一般地,用户之间的影响力可以根据用户之间的邻接关系分为两种不同的形式:显性影响力与隐性影响力。显性影响力是指由用户之间的直接邻接关系所形成的影响力;而隐性影响力则是用户之间通过若干其他用户为他们建立的间接联系而形成的影响力^[36,37]。针对在线社交网络中隐性影响力的验证与度量,目前国内外学者主要从话题之间、信息条目之间的关系进行了量化研究与分析^[38-40]。与此同时,国内外学者也积极探索将新的理论与方法应用到隐性影响力的度量模型研究上。Steeg和Galstyan^[41,42]利用信息理论中的转移熵对社交网络中用户之间的隐性影响力进行了分析。王祯骏等^[43]利用IBP(印度自助餐过程)提出了通过矩阵分解将在线社交网络的内容与结构联系到影响力传播过程,由此建立了一种隐性影响力的传播模型(InfoIBP),并利用隐马尔可夫链描述了隐性影响力的动态演化过程。

近些年来,各种实际在线社交网络获得了快速发展,它们具有规模庞大、连接复杂、结构多样等诸多特点,这就使得用户之间的连接关系变得异常复杂,此外还存在着用户之间的信息保密与关系隐藏等特殊因素,所有这些都会使得挖掘用户之间的连接关系并度量其影响力等的研究工作存在越来越大的理论困难与挑战。例如,在一个复杂的在线社交网络中,如何表达多层次以及多社团组织之间的信息传播过程?如何度量信息传播中的传播强度与信息损失率?等等。这就推动了本文提出一种新的在线社交网络及其隐性影响力评估的研究方法,即将超图结构、概率行为与信息理论三者进行合理综合,并引入到在线社交网络及其隐性影响力评估研究中,从而实现了从在线社交网络几何结构及用户行为的量化分析与数据建模研究。

针对目前由用户之间的复杂间接连接关系所引起的隐性影响力评估的理论困难与挑战,本文提出了一种新的研究方法来讨论在线社交网络及其隐性影响力评估。首先,本文以超图理论为基础描述了在线社交网络中的用户与社团组织间的拓扑结构,并利用任意两个用户之间的超路径构建它们之间的信息传播过程。其次,本文对信息在超路径上的传播过程进行概率分析,并利用信息理论中平均互信息构建隐性影响力的量化评估模型。最后,本文通过数值算例验证了这种评估方法的可计算性、易操作性与有效性。因此,本文所

给出的在线社交网络及其隐性影响力评估的新方法与相关结果, 可望能够用于研究在线社交网络中的一些重要基础理论问题, 同时也可对在线社交网络中的用户关系挖掘、信息传播与舆情控制、商业性推广以及组织变革等方面获得较多的实际应用.

2 在线社交网络的超图结构模型

首先使用超图中的顶点与超边 (即一些顶点的集合) 来表达在线社交网络中的用户、社团组织 (即一些用户的集合); 然后通过超图中的超路径来表达在线社交网络中任意两个用户之间的信息传播过程及其路径选择; 最后通过信息传播的概率行为在超图上建立了社团组织的重要度.

在下面使用超图来描述在线社交网络中的基本因素: 用户与社团组织. 其后从在线社交网络的超图结构与概率行为出发, 本文将提出针对在线社交网络的隐性影响力的一种评估方法.

用户: 在现实社会网络中, 往往用“行动者”来指代例如个人、企业、服务机构等等的社会活动交互实体^[44], 与此相应的, 在线社交网络中的用户则可以被认为是在线社交网络中能够进行信息交互的“实体”.

一个在线社交网络中包含了 n 个不同用户 $v_1, v_2, \dots, v_n$, 它们的集合为

$$V = \{v_1, v_2, \dots, v_n\}.$$

社团组织: 针对社会网络中的各个节点通过它们之间的相互连接关系可以被划分若干部分 (即子集), 而每一部分内部的节点之间存在的连接“紧密”, 不同部分之间的节点连接“稀疏”, 这些连接“紧密”的节点集合就被看成为一个社团组织^[45]. 同时, 不同社团组织之间在现实中也存在重叠现象的, 即某个节点同时属于多个社团组织^[46]. 在在线社交网络中的不同话题的引导下, 一些用户可以集聚成多个重叠的社团组织.

针对在线社交网络中 n 个不同用户 $v_1, v_2, \dots, v_n$, 它们的各个非空子集表示为

$$e_r = \{v_{i_1}^{(r)}, v_{i_2}^{(r)}, \dots, v_{i_r}^{(r)}\}, 1 \leq r \leq n,$$

其中 e_r 就是在在线社交网络中的一个社团组织.

因此, 包含 n 个不同用户的在线社交网络中最多包含 $2^n - 1$ 个不同的社团组织. 设这个在线社交网络中的社团组织集合为

$$E = \{e_1, e_2, \dots, e_m\}, 1 \leq m \leq 2^n - 1.$$

在线社交网络中的基本因素被描述为超图中的顶点、超边与超路径等. 对于超图的理论与方法, 读者可参见专著 Berge^[21].

超图: 根据 Berge 中介绍的超图理论与方法^[21], 超图可以表示为一种二元关系 $H = (V, E)$, 其中 V 表示所有顶点的集合, E 表示所有超边的集合.

在线社交网络可被表述为一个超图: 有限用户集 $V = \{v_1, v_2, \dots, v_n\}$ 为超图的顶点集合. V 中的子集合 $e_i = \{v_{j_1}^{(i)}, v_{j_2}^{(i)}, \dots, v_{j_k}^{(i)}\}, v_{j_k}^{(i)} \in V, 1 \leq k \leq i, 1 \leq i \leq m$, 而社团组织 e_i 称为超图的一个超边. 社团组织的集合 $E = \{e_1, e_2, \dots, e_m\}$ 是超图的超边集合, 它们满足如下关系:

1) $e_i \neq \emptyset, i = 1, 2, \dots, m$.

2) $\bigcup_{i=1}^m e_i = V$.

例 1 令 $V = \{v_k, 1 \leq k \leq n\}, E = \{e_l, 1 \leq l \leq 5\}$, 则如图 1 所示的二元关系 $H = (V, E)$ 就是一个超图. 图 1 所示的超图 $H = (V, E)$ 对应于一个由多个用户 (其中只标注了 9 个用户) 与 5 个社团组织所构成的一个在线社交网络.

超路径: 设在线社交网络中各条超路径的起点与终点为两个不同的用户, 而超路径内部都是由一些不

相同的社团组织通过用户信息交换而衔接在一起的, 其中在超路径上一些用户会同时属于前后衔接的两个不同的社团组织, 形象地说, 在线社交网络中的一个超路径就是由一系列相交并串联在一起的社团组织构成的.

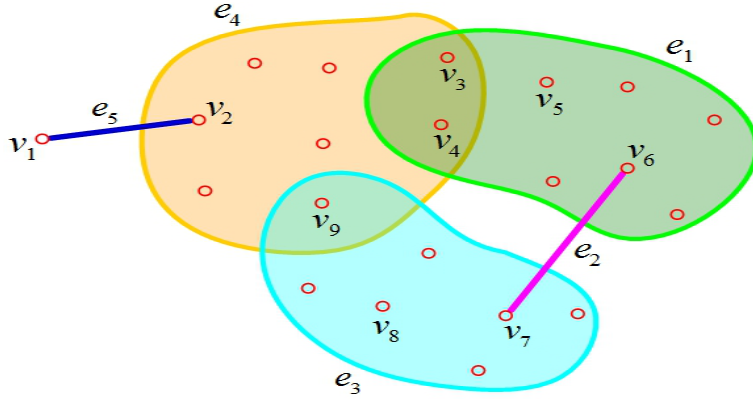


图 1 一个超图 $H = (V, E)$ 的举例

Fig. 1 An example for hypergraph $H = (V, E)$

针对在线社交网络中的超路径的数学描述如下: 如果在用户与用户之间有一条长度为 q 的超路径 (这里的单位长度表示为由两个相交且衔接的社团组织构成的虚拟距离, 不考虑它们的实际距离与地理环境等), 那么这条超路径可以表示为从首端用户到末端用户由一系列社团组织相交并衔接的一个集合序列^[47]

$$\{v_i, e_{i_1}, v_{i_1}, e_{i_2}, v_{i_2}, \dots, v_{i_{q-1}}, e_{i_q}, v_j\}, i_1, i_2, \dots, i_q \in \{1, 2, \dots, n\},$$

并且它们满足如下条件

- 1) 用户 $v_i, v_{i_1}, v_{i_2}, \dots, v_{i_{q-1}}, v_j$ 在这个在线社交网络中彼此相异;
- 2) 令 $v_{i_0} = v_i, v_{i_q} = v_j$, 则 $v_{i_k}, v_{i_{k+1}} \in e_{i_{k+1}}, k = 0, 1, \dots, q-1$;
- 3) 社团组织 $e_{i_1}, e_{i_2}, \dots, e_{i_{q-1}}, e_{i_q}$ 在这个在线社交网络中彼此相异.

从图 1 可以看出, 在线社交网络中的用户可有两种不同的类别: 第一类是用户仅仅处在一个社团组织之中, 例如 v_8 仅仅在 e_3 中, v_5 仅仅在 e_1 中, v_1 仅仅在 e_5 中, 则称这种用户为单社团组织的用户. 第二类用户可以同时属于两个或多个不同的社团组织, 例如 $v_2 \in e_5$ 并且 $v_2 \in e_4$, 则称这种用户为多重社团组织的用户. 单社团组织的用户可以保存与扩大一个信息在这个社团组织内部的影响程度; 而多重社团组织的用户既是一个信息从一个社团组织扩散到另一个社团组织的基本媒介或载体, 又可以保存与扩大一个信息在这个社团组织内部的影响程度. 从图 1 中可以容易的观察到在线社交网络中这两类不同的用户. 此外, 针对用户之间的连接关系, 也可以分为两种不同的用户对: 邻接用户对 (直接连接), 即两个用户仅属于同一社团组织; 不邻接用户对 (间接连接), 即两个用户属于不同的社团组织.

社团组织的重要度: 为了表达不同社团组织于在线社交网络中对传播信息时可能会起到的不同作用, 在这里引入社团组织的重要度. 实质上, 这对应于在超图中超边的权重. 首先, 利用关联矩阵 $(s_{k,l})_{n \times m}$ 表示在线社交网络中用户 $v_k, 1 \leq k \leq n$ 与社团组织 $e_l, 1 \leq l \leq m$ 之间的从属关系. 定义关联矩阵中的元素为

$$s_{i,j} = \begin{cases} 1, & v_i \in e_j, \\ 0, & v_i \notin e_j. \end{cases}$$

显然在矩阵 $S = (s_{i,j})$ 中, 行的序数表示用户的序数, 而列的序数表示社团组织的序数.

例2 设在线社交网络中的一个关联矩阵为

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

通过观察可知用户 v_1 同时属于社团组织 e_1, e_4, e_5 ; 用户 v_2 同时属于社团组织 e_2, e_3, e_4 ; 用户 v_3 同时属于社团组织 e_1, e_4 .

为了定量地描述与度量在线社交网络中社团组织的重要度, 首先根据关联矩阵 $(s_{k,l})_{n \times m}$ 考虑一个用户在各个社团组织中呈现的地位以及在传播信息时所发挥的作用. 用户 v_k 对应于关联矩阵 $\mathbf{S}$ 中的第 k 个 m 维行向量为

$$\mathfrak{R}_k = (s_{k,1}, s_{k,2}, \dots, s_{k,m}), \quad 1 \leq k \leq n.$$

于是, 这个 m 维行向量 $\mathfrak{R}_k$ 的范数为

$$\|\mathfrak{R}_k\| = \sqrt{\sum_{j=1}^m s_{k,j}^2}, \quad 1 \leq k \leq n.$$

注意到用户 v_k 与 m 个社团组织的结构关系由向量 $\mathfrak{R}_k$ 来表示, 因而使用向量的范数可以定义一个社团组织 e_k 在这个在线社交网络中的重要度为

$$W(e_k) = \sqrt{(\lambda_k a_k)^2 + (\mu_k b_k)^2}, \quad (1)$$

其中 $\lambda_k, \mu_k \geq 0, \lambda_k + \mu_k = 1$, 即 λ_k 与 μ_k 是一对加权系数 (注意 λ_k 与 μ_k 的设定具有主观性, 也可以来源于在线社交网络的经验数据与其现实情况的宏观分析, 并称 $\lambda_k, 1 \leq k \leq n$ 为对外传播力参数; 称 $\mu_k, 1 \leq k \leq n$ 为对内交换力参数), 并且

$$a_k = \sum_{j \neq k} \sum_{v_l \in e_j \cap e_k} \|\mathfrak{R}_l\|, \quad b_k = \sum_{v_l \in e_k / \bigcup_{j \neq k} e_j} \|\mathfrak{R}_l\|.$$

注意 a_k 表达了多重社团组织的用户在传播信息时所起的作用, 而 b_k 表示单社团组织的用户在传播信息时所起的作用, 其中 $v_l \in e_k / \bigcup_{j \neq k} e_j$ 表示只属于 e_k 这一个社团组织的用户.

注 1 1) a_k 反映了在线社交网络中社团组织 e_k 与其它社团组织的交集部分中的用户传播信息的能力 (同时也有在社团组织内部交换信息与保存信息的能力).

2) b_k 反映了在线社交网络中仅属于社团组织 e_k 的用户在社团组织 e_k 内部交换信息与保存信息的能力.

3) 注意 λ_k 与 μ_k 反映了社团组织 e_k 中两种不同能力 a_k 与 b_k 于在线社交网络中对于传播信息时所起作用的一个主观加权系数设定, 它来源于经验数据与现实情况分析. 因此, 加权系数 λ_k 与 μ_k 可由在线社交网络中的相关数据并使用统计方法来确定, 也可由用户主观的效用函数来确定.

4) 基本观察: $\lambda_k \geq \mu_k$. 这是因为在线社交网络中社团组织 e_k 的多重社团组织的用户不仅对外可以传播信息而且也在社团组织内部交换信息与保有信息, 所以它们比社团组织 e_k 中的单社团组织的用户在传播信息时所起的作用要大.

3 在不同超路径上信息传播强度及其概率分布分析

基于在线社交网络中社团组织的规模与它的重要度, 本节建立了从一个首端用户向一个末端用户传播信息时所经历的各个超路径及其拓扑结构, 并依据信息传播的概率行为给出了信息传播过程选择在不同超路径上的概率分布.

在线社交网络中, 首端用户 v_i 发出的信息必须通过超路径上的其他用户或者社团组织作为传播媒介才

能向末端用户 v_j 传播信息. 图 2 表示了一个信息传播过程及其所经历的一条超路径.

如果在线社交网络中包含有限个用户与有限个社团组织, 那么这个在线社交网络中从首端用户 v_i 到末端用户 v_j 至多包含有限条以社团组织 e_{i_1} 为开始的超路径. 将从首端用户 v_i 到末端用户 v_j 的以社团组织 e_{i_1} 为开始的所有不同超路径表示为一个集合如下

$$\Psi(v_i \xrightarrow{e_{i_1}} v_j) = \{SP^{e_{i_1}}(l_{q_1}^{e_{i_1}}), SP^{e_{i_1}}(l_{q_2}^{e_{i_1}}), \dots, SP^{e_{i_1}}(l_{q_s}^{e_{i_1}})\}, \quad (2)$$

其中 $SP^{e_{i_1}}(l_{q_i}^{e_{i_1}})$ 表示一条以社团组织 e_{i_1} 为开始的长度为 q_i 的超路径.

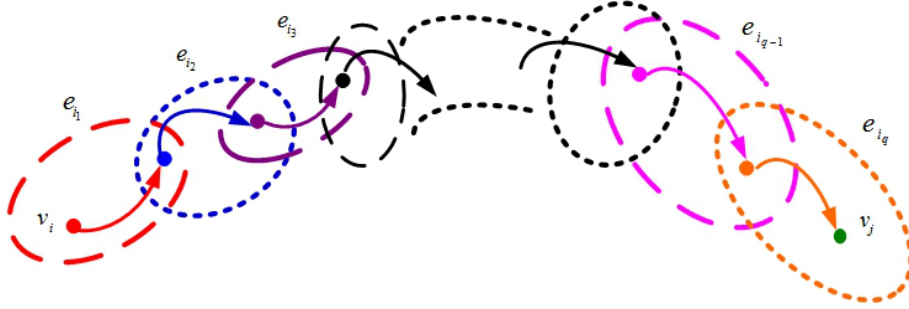


图 2 从首端用户 v_i 到末端用户 v_j 的一条长度为 q 的超路径

Fig. 2 A hyperpath of length q from sender v_i to receiver v_j

针对在线社交网络中的用户 v_i 与用户 v_j , 选取他们之间的一条长度为 q 的超路径, 在这条超路径上所包含的社团组织集合为 $E_{l_q}^{e_{i_1}} = \{e_{i_1}, e_{i_2}, \dots, e_{i_q}\}$, 用户集合为 $V_{l_q}^{e_{i_1}} = \{v_i, v_{i_1}, \dots, v_{i_{q-1}}, v_j\}$. 于是, 可以定义从首端用户 v_i 到末端用户 v_j 在超路径上传播信息强度为

$$f(SP^{e_{i_1}}(l_q^{e_{i_1}})) = \prod_{e_{i_k} \in E_{l_q}^{e_{i_1}}} \frac{|e_{i_k}|(\lambda_{i_k} a_{i_k})}{q^\alpha} \prod_{v_{i_g} \in V_{l_q}^{e_{i_1}} / \{v_i, v_j\}} \phi_{v_{i_g}}^{e_{i_1}}, \quad (3)$$

其中 1) $|e_{i_k}|$ 表示社团组织 e_{i_k} 中的用户个数. 实质上, 它反映了社团组织 e_{i_k} 在在线社交网络中的用户规模 (包括单社团组织的用户与多重社团组织的用户).

2) $\lambda_{i_k} a_{i_k}$ 表示社团组织 e_{i_k} 向外传播信息的能力 (简称传播力).

3) 由于社团组织的传播力与用户规模都对这个社团组织的传播信息强度产生影响, 所以采用一种简单度量 $|e_{i_k}|(\lambda_{i_k} a_{i_k})$ 来表达这两个因素对传播信息强度的同时影响.

4) 考虑到两个用户之间超路径长度会对信息传播产生一定的消极影响, 又为了避免将这种消极影响过分地扩大化, 因此引入了主观控制参数 $\alpha \in [0, 1]$. 例如, 若 $q = 1\,000$, $\alpha = \frac{1}{3}$, 则 $q^\alpha = 10$. 同时认为用户 v_i 与用户 v_j 之间的传播信息强度与所在超路径长度的修正值 q^α 成反比.

5) 在信息传播过程中, 如何选择信息传播的超路径依赖于这条超路径中的各个社团组织 (或用户) 对于这个要传播的信息的偏好程度, 而这种偏好性又对要传播的信息在整条超路径上的传播强度起着重要的影响. 例如, 传播信息的用户对于初始社团组织与要传播的信息有好感时就会选择积极向其他社团组织转发该信息, 而对于缺乏兴趣或反感的初始社团组织与要传播的信息转发积极性就会变小. 基于此, 针对初始社团组织以及要传播的信息, 就有必要引入一些正参数 $\phi_{v_{i_g}}^{e_{i_1}}$ 来表示这种偏好程度的一种数学度量如下: $\phi_{v_{i_g}}^{e_{i_1}}$, $v_{i_g} \in V_{l_q}^{e_{i_1}} - \{v_i, v_j\}$, 表示在超路径 $SP^{e_{i_1}}(l_q^{e_{i_1}})$ 上传播信息的用户 v_{i_g} 对于初始社团组织 e_{i_1} 与要传播信息的偏好程度.

6) 在 $f(SP^{e_{i_1}}(l_q^{e_{i_1}}))$ 中的乘积关系表示了这些不同社团组织以及它们中传播信息的用户对传播信息强度的同时影响是类似于相互独立的.

在式 (2) 中, 从首端用户 v_i 到末端用户 v_j 的所有超路径的集合为

$$\Psi(v_i \xrightarrow{e_{i_1}} v_j) = \{SP^{e_{i_1}}(l_{q_1}^{e_{i_1}}), SP^{e_{i_1}}(l_{q_2}^{e_{i_1}}), \dots, SP^{e_{i_1}}(l_{q_s}^{e_{i_1}})\};$$

再利用在长度为 q_k 的超路径 $\text{SP}^{e_{i_1}}(l_{q_k}^{e_{i_1}})$ 上的传播信息强度 $f(\text{SP}^{e_{i_1}}(l_{q_k}^{e_{i_1}}))$, 可得各个超路径上传播信息强度的比率, 即 (在信息传播过程中选择不同超路径的概率分布) 为

$$\Pr \left\{ v_i \xrightarrow{e_{i_1}} v_j, \text{SP}^{e_{i_1}}(l_{q_k}^{e_{i_1}}) \right\} = \frac{f(\text{SP}^{e_{i_1}}(l_{q_k}^{e_{i_1}}))}{\sum_{l=1}^s f(\text{SP}^{e_{i_1}}(l_{q_l}^{e_{i_1}}))}. \quad (4)$$

由于在线社交网络中的传播信息用户的兴趣具有多样性, 所以在这里假设: 一种类型的兴趣对应于在线社交网络中的一个初始社团组织. 因此, 当首端用户 v_i 所属的社团组织不唯一时, 则定义首端用户 v_i 所属的不同初始社团组织集合为

$$\Delta(v_i) = \{e_r \mid v_i \in e_r\}.$$

利用式 (2), 就可以得到从首端用户到末端用户的所有超路径集合为

$$\begin{aligned} \Phi(v_i \rightarrow v_j) &= \bigcup_{e_r \in \Delta(v_i)} \Psi(v_i \xrightarrow{e_r} v_j) \\ &= \bigcup_{e_r \in \Delta(v_i)} \left\{ \text{SP}^{e_r}(l_{q_1}^{e_r}), \text{SP}^{e_r}(l_{q_2}^{e_r}), \dots, \text{SP}^{e_r}(l_{q_{s_r}}^{e_r}) \right\}. \end{aligned}$$

对于这个超路径集合 $\Phi(v_i \rightarrow v_j)$, 图 3 描述了从首端用户 v_i 到末端用户 v_j 存在多条超路径的一个简单物理结构.

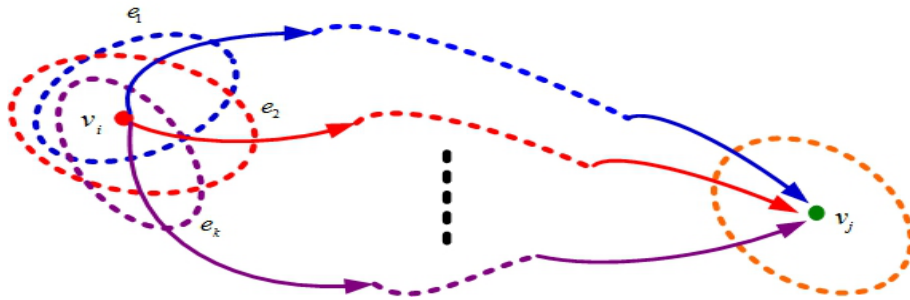


图 3 从首端用户 v_i 到末端用户 v_j 存在多条超路径集合的物理结构描述

Fig. 3 A physical structure description for multiple hyperpaths existing from sender v_i to receiver v_j

类似地, 可以给出在以 e_r 为初始社团组织的长度为 q_k 的超路径 $\text{SP}^{e_r}(l_{q_k}^{e_r})$ 上的传播信息强度为 $f(\text{SP}^{e_r}(l_{q_k}^{e_r}))$. 于是, 选取 $e_r \in \Delta(v_i)$, 可得在信息传播过程中选择不同超路径 $\text{SP}^{e_r}(l_{q_k}^{e_r})$ 的概率分布为

$$\Pr \left\{ v_i \xrightarrow{e_r} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\} = \frac{f(\text{SP}^{e_r}(l_{q_k}^{e_r}))}{\sum_{e_t \in \Delta(v_i)} \sum_{l=1}^{s_t} f(\text{SP}^{e_t}(l_{q_l}^{e_t}))}. \quad (5)$$

4 信息传播强度的波动性分析

基于在线社交网络中信息的传播过程概率行为与传播强度分布, 本节建立了首端用户发布信息的先验概率, 并对在线社交网络中信息传播过程中的强度波动性进行了概率分析.

4.1 首端用户发布信息的先验概率

在线社交网络中, 假设首端用户 v_i 向末端用户 v_j 通过初始社团组织为 e_r , 长度为 q_k 的一条超路径传播信息 C_{e_r} . 考虑到在线社交网络的结构复杂性, 末端用户 v_j 对于首端用户 v_i 所属社团组织的环境与现状并不十分了解. 为此有必要针对上一节中所建立的首端用户 v_i 所属的不同初始社团组织的集合 $\Delta(v_i)$ 给予进一步扩大这个容纳用户 v_i 的初始社团组织集合范围, 用 $\tilde{\Delta}(v_i)$ 来表示这样的扩大集合, 值

得注意: $\tilde{\Delta}(v_i)$ 中既包括已经确定的首端用户 v_i 所属的初始社团组织, 也包括需要通过统计推断等手段来判定的用户 v_i 所属于的初始社团组织. 因此 $\tilde{\Delta}(v_i) \supset \Delta(v_i)$, 并且 $\tilde{\Delta}(v_i) - \Delta(v_i)$ 中也包括还不能完全判定用户 v_i 所属的初始社团组织, 而这样的所属关系在实际在线社交网络中已经测定存在正的概率.

为了评估首端用户在这条超路径上发布信息的先验概率, 注意到这个先验概率不仅与其所属的初始社团组织 e_r 的规模以及社团组织 e_r 在整个在线社交网络中的重要度有关系, 而且还与首端用户 v_i 对初始社团组织 e_r 的偏好程度有关系. 基于此, 引入了一组关键参数: $\theta_{v_i}^{e_k}, 0 \leq \theta_{v_i}^{e_k} \leq 1, e_k \in E$, 来表示用户 v_i 对在线社交网络中第 k 个社团组织 e_k 的偏好程度.

用 $v_i \Rightarrow C_{e_r}$ 表示用户 v_i 通过初始社团组织 e_r 发布信息 C_{e_r} . 基于偏好程度参数 $\theta_{v_i}^{e_k}$ 、社团组织 e_k 的规模 $|e_k|$ 与重要度 $W(e_k)$, 就可以定义首端用户 v_i 通过初始社团组织 e_r 在任意一条超路径上发布信息 C_{e_r} 的先验概率为

$$\Pr(v_i \Rightarrow C_{e_r}) = \frac{\theta_{v_i}^{e_r} W(e_r) |e_r|}{\sum_{e_k \in \tilde{\Delta}(v_i)} \theta_{v_i}^{e_k} W(e_k) |e_k|}. \quad (6)$$

因为 $\tilde{\Delta}(v_i) \supset \Delta(v_i)$, 所以

$$\Pr(v_i \Rightarrow C_{e_r}) \leq \frac{\theta_{v_i}^{e_r} W(e_r) |e_r|}{\sum_{e_k \in \Delta(v_i)} \theta_{v_i}^{e_k} W(e_k) |e_k|}.$$

4.2 信息传播强度的波动性的概率分析

在线社交网络中, 选取一条以 e_{i_1} 为初始社团组织, 长度为 q 的从首端用户 v_i 到末端用户 v_j 的超路径 $\text{SP}^{e_{i_1}}(l_q^{e_{i_1}}) = \{v_i, e_{i_1}, v_{i_1}, e_{i_2}, v_{i_2}, \dots, v_{i_{q-1}}, e_{i_q}, v_j\}$, 其中 $v_i = v_{i_0}, v_j = v_{i_q}$. 现在描述从首端用户 v_i 到末端用户 v_j 在这条超路径上的信息传播过程如下: 首端用户 v_i 以先验概率 $\Pr(v_i \Rightarrow C_{e_{i_1}})$ 发布信息 $C_{e_{i_1}}$, 使社团组织 e_{i_1} 中的一个用户 v_{i_1} 接收到信息 $C_{e_{i_1}}$ 并转发到社团组织 e_{i_2} , 使 e_{i_2} 中的用户 v_{i_2} 接收到这一信息并转发到社团组织 e_{i_3} ; 按照这种方式, 可以一直向后续的社团组织中的用户传播信息 $C_{e_{i_1}}$; 用户 $v_{i_{q-1}}$ 将信息 $C_{e_{i_1}}$ 通过社团组织 e_{i_q} 向末端用户 v_j 传播, 由此末端用户 v_j 就接受到了信息 $C_{e_{i_1}}$. 图 4 描述了信息 $C_{e_{i_1}}$ 在超路径 $\text{SP}^{e_{i_1}}(l_q^{e_{i_1}})$ 的一个完整传播过程.

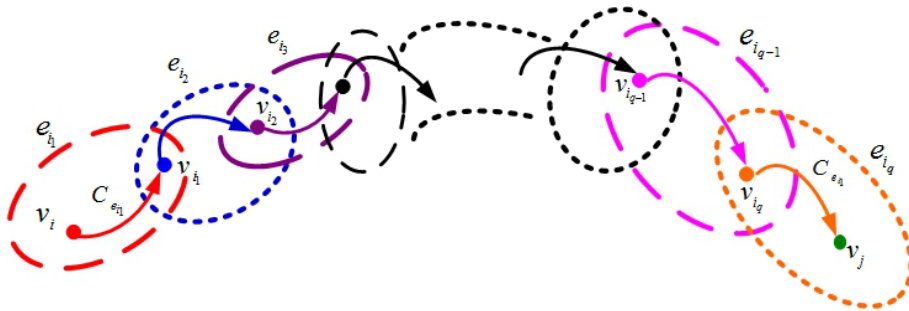


图 4 从首端用户 v_i 到末端用户 v_j 在一条超路径上的信息传播过程

Fig. 4 Information spreading processes on a hyperpath from sender v_i to receiver v_j

下面来研究在线社交网络的传播信息强度及其波动性. 由于在线社交网络的传播信息强度会依赖于信息传播过程、在超路径上的社团组织的规模与重要度等因素影响, 所以有必要建立从首端用户到末端用户在超路径上信息的传播强度, 并对它的波动性进行分析, 由此也可给出信息传播强度波动的概率分布.

首先考虑一条超路径

$$\{v_i, e_{i_1}, v_{i_1}, e_{i_2}, v_{i_2}, \dots, v_{i_{q-1}}, e_{i_q}, v_j\}, v_i = v_{i_0}, v_j = v_{i_q}, i_0, i_1, \dots, i_q \in \{1, 2, \dots, n\}.$$

对于一个社团组织 e , 它的传播力与规模分别为 λa 与 $|e|$. 为了建立信息传播强度的波动性, 引入社团

组织 e 的一个波动率

$$\beta_e = |1 - \gamma(e)| \in (0, +\infty), \quad (7)$$

其中 $\gamma(e)$ 是社团组织 e 将消息传播到下一个社团组织时信息传播强度的实际接收率.

在线社交网络中在从首端用户 v_i 到末端用户 v_j 的一条长度为 q 的超路径上, 用户 v_i 发布了一个信息, 并以社团组织集合 $E_{l_q}^{e_{i_1}} = \{e_{i_1}, e_{i_2}, \dots, e_{i_q}\}$ 为媒介, 通过传播信息的用户集合 $V_{l_q}^{e_{i_1}} - \{v_i, v_j\}$ 向用户 v_j 传播了这个信息. 于是, 可得从首端用户 v_i 到末端用户 v_j 在超路径 $\text{SP}^{e_{i_1}}(l_q^{e_{i_1}})$ 上的实际信息传播强度为

$$f_R(v_i \rightarrow v_j, \text{SP}^{e_{i_1}}(l_q^{e_{i_1}})) = \prod_{e_{i_k} \in E_{l_q}^{e_{i_1}}} \frac{\gamma(e_{i_k}) |e_{i_k}| (\lambda_{i_k} a_{i_k})}{q^\alpha} \prod_{v_{i_g} \in V_{l_q}^{e_{i_1}} - \{v_i, v_j\}} \phi_{v_{i_g}}^{e_{i_1}}. \quad (8)$$

类似地, 可以给出从用户 v_i 向用户 v_j 在超路径 $\text{SP}^{e_{i_1}}(l_q^{e_{i_1}})$ 上传播信息强度的波动值为

$$f_L(v_i \rightarrow v_j, \text{SP}^{e_{i_1}}(l_q^{e_{i_1}})) = \prod_{e_{i_k} \in E_{l_q}^{e_{i_1}}} \frac{\beta(e_{i_k}) |e_{i_k}| (\lambda_{i_k} a_{i_k})}{q^\alpha} \prod_{v_{i_g} \in V_{l_q}^{e_{i_1}} - \{v_i, v_j\}} \phi_{v_{i_g}}^{e_{i_1}}. \quad (9)$$

进一步地, 由于在线社交网络中从首端用户 v_i 向末端用户 v_j 的所有可能的超路径集合为

$$\Phi(v_i \rightarrow v_j) = \bigcup_{e_r \in \Delta(v_i)} \left\{ \text{SP}^{e_r}(l_{q_1}^{e_r}), \text{SP}^{e_r}(l_{q_2}^{e_r}), \dots, \text{SP}^{e_r}(l_{q_{s_r}}^{e_r}) \right\},$$

所以利用在线社交网络中信息传播强度的波动性, 可以得到从首端用户 v_i 向末端用户 v_j 的实际信息传播强度的概率分布为

$$\text{Pr}^R \{v_i \rightarrow v_j, \text{SP}^{e_r}(l_{q_k}^{e_r})\} = \frac{f_R(v_i \rightarrow v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}))}{\sum_{e_t \in \Delta(v_i)} \sum_{l=1}^{s_t} f_R(v_i \rightarrow v_j, \text{SP}^{e_t}(l_{q_l}^{e_t}))}. \quad (10)$$

考虑到首端用户 v_i 以先验概率 $\text{Pr}(v_i \Rightarrow C_{e_{i_1}})$ 从初始社团组织 e_r 开始发布信息 C_{e_r} , 因此, 可以表示从首端用户 v_i 在一条以初始社团组织 e_r 为开始的长度为 q_k 的超路径上向末端用户 v_j 传播信息 C_{e_r} 的概率分布如下

$$\text{Pr}^R \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\} = \text{Pr}(v_i \Rightarrow C_{e_r}) \text{Pr}^R \{v_i \rightarrow v_j, \text{SP}^{e_r}(l_{q_k}^{e_r})\}. \quad (11)$$

显然, $\text{Pr}^R \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\}$ 由传播信息 C_{e_r} 的先验概率与实际信息传播强度的概率分布这两部分联合构成.

结合第3节和第4节可知: 用户在不同超路径上的信息传播过程在各种因素的影响下有着不同的传播强度, 而这些在不同超路径上强度的比率就构成了用户在不同超路径上传播信息的概率分布.

另一方面, 从首端用户 v_i 向末端用户 v_j 传播信息强度波动的概率分布为

$$\text{Pr}^L \{v_i \rightarrow v_j, \text{SP}^{e_r}(l_{q_k}^{e_r})\} = \frac{f_L(v_i \rightarrow v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}))}{\sum_{e_t \in \Delta(v_i)} \sum_{l=1}^{s_t} f_L(v_i \rightarrow v_j, \text{SP}^{e_t}(l_{q_l}^{e_t}))}. \quad (12)$$

进一步地, 从首端用户 v_i 在超路径 $\text{SP}^{e_r}(l_{q_k}^{e_r})$ 上向末端用户 v_j 传播信息 C_{e_r} 的传播强度波动的概率分布为

$$\text{Pr}^L \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\} = \text{Pr}(v_i \Rightarrow C_{e_r}) \text{Pr}^L \{v_i \rightarrow v_j, \text{SP}^{e_r}(l_{q_k}^{e_r})\}. \quad (13)$$

5 隐性影响力的评估方法

本节利用在线社交网络中信息传播强度以及传播信息的先验概率建立了从首端用户到末端用户之间

的一个有向平均互信息. 基于平均互信息理论, 提出了在线社交网络中隐性影响力的一种有效评估方法.

5.1 在超路径上信息传播过程中的后验概率

基于在线社交网络中从首端用户到末端用户在超路径上信息的传播过程. 现在需要计算针对首端用户 v_i 向末端用户 v_j 在一条超路径上传播信息的条件概率

$$\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r} | v_j \Leftarrow C_{e_r}), \quad (14)$$

它表示从 e_r 为初始社团组织, 开始传播信息并经过长度为 q_k 的一条超路径, 在末端用户 v_j 接收到这个信息 C_{e_r} 的条件下, 信息 C_{e_r} 是由首端用户 v_i 发布的条件概率.

根据贝叶斯公式, 可得

$$\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r} | v_j \Leftarrow C_{e_r}) = \frac{\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_j \Leftarrow C_{e_r} | v_i \Rightarrow C_{e_r}) \Pr(v_i \Rightarrow C_{e_r})}{\sum_{l=1}^{S_r} \Pr^{\text{SP}^{e_r}(l_{q_l}^{e_r})}(v_j \Leftarrow C_{e_r} | v_i \Rightarrow C_{e_r}) \Pr(v_i \Rightarrow C_{e_r})}. \quad (15)$$

在式 (15) 中, 这个待定的条件概率 $\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_j \Leftarrow C_{e_r} | v_i \Rightarrow C_{e_r})$ 需要首先给出.

由于 $\Pr^R \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\}$ 与 $\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r}, v_j \Leftarrow C_{e_r})$ 在概率意义上是相同的, 所以可得

$$\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r}, v_j \Leftarrow C_{e_r}) = \Pr^R \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\}.$$

因为

$$\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r}, v_j \Leftarrow C_{e_r}) = \Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_j \Leftarrow C_{e_r} | v_i \Rightarrow C_{e_r}) \Pr(v_i \Rightarrow C_{e_r}),$$

所以

$$\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_j \Leftarrow C_{e_r} | v_i \Rightarrow C_{e_r}) = \frac{\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r}, v_j \Leftarrow C_{e_r})}{\Pr(v_i \Rightarrow C_{e_r})}.$$

于是, 可以得到

$$\Pr^{\text{SP}^{e_r}(l_{q_k}^{e_r})}(v_i \Rightarrow C_{e_r} | v_j \Leftarrow C_{e_r}) = \frac{\Pr^R \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_k}^{e_r}) \right\}}{\sum_{l=1}^{S_r} \Pr^R \left\{ v_i \xrightarrow{C_{e_r}} v_j, \text{SP}^{e_r}(l_{q_l}^{e_r}) \right\}}. \quad (16)$$

5.2 两个用户之间的平均互信息

在信息理论中, 产生信息的事物被看作为信源, 而传输信息的接收者被称为信宿. 由于在线社交网络中每个用户可以发布或接受信息, 所以在线社交网络中的首端用户可以看作为一个信源; 而末端用户可以看作为一个信宿. 于是, 在线社交网络中从首端用户到末端用户之间的任意一条超路径可看作为在信源与信宿之间的一个信道. 依据信息论 (例如, 傅祖芸^[48], 第 76 页~第 80 页) 本节提出了在线社交网络中的一种有向的平均互信息, 它可用于对信道中传输信息量的一个整体测度. 因此, 这个有向的平均互信息能够用来评估在线社交网络中的隐性影响力.

在线社交网络中, 基于在超路径上的信息传播过程, 当对首端用户 v_i 向末端用户 v_j 的传播信息量进行评估时, 首端用户 v_i 是一个信源用户, 而用户 v_j 是一个信宿用户. 在下面分析首端用户 v_i 传播信息前后实际环境发生变化的两个阶段:

a) 在信源用户 v_i 向信宿用户 v_j 发出信息之前, 用户 v_j 没有关于用户 v_i 的任何信息, 例如用户 v_j 不知道用户 v_i 在哪个社团组织、也不知道用户 v_i 会发布什么样的信息. 因此, 用户 v_j 存在对于用户 v_i 了解上的诸多不确定性. 为了描述与分析用户 v_j 获得关于用户 v_i 的有关信息量, 用熵 $H(v_i)$ 来表示这种不确定性, 其中

$$H(v_i) = - \sum_{e_s \in E} \Pr(v_i \Rightarrow C_{e_r}) \ln \Pr(v_i \Rightarrow C_{e_r}). \quad (17)$$

b) 当用户 v_i 已经向用户 v_j 发出了信息之后, 用户 v_j 可以得到用户 v_i 的部分信息, 但这只是对用户 v_i 有了一定程度 (非完全) 的了解. 因此, 在这种情况下用户 v_j 仍然存在对用户 v_i 了解上的不确定性, 而这种不确定性可用条件熵 $H(v_i | v_j)$ 来表示如下

$$H(v_i | v_j) = - \sum_{e_r \in \Delta(v_i)} \sum_{k=1}^{S_r} \Pr^{\text{SP}^{e_r}}(l_{q_k}^{e_r}) (v_i \Rightarrow C_{e_r}, v_j \Leftarrow C_{e_r}) \ln \Pr^{\text{SP}^{e_r}}(l_{q_k}^{e_r}) (v_i \Rightarrow C_{e_r} | v_j \Leftarrow C_{e_r}). \quad (18)$$

注意到两熵之差 $H(v_i) - H(v_i | v_j)$ 是在信息传播过程完成后末端用户 v_j 对于首端用户 v_i 的不确定性的减少量, 即在信息传播后用户 v_j 获得了关于用户 v_i 的信息量, 实质上这是一个有向平均互信息如下:

$$I(v_i; v_j) = H(v_i) - H(v_i | v_j). \quad (19)$$

这个有向的平均互信息 $I(v_i; v_j)$ 可用来评估在线社交网络中首端用户 v_i 对于末端用户 v_j 的隐性影响力.

综上所述, 在线社交网络中的社团组织是由各个不同兴趣话题所引导的用户连接关系的集聚, 而某个社团组织的相关信息不仅决定了这个社团组织在整个网络中的属性与地位, 还决定了该社团内部的用户发布其相关信息能力的概率分布. 就导致当不同社团组织作为每条超路径的初始社团组织时, 其内部的用户以该社团组织所决定的概率分布来发布这条信息, 就形成了不同的先验概率与后验概率, 并用来表述传播信息前后的不确定性, 最终形成了在线社交网络中的不同隐性影响力.

5.3 数值算例

本小节使用一些数值算例来验证本文所提出的在线社交网络的隐性影响力评估方法具有可计算性、易操作性与有效性.

首先利用 MATLAB 随机生成了一个 10×10 的关联矩阵 S , 它表示一个包含 10 个用户, 并由 10 个不同的兴趣话题引导了这些用户集聚形成 10 个重叠社团组织的在线社交网络,

$$S = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

针对上述在线社交网络, 使用 MATLAB 遍历得到了两个不同用户之间的超路径; 再利用本文所给出的隐性影响力评估方法计算了 90 个用户对中针对不邻接用户对的隐性影响力以及针对邻接用户对的显性影响力. 在下面提供两组数值实验 A 与 B, 由此来观察系统的主要参数对结果产生的影响.

数值实验 A: 在这个数值计算过程中, 为了避免主观因素影响过多, 系统的参数设计如下: 10 个社团组织的对外传播力参数为 $\lambda_1 = \lambda_2 = \dots = \lambda_{10} = 0.9$ 与对内交换力参数为 $\mu_1 = \mu_2 = \dots = \mu_{10} = 0.1$; 当不存在单社团组织的用户时, 设 $\lambda_1 = \lambda_2 = \dots = \lambda_{10} = 1$ 和 $\mu_1 = \mu_2 = \dots = \mu_{10} = 0$. 信源用户对 10 个社团组织的不同偏好程度为 $\theta_{v_i}^{e_r} = 1, e_r \in \Delta(v_i)$ 与 $\theta_{v_i}^{e_z} = 0.5, e_z \in E - \Delta(v_i)$. 传播信息的用户对于初始社团组织与要传播的信息具有的偏好程度为 $\phi_{v_{ig}}^{e_r} = 0.9, v_{ig} \in e_r$; $\phi_{v_{ig}}^{e_z} = 0.1, v_{ig} \in E - e_r$. 每个社团组织的波动率设为 $\beta = 0.2$. 每条超路径的长度 q 对于信息传播强度的影响系数设为 q^α , 其中 $\alpha = 0.5$.

依据上述设定的系统参数, 利用 MATLAB 进行数值计算, 得到了在线社交网络中每个用户向其他用户传播信息时产生的隐性 (或显性) 影响力, 由于篇幅有限, 在本节仅以用户 4、用户 5、用户 6、用户 7 为代表, 他们分别对其他 9 个用户的影响力评估结果如图 5 所示:

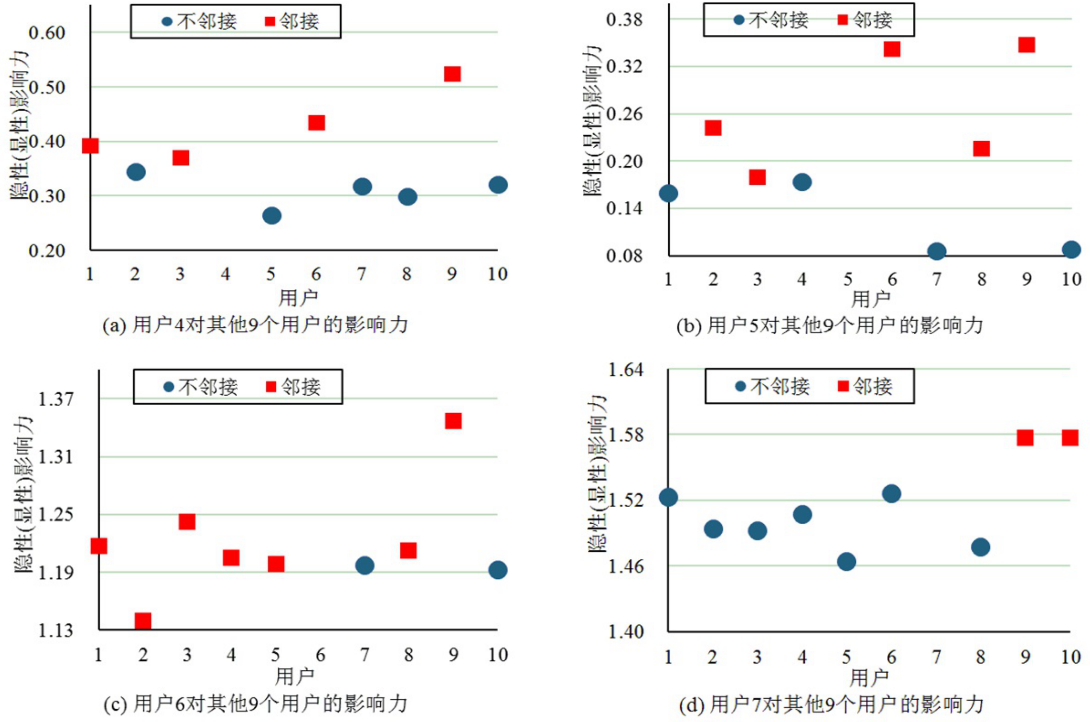


图 5 用户 4、用户 5、用户 6、用户 7 对其他 9 个用户的隐性 (或显性) 影响力

Fig. 5 Hidden (or explicit) influence of user 4, user 5, user 6, user 7 with respect to other users

根据实验全部结果可得用户 6、用户 7、用户 9、用户 10 对其他用户的隐性 (或显性) 影响力是比较大的;而那些其他用户对外影响力的峰值大多数出现在用户 6 或用户 9 上. 由此可以推断出用户 6 与用户 9 在这个在线社交网络中较为活跃, 向外传播信息以及接收信息的能力都比较强.

数值实验 B: 相比于数值实验 A, 数值实验 B 考虑了三组不同的参数变化来进一步地观察用户的影响力变化情况.

值得注意: 这里未提及的其他参数对比与数值实验 A 保持不变, 即它们仍是数值实验 A 的对应参数.

参数组 1: 10 个社团的对外传播力与对内交换力参数分别变为 $\lambda_1 = \lambda_2 = \dots = \lambda_{10} = 0.8$ 与 $\mu_1 = \mu_2 = \dots = \mu_{10} = 0.2$.

参数组 2: 每条超路径上各个传播信息的用户偏好分别变为 $\phi_{v_{i_g}}^{e_r} = 0.9, v_{i_g} \in e_r; \phi_{v_{i_g}}^{e_z} = 0.2, v_{i_g} \in E - e_r$.

参数组 3: 在每条超路径上, 各个社团组织对于信息传播的波动强度系数变为 $\beta = 0.5$.

限于篇幅, 这里仅显示用户 6 与用户 9 的数值实验结果, 如图 6 所示.

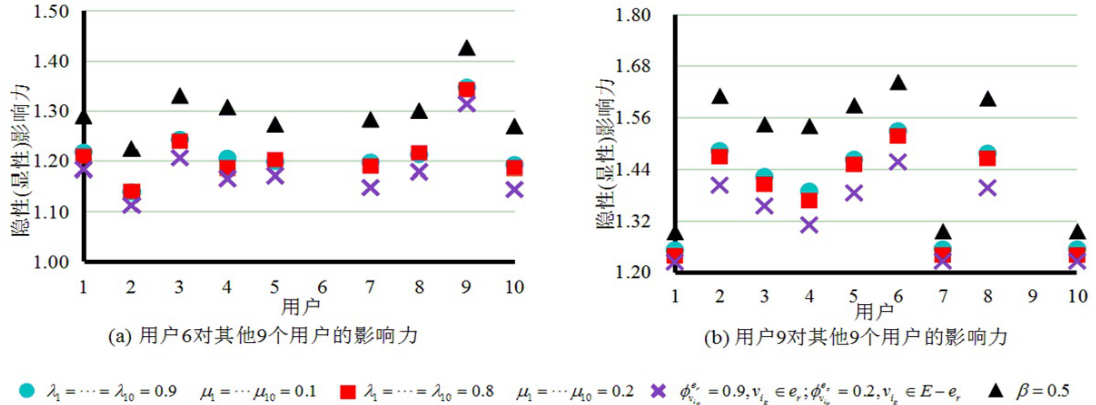


图 6 用户 6、用户 9 分别对其他用户的隐性 (或显性) 影响力

Fig. 6 Hidden (or explicit) influence of user 6, user 9 with respect to other users

虽然系统中的一些参数变化对用户之间的隐性(或显性)影响力有一定的影响,但是它们的整体变化趋势仍然保持与实验A的较高相似性。

综上所述,在包含10个用户与10个社团组织的在线社交网络中,通过对90个用户对之间的影响力评估,通过数值分析验证了本文所提出的隐性影响力评估方法具有可计算性、易操作性与有效性。同时,这种基于信息传播的影响力评估方法不仅具备量化分析影响力的作用,而且还能够检测到在线社交网络中的关键用户。

6 结束语

本文利用超图理论与方法通过超路径与概率行为建立了在线社交网络中用户或社团组织之间的信息传播过程,利用任意两个用户之间传播的信息量来量化评估隐性影响力。数值算例结果表明:本文建立的隐性影响力评估方法不仅能够不考虑在线社交网络中用户之间的邻接关系的条件下,评估它们之间的影响力(显性与隐性),还能够识别出在线社交网络中的关键用户与重要社团组织。这就为隐性影响力评估提出了一种新型的数学模型框架与新的数值计算方法。

参考文献:

- [1] Zhou B, He Z, Jiang L L, et al. Bidirectional selection between two classes in complex social networks. *Scientific Reports*, 2014, 4: 7577.
- [2] 郑双阳, 林锦贤. 基于用户行为聚类的搜索. *计算机与数字工程*, 2009, 37(12): 28–30.
Zheng S Y, Lin J X. A search engine based on user behavior clustering. *Computer & Digital Engineering*, 2009, 37(12): 28–30. (in Chinese)
- [3] Zhao Z D, Yang Z M, Zhang Z K, et al. Emergence of scaling in human-interest dynamics. *Scientific Reports*, 2013, 3: 3472.
- [4] 徐志明, 李 栋, 刘 挺, 等. 微博用户的相似性度量及其应用. *计算机学报*, 2014, 37(1): 207–218.
Xu Z M, Li D, Liu T, et al. Measuring similarity between microblog users and its application. *Chinese Journal of Computers*, 2014, 37(1): 207–218. (in Chinese)
- [5] 李金双, 李凤云. 移动社会网络中用户相似性计算的隐私保护研究. *电信科学*, 2013, 29(10): 119–122.
Li J S, Li F Y. Protected private information for proximity computing in mobile social network. *Telecommunications Science*, 2013, 29(10): 119–122. (in Chinese)
- [6] 李倩倩, 顾基发. 用户行为驱动的在线社交网络建模. *系统工程学报*, 2015, 30(1): 9–15.
Li Q Q, Gu J F. Activity driven modelling of online social network. *Journal of Systems Engineering*, 2015, 30(1): 9–15. (in Chinese)
- [7] Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks// *ACM International Conference on Knowledge Discovery and Data Mining*. ACM, 2008: 7–15.
- [8] Yue W, Huang W J, Zong L, et al. Influence maximization with limit cost in social network. *Science China*, 2013, 56(7): 1–14.
- [9] 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. *计算机学报*, 2015, 38(2): 238–248.
Cao J X, Dong D, Xu S, et al. A k-core based algorithm for influence maximization in social networks. *Chinese Journal of Computers*, 2015, 38(2): 238–248. (in Chinese)
- [10] 李国良, 楚娅萍, 冯建华, 等. 多社交网络的影响力最大化分析. *计算机学报*, 2016, 39(4): 643–656.
Li G L, Chu Y P, Feng J H, et al. Influence maximization on multiple social networks. *Chinese Journal of Computers*, 2016, 39(4): 643–656. (in Chinese)
- [11] Salnikov V, Schaub M T, Lambiotte R. Using higher-order Markov models to reveal flow-based communities in networks. *Scientific Reports*, 2016, 6: 23194.
- [12] 王 莉, 程学旗. 在线社会网络的动态社区发现及演化. *计算机学报*, 2015, 38(2): 219–237.
Wang L, Cheng X Q. Dynamic community in online social networks. *Chinese Journal of Computers*, 2015, 38(2): 219–237. (in Chinese)
- [13] 谢晓芹, 韩 帅, 陈 敏, 等. 基于社会网络的团队生成方法研究. *计算机学报*, 2017, 40(3): 712–728.
Xie X Q, Han S, Chen M, et al. Research on team formation method in social network. *Chinese Journal of Computers*, 2017, 40(3): 712–728. (in Chinese)

- [14] 陈羽中, 施 松, 朱伟平, 等. 一种基于邻域跟随关系的增量社区发现算法. 计算机学报, 2017, 40(3): 570–583.
Chen Y Z, Shi S, Zhu W P, et al. An incremental community discovery algorithm based on neighborhood following relationship. Chinese Journal of Computers, 2017, 40(3): 570–583. (in Chinese)
- [15] 林旺群, 邓 镭, 丁兆云, 等. 一种新型的层次化动态社区并行计算方法. 计算机学报, 2012, 35(8): 1712–1725.
Lin W Q, Deng L, Ding Z Y, et al. Hierarchical dynamic community detection by parallel computing. Chinese Journal of Computers, 2012, 35(8): 1712–1725. (in Chinese)
- [16] Rosvall M, Esquivel A V, Lancichinetti A, et al. Memory in network flows and its effects on spreading dynamics and community detection. Nature Communications, 2014, 5: 4630.
- [17] Bakshy E, Karrer B, Adamic L A. Social influence and the diffusion of user-created content// ACM Conference on Electronic Commerce. 2009: 325–334.
- [18] 周东浩, 韩文报. DiffRank: 一种新型社会网络信息传播检测算法. 计算机学报, 2014, 37(4): 884–893.
Zhou D H, Han W B. DiffRank: A novel algorithm for information diffusion detection in social networks. Chinese Journal of Computers, 2014, 37(4): 884–893. (in Chinese)
- [19] Pei S, Muchnik L, Andrade Jr J S, et al. Searching for superspreaders of information in real-world social media. Scientific Reports, 2014, 4:5547.
- [20] 蔡淑琴, 袁 乾, 周 鹏. 企业响应下负面口碑线性阈值传播模型研究. 系统工程学报, 2017, 32(2): 145–155.
Cai S Q, Yuan Q, Zhou P. Research on linear threshold diffusion model for negative word-of-mouth under enterprises response. Journal of Systems Engineering, 2017, 32(2): 145–155. (in Chinese)
- [21] Berge C. Hypergraphs: Combinatorics of Finite Sets. North Holland: Elsevier, 1989.
- [22] 李 威, 李宏岩, 米守防. 基于有权重超图模型的离群点发现. 大连民族大学学报, 2006, 8(5): 45–48.
Li W, Li H Y, Mi S F. Research on outlier testing method based on weighted hypergraph-model. Journal of Dalian Minzu University, 2006, 8(5): 45–48. (in Chinese)
- [23] 张 强, 李永丽, 董立岩, 等. 基于有权重超图的离群点检测. 吉林大学学报, 2007, 45(4): 611–616.
Zhang Q, Li Y L, Dong L Y, et al. Outlier testing methods based on weighted hypergraph. Journal of Jilin University, 2007, 45(4): 611–616. (in Chinese)
- [24] 胡宝丽, 游进国, 周翠莲, 等. 一种有效的加权图聚集算法. 中国科学技术大学学报, 2016(3): 180–187.
Hu B L, You J G, Zhou C L, et al. An efficient weighted graph aggregation algorithm. Journal of University of Science and Technology of China, 2016(3): 180–187. (in Chinese)
- [25] Zhou D, Huang J, Scholkopf B. Learning with hypergraphs: Clustering, classification, and embedding// Advances in Neural Information Processing Systems. 2007: 1601–1608.
- [26] 陈建斌, 宋翰涛. 基于属性分布相似度的超图高维聚类算法研究. 计算机工程与应用, 2004, 40(34): 195–198.
Chen J B, Song H T. A new clustering method in high-dimension based on sad hypergraph models. Computer Engineering and Applications, 2004, 40(34): 195–198. (in Chinese)
- [27] Suo Q, Sun S W, Hajli N, et al. User ratings analysis in social networks through a hypernetwork method. Expert Systems with Applications, 2015, 42(21): 7317–7325.
- [28] Du M. Research on information dissemination model of social network services based on probabilistic hyper-graph. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2015, 8(6): 267–274.
- [29] Gangal V, Ravindran B, Narayanam R. HEMI: Hyperedge majority influence maximization// Proceedings of the 2nd International Workshop on Social Influence Analysis. New York, 2016: 38–47.
- [30] Tao H, Wu Z, Shi J, et al. Overlapping community extraction: A link hypergraph partitioning based method// IEEE International Conference on Services Computing. New York: IEEE Press, 2014: 123–130.
- [31] Yang W, Wang G, Bhuiyan M Z A. Partitioning of hypergraph modeled complex networks based on information entropy// International Conference on Algorithms and Architectures for Parallel Processing. Springer International Publishing, 2015: 678–690.
- [32] Zhang Z K, Liu C. Hypergraph model of social tagging networks. Journal of Statistical Mechanics: Theory and Experiment, 2012, 2010, 10: P10005.
- [33] 肖玉芝, 赵海兴. 基于超图理论的在线社会网络用户行为分析. 计算机应用与软件, 2014, 31(7): 50–54.
Xiao Y Z, Zhao H X. Analysing users behaviour in online social network based on hypergraph theory. Computer Applications and Software, 2014, 31(7): 50–54. (in Chinese)
- [34] Fang Q, Sang J, Xu C, et al. Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning. IEEE Transactions on Multimedia, 2014, 16(3): 796–812.

- [35] You Z Q, Han X P, Lu L, et al. Empirical studies on the network of social groups: The case of Tencent QQ. *Plos One*, 2015, 10(7): 1–18.
- [36] 吴信东, 李 毅, 李 磊. 在线社交网络影响力分析. *计算机学报*, 2014, 37(4): 735–752.
Wu X D, Li Y, Li L. Influence analysis of online social networks. *Chinese Journal of Computers*, 2014, 37(4): 735–752. (in Chinese)
- [37] Shuai X, Ding Y, Busemeyer J, et al. Modeling indirect influence on Twitter. *International Journal on Semantic Web & Information Systems*, 2012, 8(4): 20–36.
- [38] Iwata T, Shah A, Ghahramani Z. Discovering latent influence in online social activities via shared cascade poisson processes// *ACM International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2013: 266–274.
- [39] Liu L, Tang J, Han J, et al. Mining topic-level influence in heterogeneous networks// *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. New York: ACM, 2010: 199–208.
- [40] Cui P, Wang F, Liu S, et al. Who should share what? Item-level social influence prediction for users and posts ranking// *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 2011: 185–194.
- [41] Steeg G V, Galstyan A. Information-theoretic measures of influence based on content dynamics// *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*. New York: ACM, 2013: 3–12.
- [42] Steeg G V, Galstyan A. Information transfer in social media// *Proceedings of the 21st International Conference on World Wide Web*. New York: ACM, 2012: 509–518.
- [43] 王祯骏, 王树徽, 张维刚, 等. 基于社交内容的潜在影响力传播模型. *计算机学报*, 2016, 39(8): 1528–1540.
Wang Z J, Wang S H, Zhang W G, et al. Social content based latent influence propagation model. *Chinese Journal of Computers*, 2016, 39(8): 1528–1540. (in Chinese)
- [44] Wasserman S, Faust K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994.
- [45] Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821–7826.
- [46] Clauset A. Finding local community structure in networks. *Physical Review E*, 2005, 72(2): 1–7.
- [47] 龚 勔, 程 绩. 超图的最短路径算法. *重庆大学学报*, 2005, 28(11): 106–109.
Gong Q, Cheng J. Shortest path algorithm for hypergraphs. *Journal of Chongqing University*, 2005, 28(11): 106–109. (in Chinese)
- [48] 傅祖芸. 信息论: 基础理论与应用. 第2版. 北京: 电子工业出版社, 2007.
Fu Z Y. *Information Theory: Principles and Applications*. 2nd Edition. Beijing: Publishing House of Electronics Industry, 2007. (in Chinese)

作者简介:

李泉林 (1964—), 男, 河北人, 博士, 教授, 博士生导师, 研究方向: 随机模型、计算机网络、社会网络、复杂网络、物流与供应链管理, Email: liquanlin@tsinghua.edu.cn;

李超然 (1993—), 女, 吉林人, 硕士生, 研究方向: 随机模型、社会网络、排队网络, Email: lichaoran0415@stumail.ysu.edu.cn;

马静宇 (1982—), 女, 黑龙江人, 博士生, 研究方向: 数据中心网络、排队论及其应用、马氏决策过程, Email: mjj0501@126.com.