



(12) 发明专利

(10) 授权公告号 CN 106815653 B

(45) 授权公告日 2021.01.12

(21) 申请号 201611182815.1

G06Q 50/00 (2012.01)

(22) 申请日 2016.12.19

审查员 吴琳

(65) 同一申请的已公布的文献号
申请公布号 CN 106815653 A

(43) 申请公布日 2017.06.09

(73) 专利权人 烟台中科网络技术研究所
地址 264003 山东省烟台市高新区蓝海路1号蓝海国际软件园A座

(72) 发明人 刘大伟 柯枫 刘玮 隋雪青
程学旗

(74) 专利代理机构 北京轻创知识产权代理有限公司 11212

代理人 王澎

(51) Int. Cl.

G06Q 10/04 (2012.01)

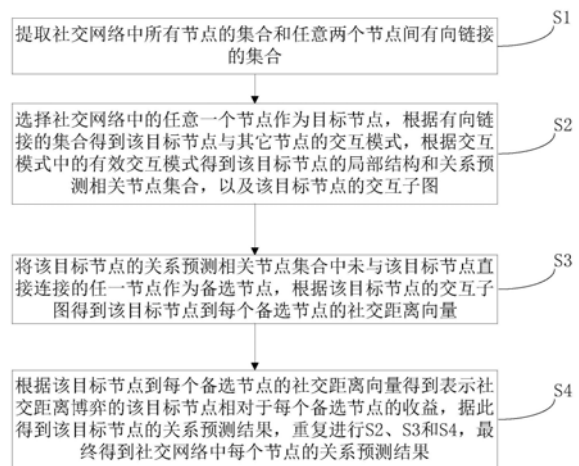
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种基于距离博弈的社交网络关系预测方法及系统

(57) 摘要

本发明涉及一种基于距离博弈的社交网络关系预测方法及系统,属于社交网络关系预测领域,能提高关系预测的速度和准确度。本发明先提取社交网络中所有节点和任意两个节点间的有向链接;选择其中一个节点作为目标节点,根据有向链接和有效交互模式得到该目标节点的局部结构、关系预测相关节点集合和交互子图;将关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,根据交互子图得到到每个备选节点的社交距离向量;根据社交距离向量得到相对于每个备选节点的收益,据此得到该目标节点的关系预测结果,最终得到社交网络中每个节点的关系预测结果。用于高效准确进行社交网络关系预测。



1. 一种基于距离博弈的社交网络关系预测方法,其特征在于,包括以下步骤:

S1,提取社交网络中所有节点的集合和任意两个节点间有向链接的集合;

S2,选择社交网络中的任意一个节点作为目标节点,根据有向链接的集合得到该目标节点与其它节点的交互模式,根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合,以及该目标节点的交互子图;

S3,将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,根据该目标节点的交互子图得到该目标节点到每个备选节点的社交距离向量;

S4,根据该目标节点到每个备选节点的社交距离向量得到表示社交距离博弈的该目标节点相对于每个备选节点的收益,据此得到该目标节点的关系预测结果,重复进行S2、S3和S4,最终得到社交网络中每个节点的关系预测结果;

所述S2中,所述有效交互模式包括:直接交互模式、间接交互模式和共性交互模式;

所述S3中,得到所述社交距离向量的过程为:将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,将除去该备选节点以外的任一节点作为中间节点,根据该目标节点的交互子图得到目标节点与该备选节点在该中间节点维度上的社交距离,进而根据在每个中间节点维度上的社交距离得到目标节点与该备选节点的社交距离向量,同理得到该目标节点与关系预测相关节点集合中未与该目标节点直接连接的其它节点作为备选节点时的社交距离向量;每一个在目标节点的局部结构中除备选节点以外的节点将作为一个社交距离概念下的中间节点 c_i ,该目标节点 x 到该备选节点 y 在该中间节点 c_i 维度上的社交距离用如下三元组来定义: $SD(x, c_i, y) = (\sum_{x_k \in N^*} w_{x_k, c_i} \cdot shortestpath(c_i, y))$,

其中权重 w_{x_k, c_i} 的和用来刻画该目标节点 x 的交互子图中的每个节点 x_k 与该中间节点 c_i 的交互特性,备选节点 y 到中间节点 c_i 的最短路径用来刻画结构特征;该社交距离为从该中间节点 c_i 的视角看,该目标节点 x 到该备选节点 y 的接近程度;该目标节点 x 的局部结构有 m 个中间节点,全维度的社交距离向量定义为: $SD_{x,y} = [SD(x, c_1, y), \dots, SD(x, c_m, y)]$,其中 c_1, c_2, \dots, c_m 在节点 x 的局部结构节点集合中。

2. 根据权利要求1所述的基于距离博弈的社交网络关系预测方法,其特征在于,所述S4中,根据所述收益得到所述关系预测结果的过程为:设定表示关系预测结果的节点集合中的节点数量,根据所述目标节点相对于每个备选节点的收益,用社会效益表示该目标节点相对于该数量的备选节点的收益之和,得到使该目标节点的社会效益最大的节点集合,即为该目标节点的关系预测结果。

3. 一种基于距离博弈的社交网络关系预测系统,其特征在于,所述系统包括:

社交网络关系提取模块,用于提取社交网络中所有节点的集合和任意两个节点间有向链接的集合;

目标节点关系提取模块,选择社交网络中的任意一个节点作为目标节点,根据有向链接的集合得到该目标节点与其它节点的交互模式,根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合,以及该目标节点的交互子图;

社交距离向量计算模块,用于将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,根据该目标节点的交互子图得到该目标节点到每个

备选节点的社交距离向量；

目标节点关系预测模块,用于根据该目标节点到每个备选节点的社交距离向量得到表示社交距离博弈的该目标节点相对于每个备选节点的收益,据此得到该目标节点的关系预测结果;

社交网络关系预测模块,用于重复调用目标节点关系提取模块、社交距离向量计算模块和目标节点关系预测模块,最终得到社交网络中每个节点的关系预测结果;

所述目标节点关系提取模块中,所述有效交互模式包括:直接交互模式、间接交互模式和共性交互模式;

所述社交距离向量计算模块用于:将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,将除去该备选节点以外的任一节点作为中间节点,根据该目标节点的交互子图得到目标节点与该备选节点在该中间节点维度上的社交距离,进而根据在每个中间节点维度上的社交距离得到目标节点与该备选节点的社交距离向量,同理得到该目标节点与关系预测相关节点集合中未与该目标节点直接连接的其它节点作为备选节点时的社交距离向量;每一个在目标节点的局部结构中除备选节点以外的节点将作为一个社交距离概念下的中间节点 c_i ,该目标节点 x 到该备选节点 y 在该中间节点 c_i 维度上的社交距离用如下三元组来定义: $SD(x, c_i, y) = (\sum_{x_k \in N'} w_{x_k, c_i} \bullet shortestpath(c_i, y))$,其中权重 w_{x_k, c_i} 的和用来刻画该目标节点 x 的交互子图中的每个节点 x_k 与该中间节点 c_i 的交互特性,备选节点 y 到中间节点 c_i 的最短路径用来刻画结构特征;该社交距离为从该中间节点 c_i 的视角看,该目标节点 x 到该备选节点 y 的接近程度;该目标节点 x 的局部结构有 m 个中间节点,全维度的社交距离向量定义为: $SD_{x,y} = [SD(x, c_1, y), \dots, SD(x, c_m, y)]$,其中 $c_1, c_2 \dots c_m$ 在节点 x 的局部结构节点集合中。

4. 根据权利要求3所述的基于距离博弈的社交网络关系预测系统,其特征在于,所述目标节点关系预测模块还用于:设定表示关系预测结果的节点集合中的节点数量,根据所述目标节点相对于每个备选节点的收益,用社会效益表示该目标节点相对于该数量的备选节点的收益之和,得到使该目标节点的社会效益最大的节点集合,即为该目标节点的关系预测结果。

一种基于距离博弈的社交网络关系预测方法及系统

技术领域

[0001] 本发明涉及社交网络关系预测领域。

背景技术

[0002] 近年来, 社会网络得到了学术界和业界的广泛关注, 尤其是针对用户交互, 以及交互的结构属性的研究层出不穷。其中关系预测是一个研究热点问题, 关系预测是指利用网络中不同节点的网络结构和/或属性信息来预测未来可能出现的链接, 如果将社交网络看做一个网络, 那么网络中的节点就是用户, 网络中的链路就是用户之间的关系, 用户之间关系的预测即网络中链路的预测。目前各国研究者提出了多种结构模型和关系模型, 包括基于特征的分类方法, 基于核的方法, 矩阵分解方法以及概率图方法等。大部分现有技术关注两个节点之间于关联模式的形成机制, 并利用监督或非监督方法对结构信息进行建模。很少技术考虑现实在线社会网络的规模巨大的实际问题。相比于学术合作、生物网络等其他网络数据集, 社会网络尤其是微博网络更为复杂, 包含了大量动态交互的关联关系, 而不同类型的关系必须分别进行建模分析, 同时要保证算法的计算可扩展性, 这也是微博网络链路预测的主要挑战。在线社会网络计算领域, 关系预测问题面临的一个挑战是网络的规模巨大。现有工作提出的度量方法在大规模网络的背景下缺乏对于计算性能的考虑。

发明内容

[0003] 本发明所要解决的技术问题是提供一种基于距离博弈的社交网络关系预测方法及系统, 目的在于使社交网络关系预测过程的计算更快速, 且预测结果更准确。

[0004] 本发明解决上述技术问题的技术方案如下:

[0005] 一种基于距离博弈的社交网络关系预测方法, 包括以下步骤:

[0006] S1, 提取社交网络中所有节点的集合和任意两个节点间有向链接的集合;

[0007] S2, 选择社交网络中的任意一个节点作为目标节点, 根据有向链接的集合得到该目标节点与其它节点的交互模式, 根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合, 以及该目标节点的交互子图;

[0008] S3, 将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点, 根据该目标节点的交互子图得到该目标节点到每个备选节点的社交距离向量;

[0009] S4, 根据该目标节点到每个备选节点的社交距离向量得到表示社交距离博弈的该目标节点相对于每个备选节点的收益, 据此得到该目标节点的关系预测结果, 重复进行S2、S3和S4, 最终得到社交网络中每个节点的关系预测结果。

[0010] 本发明的有益效果是: 利用社交网络的拓扑结构信息结合社会学理论, 提出社交距离向量等新的相似度指标, 采取非监督学习方法, 使计算更快速, 且预测结果更准确。

[0011] 在上述技术方案的基础上, 本发明还可以做如下改进:

[0012] 进一步, 所述S2中, 所述有效交互模式包括: 直接交互模式、间接交互模式和共性

交互模式。

[0013] 采用上述进一步方案的有益效果是：有效交互模式的设定和区分，有利于细致刻画社交网络用户之间的关系。

[0014] 进一步，所述S3中，得到所述社交距离向量的过程为：将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点，将除去该备选节点以外的任一节点作为中间节点，根据该目标节点的交互子图得到目标节点与该备选节点在该中间节点维度上的社交距离，进而根据在每个中间节点维度上的社交距离得到目标节点与该备选节点的社交距离向量，同理得到该目标节点与关系预测相关节点集合中未与该目标节点直接连接的其它节点作为备选节点时的社交距离向量。

[0015] 采用上述进一步方案的有益效果是：提供一种多维度社交距离向量的概念，用来度量社交网络中用户之间的接近程度。

[0016] 进一步，所述S3中，得到所述社交距离的过程为：根据所述该目标节点的交互子图得到表示各节点交互特性的权重矩阵，根据从权重矩阵获取的交互子图中的各节点到该备选节点的权重和该中间节点到该备选节点的最小跳数，得到目标节点与该备选节点在该中间节点维度上的社交距离。

[0017] 采用上述进一步方案的有益效果是：提供一种简单快速的算法用来计算局部社交距离，在非监督方法下运行，简单高效。

[0018] 进一步，所述S4中，根据所述收益得到所述关系预测结果的过程为：设定表示关系预测结果的节点集合中的节点数量，根据所述目标节点相对于每个备选节点收益，用社会效益表示该目标节点相对于该数量的备选节点收益之和，得到使该目标节点的社会效益最大的节点集合，即为该目标节点的关系预测结果。

[0019] 采用上述进一步方案的有益效果是：利用博弈论的学习方法进行关系预测，使预测结果更符合实际情况，准确度更高。

[0020] 一种基于距离博弈的社交网络关系预测系统，所述系统包括：

[0021] 社交网络关系提取模块，用于提取社交网络中所有节点的集合和任意两个节点间有向链接的集合；

[0022] 目标节点关系提取模块，选择社交网络中的任意一个节点作为目标节点，根据有向链接的集合得到该目标节点与其它节点的交互模式，根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合，以及该目标节点的交互子图；

[0023] 社交距离向量计算模块，用于将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点，根据该目标节点的交互子图得到该目标节点到每个备选节点的社交距离向量；

[0024] 目标节点关系预测模块，用于根据该目标节点到每个备选节点的社交距离向量得到表示社交距离博弈的该目标节点相对于每个备选节点收益，据此得到该目标节点的关系预测结果；

[0025] 社交网络关系预测模块，用于重复调用目标节点关系提取模块、社交距离向量计算模块和目标节点关系预测模块，最终得到社交网络中每个节点的关系预测结果。

[0026] 本发明的有益效果是：利用社交网络的拓扑结构信息结合社会学理论，提出社交距离向量等新的相似度指标，采取非监督学习方法，使计算更快速，且预测结果更准确。

[0027] 在上述技术方案的基础上,本发明还可以做如下改进:

[0028] 进一步,所述目标节点关系提取模块中,所述有效交互模式包括:直接交互模式、间接交互模式和共性交互模式。

[0029] 采用上述进一步方案的有益效果是:有效交互模式的设定和区分,有利于细致刻画社交网络用户之间的关系。

[0030] 进一步,所述社交距离向量计算模块还用于:将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,将除去该备选节点以外的任一节点作为中间节点,根据该目标节点的交互子图得到目标节点与该备选节点在该中间节点维度上的社交距离,进而根据在每个中间节点维度上的社交距离得到目标节点与该备选节点的社交距离向量,同理得到该目标节点与关系预测相关节点集合中未与该目标节点直接连接的其它节点作为备选节点时的社交距离向量。

[0031] 采用上述进一步方案的有益效果是:提供一种多维度社交距离向量的概念,用来度量社交网络中用户之间的接近程度。

[0032] 进一步,所述社交距离向量计算模块还用于:根据所述该目标节点的交互子图得到表示各节点交互特性的权重矩阵,根据从权重矩阵获取的交互子图中的各节点到该备选节点的权重和该中间节点到该备选节点的最小跳数,得到目标节点与该备选节点在该中间节点维度上的社交距离。

[0033] 采用上述进一步方案的有益效果是:提供一种简单快速的算法用来计算局部社交距离,在非监督方法下运行,简单高效。

[0034] 进一步,所述目标节点关系预测模块还用于:设定表示关系预测结果的节点集中的节点数量,根据所述目标节点相对于每个备选节点的收益,用社会效益表示该目标节点相对于该数量的备选节点的收益之和,得到使该目标节点的社会效益最大的节点集合,即为该目标节点的关系预测结果。

[0035] 采用上述进一步方案的有益效果是:利用博弈论的学习方法进行关系预测,使预测结果更符合实际情况,准确度更高。

附图说明

[0036] 图1为本发明的基于距离博弈的社交网络关系预测方法流程图;

[0037] 图2为本发明的基于距离博弈的社交网络关系预测系统的框图;

[0038] 图3为本发明中目标节点的局部结构的示意图;

[0039] 图4为本发明中目标节点的交互子图的实例图。

具体实施方式

[0040] 以下结合附图对本发明的原理和特征进行描述,所举实例只用于解释本发明,并非用于限定本发明的范围。

[0041] 实施例1

[0042] 如图1所示,一种基于距离博弈的社交网络关系预测方法,包括以下步骤:

[0043] S1,提取社交网络中所有节点的集合和任意两个节点间有向链接的集合;

[0044] S2,选择社交网络中的任意一个节点作为目标节点,根据有向链接的集合得到该

目标节点与其它节点的交互模式,根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合,以及该目标节点的交互子图;

[0045] S3,将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,根据该目标节点的交互子图得到该目标节点到每个备选节点的社交距离向量;

[0046] S4,根据该目标节点到每个备选节点的社交距离向量得到表示社交距离博弈的该目标节点相对于每个备选节点的收益,据此得到该目标节点的关系预测结果,重复进行S2、S3和S4,最终得到社交网络中每个节点的关系预测结果。

[0047] 具体的,社交网络用一个有向图表示为 $I = (N, E)$,其中 $N = \{x_1, x_2, \dots, x_n\}$ 是所有节点的集合, $E = \{e_{ij}\}$ 是任意两个节点间有向链接的集合。集合E由如下方式生成:集合中每个元素 e_{ij} 表示一条从节点 x_i 到节点 x_j 的有向链接,代表了社交网络中节点 x_i 关注节点 x_j 。对于节点 x_i ,其关注的节点集合(称为邻居)为 $F(x_i) = \{x_j \in N | e_{ji} = 1\}$,关注该节点的集合为: $A(x_i) = \{x_j \in N | e_{ji} = 1\}$ 。

[0048] 对于一个给定的目标节点 x ,可以根据有向链接的集合得到该目标节点与其它节点的交互模式,根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合,局部结构如图3所示,其中关系预测相关节点集合表示为: $N' = \{x\} \cup A(x) \cup F(A(x)) \cup F(x) \cup F(F(x)) \cup A(F(x))$,即局部结构中所有节点的集合,也可乘坐局部结构节点集合。在生成的节点集合 N' 中,当前节点 x 与未直接连接的节点 $\{y \notin F(x)\}$ 之间的社交距离由其所有局部结构节点集合的结构和交互特征构成。

[0049] 本发明利用社交网络的拓扑结构信息结合社会学理论,提出社交距离向量等新的相似度指标,采取非监督学习方法,使计算更快速,且预测结果更准确。

[0050] 进一步,S2中,有效交互模式包括:直接交互模式、间接交互模式和共性交互模式。

[0051] 具体的,节点 x 和关系预测相关节点集合中的节点 y 之间存在多种有效交互模式:

[0052] 第一种交互模式:直接交互模式,两个节点至少有一方关注另一方;

[0053] 第二种交互模式:间接交互模式,一个节点关注的节点关注另一个节点,即“朋友的朋友”可能通过中间节点与产生交互;

[0054] 第三种交互模式:共性交互模式,两个节点有共同关注的节点或者两个节点共同被一些节点关注,与节点 x 拥有共同邻居集合或关注者集合的节点集合也可能共享某些相同的兴趣或者在同一社区群体中从而产生交互。

[0055] 该改进中,有效交互模式的设定和区分,有利于细致刻画社交网络用户之间的关系。

[0056] 进一步,S3中,得到社交距离向量的过程为:将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,将除去该备选节点以外的任一节点作为中间节点,根据该目标节点的交互子图得到目标节点与该备选节点在该中间节点维度上的社交距离,进而根据在每个中间节点维度上的社交距离得到目标节点与该备选节点的社交距离向量,同理得到该目标节点与关系预测相关节点集合中未与该目标节点直接连接的其它节点作为备选节点时的社交距离向量。

[0057] 该改进中,提供一种多维度社交距离向量的概念,用来度量社交网络中用户之间的接近程度。

[0058] 进一步, S3中, 得到社交距离的过程为: 根据该目标节点的交互子图得到表示各节点交互特性的权重矩阵, 根据从权重矩阵获取的交互子图中的各节点到该备选节点的权重和该中间节点到该备选节点的最小跳数, 得到目标节点与该备选节点在该中间节点维度上的社交距离。

[0059] 具体的, 如图4所示, 为一个交互子图的具体实例。一般情况下, 定义目标节点x的交互子图 $I' = (N', E')$ 和与之匹配的权重矩阵 $W = [w_{ij}]$, 其中对于任何连接边 e_{ij} 满足 $w_{ij} \geq 0$; 且对于任何i满足 $\sum_j w_{ij} = 1$ 。每一个在目标节点的局部结构中除备选节点以外的节点将作为一个社交距离概念下的中间节点 c_i 。目标节点x到备选节点y在中间节点 c_i 维度上的社交距离用如下三元组来定义: $SD(x, c_i, y) = (\sum_{x_k \in N'} w_{x_k, c_i} \cdot \text{shortestpath}(c_i, y))$, 其中权重 w_{x_k, c_i} 的

和用来刻画目标节点x的交互子图中的每个节点 x_k 与中间节点 c_i 的交互特性, 备选节点y到中间节点 c_i 的最短路径 (shortestpath, 最小跳数) 用来刻画结构特征。社交距离的概念反映了从中间节点 c_i 的视角看, 目标节点x到备选节点y的接近程度。不妨假定目标节点x的局部结构有m个中间节点, 那么全维度的社交距离向量定义如下: $SD_{x,y} = [SD(x, c_1, y), \dots, SD(x, c_m, y)]$, 其中 c_1, c_2, \dots, c_m 在节点x的局部结构节点集合中。

[0060] 该改进中, 提供一种简单快速的算法用来计算局部社交距离, 在非监督方法下运行, 简单高效。

[0061] 进一步, S4中, 根据收益得到关系预测结果的过程为: 设定表示关系预测结果的节点集合中的节点数量, 根据目标节点相对于每个备选节点收益, 用社会效益表示该目标节点相对于该数量的备选节点收益之和, 得到使该目标节点的社会效益最大的节点集合, 即为该目标节点的关系预测结果。

[0062] 具体的, 有了相似度指标即社交距离的定义, 下面提出一个基于距离的博弈方法。定义社交距离博弈 $G(I)$ 为对 $\langle N, v \rangle$, 其中N是有向图I定义的节点集合, v代表集合N的收益向量, 对于任意 $N' \subseteq N$, $v(N') \subset R^{|N|}$, 存在一个节点y, 有 $v_y(N') = u(y, N')$, 即集合N'在节点y维度上的收益等于关系预测相关节点集合为N'的目标节点x相对于备选节点y的收益, 其中, 关系预测相关节点集合为 $N' \subseteq N$ 中的目标节点x相对于备选节点 y_k 的收益 (utility)

定义为: $u(y_k, N') = \frac{1}{m} \sum_{i=1}^m \frac{1}{SD(x, c_i, y_k)}$ 。

[0063] 定义目标节点x的社会效益 (social welfare) 为所选的k ($k \leq n$) 个节点的收益之和: $SW(x) = \sum_{k \leq n} u(y_k, N')$, n为备选节点的个数。最后依据此公式计算社交距离博弈, 得到使

得社会效益最大化时的k个节点的集合: $\{y_k\}^* = \arg \max \{SW(x) = \sum_{k \leq n} u(y_k, N')\}$, 这k个节点

就是按照本发明得到的关系预测结果。注意到本发明预测的关系是针对两个未连接的节点, 即在有向图中预测链接的产生, 但不预测该链接的方向。

[0064] 该改进中, 利用博弈论的学习方法进行关系预测, 使预测结果更符合实际情况, 准确度更高。

[0065] 如图4所示的目标节点 x_0 的交互子图所示, 用具体计算实例说明上述过程, 目的是找到与目标节点 x_0 最可能产生联系的节点。

[0066] 1.由公式 $N' = \{x\} \cup A(x) \cup F(A(x)) \cup F(x) \cup F(F(x)) \cup A(F(x))$ 得到 x_0 的局部结构集合为 $\{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$,具有9个节点,权重矩阵为:

$$[0067] \quad W = [w_{ij}] = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}。$$

[0068] 2.计算社交距离SD

[0069] 对于其中的中间节点 x_1 ,目标节点 x_0 和备选节点 x_5 的社交距离定义为

$$SD(x_0, x_1, x_5) = \left(\sum_{x_k \in N'} w_{x_k, x_1} \cdot \text{shortestpath}(x_1, x_5) \right) = \frac{7}{3};$$

[0070] 对于其中的中间节点 x_2 ,目标节点 x_0 和备选节点 x_5 的社交距离定义为

$$SD(x_0, x_2, x_5) = \left(\sum_{x_k \in N'} w_{x_k, x_2} \cdot \text{shortestpath}(x_2, x_5) \right) = \frac{7}{3};$$

[0071] 类似可得对于其他中间节点维度上目标节点 x_0 和备选节点 x_5 的社交距离,从而得到全维度(除去备选节点本身的维度,剩余8个维度)的社交距离向量 $SD_{x_0, x_5} = [\frac{7}{6}, \frac{7}{3}, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{2}, \frac{1}{2}]$ 。

[0072] 同理可得目标节点 x_0 与其他备选节点 x_6, \dots, x_8 的社交距离向量分别为:

$$[0073] \quad SD_{x_0, x_6} = [\frac{7}{6}, \frac{14}{3}, \frac{1}{2}, 1, \frac{3}{4}, 1, \frac{3}{4}, \frac{3}{4}],$$

$$[0074] \quad SD_{x_0, x_7} = [\frac{7}{2}, 7, \frac{3}{2}, \frac{1}{2}, \frac{1}{4}, 2, 2, \frac{1}{2}],$$

$$[0075] \quad SD_{x_0, x_8} = [\frac{7}{6}, \frac{7}{3}, 1, \frac{1}{2}, \frac{1}{2}, 2, 1, \frac{1}{2}]。$$

[0076] 3.进行基于距离的博弈,得到预测结果节点集合

[0077] 关系预测相关节点集合为 $N' \subseteq N$ 的目标节点 x_0 相对于备选节点 x_5 的收益为:

$$u(x_5, N') = \frac{1}{m} \sum_{i=1}^m \frac{1}{SD(x, c_i, y_k)} = \frac{1}{8} \sum_{i=1}^8 \frac{1}{SD(x_0, c_i, x_5)} = \frac{1}{8} (\frac{6}{7} + \frac{3}{7} + 1 + 2 + 2 + 3 + 2 + 2) = \frac{93}{56};$$

$$[0078] \quad \text{类似可得: } u(x_6, N') = \frac{127}{112}, \quad u(x_7, N') = \frac{53}{42}, \quad u(x_8, N') = \frac{141}{896}。$$

[0079] 目标节点 x_0 社会效益为: $SW(x_0) = \sum_{k \leq n} u(x_k, N')$, n 为备选节点的个数。

[0080] 若我们要预测1个最有可能出现的关系,即与目标节点 x_0 连接的节点,那么令 $k=1$,有 $\text{argmax} SW(x_0)$ 时,解为 x_5 。

[0081] 重复以上步骤,得到所有节点的关系预测结果即可。

[0082] 实施例2

[0083] 如图2所示,一种基于距离博弈的社交网络关系预测系统,包括:

[0084] 社交网络关系提取模块,用于提取社交网络中所有节点的集合和任意两个节点间有向链接的集合;

[0085] 目标节点关系提取模块,选择社交网络中的任意一个节点作为目标节点,根据有向链接的集合得到该目标节点与其它节点的交互模式,根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合,以及该目标节点的交互子图;

[0086] 社交距离向量计算模块,用于将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,根据该目标节点的交互子图得到该目标节点到每个备选节点的社交距离向量;

[0087] 目标节点关系预测模块,用于根据该目标节点到每个备选节点的社交距离向量得到表示社交距离博弈的该目标节点相对于每个备选节点的收益,据此得到该目标节点的关系预测结果;

[0088] 社交网络关系预测模块,用于重复调用目标节点关系提取模块、社交距离向量计算模块和目标节点关系预测模块,最终得到社交网络中每个节点的关系预测结果。

[0089] 具体的,具体的,社交网络用一个有向图表示为 $I = (N, E)$,其中 $N = \{x_1, x_2, \dots, x_n\}$ 是所有节点的集合, $E = \{e_{ij}\}$ 是任意两个节点间有向链接的集合。集合 E 由如下方式生成:集合中每个元素 e_{ij} 表示一条从节点 x_i 到节点 x_j 的有向链接,代表了社交网络中节点 x_i 关注节点 x_j 。对于节点 x_i ,其关注的节点集合(称为邻居)为 $F(x_i) = \{x_j \in N | e_{ji} = 1\}$,关注该节点的集合为: $A(x_i) = \{x_j \in N | e_{ji} = 1\}$ 。

[0090] 对于一个给定的目标节点 x ,可以根据有向链接的集合得到该目标节点与其它节点的交互模式,根据交互模式中的有效交互模式得到该目标节点的局部结构和关系预测相关节点集合,局部结构如图3所示,其中关系预测相关节点集合表示为: $N' = \{x\} \cup A(x) \cup F(A(x)) \cup F(x) \cup F(F(x)) \cup A(F(x))$,即局部结构中所有节点的集合,也可乘坐局部结构节点集合。在生成的节点集合 N' 中,当前节点 x 与未直接连接的节点 $\{y \notin F(x)\}$ 之间的社交距离由其所有局部结构节点集合的结构和交互特征构成。

[0091] 本发明利用社交网络的拓扑结构信息结合社会学理论,提出社交距离向量等新的相似度指标,采取非监督学习方法,使计算更快速,且预测结果更准确。

[0092] 进一步,目标节点关系提取模块中,有效交互模式包括:直接交互模式、间接交互模式和共性交互模式。

[0093] 具体的,节点 x 和关系预测相关节点集合中的节点 y 之间存在多种有效交互模式:

[0094] 第一种交互模式:直接交互模式,两个节点至少有一方关注另一方;

[0095] 第二种交互模式:间接交互模式,一个节点关注的节点关注另一个节点,即“朋友的朋友”可能通过中间节点与产生交互;

[0096] 第三种交互模式:共性交互模式,两个节点有共同关注的节点或者两个节点共同被一些节点关注,与节点 x 拥有共同邻居集合或关注者集合的节点集合也可能共享某些相同的兴趣或者在同一社区群体中从而产生交互。

[0097] 该改进中,有效交互模式的设定和区分,有利于细致刻画社交网络用户之间的关

系。

[0098] 进一步,社交距离向量计算模块还用于:将该目标节点的关系预测相关节点集合中未与该目标节点直接连接的任一节点作为备选节点,将除去该备选节点以外的任一节点作为中间节点,根据该目标节点的交互子图得到目标节点与该备选节点在该中间节点维度上的社交距离,进而根据在每个中间节点维度上的社交距离得到目标节点与该备选节点的社交距离向量,同理得到该目标节点与关系预测相关节点集合中未与该目标节点直接连接的其它节点作为备选节点时的社交距离向量。

[0099] 该改进中,提供一种多维度社交距离向量的概念,用来度量社交网络中用户之间的接近程度。

[0100] 进一步,社交距离向量计算模块还用于:根据该目标节点的交互子图得到表示各节点交互特性的权重矩阵,根据从权重矩阵获取的交互子图中的各节点到该备选节点的权重和该中间节点到该备选节点的最小跳数,得到目标节点与该备选节点在该中间节点维度上的社交距离。

[0101] 具体的,如图4所示,为一个交互子图的具体实例。一般情况下,定义目标节点x的交互子图 $I' = (N', E')$ 和与之匹配的权重矩阵 $W = [w_{ij}]$,其中对于任何连接边 e_{ij} 满足 $w_{ij} \geq 0$;且对于任何i满足 $\sum_j w_{ij} = 1$ 。每一个在目标节点的局部结构中除备选节点以外的节点将作为一个社交距离概念下的中间节点 c_i 。目标节点x到备选节点y在中间节点 c_i 维度上的社交距离用如下三元组来定义: $SD(x, c_i, y) = (\sum_{x_k \in N'} w_{x_k, c_i} \cdot \text{shortestpath}(c_i, y))$,其中权重 w_{x_k, c_i} 的

和用来刻画目标节点x的交互子图中的每个节点 x_k 与中间节点 c_i 的交互特性,备选节点y到中间节点 c_i 的最短路径(shortestpath,最小跳数)用来刻画结构特征。社交距离的概念反映了从中间节点 c_i 的视角看,目标节点x到备选节点y的接近程度。不妨假定目标节点x的局部结构有m个中间节点,那么全维度的社交距离向量定义如下: $SD_{x,y} = [SD(x, c_1, y), \dots, SD(x, c_m, y)]$,其中 c_1, c_2, \dots, c_m 在节点x的局部结构节点集合中。

[0102] 该改进中,提供一种简单快速的算法用来计算局部社交距离,在非监督方法下运行,简单高效。

[0103] 进一步,目标节点关系预测模块还用于:设定表示关系预测结果的节点集合中的节点数量,根据目标节点相对于每个备选节点的收益,用社会效益表示该目标节点相对于该数量的备选节点的收益之和,得到使该目标节点的社会效益最大的节点集合,即为该目标节点的关系预测结果。

[0104] 具体的,有了相似度指标即社交距离的定义,下面提出一个基于距离的博弈方法。定义社交距离博弈 $G(I)$ 为对 $\langle N, v \rangle$,其中N是有向图I定义的节点集合,v代表集合N的收益向量,对于任意 $N' \subseteq N$, $v(N') \subset R^{|N' |}$,存在一个节点y,有 $v_y(N') = u(y, N')$,即集合N'在节点y维度上的收益等于关系预测相关节点集合为N'的目标节点x相对于备选节点y的收益,其中,关系预测相关节点集合为 $N' \subseteq N$ 中的目标节点x相对于备选节点 y_k 的收益(utility)

定义为: $u(y_k, N') = \frac{1}{m} \sum_{i=1}^m \frac{1}{SD(x, c_i, y_k)}$ 。

[0105] 定义目标节点x的社会效益(social welfare)为所选的k($k \leq n$)个节点的收益之

和： $SW(x) = \sum_{k \leq n} u(y_k, N')$ ， n 为备选节点的个数。最后依据此公式计算社交距离博弈，得到使

得社会效益最大化时的 k 个节点的集合： $\{y_k\}^* = \arg \max \{SW(x) = \sum_{k \leq n} u(y_k, N')\}$ ，这 k 个节点

就是按照本发明得到的关系预测结果。注意到本发明预测的关系是针对两个未连接的节点，即在有向图中预测链接的产生，但不预测该链接的方向。

[0106] 该改进中，利用博弈论的学习方法进行关系预测，使预测结果更符合实际情况，准确度更高。

[0107] 本发明提出一种多维度社交距离的概念，用来度量社交网络中一群用户之间的接近程度，进而提出一种快速哈希算法：局部敏感社交距离哈希算法，该算法在非监督配置下运行，完成近似最近邻搜索同时避免了高维距离计算。

[0108] 本发明利用社交网络的拓扑结构信息结合社会学理论，提出新的相似度指标，并利用博弈论的学习方法进行关系预测。

[0109] 计算社交距离时需要根据社交网络的交互模式得到局部结构，其计算在网络拓扑结构已知的基础上，至少是对于预测节点的2跳范围内的邻接网络已知。在进行基于距离的博弈时，可能会得到一个或多个结果作为关系预测的结果，多个结果时各个结果之间同等对待，没有排序。本发明定义了新的相似度指标定义：社交距离，和基于距离的博弈方法。

[0110] 本发明本质上是一种非监督学习方法，根据已知的社交网络拓扑结果即可以计算相似度指标并进行基于距离的博弈。相比与有监督的，如基于路径、随机游走等方法，具有计算快速的优势；相比于其他非监督方法，如基于近邻的各种指标相比，具有对关系更细致的刻画，能够得到更为准确的预测结果。

[0111] 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

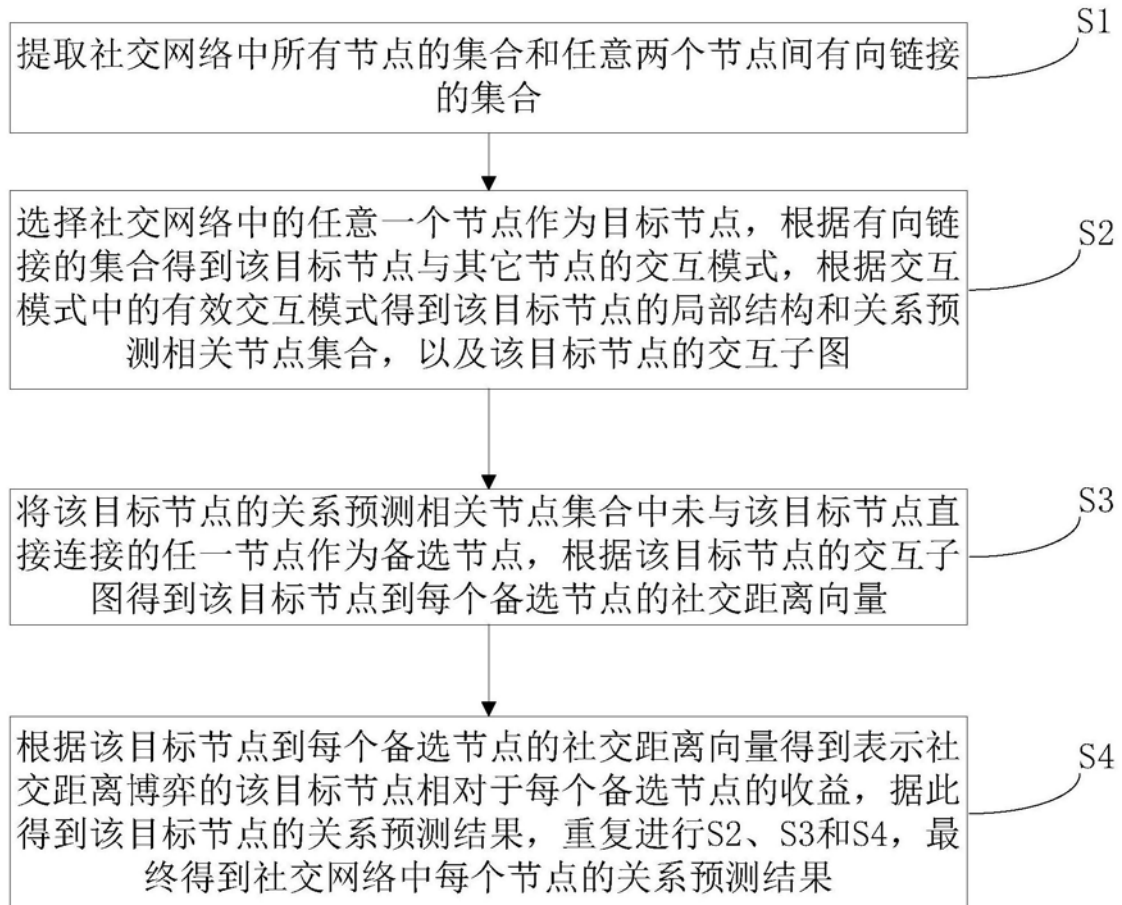


图1

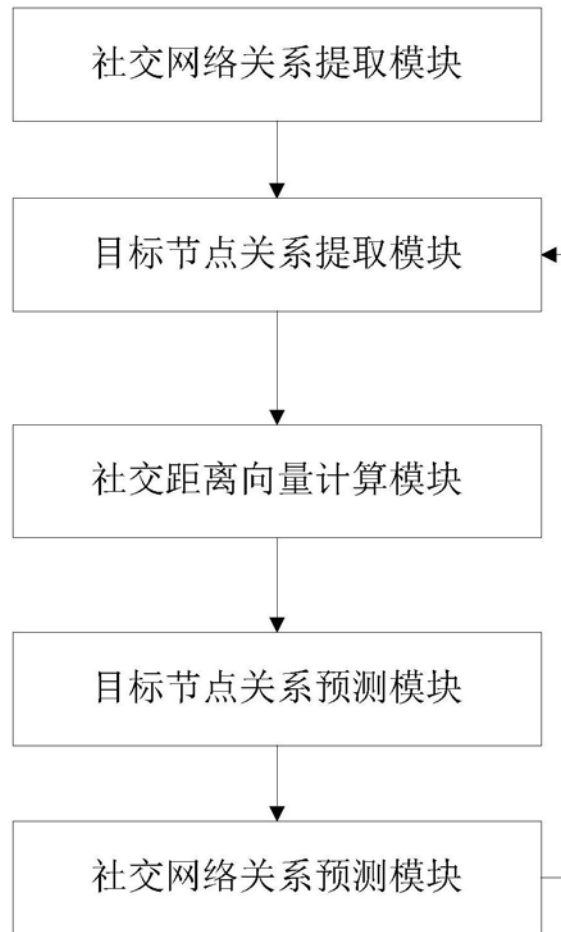


图2

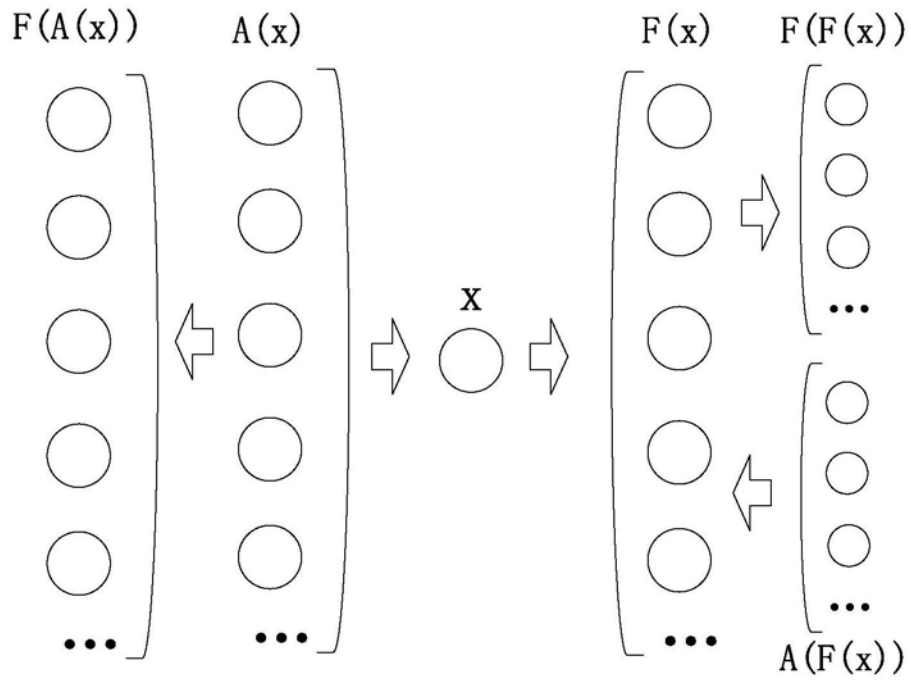


图3

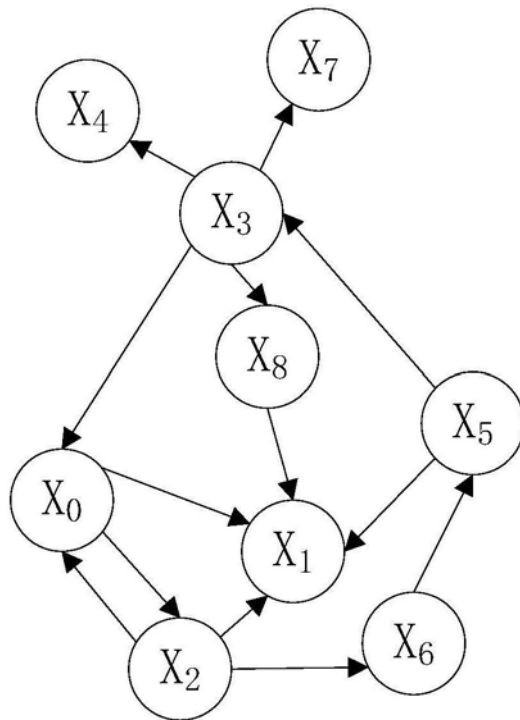


图4