

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 1 月 23 日 (23.01.2020)



(10) 国际公布号
WO 2020/015464 A1

(51) 国际专利分类号:
G06F 16/901 (2019.01)

(21) 国际申请号: PCT/CN2019/089022

(22) 国际申请日: 2019 年 5 月 29 日 (29.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810784744.5 2018 年 7 月 17 日 (17.07.2018) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 向彪(XIANG, Biao); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 刘子奇(LIU, Ziqi); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 周俊(ZHOU, Jun); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 李小龙(LI, Xiaolong); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(54) Title: METHOD AND APPARATUS FOR EMBEDDING RELATIONAL NETWORK DIAGRAM

(54) 发明名称: 关系网络图嵌入的方法及装置

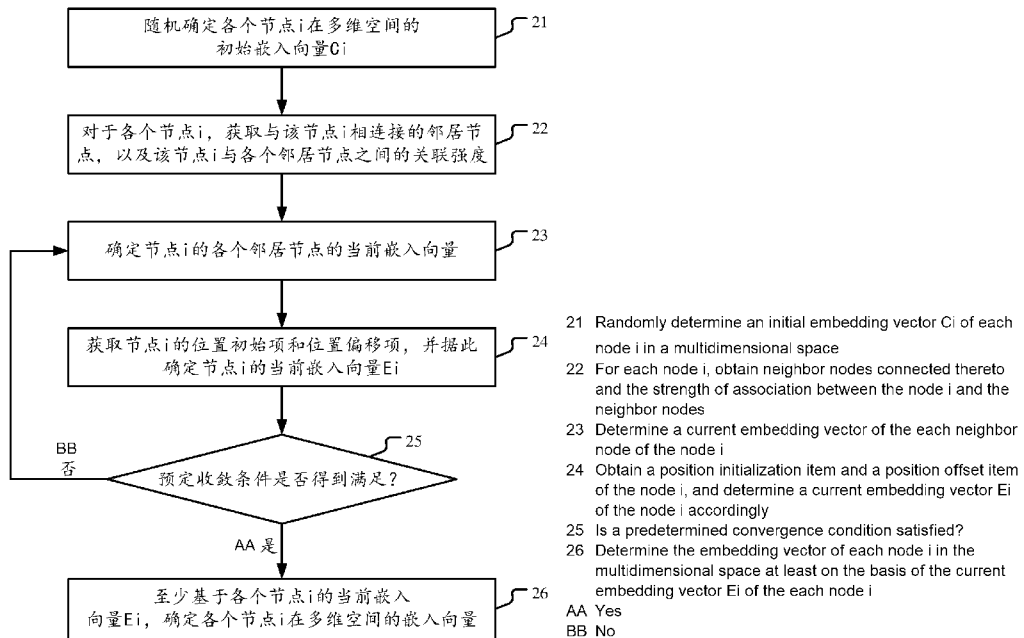


图 2

(57) Abstract: Provided by the embodiments of the present description are a method and apparatus for embedding a relational network diagram into a multidimensional space, the method comprising: randomly determining an initial embedding vector C_i of each node i in a multidimensional space; obtaining neighbor nodes of each node i and the strength of association thereof with each neighbor node; determining a current embedding vector of the each neighbor node of the node i ; forming a position initialization item and a position offset item of the nodes i on the basis of a predetermined attenuation coefficient, the strength of association and the current position

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

of the neighbor nodes, and determining a current embedding vector of the nodes i accordingly; repeating the described steps until a convergence condition is satisfied, at which time the embedding vector of each node i in the multidimensional space may be determined. Thus, a relational network graph is efficiently embedded into the multi-dimensional space.

(57) 摘要: 本说明书实施例提供一种将关系网络图嵌入到多维空间的方法和装置, 在该方法中, 首先随机确定各个节点 i 在多维空间的初始嵌入向量 C_i , 然后对于各个节点 i , 获取其邻居节点, 以及与各个邻居节点之间的关联强度。此外, 还确定节点 i 的各个邻居节点的当前嵌入向量。基于预定的衰减系数、上述关联强度以及邻居节点的当前位置, 形成节点 i 的位置初始项和位置偏移项, 并据此确定节点 i 的当前嵌入向量。反复执行上述步骤直到收敛条件得到满足, 此时即可以确定各个节点 i 在多维空间的嵌入向量。如此, 高效地将关系网络图嵌入到多维空间中。

关系网络图嵌入的方法及装置

技术领域

[01]本说明书一个或多个实施例涉及计算机信息处理领域，尤其涉及关系网络图嵌入的方法和装置。

5 背景技术

[02]关系网络图是对现实世界中实体之间的关系的描述，目前广泛地应用于各种计算机信息处理中。一般地，关系网络图包含一个节点集合和一个边集合，节点表示现实世界中的实体，边表示现实世界中实体之间的联系。例如，在社交网络中，人就是实体，人和人之间的关系或联系就是边。

- 10 [03]在许多情况下，希望将关系网络图中的每个节点（实体）用多维空间中的坐标值来表示，也就是将各个节点映射到一个多维空间中，用多维空间中的点代表图中的节点。多维空间可以是2维、3维空间，也可以是更高维空间。用多维空间的坐标来表达图中的节点，可以应用于计算节点和节点之间的相似度，发现图中的社团结构，预测未来可能形成的边联系，以及对图进行可视化等。将图中的节点映射到多维空间的过程称为图
- 15 嵌入。

[04]图嵌入是一种非常重要的基础技术能力。当前学术界已研究出多种图嵌入方法，如DeepWalk, node2vec, GraphRep等。但由于这些算法内部均采用了蒙特卡洛采样方法，计算效率比较低。当图的规模变得很大时（如支付宝朋友关系网络有5亿以上节点），进行图嵌入计算将耗费巨大的计算资源。

- 20 [05]因此，希望能有改进的方案，更加快速有效地进行关系网络图的图嵌入过程。

发明内容

[06]本说明书一个或多个实施例描述了一种关系网络图的图嵌入方法，可以高效地将复杂关系网络图中的节点嵌入到多维空间中，以便于后续的信息处理。

- [07]根据第一方面，提供了一种将关系网络图嵌入到多维空间的方法，所述关系网络图
- 25 包括多个节点，所述多个节点中具有关联关系的节点以一定关联强度互相连接，所述方法包括：

[08]随机确定所述多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i ;

[09]对于各个节点 i , 获取与该节点 i 相连接的邻居节点, 以及该节点 i 与各个邻居节点之间的关联强度;

[10]确定该节点 i 的各个邻居节点的当前嵌入向量;

5 [11]获取该节点 i 的位置初始项和位置偏移项, 并根据所述位置初始项和位置偏移项, 确定节点 i 的当前嵌入向量 E_i , 其中所述位置初始项基于所述初始嵌入向量 C_i 而确定, 所述位置偏移项根据预定衰减系数 α 、所述各个邻居节点的当前嵌入向量以及该节点 i 与各个邻居节点之间的关联强度而确定;

10 [12]判断预定收敛条件是否得到满足, 在不满足该预定收敛条件的情况下, 再次确定该节点 i 的各个邻居节点的当前嵌入向量, 以及再次确定节点 i 的当前嵌入向量 E_i , 直到该预定收敛条件得到满足;

[13]至少基于满足所述预定收敛条件的各个节点 i 的当前嵌入向量 E_i , 确定各个节点 i 在所述多维空间的嵌入向量。

[14]根据一种实施方式, 通过以下方式获取节点 i 的邻居节点信息:

15 [15]获取记录所述关系网络图的网络关系的邻接矩阵, 邻接矩阵中第 m 行第 k 列的元素对应于第 m 节点与第 k 节点之间的关联强度;

[16]通过所述邻接矩阵, 确定节点 i 的邻居节点, 以及节点 i 与各个邻居节点之间的关联强度。

[17]进一步地, 通过邻接矩阵确定节点 i 的各个邻居节点, 以及各个关联强度包括:

20 [18]获取邻接矩阵中与节点 i 对应的第 i 行元素或第 i 列元素;

[19]将所述第 i 行元素或第 i 列元素中非零元素对应的节点确定为节点 i 的邻居节点; 将所述非零元素的值确定为节点 i 与对应邻居节点之间的关联强度。

[20]根据一个实施例, 位置初始项基于初始嵌入向量 C_i 以及所述预定衰减系数而确定。

[21]在一个实施例中, 通过以下方式获取节点 i 的位置偏移项:

25 [22]以节点 i 与各个邻居节点之间的关联强度为权重, 对各个邻居节点的当前嵌入向量求和, 确定邻居中心位置;

[23]至少基于所述预定衰减系数 α , 所述邻居中心位置, 确定所述位置偏移项。

[24]在另一实施例中，通过以下方式获取节点 i 的位置偏移项：

[25]确定节点 i 与其所有邻居节点的关联强度的和值；

[26]确定节点 i 与各个邻居节点之间的关联强度与所述和值的比例，作为相对关联强度；

[27]以所述相对关联强度为权重，对各个邻居节点的当前嵌入向量求和，确定邻居中心位置；

[28]将邻居中心位置与所述预定衰减系数 α 的乘积，作为所述位置偏移项。

[29]根据一种可能的设计，上述预定收敛条件可以是：对于每个节点，本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值小于第一预定值；或者，各个节点的本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值的总和小于第二预定值。

10 [30]根据另一种可能的设计，上述预定收敛条件可以是，确定各个节点 i 的当前嵌入向量 E_i 的次数达到预定次数阈值。

[31]在一个实施例中，将节点 i 的嵌入向量确定为，满足所述预定收敛条件时节点 i 的当前嵌入向量 E_i 与其位置初始项之差。

15 [32]根据第二方面，提供一种将关系网络图嵌入到多维空间的装置，所述关系网络图包括多个节点，所述多个节点中具有关联关系的节点以一定关联强度互相连接，所述装置包括：

[33]初始位置确定单元，配置为随机确定所述多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i ；

20 [34]邻居节点确定单元，配置为对于各个节点 i ，获取与该节点 i 相连接的邻居节点，以及该节点 i 与各个邻居节点之间的关联强度；

[35]邻居位置确定单元，配置为确定该节点 i 的各个邻居节点的当前嵌入向量；

25 [36]节点位置确定单元，配置为获取该节点 i 的位置初始项和位置偏移项，并根据所述位置初始项和位置偏移项，确定节点 i 的当前嵌入向量 E_i ，其中所述位置初始项基于所述初始嵌入向量 C_i 而确定，所述位置偏移项根据预定衰减系数 α 、所述各个邻居节点的当前嵌入向量以及该节点 i 与各个邻居节点之间的关联强度而确定；

[37]条件判定单元，配置为判断预定收敛条件是否得到满足，在不满足该预定收敛条件的情况下，使得所述邻居位置确定单元再次确定该节点 i 的各个邻居节点的当前嵌入向量，所述节点位置确定单元再次确定节点 i 的当前嵌入向量 E_i ，直到该预定收敛条件得

到满足;

[38]嵌入位置确定单元,配置为至少基于满足所述预定收敛条件的各个节点 i 的当前嵌入向量 E_i ,确定各个节点 i 在所述多维空间的嵌入向量。

[39]根据第三方面,提供了一种计算机可读存储介质,其上存储有计算机程序,当所述
5 计算机程序在计算机中执行时,令计算机执行第一方面的方法。

[40]根据第四方面,提供了一种计算设备,包括存储器和处理器,其特征在于,所述存储器中存储有可执行代码,所述处理器执行所述可执行代码时,实现第一方面的方法。

[41]通过本说明书实施例提供的方法和装置,可以高效地将关系网络图嵌入到多维空间中,便于后续的节点信息处理。

10 附图说明

[42]为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它的附图。

15 [43]图 1 为本说明书披露的一个实施例的关系网络图的示意图;

[44]图 2 示出根据一个实施例的将关系网络图嵌入到多维空间的方法;

[45]图 3 示出嵌入到二维空间的关系网络图的示例;

[46]图 4 示出根据一个实施例的图嵌入装置的示意性框图。

具体实施方式

20 [47]下面结合附图,对本说明书提供的方案进行描述。

[48]图 1 为本说明书披露的一个实施例的关系网络图的示意图。如图 1 所示,该关系网络图包括多个节点,为了清楚起见,图 1 中为这些节点进行了编号。在这些节点中,具有关联关系的节点之间用边进行连接。在一个例子中,图 1 中的节点表示社交网络中的人或用户,两个节点之间通过边连接起来,即表示对应的两个用户存在社交上的关联,
25 例如转账、留言、通讯等等。

[49]在一个实施例中,节点之间的关联关系还具有不同的关联强度。例如,在一个例子

中，针对不同的社交交互行为，设定不同的关联强度，比如，进行转账交互的用户的关联强度为 0.8，进行留言操作的用户的关联强度为 0.5，等等。在一个实施例中，在关联关系具有不同关联强度的情况下，可以利用边的属性或边的权值来表示该边所连接的两个用户之间的关联强度。

5 [50]在图 1 中的关系网络图中，为了示出各个节点以及节点间的连接关系，而示意性地示出各个节点的位置。实际上，网络关系图并不对节点的位置进行设置。对于节点的位置，需要采用图嵌入的方法，将各个节点映射到多维空间中。下面描述本说明书实施例所提供的图嵌入的方法。

[51]图 2 示出根据一个实施例的将关系网络图嵌入到多维空间的方法，其中关系网络图
10 包括多个节点，多个节点中具有关联关系的节点以一定关联强度互相连接。上述方法可以由任何具有计算、处理能力的装置、设备、平台、设备集群执行。如图 2 所示，所述方法包括：步骤 21，随机确定多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i ；步骤 22，对于各个节点 i ，获取与该节点 i 相连接的邻居节点，以及该节点 i 与各个邻居节点之间的关联强度；步骤 23，确定该节点 i 的各个邻居节点的当前嵌入向量；步骤 24，
15 获取该节点 i 的位置初始项和位置偏移项，并根据所述位置初始项和位置偏移项，确定节点 i 的当前嵌入向量 E_i ，其中所述位置初始项基于所述初始嵌入向量 C_i 而确定，所述位置偏移项根据预定衰减系数 α 、所述各个邻居节点的当前嵌入向量以及该节点 i 与各个邻居节点之间的关联强度而确定；步骤 25，判断预定收敛条件是否得到满足；如果不满足预定收敛条件，则再次确定该节点 i 的各个邻居节点的当前嵌入向量，以及再次
20 确定节点 i 的当前嵌入向量 E_i ，直到该预定收敛条件得到满足；步骤 26，至少基于满足所述预定收敛条件的各个节点 i 的当前嵌入向量 E_i ，确定各个节点 i 在多维空间的嵌入向量。下面描述以上各个步骤的执行方式。

[52]首先，在步骤 21，随机确定关系网络图的多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i 。假定关系网络图包含 N 个节点，要嵌入的多维空间的维数为 s ，那么针对
25 N 个节点中的每个节点 i ，为其随机生成一个 s 维向量 C_i ，作为其初始嵌入向量。

[53]另一方面，在步骤 22，对于各个节点 i ，获取与该节点 i 相连接的邻居节点，以及该节点 i 与各个邻居节点之间的关联强度。

[54]可以理解，在关系网络图中，具有关联关系的节点之间会互相连接，互相连接的节点之间，互为邻居节点。另外，可以理解，关系网络图的拓扑结构可以通过多种方式记录。例如，在一个示例中，通过图表记录关系网络图的连接关系。此时，可以从上述图
30

表中读取各个节点 i 的邻居节点信息, 以及节点 i 与邻居节点之间的关联强度。

[55] 在一个实施例中, 通过矩阵来记录关系网络图的连接关系。例如, 描述一个关系网络图的矩阵可以有邻接矩阵, 度矩阵, 拉普拉斯矩阵等。在一个示例中, 通过记录关系网络图的网络关系的邻接矩阵来获取节点的邻居信息和关联强度信息。

5 [56] 具体地, 假定矩阵 A 是关系网络图 G 的邻接矩阵, 矩阵 A 可以表示为:

$$A=[a_{mk}]_{N*N},$$

[57] 其中, 第 m 行第 k 列的元素 a_{mk} 对应于节点 m 与节点 k 之间的关联强度。

[58] 如果两个节点之间没有连接, 不存在关联关系, 那么它们之间的关联强度为 0。

10 [59] 通过这样的邻接矩阵, 可以简单地获取各个节点的邻居信息和关联强度信息。具体地, 对于节点 i , 获取邻接矩阵 A 中与节点 i 对应的第 i 行元素或第 i 列元素, 即 a_{ij} 或 a_{ji} ; 将第 i 行元素或第 i 列元素中非零元素对应的节点 j 确定为节点 i 的邻居节点, 并将非零元素的值确定为节点 i 与对应邻居节点之间的关联强度。

[60] 在确定出各个节点 i 的邻居节点 j 的基础上, 在步骤 23, 确定节点 i 的各个邻居节点 j 的当前嵌入向量 E_j 。

15 [61] 可以理解, 由于在步骤 21 为每个节点都随机产生了初始嵌入向量, 在第一次执行该步骤 23 时, 对于未更新过当前嵌入向量的邻居节点 j , 其当前嵌入向量 E_j 即其对应的初始嵌入向量 C_j 。各个节点的当前嵌入向量会在后续进行迭代更新, 这将在后续步骤中展开描述。

20 [62] 基于步骤 22 和步骤 23 针对节点 i 获得的相关信息, 在步骤 24, 确定节点 i 的当前嵌入向量 E_i 。具体地, 节点 i 的当前嵌入向量 E_i 可以认为由两部分组成: 位置初始项 VI 和位置偏移项 VD :

$$E_i=VI+VD,$$

25 [63] 其中位置初始项 VI 基于初始嵌入向量 C_i 而确定, 位置偏移项 VD 根据预定衰减系数 α 、各个邻居节点 j 的当前嵌入向量 E_j 以及该节点 i 与各个邻居节点之间的关联强度 a_{ij} 而确定。

[64] 在一个实施例中, 节点 i 的位置初始项 VI 即为其初始嵌入向量 C_i , 即:

$$VI=C_i.$$

[65]在另一实施例中，初始位置项可以是初始嵌入向量 C_i 乘以一定系数。例如，该系数可以与在位置偏移项中引入的衰减系数 α 有关。因此，在一个实施例中，可以基于初始嵌入向量 C_i 以及该衰减系数 α ，确定位置初始项。具体地，在一个例子中，将位置初始项 VI 确定为：

$$5 \quad VI = (1 - \alpha) C_i$$

[66]一般地，位置初始项一旦确定，在后续更新迭代过程中固定不变。

[67]另一方面，还要确定节点 i 的位置偏移项 VD 。根据说明书的至少一个实施例，根据预定衰减系数 α 、各个邻居节点 j 的当前嵌入向量 E_j 以及该节点 i 与各个邻居节点之间的关联强度 a_{ij} 来确定位置偏移项 VD 。

10 [68]其中衰减系数 α 用于调节位置偏移调整的步长或大小，一般被预设为 0 到 1 之间的数值。

[69]在一个实施例中，以节点 i 与各个邻居节点 j 之间的关联强度 a_{ij} 为权重，对各个邻居节点 j 的当前嵌入向量 E_j 求和，确定邻居中心位置；然后基于预定衰减系数 α ，上述邻居中心位置，确定位置偏移项 VD 。

15 [70]在一个例子中，根据上述思想，将位置偏移项 VD 确定为：

$$VD = \alpha \sum_{j \in N(i)} a_{ij} E_j$$

[71]其中 $N(i)$ 表示节点 i 的邻居节点集合。

[72]以上 VD 的计算方式比较适合于关联强度 a_{ij} 本身就定义在 0 到 1 之间的情况。如果关联强度 a_{ij} 的范围较大，可以在预先设定衰减系数的时候，将其设定为较小的数值。

20 [73]在另一实施例中，通过以下方式确定节点 i 的位置偏移项 VD ：确定节点 i 与其所有邻居节点 j 的关联强度的和值 d_i ；确定节点 i 与各个邻居节点 j 之间的关联强度 a_{ij} 与所述和值 d_i 的比例，作为相对关联强度；以所述相对关联强度为权重，对各个邻居节点 j 的当前嵌入向量 E_j 求和，确定邻居中心位置；将邻居中心位置与预定衰减系数 α 的乘积，作为位置偏移项 VD 。

25 [74]在一个例子中，根据上述思想，将位置偏移项 VD 确定为：

$$VD = \alpha \sum_{j \in N(i)} \frac{a_{ij}}{d_i} E_j$$

[75] 其中:

$$d_i = \sum_{j=1}^N a_{ij}$$

5 [76] 如此, 考虑节点 i 与各个邻居节点 j 的关联强度, 确定节点 i 的邻居中心位置, 然后以衰减系数作为调节, 确定位置偏移项 VD , 如此, 该位置偏移项 VD 可以反映向邻居中心偏移的距离。

[77] 根据一个具体例子, 结合前述的位置初始项, 以及如上所述根据相对关联强度确定的位置偏移项, 可以将节点 i 的当前嵌入向量 E_i 确定为:

$$E_i = (1 - \alpha) C_i + \alpha \sum_{j \in N(i)} \frac{a_{ij}}{d_i} E_j$$

10 [78] 以上描述了多种确定节点 i 的当前嵌入向量 E_i 的方式。

[79] 根据任一方式, 对于关系网络图中的每个节点 i 执行以上的步骤 23 和 24, 从而为每个节点确定出当前嵌入向量。

15 [80] 接着, 在步骤 25, 判断预定收敛条件是否得到满足。如果不满足预定收敛条件, 则返回到步骤 23 和步骤 24, 再次确定节点 i 的各个邻居节点的当前嵌入向量, 以及再次确定节点 i 的当前嵌入向量 E_i 。

20 [81] 可以理解, 以上的步骤 23 和 24 是针对关系网络图中每个节点执行, 因此每次执行步骤 23 和 24 的循环, 各个节点的当前嵌入向量都会得到更新。相应地, 在第 $n+1$ 次执行步骤 23 时, 对于同一节点 i , 其邻居节点 j 的当前嵌入向量 E_j 与第 n 次执行时有所不同, 实际上在第 $n+1$ 次执行时使用是, 第 n 次执行完步骤 24 时各个节点的当前嵌入向量。这样, 步骤 24 中的位置偏移项在每次执行上述循环时都会发生改变, 进而使得各个节点 i 的当前嵌入向量不断得到更新。

[82] 这样的循环反复执行, 直到预定收敛条件得到满足。

[83] 在一个实施例中, 预定收敛条件根据偏移调整量而设置, 所述偏移调整量对应于本次确定的位置与前次确定的位置之间的偏移。

[84]具体地，在一个实施例中，预定收敛条件可以设置为，对于每个节点，本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值小于第一预定值。例如，对于关系网络图中的 N 个节点，如果每个节点的当前嵌入向量相对于上一次确定的嵌入向量之间的差值，也就是偏移距离，都小于一个距离阈值，那么说明，节点的位置调整已经小到一定程度，节点的位置趋于稳定和收敛，从而达到收敛条件。

[85]在另一实施例中，预定收敛条件可以设置为，各个节点的本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值的总和小于第二预定值。也就是说，考虑 N 个节点的偏移距离的总和 DT:

$$DT = \sum_{i=1}^N D_i$$

[86]其中 D_i 为节点 i 的偏移距离，即当前嵌入向量相对于上一次确定的嵌入向量之间的差值。

[87]当偏移距离的总和 DT 小于某个阈值，那么说明，节点的总体位置调整较小，节点的位置趋于稳定和收敛，从而达到收敛条件。

[88]在另一实施例中，还可以根据经验，预设循环的执行次数作为收敛条件。也就是说，当确定各个节点 i 的当前嵌入向量 E_i 的次数达到预定次数阈值，即认为满足收敛条件。根据经验，上述执行次数一般可以设置为 10-20 次之间。

[89]如果收敛条件得到满足，那么退出循环，进入步骤 26，至少基于满足预定收敛条件的各个节点 i 的当前嵌入向量 E_i ，确定各个节点 i 在多维空间的嵌入向量 Q_i 。

[90]在一个实施例中，将满足收敛条件的各个节点 i 的当前嵌入向量 E_i ，作为其嵌入向量 Q_i ，即 $Q_i = E_i$ 。

[91]在另一实施例中，为了减少最初随机产生的初始嵌入向量的影响，将节点 i 的嵌入向量确定为，满足预定收敛条件时节点 i 的当前嵌入向量 E_i 与其位置初始项之差，即：

$$Q_i = E_i - VI$$

[92]其中 VI 与初始嵌入向量 C_i 相关联，例如等于 C_i ，或者等于 C_i 乘以一定系数，比如 $(1 - \alpha) C_i$ 。

[93]如此，确定了各个节点 i 在多维空间的嵌入向量。

[94]基于这样确定的嵌入向量，就可以将关系网络图中的节点嵌入到多维空间中。嵌入

到多维空间中的节点具有了位置信息，并且由于嵌入过程中考虑了节点之间的连接关系和连接强度，因此其位置信息中也体现了节点之间的关联关系。例如，多维空间中位置相近的节点之间，关联关系更强。如此，非常有利于后续对节点关系信息的进一步处理，例如对节点进行聚类，发现节点形成的团体，计算节点之间的相似度，预测节点潜在的边联系，等等。当将关系网络图嵌入到二维空间或三维空间时，还非常有利于对关系网络的可视化呈现。

[95]图 3 示出嵌入到二维空间的关系网络图的示例。更具体地，图 3 是采用图 2 所示的方法，将图 1 的关系网络图嵌入到二维空间的示例。相比于图 1 中出于示意而随意摆放的节点，图 3 中节点的位置含有更多的信息量，体现了节点之间的关联关系。一些节点之间位置非常靠近，意味着，这些节点之间具有更强的关联关系。并且，从节点位置分布上也可以看出，节点会呈现潜在的节点簇。这样的信息都会有利于对关系网络中节点信息的进一步处理。

[96]根据另一方面，本说明书的实施例还提供一种将关系网络图嵌入到多维空间的装置，其中有待嵌入的关系网络图包括多个节点，多个节点中具有关联关系的节点以一定关联强度互相连接。图 4 示出根据一个实施例的图嵌入装置的示意性框图。如图 4 所示，图嵌入装置 400 包括：初始位置确定单元 41，配置为随机确定所述多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i ；邻居节点确定单元 42，配置为对于各个节点 i ，获取与该节点 i 相连接的邻居节点，以及该节点 i 与各个邻居节点之间的关联强度；邻居位置确定单元 43，配置为确定该节点 i 的各个邻居节点的当前嵌入向量；节点位置确定单元 44，配置为获取该节点 i 的位置初始项和位置偏移项，并根据所述位置初始项和位置偏移项，确定节点 i 的当前嵌入向量 E_i ，其中所述位置初始项基于所述初始嵌入向量 C_i 而确定，所述位置偏移项根据预定衰减系数 α 、所述各个邻居节点的当前嵌入向量以及该节点 i 与各个邻居节点之间的关联强度而确定；条件判定单元 45，配置为判断预定收敛条件是否得到满足，在不满足该预定收敛条件的情况下，使得所述邻居位置确定单元再次确定该节点 i 的各个邻居节点的当前嵌入向量，所述节点位置确定单元再次确定节点 i 的当前嵌入向量 E_i ，直到该预定收敛条件得到满足；以及嵌入位置确定单元 46，配置为至少基于满足所述预定收敛条件的各个节点 i 的当前嵌入向量 E_i ，确定各个节点 i 在所述多维空间的嵌入向量。

[97]根据一种实施方式，邻居节点确定单元 42 配置为：获取记录所述关系网络图的网络关系的邻接矩阵，所述邻接矩阵中第 m 行第 k 列的元素对应于第 m 节点与第 k 节点之

间的关联强度；通过所述邻接矩阵，确定节点 i 的邻居节点，以及节点 i 与各个邻居节点之间的关联强度。

[98]进一步地，在一个具体例子中，所述邻居节点确定单元 42 通过以下方式确定邻居节点信息：获取所述邻接矩阵中与节点 i 对应的第 i 行元素或第 i 列元素；将所述第 i 行元素或第 i 列元素中非零元素对应的节点确定为节点 i 的邻居节点；将所述非零元素的值确定为节点 i 与对应邻居节点之间的关联强度。

[99]在一个实施例中，节点位置确定单元 44 包括，初始项确定模块 441，配置为基于所述初始嵌入向量 C_i 以及所述预定衰减系数，确定所述位置初始项。

[100] 在一个实施例中，节点位置确定单元 44 包括偏移项确定模块 442，用于确定偏移项。

[101] 在一个例子中，偏移项确定模块 442 配置为：以节点 i 与各个邻居节点之间的关联强度为权重，对各个邻居节点的当前嵌入向量求和，确定邻居中心位置；至少基于所述预定衰减系数 α ，所述邻居中心位置，确定所述位置偏移项。

[102] 在另一例子中，偏移项确定模块 442 配置为：确定节点 i 与其所有邻居节点的关联强度的和值；确定节点 i 与各个邻居节点之间的关联强度与所述和值的比例，作为相对关联强度；以所述相对关联强度为权重，对各个邻居节点的当前嵌入向量求和，确定邻居中心位置；将邻居中心位置与所述预定衰减系数 α 的乘积，作为所述位置偏移项。

[103] 根据一种可能的设计，条件判定单元 45 所依据的预定收敛条件可以是：对于每个节点，本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值小于第一预定值；或者，各个节点的本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值的总和小于第二预定值。

[104] 根据一种可能的设计，预定收敛条件也可以是，确定各个节点 i 的当前嵌入向量 E_i 的次数达到预定次数阈值。

[105] 在一个实施例中，嵌入位置确定单元 46 配置为，将节点 i 的嵌入向量确定为，满足所述预定收敛条件时节点 i 的当前嵌入向量 E_i 与其位置初始项之差。

[106] 通过以上的方法和装置，可以快速有效地将复杂的关系网络图嵌入到任意维度的多维空间中，从而便于后续的节点信息处理。

[107] 根据另一方面的实施例，还提供一种计算机可读存储介质，其上存储有计算机

程序，当所述计算机程序在计算机中执行时，令计算机执行结合图 2 所描述的方法。

[108] 根据再一方面的实施例，还提供一种计算设备，包括存储器和处理器，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现结合图 2 所述的方法。

- 5 [109] 本领域技术人员应该可以意识到，在上述一个或多个示例中，本发明所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。

- 10 [110] 以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的技术方案的基础之上，所做的任何修改、等同替换、改进等，均应包括在本发明的保护范围之内。

权利要求书

1.一种将关系网络图嵌入到多维空间的方法,所述关系网络图包括多个节点,所述多个节点中具有关联关系的节点以一定关联强度互相连接,所述方法包括:

随机确定所述多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i ;

5 对于各个节点 i , 获取与该节点 i 相连接的邻居节点, 以及该节点 i 与各个邻居节点之间的关联强度;

确定该节点 i 的各个邻居节点的当前嵌入向量;

获取该节点 i 的位置初始项和位置偏移项, 并根据所述位置初始项和位置偏移项, 确定节点 i 的当前嵌入向量 E_i , 其中所述位置初始项基于所述初始嵌入向量 C_i 而确定,

10 所述位置偏移项根据预定衰减系数 α 、所述各个邻居节点的当前嵌入向量以及该节点 i 与各个邻居节点之间的关联强度而确定;

判断预定收敛条件是否得到满足, 在不满足该预定收敛条件的情况下, 再次确定该节点 i 的各个邻居节点的当前嵌入向量, 以及再次确定节点 i 的当前嵌入向量 E_i , 直到该预定收敛条件得到满足;

15 至少基于满足所述预定收敛条件的各个节点 i 的当前嵌入向量 E_i , 确定各个节点 i 在所述多维空间的嵌入向量。

2.根据权利要求 1 所述的方法, 其中获取与该节点 i 相连接的邻居节点, 以及该节点 i 与各个邻居节点之间的关联强度包括:

20 获取记录所述关系网络图的网络关系的邻接矩阵, 所述邻接矩阵中第 m 行第 k 列的元素对应于第 m 节点与第 k 节点之间的关联强度;

通过所述邻接矩阵, 确定节点 i 的邻居节点, 以及节点 i 与各个邻居节点之间的关联强度。

3.根据权利要求 2 所述的方法, 其中通过所述邻接矩阵, 确定节点 i 的邻居节点, 以及节点 i 与各个邻居节点之间的关联强度包括:

25 获取所述邻接矩阵中与节点 i 对应的第 i 行元素或第 i 列元素;

将所述第 i 行元素或第 i 列元素中非零元素对应的节点确定为节点 i 的邻居节点; 将所述非零元素的值确定为节点 i 与对应邻居节点之间的关联强度。

4.根据权利要求 1 所述的方法, 其中所述获取该节点 i 的位置初始项包括, 基于所述初始嵌入向量 C_i 以及所述预定衰减系数, 确定所述位置初始项。

30 5.根据权利要求 1 所述的方法, 其中获取该节点 i 的位置偏移项包括:

以节点 i 与各个邻居节点之间的关联强度为权重, 对各个邻居节点的当前嵌入向量

求和，确定邻居中心位置；

至少基于所述预定衰减系数 α ，所述邻居中心位置，确定所述位置偏移项。

6.根据权利要求1所述的方法，其中获取该节点 i 的位置偏移项包括：

确定节点 i 与其所有邻居节点的关联强度的和值；

- 5 确定节点 i 与各个邻居节点之间的关联强度与所述和值的比例，作为相对关联强度；
以所述相对关联强度为权重，对各个邻居节点的当前嵌入向量求和，确定邻居中心位置；

将邻居中心位置与所述预定衰减系数 α 的乘积，作为所述位置偏移项。

7.根据权利要求1所述的方法，其中所述预定收敛条件包括：

- 10 对于每个节点，本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值小于第一预定值；或者

各个节点的本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值的总和小于第二预定值。

- 15 8.根据权利要求1所述的方法，其中所述预定收敛条件包括：确定各个节点 i 的当前嵌入向量 E_i 的次数达到预定次数阈值。

9.根据权利要求1所述的方法，其中确定各个节点 i 在所述多维空间的嵌入向量包括，将节点 i 的嵌入向量确定为，满足所述预定收敛条件时节点 i 的当前嵌入向量 E_i 与其位置初始项之差。

- 20 10.一种将关系网络图嵌入到多维空间的装置，所述关系网络图包括多个节点，所述多个节点中具有关联关系的节点以一定关联强度互相连接，所述装置包括：

初始位置确定单元，配置为随机确定所述多个节点中各个节点 i 在多维空间的初始嵌入向量 C_i ；

- 25 邻居节点确定单元，配置为对于各个节点 i ，获取与该节点 i 相连接的邻居节点，以及该节点 i 与各个邻居节点之间的关联强度；

邻居位置确定单元，配置为确定该节点 i 的各个邻居节点的当前嵌入向量；

- 30 节点位置确定单元，配置为获取该节点 i 的位置初始项和位置偏移项，并根据所述位置初始项和位置偏移项，确定节点 i 的当前嵌入向量 E_i ，其中所述位置初始项基于所述初始嵌入向量 C_i 而确定，所述位置偏移项根据预定衰减系数 α 、所述各个邻居节点的当前嵌入向量以及该节点 i 与各个邻居节点之间的关联强度而确定；

条件判定单元，配置为判断预定收敛条件是否得到满足，在不满足该预定收敛条件

的情况下,使得所述邻居位置确定单元再次确定该节点 i 的各个邻居节点的当前嵌入向量,所述节点位置确定单元再次确定节点 i 的当前嵌入向量 E_i ,直到该预定收敛条件得到满足;

5 嵌入位置确定单元,配置为至少基于满足所述预定收敛条件的各个节点 i 的当前嵌入向量 E_i ,确定各个节点 i 在所述多维空间的嵌入向量。

11.根据权利要求10所述的装置,其中所述邻居节点确定单元配置为:

获取记录所述关系网络图的网络关系的邻接矩阵,所述邻接矩阵中第 m 行第 k 列的元素对应于第 m 节点与第 k 节点之间的关联强度;

10 通过所述邻接矩阵,确定节点 i 的邻居节点,以及节点 i 与各个邻居节点之间的关联强度。

12.根据权利要求11所述的装置,其中所述邻居节点确定单元配置为:

获取所述邻接矩阵中与节点 i 对应的第 i 行元素或第 i 列元素;

将所述第 i 行元素或第 i 列元素中非零元素对应的节点确定为节点 i 的邻居节点;

将所述非零元素的值确定为节点 i 与对应邻居节点之间的关联强度。

15 13.根据权利要求10所述的装置,其中所述节点位置确定单元包括,初始项确定模块,配置为基于所述初始嵌入向量 C_i 以及所述预定衰减系数,确定所述位置初始项。

14.根据权利要求10所述的装置,其中所述节点位置确定单元包括偏移项确定模块,配置为:

20 以节点 i 与各个邻居节点之间的关联强度为权重,对各个邻居节点的当前嵌入向量求和,确定邻居中心位置;

至少基于所述预定衰减系数 α ,所述邻居中心位置,确定所述位置偏移项。

15.根据权利要求10所述的装置,其中所述节点位置确定单元包括偏移项确定模块,配置为:

确定节点 i 与其所有邻居节点的关联强度的和值;

25 确定节点 i 与各个邻居节点之间的关联强度与所述和值的比例,作为相对关联强度;以所述相对关联强度为权重,对各个邻居节点的当前嵌入向量求和,确定邻居中心位置;

将邻居中心位置与所述预定衰减系数 α 的乘积,作为所述位置偏移项。

16.根据权利要求10所述的装置,其中所述预定收敛条件包括:

30 对于每个节点,本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值小于第一预定值;或者

各个节点的本次确定的当前嵌入向量与前次确定的当前嵌入向量的差值的总和小于第二预定值。

17.根据权利要求 10 所述的装置，其中所述预定收敛条件包括：确定各个节点 i 的当前嵌入向量 E_i 的次数达到预定次数阈值。

5 18.根据权利要求 10 所述的装置，其中所述嵌入位置确定单元配置为，将节点 i 的嵌入向量确定为，满足所述预定收敛条件时节点 i 的当前嵌入向量 E_i 与其位置初始项之差。

19.一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行权利要求 1-9 中任一项的所述的方法。

10 20.一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-9 中任一项所述的方法。

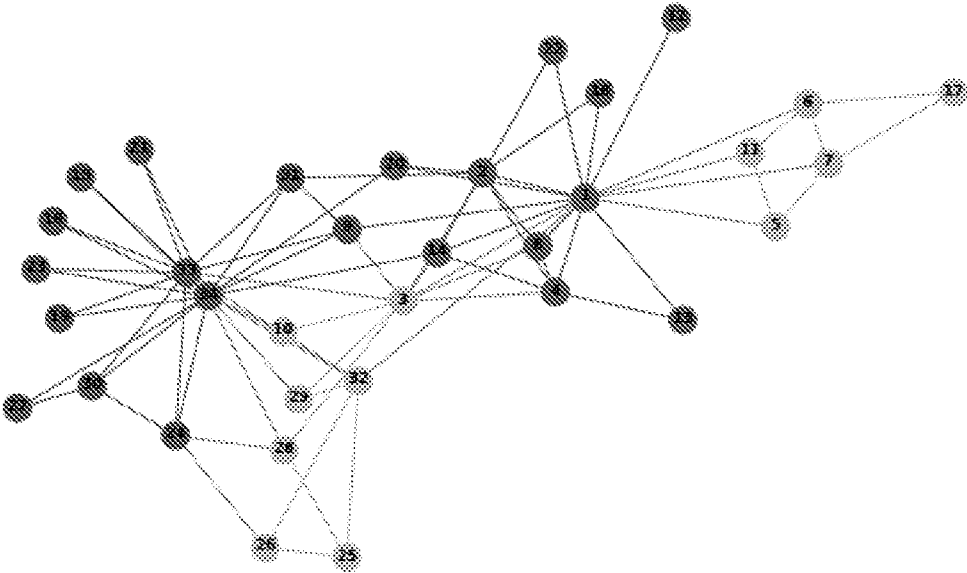


图 1

2 / 3

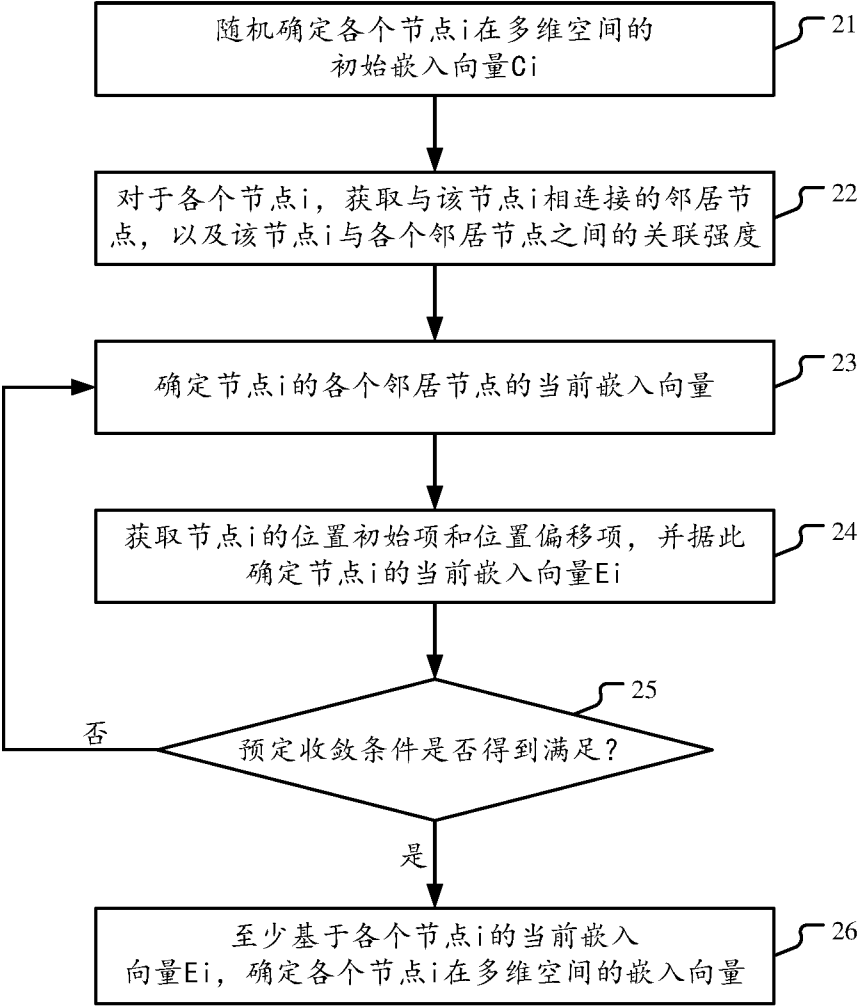


图 2

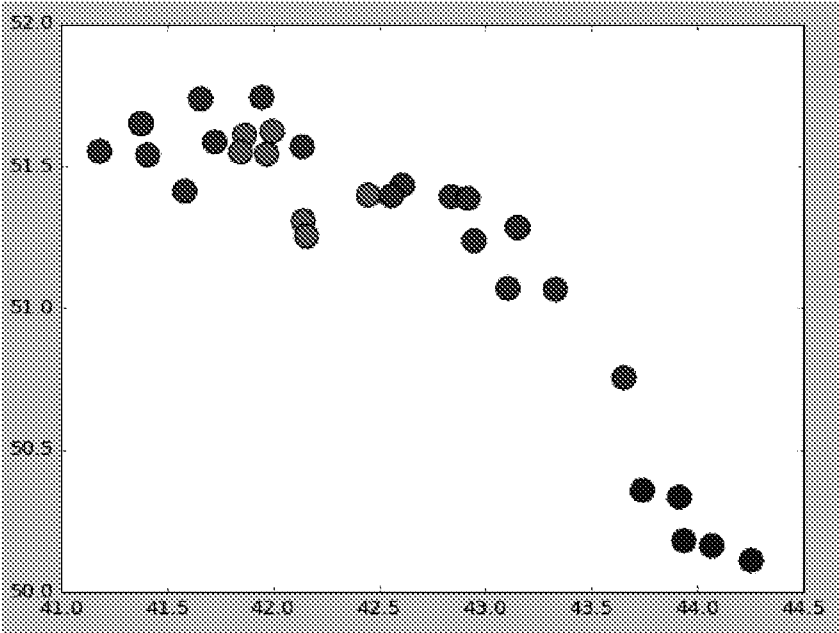


图 3

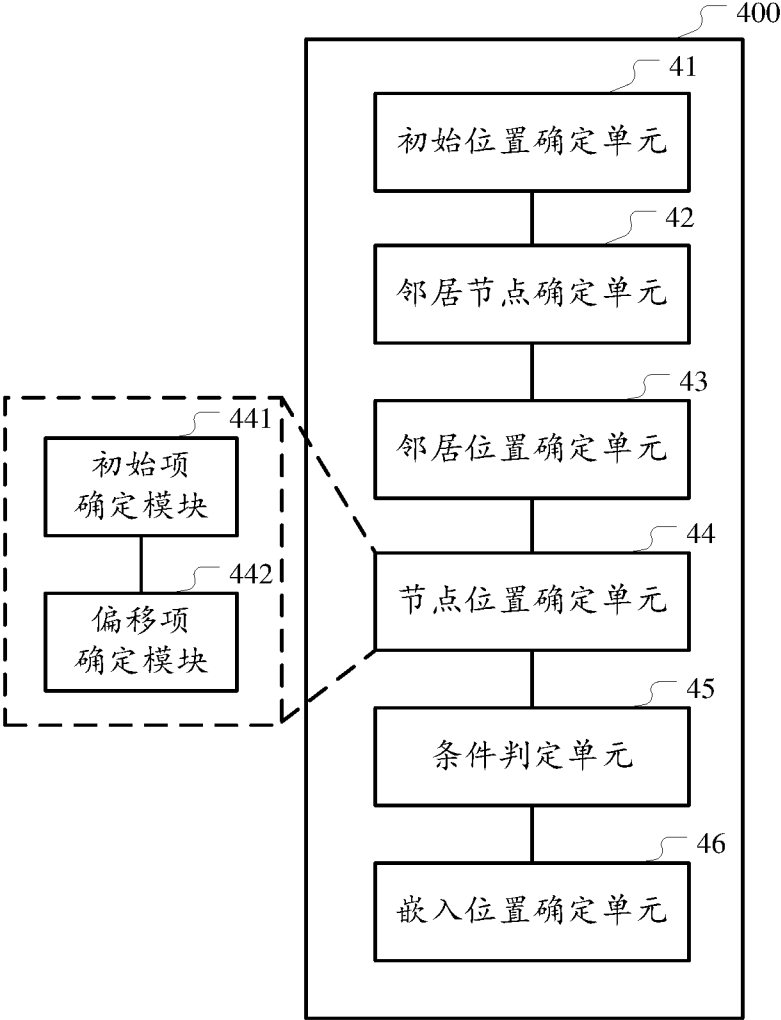


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/089022

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/901(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F16/-;G06F17/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI; SIPOABS; CNABS; CNTXT; CNKI: 图嵌入, 顶点, 节点, 向量, 嵌入, 映射; graph embed????, node?, vertice?, vertex??, vector, map, embed????,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109063041 A (ALIBABA GROUP HOLDING LIMITED) 21 December 2018 (2018-12-21) claims 1-20	1-20
A	US 2010121792 A1 (MICROSOFT CORP.) 13 May 2010 (2010-05-13) entire document	1-20
A	CN 107633263 A (TSINGHUA UNIVERSITY) 26 January 2018 (2018-01-26) entire document	1-20
A	CN 107145977 A (UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA) 08 September 2017 (2017-09-08) entire document	1-20
A	CN 107392782 A (PHICOMM (SHANGHAI) CO., LTD.) 24 November 2017 (2017-11-24) entire document	1-20

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

08 August 2019

Date of mailing of the international search report

20 August 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/089022

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109063041	A	21 December 2018	None			
US	2010121792	A1	13 May 2010	WO	2008086323	A1	17 July 2008
				EP	2100228	A1	16 September 2009
CN	107633263	A	26 January 2018	None			
CN	107145977	A	08 September 2017	None			
CN	107392782	A	24 November 2017	None			

国际检索报告

国际申请号

PCT/CN2019/089022

A. 主题的分类

G06F 16/901(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F16/-;G06F17/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

DWPI;SIPOABS;CNABS;CNTXT;CNKI: 图嵌入, 顶点, 节点, 向量, 嵌入, 映射; graph embed????, node?, vertice?, vertex??, vector, map, embed????,

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109063041 A (阿里巴巴集团控股有限公司) 2018年 12月 21日 (2018 - 12 - 21) 权利要求第1-20项	1-20
A	US 2010121792 A1 (MICROSOFT CORP) 2010年 5月 13日 (2010 - 05 - 13) 全文	1-20
A	CN 107633263 A (清华大学) 2018年 1月 26日 (2018 - 01 - 26) 全文	1-20
A	CN 107145977 A (电子科技大学) 2017年 9月 8日 (2017 - 09 - 08) 全文	1-20
A	CN 107392782 A (上海斐讯数据通信技术有限公司) 2017年 11月 24日 (2017 - 11 - 24) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 8月 8日

国际检索报告邮寄日期

2019年 8月 20日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

郝晓丽

传真号 (86-10)62019451

电话号码 62412331

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/089022

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109063041	A	2018年 12月 21日	无			
US	2010121792	A1	2010年 5月 13日	W0	2008086323	A1	2008年 7月 17日
				EP	2100228	A1	2009年 9月 16日
CN	107633263	A	2018年 1月 26日	无			
CN	107145977	A	2017年 9月 8日	无			
CN	107392782	A	2017年 11月 24日	无			

表 PCT/ISA/210 (同族专利附件) (2015年1月)