



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201923631 A

(43)公開日：中華民國 108 (2019) 年 06 月 16 日

(21)申請案號：107132215

(22)申請日：中華民國 107 (2018) 年 09 月 13 日

(51)Int. Cl. : **G06F17/30 (2006.01)**

(30)優先權：2017/11/17 中國大陸 201711144728.1

(71)申請人：香港商阿里巴巴集團服務有限公司 (香港地區) ALIBABA GROUP SERVICES LIMITED (HK)

香港

(72)發明人：曹紹升 (CN)；楊新星 (CN)；周俊 (CN)

(74)代理人：林志剛

申請實體審查：有 申請專利範圍項數：28 項 圖式數：7 共 43 頁

(54)名稱

隨機漫步、基於叢集的隨機漫步方法、裝置以及設備

(57)摘要

本說明書實施例公開了隨機漫步、基於叢集的隨機漫步方法、裝置以及設備，方案包括：獲取圖資料包含的各節點的資訊，根據各節點的資訊產生反應節點及其相鄰節點之間的對應關係的二維陣列，根據二維陣列產生隨機序列，實現在圖資料中的隨機漫步；該方案可以既適用於叢集也適用於單機。

指定代表圖：

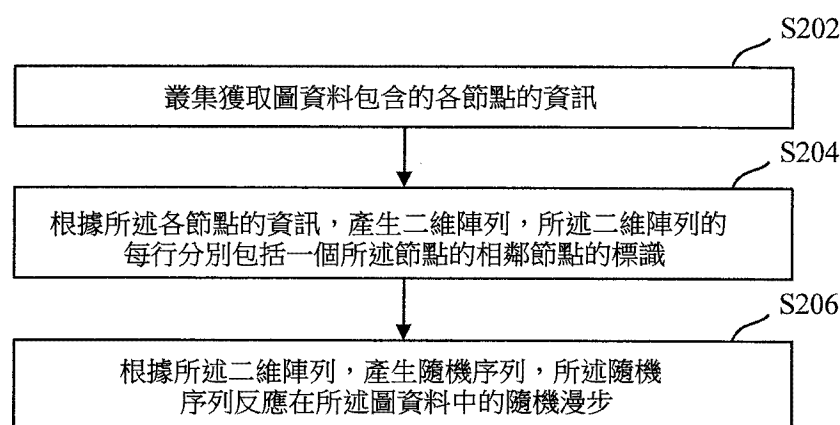


圖 2

【發明說明書】

【中文發明名稱】

隨機漫步、基於叢集的隨機漫步方法、裝置以及設備

【技術領域】

本說明書涉及電腦軟體技術領域，尤其涉及隨機漫步、基於叢集的隨機漫步方法、裝置以及設備。

【先前技術】

隨著電腦和網際網路技術的迅速發展，很多業務都可以在網上進行，圖計算是處理社交方面的網上業務的一種常用手段。

例如，對於社交風控業務中的帳戶欺詐識別：每個使用者分別作為一個節點，若兩個使用者之間存在轉帳關係，則對應的兩個節點之間存在一條邊，邊可以是無向的，也可以是根據轉帳方向定義了方向的；以此類推，可以得到包含多個節點和多條邊的圖資料，進而基於圖資料進行圖計算以實現風控。

隨機漫步演算法是圖計算中比較基礎和重要的一環，其為上層複雜演算法提供支持。在現有技術中，一般採用這樣的隨機漫步演算法：在資料庫中隨機讀取圖資料包含的一個節點，再繼續在該資料庫中隨機讀取該節點的一個相鄰節點，以此類推，實現在圖資料中的隨機漫步。

基於現有技術，需要能夠應用於大規模圖資料的更為

高效的隨機漫步方案。

【發明內容】

本說明書實施例提供隨機漫步、基於叢集的隨機漫步方法、裝置以及設備，用以解決如下技術問題：需要能夠應用於大規模圖資料的更為高效的隨機漫步方案。

為解決上述技術問題，本說明書實施例是這樣實現的：

本說明書實施例提供的一種基於叢集的隨機漫步方法，包括：

所述叢集獲取圖資料包含的各節點的資訊；

根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

本說明書實施例提供的一種隨機漫步方法，包括：

獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

本說明書實施例提供的一種基於叢集的隨機漫步裝置，所述裝置屬於所述叢集，包括：

獲取模組，獲取圖資料包含的各節點的資訊；

第一產生模組，根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

第二產生模組，根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

本說明書實施例提供的一種隨機漫步裝置，包括：

獲取模組，獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

產生模組，根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

本說明書實施例提供的一種基於叢集的隨機漫步設備，所述設備屬於所述叢集，包括：

至少一個處理器；以及，

與所述至少一個處理器通信連接的記憶體；其中，

所述記憶體儲存有可被所述至少一個處理器執行的指令，所述指令被所述至少一個處理器執行，以使所述至少一個處理器能夠：

獲取圖資料包含的各節點的資訊；

根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

本說明書實施例提供的一種隨機漫步設備，包括：

至少一個處理器；以及，
與所述至少一個處理器通信連接的記憶體；其中，
所述記憶體儲存有可被所述至少一個處理器執行的指令，所述指令被所述至少一個處理器執行，以使所述至少一個處理器能夠：

獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

本說明書實施例採用的上述至少一個技術方案能夠達到以下有益效果：有利於減少對原始保存圖資料的資料庫的存取，二維陣列在產生後無需依賴該資料庫，通過二維陣列能夠快速索引節點的相鄰節點，該方案能夠適用於大規模圖資料且效率較高，在基於叢集實施該方案的情況下，還能夠進一步地提高效率。

【圖式簡單說明】

為了更清楚地說明本說明書實施例或現有技術中的技術方案，下面將對實施例或現有技術描述中所需要使用的附圖作簡單地介紹，顯而易見地，下面描述中的附圖僅僅是本說明書中記載的一些實施例，對於本領域普通技術人員來講，在不付出創造性勞動性的前提下，還可以根據這些附圖獲得其他的附圖。

圖1為本說明書的方案在一種實際應用情境下涉及的一種整體架構示意圖；

圖2為本說明書實施例提供的一種基於叢集的隨機漫步方法的流程示意圖；

圖3為本說明書實施例提供的一種實際應用情境下，基於叢集的二維陣列產生流程示意圖；

圖4為本說明書實施例提供的一種實際應用情境下，基於叢集的隨機序列產生流程示意圖；

圖5為本說明書實施例提供的一種隨機漫步方法的流程示意圖；

圖6為本說明書實施例提供的對應於圖2的一種基於叢集的隨機漫步裝置的結構示意圖；

圖7為本說明書實施例提供的對應於圖5的一種隨機漫步裝置的結構示意圖。

【實施方式】

本說明書實施例提供隨機漫步、基於叢集的隨機漫步方法、裝置以及設備。

為了使本技術領域的人員更好地理解本說明書中的技術方案，下面將結合本說明書實施例中的附圖，對本說明書實施例中的技術方案進行清楚、完整地描述，顯然，所描述的實施例僅僅是本申請一部分實施例，而不是全部的實施例。基於本說明書實施例，本領域普通技術人員在沒有作出創造性勞動前提下所獲得的所有其他實施例，都應

當屬於本申請保護的範圍。

本說明書的方案既適用於叢集，也適用於單機。在叢集下對於大規模圖資料的處理效率更高，原因在於：可以拆分任務(比如，資料讀取任務、資料同步任務等)，進而由叢集中的多個機器並行執行被分配給自己的一部分任務。以下各實施例主要基於叢集情境進行說明。

方案涉及的叢集可以有一個或者多個，以圖1為例，涉及了兩個叢集。

圖1為本說明書的方案在一種實際應用情境下涉及的一種整體架構示意圖。該整體架構中，主要涉及三部分：伺服器叢集、工作機叢集、資料庫。資料庫保存有圖資料，供叢集讀取，伺服器叢集與工作機叢集相互配合，根據從資料庫讀取的資料，實現在圖資料中的隨機漫步。

圖1中的架構是示例性的，並非唯一。比如，方案可以涉及一個叢集，該叢集中包含至少一個調度機和多個工作機；再比如，方案也可以涉及一個工作機叢集和一個伺服器；等等；方案涉及的機器相互配合，實現在圖資料中的隨機漫步。

下面對本說明書的方案進行詳細說明。

圖2為本說明書實施例提供的一種基於叢集的隨機漫步方法的流程示意圖。圖2中各步驟由叢集中的至少一個機器(或者機器上的程式)執行，不同步驟的執行主體可以不同。

圖2中的流程包括以下步驟：

S202：所述叢集獲取圖資料包含的各節點的資訊。

在本說明書實施例中，節點的資訊可以包括：節點的標識、節點的相鄰節點的標識(以下以此為例)、或者標識以外的能夠指示節點的相鄰節點的資訊等。各節點的資訊可以是一次性獲取的，也可以是分多次獲取的。

一般地，原始的圖資料保存於資料庫中，在這種情況下，需要通過存取資料庫，讀取得到各節點的資訊。為了避免重複讀取資料增加資料庫的負擔，叢集中的多個機器可以分別讀取不重複的一部分節點的資訊，進一步地，多個機器可以並行讀取資料庫，以快速獲取節點的資訊。

例如，可以由工作機叢集中的各工作機並行地、分別從資料庫讀取一部分節點的資訊並進行處理，再將處理後得到的資料同步至伺服器叢集。或者，各工作機也可以將讀取的節點的資訊，直接同步至伺服器叢集，由伺服器叢集進一步地處理。所述處理至少包括產生二維陣列。

S204：根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識。

在本說明書實施例中，二維陣列可以視為矩陣，其每行分別為一個一維陣列。

每行可以分別對應一個節點，該行至少包括其對應的節點的相鄰節點的標識，每個相鄰節點的標識即可以是該行的一個一維陣列元素。為了便於索引，該對應的節點的標識自身也可以是該行的一個一維陣列元素，比如，該對

應的節點的標識為該行的第0個一維陣列元素，之後的一維陣列元素依次為該節點的各相鄰節點的標識。或者，該節點的標識自身可以不包含在該行內，而只是與該行具有關聯關係，通過該關聯關係能夠用該節點的標識索引到該行。

根據二維陣列以及任意節點的標識，能夠快速索引到該節點的任意相鄰節點的標識，從而，有利於高效地在圖資料中隨機漫步。

為了便於索引，各節點的標識首選地為數字。比如，用各節點的標識大小定義各節點的順序，從0開始計數，順序最先的節點的標識為0、順序第二的節點的標識為1，依次類推。以下各實施例基於該例中的定義進行說明。

當然，若節點原本的標識並非數字，也可以基於一一映射的規則，將所述原本的標識映射為數字後，作為節點的標識用於產生二維陣列。

S206：根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

在本說明書實施例中，隨機序列為多個節點的標識構成的序列，各標識在該隨機序列中的順序即為隨機漫步順序，隨機序列的最大長度一般由預定的隨機漫步步數決定。

在得到二維陣列後，可以相互獨立地多次執行步驟**S206**，進而得到多個相互獨立的隨機序列。比如，各工作機分別根據二維陣列，產生一個或者多個隨機序列。

通過圖2的方法，有利於減少對原始保存圖資料的資料庫的存取，二維陣列在產生後無需依賴該資料庫，通過二維陣列能夠快速索引節點的相鄰節點，該方案能夠適用於大規模圖資料且效率較高，由於基於叢集實施該方法，因此，還能夠進一步地提高效率。

基於圖2的方法，本說明書實施例還提供了該方法的一些具體實施方案，以及擴展方案，下面以圖1中的架構為例，進行說明。

在本說明書實施例中，如前所述，所述叢集可以包括伺服器叢集和工作機叢集，對於步驟S202，所述叢集獲取圖資料包含的各節點的資訊，具體可以包括：

所述工作機叢集從資料庫讀取圖資料包含的各節點的相鄰節點的標識，其中，每個工作機讀取一部分節點的相鄰節點的標識。需要說明的是，若對於工作機叢集，節點的標識本身也是未知的，則工作機叢集可以讀取節點的標識，並根據節點的標識(在資料庫中作為主鍵)，讀取節點的相鄰節點的標識。

例如，假定有標識分別為0~4的5個節點。工作機叢集包括工作機0、工作機1、工作機2，每個工作機分別從資料庫讀取一部分節點的相鄰節點的標識；比如，工作機0讀取節點1的相鄰節點的標識(分別為0、2)，以及節點2的標識(分別為1、3、4)；工作機1讀取節點0的相鄰節點的標識(為1)；工作機2讀取節點3的相鄰節點的標識(分別為2、4)，以及節點4的標識(分別為2、3)。

在本說明書實施例中，各工作機可以根據自己讀取標識的相鄰節點及其對應節點的標識，產生非全量的二維陣列。

進一步地，工作機叢集可以將這些非全量的二維陣列同步給伺服器叢集。從而，伺服器叢集能夠得到由這些非全量的二維陣列構成的全量的二維陣列，具體地：伺服器叢集可以通過專門整合(比如，拆分二維陣列、合併二維陣列等)這些非全量的二維陣列，得到全量的二維陣列；也可以不專門整合而只是在工作機叢集同步完畢後，將同步得到的全部資料視為一個整體，即全量的二維陣列。伺服器叢集中的各伺服器可以分別保存全量的二維陣列，也可以只保存全量的二維陣列的一部分。

步驟 S204 所述的二維陣列可以是所述全量的二維陣列，也可以是所述非全量的二維陣列，也可以是對所述全量的二維陣列進一步地處理(比如，重排序等)後得到的二維陣列，下面各實施例主要以第三種情況為例進行說明。需要說明的是，若是非全量的二維陣列，則後續的隨機漫步相應地在該非全量的二維陣列涉及的節點(這些節點只是圖資料中的一部分節點)中進行，任意工作機可以根據自己產生的非全量的二維陣列，產生隨機序列，而未必要依賴於上述同步和伺服器叢集。

對上述同步後的動作繼續說明。伺服器叢集可以進一步地將全量的二維陣列再向各工作機分別同步，以便各工作機能夠根據全量的二維陣列，產生隨機序列。上述的第

三種情況已經提到，各工作機可以對全量的二維陣列進一步處理後，再用於生產隨機序列。

例如，可以根據節點標識順序，對所述全量的二維陣列中的各行進行排序；根據排序後的二維陣列，產生隨機序列。比如，將節點0及其相鄰節點的標識所在的行排在第一行，將節點1及其相鄰節點的標識所在的行排在第二行，以此類推；進一步地，還可以將第一行中節點0的標識剔除，只保留其相鄰節點的標識，並建立節點0與處理後的第一行之間的關聯關係，以便於後續根據節點0的標識索引第一行中的任意相鄰節點的標識，以此類推，處理後的每行中可以只保留相鄰節點的標識。

在本說明書實施例中，在處理後各行中，一維陣列元素數量相等，該元素數量一般不小於各節點中鄰居節點最多的節點的鄰居節點數量。對於用鄰居節點的標識填不滿的行，可以用空元素(也即“null”元素)在行的尾部進行填充。另外，若只有個別節點的鄰居節點數量很多，而其他節點的鄰居數量相比而言少很多，而也可以參照所述其他節點定義處理後各行中一維陣列元素數量，而對於該個別節點的鄰居節點，可以只取一部分鄰居節點作為該個別節點對應行的元素，以免無謂地浪費大量內部記憶體。

根據上面的說明，本說明書實施例提供的一種實際應用情境下，基於叢集的二維陣列產生流程示意圖，如圖3所示。

在圖3中，資料庫中的資料表以節點的標識作為主

鍵，記錄了各節點的相鄰節點的標識，其中，節點0的相鄰節點為節點1，節點1的相鄰節點為節點0、節點2，節點2的相鄰節點為節點1、節點3、節點4，節點3的相鄰節點為節點2、節點4，節點4的相鄰節點為節點2、節點3。工作機0~2如前所述，優選地可以並行分別從資料庫讀取一部分節點的相鄰節點的標識。

每個工作機根據自己讀取的標識，對應地產生非全量的二維陣列。工作機0產生的二維陣列中包含兩行，工作機1產生的二維陣列中包含一個行，工作機2產生的二維陣列中包含兩行。在非全量的二維陣列的每行中，既包括節點的標識，也包括該節點的各相鄰節點的標識。

工作機叢集將產生的各非全量的二維陣列都同步至伺服器叢集，可以看到伺服器叢集得到了全量的二維陣列並分部分保存在伺服器0~2中。

伺服器叢集將全量的二維陣列分別同步給各工作機。則各工作機可以分別獨立地對全量的二維陣列進行排序和節點剔除處理，得到有序的只包含相鄰節點的標識的二維陣列，用於產生隨機序列。

在本說明書實施例中，對於步驟S206，所述根據所述二維陣列，產生隨機序列，具體可以包括：

所述工作機在所述各節點的標識中，隨機確定一個標識，作為目標節點的標識；根據所述目標節點的標識，在所述二維陣列中索引得到對應的行，所述對應的行包括所述目標節點的標識，以及所述目標節點的相鄰節點的標

識；確定所述對應的行包括的相鄰節點的標識的數量；隨機確定一個小於所述數量的非負整數，並獲取所述對應的行包括的第所述非負整數個相鄰節點的標識；通過將該第所述非負整數個相鄰節點重新作為目標節點進行迭代計算，產生由依次得到的各目標節點的標識構成的隨機序列。

進一步地沿用圖3的例子，結合圖4說明。圖4為本說明書實施例提供的一種實際應用情境下，基於叢集的隨機序列產生流程示意圖。

假定圖資料共包含 N 個節點，第 m 個所述節點的標識為 m ， $0 \leq m \leq N-1$ ，所述目標節點為第 i 個節點，所述對應的行為所述二維陣列的第 i 行。所述對應的行為一維陣列，所述目標節點的第 n 個相鄰節點的標識為該一維陣列的第 n 個元素， n 從 0 開始計數，所述非負整數記作 j 。

在圖5中， $N=5$ ，工作機根據伺服器叢集所同步的全量的二維陣列，處理後得到的二維陣列(稱為相鄰節點陣列)中相應地包含5行，依次對應於節點0~4，每行分別為一個一維陣列，一維陣列包括其對應的節點的各相鄰節點的標識，不足部分用“null”元素填充。

工作機隨機產生一個屬於 $[0, N-1=4]$ 的整數，即工作機在各節點的標識中，隨機確定的目標節點的標識；根據目標節點的標識 i ，在相鄰節點陣列索引到第 i 行(為一維陣列)；確定該第 i 行包含的非“null”元素的元素數量；隨機確定一個小於該元素數量的非負整數 j ；通過讀取該第 i 行

的第 j 個元素，得到目標節點的第 j 個相鄰節點的標識。

假定目標節點的標識為 2， $j=1$ 。則目標節點為節點 2，該第 i 行為 [1, 3, 4]，獲取的目標節點的第 1 個相鄰節點的標識為該陣列的第 1 個元素，即 3。從而，實現從節點 2 隨機漫步到節點 3，進而將節點 3 作為目標節點迭代計算，繼續隨機漫步，如此，依次經過的多個節點的標識構成隨機序列。

在圖 4 中，預先設定隨機漫步步數為 8，批數為 5。用矩陣進行表示，隨機漫步步數比如為該矩陣的列數，批數為該矩陣的行數，該矩陣的每一行可以儲存一個隨機序列。

隨機漫步步數定義了一個隨機序列的最大長度，每當隨機序列達到該最大長度時，可以不依賴該隨機序列而開始產生下一個隨機序列。

批數定義了每個工作機在向資料庫寫入已產生前，產生隨機序列的最大個數，到達該最大個數時，工作機可以將自己已產生未寫入的多個隨機序列(表示為對應的矩陣)寫入資料庫。比如，圖 5 中工作機 2 當前已產生未寫入的隨機序列已經到達最大個數 5，則可以將對應的矩陣寫入資料庫。

以圖 4 中工作機 0 產生的第一個隨機序列 (3, 4, 3, 2, 4, 2, 3, 2) 為例，該隨機序列即表示依次經過下列節點的隨機漫步過程：節點 3、節點 4、節點 3、節點 2、節點 4、節點 2、節點 3、節點 2。

進一步地，還可以預先設定臨限值，用於限定整個工作機叢集產生的隨機序列的最大總數量。當到達該設定臨限值時，各工作機可以停止產生隨機序列。

另外，在實際應用中，工作機叢集中的某些工作機可能會出現異常，導致之前用於產生隨機序列的二維陣列丟失。比如，若工作機將該二維陣列只儲存在內部記憶體中，則當機後內部記憶體中的資料會丟失。在這種情況下，當這些工作機恢復正常時，可以從伺服器叢集重新獲取全量的二維陣列並進行處理後用於產生隨機序列。圖4中通過工作機2示出了這種情況。

上面主要是基於叢集情境，對本說明書的方案進行說明的，本說明書的方案也可以脫離叢集情境。比如，基於同樣的思路，本說明書實施例還提供了一種隨機漫步方法的流程示意圖，如圖5所示。

圖5中的流程的執行主體可以是單一的計算設備，也可以是多個計算設備，該流程包括以下步驟：

S502：獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識。

在步驟S502中，二維陣列具體由誰產生，本申請並不做限定。一般地，只要圖資料未發生變化，根據該圖資料已產生的二維陣列可以一直複用。

S504：根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

基於同樣的思路，本說明書實施例還提供了上面各方法的對應裝置，如圖6、圖7所示。

圖6為本說明書實施例提供的對應於圖2的一種基於叢集的隨機漫步裝置的結構示意圖，該裝置屬於所述叢集，包括：

獲取模組601，獲取圖資料包含的各節點的資訊；

第一產生模組602，根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

第二產生模組603，根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

可選地，所述叢集包括伺服器叢集和工作機叢集；

所述獲取模組601獲取圖資料包含的各節點的資訊，具體包括：

所述工作機叢集從資料庫讀取圖資料包含的各節點的相鄰節點的標識，其中，每個工作機讀取一部分節點的相鄰節點的標識。

可選地，所述第一產生模組602根據所述各節點的資訊，產生二維陣列，具體包括：

各所述工作機分別根據自己讀取標識的相鄰節點及其對應節點的標識，產生非全量的二維陣列；

所述工作機叢集將各所述非全量的二維陣列向所述伺服器叢集同步；

所述伺服器叢集根據各所述非全量的二維陣列，得到

全量的二維陣列。

可選地，所述第二產生模組 603 根據所述二維陣列，產生隨機序列前，所述伺服器叢集將所述全量的二維陣列向各所述工作機同步，以便各所述工作機根據所述全量的二維陣列，產生隨機序列。

可選地，所述第二產生模組 603 根據所述二維陣列，產生隨機序列，具體包括：

所述工作機根據節點標識順序，對所述全量的二維陣列中的各行進行排序；

根據排序後的二維陣列，產生隨機序列。

可選地，所述第二產生模組 603 根據所述二維陣列，產生隨機序列，具體包括：

所述工作機在所述各節點的標識中，隨機確定一個標識，作為目標節點的標識；

根據所述目標節點的標識，在所述二維陣列中索引得到對應的行，所述對應的行包括所述目標節點的標識，以及所述目標節點的相鄰節點的標識；

確定所述對應的行包括的相鄰節點的標識的數量；

隨機確定一個小於所述數量的非負整數，並獲取所述對應的行包括的第所述非負整數個相鄰節點的標識；

通過將該第所述非負整數個相鄰節點重新作為目標節點進行迭代計算，產生由依次得到的各目標節點的標識構成的隨機序列。

可選地，所述節點總數量為 N ，第 m 個所述節點的標

識為 m ， $0 \leq m \leq N-1$ ，所述目標節點為第 i 個節點，所述對應的行為所述二維陣列的第 i 行。

可選地，所述對應的行為一維陣列，所述目標節點的第 n 個相鄰節點的標識為該一維陣列的第 n 個元素， n 從 0 開始計數；

所述非負整數記作 j ，所述工作機獲取所述對應的行包括的第所述非負整數個相鄰節點的標識，具體包括：

所述工作機通過讀取該一維陣列的第 j 個元素，得到所述目標節點的第 j 個相鄰節點的標識。

可選地，所述一維陣列的元素總數量等於所述各節點中鄰居節點最多的節點的鄰居節點數量。

可選地，所述工作機產生由依次得到的各目標節點的標識構成的隨機序列，具體包括：

所述工作機當依次得到的各目標節點總數量達到預設的隨機漫步步數時，產生由所述依次得到的各目標節點的標識構成的隨機序列。

可選地，所述第二產生模組 603 產生隨機序列，具體包括：

各所述工作機分別產生隨機序列，直至產生的隨機序列總數量達到設定臨限值。

可選地，所述工作機若本地已有的所述二維陣列丟失，則重新從所述伺服器叢集獲取。

圖 7 為本說明書實施例提供的對應於圖 5 的一種隨機漫步裝置的結構示意圖，該裝置包括：

獲取模組 701，獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

產生模組 702，根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

基於同樣的思路，本說明書實施例還提供了對應於圖 2 的一種基於叢集的隨機漫步設備，該設備屬於所述叢集，包括：

至少一個處理器；以及，

與所述至少一個處理器通信連接的記憶體；其中，

所述記憶體儲存有可被所述至少一個處理器執行的指令，所述指令被所述至少一個處理器執行，以使所述至少一個處理器能夠：

獲取圖資料包含的各節點的資訊；

根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

基於同樣的思路，本說明書實施例還提供了對應於圖 5 的一種隨機漫步設備，包括：

至少一個處理器；以及，

與所述至少一個處理器通信連接的記憶體；其中，

所述記憶體儲存有可被所述至少一個處理器執行的指令，所述指令被所述至少一個處理器執行，以使所述至少

一個處理器能夠：

獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

基於同樣的思路，本說明書實施例還提供了對應於圖2的一種非揮發性電腦儲存媒體，儲存有電腦可執行指令，所述電腦可執行指令設置為：

獲取圖資料包含的各節點的資訊；

根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

基於同樣的思路，本說明書實施例還提供了對應於圖5的一種非揮發性電腦儲存媒體，儲存有電腦可執行指令，所述電腦可執行指令設置為：

獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

上述對本說明書特定實施例進行了描述。其它實施例在所附申請專利範圍的範圍內。在一些情況下，在申請專

利範圍中記載的動作或步驟可以按照不同於實施例中的順序來執行並且仍然可以實現期望的結果。另外，在附圖中描繪的過程不一定要求示出的特定順序或者連續順序才能實現期望的結果。在某些實施方式中，多任務處理和並行處理也是可以的或者可能是有利的。

本說明書中的各個實施例均採用遞進的方式描述，各個實施例之間相同相似的部分互相參見即可，每個實施例重點說明的都是與其他實施例的不同之處。尤其，對於裝置、設備、非揮發性電腦儲存媒體實施例而言，由於其基本相似於方法實施例，所以描述的比較簡單，相關之處參見方法實施例的部分說明即可。

本說明書實施例提供的裝置、設備、非揮發性電腦儲存媒體與方法是對應的，因此，裝置、設備、非揮發性電腦儲存媒體也具有與對應方法類似的有益技術效果，由於上面已經對方法的有益技術效果進行了詳細說明，因此，這裡不再贅述對應裝置、設備、非揮發性電腦儲存媒體的有益技術效果。

在20世紀90年代，對於一個技術的改進可以很明顯地區分是硬體上的改進(例如，對二極體、電晶體、開關等電路結構的改進)還是軟體上的改進(對於方法流程的改進)。然而，隨著技術的發展，當今的很多方法流程的改進已經可以視為硬體電路結構的直接改進。設計人員幾乎都通過將改進的方法流程編程到硬體電路中來得到相應的硬體電路結構。因此，不能說一個方法流程的改進就不能

用硬體實體模組來實現。例如，可程式設計邏輯裝置 (Programmable Logic Device, PLD)(例如現場可程式閘陣列 (Field Programmable Gate Array, FPGA))就是這樣一種積體電路，其邏輯功能由使用者對裝置編程來確定。由設計人員自行編程來把一個數字系統“集成”在一片PLD上，而不需要請晶片製造廠商來設計和製作專用的積體電路晶片。而且，如今，取代手工地製作積體電路晶片，這種編程也多半改用“邏輯編譯器(logic compiler)”軟體來實現，它與程式開發撰寫時所用的軟體編譯器相類似，而要編譯之前的原始代碼也得用特定的編程語言來撰寫，此稱之為硬體描述語言(Hardware Description Language, HDL)，而HDL也並非僅有一種，而是有許多種，如ABEL(Advanced Boolean Expression Language)、AHDL(Altera Hardware Description Language)、Confluence、CUPL(Cornell University Programming Language)、HDCal、JHDL(Java Hardware Description Language)、Lava、Lola、MyHDL、PALASM、RHDL(Ruby Hardware Description Language)等，目前最普遍使用的是VHDL(Very-High-Speed Integrated Circuit Hardware Description Language)與Verilog。本領域技術人員也應該清楚，只需要將方法流程用上述幾種硬體描述語言稍作邏輯編程並編程到積體電路中，就可以很容易得到實現該邏輯方法流程的硬體電路。

控制器可以按任何適當的方式實現，例如，控制器可以採取例如微處理器或處理器以及儲存可由該(微)處理器

執行的電腦可讀程式代碼(例如軟體或韌體)的電腦可讀媒體、邏輯閘、開關、特殊應用積體電路積體電路(Application Specific Integrated Circuit, ASIC)、可程式邏輯控制器和嵌入微控制器的形式，控制器的例子包括但不限於以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20以及Silicone Labs C8051F320，記憶體控制器還可以被實現為記憶體的控制邏輯的一部分。本領域技術人員也知道，除了以純電腦可讀程式代碼方式實現控制器以外，完全可以通過將方法步驟進行邏輯編程來使得控制器以邏輯閘、開關、特殊應用積體電路積體電路、可程式邏輯控制器和嵌入微控制器等的形式來實現相同功能。因此這種控制器可以被認為是一種硬體部件，而對其內包括的用於實現各種功能的裝置也可以視為硬體部件內的結構。或者甚至，可以將用於實現各種功能的裝置視為既可以是實現方法的軟體模組又可以是硬體部件內的結構。

上述實施例闡明的系統、裝置、模組或單元，具體可以由電腦晶片或實體實現，或者由具有某種功能的產品來實現。一種典型的實現設備為電腦。具體的，電腦例如可以為個人電腦、膝上型電腦、蜂窩電話、相機電話、智慧型電話、個人數位助理、媒體播放器、導航設備、電子郵件設備、遊戲控制台、平板電腦、可穿戴設備或者這些設備中的任何設備的組合。

為了描述的方便，描述以上裝置時以功能分為各種單

元分別描述。當然，在實施本說明書時可以把各單元的功能在同一個或多個軟體和/或硬體中實現。

本領域內的技術人員應明白，本說明書實施例可提供為方法、系統、或電腦程式產品。因此，本說明書實施例可採用完全硬體實施例、完全軟體實施例、或結合軟體和硬體方面的實施例的形式。而且，本說明書實施例可採用在一個或多個其中包含有電腦可用程式代碼的電腦可用儲存媒體(包括但不限於磁盤記憶體、**CD-ROM**、光學記憶體等)上實施的電腦程式產品的形式。

本說明書是參照根據本說明書實施例的方法、設備(系統)、和電腦程式產品的流程圖和/或方塊圖來描述的。應理解可由電腦程式指令實現流程圖和/或方塊圖中的每一流程和/或方塊、以及流程圖和/或方塊圖中的流程和/或方塊的結合。可提供這些電腦程式指令到通用電腦、專用電腦、嵌入式處理機或其他可程式資料處理設備的處理器以產生一個機器，使得通過電腦或其他可程式資料處理設備的處理器執行的指令產生用於實現在流程圖一個流程或多個流程和/或方塊圖一個方塊或多個方塊中指定的功能的裝置。

這些電腦程式指令也可儲存在能引導電腦或其他可程式資料處理設備以特定方式工作的電腦可讀記憶體中，使得儲存在該電腦可讀記憶體中的指令產生包括指令裝置的製造品，該指令裝置實現在流程圖一個流程或多個流程和/或方塊圖一個方塊或多個方塊中指定的功能。

這些電腦程式指令也可裝載到電腦或其他可程式資料處理設備上，使得在電腦或其他可程式設備上執行一系列操作步驟以產生電腦實現的處理，從而在電腦或其他可程式設備上執行的指令提供用於實現在流程圖一個流程或多個流程和/或方塊圖一個方塊或多個方塊中指定的功能的步驟。

在一個典型的配置中，計算設備包括一個或多個處理器(CPU)、輸入/輸出介面、網路介面和內部記憶體。

內部記憶體可能包括電腦可讀媒體中的非永久性記憶體，隨機存取記憶體(RAM)和/或非揮發性內部記憶體等形式，如唯讀記憶體(ROM)或閃存(flash RAM)。內部記憶體是電腦可讀媒體的示例。

電腦可讀媒體包括永久性和非永久性、可移動和非可移動媒體可以由任何方法或技術來實現資訊儲存。資訊可以是電腦可讀指令、資料結構、程式的模組或其他資料。電腦的儲存媒體的例子包括，但不限於相變記憶體(PRAM)、靜態隨機存取記憶體(SRAM)、動態隨機存取記憶體(DRAM)、其他類型的隨機存取記憶體(RAM)、唯讀記憶體(ROM)、電可抹除可編程唯讀記憶體(EEPROM)、快閃記憶體或其他內部記憶體技術、唯讀光碟唯讀記憶體(CD-ROM)、數位多功能光碟(DVD)或其他光學儲存、磁盒式磁帶，磁帶式磁碟儲存器或其他磁性儲存設備或任何其他非傳輸媒體，可用於儲存可以被電腦設備存取的資訊。按照本文中的界定，電腦可讀媒體不包括暫存電腦可

讀媒體(transitory media)，如調變的資料信號和載波。

還需要說明的是，術語“包括”、“包含”或者其任何其他變體意在涵蓋非排他性的包含，從而使得包括一系列要素的過程、方法、商品或者設備不僅包括那些要素，而且還包括沒有明確列出的其他要素，或者是還包括為這種過程、方法、商品或者設備所固有的要素。在沒有更多限制的情況下，由語句“包括一個……”限定的要素，並不排除在包括所述要素的過程、方法、商品或者設備中還存在另外的相同要素。

本說明書可以在由電腦執行的電腦可執行指令的一般上下文中描述，例如程式模組。一般地，程式模組包括執行特定任務或實現特定抽象資料類型的例程、程式、物件、組件、資料結構等等。也可以在分布式計算環境中實踐本說明書，在這些分布式計算環境中，由通過通信網路而被連接的遠端處理設備來執行任務。在分布式計算環境中，程式模組可以位於包括儲存設備在內的本地和遠端電腦儲存媒體中。

本說明書中的各個實施例均採用遞進的方式描述，各個實施例之間相同相似的部分互相參見即可，每個實施例重點說明的都是與其他實施例的不同之處。尤其，對於系統實施例而言，由於其基本相似於方法實施例，所以描述的比較簡單，相關之處參見方法實施例的部分說明即可。

以上所述僅為本說明書實施例而已，並不用於限制本申請。對於本領域技術人員來說，本申請可以有各種更改

和變化。凡在本申請的精神和原理之內所作的任何修改、等同替換、改進等，均應包含在本申請的申請專利範圍之內。

【符號說明】

S202：步驟

S204：步驟

S206：步驟

S502：步驟

S504：步驟

601：獲取模組

602：第一產生模組

603：第二產生模組

701：獲取模組

702：產生模組



201923631

【發明摘要】

【中文發明名稱】

隨機漫步、基於叢集的隨機漫步方法、裝置以及設備

【中文】

本說明書實施例公開了隨機漫步、基於叢集的隨機漫步方法、裝置以及設備，方案包括：獲取圖資料包含的各節點的資訊，根據各節點的資訊產生反應節點及其相鄰節點之間的對應關係的二維陣列，根據二維陣列產生隨機序列，實現在圖資料中的隨機漫步；該方案可以既適用於叢集也適用於單機。

【指定代表圖】第(2)圖。

【代表圖之符號簡單說明】無

【特徵化學式】無

【發明申請專利範圍】

【第1項】

一種基於叢集的隨機漫步方法，包括：

所述叢集獲取圖資料包含的各節點的資訊；

根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

【第2項】

根據請求項1所述的方法，所述叢集包括伺服器叢集和工作機叢集；

所述叢集獲取圖資料包含的各節點的資訊，具體包括：

所述工作機叢集從資料庫讀取圖資料包含的各節點的相鄰節點的標識，其中，每個工作機讀取一部分節點的相鄰節點的標識。

【第3項】

根據請求項2所述的方法，所述根據所述各節點的資訊，產生二維陣列，具體包括：

各所述工作機分別根據自己讀取標識的相鄰節點及其對應節點的標識，產生非全量的二維陣列；

所述工作機叢集將各所述非全量的二維陣列向所述伺服器叢集同步；

所述伺服器叢集根據各所述非全量的二維陣列，得到

全量的二維陣列。

【第4項】

根據請求項3所述的方法，所述根據所述二維陣列，產生隨機序列前，所述方法還包括：

所述伺服器叢集將所述全量的二維陣列向各所述工作機同步，以便各所述工作機根據所述全量的二維陣列，產生隨機序列。

【第5項】

根據請求項4所述的方法，所述根據所述二維陣列，產生隨機序列，具體包括：

根據節點標識順序，對所述全量的二維陣列中的各行進行排序；

根據排序後的二維陣列，產生隨機序列。

【第6項】

根據請求項2所述的方法，所述根據所述二維陣列，產生隨機序列，具體包括：

所述工作機在所述各節點的標識中，隨機確定一個標識，作為目標節點的標識；

根據所述目標節點的標識，在所述二維陣列中索引得到對應的行，所述對應的行包括所述目標節點的標識，以及所述目標節點的相鄰節點的標識；

確定所述對應的行包括的相鄰節點的標識的數量；

隨機確定一個小於所述數量的非負整數，並獲取所述對應的行包括的第所述非負整數個相鄰節點的標識；

通過將該第所述非負整數個相鄰節點重新作為目標節點進行迭代計算，產生由依次得到的各目標節點的標識構成的隨機序列。

【第7項】

根據請求項6所述的方法，所述節點總數量為 N ，第 m 個所述節點的標識為 m ， $0 \leq m \leq N-1$ ，所述目標節點為第 i 個節點，所述對應的行為所述二維陣列的第 i 行。

【第8項】

根據請求項6所述的方法，所述對應的行為一維陣列，所述目標節點的第 n 個相鄰節點的標識為該一維陣列的第 n 個元素， n 從 0 開始計數；

所述非負整數記作 j ，所述獲取所述對應的行包括的第所述非負整數個相鄰節點的標識，具體包括：

通過讀取該一維陣列的第 j 個元素，得到所述目標節點的第 j 個相鄰節點的標識。

【第9項】

根據請求項8所述的方法，所述一維陣列的元素總數量等於所述各節點中鄰居節點最多的節點的鄰居節點數量。

【第10項】

根據請求項6所述的方法，所述產生由依次得到的各目標節點的標識構成的隨機序列，具體包括：

當依次得到的各目標節點總數量達到預設的隨機漫步步數時，產生由所述依次得到的各目標節點的標識構成的

隨機序列。

【第11項】

根據請求項2所述的方法，所述產生隨機序列，具體包括：

各所述工作機分別產生隨機序列，直至產生的隨機序列總數量達到設定臨限值。

【第12項】

根據請求項4所述的方法，所述方法還包括：

所述工作機若本地已有的所述二維陣列丟失，則重新從所述伺服器叢集獲取。

【第13項】

一種隨機漫步方法，包括：

獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

【第14項】

一種基於叢集的隨機漫步裝置，所述裝置屬於所述叢集，包括：

獲取模組，獲取圖資料包含的各節點的資訊；

第一產生模組，根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

第二產生模組，根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

【第15項】

根據請求項14所述的裝置，所述叢集包括伺服器叢集和工作機叢集；

所述獲取模組獲取圖資料包含的各節點的資訊，具體包括：

所述工作機叢集從資料庫讀取圖資料包含的各節點的相鄰節點的標識，其中，每個工作機讀取一部分節點的相鄰節點的標識。

【第16項】

根據請求項15所述的裝置，所述第一產生模組根據所述各節點的資訊，產生二維陣列，具體包括：

各所述工作機分別根據自己讀取標識的相鄰節點及其對應節點的標識，產生非全量的二維陣列；

所述工作機叢集將各所述非全量的二維陣列向所述伺服器叢集同步；

所述伺服器叢集根據各所述非全量的二維陣列，得到全量的二維陣列。

【第17項】

根據請求項16所述的裝置，所述第二產生模組根據所述二維陣列，產生隨機序列前，所述伺服器叢集將所述全量的二維陣列向各所述工作機同步，以便各所述工作機根據所述全量的二維陣列，產生隨機序列。

【第18項】

根據請求項17所述的裝置，所述第二產生模組根據所述二維陣列，產生隨機序列，具體包括：

所述工作機根據節點標識順序，對所述全量的二維陣列中的各行進行排序；

根據排序後的二維陣列，產生隨機序列。

【第19項】

根據請求項15所述的裝置，所述第二產生模組根據所述二維陣列，產生隨機序列，具體包括：

所述工作機在所述各節點的標識中，隨機確定一個標識，作為目標節點的標識；

根據所述目標節點的標識，在所述二維陣列中索引得到對應的行，所述對應的行包括所述目標節點的標識，以及所述目標節點的相鄰節點的標識；

確定所述對應的行包括的相鄰節點的標識的數量；

隨機確定一個小於所述數量的非負整數，並獲取所述對應的行包括的第所述非負整數個相鄰節點的標識；

通過將該第所述非負整數個相鄰節點重新作為目標節點進行迭代計算，產生由依次得到的各目標節點的標識構成的隨機序列。

【第20項】

根據請求項19所述的裝置，所述節點總數量為 N ，第 m 個所述節點的標識為 m ， $0 \leq m \leq N-1$ ，所述目標節點為第 i 個節點，所述對應的行為所述二維陣列的第 i 行。

【第21項】

根據請求項19所述的裝置，所述對應的行為一維陣列，所述目標節點的第 n 個相鄰節點的標識為該一維陣列的第 n 個元素， n 從0開始計數；

所述非負整數記作 j ，所述工作機獲取所述對應的行包括的第所述非負整數個相鄰節點的標識，具體包括：

所述工作機通過讀取該一維陣列的第 j 個元素，得到所述目標節點的第 j 個相鄰節點的標識。

【第22項】

根據請求項21所述的裝置，所述一維陣列的元素總數量等於所述各節點中鄰居節點最多的節點的鄰居節點數量。

【第23項】

根據請求項19所述的裝置，所述工作機產生由依次得到的各目標節點的標識構成的隨機序列，具體包括：

所述工作機當依次得到的各目標節點總數量達到預設的隨機漫步步數時，產生由所述依次得到的各目標節點的標識構成的隨機序列。

【第24項】

根據請求項15所述的裝置，所述第二產生模組產生隨機序列，具體包括：

各所述工作機分別產生隨機序列，直至產生的隨機序列總數量達到設定臨限值。

【第25項】

根據請求項17所述的裝置，所述工作機若本地已有的所述二維陣列丟失，則重新從所述伺服器叢集獲取。

【第26項】

一種隨機漫步裝置，包括：

獲取模組，獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

產生模組，根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

【第27項】

一種基於叢集的隨機漫步設備，所述設備屬於所述叢集，包括：

至少一個處理器；以及，

與所述至少一個處理器通信連接的記憶體；其中，

所述記憶體儲存有可被所述至少一個處理器執行的指令，所述指令被所述至少一個處理器執行，以使所述至少一個處理器能夠：

獲取圖資料包含的各節點的資訊；

根據所述各節點的資訊，產生二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

【第28項】

一種隨機漫步設備，包括：

至少一個處理器；以及，
與所述至少一個處理器通信連接的記憶體；其中，
所述記憶體儲存有可被所述至少一個處理器執行的指令，所述指令被所述至少一個處理器執行，以使所述至少一個處理器能夠：

獲取根據圖資料包含的各節點的資訊產生的二維陣列，所述二維陣列的每行分別包括一個所述節點的相鄰節點的標識；

根據所述二維陣列，產生隨機序列，所述隨機序列反應在所述圖資料中的隨機漫步。

