

图数据生成的方法及装置

技术领域

[01]本说明书实施例通常涉及基准测试领域，尤其涉及应用于基准测试的图数据生成的方法及装置。

背景技术

[02]随着图计算技术逐渐成熟，图数据库和图计算被越来越广泛地应用于金融、客服、医疗等领域，尤其是金融领域。在基于图数据实现的应用投入使用之前，需要使用图数据来对该应用进行基准测试，并且只有通过基准测试后的应用才被允许投入使用。如何高效地生成用于基准测试的图数据成为亟待解决的问题。

发明内容

[03]鉴于上述，本说明书实施例提供用于生成应用于基准测试的图数据的方法及装置。利用该方法及装置，可以高效地生成用于基准测试的图数据。

[04]根据本说明书实施例的一个方面，提供一种用于生成应用于基准测试的图数据的方法，包括：创建多个实体顶点以及各个实体顶点的对应实体账户顶点；在各个实体顶点以及对应的实体账户顶点之间创建拥有关系；根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集，所述起点实体账户顶点集和所述终点实体账户顶点集之间不具有重合的实体账户顶点；以及基于所述起点实体账户顶点集和所述终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[05]在上述方面的一个示例中，每个实体账户顶点的账户顶点属性包括账户关联属性，所述方法还可以包括：基于各个实体账户顶点的账户关联属性创建账户属性顶点；以及根据账户关联属性来在各个账户属性顶点之间以及各个账户属性顶点与对应的实体账户顶点之间创建账户属性关系。

[06]在上述方面的一个示例中，所述实体顶点包括个人顶点和组织顶点，所述实体账户顶点包括个人账户顶点和组织账户顶点，以及所述账户属性顶点包括账户注册地址、注册地址、登录网络地址和登录物理地址中的至少一个，其中，所述账户属性关系包括位于关系、电话注册关系、登录网络地址关系和登录物理地址关系中的至少一个。

[07]在上述方面的一个示例中，所述方法还可以包括：获取实体顶点的顶点出度分布信息。此外，创建各个实体顶点的对应实体账户顶点可以包括：根据所述顶点出度分布信息，创建各个实体顶点的对应实体账户顶点。

[08]在上述方面的一个示例中，每个实体账户顶点的账户顶点属性包括顶点出度和顶点入度，基于所述起点实体账户顶点集和所述终点实体账户顶点集创建实体账户顶点之间的账户关联关系可以包括：根据所述起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及所述终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率；基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从所述起点实体账户顶点集和所述终点实体账户顶点集中选择至少一个起点实体账户顶点以及对应的终点实体账户顶点；计算所选择的起点实体账户顶点和对应的终点实体账户顶点之间的属性距离；基于所计算出的属性距离，确定所选择的起点实体账户顶点和对应的终点实体账户顶点之间的关系创建概率；以及根据所述关系创建概率，在所选择的起点实体账户顶点和对应的终点实体账户顶点之间创建账户关联关系。

[09]在上述方面的一个示例中，所述账户关联关系的创建过程被循环执行，直到未创建出新的账户关联关系为止，其中，每次循环过程所使用的关系创建概率通过对上一循环过程的关系创建概率进行衰减处理得到。

[10]在上述方面的一个示例中，从所述起点实体账户顶点和对应的终点实体账户顶点的选择过程到所述账户关联关系的创建过程被循环执行，直到所创建的账户关联关系的数目达到预定数目。

[11]在上述方面的一个示例中,所述方法还可以包括:获取实体账户顶点的顶点出度/入度分布信息;以及根据所述顶点出度/入度分布信息,确定各个实体账户顶点的顶点出度和顶点入度。

[12]在上述方面的一个示例中,所述方法还可以包括:获取社交网络出度/入度分布信息;以及根据所述社交网络出度/入度分布信息来在所述实体顶点之间创建认识/从属关系。此外,基于所计算出的属性距离,确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率可以包括:基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系,确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

[13]在上述方面的一个示例中,根据所述顶点出度分布信息,创建所述多个实体顶点的对应实体账户顶点可以包括:根据所述顶点出度分布信息,创建各个实体顶点的对应实体账户顶点以及业务申请顶点;以及在各个业务申请顶点与对应的实体顶点之间创建申请关系。

[14]在上述方面的一个示例中,所述方法还可以包括:从所述多个实体顶点中抽取多个第一实体顶点。此外,创建各个实体顶点的对应实体账户顶点可以包括:创建各个第一实体顶点的对应实体账户顶点。

[15]根据本说明书的另一实施例,提供一种用于生成应用于基准测试的图数据的方法,包括:经由各个顶点生成框架,分别创建多个实体顶点;经由顶点分块框架,从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点;经由各个顶点生成框架,分别创建所抽取的各个第一实体顶点的对应实体账户顶点,并且在各个实体账户顶点与对应的实体顶点之间创建拥有关系;经由所述顶点分块框架来从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集;以及经由各个顶点关系生成框架,分别基于所抽取的起点实体账户顶点集和终点实体账户顶点集,创建实体账户顶点之间的账户关联关系。

[16]在上述方面的一个示例中,每个实体账户顶点的账户顶点属性包括账户关联属性,所述方法还可以包括:经由各个顶点生成框架,基于各自的实体账户顶点的账户关联属性创建账户属性顶点,并且基于账户关联属性来在各个账户属性顶点之间以及各个账户属性顶点与对应的实体账户顶点之间创建账户属性关系。

[17]在上述方面的一个示例中,从所述顶点分块框架的实体顶点抽取过程到所述各个顶点关系生成框架的账户关联关系创建过程被循环执行。

[18]在上述方面的一个示例中,所述顶点分块框架的顶点抽取过程是不放回抽取过程,并且直到所有顶点被抽取完毕为止。

[19]在上述方面的一个示例中,每个实体账户顶点的账户顶点属性包括顶点出度和顶点入度,经由各个顶点关系生成框架,基于所述起点实体账户顶点集和所述终点实体账户顶点集创建实体账户顶点之间的账户关联关系可以包括:根据所述起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及所述终点实体账户顶点集中的各个终点实体账户顶点的顶点入度,确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率;循环执行下述过程,直到所创建的账户关联关系达到第一预定数目 M : 基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率,从所述起点实体账户顶点集和所述终点实体账户顶点集中选择至少一个起点实体账户顶点以及对应的终点实体账户顶点;计算所选择的起点实体账户顶点和终点实体账户顶点之间的属性距离;基于所计算出的属性距离,确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率;以及基于所述关系创建概率来在所选择的起点实体账户顶点和终点实体账户顶点之间创建账户关联关系。

[20]在上述方面的一个示例中,所述第一预定数目 $M = P/K$, 其中, P 为所述多个实体账户顶点的总出度数量, 以及 K 为循环执行次数。

[21]在上述方面的一个示例中，所述账户关联关系的创建过程被循环执行，直到未创建出新的账户关联关系为止，其中，每次循环过程所使用的关系创建概率通过对上一循环过程的关系创建概率进行衰减处理得到。

[22]在上述方面的一个示例中，所述方法还可以包括：经由各个顶点生成框架的对应数据分布接口获取实体账户顶点的顶点出度/入度分布信息；以及经由各个顶点生成框架根据所获取的顶点出度/入度分布信息，确定各个实体账户顶点的顶点出度和顶点入度。

[23]在上述方面的一个示例中，所述方法还可以包括：经由各个顶点生成框架的对应数据分布接口获取社交网络出度/入度分布信息；以及经由各个顶点生成框架根据所获取社交网络出度/入度分布信息，在所述实体顶点之间创建认识/从属关系。此外，基于所计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率可以包括：基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

[24]在上述方面的一个示例中，所述方法还可以包括：经由各个顶点生成框架的对应数据分布接口获取实体顶点的顶点出度分布信息，以及经由各个顶点生成框架根据所获取的顶点出度分布信息，确定各个实体顶点的顶点出度。此外，经由各个顶点生成框架，分别创建所抽取的各个第一实体顶点的对应实体账户顶点可以包括：经由各个顶点生成框架，分别基于所抽取的各个第一实体顶点的顶点出度，创建所述各个第一实体顶点的对应实体账户顶点。

[25]根据本说明书的实施例的另一方面，提供一种用于生成应用于基准测试的图数据的装置，包括：顶点生成单元，创建多个实体顶点以及各个实体顶点的对应实体账户顶点；拥有关系生成单元，在各个实体顶点以及对应的实体账户顶点之间创建拥有关系；顶点分块单元，根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集，所述起点实体账户顶点集和所述终点实体账户顶点集之间不具有重合的实体账户顶点；以及关联关系生成单元，基于所述起点实体账户顶点集和所述终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[26]根据本说明书的实施例的另一方面，提供一种用于生成应用于基准测试的图数据的装置，包括：至少两个顶点生成框架，每个顶点生成框架部署在一个第一设备处；至少两个顶点关系生成框架，每个顶点关系生成框架部署在一个第二设备处；以及顶点分块框架，部署在第三设备处，其中，各个顶点生成框架被配置为：创建多个实体顶点；创建所述顶点分块框架所抽取的各个第一实体顶点的对应实体账户顶点；以及在各个实体账户顶点和对应的实体顶点之间创建拥有关系，所述顶点分块框架被配置为从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点；以及从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集，各个顶点关系生成框架被配置为基于所抽取的起点实体账户顶点集和终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[27]在上述方面的一个示例中，所述装置还可以包括：部署在各个第一设备处的数据分布接口，获取顶点出度信息，其中，各个实体顶点的顶点出度基于对应的顶点出度分布信息确定。

[28]在上述方面的一个示例中，每个实体账户顶点的账户顶点属性包括顶点出度和顶点入度。各个顶点关系生成框架被配置为：根据所述起点实体账户顶点集中的各个起点实体账户顶点的顶点出度及所述终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率；循环执行下述过程，直到所创建的账户关联关系达到第一预定数目 M：基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从所述起点实体账户顶点集和所述终点实体账户顶点集中选择出至少一个起点实体账户顶点及对应的终点实体账户顶点；计算所选择的起

点实体账户顶点和终点实体账户顶点之间的属性距离；基于计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率；以及基于所述关系创建概率来在所选择的起点实体账户顶点和终点实体账户顶点之间创建账户关联关系。

[29]在上述方面的一个示例中，所述装置还可以包括：部署在各个第一设备处的数据分布接口，获取实体账户顶点的顶点出度/入度分布信息；其中，各个实体账户顶点的顶点出度和顶点入度根据对应的顶点出度/入度分布信息确定。

[30]在上述方面的一个示例中，所述装置还可以包括：部署在各个第一设备处的数据分布接口，获取社交网络出度/入度分布信息；各个顶点生成框架根据所获取的社交网络出度/入度分布信息来在所述实体顶点之间创建认识/从属关系，并且基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

[31]在上述方面的一个示例中，所述多个第一设备中的部分第一设备或每个第一设备分别与所述多个第二设备中的一个第二设备相同，和/或所述第三设备与所述多个第一设备和/或所述多个第二设备中的一个设备相同。

[32]根据本说明书的实施例的另一方面，提供一种用于生成应用于基准测试的图数据的系统，包括：至少两个第一设备，每个第一设备部署有顶点生成框架；至少两个第二设备，每个第二设备部署有顶点关系生成框架；以及第三设备，部署有顶点分块框架。各个顶点生成框架被配置为：创建多个实体顶点；创建所述顶点分块框架所抽取的各个第一实体顶点的对应实体账户顶点；以及在各个实体账户顶点与对应的实体顶点之间创建拥有关系。所述顶点分块框架被配置为从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点；以及从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集。各个顶点关系生成框架被配置为基于所抽取的起点实体账户顶点集和终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[33]根据本说明书的实施例的另一方面，提供一种用于生成应用于基准测试的图数据的装置，包括：至少一个处理器，与所述至少一个处理器耦合的存储器，以及存储在所述存储器中的计算机程序，所述至少一个处理器执行所述计算机程序来实现如上所述的方法。

[34]根据本说明书的实施例的另一方面，提供一种计算机可读存储介质，其存储有可执行指令，所述指令当被执行时使得处理器执行如上所述的方法。

[35]根据本说明书的实施例的另一方面，提供一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行来实现如上所述的方法。

附图说明

[36]通过参照下面的附图，可以实现对于本说明书内容的本质和优点的进一步理解。在附图中，类似组件或特征可以具有相同的附图标记。

[37]图1示出了根据本说明书的第一实施例的图数据生成方法的示例流程图。

[38]图2示出了根据本说明书的第一实施例的账户关联关系创建过程的一个示例流程图。

[39]图3示出了根据本说明书的第一实施例的账户关联关系创建过程的另一示例流程图。

[40]图4示出了根据本说明书的第一实施例的图数据生成过程的示例示意图。

[41]图5示出了根据本说明书的第一实施例的图数据的数据结构的示例示意图。

[42]图6示出了根据本说明书的第一实施例的用于生成应用于基准测试的图数据的装置的方框图。

[43]图7示出了根据本说明书的第二实施例的用于生成应用于基准测试的图数据的系统的方框图。

[44]图8示出了根据本说明书的第二实施例的图数据生成方法的示例流程图。

[45]图9示出了根据本说明书的第二实施例的账户关联关系创建过程的示例流程图。

[46]图10示出了根据本说明书的第二实施例的图数据生成装置的方框图。

[47]图 11 示出了根据本说明书的第二实施例的顶点生成框架的示例方框图。

[48]图 12 示出了根据本说明书的第二实施例的顶点关系生成框架的示例方框图。

[49]图 13 示出了根据本说明书的实施例的基于计算机系统实现的图数据生成装置的示例示意图。

具体实施方式

[50]现在将参考示例实施方式讨论本文描述的主题。应该理解，讨论这些实施方式只是为了使得本领域技术人员能够更好地理解从而实现本文描述的主题，并非是对权利要求书中所阐述的保护范围、适用性或者示例的限制。可以在不脱离本说明书内容的保护范围的情况下，对所讨论的元素的功能和排列进行改变。各个示例可以根据需要，省略、替代或者添加各种过程或组件。例如，所描述的方法可以按照与所描述的顺序不同的顺序来执行，以及各个步骤可以被添加、省略或者组合。另外，相对一些示例所描述的特征在其它例子中也可以进行组合。

[51]如本文中使用的，术语“包括”及其变型表示开放的术语，含义是“包括但不限于”。术语“基于”表示“至少部分地基于”。术语“一个实施例”和“一实施例”表示“至少一个实施例”。术语“另一个实施例”表示“至少一个其他实施例”。术语“第一”、“第二”等可以指代不同的或相同的对象。下面可以包括其他的定义，无论是明确的还是隐含的。除非上下文中明确地指明，否则一个术语的定义在整个说明书中是一致的。

[52]在基于图数据实现的应用投入使用之前，需要使用图数据来对该应用进行基准测试，并且只有通过基准测试后的应用才被允许投入使用。基准测试是指通过设计科学的测试方法、测试工具和测试系统，实现对一类测试对象的某项性能指标进行定量和可对比的测试。例如，对计算机 CPU 进行浮点运算、数据访问的带宽和延迟等指标的基准测试，可以使用户清楚地了解每款 CPU 的运算性能及作业吞吐能力是否满足应用程序的要求。对数据库管理系统的 ACID（Atomicity, Consistency, Isolation, Durability，原子性、一致性、独立性和持久性）、查询时间和联机事务处理能力等方面的性能指标进行基准测试，也有助于使用者挑选最符合自己需求的数据库系统。

[53]LDBC（Linked Data Benchmark Council）提出的 LDBC SNB DATAGEN 是一种基于社交网络的基准测试 SNB（Social Network Benchmark）。LDBC SNB DATAGEN 所生成的数据规模范围为 100MB-1TB。然而，LDBC SNB DATAGEN 所生成的数据场景过于定制化，不易修改，与一些应用场景（例如，金融应用场景）的需求相差较大。此外，LDBC SNB DATAGEN 采用两个顶点属性的属性距离作为关系创建概率的影响因素，关系生成逻辑较为简单。此外，采用 LDBC SNB DATAGEN 方案，在由于计算机硬件物理瓶颈等因素导致在关系生成时会对顶点进行分块的情况下，会导致分块与分块之间的顶点无法生成关系。

[54]鉴于上述，本说明书的实施例提供了一种用于生成应用于基准测试的图数据的方案。在该方案中，经由顶点生成框架创建多个实体顶点以及各个实体顶点的对应实体账户顶点，并且在各个实体顶点以及对应的实体账户顶点之间创建拥有关系。经由顶点分块框架根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集，起点实体账户顶点集和终点实体账户顶点集之间不具有重合的实体账户顶点。然后，经由顶点关系生成框架基于起点实体账户顶点集和终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[55]在本说明书中，术语“账户”是指用于反映资产数据的增减变动情况及其结果的载体，例如，金融资产账户、数字资产账户或者其它类型的数据资产账户等。术语“账户数据”可以包括金融资产数据（例如，资金数据、借贷数据、负债数据等）、数字资产数据或者其它类型的资产数据等。术语“账户关联关系”是指两个账户之间可能发生的所有类型的关系，例如，账户数据转移关系、账户绑定关系、账户从属关系以及账户之间可以发生的其它类型的关联关系等。

[56]下面参照图 1 到图 12 描述根据本说明书的实施例的图数据生成的系统、方法及装置。

[57]图 1 示出了根据本说明书的第一实施例的图数据生成方法 100 的示例流程图。图 1 中示出的图数据生成方法由图数据生成装置执行，该图数据生成装置的各个组件可以部署在同一设备处或不同设备处。

[58]如图 1 所示，在 110，创建多个实体顶点以及各个实体顶点的对应实体账户顶点。在一个示例中，每个实体顶点可具有实体顶点属性。实体顶点属性可包括顶点出度。相应地，可基于各个实体顶点的顶点出度创建对应的实体账户顶点。此外，实体顶点属性可包括实体标识。实体标识用于唯一标识实体顶点。实体标识例如可是全局唯一标识，例如，基于所对应的分块编号创建的全局唯一整数。在一个示例（例如，金融应用场景示例）中，实体可包括个人实体和组织实体。相应地，实体顶点可包括个人顶点（Person）和组织顶点（Organization）。在一个示例中，各个实体顶点的顶点出度可是预先设置的固定值。在另一示例中，各个实体顶点的顶点出度可基于例如经由数据分布接口输入的顶点出度分布信息确定。例如，可以基于顶点出度分布信息（例如，幂率分布）随机产生整数。在另一示例中，实体顶点属性还可包括顶点入度。相应地，各个实体顶点的顶点出度和顶点入度可预先设置，或者基于例如经由数据分布接口输入的顶点出度/入度分布信息确定。此外，实体顶点属性还可包括实体名称。例如，在实体顶点是个人顶点的情况下，实体名称可包括姓（First Name）和名（Last Name）。在实体顶点是组织顶点的情况下，实体名称可包括组织名称（Organization Name）。

[59]在一个示例中，所创建的实体账户顶点可以包括个人账户顶点（PersonalAccount）和组织账户顶点（OrganizationalAccount）。此外，在一个示例中，每个实体账户顶点的账户顶点属性可以包括顶点标识、账户创建日期（CreateDate）和账户有效性标识（IsBlocked）等。账户有效性标识 IsBlocked 可以采用布尔值（Boolean）表示，用于指示账户是否有效。例如，可以采用布尔值“1”表示有效，以及布尔值“0”表示无效。在另一示例中，也可以反向表示。在一个示例中，CreateDate 的取值 DateTime 可以通过随机生成器在限定时间范围内产生。IsBlocked 的取值可以通过随机生成器产生。

[60]此外，在另一示例中，针对各个实体顶点，还可以创建业务申请顶点。业务申请顶点的具体形式可以基于具体的应用场景决定。例如，在金融应用场景下，业务申请顶点的示例可以包括贷款申请（LoanApplication）顶点、融资申请顶点等。LoanApplication 顶点的实体顶点属性可以具有顶点标识和 LoanAmount。LoanAmount 的取值为 Decimal 值。相应地，基于各个实体顶点的顶点出度为该各个实体顶点创建对应的实体账户顶点以及业务申请顶点。这里，实体账户顶点和业务申请顶点例如可以统称为实体关联顶点。

[61]在 120，在各个实体顶点以及对应的实体账户顶点之间创建拥有关系（Owe）。在另一示例中，在还创建有业务申请顶点的情况下，除了在各个实体账户顶点与对应的实体顶点之间创建拥有关系之外，还可以在各个业务申请顶点与对应的实体顶点之间创建申请关系（Apply）。该申请关系还可以具有关系属性（ApplyDate）。ApplyDate 的取值通过随机生成器在限定时间范围内产生。

[62]在另一示例中，每个实体账户顶点还可具有账户顶点属性。账户顶点属性可包括账户关联属性。在实体账户顶点包括个人账户顶点（PersonalAccount）和组织账户顶点（OrganizationalAccount）的情况下，账户关联属性的示例例如可包括但不限于账户注册地址、注册电话（Phone）、登录网络地址（IP）和登录物理地址（MAC）。账户注册地址例如可是账户注册城市（City）。登录网络地址（IP）例如可是登录账户时所使用的 IP 地址。登录物理地址（MAC）可是登录账户时所使用设备的设备物理地址，例如，MAC 地址等。

[63]个人账户 PersonalAccount 或组织账户 OrganizationalAccount 的注册电话（Phone）、登录网络地址（IP）、登录物理地址（MAC）和注册地址（City）会在创建个人账户或组织账户时创建。City 的取值在城市数据资源库中随机抽取，Phone 的取值在电话数

据资源库中随机抽取, IP 地址的数量通过随机生成器生成, 然后从网络地址数据资源库中随机抽取相应数量的 IP 地址。MAC 地址的数量通过随机生成器生成, 然后从物理地址数据资源库中随机抽取相应数量的 MAC 地址。

[64]在一个示例中, 在实体账户顶点具有账户关联属性的情况下, 还可以基于各个实体账户顶点的账户关联属性创建账户属性顶点; 并且根据账户关联属性来在各个账户属性顶点之间以及各个账户属性顶点与对应的实体账户顶点之间创建账户属性关系。账户属性关系的示例例如可以包括但不限于: 位于关系 (IsLocatedIn)、电话注册关系 (SignUpDate)、登录网络地址关系 (SignInWithIP) 和登录物理地址关系 (SignInWithMAC) 中的至少一个。例如, 在 PersonalAccount 与账户属性顶点 IP 之间创建账户属性关系 SignInWithIP, 该账户属性关系具有关系属性 SignInDate。SignInDate 的取值通过随机生成器在限定时间范围内产生。在 PersonalAccount 与账户属性顶点 MAC 之间创建账户属性关系 SignInWithMAC, 该账户属性关系具有关系属性 SignInDate。SignInDate 的取值通过随机生成器在限定时间范围内产生。在 PersonalAccount 与账户属性顶点 Phone 之间创建账户属性关系 SignUpWithPhone, 该账户属性关系具有关系属性 SignUpDate。SignUpDate 的取值通过随机生成器在限定时间范围内产生。在 PersonalAccount 与账户属性顶点 City 之间创建账户属性关系 IsLocatedIn。在账户属性顶点 Phone 与账户属性顶点 City 之间创建账户属性关系 IsLocatedIn。

[65]在如上创建完实体账户顶点后, 在 130, 根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集, 起点实体账户顶点集和终点实体账户顶点集之间不具有重合的实体账户顶点。在本说明书中, 起点实体账户顶点作为图数据的边关系的起点, 以及终点实体账户顶点作为图数据的边关系的终点。在一个示例中, 可以将所创建的实体账户顶点分类为起点实体账户顶点集和终点实体账户顶点集。在另一示例中, 也可以从所创建的实体账户顶点中抽取出起点实体账户顶点集和终点实体账户顶点集。在本说明书中, 图数据是指有向图数据。

[66]在 140, 基于起点实体账户顶点集和终点实体账户顶点集, 创建实体账户顶点之间的账户关联关系, 由此创建出所需的图数据。在本说明书中, 两个账户的账户关联关系的示例例如可以包括但不限于账户数据转移关系、账户绑定关系以及账户之间可以发生的其它类型的关联关系等。账户数据转移关系的示例可以包括但不限于账户资金转移关系、借贷数据转移关系、负债数据转移关系等。在一个示例中, 所创建的图数据可以是金融图数据, 以及账户关联关系可以是转账关系。

[67]在一个示例中, 针对图 1 示出的图数据生成方法, 还可从多个实体顶点中抽取多个第一实体顶点。然后, 创建各个所抽取出的第一实体顶点的对应实体账户顶点。

[68]图 2 示出了根据本说明书的第一实施例的账户关联关系创建过程 200 的一个示例流程图。在图 2 的示例中, 实体账户顶点的账户属性包括顶点出度和顶点入度。

[69]如图 2 所示, 在 210, 根据起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及终点实体账户顶点集中的各个终点实体账户顶点的顶点入度, 确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率。例如, 对于起点实体账户顶点, 基于该起点实体账户顶点的顶点出度除以该起点实体账户顶点集的总顶点出度, 确定该起点实体账户顶点的选中概率。每个起点实体账户顶点集中的各个起点实体账户顶点的选中概率和为 1。对于终点实体账户顶点, 基于该终点实体账户顶点的顶点入度除以该终点实体账户顶点集的总顶点入度, 确定该终点实体账户顶点的选中概率。每个终点实体账户顶点集中的各个终点实体账户顶点的选中概率和为 1。在一个示例中, 选中概率确定过程中所使用的顶点入度是终点实体账户顶点的顶点属性信息中的顶点入度。在另一示例中, 选中概率确定过程中所使用的顶点入度是从终点实体账户顶点的顶点属性信息中的顶点入度中去除来自实体顶点的顶点入度之后得到的顶点入度。

[70]在确定出各个起点实体账户顶点和各个终点实体账户顶点的选中概率后, 在 220,

基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从起点实体账户顶点集和终点实体账户顶点集中选择至少一个起点实体账户顶点以及对应的终点实体账户顶点。这里，实体账户顶点的选择过程是基于选中概率的随机选择过程。所选择出的起点实体账户顶点可以包括一个或多个起点实体账户顶点，每个起点实体账户顶点包括一个对应的终点实体账户顶点。

[71]在 230，计算所选择的起点实体账户顶点和对应的终点实体账户顶点之间的属性距离。例如，在所选择的起点实体账户顶点和终点实体账户顶点之间存在多种相同类型的属性时，可以计算该多种相同类型的属性之间的属性距离 D 。例如，假设所选择的起点实体账户顶点和终点实体账户顶点都具有注册地址、注册电话、登录网络地址，可以基于注册地址、注册电话、登录网络地址分别计算出相应的属性距离 D_1 到 D_3 。

[72]在 240，基于所计算出的属性距离，确定所选择的各个起点实体账户顶点和对应的终点实体账户顶点之间的关系创建概率。例如，可以利用属性距离 D 与关系创建概率 P 之间的函数关系 $P = f(D)$ 来确定。在属性距离包括多个属性距离的情况下，在一个示例中，可以基于多个属性距离确定出整合后的属性距离，然后基于该整合后的属性距离来确定关系创建概率。或者，可以基于函数关系 $P = f(D_1, \dots, D_i)$ 来确定关系创建概率，其中， i 为属性个数。针对各个属性距离，还可以分配不同的权重，然后基于各个属性距离及其权重来确定关系创建概率。

[73]在如上确定出各个起点实体账户顶点和对应的终点实体账户顶点之间的关系创建概率后，在 250，根据关系创建概率，在所选择的各个起点实体账户顶点和对应的终点实体账户顶点之间创建账户关联关系。在本说明书中，所创建的账户关联关系例如可包括账户数据转移关系、账户绑定关系、账户从属关系以及账户之间可发生的其它类型的关联关系等。账户数据转移关系例如可是账户数据转移行为。例如，假设起点实体账户顶点是“张三”，终点实体账户顶点是“李四”，则实体账户顶点“张三”和“李四”之间的一条账户数据转移关系可是“张三在 02 月 18 日向李四转账 XX 元”。此外，相较于该条账户数据转移关系，“张三在 08 月 20 日向李四转账 XX 元”则是另一条账户数据转移关系。

[74]在一个示例中，可以根据关系创建概率，在所选择的各个起点实体账户顶点和对应的终点实体账户顶点之间创建多条账户关联关系，使得所创建的账户关联关系达到预定数目条账户关联关系。

[75]在另一示例中，上述账户关联关系的创建过程可以是循环过程。具体地，针对每个起点实体账户顶点和对应的终点实体账户顶点，将 240 中创建的关系创建概率作为初始关系创建概率，循环执行下述过程，直到未创建出账户关联关系为止：在每次循环时，基于当前关系创建概率来在起点实体账户顶点和对应的终点实体账户顶点创建账户关联关系。然后，判断当前是否创建出账户关联关系。如果当前创建出账户关联关系，则对当前循环过程所使用的关系创建概率进行衰减处理得到下一循环过程的当前关系创建概率，然后执行下一循环过程。如果当前未创建出账户关联关系，则循环结束。所述衰减处理的示例例如可以包括但不限于：根据线性衰减函数或者非线性衰减函数来对关系创建概率进行衰减处理。线性衰减函数或者非线性衰减函数的函数表达式可以是基于具体应用场景确定出的任何合适的函数表达式。

[76]图 3 示出了根据本说明书的第一实施例的账户关联关系创建过程 300 的另一示例流程图。在图 3 的示例中，实体账户顶点的账户顶点属性包括顶点出度和顶点入度。

[77]在 310，根据起点实体账户顶点集中的各个起点实体账户顶点的顶点出度及终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率。选中概率的确定过程可参考上面参照图 2 描述的过程。

[78]在如上确定出各个起点实体账户顶点和各个终点实体账户顶点的选中概率后，循环执行 320 到 380，直到所创建的账户关联关系达到预定数目。

[79]具体地，在每次循环时，在 320，基于各个起点实体账户顶点和各个终点实体账户

顶点的选中概率,从起点实体账户顶点集和终点实体账户顶点集中选择至少一个起点实体账户顶点以及对应的终点实体账户顶点。这里,实体账户顶点的选择过程是基于选中概率的随机选择过程。所选择出的起点实体账户顶点可以包括一个或多个起点实体账户顶点,每个起点实体账户顶点包括一个对应的终点实体账户顶点。

[80]在 330,计算所选择的起点实体账户顶点和对应的终点实体账户顶点之间的属性距离。属性距离的计算过程可以参考上面参照图 2 的 230 描述的过程。

[81]在 340,基于所计算出的属性距离,确定所选择的各个起点实体账户顶点和对应的终点实体账户顶点之间的初始关系创建概率。初始关系创建概率可以参考上面参照图 2 的 240 描述的过程。

[82]在如上确定出各个起点实体账户顶点和对应的终点实体账户顶点之间的初始关系创建概率后,针对每个起点实体账户顶点和对应的终点实体账户顶点,循环执行 350 到 370,直到未创建出账户关联关系。

[83]具体地,在每次循环时,在 350,根据当前关系创建概率,在所选择的各个起点实体账户顶点和对应的终点实体账户顶点之间创建账户关联关系。在首次循环时,当前关系创建概率是初始关系创建概率。接着,在 360,判断当前是否创建出账户关联关系。如果当前创建出账户关联关系,则在 370,对当前循环过程所使用的关系创建概率进行衰减处理得到下一循环过程的当前关系创建概率,然后返回到 350,执行下一循环过程。如果当前未创建出账户关联关系,则过程进行到 380。

[84]在 380,判断所创建出的账户关联关系的关系数目是否达到预定数目。如果达到预定数目,则流程结束。如果未达到预定数目,则返回到 320,执行下一循环过程。

[85]在另一示例中,在图 2 或图 3 中示出的账户关联关系创建过程中,还可以包括获取实体账户顶点的顶点出度/入度分布信息;并且根据所获取的顶点出度/入度分布信息,确定各个实体账户顶点的顶点出度和顶点入度。

[86]在另一示例中,在图 2 或图 3 中示出的账户关联关系创建过程中,还可包括获取社交网络出度/入度分布信息;并且根据所获取的社交网络出度/入度分布信息来在实体顶点之间创建认识/从属关系。然后,在确定关系创建概率时,基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系,确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

[87]图 4 示出了根据本说明书的实施例的图数据生成过程 400 的示例示意图。图 5 示出了根据本说明书的实施例的图数据的数据结构的示例示意图。

[88]如图 4 所示,在该图数据生成过程中,在顶点生成框架中创建实体顶点、实体账户顶点和账户属性顶点,并且实体顶点、实体账户顶点和账户属性顶点的创建机制不同。实体顶点的创建不需要任何数据输入。实体账户顶点的创建需要输入已经创建出的实体顶点,以及账户属性顶点的创建需要已经创建的实体账户顶点的账户关联属性。此外,在顶点生成框架中,还分别创建各个实体账户顶点与对应的实体顶点之间的拥有关系,以及各个账户属性顶点之间以及各个账户属性顶点与对应的实体账户顶点之间的账户属性关系。在顶点关系生成框架,创建各个实体账户顶点之间的账户关联关系,例如,转账关系(Transfer)。如图 5 中所示,转账关系具有关系属性 TransferAmount。TransferAmount 的取值为 Decimal 值。

[89]图 6 示出了根据本说明书的第一实施例的用于生成应用于基准测试的图数据的装置 600 的方框图。如图 6 所示,装置 600 包括顶点生成单元 610、拥有关系生成单元 620、顶点分块单元 630 和关联关系生成单元 640。

[90]顶点生成单元 610 被配置为创建多个实体顶点以及各个实体顶点的对应实体账户顶点。顶点生成单元 610 的操作可以参考上面参照图 1 的 110 描述的操作。

[91]拥有关系生成单元 620 被配置为在各个实体顶点以及对应的实体账户顶点之间创建拥有关系。拥有关系生成单元 620 的操作可以参考上面参照图 1 的 120 描述的操作。

[92]顶点分块单元 630 被配置为根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集,所述起点实体账户顶点集和所述终点实体账户顶点集之间不具有重合的实体账户顶点。顶点分块单元 630 的操作可参考上面参照图 1 的 130 描述的操作。

[93]关联关系生成单元 640 被配置为基于起点实体账户顶点集和终点实体账户顶点集,创建实体账户顶点之间的账户关联关系。关联关系生成单元 640 的操作可以参考上面参照图 1 的 140 描述的操作以及参照图 2 或图 3 描述的操作。

[94]在另一示例中,拥有关系生成单元 620 和关联关系生成单元 640 可以采用同一关系生成单元实现。

[95]另一示例中,顶点生成单元 610 还可被配置为从多个实体顶点中抽取多个第一实体顶点。然后,顶点生成单元 610 创建各个所抽取出的第一实体顶点的对应实体账户顶点。

[96]在另一示例中,顶点生成单元 610 还可以被配置为针对各个实体顶点创建业务申请顶点。相应地,装置 600 还可以包括申请关系生成单元(未示出)。申请关系生成单元被配置为在各个业务申请顶点与对应的实体顶点之间创建申请关系(Apply)。申请关系生成单元可以与拥有关系生成单元 620 和关联关系生成单元 640 采用同一单元实现,也可以采用不同单元实现。

[97]在另一示例中,装置 600 还可以包括数据分布信息获取单元(未示出)。数据分布信息获取单元可以被配置为获取实体顶点的顶点出度分布信息。相应地,顶点生成单元 610 根据所获取的顶点出度分布信息,创建各个实体顶点的对应实体账户顶点。数据分布信息获取单元还可以被配置为获取实体账户顶点的顶点出度/入度分布信息。相应地,顶点生成单元 610 根据所获取的顶点出度/入度分布信息,确定各个实体账户顶点的顶点出度和顶点入度。数据分布信息获取单元还可以被配置为获取社交网络出度/入度分布信息。相应地,装置 600 还可以包括实体顶点关系生成单元(未示出)。实体顶点关系生成单元根据所获取的社交网络出度/入度分布信息来在实体顶点之间创建认识/从属关系。然后,关联关系生成单元 640 基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系,确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。同样,实体顶点关系生成单元可以与申请关系生成单元、拥有关系生成单元 620 和关联关系生成单元 640 采用同一单元实现,也可以采用不同单元实现。

[98]利用本说明书的第一实施例示出的图数据生成方案,可以生成具有真实图数据结构的测试图数据,由此应用于基准测试。该图数据生成方案尤其适用于生成金融图数据。

[99]图 7 示出了根据本说明书的第二实施例的用于生成应用于基准测试的图数据的系统 700 的方框图。

[100]如图 7 所示,系统 700 包括 M 个第一设备 710-1 到 710-M、N 个第二设备 720-1 到 720-N 以及第三设备 730。这里,M 和 N 的取值可以相同,也可以不同。M 和 N 的具体取值可以根据具体的应用场景决定,例如,可以基于应用场景所需要生成的图数据规模决定。第一设备、第二设备和第三设备可以是任意类型的具有计算能力或处理能力的服务器设备或终端设备。例如,服务器设备的示例可以包括但不限于:单台服务器、服务器集群、云端服务器或云端服务器集群等。终端设备的示例可以包括但不限于:智能手机、个人电脑(personal computer, PC)、笔记本电脑、平板电脑、电子阅读器、网络电视、可穿戴设备等智能终端设备中的任一种。

[101]第一设备、第二设备和第三设备之间可以直接通信或者经由网络通信来进行数据传输。在一些实施例中,网络可以是有线网络或无线网络中的任意一种或多种。网络的示例例如可以包括但不限于电缆网络、光纤网络、电信网络、企业内部网络、互联网、局域网络(LAN)、广域网络(WAN)、无线局域网络(WLAN)、城域网(MAN)、公共交换电话网络(PSTN)、蓝牙网络、紫蜂网络(ZigZee)、近场通讯(NFC)、设备内总线、设备内线路等或其任意组合。

[102] 第一设备 710-1 到 710-M 中的每个第一设备可以部署有数据分布接口 711 和顶点生成框架 712。第二设备 720-1 到 720-N 中的每个第二设备可以部署有顶点关系生成框架 721。第三设备 730 可以部署有顶点分块框架 731。在本说明书中，术语“框架”可以等同于“单元”、“模块”、“平台”等。

[103] 数据分布接口 711 可以被配置为获取（例如，供用户输入）顶点出度分布信息或者顶点出度/入度分布信息。这里，顶点出度是指以该顶点为起点的边的数量。顶点入度是指以该顶点为终点的边的数量。顶点出度分布信息可以被顶点生成框架 712 使用来确定所创建的各个实体顶点的顶点出度。此外，数据分布接口 711 还可以被配置为获取实体账户顶点的顶点出度/入度分布信息。相应地，顶点生成框架 712 根据实体账户顶点的顶点出度/入度分布信息，确定各个实体账户顶点的顶点出度和顶点入度。此外，数据分布接口 711 还可以被配置为获取社交网络出度/入度分布信息。所获取的社交网络出度/入度分布信息被顶点生成框架 712 使用来在所创建的实体顶点之间创建认识/从属关系。

[104] 第一设备 710-1 到 710-M 中的每个第一设备可以对应于顶点分块框架 731 所分块出的多个顶点分块中的每个顶点分块，每个第一设备中的顶点生成框架 712 被配置为对从顶点分块框架 731 接收的顶点分块进行处理。

[105] 具体地，各个第一设备上的顶点生成框架 712 被配置为创建多个实体顶点。各个顶点生成框架 712 所创建的实体顶点可以发送给顶点分块框架 731，也可以存储到相同的数据存储空间（数据存储器或数据存储单元）中，以供顶点分块框架 731 从该数据存储空间获取。

[106] 顶点分块框架 731 被配置为从所创建的实体顶点中为各个顶点生成框架 712 抽取实体顶点分块，每个顶点生成框架 712 对应一个实体顶点分块，每个实体顶点分块包括多个第一实体顶点。这里，顶点分块框架 731 所执行的实体顶点抽取是不放回的随机抽取，并且每次抽取处理时，需要将所创建的所有实体顶点都抽取完毕为止。例如，假设各个顶点生成框架所创建的实体顶点为 100 个实体顶点，并且顶点生成框架的数目为 10，则顶点分块框架 731 需要执行 10 次随机抽取处理，将该 100 个实体顶点抽取为 10 个实体顶点分块，每个实体顶点分块所包括的实体顶点的数目可以相同或者不同。而且，在随机抽取处理时，前一抽取处理所抽取出的实体顶点不再放回当前抽取处理的实体顶点池。所抽取出的 10 个实体顶点分块例如可以被分发到各个顶点生成框架 712。

[107] 在各个顶点生成框架 712 得到顶点分块框架 731 抽取出的多个第一实体顶点（实体顶点分块）后，各个顶点生成框架 712 还被配置为基于所得到的各个第一实体顶点的顶点出度为该各个第一实体顶点创建对应的实体账户顶点。此外，在另一示例中，各个顶点生成框架 712 还可以生成业务申请顶点。业务申请顶点的具体形式可以基于具体的应用场景决定。例如，在金融应用场景下，业务申请顶点的示例可以包括贷款申请（LoanApplication）顶点、融资申请顶点等。相应地，各个顶点生成框架 712 被配置为基于所得到的各个第一实体顶点的顶点出度为该各个第一实体顶点创建对应的实体账户顶点以及业务申请顶点。这里，实体账户顶点和业务申请顶点例如可以统称为实体关联顶点。同样，所创建的实体账户顶点可以发送给顶点分块框架 731，也可以存储到相同的数据存储空间中，以供顶点分块框架 731 从该数据存储空间获取。

[108] 在如上创建出实体账户顶点后，各个顶点生成框架 712 被配置为在各个实体账户顶点与对应的第一实体顶点之间创建拥有关系（Owe）。在另一示例中，在各个顶点生成框架 712 还创建业务申请顶点的情况下，除了在各个实体账户顶点与对应的第一实体顶点之间创建拥有关系之外，各个顶点生成框架 712 还在各个业务申请顶点与对应的第一实体顶点之间创建申请关系（Apply）。

[109] 此外，在各个实体账户顶点具有账户关联属性的情况下，各个顶点生成框架 712 还被配置为基于各个实体账户顶点的账户关联属性创建账户属性顶点，并且基于账户关联属性，在各个账户属性顶点之间以及各个账户属性顶点与对应的实体账户顶点之间创

建账户属性关系。

[110] 在各个顶点生成框架 712 创建完实体账户顶点后，顶点分块框架 731 还可以被配置为从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集。同样，顶点分块框架 731 的起点实体账户顶点集和终点实体账户顶点集的抽取过程是不放回抽取。此外，顶点分块框架的上述抽取过程可以是直到所有实体账户顶点被抽取完为止。

[111] 在得到起点实体账户集和终点实体账户集后，各个顶点关系生成框架 721 被配置为基于所接收的起点实体账户顶点集和终点实体账户顶点集来创建实体账户顶点之间的账户关联关系，由此创建出所需的图数据。在一个示例中，所述图数据可以是金融图数据，以及账户关联关系可以是转账关系。顶点关系生成框架 721 的账户关联关系创建过程将会在下面参照附图具体描述。

[112] 在本说明书的其它实施例中，各个第一设备也可以不包括数据分布接口 711。

[113] 此外，在图 7 的示例中，第一设备、第二设备和第三设备被示出为不同的设备。在本说明书的其它实施例中，第一设备 710-1 到 710-M 中的部分第一设备或每个第一设备可以分别与第二设备 720-1 到 720-N 中的一个第二设备相同。换言之，一个设备上可以同时部署顶点生成框架和顶点关系生成框架。在另一示例中，第三设备 730 可以与第一设备 710-1 到 710-M 和/或第二设备 720-1 到 720-N 中的一个设备相同。换言之，一个设备上可以同时部署顶点生成框架和顶点分块框架，同时部署顶点关系生成框架和顶点分块框架，或者同时部署顶点生成框架、顶点关系生成框架和顶点分块框架。

[114] 图 8 示出了根据本说明书的实施例的图数据生成方法 800 的示例流程图。

[115] 如图 8 所示，在 810，在各个第一设备的顶点生成框架处，分别创建多个实体顶点。在一个示例中，每个实体顶点的实体顶点属性可以包括顶点出度。这里，各个实体顶点的顶点出度可以是基于经由该顶点生成框架所位于的第一设备处的数据分布接口获取的顶点出度分布信息确定出。在一个示例中，所创建的实体顶点可以发送给顶点分块框架，也可以存储到共同的数据存储空间中，以供顶点分块框架获取。在实体顶点的实体顶点属性包括顶点出度和顶点入度的情况下，可以经由该顶点生成框架所位于的第一设备处的数据分布接口获取顶点出度/入度分布信息。

[116] 在各个顶点生成框架创建出实体顶点后，循环执行 820 到 860 的操作过程，直到循环执行预定次数，例如，K 次。

[117] 具体地，在每次循环过程中，在 820，第三设备处的顶点分块框架从所创建的实体顶点中为各个顶点生成框架抽取实体顶点分块，每个顶点生成框架对应一个实体顶点分块，每个实体顶点分块包括多个第一实体顶点。在顶点分块框架与顶点生成框架位于不同设备主体的情况下，顶点分块框架所抽取的多个第一实体顶点分块可以分发到对应的顶点生成框架。要说明的是，在每次循环过程中，用于实体顶点抽取的实体顶点包括步骤 810 中创建的所有实体顶点。此外，顶点分块框架的实体顶点抽取过程采用如上参照图 7 所述的实体顶点抽取过程。

[118] 在 830，在各个顶点生成框架处，分别基于所抽取的各个第一实体顶点的顶点出度为各个第一实体顶点创建对应的实体账户顶点，并且在各个实体账户顶点与对应的实体顶点之间创建拥有关系。在一个示例中，所创建的实体账户顶点可以发送给顶点分块框架，也可以存储到共同的数据存储空间中，以供顶点分块框架获取。

[119] 在 840，顶点分块框架从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集。

[120] 在 850，在各个顶点关系生成框架处，分别基于所抽取的起点实体账户顶点集和终点实体账户顶点集来创建实体账户顶点之间的账户关联关系。账户关联关系的创建过程将在下面参照图 9 详细描述。

[121] 在 860，判断是否达到预定循环次数（例如，K 次）。如果达到预定循环次数，则

流程结束。如果未达到预定循环次数，则返回到 820，执行下一循环过程。

[122] 针对图 8 描述的图数据生成方法，也可以采用与针对图 1 描述的图数据生成方法的修改相应的修改方式进行修改。

[123] 图 9 示出了根据本说明书的实施例的账户关联关系创建过程 850 的示例流程图。该账户关联关系创建过程是单个顶点关系生成框架所执行的过程。

[124] 如图 8 所示，在 851，根据起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率。

[125] 在如上确定出各个起点实体账户顶点和各个终点实体账户顶点的选中概率后，循环执行 852 到 858，直到所创建的账户关联关系达到第一预定数目 M 。在一个示例中，在从顶点分块框架的实体顶点抽取过程到各个顶点关系生成框架的账户关联关系创建过程被循环执行 K 次时，第一预定数目 $M = P/K$ ，其中， P 为所创建的多个实体账户顶点（所有实体账户顶点）的总出度数量。在另一示例中， P 也可以是预定设置的用于指示需要创建的账户关联关系总数的预定值。

[126] 具体地，在每次循环过程中，在 852，基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从起点实体账户顶点集和终点实体账户顶点集中选择出至少一个起点实体账户顶点以及对应的终点实体账户顶点。在一个示例中，每次选择出一个起点实体账户顶点和一个终点实体账户顶点。在另一示例中，每次也可以选择出多个起点实体账户顶点以及对应的终点实体账户顶点。这里，实体账户顶点的选择过程是基于选中概率的随机选择过程。

[127] 在 853，计算所选择的起点实体账户顶点和终点实体账户顶点之间的属性距离。属性距离的计算过程可以参考上面参照图 2 的 230 描述的过程。

[128] 在 854，基于所计算出的属性距离 D ，确定所选择的起点实体账户顶点和终点实体账户顶点之间的初始关系创建概率。初始关系创建概率的确定过程可以参考上面参照图 2 的 240 描述的过程。

[129] 接着，循环执行 855 到 857，直到未创建出新的账户关联关系为止。在 855，基于当前关系创建概率来在所选择的起点实体账户顶点和终点实体账户顶点之间创建账户关联关系。在 856，判断当前是否创建出账户关联关系。如果当前创建出账户关联关系，则在 857，对当前循环过程所使用的关系创建概率进行衰减处理得到下一循环过程的前关系创建概率，然后返回到 855，执行下一循环过程。

[130] 如果当前未创建出账户关联关系，则进行到 858。在 858，判断所创建的账户关联关系的关系数量是否达到第一预定数目 M 。如果达到第一预定数目 M ，则流程进行到图 8 的 860。如果未达到第一预定数目 M ，则返回到 852，执行下一循环过程。

[131] 如上参照图 7 到图 9，描述了根据本说明书的第二实施例的图数据生成方法。要说明的是，上述参照附图描述的实施例仅仅是例示性的，在其它实施例中，还可以对上述实施例进行各种适应性修改。

[132] 例如，在其它实施例中，还可以经由各个顶点生成框架的对应数据分布接口获取社交网络出度/入度分布信息。然后，在各个顶点生成框架处，根据所获取的社交网络出度/入度分布信息来在实体顶点之间创建认识/从属关系。例如，在个人顶点和/组织顶点之间创建认识/从属关系。相应地，在确定初始关系创建概率时，除了考虑所选择的起点实体账户顶点和终点实体账户顶点之间的属性距离之外，还需要考虑所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系。换言之，基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定所选择的起点实体账户顶点和终点实体账户顶点之间的初始关系创建概率。

[133] 此外，在图 9 的示例中，基于关系创建概率的账户关联关系的创建过程被示出为

循环过程。在其它实施例中,也可一次性创建出多条账户关联关系,而不执行循环过程。

[134] 下面将结合一个示例来说明根据本说明书的第二实施例的图数据生成过程。

[135] 该示例中,假设存在 10 个顶点生成框架,10 个顶点关系生成框架及 1 个顶点分块框架。在 10 个顶点生成框架总共生成 100 个实体顶点后,执行 5 次循环过程来生成图数据。在每次循环时,顶点分块框架将所有 100 个实体顶点随机分块为 10 个实体顶点分块,每个实体顶点分块包括 10 个实体顶点。然后,顶点分块框架向每个顶点生成框架分发一个实体顶点分块。在接收到实体顶点分块后,各个顶点生成框架根据各个实体顶点的顶点出度创建对应的实体账户顶点,并且在所创建的实体账户顶点和对应的实体顶点之间创建拥有关系。

[136] 随后,顶点分块框架将所创建的所有实体账户顶点随机分块为 10 个实体账户分块,每个实体账户分块包括一个起点实体账户顶点集和一个终点实体账户顶点集。所分块出的各个实体账户顶点集之间不具有公共实体账户顶点。然后,顶点分块框架向每个顶点关系生成框架分发一个实体账户顶点分块。在接收到实体账户顶点分块后,各个顶点关系生成框架根据起点实体账户顶点集和终点实体账户顶点集来创建实体账户顶点之间的账户关联关系。如此循环 5 次,直到创建出预定数目条账户关联关系。

[137] 利用根据本说明书的第二实施例的图数据生成方案,顶点生成过程和顶点关系生成过程被分发到多个顶点生成框架和多个顶点关系生成框架中执行,从而使得可以容易地生成任一数据规模的图数据。此外,在上述图数据生成方案中,通过将与应用场景相关的顶点生成过程、顶点关系生成过程、属性关系生成过程和与应用场景无关的顶点分块过程部署在不同的处理框架上执行,从而将与应用场景相关的顶点生成过程、顶点关系生成过程、属性关系生成过程和与应用场景无关的数据分块过程解耦,从而使得应用场景修改和扩展成为可能。而且,在进行账户关联关系创建时,从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集,这种抽取过程是随机抽取,从而确保分块与分块之间的顶点可以生成关系。

[138] 此外,利用上述图数据生成方案,在创建账户关联关系时,通过确定出初始关系创建概率,基于该初始关系创建概率来创建出账户关联关系,并且在创建出账户关联关系后,衰减该初始关系创建概率来进一步创建账户关联关系,如此循环多次,由此使得所创建出的账户关联关系更加符合实际的应用场景。

[139] 图 10 示出了根据本说明书的实施例的图数据生成装置 1000 的方框图。如图 10 所示,图数据生成装置 1000 包括多个(例如,M 个)数据分布接口 1010、多个(例如,M 个)顶点生成框架 1020、多个(例如,N 个)顶点关系生成框架 1030 和顶点分块框架 1040。这里,M 和 N 的取值可以相同,也可以不同。每个数据分布接口 1010 和一个顶点生成框架 1020 部署在一个第一设备上,以及每个顶点关系生成框架 1030 部署在一个第二设备上。顶点分块框架 1040 部署在第三设备上。

[140] 数据分布接口 1010 被配置为获取实体顶点的顶点出度分布信息。各个顶点生成框架 1020 被配置为创建多个实体顶点,每个实体顶点的实体顶点属性包括顶点出度,其中,各个实体顶点的顶点出度可以基于所获取的顶点出度分布信息确定。

[141] 顶点分块框架 1040 被配置为从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点。然后,各个顶点生成框架 1020 还被配置为基于顶点分块框架所抽取的各个第一实体顶点的顶点出度为该各个第一实体顶点创建对应实体账户顶点,并且在各个实体账户顶点与对应的第一实体顶点之间创建拥有关系。

[142] 顶点分块框架 1040 还被配置为从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集。

[143] 各个顶点关系生成框架 1030 被配置为基于所抽取的起点实体账户顶点集和终点实体账户顶点集来创建实体账户顶点之间的账户关联关系。

[144] 数据分布接口 1010 还可以被配置为获取实体账户顶点的顶点出度/入度分布信息。

相应地，各个顶点生成框架 1020 可以基于所获取的顶点出度/入度分布信息，确定各个实体账户顶点的顶点出度和顶点入度。

[145] 图 11 示出了根据本说明书的实施例的顶点生成框架 1100 的示例方框图。如图 11 所示，顶点生成框架 1100 包括实体顶点创建单元 1110、实体顶点接收单元 1120、关联顶点创建单元 1130、账户属性顶点创建单元 1140 和关系创建单元 1150。

[146] 实体顶点创建单元 1110 被配置为创建多个实体顶点。在一个示例中，可以经由数据分布接口获取实体顶点的顶点出度分布信息，并且实体顶点创建单元 1110 可以基于所获取的顶点分布信息确定各个实体顶点的顶点出度。

[147] 在顶点分块框架对所创建的多个实体顶点进行实体顶点抽取后，实体顶点接收单元 1120 被配置为从顶点分块框架接收对应的多个第一实体顶点。在顶点分块框架与顶点生成框架位于同一设备主体时，可以无需实体顶点接收单元 1120。

[148] 关联顶点创建单元 1130 被配置为基于从顶点分块框架接收的各个第一实体顶点的顶点出度为该各个第一实体顶点创建对应的实体账户顶点。关系创建单元 1150 被配置为在所创建的实体账户顶点与对应的实体顶点之间创建拥有关系。

[149] 此外，在存在业务申请顶点的情况下，关联顶点创建单元 1130 被配置为基于从顶点分块框架接收的各个第一实体顶点的顶点出度为该各个第一实体顶点创建对应的实体账户顶点和业务申请顶点。相应地，关系创建单元 1150 被配置为在所创建的实体账户顶点与对应的实体顶点之间创建拥有关系以及在各个业务申请顶点与对应的实体顶点之间创建申请关系。

[150] 在实体账户顶点具有账户关联属性的情况下，账户属性顶点创建单元 1140 被配置为基于各个实体账户顶点的账户关联属性创建账户属性顶点。相应地，关系创建单元 1150 被配置为根据账户关系属性在各个账户属性顶点以及各个账户属性顶点与对应的实体账户顶点之间创建账户属性关系。在实体账户顶点不具有账户关联属性的情况下，可以无需账户属性顶点创建单元 1140。

[151] 要说明的是，在其它实施例中，实体顶点创建单元 1110、关联顶点创建单元 1130 和账户属性顶点创建单元 1140 中的部分单元或全部单元可以采用同一单元实现。

[152] 图 12 示出了根据本说明书的实施例的顶点关系生成框架 1200 的示例方框图。如图 12 所示，顶点关系生成框架 1200 包括选中概率确定单元 1210、实体账户顶点选择单元 1220、属性距离计算单元 1230、关系创建概率确定单元 1240 和关系创建单元 1250。

[153] 选中概率确定单元 1210 被配置为根据起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率。

[154] 实体账户顶点选择单元 1220、属性距离计算单元 1230、关系创建概率确定单元 1240 和关系创建单元 1250 循环执行操作，直到所创建的账户关联关系达到第一预定数目 M。

[155] 具体地，在每次循环过程中，实体账户顶点选择单元 1220 被配置为基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从起点实体账户顶点集和终点实体账户顶点集中选择出至少一个起点实体账户顶点以及对应的终点实体账户顶点。

[156] 属性距离计算单元 1230 被配置为计算所选择的起点实体账户顶点和终点实体账户顶点之间的属性距离。

[157] 关系创建概率确定单元 1240 被配置为基于所计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账户顶点之间的初始关系创建概率。

[158] 关系创建单元 1250 被配置为循环执行下述过程，直到未创建出新的账户关联关系为止：基于当前关系创建概率来在所选择的起点实体账户顶点和终点实体账户顶点之间创建账户关联关系，其中，每次循环过程所使用的关系创建概率通过对上一循环过程的关系创建概率进行衰减处理得到。

[159] 此外，数据分布接口可以被配置为获取社交网络出度/入度分布信息。在这种情况下

下，关系创建单元 1250 可以被配置为根据所获取的社交网络出度/入度分布信息来在实体顶点之间创建认识/从属关系。此外，关系创建概率确定单元 1240 被配置为基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定所选择的起点实体账户顶点和终点实体账户顶点之间的初始关系创建概率。

[160] 在本说明书中，在一个示例中，顶点生成框架与顶点关系生成框架之间可以具有一一对应关系。在顶点生成框架与顶点关系生成框架之间具有一一对应关系的情况下，该顶点生成框架可以与对应的顶点关系生成框架部署在同一设备处。在这种情况下，关系创建单元 1150 也可以作为顶点关系生成框架的组件包含在顶点关系生成框架中，而不作为顶点生成框架的组件。

[161] 如上参照图 1 到图 12，对根据本说明书实施例的图数据生成的方法、装置和系统进行了描述。上面的图数据生成装置可以采用硬件实现，也可以采用软件或者硬件和软件的组合来实现。

[162] 图 13 示出了根据本说明书的实施例的基于计算机系统实现的图数据生成装置 1300 的示意图。如图 13 所示，图数据生成装置 1300 可包括至少一个处理器 1310、存储器（例如非易失性存储器）1320、内存 1330 和通信接口 1340，且至少一个处理器 1310、存储器 1320、内存 1330 和通信接口 1340 经由总线 1360 连接在一起。至少一个处理器 1310 执行存储器中存储或编码的至少一个计算机可读指令（即上述以软件形式实现的元素）。

[163] 在一个实施例中，在存储器中存储计算机可执行指令，其当执行时使得至少一个处理器 1310：创建多个实体顶点以及各个实体顶点的对应实体账户顶点；在各个实体顶点以及对应的实体账户顶点之间创建拥有关系；根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集，起点实体账户顶点集和终点实体账户顶点集之间不具有重合的实体账户顶点；以及基于起点实体账户顶点集和终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[164] 在另一实施例中，在存储器中存储计算机可执行指令，其当执行时使得至少一个处理器 1310：经由各个顶点生成框架，分别创建多个实体顶点；经由顶点分块框架，从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点；经由各个顶点生成框架，分别创建所抽取的各个第一实体顶点的对应实体账户顶点，并且在各个实体账户顶点与对应的实体顶点之间创建拥有关系；经由顶点分块框架来从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集；以及经由各个顶点关系生成框架，分别基于所抽取的起点实体账户顶点集和终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

[165] 应该理解，在存储器中存储的计算机可执行指令当执行时使得至少一个处理器 1310 进行本说明书的各个实施例中以上结合图 1-图 12 描述的各种操作和功能。

[166] 根据一个实施例，提供了一种比如机器可读介质（例如，非暂时性机器可读介质）的程序产品。机器可读介质可以具有指令（即，上述以软件形式实现的元素），该指令当被机器执行时，使得机器执行本说明书的各个实施例中以上结合图 1-图 12 描述的各种操作和功能。具体地，可以提供配有可读存储介质的系统或者装置，在该可读存储介质上存储着实现上述实施例中任一实施例的功能的软件程序代码，且使该系统或者装置的计算机或处理器读出并执行存储在可读存储介质中的指令。

[167] 在这种情况下，从可读介质读取的程序代码本身可实现上述任何一项实施例的功能，因此机器可读代码和存储机器可读代码的可读存储介质构成了本发明的一部分。

[168] 可读存储介质的实施例包括软盘、硬盘、磁光盘、光盘（如 CD-ROM、CD-R、CD-RW、DVD-ROM、DVD-RAM、DVD-RW、DVD-RW）、磁带、非易失性存储卡和 ROM。可选择地，可以由通信网络从服务器计算机上或云上下载程序代码。

[169] 根据一个实施例，提供一种计算机程序产品，该计算机程序产品包括计算机程序，

该计算机程序当被处理器执行时,使得处理器执行本说明书的各个实施例中以上结合图1-图12描述的各种操作和功能。

[170] 本领域技术人员应当理解,上面公开的各个实施例可以在不偏离发明实质的情况下做出各种变形和修改。因此,本发明的保护范围应当由所附的权利要求书来限定。

[171] 需要说明的是,上述各流程和各系统结构图中不是所有的步骤和单元都是必须的,可以根据实际的需要忽略某些步骤或单元。各步骤的执行顺序不是固定的,可以根据需要进行确定。上述各实施例中描述的装置结构可以是物理结构,也可以是逻辑结构,即,有些单元可能由同一物理实体实现,或者,有些单元可能分由多个物理实体实现,或者,可以由多个独立设备中的某些部件共同实现。

[172] 以上各实施例中,硬件单元或模块可通过机械或电气方式实现。例如,一个硬件单元、模块或处理器可包括永久性专用的电路或逻辑(如专门的处理器,FPGA或ASIC)来完成相应操作。硬件单元或处理器还可包括可编程逻辑或电路(如通用处理器或其它可编程处理器),可由软件进行临时的设置以完成相应操作。具体的实现方式(机械方式、或专用的永久性电路、或者临时设置的电路)可基于成本和时间上的考虑来确定。

[173] 上面结合附图阐述的具体实施方式描述了示例性实施例,但并不表示可以实现的或者落入权利要求书的保护范围的所有实施例。在整个本说明书中使用的术语“示例性”意味着“用作示例、实例或例示”,并不意味着比其它实施例“优选”或“具有优势”。出于提供对所描述技术的理解的目的,具体实施方式包括具体细节。然而,可以在没有这些具体细节的情况下实施这些技术。在一些实例中,为了避免对所描述的实施例的概念造成难以理解,公知的结构和装置以框图形式示出。

[174] 本公开内容的上述描述被提供来使得本领域任何普通技术人员能够实现或者使用本公开内容。对于本领域普通技术人员来说,对本公开内容进行的各种修改是显而易见的,并且,也可以在不脱离本公开内容的保护范围的情况下,将本文所定义的一般性原理应用于其它变型。因此,本公开内容并不限于本文所描述的示例和设计,而是与符合本文公开的原理和新颖性特征的最广范围相一致。

权利要求书

- 1、一种用于生成应用于基准测试的图数据的方法，包括：
创建多个实体顶点以及各个实体顶点的对应实体账户顶点；
在各个实体顶点以及对应的实体账户顶点之间创建拥有关系；
根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集，所述起点实体账户顶点集和所述终点实体账户顶点集之间不具有重合的实体账户顶点；以及
基于所述起点实体账户顶点集和所述终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。
- 2、如权利要求 1 所述的方法，其中，每个实体账户顶点的账户顶点属性包括账户关联属性，所述方法还包括：
基于各个实体账户顶点的账户关联属性创建账户属性顶点；以及
根据账户关联属性来在各个账户属性顶点之间和/或各个账户属性顶点与对应的实体账户顶点之间创建账户属性关系。
- 3、如权利要求 2 所述的方法，其中，所述实体顶点包括个人顶点和组织顶点，所述实体账户顶点包括个人账户顶点和组织账户顶点，以及所述账户属性顶点包括账户注册地址、注册电话、登录网络地址和登录物理地址中的至少一个，
其中，所述账户属性关系包括位于关系、电话注册关系、登录网络地址关系和登录物理地址关系中的至少一个。
- 4、如权利要求 1 所述的方法，还包括：
获取实体顶点的顶点出度分布信息，
创建各个实体顶点的对应实体账户顶点包括：
根据所述顶点出度分布信息，创建各个实体顶点的对应实体账户顶点。
- 5、如权利要求 1 所述的方法，其中，每个实体账户顶点的账户顶点属性包括顶点出度和顶点入度，基于所述起点实体账户顶点集和所述终点实体账户顶点集创建实体账户顶点之间的账户关联关系包括：
根据所述起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及所述终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率；
基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从所述起点实体账户顶点集和所述终点实体账户顶点集中选择至少一个起点实体账户顶点以及对应的终点实体账户顶点；
计算所选择的起点实体账户顶点和对应的终点实体账户顶点之间的属性距离；
基于所计算出的属性距离，确定所选择的起点实体账户顶点和对应的终点实体账户顶点之间的关系创建概率；以及
根据所述关系创建概率，在所选择的起点实体账户顶点和对应的终点实体账户顶点之间创建账户关联关系。
- 6、如权利要求 5 所述的方法，其中，所述账户关联关系的创建过程被循环执行，直到未创建出新的账户关联关系为止，其中，每次循环过程所使用的关系创建概率通过对上一循环过程的关系创建概率进行衰减处理得到。
- 7、如权利要求 5 所述的方法，其中，从所述起点实体账户顶点和对应的终点实体账户顶点的选择过程到所述账户关联关系的创建过程被循环执行，直到所创建的账户关联关系的数目达到预定数目。
- 8、如权利要求 5 所述的方法，还包括：
获取实体账户顶点的顶点出度/入度分布信息；以及
根据所述顶点出度/入度分布信息，确定各个实体账户顶点的顶点出度和顶点入度。
- 9、如权利要求 5 所述的方法，还包括：

获取社交网络出度/入度分布信息；以及

根据所述社交网络出度/入度分布信息来在所述实体顶点之间创建认识/从属关系，

基于所计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率包括：

基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

10、如权利要求 4 所述的方法，其中，根据所述顶点出度分布信息，创建所述多个实体顶点的对应实体账户顶点包括：

根据所述顶点出度分布信息，创建各个实体顶点的对应实体账户顶点以及业务申请顶点；以及

在各个业务申请顶点与对应的实体顶点之间创建申请关系。

11、如权利要求 1 所述的方法，还包括：

从所述多个实体顶点中抽取多个第一实体顶点；

创建各个实体顶点的对应实体账户顶点包括：

创建各个第一实体顶点的对应实体账户顶点。

12、一种用于生成应用于基准测试的图数据的方法，包括：

经由各个顶点生成框架，分别创建多个实体顶点；

经由顶点分块框架，从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点；

经由各个顶点生成框架，分别创建所抽取的各个第一实体顶点的对应实体账户顶点，并且在各个实体账户顶点与对应的实体顶点之间创建拥有关系；

经由所述顶点分块框架，从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集；以及

经由各个顶点关系生成框架，分别基于所抽取的起点实体账户顶点集和终点实体账户顶点集创建实体账户顶点之间的账户关联关系。

13、如权利要求 12 所述的方法，其中，每个实体账户顶点的账户顶点属性包括账户关联属性，所述方法还包括：

经由各个顶点生成框架，基于各自的实体账户顶点的账户关联属性创建账户属性顶点；并且根据账户关联属性来在各个账户属性顶点之间以及各个账户属性顶点与对应的实体账户顶点之间创建账户属性关系。

14、如权利要求 12 所述的方法，其中，从所述顶点分块框架的实体顶点抽取过程到所述各个顶点关系生成框架的账户关联关系创建过程被循环执行。

15、如权利要求 12 所述的方法，其中，所述顶点分块框架的顶点抽取过程是不放回抽取过程，并且直到所有顶点被抽取完毕为止。

16、如权利要求 12 或 14 所述的方法，其中，每个实体账户顶点的账户顶点属性包括顶点出度和顶点入度，

经由各个顶点关系生成框架，基于所述起点实体账户顶点集和所述终点实体账户顶点集创建实体账户顶点之间的账户关联关系包括：

根据所述起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及所述终点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点和各个终点实体账户顶点的选中概率；

循环执行下述过程，直到所创建的账户关联关系达到第一预定数目 M：

基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从所述起点实体账户顶点集和所述终点实体账户顶点集中选择至少一个起点实体账户顶点以及对应的终点实体账户顶点；

计算所选择的起点实体账户顶点和终点实体账户顶点之间的属性距离；

基于所计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率；以及

基于所述关系创建概率来在所选择的起点实体账户顶点和终点实体账户顶点之间创建账户关联关系。

17、如权利要求 16 所述的方法，其中，所述第一预定数目 $M = P/K$ ，其中， P 为所述多个实体账户顶点的总出度数量，以及 K 为循环执行次数。

18、如权利要求 16 所述的方法，其中，所述账户关联关系的创建过程被循环执行，直到未创建出新的账户关联关系为止，其中，每次循环过程所使用的关系创建概率通过上一循环过程的关系创建概率进行衰减处理得到。

19、如权利要求 16 所述的方法，还包括：

经由各个顶点生成框架的对应数据分布接口获取实体账户顶点的顶点出度/入度分布信息；以及

经由各个顶点生成框架，根据所获取的顶点出度/入度分布信息确定各个实体账户顶点的顶点出度和顶点入度。

20、如权利要求 16 所述的方法，还包括：

经由各个顶点生成框架的对应数据分布接口获取社交网络出度/入度分布信息；以及

经由各个顶点生成框架，根据所获取的社交网络出度/入度分布信息在所述实体顶点之间创建认识/从属关系，

基于所计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率包括：

基于所计算出的属性距离以及所选择的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

21、如权利要求 12 所述的方法，还包括：

经由各个顶点生成框架的对应数据分布接口获取实体顶点的顶点出度分布信息，及

经由各个顶点生成框架，根据所获取的顶点出度分布信息确定各个实体顶点的顶点出度，

经由各个顶点生成框架，分别创建所抽取的各个第一实体顶点的对应实体账户顶点包括：

经由各个顶点生成框架，分别基于所抽取的各个第一实体顶点的顶点出度，创建所述各个第一实体顶点的对应实体账户顶点。

22、一种用于生成应用于基准测试的图数据的装置，包括：

顶点生成单元，创建多个实体顶点以及各个实体顶点的对应实体账户顶点；

拥有关系生成单元，在各个实体顶点以及对应的实体账户顶点之间创建拥有关系；

顶点分块单元，根据所创建的实体账户顶点确定起点实体账户顶点集和终点实体账户顶点集，所述起点实体账户顶点集和所述终点实体账户顶点集之间不具有重合的实体账户顶点；以及

关联关系生成单元，基于所述起点实体账户顶点集和所述终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

23、一种用于生成应用于基准测试的图数据的装置，包括：

至少两个顶点生成框架，每个顶点生成框架部署在一个第一设备处；

至少两个顶点关系生成框架，每个顶点关系生成框架部署在一个第二设备处；以及

顶点分块框架，部署在第三设备处，

其中，各个顶点生成框架被配置为：

创建多个实体顶点；

创建所述顶点分块框架所抽取的各个第一实体顶点的对应实体账户顶点；及
在各个实体账户顶点和对应的实体顶点之间创建拥有关系；

所述顶点分块框架被配置为从所创建的实体顶点中为各个顶点生成框架抽取多个
第一实体顶点；以及从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体
账户顶点集和终点实体账户顶点集；

各个顶点关系生成框架被配置为基于所抽取的起点实体账户顶点集和终点实体账
户顶点集，创建实体账户顶点之间的账户关联关系。

24、如权利要求 23 所述的装置，还包括：

部署在各个第一设备处的数据分布接口，获取顶点出度信息，

其中，各个实体顶点的顶点出度基于对应的顶点出度分布信息确定。

25、如权利要求 23 所述的装置，其中，每个实体账户顶点的账户顶点属性包括顶
点出度和顶点入度，

各个顶点关系生成框架被配置为：

根据所述起点实体账户顶点集中的各个起点实体账户顶点的顶点出度以及所述终
点实体账户顶点集中的各个终点实体账户顶点的顶点入度，确定各个起点实体账户顶点
和各个终点实体账户顶点的选中概率；

循环执行下述过程，直到所创建的账户关联关系达到第一预定数目 M：

基于各个起点实体账户顶点和各个终点实体账户顶点的选中概率，从所述起
点实体账户顶点集和所述终点实体账户顶点集中选择至少一个起点实体账户顶点以及
对应的终点实体账户顶点；

计算所选择的起点实体账户顶点和终点实体账户顶点之间的属性距离；

基于所计算出的属性距离，确定所选择的起点实体账户顶点和终点实体账
户顶点之间的关系创建概率；以及

基于所述关系创建概率来在所选择的起点实体账户顶点和终点实体账户顶
点之间创建账户关联关系。

26、如权利要求 25 所述的装置，还包括：

部署在各个第一设备处的数据分布接口，获取实体账户顶点的顶点出度/入度分布信
息；

其中，各个实体账户顶点的顶点出度和顶点入度根据对应的顶点出度/入度分布信息
确定。

27、如权利要求 25 所述的装置，还包括：

部署在各个第一设备处的数据分布接口，获取社交网络出度/入度分布信息；

各个顶点生成框架根据所获取的社交网络出度/入度分布信息来在所述实体顶点之
间创建认识/从属关系，并且各个顶点关系生成框架基于所计算出的属性距离以及所选择
的起点实体账户顶点和终点实体账户顶点各自所属实体顶点之间的认识/从属关系，确定
所选择的起点实体账户顶点和终点实体账户顶点之间的关系创建概率。

28、如权利要求 23 所述的装置，其中，所述多个第一设备中的部分第一设备或每
个第一设备分别与所述多个第二设备中的一个第二设备相同，和/或所述第三设备与所
述多个第一设备和/或所述多个第二设备中的一个设备相同。

29、一种用于生成应用于基准测试的图数据的系统，包括：

至少两个第一设备，每个第一设备部署有顶点生成框架；

至少两个第二设备，每个第二设备部署有顶点关系生成框架；以及

第三设备，部署有顶点分块框架，

其中，各个顶点生成框架被配置为：

创建多个实体顶点；

创建所述顶点分块框架所抽取的各个第一实体顶点的对应实体账户顶点；及

在各个实体账户顶点与对应的实体顶点之间创建拥有关系，以及；

所述顶点分块框架被配置为从所创建的实体顶点中为各个顶点生成框架抽取多个第一实体顶点；以及从所创建的实体账户顶点中为各个顶点关系生成框架抽取起点实体账户顶点集和终点实体账户顶点集；

各个顶点关系生成框架被配置为基于所抽取的起点实体账户顶点集和终点实体账户顶点集，创建实体账户顶点之间的账户关联关系。

30、一种用于生成应用于基准测试的图数据的装置，包括：

至少一个处理器，

与所述至少一个处理器耦合的存储器，以及

存储在所述存储器中的计算机程序，所述至少一个处理器执行所述计算机程序来实现如权利要求 1 到 11 中任一所述的方法或者实现如权利要求 12 到 21 中任一所述的方法。

31、一种计算机可读存储介质，其存储有可执行指令，所述指令当被执行时使得处理器执行如权利要求 1 到 11 中任一所述的方法或者实现如权利要求 12 到 21 中任一所述的方法。

32、一种计算机程序产品，包括计算机程序，所述计算机程序被处理器执行来实现如权利要求 1 到 11 中任一所述的方法或者实现如权利要求 12 到 21 中任一所述的方法。