**说 明 书 摘 要**

本说明书公开了一种图数据处理系统、方法、设备及存储介质，可以根据预设的图模型匹配算法，从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，针对每个集合操作，根据执行该集合操作所涉及的两个集合中元素的数量，确定按照不同的执行策略执行该集合操作所占用的处理单元的性能对应的代价值，进而可以选择对处理单元的性能占用最小的执行策略来执行该集合操作，从而可以提升图数据处理效率。

**摘 要 附 图**



**权 利 要 求 书**

1、一种图数据处理系统，其特征在于，所述图数据处理系统包括：各处理单元以及存储器，其中，每个处理单元中还设有：决策模块；

所述处理单元用于，针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，所述集合操作用于表示对所述目标图数据中的两个节点的邻居节点集执行指定类型的操作，所述指定类型的操作包括：取交集、取差集中的至少一种；

通过所述决策模块，所述处理单元针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据所述代价值，从各执行策略中选取目标策略，以使所述处理单元根据所述目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在所述存储器中；响应于得到每个集合操作对应的执行结果后，从所述存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定所述目标图数据中与所述指定图模式相匹配的子图，以根据所述子图，执行任务；

所述存储器用于，对各集合操作对应的执行结果进行存储。

2、如权利要求1所述的图数据处理系统，其特征在于，所述处理单元中还设有：检测模块；

所述检测模块用于，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值，若是，则确定该集合操作为目标集合操作，并将所述目标集合操作的执行结果持久化保存，以在需要再次执行该集合操作时重复使用。

3、如权利要求1所述的图数据处理系统，其特征在于，所述图数据处理系统还包括：动态划分模块；

所述动态划分模块用于，获取原始图数据，针对所述原始图数据中的每个节点，判断该节点的度数是否超过预设阈值，若是，则确定该节点为中心节点，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点；

根据每个中心节点以及每个中心节点的关联节点，确定各图数据分块，并针对每个图数据分块，将该图数据分块作为目标图数据，以使所述处理单元对该图数据分块进行处理。

4、如权利要求3所述的图数据处理系统，其特征在于，所述动态划分模块用于，针对每个图数据分块，生成用于对该图数据分块进行处理的处理任务，并将所述处理任务添加到预设的任务队列中，以使所述处理单元从所述任务队列中获取所述处理任务，并将所述处理任务对应的图数据分块作为目标图数据。

5、如权利要求3所述的图数据处理系统，其特征在于，所述动态划分模块用于，针对每个中心节点，判断该中心节点是否为已访问节点，若否，则通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，并将该中心节点设为已访问节点。

6、如权利要求2所述的图数据处理系统，其特征在于，所述检测模块用于，针对每个集合操作，判断是否存在该集合操作对应的唯一标识，若否，则根据该集合操作所述涉及的两个集合以及该集合操作对应的指定类型，生成该集合操作对应的唯一标识并保存。

7、如权利要求1所述的图数据处理系统，其特征在于，所述决策模块用于针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及执行该集合操作的处理单元的处理数据，确定按照每种执行策略执行该集合操作所需的计算时间和访存时间，根据所述计算时间和所述访存时间，确定每种执行策略对应的代价值，所述处理单元的处理数据包括：处理单元的带宽、处理单元访存延迟。

8、一种图数据处理方法，其特征在于，所述图数据处理系统包括：各处理单元以及存储器，其中，每个处理单元中还设有：决策模块，所述方法包括：

所述处理单元针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，所述集合操作用于表示对所述目标图数据中的两个节点的邻居节点集执行指定类型的操作，所述指定类型的操作包括：取交集、取差集中的至少一种；

通过所述决策模块，针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据所述代价值，从各执行策略中选取目标策略，以使所述处理单元根据所述目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在所述存储器中；

响应于得到每个集合操作对应的执行结果后，从所述存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定所述目标图数据中与所述指定图模式相匹配的子图，以根据所述子图，执行任务。

9、如权利要求8所述的方法，其特征在于，所述处理单元中还设有：检测模块，所述方法还包括：

通过所述检测模块，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值，若是，则确定该集合操作为目标集合操作，并将所述目标集合操作的执行结果持久化保存，以在需要再次执行该集合操作时重复使用。

10、如权利要求8所述的方法，其特征在于，所述图数据处理系统还包括：动态划分模块；

针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，具体包括：

将获取到的图数据分块作为目标图数据，根据预设的图模式匹配算法，确定从所述目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，所述图数据分块是通过所述动态划分模块获取原始图数据，针对所述原始图数据中的每个节点，判断该节点的度数是否超过预设阈值，若是，则确定该节点为中心节点，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，根据每个中心节点以及每个中心节点的关联节点确定的。

11、如权利要求10所述的方法，其特征在于，将获取到的图数据分块作为目标图数据，具体包括：

从预设的任务队列中获取处理任务，并将所述处理任务对应的图数据分块作为目标图数据，所述处理任务是所述动态划分模块针对每个图数据分块，生成并添加到预设的任务队列中的。

12、如权利要求10所述的方法，其特征在于，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点之前，所述方法还包括：

针对每个中心节点，判断该中心节点是否为已访问节点；

针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，具体包括：

若该中心节点不为已访问节点，则通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，并将该中心节点设为已访问节点。

13、如权利要求9所述的方法，其特征在于，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值之前，所述方法还包括：

针对每个集合操作，判断是否存在该集合操作对应的唯一标识，若否，则根据该集合操作所述涉及的两个集合以及该集合操作对应的指定类型，生成该集合操作对应的唯一标识并保存。

14、如权利要求8所述的方法，其特征在于，根据该集合操作所涉及的两个集合中包含的元素的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，具体包括：

针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及执行该集合操作的处理单元的处理数据，确定按照每种执行策略执行该集合操作所需的计算时间和访存时间，根据所述计算时间和所述访存时间，确定每种执行策略对应的代价值，所述处理单元的处理数据包括：处理单元的带宽、处理单元访存延迟。

15、一种计算机可读存储介质，其特征在于，所述存储介质存储有计算机程序，所述计算机程序被处理器执行时实现上述权利要求8~14任一项所述的方法。

16、一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时实现上述权利要求8~14任一项所述的方法。

**说 明 书**

一种图数据处理系统、方法、设备及存储介质

**技术领域**

本说明书涉及图计算技术领域，尤其涉及一种图数据处理系统、方法、设备及存储介质。

**背景技术**

随着大数据技术的发展，图数据的规模在不断增长, 图数据的类型也在不断增多, 并且图数据中的每个节点对应的实体对象之间的关系也变得更加复杂。而如何分析和挖掘图数据中蕴含的复杂关系, 成为当前的研究热点。

现有技术中，通常采用图模式匹配算法从图数据中提取出与指定图模式相匹配的子图，并基于提取出的子图进行任务执行。其中，这里的图模式用于表示在实际场景中的各特定实体之间所呈现的关联规则，例如：医药领域中，存在一种特定的分子结构可以用于医疗，这种特征的分子结构中包含的各分子之间的连接关系可以被设计为一种图模式，并从其他大分子物质对应的图数据中，查询与这种图模式相匹配的子图，以确定该其他大分子物质中是否包含这种特定的分子结构。但是，由于目前图模式匹配的算法的执行效率较低，无法满足规模较大的图数据的处理需求。

因此，如何能够提升图模式匹配算法的执行效率，则是一个亟待解决的问题。

**发明内容**

本说明书提供一种图数据处理方法、装置、设备及存储介质，以部分的解决现有技术存在的上述问题。

本说明书采用下述技术方案：

本说明书提供了一种图数据处理系统，所述图数据处理系统包括：各处理单元以及存储器，其中，每个处理单元中还设有：决策模块；

所述处理单元用于，针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，所述集合操作用于表示对所述目标图数据中的两个节点的邻居节点集执行指定类型的操作，所述指定类型的操作包括：取交集、取差集中的至少一种；

通过所述决策模块，所述处理单元针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据所述代价值，从各执行策略中选取目标策略，以使所述处理单元根据所述目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在所述存储器中；响应于得到每个集合操作对应的执行结果后，从所述存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定所述目标图数据中与所述指定图模式相匹配的子图，以根据所述子图，执行任务；

所述存储器用于，对各集合操作对应的执行结果进行存储。

可选地，所述处理单元中还设有：检测模块；

所述检测模块用于，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值，若是，则确定该集合操作为目标集合操作，并将所述目标集合操作的执行结果持久化保存，以在需要再次执行该集合操作时重复使用。

可选地，所述图数据处理系统还包括：动态划分模块；

所述动态划分模块用于，获取原始图数据，针对所述原始图数据中的每个节点，判断该节点的度数是否超过预设阈值，若是，则确定该节点为中心节点，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点；

根据每个中心节点以及每个中心节点的关联节点，确定各图数据分块，并针对每个图数据分块，将该图数据分块作为目标图数据，以使所述处理单元对该图数据分块进行处理。

可选地，所述动态划分模块用于，针对每个图数据分块，生成用于对该图数据分块进行处理的处理任务，并将所述处理任务添加到预设的任务队列中，以使所述处理单元从所述任务队列中获取所述处理任务，并将所述处理任务对应的图数据分块作为目标图数据。

可选地，所述动态划分模块用于，针对每个中心节点，判断该中心节点是否为已访问节点，若否，则通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，并将该中心节点设为已访问节点。

可选地，所述检测模块用于，针对每个集合操作，判断是否存在该集合操作对应的唯一标识，若否，则根据该集合操作所述涉及的两个集合以及该集合操作对应的指定类型，生成该集合操作对应的唯一标识并保存。

可选地，所述决策模块用于针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及执行该集合操作的处理单元的处理数据，确定按照每种执行策略执行该集合操作所需的计算时间和访存时间，根据所述计算时间和所述访存时间，确定每种执行策略对应的代价值，所述处理单元的处理数据包括：处理单元的带宽、处理单元访存延迟。

本说明书提供了一种图数据处理方法，所述图数据处理系统包括：各处理单元以及存储器，其中，每个处理单元中还设有：决策模块，所述方法包括：

所述处理单元针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，所述集合操作用于表示对所述目标图数据中的两个节点的邻居节点集执行指定类型的操作，所述指定类型的操作包括：取交集、取差集中的至少一种；

通过所述决策模块，针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据所述代价值，从各执行策略中选取目标策略，以使所述处理单元根据所述目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在所述存储器中；

响应于得到每个集合操作对应的执行结果后，从所述存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定所述目标图数据中与所述指定图模式相匹配的子图，以根据所述子图，执行任务。

可选地，所述处理单元中还设有：检测模块，所述方法还包括：

通过所述检测模块，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值，若是，则确定该集合操作为目标集合操作，并将所述目标集合操作的执行结果持久化保存，以在需要再次执行该集合操作时重复使用。

可选地，所述图数据处理系统还包括：动态划分模块；

针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，具体包括：

将获取到的图数据分块作为目标图数据，根据预设的图模式匹配算法，确定从所述目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，所述图数据分块是通过所述动态划分模块获取原始图数据，针对所述原始图数据中的每个节点，判断该节点的度数是否超过预设阈值，若是，则确定该节点为中心节点，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，根据每个中心节点以及每个中心节点的关联节点确定的。

可选地，将获取到的图数据分块作为目标图数据，具体包括：

从预设的任务队列中获取处理任务，并将所述处理任务对应的图数据分块作为目标图数据，所述处理任务是所述动态划分模块针对每个图数据分块，生成并添加到预设的任务队列中的。

可选地，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点之前，所述方法还包括：

针对每个中心节点，判断该中心节点是否为已访问节点；

针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，具体包括：

若该中心节点不为已访问节点，则通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，并将该中心节点设为已访问节点。

可选地，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值之前，所述方法还包括：

针对每个集合操作，判断是否存在该集合操作对应的唯一标识，若否，则根据该集合操作所述涉及的两个集合以及该集合操作对应的指定类型，生成该集合操作对应的唯一标识并保存。

可选地，根据该集合操作所涉及的两个集合中包含的元素的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，具体包括：

针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及执行该集合操作的处理单元的处理数据，确定按照每种执行策略执行该集合操作所需的计算时间和访存时间，根据所述计算时间和所述访存时间，确定每种执行策略对应的代价值，所述处理单元的处理数据包括：处理单元的带宽、处理单元访存延迟。

本说明书提供了一种计算机可读存储介质，所述存储介质存储有计算机程序，所述计算机程序被处理器执行时实现上述图数据处理方法。

本说明书提供了一种电子设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现上述图数据处理方法。

本说明书采用的上述至少一个技术方案能够达到以下有益效果：

在本说明书提供的图数据处理方法，处理单元针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，集合操作用于表示对目标图数据中的两个节点的邻居节点集执行指定类型的操作，指定类型的操作包括：取交集、取差集中的至少一种，通过决策模块，针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据代价值，从各执行策略中选取目标策略，以使处理单元根据目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在存储器中，响应于得到每个集合操作对应的执行结果后，从存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定目标图数据中与指定图模式相匹配的子图，以根据子图，执行任务。

从上述方法中可以看出，可以根据预设的图模型匹配算法，从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，针对每个集合操作，根据执行该集合操作所涉及的两个集合中元素的数量，确定按照不同的执行策略执行该集合操作所占用的处理单元的性能对应的代价值，进而可以选择对处理单元的性能占用最小的执行策略来执行该集合操作，从而可以提升图数据处理效率。

**附图说明**

此处所说明的附图用来提供对本说明书的进一步理解，构成本说明书的一部分，本说明书的示意性实施例及其说明用于解释本说明书，并不构成对本说明书的不当限定。在附图中：

图1为本说明书中提供的一种图数据处理系统的示意图；

图2为本说明书中提供的图模式匹配算法的示意图；

图3为本说明书提供的动态划分模块的示意图；

图4为本说明书中提供的一种图数据处理方法的示意图；

图5为本说明书提供的一种对应于图4的电子设备的示意图。

**具体实施方式**

为使本说明书的目的、技术方案和优点更加清楚，下面将结合本说明书具体实施例及相应的附图对本说明书技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本说明书一部分实施例，而不是全部的实施例。基于本说明书中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本说明书保护的范围。

以下结合附图，详细说明本说明书各实施例提供的技术方案。

本说明书中提供了一种图数据处理系统，如图1所示：

图1为本说明书中提供的一种图数据处理系统的示意图。

从图1中可以看出，图数据处理系统包括：各处理单元以及存储器，其中，每个处理单元中还设有：决策模块。

其中，处理单元用于针对获取到的目标图数据，根据预设的图模式匹配算法，确定从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，这里的集合操作用于对目标图数据中的两个节点的邻居节点集进行指定类型的操作，这里的指定类型的操作包括：取交集、取差集等操作。

进一步地，决策模块用于，针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据代价值，从各执行策略中选取目标策略（如：确定代价值最小的执行策略为目标策略），以使处理单元根据目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在存储器中，响应于得到每个集合操作对应的执行结果后，从存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定目标图数据中与指定图模式相匹配的子图，以根据子图，执行任务。

其中，上述的执行策略可以根据实际需求设置，诸如：基于合并算法的执行策略、基于二分查找算法的执行策略、基于哈希算法的执行策略、基于比特数组算法的执行策略。

在不同的应用场景中，需要根据子图执行的任务也并不相同，例如：在电商场景中，可以根据用户之间的共同点，确定出具有相近需求的用户之间的关联关系所对应的图模式，进而可以从用户图数据中，确定与上述的图模式相匹配的用户子图，并根据用户子图，为用户子图中的每个用户进行商品推荐。

需要说明的是，从目标图数据中提取出不同的图模式所需的图模式匹配算法也不相同，处理单元可以按照预设的图模式匹配算法，确定从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，具体如图2所示。

图2为本说明书中提供的图模式匹配算法的示意图。

结合图2可以看出，若需要匹配的图模式为互相之间通过一条边相连的三个节点组成三角形结构时，对应的图模式匹配算法可以为针对目标图数据中的每个节点，判断该节点的邻居节点集和该节点的每个邻居节点的邻居节点集之间是否存在交集，若是，则说明存在上述的三角形结构的图模式，若否，则说明不存在上述的三角形结构的图模式。

例如：假设节点A的邻居节点集包括节点B、节点C，节点B的邻居节点集包括节点C、节点A、节点D，此时节点A的邻居节点集和节点B的邻居节点集中均包含有节点C，此时可以说明节点C为节点A和节点B的公共邻居节点，也就是说，节点A、节点B、节点C之间均通过一条边相连，即满足上述的三角形结构的图模式，此时，可以将节点A、节点B、节点C作为图模式匹配算法匹配出的子图。

进一步地，从上述内容中可以看出，在执行图模式匹配算法的过程中存在大量的冗余集合操作，例如：在实际应用场景中，会先取节点A的邻居节点集和节点A的邻居节点B的邻居节点集之间的交集，从而得到节点A、节点B、节点C对应的子图，除此之外，由于图模式匹配算法需要从每个节点出发，来确定该节点的邻居节点集和该节点的邻居节点的邻居节点集之间的交集，因此，除了从节点A出发，来确定节点A的邻居节点集和节点A的邻居节点B的邻居节点集之间的交集之外，还会从节点B出发，确定节点B的邻居节点集和节点B的邻居节点A的邻居节点集之间的交集，从而得到节点A、节点B、节点C对应的子图，因此，降低了图模式匹配算法的执行效率。

基于此，在上述的处理单元中还可以设有：检测模块，其中，检测模块用于针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值，若是，则确定该集合操作为目标集合操作，并将目标集合操作的执行结果进行持久化保存（可以理解为将目标集合操作的执行结果保存到指定内存空间中，以避免在执行图模式匹配算法的过程中将目标结合操作的执行结果覆盖），以在需要再次执行该集合操作时重复使用。

当然，检测模块在判断该集合操作被执行的次数是否超过预设阈值之前，还可以针对每个集合操作，判断是否存在该集合操作对应的唯一标识，若否，则根据该集合操作涉及的两个集合以及该集合操作对应的指定类型，生成该集合操作对应的唯一标识并保存，若是，则可以根据该集合操作对应的唯一标识，确定该集合操作被执行的次数，并判断该集合操作被执行的次数是否超过预设阈值。

需要说明的是，针对每个节点，可以使用该节点对应的唯一标识对该节点的邻居节点集进行标识。

针对中间数据（即，二个集合取交集或取差集后的结果）组成的节点集合，由于中间数据与其他节点集合再次进行集合操作的频率较低，并且由于由中间数据组成的节点集合中节点的数量较少，所以检测模块并不会针对中间数据组成的节点集合对应的集合操作进行检测。

进一步地，检测模块可以针对每个集合操作，判断该集合操作是否为目标集合操作，若是，则可以返回预先保存的该目标集合操作的执行结果在存储器中的地址，若否，则可以将该集合操作传输给决策模块进行执行。

在实际应用场景中，图数据处理系统所获取到原始图数据的规模往往较大，从而使得在直接将原始图数据作为目标图数据进行图模式匹配时，往往会导致由于数据规模较大，且较为分散，从而使得执行一个集合操作会被多个处理单元执行，进而使处理单元之间针对该集合操作的被执行次数的统计值需要进行同步，进而导致大量的额外性能开销。

基于此，上述的图数据处理系统中还可以包含有动态划分模块，其中，动态划分模块用于，获取原始图数据，针对原始图数据中的每个节点，判断该节点的度数是否超过预设阈值，若是，则确定该节点为中心节点，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，根据每个中心节点以及每个中心节点的关联节点，确定各图数据分块，并针对每个图数据分块，将该图数据分块作为目标图数据，以使处理单元对该图数据分块进行处理，具体如图3所示。

图3为本说明书提供的动态划分模块的示意图。

结合图3可以看出，动态划分模块可以针对每个中心节点，判断该中心节点是否为已访问节点，若否，则通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，并将该中心节点设为已访问节点。

其中，上述的邻居节点遍历的方法可以为诸如：深度优先遍历，广度优先遍历。针对每个中心节点，当访问该中心节点的邻居节点的深度达到预设阈值，或者，没有可以访问的节点后，则结束上述的多轮邻居节点遍历，得到图数据分块。

进一步地，动态划分模块可以针对每个图数据分块，生成用于对该图数据分块进行处理的处理任务，并将处理任务添加到预设的任务队列中，以使处理单元从任务队列中获取处理任务，并将处理任务对应的图数据分块作为目标图数据。

需要说明的是，由于动态划分模块可以将原始图数据划分为若干个图数据分块，并将每个图数据分块的处理任务分配给一个处理单元来执行，所以，使得一个集合操作往往集中在一个处理单元中，进而可以避免个处理单元之间的数据同步而造成了处理单元的性能损耗。

另外，上述的图数据处理系统可以为每个处理单元均分配有相互独立的存储空间用于保存集合操作的执行结果，以及集合操作被执行的次数等数据。

需要说明的是，通过在动态划分模块以及各处理单元之间设置的任务队列，可以使得各处理单元执行各处理任务，以及动态划分模块对原始图数据进行划分，并生成各处理任务可以是并行执行的，从而可以提升图数据的处理效率。

进一步地，决策模块在确定每个集合操作对应的目标执行策略时，可以针对每个集合操作，根据该集合操作所涉及的两个集合中包含的元素的数量，以及执行该集合操作的处理单元的处理数据，确定按照每种执行策略执行该集合操作所需的计算时间和访存时间，根据计算时间和访存时间，确定每种执行策略对应的代价值，这里的处理单元的处理数据包括：处理单元的带宽、处理单元访存延迟。

另外，在本说明书中，针对计算-访存比大的执行策略（如：基于合并算法的执行策略），可以在处理单元中设置专用逻辑电路模块进行计算。针对计算-访存比小的执行策略（如：基于比特数组算法的执行策略），则可以使用近内存架构，即在存储器内设置位操作，以执行集合操作，可以理解为在集合操作在存储器的内部完成，减少了存储器与处理单元之间进行数据传输所需要的访存时间，从而可以提升的图数据处理效率。

从上述内容中可以看出，可以根据预设的图模型匹配算法，从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，针对每个集合操作，根据执行该集合操作所涉及的两个集合中元素的数量，确定按照不同的执行策略执行该集合操作所占用的处理单元的性能对应的代价值，进而可以选择对处理单元的性能占用最小的执行策略来执行该集合操作，从而可以提升图数据处理效率。

在本说明书中，上述的处理单元可以是指多核处理器的各处理核心，上述的检测模块、决策模块、动态划分模块可以为设置在多核处理器上的硬件单元。

为了进一步地对上述图数据处理系统进行详细说明，本说明书还提供了通过上述的图数据处理系统进行图数据处理的方法，具体如图4所示。

图4为本说明书中提供的一种图数据处理方法的示意图，包括以下步骤：

S401：所述处理单元针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从所述目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，所述集合操作用于表示对所述目标图数据中的两个节点的邻居节点集执行指定类型的操作，所述指定类型的操作包括：取交集、取差集中的至少一种；

S402：通过所述决策模块，针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据所述代价值，从各执行策略中选取目标策略，以使所述处理单元根据所述目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在所述存储器中；

S403：响应于得到每个集合操作对应的执行结果后，从所述存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定所述目标图数据中与所述指定图模式相匹配的子图，以根据所述子图，执行任务。

处理单元针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，集合操作用于表示对目标图数据中的两个节点的邻居节点集执行指定类型的操作，指定类型的操作包括：取交集、取差集中的至少一种；

通过决策模块，针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及预设的代价函数，确定按照每种执行策略执行该集合操作对应的代价值，并根据代价值，从各执行策略中选取目标策略，以使处理单元根据目标策略执行该集合操作，得到该集合操作对应的执行结果，并存储在存储器中；

响应于得到每个集合操作对应的执行结果后，从存储器中读取各集合操作对应的执行结果，并根据各集合操作对应的执行结果，确定目标图数据中与指定图模式相匹配的子图，以根据子图，执行任务。

通过检测模块，针对每个集合操作，判断该集合操作被执行的次数是否超过预设阈值，若是，则确定该集合操作为目标集合操作，并将目标集合操作的执行结果持久化保存，以在需要再次执行该集合操作时重复使用。

针对获取到的目标图数据，根据预设的图模式匹配算法，确定在从目标图数据中提取与指定图模式相匹配的子图时所需的各集合操作，具体包括：

将获取到的图数据分块作为目标图数据，根据预设的图模式匹配算法，确定从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，图数据分块是通过动态划分模块获取原始图数据，针对原始图数据中的每个节点，判断该节点的度数是否超过预设阈值，若是，则确定该节点为中心节点，针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，根据每个中心节点以及每个中心节点的关联节点确定的。

从预设的任务队列中获取处理任务，并将处理任务对应的图数据分块作为目标图数据，处理任务是动态划分模块针对每个图数据分块，生成并添加到预设的任务队列中的。

针对每个中心节点，判断该中心节点是否为已访问节点；

针对每个中心节点，通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，具体包括：

若该中心节点不为已访问节点，则通过多轮邻居节点遍历，确定与该中心节点之间存在连接关系的各节点，作为该中心节点的关联节点，并将该中心节点设为已访问节点。

针对每个集合操作，判断是否存在该集合操作对应的唯一标识，若否，则根据该集合操作涉及的两个集合以及该集合操作对应的指定类型，生成该集合操作对应的唯一标识并保存。

针对每个集合操作，根据该集合操作所涉及的两个节点集合中包含的节点的数量，以及执行该集合操作的处理单元的处理数据，确定按照每种执行策略执行该集合操作所需的计算时间和访存时间，根据计算时间和访存时间，确定每种执行策略对应的代价值，处理单元的处理数据包括：处理单元的带宽、处理单元访存延迟。

从上述内容中可以看出，可以根据预设的图模型匹配算法，从目标图数据中提取与指定图模式相匹配的子图所需的各集合操作，针对每个集合操作，根据执行该集合操作所涉及的两个集合中元素的数量，确定按照不同的执行策略执行该集合操作所占用的处理单元的性能对应的代价值，进而可以选择对处理单元的性能占用最小的执行策略来执行该集合操作，从而可以提升图数据处理效率。

本说明书还提供了一种计算机可读存储介质，该存储介质存储有计算机程序，计算机程序可用于执行上述图1提供的一种的方法。

本说明书还提供了图5所示的一种对应于图1的电子设备的示意结构图。如图5所示，在硬件层面，该电子设备包括处理器、内部总线、网络接口、内存以及非易失性存储器，当然还可能包括其他业务所需要的硬件。处理器从非易失性存储器中读取对应的计算机程序到内存中然后运行，以实现上述图1的方法。

当然，除了软件实现方式之外，本说明书并不排除其他实现方式，比如逻辑器件抑或软硬件结合的方式等等，也就是说以下处理流程的执行主体并不限定于各个逻辑单元，也可以是硬件或逻辑器件。

在20世纪90年代，对于一个技术的改进可以很明显地区分是硬件上的改进（例如，对二极管、晶体管、开关等电路结构的改进）还是软件上的改进（对于方法流程的改进）。然而，随着技术的发展，当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此，不能说一个方法流程的改进就不能用硬件实体模块来实现。例如，可编程逻辑器件（Programmable Logic Device, PLD）（例如现场可编程门阵列（Field Programmable Gate Array，FPGA））就是这样一种集成电路，其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片PLD上，而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且，如今，取代手工地制作集成电路芯片，这种编程也多半改用“逻辑编译器（logic compiler）”软件来实现，它与程序开发撰写时所用的软件编译器相类似，而要编译之前的原始代码也得用特定的编程语言来撰写，此称之为硬件描述语言（Hardware Description Language，HDL），而HDL也并非仅有一种，而是有许多种，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）与Verilog。本领域技术人员也应该清楚，只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中，就可以很容易得到实现该逻辑方法流程的硬件电路。

控制器可以按任何适当的方式实现，例如，控制器可以采取例如微处理器或处理器以及存储可由该（微）处理器执行的计算机可读程序代码（例如软件或固件）的计算机可读介质、逻辑门、开关、专用集成电路（Application Specific Integrated Circuit，ASIC）、可编程逻辑控制器和嵌入微控制器的形式，控制器的例子包括但不限于以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及Silicone Labs C8051F320，存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的，计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本说明书时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

本领域内的技术人员应明白，本说明书的实施例可提供为方法、系统、或计算机程序产品。因此，本说明书可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本说明书可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本说明书是参照根据本说明书实施例的方法、设备（系统）、和计算机程序产品的流程图和／或方框图来描述的。应理解可由计算机程序指令实现流程图和／或方框图中的每一流程和／或方框、以及流程图和／或方框图中的流程和／或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程图数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程图数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程图数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程图数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能的步骤。

在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

本领域技术人员应明白，本说明书的实施例可提供为方法、系统或计算机程序产品。因此，本说明书可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本说明书可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本说明书可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本说明书，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

以上所述仅为本说明书的实施例而已，并不用于限制本说明书。对于本领域技术人员来说，本说明书可以有各种更改和变化。凡在本说明书的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本说明书的权利要求范围之内。

**说 明 书 附 图**



图1

图2



图3



图4



图5