(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102053988 A
(43)申请公布日 2011.05.11

(21)申请号 200910211313.0

(22)申请日 2009.10.30

(71)申请人 国际商业机器公司
地址 美国纽约

(72)发明人 曹楠 时磊 孙冀萌 钱伟江 刘世霞

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 于静 周春燕

(51)Int.Cl.
G06F 17/30(2006.01)

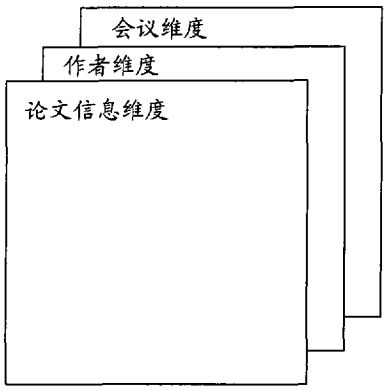
权利要求书 4 页 说明书 10 页 附图 8 页

(54)发明名称

数据集的可视化方法和系统

(57)摘要

本发明提供一种数据集的可视化方法和系统，该方法包括：将数据集基于不同信息维度划分为多个信息层；以及分别将基于不同信息维度的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。在本发明中，通过分别从数据集的不同信息维度呈现数据集的不同概况来可视化数据集，在确保向数据集分析人员呈递数据集的全面信息的同时，防止呈现内容的失真以及视觉混乱。



CN 102053988 A

1. 一种数据集的可视化方法，包括：

将数据集基于不同信息维度划分为多个信息层；以及

分别对基于不同信息维度的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。

2. 根据权利要求 1 所述的方法，其中进行可视化处理的步骤进一步包括：

利用透明色处理技术对上述多个信息层的各自的视图进行处理，以将其组合为一个视图，并且使得上述多个信息层的各自的视图之间能够进行切换。

3. 根据权利要求 1 所述的方法，其中进行可视化处理的步骤进一步包括：

从上述基于不同信息维度的多个信息层中，选择主信息层；

对上述主信息层所包含的数据集进行概括，以构成包含中心节点及其之间的链接关系的样本数据集；以及

以上述样本数据集为布局样本，为上述主信息层生成基于密度的等高线图。

4. 根据权利要求 3 所述的方法，其中上述概括的步骤进一步包括：

对上述主信息层的数据集进行节点概括，以获得包含多个中心节点的中心节点集；以及

根据上述主信息层的数据集，为上述中心节点集中的中心节点进行链接概括，以获得包含中心节点之间的链接关系的中心节点链接集。

5. 根据权利要求 4 所述的方法，其中对上述主信息层的数据集进行节点概括的步骤进一步包括：

从上述主信息层的数据集中，根据节点的中心度，选择出一个最重要的节点，将其移动到中心节点集中；

依次执行以下步骤，直到中心节点集中的中心节点数达到预定的值：

对于中心节点集中的各个中心节点，计算其与上述主信息层的数据集中未被选择到中心节点集中的节点之间的最短距离向量；以及

从主信息层的数据集中未被选择到中心节点集中的节点中选择出一个与中心节点的最短距离是最短的这样的节点，移动到中心节点集中。

6. 根据权利要求 4 所述的方法，其中为上述中心节点集中的中心节点进行链接概括的步骤进一步包括：

对于上述中心节点集中的任意两个中心节点：

利用广度优先搜索算法在主信息层的数据集中寻找所有连接这两个中心节点的路径；以及

对上述路径中长度小于预定的最大长度的路径进行加权合并，作为直接连接上述任意两个中心节点的链接，添加到上述中心节点链接集中。

7. 根据权利要求 3 所述的方法，其中为上述主信息层生成基于密度的等高线图的步骤进一步包括：

为上述样本数据集中的各个中心节点，以其周围的未被选择到上述样本数据集中的节点的数量作为该中心节点的质量，计算该中心节点的密度分布；

将上述样本数据集中的各个中心节点的密度分布结合到用于生成等高线的高度矩阵中；

利用上述高度矩阵，为上述各个中心节点生成等高线并填充颜色，以为上述主信息层生成基于密度的等高线图；以及

将上述主信息层中、与上述各个中心节点相对应的背景信息布局到上述基于密度的等高线图上。

8. 根据权利要求 3 所述的方法，其中进行可视化处理的步骤还包括：

将非主信息层中与上述主信息层中的中心节点的背景信息对应的信息布局到非主信息层的等高线图上，其中非主信息层的等高线图与主信息层的等高线图一致。

9. 根据权利要求 7 所述的方法，其中上述计算中心节点的密度分布的步骤进一步包括：

对于上述样本数据集中的各个中心节点，根据下式计算密度分布：

$$f(x)=\frac{1}{n}\sum_{i=1}^{m}\frac{m_i}{h}K\left(\frac{x-X_i}{h}\right)$$

其中，x 表示屏幕上的某个位置的二维坐标，X_i 表示中心节点 i 在屏幕上的二维坐标，n 是上述主信息层中的总节点数，m 是上述样本数据集中的中心节点数，m_i 是上述主信息层中未被选择到样本数据集中的、中心节点 i 的周围节点的数量，h 是带宽，K() 是核函数。

10. 根据权利要求 9 所述的方法，其中上述带宽 h 是通过交叉验证而得到的、使下式的结果最小的值：

$$\hat{J}(h)\approx\frac{1}{hn^2}\sum_i\sum_j K^*\left(\frac{X_i-X_j}{h}\right)+\frac{2}{hn}K(0)$$

其中，$K^*(x)=K^{(2)}(x)-2K(x)$，$K^{(2)}(x)=\int K(x-y)K(y)dy$，K(x) 为高斯分布函数 N(0,1)，$K^{(2)}(x)$ 为高斯分布函数 N(0,2)。

11. 根据权利要求 9 所述的方法，其中上述密度分布结合步骤进一步包括：

根据下式对上述样本数据集中的各个中心节点的密度分布进行合成，以生成高度矩阵的每一坐标处的合成密度分布：

$$f(x)=\sum_{G\text{中的所有}P_s}f_s(x)$$

其中，G 表示上述样本数据集，p_s 表示样本数据集 G 中的某个中心节点，$f_s(x)$ 是中心节点 p_s 的密度分布。

12. 一种数据集的可视化系统，包括：

分层单元，其将数据集基于不同信息维度划分为多个信息层；以及

可视化单元，其分别对基于不同信息维度的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。

13. 根据权利要求 12 所述的系统，其中上述可视化单元，利用透明色处理技术对该多个信息层的各自的视图进行处理，以将其组合为一个视图，并且使得上述多个信息层的各自的视图之间能够进行切换。

14. 根据权利要求 12 所述的系统，其中上述可视化单元进一步包括：

主信息层选择单元，其从上述基于不同信息维度的多个信息层中，选择主信息层；

数据集概括单元，其对上述主信息层所包含的数据集进行概括，以构成包含中心节点及其之间的链接关系的样本数据集；以及

视图生成单元，其以上述样本数据集为布局样本，为上述主信息层生成基于密度的等高线图。

15. 根据权利要求 14 所述的系统，其中上述数据集概括单元进一步包括：

节点概括单元，其对上述主信息层的数据集进行节点概括，以获得包含多个中心节点的中心节点集；以及

链接概括单元，其根据上述主信息层的数据集，为上述中心节点集中的中心节点进行链接概括，以获得包含中心节点之间的链接关系的中心节点链接集。

16. 根据权利要求 15 所述的系统，其中上述节点概括单元：

从上述主信息层的数据集中，根据节点的中心度，选择出一个最重要的节点，将其移动到中心节点集中；

依次进行以下处理，直到中心节点集中的中心节点数达到预定的值：

对于中心节点集中的各个中心节点，计算其与上述主信息层的数据集中未被选择到中心节点集中的节点之间的最短距离向量；以及

从主信息层的数据集中未被选择到中心节点集中的节点中选择出一个与中心节点的最短距离是最短的这样的节点，移动到中心节点集中。

17. 根据权利要求 15 所述的系统，其中上述链接概括单元对于上述中心节点集中的任意两个中心节点：

利用广度优先搜索算法在主信息层的数据集中寻找所有连接这两个中心节点的路径；以及

对上述路径中长度小于预定的最大长度的路径进行加权合并，作为直接连接上述任意两个中心节点的链接，添加到上述中心节点链接集中。

18. 根据权利要求 14 所述的系统，其中上述视图生成单元进一步包括：

密度分布计算单元，其为上述样本数据集中的各个中心节点，以其周围的未被选择到上述样本数据集中的节点的数量作为该中心节点的质量，计算该中心节点的密度分布；

密度分布结合单元，其将上述密度分布计算单元所计算出的各个中心节点的密度分布结合到用于生成等高线的高度矩阵中；

等高线生成单元，其利用上述高度矩阵，为上述各个中心节点生成等高线并填充颜色，以为上述主信息层生成基于密度的等高线图；以及

信息布局单元，其将上述主信息层中与上述各个中心节点相对应的背景信息布局到上述基于密度的等高线图上。

19. 根据权利要求 14 所述的系统，其中上述视图生成单元，将非主信息层中与上述主信息层中的中心节点的背景信息对应的信息布局到非主信息层的等高线图上，其中非主信息层的等高线图与主信息层的等高线图一致。

20. 根据权利要求 18 所述的系统，其中上述密度分布计算单元，对于上述样本数据集中的各个中心节点，根据下式计算密度分布：

$$f(x)=\frac{1}{n}\sum_{i=1}^{m}\frac{m_i}{h}K\left(\frac{x-X_i}{h}\right)$$

其中，x 表示屏幕上的某个位置的二维坐标，X_i 表示中心节点 i 在屏幕上的二维坐标，n 是上述主信息层中的总节点数，m 是上述样本数据集中的中心节点数，m_i 是上述主信息层

中未被选择到样本数据集中的、中心节点 i 的周围节点的数量，h 是带宽，K() 是核函数。

21. 根据权利要求 20 所述的系统，其中上述带宽 h 是通过交叉验证而得到的、使下式的结果最小的值：

$$\hat{J}(h) \approx \frac{1}{hn^2}\sum_i\sum_j K^*\left(\frac{X_i - X_j}{h}\right) + \frac{2}{hn}K(0)$$

其中，$K^*(x) = K^{(2)}(x) - 2K(x)$，$K^{(2)}(x) = \int K(x-y)K(y)dy$，K(x) 为高斯分布函数 N(0,1)，$K^{(2)}(x)$ 为高斯分布函数 N(0,2)。

22. 根据权利要求 20 所述的系统，其中上述密度分布结合单元根据下式对上述样本数据集中的各个中心节点的密度分布进行合成，以生成高度矩阵的每一坐标处的合成密度分布：

$$f(x) = \sum_{G\text{中的所有}P_s} f_s(x)$$

其中，G 表示上述样本数据集，p_s 表示样本数据集 G 中的某个中心节点，$f_s(x)$ 是中心节点 p_s 的密度分布。

数据集的可视化方法和系统

技术领域

[0001] 本发明涉及数据处理领域，具体地，涉及数据集的可视化方法和系统。

背景技术

[0002] 社会网络是由多个节点（通常代表个人或组织）构成的社会结构，其中的节点相互之间通过一种或多种特定类型的依赖关系联结起来。节点之间的依赖关系例如是金融关系、人际关系、社会关系等。社会网络，作为自然结构出现在我们的日常生活中，节点之间的关系能够揭示关于该结构的诸多信息。

[0003] 1964 年以来，社会网络分析便成为一个重要的研究方向，目前已经发展成为具有其自身的理论说明、方法、社会网络分析软件及研究人员等的范型。

[0004] 对于社会网络分析来说，可视化是能够提供极大便利的重要技术。目前，社会网络的可视化主要分为两种类型：第一种类型是如图 1(a) 所示的节点链接图那样仅呈现节点之间的依赖关系而忽视了节点的背景信息的可视化方法，第二种类型是如图 1(b) 所示的那样不仅呈现了节点之间的依赖关系而且还呈现了节点的背景信息的可视化方法。

[0005] 在社会网络分析中，分析人员对于社会网络的研究不仅专注于社会网络的拓扑，而且还要考虑社会网络中各个节点背后的背景信息。

[0006] 因此，上述第一种类型的社会网络的可视化方法，由于不能够呈现节点背后的背景信息，所以存在着不利于社会网络分析的顺利开展的问题。

[0007] 此外，在上述第二种类型的社会网络的可视化方法中，即使呈现了节点的背景信息，但是也会由于呈现方式的混乱，而存在着不能够有效地引导社会网络分析的顺利开展的问题。例如就图 1(b) 所示的可视化方法而言，可以看出，由于呈现方式的不适当，在单个视图上同时呈现了大量节点的多种背景信息，引起了极度的视觉混乱。

[0008] 此外，网络的数据集通常是多维的，即包含多种属性的信息，但在上述第二种类型的可视化方法中，除了可能出现图 1(b) 所示的呈现方式混乱的情况之外，还存在着将高维度（多种属性）的背景信息压缩为低维度（少数一种或几种属性）的背景信息的情况。在此情况下，由于节点的大部分背景信息的省略，将引起呈现内容的失真。

[0009] 上述这些问题不仅仅存在于社会网络的可视化的情况，而且还存在于其他诸如 SMS(Short Message Service，短消息服务）网络、互联网等基于内容的网络的可视化情况。

发明内容

[0010] 鉴于上述问题，本发明提供一种数据集的可视化方法和系统，以便通过分别从数据集的不同信息维度呈现数据集的不同概况来可视化数据集，在确保向数据集分析人员呈递数据集的全面信息的同时，防止呈现内容的失真以及视觉混乱。

[0011] 根据本发明的一个方面，提供了一种数据集的可视化方法，包括：将数据集基于不同信息维度划分为多个信息层；以及分别将基于不同信息维度划分的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。

[0012] 根据本发明的另一个方面，提供了一种数据集的可视化系统，包括：分层单元，其将数据集基于不同信息维度划分为多个信息层；以及可视化单元，其分别将基于不同信息维度的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。

[0013] 如果采用本发明，则通过分别从数据集的不同信息维度呈现数据集的不同概况来可视化数据集，使数据集分析人员能够根据自身的需要从不同的角度获得关于数据集的不同信息，从而有助于数据集分析的开展。

附图说明

[0014] 相信通过以下结合附图对本发明具体实施方式的说明，能够使人们更好地了解本发明上述的特点、优点和目的。

[0015] 图 1 是现有的社会网络的可视化方法的图示说明；

[0016] 图 2 是根据本发明实施例的网络的可视化方法的流程图；

[0017] 图 3 是图 2 中的步骤 205 的图示说明；

[0018] 图 4 是图 2 中的步骤 210 的图示说明；

[0019] 图 5 是图 2 中的步骤 210 的详细流程图；

[0020] 图 6 是图 5 中的步骤 510 的详细流程图；

[0021] 图 7 是图 6 中的步骤 605 的详细流程图；

[0022] 图 8 是图 5 中的步骤 515 的详细流程图；以及

[0023] 图 9 是根据本发明实施例的网络的可视化系统的方框图。

具体实施方式

[0024] 下面就结合附图对本发明的各个优选实施例进行详细说明。

[0025] 图 2 是根据本发明实施例的网络的可视化方法的流程图。

[0026] 如图 2 所示，本实施例的网络的可视化方法，在步骤 205，将网络的数据集基于不同信息维度划分为多个信息层。其中，每一信息维度的信息层是由上述网络的数据集中该信息维度的数据组成的。

[0027] 在本步骤中，可以根据网络的数据集中所包含的任何信息维度，来对网络进行信息层的划分。例如，在一个与论文有关的网络的情况下，可以理解，论文数据集中将会包含诸如论文信息、作者、会议等多种维度的信息。在此情况下，可以如图 3(a) 所示，将与论文有关的网络划分为基于论文信息维度的信息层、基于作者维度的信息层和基于会议维度的信息层。

[0028] 此外，在一个实施例中，在网络的数据集中包含较少的信息维度时，在本步骤中，也可以如图 3(b) 所示，简单地将网络划分为基于网络拓扑的信息层和在网络拓扑的基础上附加了背景信息的信息层。其中，基于网络拓扑的信息层仅包含与网络拓扑有关的信息，即网络中的各个节点以及各个节点之间的链接关系。此外，在网络拓扑的基础上附加了背景信息的信息层，则除了包含与网络拓扑有关的信息之外，还包含该网络中的多个节点的属性描述。

[0029] 此外，在本步骤中，也可以基于网络的数据集中隐含而非直接存在的信息维度来生成信息层。例如，在与文档有关的网络的情况下，可以根据文档中所隐含的关键字，生成

基于关键字维度的信息层。在此情况下，如果与文档有关的网络的数据集仅给出文档而并没有直接给出文档中所包含的关键字，则在本步骤中，需要首先采用适合的内容提取模型、诸如 TF-IDF 和 LDA 等，从各个文档中提取出关键字信息，然后再根据所提取的关键字信息，划分成基于关键字维度的信息层。

[0030] 在步骤 210，分别对基于不同信息维度的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。

[0031] 在本步骤中，可以采用本领域中任何一种已有的可视化方法来分别为上述多个信息层生成视图。例如，可以采用等高线图生成方法来分别生成上述多个信息层的等高线图。

[0032] 此外，在分别将上述多个信息层可视化时，该多个信息层的各自的视图的呈现方式也可以是多种的。

[0033] 例如在一个实施例中，可以将该多个信息层的各自的视图组合为一个视图，来呈现给分析人员，并且使得分析人员能够在上述多个信息层的各自的视图之间进行切换。

[0034] 在此情况下，可以利用 alpha bending（透明色处理）技术，来进行多个信息层的视图的组合。具体地，当分析人员聚焦于整个网络时，调整各个信息层的色彩 alpha 值，被聚焦的信息层采用较大的 alpha 值，而其他信息层采用较小的 alpha 值，从而使各个信息层能够重叠在一起，而在分析人员希望聚焦于多个信息层中的某一个信息层从而切换到该信息层时，改变该信息层的视图的色彩 alpha 值，将其设置为非透明，同时将其他信息层的视图设置为背景视图而不可见。

[0035] 此外，关于该多个信息层的视图之间的切换，可以通过提供切换按钮或菜单来实现瞬时切换，也可以通过提供滚动条，采用过渡的色彩 alpha 值的平滑方式来实现。通过提供滚动条，能够实现多个信息层的视图之间的平滑切换。

[0036] 此外，在另一个实施例中，在本步骤中，也可以将上述多个信息层的视图作为单独的视图呈现给分析人员，使分析人员无需切换便能够同时观看到网络的不同信息层的视图。

[0037] 此外，在本步骤中，除了能够采用本领域中任何一种已有的可视化方法来分别为上述多个信息层生成视图之外，也可以采用根据本发明一个实施例的基于密度的等高线图生成方法，来为上述多个信息层中的至少一个生成基于密度的等高线图。

[0038] 关于根据本发明一个实施例的基于密度的等高线图生成方法，为了能够直观地理解，图 4(a) ～ (c) 分别示出了对于某一与论文有关的网络，在将该网络分别划分为基于会议维度的信息层、基于作者维度的信息层和基于关键字维度的信息层的情况下，采用根据本发明一个实施例的该基于密度的等高线图生成方法，分别将各个信息层可视化而得到的示例性视图。如上所述，这些视图可以利用 alpha bending 技术进行处理，组合为一个视图，并使分析人员能够通过切换来观看各个视图。此外，这些视图也可以作为单独的视图分别呈现给分析人员。

[0039] 此外，图 4(d) ～ (e) 示出了对于某一网络，在简单地将该网络划分为基于网络拓扑的信息层和在网络拓扑的基础上附加了背景信息的信息层的情况下，采用根据本发明一个实施例的该基于密度的等高线图生成方法分别将各个信息层可视化而得到的示例性视图。同样，这些视图可以组合地呈现，也可以单独地呈现。

[0040] 在根据本发明一个实施例的该基于密度的等高线图生成方法中，采用等高线以及

颜色的结合来表示节点之间的关系。具体地，在该方法中，仅提取并布局重要的节点，并且利用等高线来表示未提取出的潜在节点及其之间的关系，而等高线内的填充颜色则用于表示节点之间的不同等级的关系。例如，等高线内的填充颜色越深，表示该等高线内的节点之间的关系越紧密。此外，等高线内的填充颜色还用于表示等高线内的信息密度，该信息密度是利用等高线内围绕着被布局的重要节点的、其他未呈现节点计算出的。

[0041] 下面关于根据本发明一个实施例的该基于密度的等高线图生成方法，结合图 5-8 进行详细描述。图 5-8 是示出在图 1 的步骤 210 中采用根据本发明一个实施例的该基于密度的等高线图生成方法将上述基于不同信息维度的多个信息层中的至少一个可视化的过程的详细流程图。

[0042] 具体地，如图 5 所示，首先在步骤 505，从上述基于不同信息维度的多个信息层中，选择主信息层。

[0043] 在本步骤中，可以采用本领域中任何一种已有的布局方法，对于上述多个信息层的每一个，分别根据该信息层所包含的数据集，生成视图，进而根据所生成的视图选择出能够得到最佳布局效果的视图的信息层，作为主信息层。具体地，可以根据以下条件来衡量视图的布局效果：

[0044] a) 具有较佳的拓扑结构，能够清晰地划分为几个部分；

[0045] b) 具有良好的对称结构，所谓良好的对称结构，是这样来评价的：选择视图的中心点（到视图的四周距离都相同或近似的节点），以该中心点为中心画一个十字，将视图分成四份，如果每一份中节点的数量都相同，那么视图就具有良好的对称结构；

[0046] c) 平均路径长度短，所谓平均路径长度，是这样计算得到的：在视图中选择任意两个节点组成一个节点对，计算它们之间的最短距离，进而计算视图中所存在的所有节点对的最短距离的平均值；

[0047] d) 视图的规模较小，即视图中所包含的节点的数目较少。

[0048] 在步骤 510，对上述主信息层所包含的数据集进行概括，以构成包含中心节点及其之间的链接关系的样本数据集。该样本数据集，用作为在为各个信息层生成视图时的布局样本。

[0049] 一般而言，网络的数据集的信息量都是非常大的，进而根据网络的数据集所得到的各个信息层的信息量也都是非常大的，这样，如果将各个信息层的所有信息都直接呈现在视图上，则会造成视觉混乱。所以，在本步骤中，在生成视图之前，对作为各个信息层的视图的布局样本的主信息层的数据集进行采样。当然，采样后的样本数据集，应该由能够体现原主信息层的数据集概况的典型数据、即重要的节点及其之间的链接构成。

[0050] 关于该步骤，结合图 6 进行详细描述。

[0051] 如图 6 所示，首先，在步骤 605，对上述主信息层的数据集进行节点概括，以获得包含多个中心节点的中心节点集。

[0052] 在一个实施例中，在本步骤中，根据节点的中心度对上述主信息层的数据集进行节点概括。也就是说，从该主信息层的数据集中提取出多个分别处于其他节点所包围的中心的中心节点，构成中心节点集。

[0053] 具体地，首先，根据节点的中心度，确定一个最重要的节点，然后以该最重要的节点为基准，计算节点之间的最短距离，来选择相互之间距离最远的多个节点，将这些节点作

为中心节点。也就是说，可以认为相互之间距离最远的多个节点是均匀地分布在视图的不同部分上的，所以通过提取这些节点作为中心节点，不会导致某一部分信息的丢失，从而不会导致所生成的视图的极大失真。本领域技术人员可以理解，上述节点的中心度，可以是等级 (degree) 中心度、接近性 (closeness) 中心度、中间性 (betweenness) 中心度等。

[0054] 关于该步骤，可以利用图 7 所示的过程来实现。在图 7 所示的过程中，假设需要从上述主信息层的数据集 V 中概括出包含 m 个中心节点的中心节点集 P。

[0055] 如图 7 所示，首先在步骤 705，根据节点的中心度，从上述主信息层的数据集 V 中选择出一个最重要的节点 p_1，将其移动到中心节点集 P 中。

[0056] 接着，在步骤 710，对于中心节点集 P 中的中心节点 p_i，计算其与当前主信息层的数据集 V 中的各个节点的最短距离向量 $d_i[1,...n]$，其中 n 是当前主信息层的数据集 V 中的节点数量。

[0057] 在此，在各个中心节点 p_i 的最短距离向量 $d_i[1,...n]$ 中，分别保存了该中心节点 p_i 到数据集 V 中的各个节点的最短距离，即 $d_i[1]$ 保存了 p_i 到数据集 V 中的第 1 个节点的最短距离，$d_i[2]$ 保存了 p_i 到数据集 V 中的第 2 个节点的最短距离，等等。

[0058] 在步骤 715，在中心节点集 P 中的所有中心节点相互之间，进行最短距离向量的比较，以从当前主信息层的数据集 V 中选择出一个节点，将其从 V 移动到 P 中，该选择的节点到中心节点集 P 中的中心节点的最短距离大于数据集 V 中的其他节点。

[0059] 具体而言，首先针对中心节点集 P 中的各个中心节点 p_i，根据其最短距离向量 $d_i[1,...n]$，在数据集 V 中确定一个距离该中心节点 p_i 最远的节点 x，即与 p_i 的最短距离 $d_i[x]$ 最大的节点，进而在各个中心节点 p_i 的最远节点 x 相互之间，进行最短距离 $d_i[x]$ 的比较，从而最终确定出一个最短距离 $d_i[x]$ 最大的节点 x，将其从数据集 V 移动到 P 中。

[0060] 例如，假设中心节点集 P 中存在 a 和 b 两个节点，则首先根据节点 a、b 的最短距离向量，在数据集 V 中为节点 a 确定一个最远的节点 a1，为节点 b 确定一个最远的节点 b1，然后对节点 a、a1 之间的距离与节点 b、b1 之间的距离进行比较，选择其中较大的距离所对应的那个节点 (a1 或 b1)，将其从数据集 V 移动到中心节点集 P 中。

[0061] 在步骤 720，判断中心节点集 P 中的中心节点数是否达到 m，如果是，则该过程结束，否则返回到步骤 710。

[0062] 以上图 7 的过程就是对图 6 中的步骤 605 的进一步详细化。

[0063] 接着，返回到图 6，在步骤 610，根据主信息层的原始数据集，为中心节点集中的各个中心节点进行链接概括，以获得包含中心节点之间的链接关系的中心节点链接集。

[0064] 由于通过步骤 605 中的节点的概括，使中心节点集中的中心节点作为与其相关的周围节点的代表而被选择出，所以也应该将这些相关的周围节点之间的链接概括并绑定到其相应的中心节点上。

[0065] 具体地，在本步骤中，对于中心节点集中的任意两个中心节点 p_1 和 p_2，利用广度优先搜索 (Breadth-First-Search，BFS) 算法在上述主信息层的原始数据集中寻找所有连接这两个中心节点的路径、即边，并且对这些边中长度小于预定的最大长度 λ 的边进行加权合并，作为直接连接中心节点集中的这两个中心节点 p_1 和 p_2 的边，添加到中心节点链接集中。例如，假设中心节点 p1、p2 之间有 10 条边 e1，e2，...en，每条边的权值为 w1，w2，...wn，则利用一条权值为 w1+w2...+wn 的边 e 来代替这 10 条边，将该边 e 添加到中心节点链接集

中，同时将上述 10 条边 e1，e2，...en 从主信息层的原始数据集中删除。

[0066] 并且，在获得了中心节点链接集之后，该中心节点链接集与上述的中心节点集一起构成了样本数据集。

[0067] 以上图 6 的过程就是对图 5 中的步骤 510 的进一步详细化。

[0068] 接着，返回到图 5，在步骤 515，以上述概括出的样本数据集为布局样本，为上述主信息层生成基于密度的等高线图。

[0069] 关于该步骤，下面结合图 8 进行详细描述。

[0070] 如图 8 所示，首先在步骤 805，计算生成等高线图所需的高度矩阵的维数。

[0071] 高度矩阵是任何一种等高线生成算法都需要的输入。为了生成 N×N 维高度矩阵，在本步骤中，根据屏幕的尺寸，基于下式 (1) 来计算高度矩阵的维数 N：

[0072] $$N = \frac{\sqrt{width * height}}{ratio} \qquad (1)$$

[0073] 其中，width 和 height 分别是屏幕的宽度和高度，ratio 是常量。

[0074] 考虑到高度矩阵的维数 N 越大，所生成的等高线越平滑，但所花费的计算时间也越多这一事实，根据本发明的发明人的经验，将上面的常量 ratio 设置为 10 是适宜的。

[0075] 接着，在步骤 810，将上述样本数据集中的各个中心节点布局到屏幕上。也就是说，根据样本数据集中所包含的中心节点和中心节点之间的链接关系，确定各个中心节点在屏幕上的布局。

[0076] 在该步骤中，可以采用本领域中任何一种已有的布局方法将上述样本数据集中的中心节点布局到屏幕上。

[0077] 在步骤 815，为上述样本数据集中的各个中心节点，以其周围未被选择到上述样本数据集中的节点的数量作为该中心节点的质量，计算该中心节点的密度分布。

[0078] 由于样本数据集中的各个中心节点是从原始的主信息层的数据集中、作为其周围节点的代表被概括出来的，所以在本步骤中，将围绕着中心节点的周围节点的数量作为中心节点的质量，计算出中心节点的密度分布，以便将周围节点体现在中心节点的密度分布中。

[0079] 具体地，将主信息层的数据集中未被选择到样本数据集中的各个节点分别指派给距离该节点最近的中心节点，在此，假设主信息层的数据集中指派给中心节点 i 的未选择节点的数量为 m_i，则利用下式 (2) 来计算中心节点 i 的密度分布 f(x)：

[0080] $$f(x) = \frac{1}{n}\sum_{i=1}^{m}\frac{m_i}{h}K\left(\frac{x - X_i}{h}\right) \qquad (2)$$

[0081] 其中，x 表示屏幕上的某个位置的二维坐标，X_i 表示中心节点 i 在屏幕上的二维坐标，n 是原始的主信息层中的总节点数，m 是样本数据集中的中心节点数，h 是带宽，K() 是核函数。

[0082] 对于上式 (2) 中的核函数 K()，可以使用本领域中已有的分布函数，例如具有 0 平均数和最小的整数变量的高斯分布函数，即 N(0,1)。

[0083] 此外，上式 (2) 中的带宽 h，是用于控制所获得的密度分布 f(x) 的平滑程度的常量。h 越小，所得到的分布 f(x) 越将出现窄而陡峭的波峰，h 越大，f(x) 的分布越均匀及平滑。对于带宽 h，可以通过交叉验证来得到。

[0084] 在此，在优选实施例中，根据下式 (3) 所示的评估器，通过弃一法交叉验证来评估出带宽 h 的最佳值：

[0085] $$\hat{J}(h) \approx \frac{1}{hn^2}\sum_i\sum_j K^*\left(\frac{X_i - X_j}{h}\right) + \frac{2}{hn}K(0) \qquad (3)$$

[0086] 其中，$K^*(x) = K^{(2)}(x) - 2K(x)$，$K^{(2)}(x) = \int K(x-y)K(y)dy$，K(x) 为高斯分布函数 N(0,1)，$K^{(2)}(x)$ 为高斯分布函数 N(0,2)。也就是说，根据上式 (3)，利用弃一法交叉验证获得使$\hat{J}(h)$最小的 h 值，作为上述带宽常量 h。

[0087] 为了便于理解，下面说明用于评估出最佳带宽 h 的上式 (3) 的推导过程。

[0088] 首先，定义密度分布 f(x) 与其评估器$\hat{f}(x)$之间的损失函数如下：

[0089] $$L(h) = \int\left(f(x) - \hat{f}(x)\right)^2 dx = \int f^2(x)dx + \int \hat{f}^2(x)dx - 2\int f(x)\hat{f}(x)dx \qquad (4)$$

[0090] 其中，评估器$\hat{f}(x)$是正态分布，其定义为：

[0091] $$\frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}}$$

[0092] 也就是说，正态分布密度函数$\hat{f}(x)$是本发明中的密度分布 f(x) 的理想目标，因为本发明中的密度分布 f(x) 是基于所概括出的中心节点并且在考虑了中心节点的周围节点的情况下而获得的，所以其并不满足正态分布，但是因为正态分布密度函数曲线的形状是完美的发散对称形的，是可视化的追求目标，所以应该使本发明中的密度分布 f(x) 尽可能逼近于正态分布。

[0093] 从而，评估出带宽 h 的最佳值也就是评估出使本发明中的密度分布 f(x) 尽可能逼近于正态分布密度函数$\hat{f}(x)$的最佳 h 值。

[0094] 上述损失函数通过求取$\left(f(x) - \hat{f}(x)\right)$的、关于带宽 h 的一阶导数，来获得使$\left(f(x) - \hat{f}(x)\right)$最小化的最佳 h 值。在上式 (4) 中，由于右侧的第一项 $\int f^2(x)dx$ 与带宽 h 无关，所以可以不考虑这一项，从而简化得到下式 (5)，以通过最小化下式 (5) 来评估出最佳带宽 h。

[0095] $$J(h) = \int \hat{f}^2(x)dx - 2\int f(x)\hat{f}(x)dx \qquad (5)$$

[0096] 进而，在上式 (5) 的基础上，为了加快求解速度，将积分离散化，从而得到上式 (3) 的评估器来评估出最佳带宽 h。

[0097] 接着，在步骤 820，对上述样本数据集中的各个中心节点的密度分布进行合成，以生成高度矩阵的每一坐标处的合成密度分布，结合在高度矩阵中。

[0098] 具体地，在本步骤中，根据下式 (6) 来进行各个中心节点的密度分布的合成：

[0099] $$f(x) = \sum_{G\text{中的所有}P_s} f_s(x) \qquad (6)$$

[0100] 其中，G 表示样本数据集，p_s 表示样本数据集 G 中的某个中心节点，$f_s(x)$ 是中心节点 p_s 的密度分布。

[0101] 也就是说，在上式 (6) 中，对于屏幕上的某个位置 x(x 表示该位置的二维坐标)，由于各个中心节点在该位置处都可能有密度分布，所以通过合并各个中心节点在该位置处的密度分布来得到该位置处的合成密度分布。

[0102] 此外，关于合成密度分布与高度矩阵的结合进行描述。如本领域技术人员所公知

的，高度矩阵是由多个具有二维坐标的小格子构成的，所以合成密度分布与高度矩阵的结合，就是指将高度矩阵中的各个小格子所具有的二维坐标代入上式 (6) 中，从而为各个小格子计算出其二维坐标的位置处的合成密度分布值，并存储到相应的小格子中。

[0103] 接着，在步骤 825，利用所生成的高度矩阵，为上述样本数据集中已经布局在了屏幕上的各个中心节点生成等高线，并填充颜色，以为主信息层生成基于密度的等高线图。

[0104] 考虑到所生成的等高线要体现出各个中心节点的基于密度的形状，所以在本步骤中，优选使用在 http://members.bellatlantic.net/ ～ vze2vrva/thesis.html 处公开的已有跟踪算法来生成等高线。

[0105] 在步骤 830，将主信息层中与上述中心节点相对应的背景信息布局到上述基于密度的等高线图上。

[0106] 在该步骤中，优选采用力矢量布局模型 (force directed model)，来将相应的背景信息布局到上述基于密度的等高线图上，同时避免背景信息与中心节点的重叠。

[0107] 返回到图 5，还包括可选的步骤 520。在可选的步骤 520，通过在上述主信息层的基于密度的等高线图上改变相应的背景信息，为上述多个信息层中上述主信息层之外的至少一个信息层生成基于密度的等高线图。

[0108] 也就是说，在该可选的步骤 520，将非主信息层中与上述主信息层中的中心节点的背景信息对应的信息布局到非主信息层的等高线图上，其中非主信息层的等高线图与主信息层的等高线图一致。

[0109] 在此，由于上述多个信息层都是用于表示同一网络的，并且其中的主信息层是能够生成良好布局的视图的信息层，所以可以考虑将根据主信息层的数据集生成的基于密度的等高线图的布局直接用于其他的信息层，这样既能够保证各个信息层的视图对网络概况的忠实呈现，又能够保证各个信息层的视图都成为良好布局的视图。这样，对于不同的信息层而言，由于其呈递的网络的背景信息是不同的，所以只需要在主信息层的基于密度的等高线图上改变相应的背景信息即可。

[0110] 但是，各个信息层的节点以及背景信息并不是一一对应的，例如在与论文有关的网络的情况下，基于会议维度的信息层中的一个节点（一个背景信息）可能对应着论文信息维度的信息层中的多个节点（多个背景信息），这样，就需要以主信息层为基准，进行背景信息的对应和提取。本领域技术人员可以理解，可以采用本领域中的多种方法来实现多个信息层之间的背景信息的对应和提取，例如当把会议维度作为主信息层，把作者维度作为非主信息层时，主信息层中的一个节点，即一个会议可能与多个会议论文的作者相对应，这时限于布局的要求可能需要从众多的作者中选择出有代表性的作者布局到非主信息层中。选择的策略可以有很多种方式，包括按照作者的出现频率；按照作者的重要程度，如被引用的次数；按照作者的顺序，如是否是第一作者等。

[0111] 以上就是对本实施例的数据集的可视化方法的详细描述。在本实施例中，通过分别从网络的不同信息维度呈现网络的不同概况来可视化网络，在确保向网络分析人员呈递网络的全面信息的同时，能够防止呈现内容的失真以及视觉混乱，使网络的分析人员能够根据自身的需要从不同的角度清晰地获得关于网络的不同信息，从而获得网络分析的极大便利。此外，在本实施例中，对于网络的基于不同信息维度的各个信息层，通过基于从主信息层的数据集中概括出的少量重要节点，生成基于密度分布的等高线图，能够在不失真的

情况下极大地简化所生成的视图的整体布局。

[0112] 在同一发明构思下，本发明提供一种网络的可视化系统。下面结合附图对其进行描述。

[0113] 图 9 是根据本发明实施例的网络的可视化系统的方框图。如图 9 所示，本实施例的网络的可视化系统 90 包括：分层单元 91、可视化单元 92。

[0114] 具体地，分层单元 91 将网络的数据集基于不同信息维度划分为多个信息层。其中，每一信息维度的信息层是由网络的数据集中该信息维度的数据组成的。

[0115] 可视化单元 92 分别对基于不同信息维度的上述多个信息层进行可视化处理，以用于呈现该多个信息层的各自的视图。在一个实施例中，可视化单元 92 利用透明色处理技术对该多个信息层的各自的视图进行处理，以将其组合为一个视图，并且使得上述多个信息层的各自的视图之间能够进行切换。

[0116] 如图 9 所示，可视化单元 92 可进一步包括：主信息层选择单元 921、数据集概括单元 922 和视图生成单元 923。

[0117] 主信息层选择单元 921 从上述基于不同信息维度的多个信息层中，选择能够生成良好布局的视图的信息层，作为主信息层。

[0118] 数据集概括单元 922 对上述主信息层所包含的数据集进行概括，以构成包含中心节点及其之间的链接关系的样本数据集。

[0119] 如图 9 所示，数据集概括单元 922 可进一步包括：节点概括单元 9221 和链接概括单元 9222。

[0120] 节点概括单元 9221 对上述主信息层的数据集进行节点概括，以获得包含多个中心节点的中心节点集。具体地，节点概括单元 9221 从上述主信息层的数据集中，根据节点的中心度，选择出一个最重要的节点，将其移动到中心节点集中，并且依次进行以下处理，直到中心节点集中的中心节点数达到预定的值：对于中心节点集中的各个中心节点，计算其与上述主信息层的数据集中未被选择到中心节点集中的节点之间的最短距离向量；以及从主信息层的数据集中未被选择到中心节点集中的节点中选择出一个与中心节点的最短距离是最短的这样的节点，移动到中心节点集中。

[0121] 链接概括单元 9222 根据上述主信息层的数据集，为上述中心节点集中的中心节点进行链接概括，以获得包含中心节点之间的链接关系的中心节点链接集。具体地，链接概括单元 9222 对于上述中心节点集中的任意两个中心节点：利用广度优先搜索算法在主信息层的数据集中寻找所有连接这两个中心节点的路径；以及对上述路径中长度小于预定的最大长度的路径进行加权合并，作为直接连接上述任意两个中心节点的链接，添加到中心节点链接集中。

[0122] 接着，视图生成单元 923 以上述样本数据集为布局样本，为上述主信息层生成基于密度的等高线图。

[0123] 如图 9 所示，视图生成单元 923 可进一步包括：节点布局单元 9231、高度矩阵生成单元 9232、密度分布计算单元 9233、密度分布结合单元 9234、等高线生成单元 9235、信息布局单元 9236。

[0124] 节点布局单元 9231 将上述样本数据集中的各个中心节点布局到屏幕上。

[0125] 高度矩阵生成单元 9232 生成在等高线的生成中所用的高度矩阵。

[0126] 密度分布计算单元 9233 为上述样本数据集中的、被布局到了屏幕上的各个中心节点，以其周围的未被选择到上述样本数据集中的节点的数量作为该中心节点的质量，计算该中心节点的密度分布。具体地，密度分布计算单元 9233 对于上述样本数据集中的、被布局到了屏幕上的各个中心节点，根据上式 (2) 计算密度分布。

[0127] 密度分布结合单元 9234 将密度分布计算单元 9233 所计算出的各个中心节点的密度分布结合到用于生成等高线的高度矩阵中。具体地，密度分布结合单元 9234 根据上式 (6) 对上述样本数据集中的各个中心节点的密度分布进行合成，以生成高度矩阵的每一坐标处的合成密度分布。

[0128] 等高线生成单元 9235 利用上述高度矩阵，为上述被布局到了屏幕上的各个中心节点生成等高线并填充颜色，以为上述主信息层生成基于密度的等高线图。

[0129] 信息布局单元 9236 将上述主信息层中与上述各个中心节点相对应的背景信息布局到上述基于密度的等高线图上。

[0130] 视图生成单元 923，还通过在主信息层的基于密度的等高线图上改变相应的背景信息，为上述多个信息层中上述主信息层之外的至少一个非主信息层生成基于密度的等高线图。具体而言，上述视图生成单元 923 将非主信息层中与主信息层中的中心节点的背景信息对应的信息布局到非主信息层的等高线图上，其中非主信息层的等高线图与主信息层的等高线图一致。

[0131] 以上就是对本发明的网络的可视化系统的详细描述。其中，该系统及其各个组成部分，可以由专用的电路或芯片构成，也可以通过计算机（处理器）执行相应的程序来实现。

[0132] 以上虽然通过一些示例性的实施例对本发明的数据集的可视化方法和系统进行了详细的描述，但是以上这些实施例并不是穷举的，本领域技术人员可以在本发明的精神和范围内实现各种变化和修改。因此，本发明并不限于这些实施例，本发明的范围仅以所附权利要求为准。

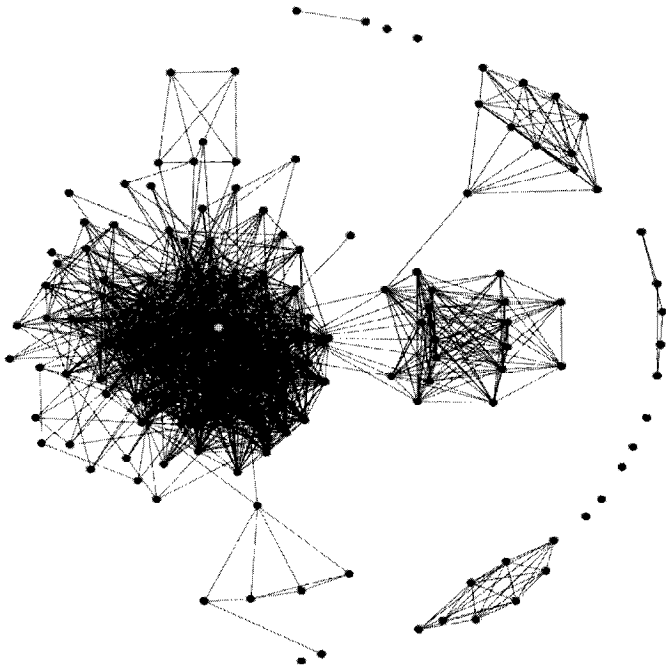


图 1(a)

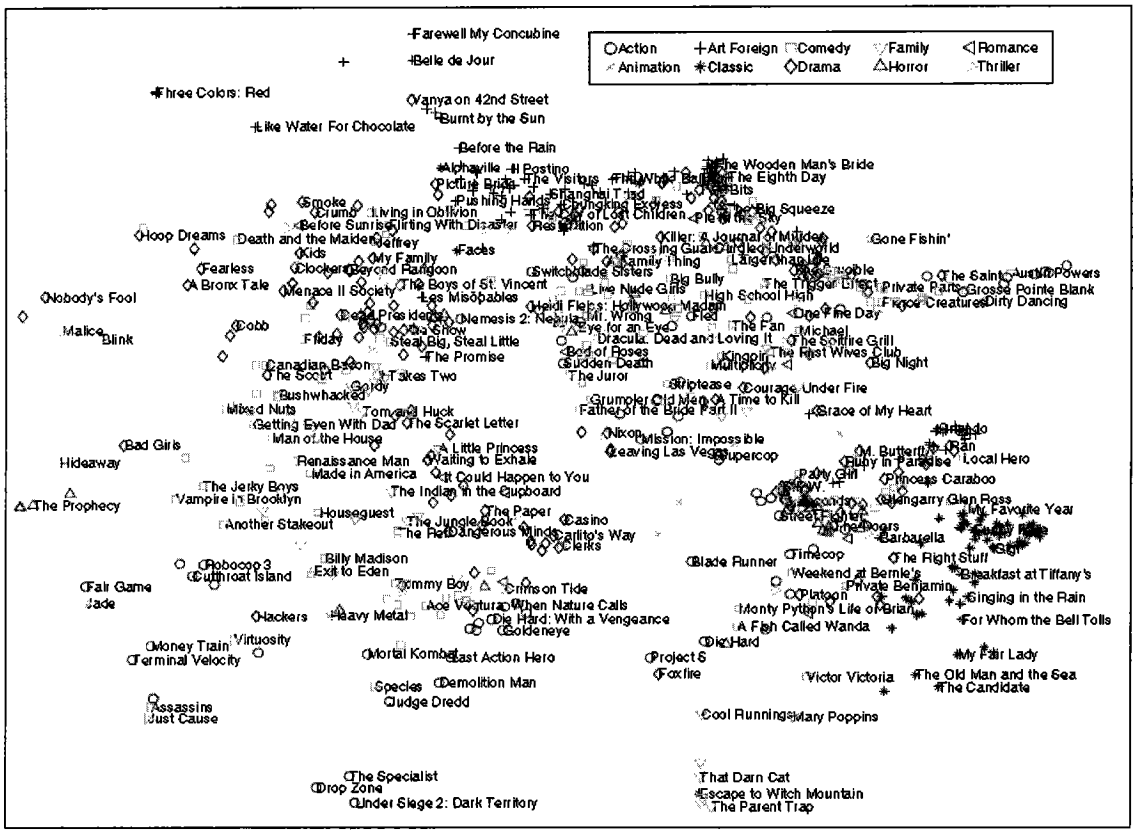

Action
Animation
Art Foreign
Classic
Comedy
Drama
Family
Horror
Romance
Thriller
Farewell My Concubine
Belle de Jour
Three Colors: Red
Vanya on 42nd Street
Like Water For Chocolate
Burnt by the Sun
Before the Rain
Alphaville
Postino
The Visitors
The Wooden Man's Bride
The Eighth Day
Picture Bride
Shanghai Triad
Pushing Hands
The Howling Fortress
The Big Squeeze
Smoke
Crumb
Living in Oblivion
The City of Lost Children
Before Sunrise
Flirting With Disaster
Hoop Dreams
Death and the Maiden
Killer: A Journal of Murder
Gone Fishin'
Kids
My Family
Faces
The Crossing Guard
Family Thing
Fearless
Clockers
Beyond Rangoon
Switchblade Sisters
Big Bully
The Saint
Nobody's Fool
A Bronx Tale
Menace II Society
The Boys of St. Vincent
Live Nude Girls
High School High
Private Parts
Grosse Pointe Blank
Les Misérables
Heidi Fleiss: Hollywood Madam
Dirty Dancing
Malice
Blink
Dobb
Nemesis 2: Nebula
Mr. Wrong
Eye for an Eye
The Fan
Michael
Friday
Steal Big, Steal Little
Dracula: Dead and Loving It
The Sunset Grill
The Promise
Bed of Roses
Kingpin
The Rest Wives Club
Canadian Bacon
Sudden Death
Multiplicity
Big Night
The Scout
Takes Two
The Juror
Striptease
Courage Under Fire
Bushwhacked
Grumpier Old Men
Time to Kill
Mixed Nuts
Tom and Huck
Father of the Bride Part II
Grace of My Heart
Getting Even With Dad
The Scarlet Letter
Nixon
Man of the House
Mission: Impossible
Bad Girls
Little Princess
Leaving Las Vegas
Supercop
Hideaway
Renaissance Man
Waiting to Exhale
Ruby in Paradise
Local Hero
Made in America
It Could Happen to You
M. Butterfly
The Jerky Boys
The Indian in the Cupboard
Princess Caraboo
Vampire in Brooklyn
The Prophecy
Houseguest
The Paper
Casino
Another Stakeout
The Jungle Book
Dangerous Minds
Carlito's Way
Clerks
Barbarella
Billy Madison
Blade Runner
Timecop
Robocop 3
Exit to Eden
The Right Stuff
Cutthroat Island
Weekend at Bernie's
Breakfast at Tiffany's
Fair Game
Jade
Tommy Boy
Crimson Tide
Private Benjamin
Platoon
Singing in the Rain
Ace Ventura
When Nature Calls
Monty Python's Life of Brian
Hackers
Heavy Metal
Die Hard: With a Vengeance
A Fish Called Wanda
For Whom the Bell Tolls
Goldeneye
Money Train
Virtuosity
Die Hard
Terminal Velocity
Mortal Kombat
Last Action Hero
Project S
My Fair Lady
Foxfire
Victor Victoria
The Old Man and the Sea
Species
Demolition Man
The Candidate
Judge Dredd
Assassins
Just Cause
Cool Running
Mary Poppins
The Specialist
That Darn Cat
Drop Zone
Under Siege 2: Dark Territory
Escape to Witch Mountain
The Parent Trap

图 1(b)

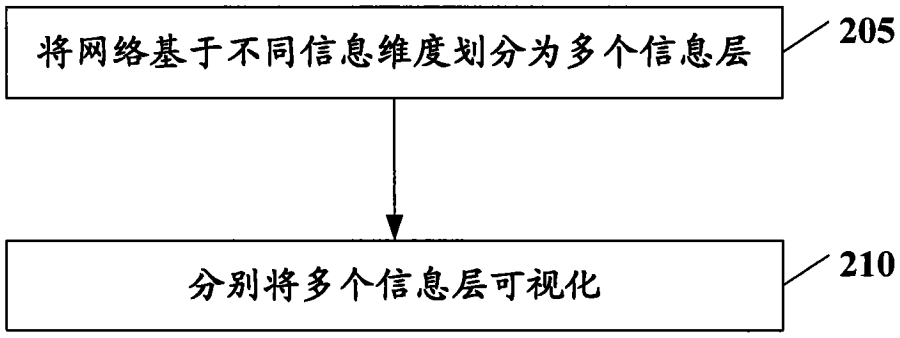

将网络基于不同信息维度划分为多个信息层
205
分别将多个信息层可视化
210

图 2

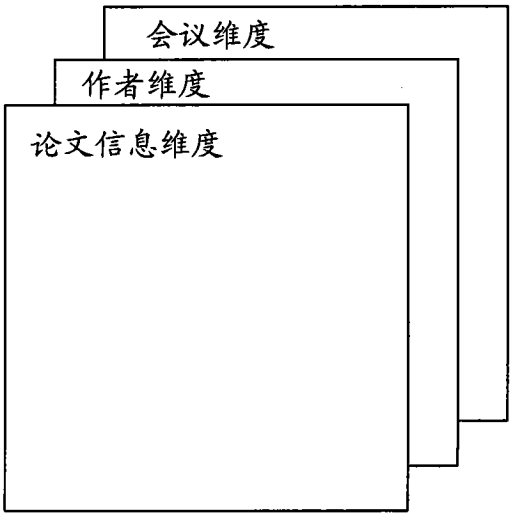

会议维度
作者维度
论文信息维度

图 3(a)

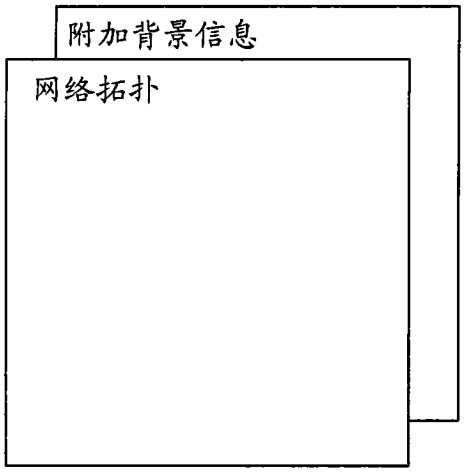

附加背景信息
网络拓扑

图 3(b)

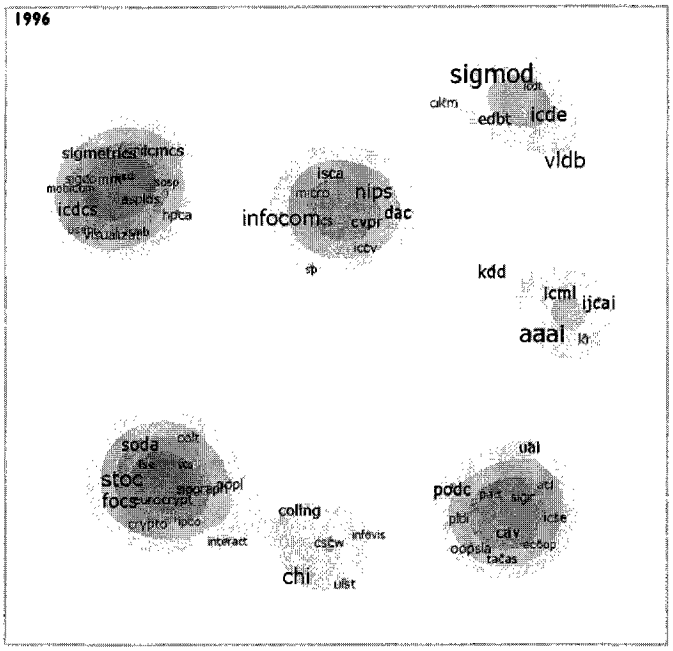

1996
sigmod
cikm
edbt
icde
vldb
sigmetrics
icmcs
sigcomm
sosp
mobicom
asplos
icdcs
hpca
isca
micro
nips
infocom
cvpr
dac
iccv
sp
kdd
icml
ijcai
aaai
soda
soft
stoc
focs
sigcse
ppl
crypto
ipco
interact
coling
infovis
cscw
chi
uist
uai
podc
pldi
icse
caiv
oopsla
ecoop
tacas

图 4(a)

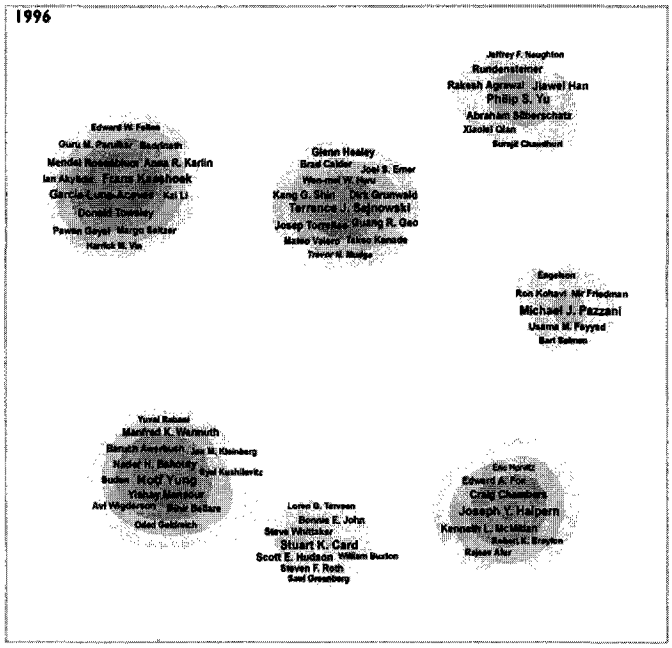

1996
Jeffrey F. Naughton
Rundensteiner
Rakesh Agrawal
Jiawei Han
Philip S. Yu
Abraham Silberschatz
Xiaolei Qian
Sunjit Chaudhuri
Edward W. Felten
Guru M. Parulkar
Sanjeeb
Mendel Rosenblum
Anna R. Karlin
Ian Akyildiz
Frans Kaashoek
Garcia-Luna-Aceves
Kai Li
Donald Sterbenz
Pawan Goyal
Margo Seltzer
Hariluk M. Vin
Glenn Healey
Brad Calder
Joel S. Emer
Wen-mei W. Hwu
Kang G. Shin
Dirk Grunwald
Terrance J. Sejnowski
Josep Torrellas
Guang R. Gao
Mateo Valero
Takao Kanade
Trevor N. Mudge
Engelson
Ron Kohavi
Mir Friedman
Michael J. Pazzani
Usama M. Fayyad
Bart Selman
Yuval Rabani
Manfred K. Warmuth
Baruch Awerbuch
Jon M. Kleinberg
Nathan H. Bshouty
Ravi Kashiberitz
Boston
Moti Yung
Yishay Mansour
Avi Wigderson
Oded Goldreich
Leno G. Tervyan
Bonnie E. John
Steve Whittaker
Stuart K. Card
Scott E. Hudson
William Buxton
Steven F. Roth
Saul Greenberg
Eric Horvitz
Edward A. Fox
Craig Chambers
Joseph Y. Halpern
Kenneth L. McMillan
Robert E. Brayton
Rajeev Alur

图 4(b)

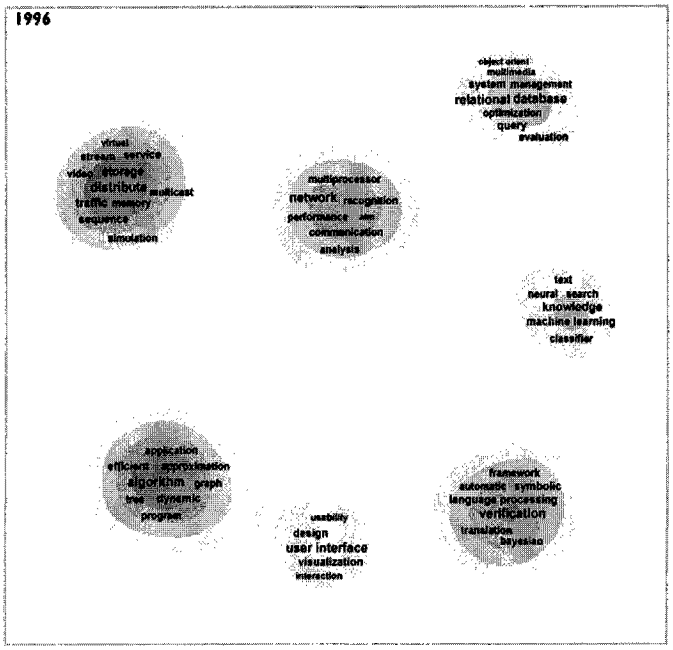

1996
object orient
multimedia
system management
relational database
optimization
query
evaluation
virtual
stream service
video storage
distribute multicast
traffic memory
sequence
simulation
multiprocessor
network recognition
performance
communication
analysis
text
neural search
knowledge
machine learning
classifier
application
efficient approximation
algorithm graph
tree dynamic
program
usability
design
user interface
visualization
interaction
framework
automatic symbolic
language processing
verification
translation
bayesian

图 4(c)

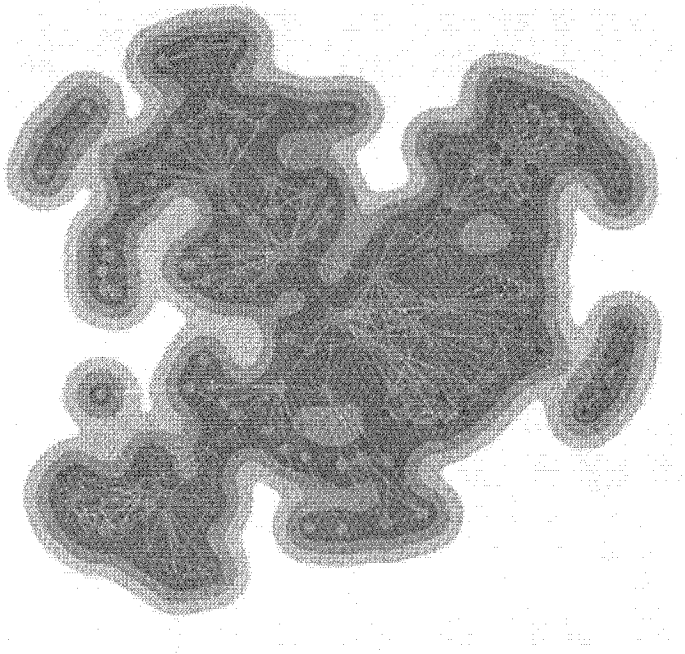


图 4(d)

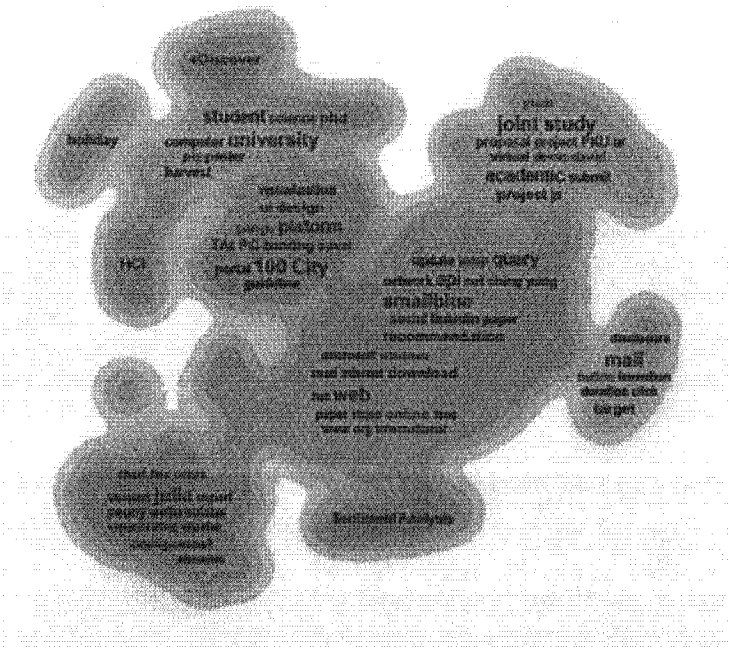

student
computer university
joint study
mail

图 4(e)

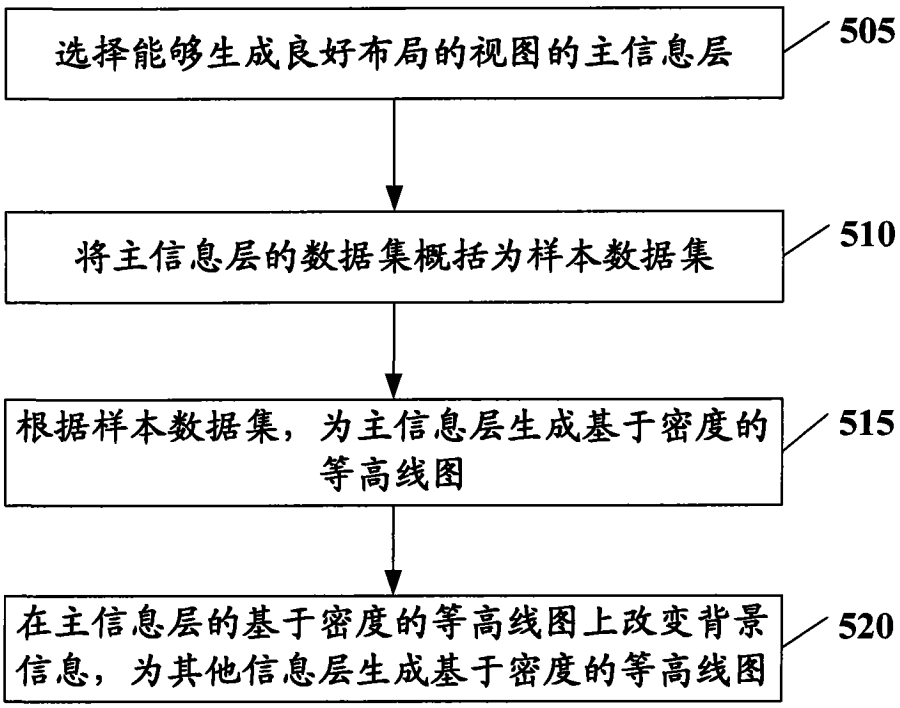

选择能够生成良好布局的视图的主信息层
505
将主信息层的数据集概括为样本数据集
510
根据样本数据集，为主信息层生成基于密度的等高线图
515
在主信息层的基于密度的等高线图上改变背景信息，为其他信息层生成基于密度的等高线图
520

图 5

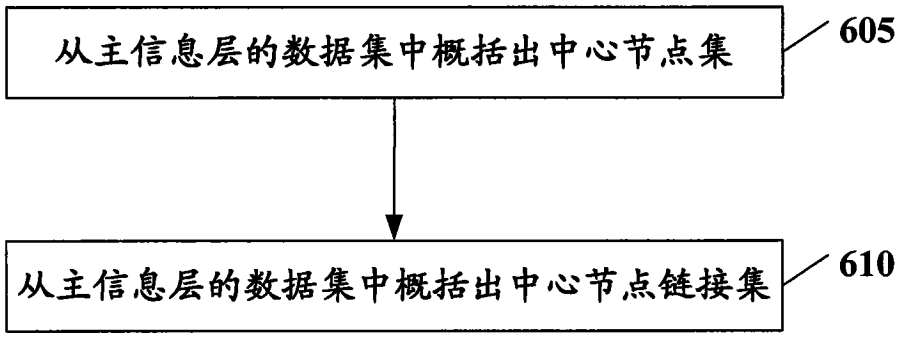

从主信息层的数据集中概括出中心节点集
605
从主信息层的数据集中概括出中心节点链接集
610

图 6

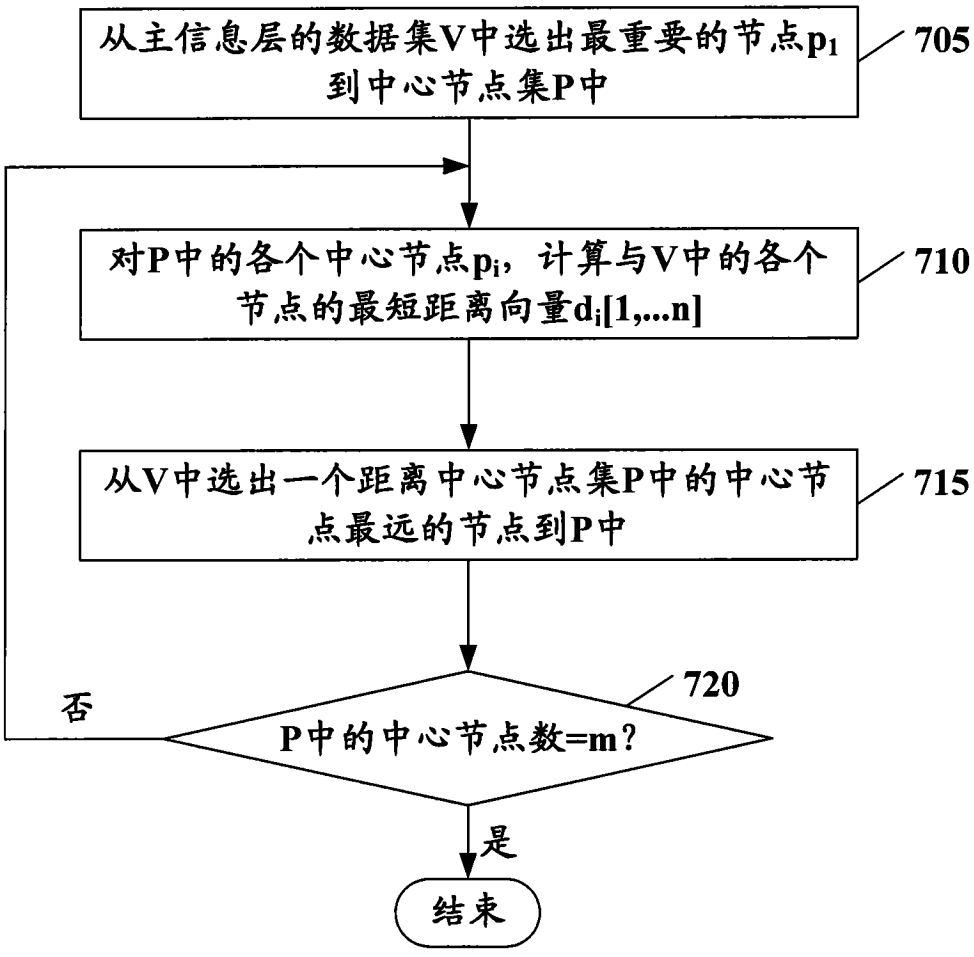

从主信息层的数据集V中选出最重要的节点p_1到中心节点集P中
705
对P中的各个中心节点p_i，计算与V中的各个节点的最短距离向量$d_i[1,...n]$
710
从V中选出一个距离中心节点集P中的中心节点最远的节点到P中
715
P中的中心节点数=m?
720
否
是
结束

图 7

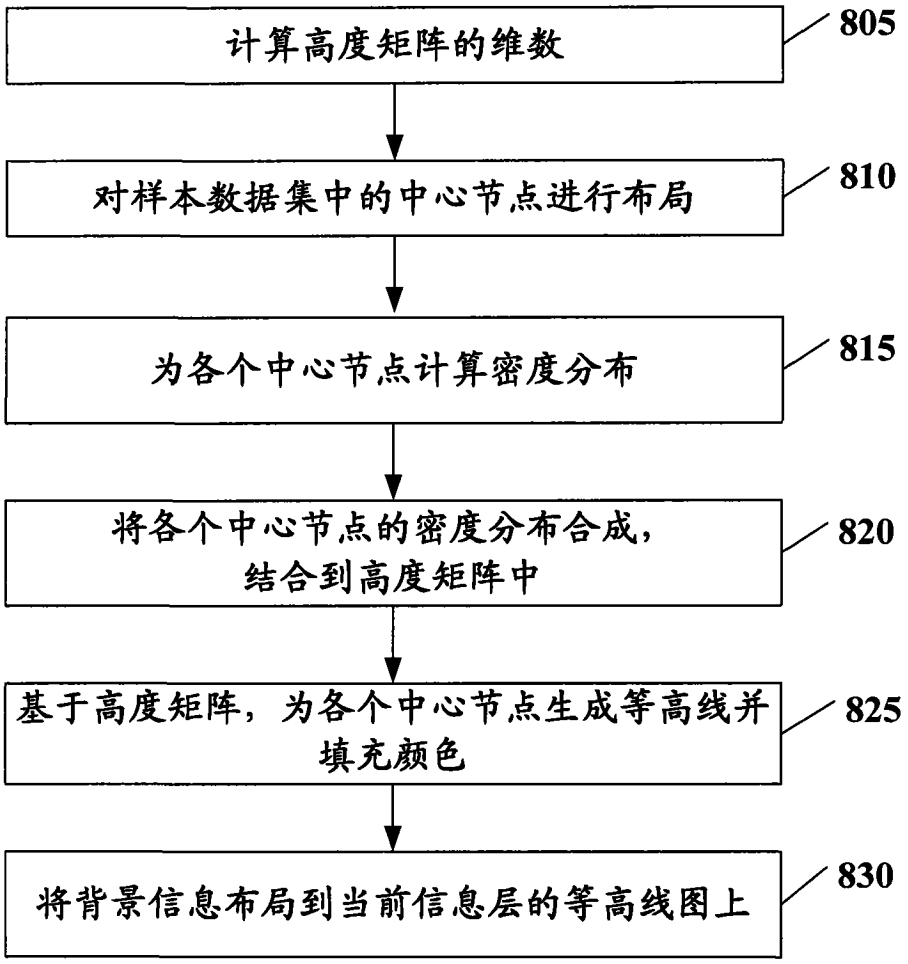

计算高度矩阵的维数
805
对样本数据集中的中心节点进行布局
810
为各个中心节点计算密度分布
815
将各个中心节点的密度分布合成，结合到高度矩阵中
820
基于高度矩阵，为各个中心节点生成等高线并填充颜色
825
将背景信息布局到当前信息层的等高线图上
830

图 8

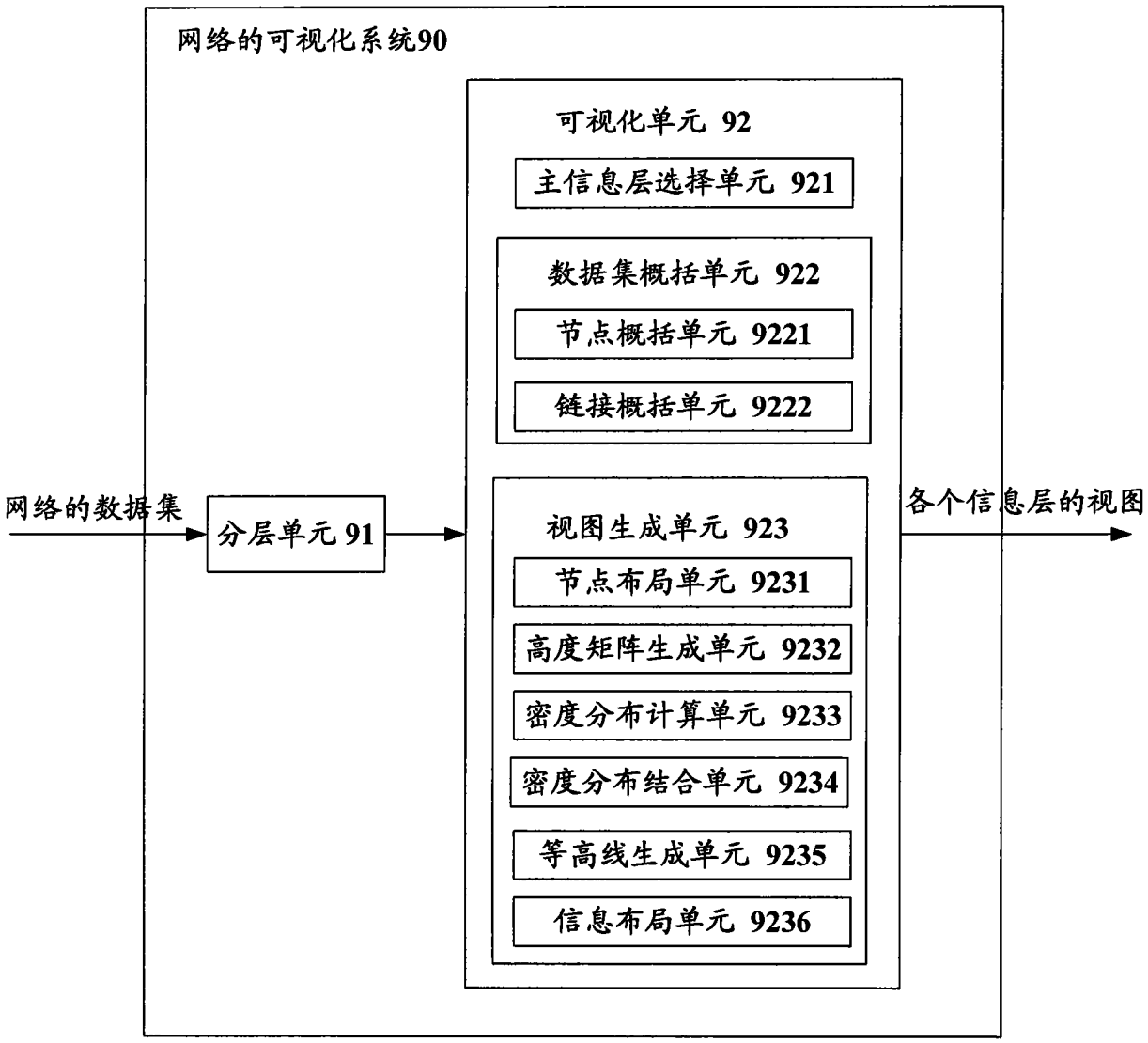

网络的可视化系统90
可视化单元 92
主信息层选择单元 921
数据集概括单元 922
节点概括单元 9221
链接概括单元 9222
网络的数据集
分层单元 91
视图生成单元 923
节点布局单元 9231
高度矩阵生成单元 9232
密度分布计算单元 9233
密度分布结合单元 9234
等高线生成单元 9235
信息布局单元 9236
各个信息层的视图

图 9