

7-15-2020

## A Survey on Graph Representation and Visualization Techniques

Chen Yi

*Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;*

Menglu Zhang

*Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;*

Yuchai Wan

*Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

# A Survey on Graph Representation and Visualization Techniques

## Abstract

**Abstract:** A graph consists of nodes and edges, and represents the relationship between two or more entities. Graph-based analysis can help understand the structure and nature of the entity relationships and reveal the implicit relationships in graphs. The representation and visualization methods of graphs play an important role in the graph analysis. In graph visualization research, the accuracy of knowledge transfer and people's mental map, etc. should be considered first, then the graphs beauty, the time requirement, and the computer performance. *The graph representation methods, graph layout algorithms, visualization methods based on node-link-graphs, adjacency matrices, and graph embedding are reviewed, summarized and contrasted. The future work and challenges are discussed.*

## Keywords

Graph representation, visualization, node-link-graph, adjacency matrix, graph embedding

## Recommended Citation

Chen Yi, Zhang Menglu, Wan Yuchai. A Survey on Graph Representation and Visualization Techniques[J]. Journal of System Simulation, 2020, 32(7): 1232-1243.

# 图的表示与可视化方法综述

陈谊\*, 张梦录, 万玉钗

(北京工商大学计算机与信息工程学院食品安全大数据技术北京市重点实验室, 北京 100048)

**摘要:** 图是由节点和边组成的图形, 通常用于表示两个或多个实体之间的关系。基于图的分析可以帮助人们理解实体关系的结构和本质, 探索图中的隐含关联。图的表示与可视化方法在图分析中起着重要作用, 在图可视化研究中首先要考虑知识传达是否准确、人们的思维地图等方面, 同时还要考虑图形是否美观、构建图所需的时间、以及计算机的性能等问题。综述了基于节点-链接、邻接矩阵以及图嵌入的图表示方法、图布局算法以及可视化方法, 并对这些方法进行归纳与对比。最后对图表示与可视化技术的未来发展趋势进行了展望。

**关键词:** 图表示; 可视化; 节点-链接图; 邻接矩阵; 图嵌入

中图分类号: TP391

文献标识码: B

文章编号: 1004-731X (2020) 07-1232-12

DOI: 10.16182/j.issn1004731x.joss.19-VR0444

## A Survey on Graph Representation and Visualization Techniques

Chen Yi\*, Zhang Menglu, Wan Yuchai

(Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

**Abstract:** A graph consists of nodes and edges, and represents the relationship between two or more entities. Graph-based analysis can help understand the structure and nature of the entity relationships and reveal the implicit relationships in graphs. The representation and visualization methods of graphs play an important role in the graph analysis. In graph visualization research, the accuracy of knowledge transfer and people's mental map, etc. should be considered first, then the graphs beauty, the time requirement, and the computer performance. *The graph representation methods, graph layout algorithms, visualization methods based on node-link-graphs, adjacency matrices, and graph embedding are reviewed, summarized and contrasted. The future work and challenges are discussed.*

**Keywords:** Graph representation; visualization; node-link-graph; adjacency matrix; graph embedding

## 引言

现实世界中的大多数实体都不是相互独立的, 它们通过不同的链接方式组合在一起形成了图(也称作网络), 通过对图的挖掘与分析, 有助于人们

理解嵌入在抽象网络数据中的复杂关系, 如通信网络、社交网络、文献互引网络、营销网络<sup>[1]</sup>等。近年来更是有许多学者将其应用在食品安全预警和追溯方面<sup>[2]</sup>。

在对图中的信息进行挖掘的过程中都会存在一个问题, 即用户只看到了输入和输出, 这就大大限制了人们对数据的理解程度。通过可视化的方法展示数据分析的整个流程<sup>[3-4]</sup>, 对人们理解数据起到了至关重要的作用<sup>[5]</sup>。



收稿日期: 2019-08-24 修回日期: 2019-10-29;  
基金项目: 国家自然科学基金面上项目(61972010),  
国家重点研发计划(2018YFC1603602), 国家科技基础性工作专项(2015FY111200);  
作者简介: 陈谊(通讯作者 1963-), 女, 北京, 博士, 教授, 研究方向为信息可视化与可视分析。

<http://www.china-simulation.com>

• 1232 •

传统的图布局方法大多是基于节点-链接的形式构建,如使用力导向模型<sup>[6]</sup>(force-directed model)和应力模型<sup>[7]</sup>(stress model)等。它们将图看作是一个力系统或能量系统,通过优化算法使系统中的合力最小或能量最低以达到图布局的目的。其关键任务是节点的布局效果需符合人类的分析习惯和审美观点,如布局的对称性、节点分布的均匀程度、尽量减少边交叉等,以此更好的帮助人们发现数据之间蕴含的关联关系<sup>[8]</sup>。然而当节点之间具有大量的链接时,图的布局就会显得十分的混乱。为了解决节点-链接图无法适应于稠密图布局的情况,研究人员提出了基于邻接矩阵的图可视化方法,它以方阵的形式存储节点之间的链接关系,解决了图型布局混乱和边交叉的问题。但是使用节点-链接图和邻接矩阵在存储图数据时需要保留节点之间的链接信息,不便于计算,为了解决这个问题研究人员提出了将图中节点以向量形式表达的方法,即图嵌入方法。该方法便于对大图<sup>[9]</sup>进行处理,更好的支持对大图的可视分析<sup>[10]</sup>。

本文归纳总结了现有图的表示与可视化方法,包括:节点-链接图、邻接矩阵、以及刚刚兴起的基于图嵌入的方法,对这些技术进行了分类和对比。

## 1 基于节点-链接的图可视化方法

用节点表示对象,用边表示关系的节点-链接布局是最自然的图可视化布局表达。它容易被用户理解,帮助人们快速建立事物之间的联系,显式地表达事物之间的关系<sup>[11]</sup>。

### 1.1 基于节点-链接的图表示方法

节点-链接的图通常被表述为二元组形式,即  $G=(V,E)$ , 其中  $V$  为顶点集合,  $E$  为顶点间所有的边集合。表达具有相同节点和关系的图(同构图)有多种布局形式,给人以不同的视觉表达。如图 1 所示,3 种布局方法表达的是同一组关联数据,其二元组形式表示为  $V=\{1, 2, 3, 4, 5\}$ ,  $E=\{(1,2), (1,3),$

$(2,3), (3,4), (4,1), (5,4)\}$ 。节点-链接布局算法要解决的主要问题是在给定关系数据的情况下寻找到特定场景中更好的布局形式。

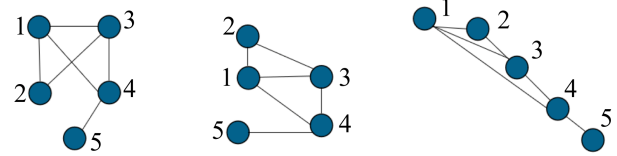


图 1 同构图的 3 种布局方法

Fig. 1 Three layout methods of isomorphism

图 2 为基于节点-链接图的生成过程:

- (1) 将具有关联关系的原始数据集结构化。
- (2) 设置节点初始位置,调整节点位置以趋近理想布局状态,若达到理想布局状态则算法结束,否则继续调整节点位置以完成图布局。
- (3) 用户感知已完成的图布局是否可以达到预期以评估算法的有效性。

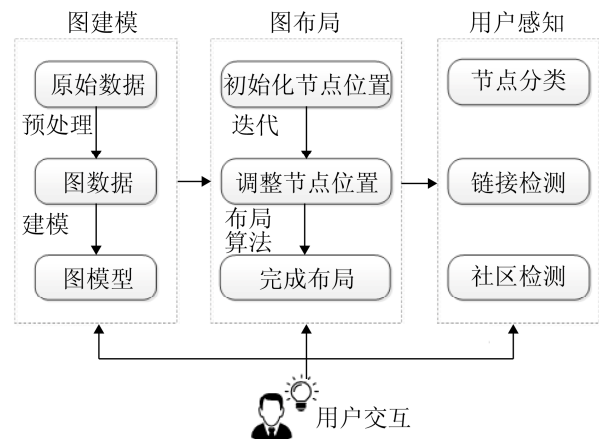


图 2 节点-链接图的生成过程

Fig. 2 Generation process of node-link graph

### 1.2 弹簧电力模型

受胡克定律和库仑定律<sup>[12]</sup>的启发, Peter Eades 提出了力导向布局算法<sup>[6]</sup>。这种算法的基本思想是将图形看作是一个物理系统,各个节点之间互相具有引力和斥力,并凭借力的作用四处移动。直到所有节点间的作用力达到平衡状态,节点停止移动。Feng 等<sup>[13]</sup>通过运用为边增加权重的思想提出了集群图定理,通过改变不同边的权重可以实

现节点-链接图的区域缩放。除此之外，还可以将选择的路径约束为直线、实现符合逻辑的鱼眼镜头<sup>[14]</sup>、将节点聚集在一起、对特定区域的形状进行调整等。最为大众所熟知的是力导向模型是弹簧电子力模型(Spring-Electrical Model)<sup>[15]</sup>也被称为 FR 模型，该模型把每个节点都看做是一个电荷，电荷之间存在斥力，也就是库仑力，将节点之间存在的边模拟为胡克定律中的弹力，利用优化算法使节点间所受到的力趋近于平衡以达到图布局的目的。虽然这种算法的布局时间较长，但是布局过程中节点位置的变化很流畅。为了加速图布局算法，多级力导向布局算法<sup>[16]</sup>被提出，此后的多数算法都参照了上述算法的思想并取得了不错的效果。

### 1.3 应力模型

Kamada 和 Kawai 提出的应力模型<sup>[7]</sup>(stress model)通过最小化节点的理想间距与欧式距离之间差的平方，将图布局转化成一个能量优化问题，其中应力模型与多维尺度分析(Multidimensional scaling, 简称 MDS)中的算法具有相同的公式为：

$$S(\mathbf{X}) = \sum_{i < j} w_{ij} (\|\mathbf{X}_i - \mathbf{X}_j\| - d_{ij})^2 \quad (1)$$

式中： $\mathbf{X}_i$  和  $\mathbf{X}_j$  分别为二维可视化空间中第  $i$  个和第  $j$  个节点； $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_n\}^T$  为所有节点的  $n \times 2$  矩阵； $d_{ij}$  为节点  $\mathbf{X}_i$  与  $\mathbf{X}_j$  之间的目标距离；用  $w_{ij} = d_{ij}^{-2}$  作为归一化常数。通过使应力函数  $S(\mathbf{X})$  最小化，来达到优化可视化布局的目的。由于应力模型和 MDS 具有相同的公式，则可通过多维尺度分析的一些算法对应力模型进行优化以得到近似解，这种方法又被称为应力优化(Stress Majorization)。Marc 等<sup>[17]</sup>对应力优化中所涉及矩阵的线性代数性质进行了实验分析，结果表明，低秩矩阵可以很好地表示图的节点以及边之间的关系，由此提出了一种新的求解应力优化的算法，该算法大大降低了图形绘制所需要的时间。Yu-Jung 等<sup>[18]</sup>提出了一种聚类图布局算法，该算法采用了应力模型来

绘制簇内图，采用弹簧电力模型放置簇，并结合力矩平衡思想实现了图的布局。

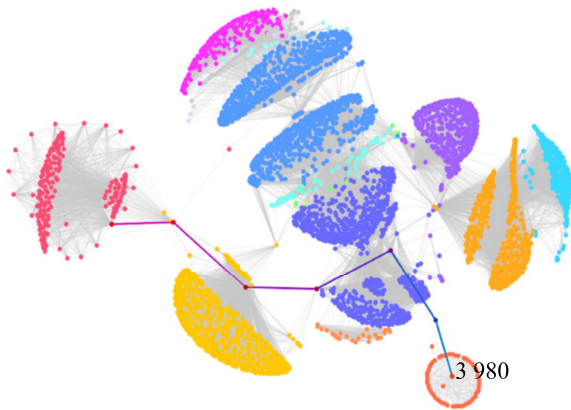
弹簧电力模型以及应力模型在原理上是相似的<sup>[19]</sup>，它们都优化了目标函数<sup>[20]</sup>。Wang 等<sup>[21]</sup>通过将边构造为向量，对边的方向和长度进行约束并将其方法扩展为一个可应用于约束图和非约束图的框架，并采用了 GPU 并行化计算技术绘制图形，可支持绘制上万个节点的约束图。

### 1.4 受约束的图布局算法

为了更好的揭示数据中蕴含的关系、使图的布局更加的清晰美观，关于图绘制技术的约束问题被提了出来。受约束的布局算法包括对节点、边和集群的约束。如给节点添加标签时对节点施加约束预留足够的位置以避免标签重叠；以向量的形式对边施加约束可控制边的方向以减少边交叉；对数据中具有集群属性的子图施加约束、合理的放置其位置以凸显集群内与集群间的关系<sup>[22]</sup>。

在力导向模型的基础上，在绘制图形的过程中对节点施加约束这一思想首先由 He 等<sup>[23]</sup>提出。此后 Dwyer 在受约束的有向图绘制方面取得了很多成果，通过应力优化技术并分别使用正交排序<sup>[24]</sup>、二次规划<sup>[25]</sup>以及提出了一种基于梯度投影的增量算法 IPSEP-COLA<sup>[22]</sup>来约束图的绘制等。最近 Zheng 等<sup>[26]</sup>运用了随机梯度下降(SGD)的方法，通过每次移动一对顶点来最小化其能量函数，该方法对节点的初始化布局要求不高且更容易生成约束布局。

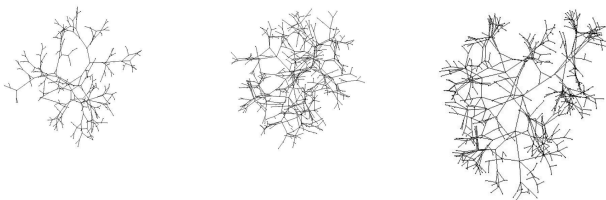
绘制聚类网络可显著提高隐藏在相关抽象关系下的网络语义，使用文本标签注释网络节点可增加网络的可读性，Wu 等<sup>[27]</sup>提出了一种通过引入空间分区技术在美学上保留聚类网络节点周围空间的方法实现了在聚类网络中插入标签的布局。相比于对集群和节点的约束，对边约束的研究比较少，Wang 等<sup>[21]</sup>将边构造为向量以实现对其布局方向的约束，如图 3 所示。

图 3 基于边约束生成的图<sup>[21]</sup>Fig. 3 A graph generated based on edge constraints<sup>[21]</sup>

### 1.5 多层迭代算法

随着数据量不断地增大, 先前被提出的布局方法已经无法满足大规模图形的绘制, 为此多层迭代算法<sup>[28]</sup> (Multilevel Algorithm) 被提出。这种方法从绘制简易图形到逐步增加结点来完成图形的绘制, 常被应用于大图的绘制<sup>[29-30]</sup>。

多层迭代算法可分为 3 个阶段: 粗化、初始化和单层布局(细化)。粗化步骤是对顶点进行聚类以定义一个更小的图(简化图), 并递归地迭代这个过程, 直到图的大小低于阈值, 从而得到一个粗糙的图层次结构,  $G_0, G_1, \dots, G_L$  逐渐简化, 首先绘制最简图  $G_L$ , 之后在  $G_L$  中插入  $G_{L-1}$  中拥有而  $G_L$  中没有的节点, 并不断迭代获得插入节点最佳位置, 即完成  $G_{L-1}$  的绘制。以此类推, 最终完成整个图形的绘制。由简到复杂的绘制过程如图 4 所示。

图 4 节点数为 511、1093、1365 的树<sup>[30]</sup>Fig. 4 Trees with 511, 1093, and 1365 nodes<sup>[30]</sup>

FM<sup>[31]</sup>粗化过程是将具有小直径的子图分割和折叠以获得多级表示。An 等<sup>[32]</sup>基于 k 核法(k-core)实现了层次分明的图布局。最近, 一种具

有共享内存并行化的 Maxent stress<sup>[33]</sup>优化的多级版本被提了出来<sup>[29]</sup>, 他们的粗化方法使用标签传播方法进行图聚类。为了提高图型绘制的速度, 大部分基于 GPU 的解决方案<sup>[34]</sup>被提出。但是目前大多数算法对计算机硬件的要求比较高, 为了解决这一问题, Alessio 等<sup>[16]</sup>提出了一种多级力导向图可视化算法 MULTI-GILA。该方法可以在廉价的云计算平台上实现超大图的高质量布局。

### 1.6 动态图

静态图是指在图形绘制的各个阶段, 未加入时间维度, 图中的节点、关系以及节点或关系的权重不随时间的变化而变化, 即图在完成最终布局后保持为静止状态。动态图可以理解为在图的节点、关系(边)、节点或关系的权重会随时间而发生变化<sup>[11]</sup>。

动态图被定义为:

$$D = \{G_1, G_2, \dots, G_i, \dots, G_n\}$$

式中:  $G_t = (V_t, E_t)$ , 表示动态图在时间  $t_i$  的静态表示。动画可以帮助用户无需过多思考便可以很好的理解数据<sup>[35]</sup>, 可有效的可视化不断发展的网络系统, 是静态图所不能做到的。对于动态图的绘制, 最重要的一点是在图形的变换过程中保持可读性, 即在切换时间帧时, 保持思维地图<sup>[36]</sup>, 动态图和静态图的布局方式类似, 分为两个过程: (1)节点位置的初始化, 图 5 展示了三种初始化算法<sup>[37]</sup>, 其中 BM(Barycenter Merging)算法将所有新加入的节点分别置于图的重心, 如图 5(a)所示; SBM(Sequential Barycenter Merging)算法在 BM 算法的基础上按照时间序列依次将节点置于图中, 如图 5(b)所示; SSBM(Sorted Sequential Barycenter Merging)算法又在 SBM 算法的基础上对节点进行了排序, 如图 5(c)所示; (2)节点位置的不断迭代, 图 6 展示了 AER(Additional Edge Resizing)方案<sup>[37]</sup>中节点位置初始化的过程。当 AER 向已有图中添加新边时, 首先移动当前距离最大的节点 2 和节点 10, 然后移动节点 1 和节点 10, 当节点之间的距



离满足一定条件后节点停止移动并完成布局。这两个过程共同完成了当构成图的数据发生变化时, 图布局变化过程。

基于节点-链接的图简单易懂, 容易被人们所理解, 但是当节点间的关系(边)比较稠密时会产生较大的视觉杂乱。邻接矩阵克服了这一缺点, 常被用作稠密图的可视化。

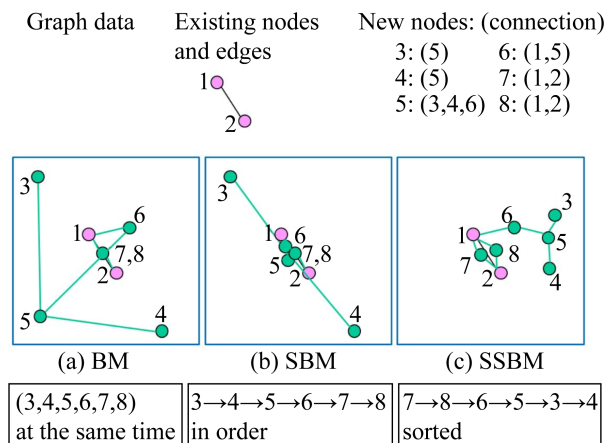


图 5 节点位置的初始化<sup>[37]</sup>

Fig. 5 Initialization of node locations<sup>[37]</sup>

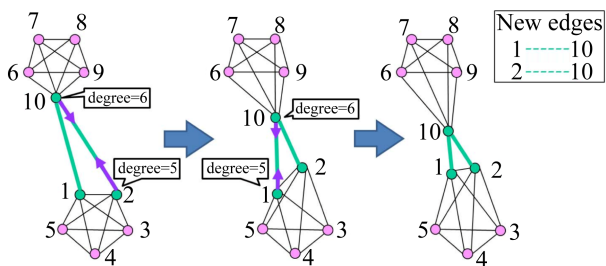


图 6 节点位置的调整过程<sup>[37]</sup>

Fig. 6 Nodes location adjustment process<sup>[37]</sup>

## 2 基于邻接矩阵的图分析方法

与节点-链接方法相比, 邻接矩阵可以很好的规避节点遮挡及边的交叉问题, 特别是在处理稠密图时, 邻接矩阵比节点-链接图更具可读性<sup>[38]</sup>。

### 2.1 基于邻接矩阵的表示方法

邻接矩阵通常用一个方阵的行向量和列向量来表达节点, 用行列交叉的元素来表达节点之间的关系, 方阵中的元素数值大小可用来表示节点之间关系的有无或强弱。

同节点-链接图相同, 表达具有相同节点和关系的邻接矩阵也有多种布局形式, 图 7 为具有相同关系却有不同节点排序的 4 种(远不止 4 种)可视化形式。由于邻接矩阵的可视化效果受节点排列顺序影响较大, 需要进行恰当地对节点进行排序以发现图中许多有价值的子结构。目前有大量关于矩阵的研究专注于节点顺序排列以及基于邻接矩阵的图模式挖掘<sup>[39]</sup>。

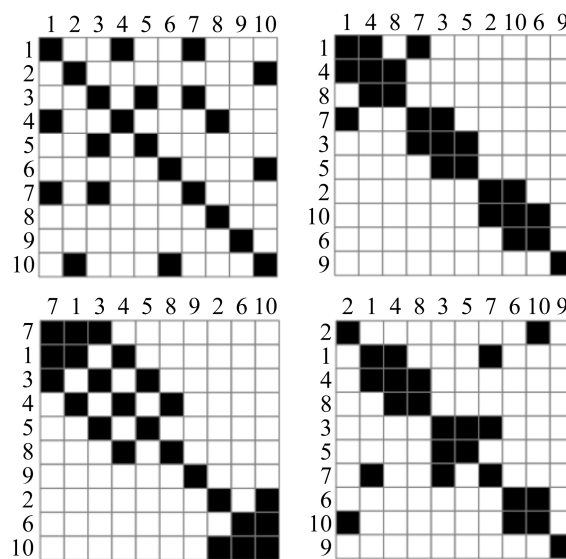


图 7 具有相同关系的 4 种邻接矩阵<sup>[39]</sup>

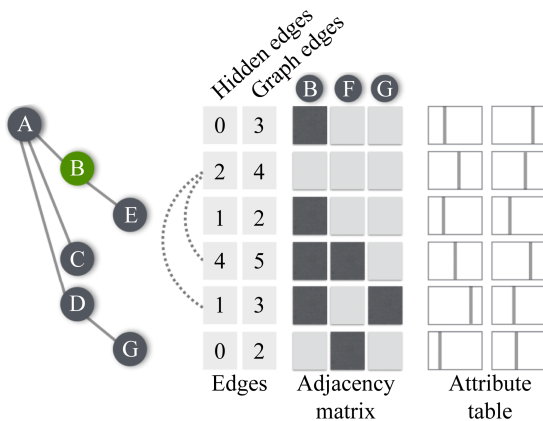
Fig. 7 Four adjacency matrices of same relationship<sup>[39]</sup>

总的来讲, 节点-链接图一般适用于稀疏网络, 而邻接矩阵更适用于稠密网络<sup>[40]</sup>。且邻接矩阵通常无法直观呈现网络拓扑结构, 也难以直观表达网络的中心和关系的传递性, 不适合用以完成路径跟踪类的任务。为了解决这一问题改进的邻接矩阵以及混合布局<sup>[41]</sup>的邻接矩阵被提了出来。

### 2.2 基于邻接矩阵的可视化方法

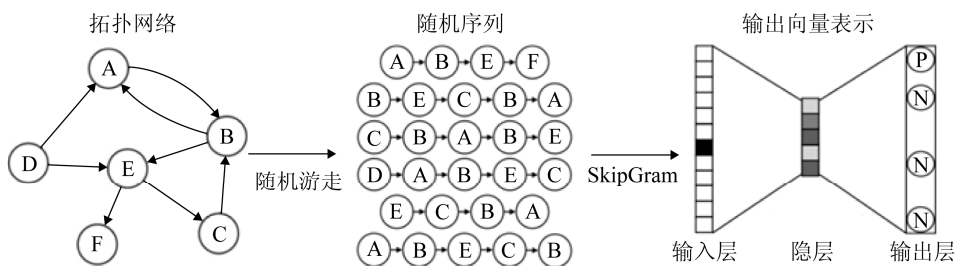
标准的邻接矩阵仅通过矩阵的行和列来表示矩阵中的信息。Ghoniem 等<sup>[40]</sup>的研究表明, 在一些低层次的阅读任务中, 除了路径查找外, 对于大图或稠密图, 邻接矩阵的表现要优于节点-链接图。在实际应用中为了表达更多的信息往往通过将邻接矩阵与其他可视化的视图结合或嵌入到其他布局中以表达更多信息。

MatrixExplorer<sup>[41]</sup>通过并置视图将同一个网络的节点-链接图和邻接矩阵表示形式耦合在一起。MatLink<sup>[42]</sup>在邻接矩阵的边界上覆盖了链接,是邻接矩阵与节点-链接图的结合,它进一步的表达了矩阵中的关联关系。StratomeX<sup>[43]</sup>通过将矩阵和桑基图柱状图等的结合,实现了一个可交互的多视图可视化系统,可以实现对基因组的可视分析。AniMatrix<sup>[44]</sup>是一个基于矩阵的交互式可视化系统,它与动画相结合,描述了软件设计是如何演进的。例如,它显示随着时间的推移添加和删除哪些新模块和耦合。目前邻接矩阵经常作为可视化的一部分与其他视图协同交互以更好地传递数据中存在的信息。Juniper<sup>[45]</sup>通过将树与邻接矩阵结合的方法实现了多变量图的可视分析,如图 8 所示,它可以很好的分析由学者、论文和引文度量组成的网络。

图 8 树与矩阵的结合<sup>[45]</sup>Fig. 8 The tree combines with adjacency matrix<sup>[45]</sup>

### 3 基于图嵌入的图分析方法

相比于传统的节点-链接和邻接矩阵,基于图嵌入的图分析方法可以对图实现更好的并行化处

图 9 基于 DeepWalk 的图嵌入过程<sup>[47]</sup>Fig. 9 A graph embedding process based on DeepWalk<sup>[47]</sup>

理,对图的计算以及查询做出了极大地优化,使得机器学习方法可以运用到图分析中。

#### 3.1 基于图嵌入的图表示方法

基于图嵌入的图表示方法可将图中的节点等信息以向量的形式表示,这些向量与向量之间在高维空间中保存了节点之间距离等信息,也保存了原始图的结构特征,在同质图和异质图的研究中都取得了较好的效果。同质图是指节点和边的类型数都为 1 的图,如论文互引网络中的节点和边仅分别表示单篇论文和引用关系;异质图是指节点类型数大于 1 或者边的类型数大于 1 的图,如计算机网络中的节点由计算机、路由器、打印机等多种类型组成。

##### 3.1.1 同质图嵌入方法

同质图嵌入方法被用作同质图的向量化处理,即在对图进行处理时,不考虑图中的节点以及链接关系的类型。

受 word2vec<sup>[46]</sup>的启发,2014 年图嵌入算法 Deepwalk<sup>[47]</sup>被提出,该算法首先在图中随机选择起始点,采用随机游走的方式生成随机序列并作为训练样本,然后将随机序列输入到 SkipGram 模型<sup>[46]</sup>中进行训练,最终输出向量形式的节点表示,如图 9 所示。随后有大量对 Deepwalk 的改进算法都取得了很好的效果,其中 node2vec<sup>[48]</sup>通过选择深度优先(DFS)或广度优先(BFS)随机游走的方式进行图训练样本采样,实现了在高维空间中保持图的同质性和结构等价性。图的同质性是指在原始图中距离相近的节点,其嵌入向量的距离也应该尽量接近。



如图 10 所示, 节点  $u$  与其相连的节点  $s_1$ 、 $s_2$ 、 $s_3$ 、 $s_4$  的嵌入向量的距离应该是接近的。结构等价性指的是结构上相似的节点, 其嵌入向量间的距离也应该尽量接近。如图 10 中, 节点  $u$  和节点  $s_6$  都是各自局域网络的中心节点, 结构上相似, 其节点嵌入向量之间的距离应该接近。

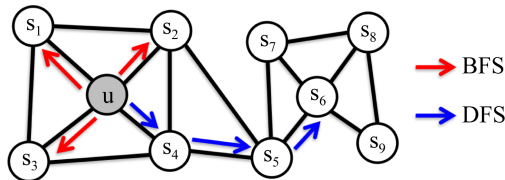


图 10 节点  $u$  的广度优先和深度优先游走策略<sup>[48]</sup>  
Fig. 10 BFS and DFS walk strategies from node  $u$ <sup>[48]</sup>

在此之后 Tang 等<sup>[49]</sup>提出了考虑一阶临近性和二阶邻近性的图嵌入表示方法。除了对节点的向量化表示外, 还实现了对子图和整个图的向量化表示<sup>[50-51]</sup>。

### 3.1.2 异质图嵌入方法

异质图的图嵌入方法是一个非常新的研究方向。它考虑到了现实世界的网络往往都是复杂多元的, 往往拥有多个类型的节点和链接关系。

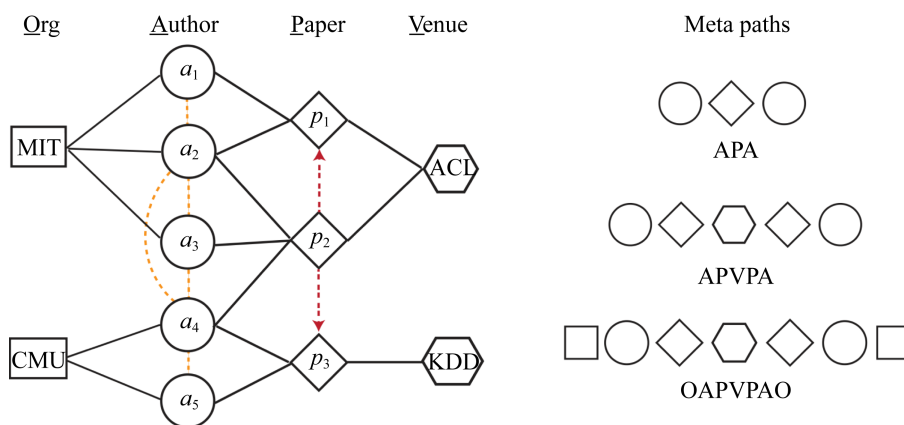
Dong 等<sup>[52]</sup>在 2017 年提出了异质网络的图嵌入框架 `metapath2vce` 和 `metapath2vec++`, 这两个

框架分别提出了考虑节点和链接类型的基于元路径<sup>[53]</sup>(meta-path)的随机游走如图 11(a)所示, 和考虑节点和链接类型的 SkipGram 模型如图 11(b)所示, 这两个过程分别对训练样本的采样阶段和训练阶段进行了优化。同年 Fu 等<sup>[54]</sup>提出的 HIN2Vec 分别对图中的节点和关系(元路径)进行向量表示。石川等<sup>[55]</sup>将异质图嵌入方法应用到推荐系统上面, 取得了非常好的效果。Wang 等<sup>[56]</sup>尝试了在双曲空间中嵌入异质图, 实验结果表明, 该模型在网络重构和链路预测任务上有较好的性能。

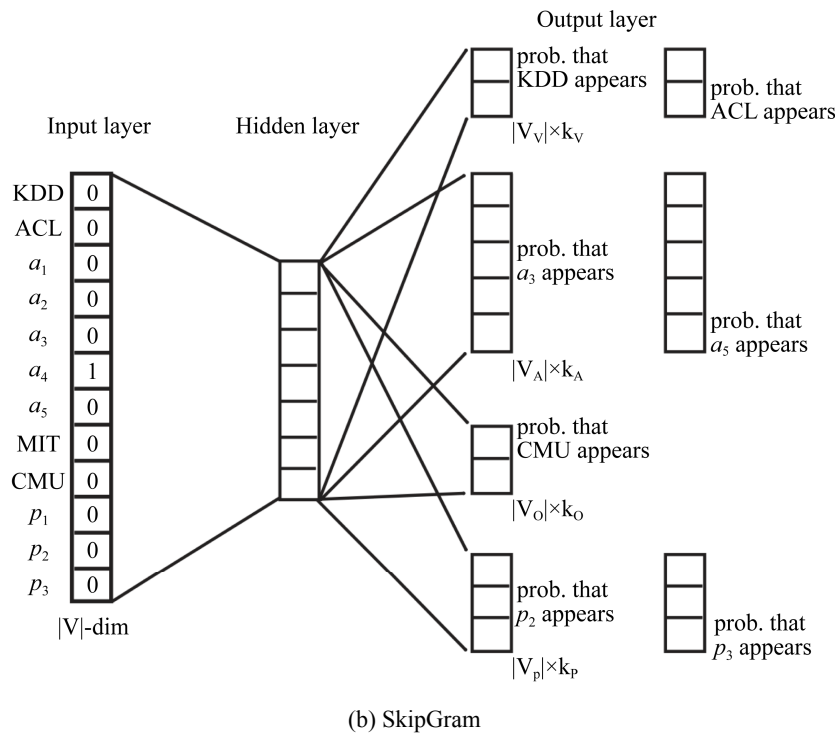
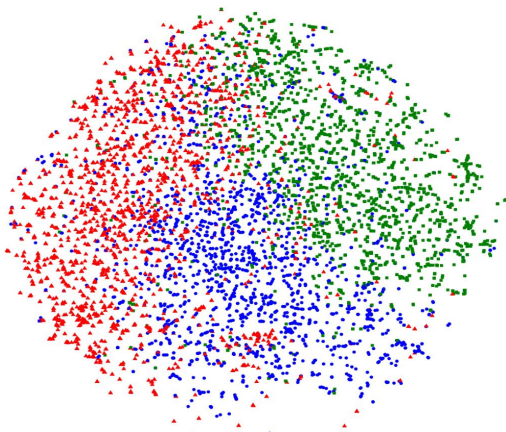
### 3.2 基于图嵌入的可视化方法

基于图嵌入的方法将网络中的节点等信息表示为向量形式(也称为高维数据), 可通过降维、聚类等技术对不同的实体在低维空间上进行可视化。

目前已经有很多方法被用作对高维数据进行降维处理, 如 T-SNE<sup>[57]</sup>和 LargeVis<sup>[58]</sup>等, 他们可以将高维数据映射到人们可以感知的低维空间, 如图 12 所示为先通过 Line 将合著者网络中的节点表示为高维向量, 然后采用 T-SNE 将高维向量降维到二维平面并以散点图的形式展示<sup>[49]</sup>。但是该方法在形式上还是很难被人们所理解的, 之后人们又提出了一些带有交互的可视化方法或系统。



(a) 基于元路径的随机游走

图 11 Dong 等提出的异质图嵌入方法<sup>[52]</sup>Fig. 11 The heterogeneous embedding approach of Dong et al<sup>[52]</sup>图 12 基于 Line 的论文合著者网络可视化<sup>[49]</sup>Fig. 12 Line-based co-author network visualization<sup>[49]</sup>

Pienta 等<sup>[59]</sup>开发了 VIGOR, 以实现对不同研究领域中具有关系的作者、文章和会议构建为一个子图。通过查询, 将符合条件的所有子图都映射到二维平面, 然后将相似的结果聚类, 分析人员可以快速地检测并理解结果中的潜在模式。Yan 等<sup>[60]</sup>使用挖掘频繁子图的算法 GRAMI<sup>[61]</sup>提取出图中的频繁子图, 随后通过图嵌入算法 graph2vec<sup>[51]</sup>将各个频繁子图做向量化表示, 随后以散点图的形式映

射到二维平面, 通过分析可以发现二维平面的散点图保存了很多语义信息。Chen 等<sup>[62]</sup>更是将图嵌入的方法应用在了子图的查询与子图模式挖掘上面, 他们首先以节点-链接的布局方式展示了大型网络的结构, 并预先对整个网络的节点进行向量化表示, 可以通过预先给定的子图模式范例, 以及自己随机设定的图, 在整个网络中探索与之相似的图结构, 如图 13(b)所示可以通过人为指定需要查询的网络模式, 图 13(c)为展示网络结构的节点-链接视图, 图 13(d)与子图范例结构类似的建议查询结果。虽然这种探索方式需要预先花费大量时间对网络进行向量化表示, 但是却大大减少了用户在交互式探索过程中所花费的查询时间。

#### 4 图的表示与可视化方法对比

图的表示与可视化方法在图分析中起着重要作用, 目前主要有节点-链接图、邻接矩阵以及图嵌入 3 类方法, 表 1 对这 3 类方法的特点进行了总结和对比。

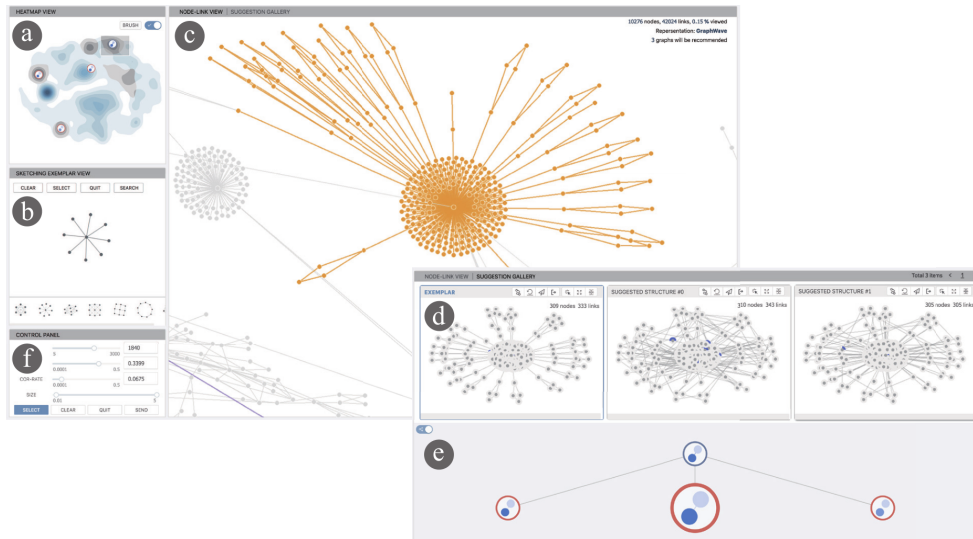


图 13 Chen 等提出的基于结构的大型网络探索方法<sup>[62]</sup>

Fig. 13 The structure-based large-scale network exploration approach proposed by Chen et al.<sup>[62]</sup>

表 1 图的表示与可视化方法小结与对比

Tab. 1 Graph representation and visualization methods summary and comparison

图的表示与可视化方法	方法分类	特点	典型算法举例	典型文献
节点-链接	弹簧电力模型	易于理解，但难以适应大型网络	FR <sup>[15]</sup>	[15-16]
	应力模型	可定量分析可视化的质量，但易形成局部最优布局	LRSM <sup>[17]</sup>	[17-21]
	受约束的图布局	图形美观，易于感知，可较好的理解数据间的关系	IPSEP-COLA <sup>[22]</sup>	[23-27]
	多层迭代算法	可视化速度快，可处理大规模图	FM <sup>[31]</sup>	[29-34]
	动态图	有利于分析现实世界中网络节点以及关系的建立与消失过程中的细节	SSBM <sup>[37]</sup>	[35-37]
邻接矩阵	矩阵表示法	避免了边交叉带来的视觉混乱，可视化效果受节点排列顺序影响较大	MatrixExplorer <sup>[42]</sup>	[39-45]
图嵌入	同质图	支持超大规模的图分析，可将机器学习的方法应用到图的分析	Deepwalk <sup>[47]</sup>	[47-51], [59-60]
	异质图	相比于同质图的图嵌入方法，生成表示节点的嵌入向量可更好的保存图中信息，但泛化能力较差	metapath2vec <sup>[52]</sup>	[52-56], [62]

## 5 结论

随着科学技术的发展，数据规模的不断增大，实体间的关联也越加复杂，形成的图的规模也越来越大，从而图分析技术也面临着巨大的机遇与挑战，具体如下。

(1) 随着图嵌入技术的逐渐成熟，对图的探索方式也逐渐的多元化，对异质网络以及动态网络的嵌入方法研究成为被关注的方向。以向量形式来表示网络中的节点，网络中的子图，甚至是整个网络的技术不断地出现，为大规模图的布局、探索以及

可视化提供了新方向。

(2) 网络规模的扩大，维度的增高，使其语义更为丰富。如何建立有效的交互方式帮助用户正确的理解和探索网络，进一步将人的智能与机器智能有机结合，辅助用户发现动态网络的变化规律和模式并预测将来的发展趋势，允许探索动态网络的历史以达到和当前网络对比的目的等都是未来研究中需要着重考虑的问题。

(3) 基于 GPU 的加速布局可以在几秒钟内绘制具有数十万个节点的一般图形，有效的解决了基

于力导向布局方法绘制图形速度慢的短板,也使得基于节点-链接的图布局可以应用到更多的领域。

(4) 新的可视化技术不断出现,但是评估还处于初级阶段,如何用数学方法定量的评估图布局的美观效果还没有被解决,人类审美观的差异对建立一个普适的评估方法带来了很大的挑战。

## 参考文献:

- [1] Heer J, Boyd D. Vizster: Visualizing Online Social Networks[C]// IEEE Symposium on Information Visualization. Washington: IEEE, 2005: 32-39.
- [2] 江美辉, 安海忠, 高湘昀, 等. 基于复杂网络的食物安全事件新闻文本可视化及分析[J]. 情报杂志, 2015, 34(12): 121-127.  
Jiang Meihui, An Haizhong, Gao Xiangyun, et al. The Visualization and Analysis of News Texts about Food Safety Incidents Based on Complex Networks[J]. Journal of Intelligence, 2015, 34(12): 121-127.
- [3] 杜晓敏, 陈谊, 李玥. TransGraph: 一种基于变换的可视分析关联图[J]. 计算机辅助设计与图形学学报, 2018(1): 79-89.  
Du Xiaomin, Chen Yi, Li Yue. TransGraph: A Transformation-Based Graph for Analyzing Relations in Data Set[J]. Journal of Computer-Aided Design & Computer Graphics, 2018(1): 79-89.
- [4] Beck F, Burch M, Diehl S, et al. A Taxonomy and Survey of Dynamic Graph Visualization[J]. Computer Graphics Forum (S0167-7055), 2017, 36(1): 133-159.
- [5] 李志刚, 陈谊, 张鑫跃, 等. 一种基于力导向布局的层次结构可视化方法[J]. 计算机仿真, 2014, 31(3): 283-288.  
Li Zhigang, Chen Yi, Zhang Xinyue, et al. Hierarchical Structure Visualization Methods Research[J]. Computer Simulation, 2014, 31(3): 283-288.
- [6] Eades P. A heuristic for graph drawing[J]. Congressus Numerantium (S0316-1382), 1984, 42: 149-160.
- [7] Kamada T, Kawai S. An algorithm for drawing general undirected graphs[J]. Information Processing Letters (S0020-0190), 1989, 31(1): 7-15.
- [8] Chen Y, Guan Z, Zhang R, et al. A survey on visualization approaches for exploring association relationships in graph data[J]. Journal of Visualization (S1343-8875), 2019, 22(3): 625-639.
- [9] 时磊, 廖琦, 林闯. 基于变换的大图点边可视化综述[J]. 计算机辅助设计与图形学学报, 2013, 25(3): 304-311.  
Shi Lei, Liao Qi, Lin Chuang. Survey on Transformation-based Large Graph Visualization[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(3): 304-311.
- [10] Pienta R, Abello J, Kahng M, et al. Scalable graph exploration and visualization: Sensemaking challenges and opportunities[C]// International Conference on Big Data & Smart Computing. Bangkok, Thailand: IEEE, 2014: 271-278.
- [11] 陈为, 沈则潜, 陶煜波. 数据可视化[M]. 北京: 电子工业出版社, 2013.  
Chen Wei, Shen Zeqian, Tao Yubo. Data visualization[M]. Beijing: Publishing House of Electronics Industry, 2013.
- [12] Battista G D, Eades P, Tamassia R, et al. Graph Drawing: Algorithms for the Visualization of Graphs[M]. Upper Saddle River: Prentice Hall, 1998.
- [13] Eades P, Feng Q W. Multilevel visualization of clustered graphs[C]// Graph drawing. Berlin Heidelberg: Springer, 1997: 101-112.
- [14] Wang Y H, Wang Y Y, Zhang H F, et al. Structure-aware Fisheye Views for Efficient Large Graph Exploration[J]. IEEE Transactions on Visualization and Computer Graphics (S1077-2626), 2018, 25(1): 566-575.
- [15] Fruchterman T M J, Reingold E M. Graph drawing by force-directed placement[J]. Software-Practice and Experience (S0038-0644), 1991, 21(11): 1129-1164.
- [16] Arleo A, Didimo W, Liotta G, et al. A distributed multilevel force-directed algorithm[J]. IEEE Transactions on Parallel and Distributed Systems (S1045-9219), 2019, 30(4): 754-765.
- [17] Khoury M, Hu Y, Krishnan S, et al. Drawing Large Graphs by Low-Rank Stress Majorization[J]. Computer Graphics Forum (S0167-7055), 2012, 31(3): 975-984.
- [18] Ko Y J, Yen H C. Drawing clustered graphs using stress majorization and force-directed placements[C]// 20<sup>th</sup> International Conference Information Visualisation. PORTUGAL: IEEE, 2016: 69-74.
- [19] Hu Y. Algorithms for Visualizing Large Networks[J]. Combinatorial Scientific Computing (S1862-4405), 2011, 5(3): 180-186.
- [20] Gibson H, Faith J, Vickers P. A survey of two-dimensional graph layout techniques for information visualisation[J]. Information Visualization (S1473-8716), 2013, 12(3/4): 324-357.
- [21] Wang Y, Wang Y, Sun Y, et al. Revisiting Stress Majorization as a Unified Framework for Interactive Constrained Graph Visualization[J]. IEEE Trans Vis

- Comput Graph (S1077-2626), 2018, 24(1): 489-499.
- [22] Tim D, Yehuda K, Kim M. IPSep-CoLa: An Incremental Procedure for Separation Constraint Layout of Graphs[J]. IEEE Transactions on Visualization and Computer Graphics (S1077-2626), 2006, 12(5): 821-828.
- [23] He W, Marriott K. Constrained graph layout[C]// International Symposium on Graph Drawing. Berlin Heidelberg: Springer, 1998: 289-314.
- [24] Dwyer T, Koren Y, Marriott K. Stress Majorization with Orthogonal Ordering Constraints[C]// Graph Drawing. Berlin Heidelberg: Springer, 2006: 141-152.
- [25] Didimo W, Montecchiani F. Fast layout computation of clustered networks: Algorithmic advances and experimental analysis[J]. Information Sciences (S0020-0255), 2014, 260(15): 185-199.
- [26] Xiang-Sheng Z J, Samraat P, Matthew G D F. Graph Drawing by Stochastic Gradient Descent[J]. IEEE Transactions on Visualization and Computer Graphics (S1077-2626), 2018, 25(9): 2738-2748.
- [27] Wu H Y, Takahashi S, Ishida R. Overlap-free labeling of clustered networks based on Voronoi tessellation[J]. Journal of Visual Languages & Computing (S1045-926X), 2018, 44: 106-119.
- [28] Barnard S T, Simon H D. Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems[J]. Concurrency & Computation Practice & Experience (S1532-0634), 1994, 6(2): 101-117.
- [29] Meyerhenke H, Nollenburg M, Schulz C. Drawing Large Graphs by Multilevel Maxent-Stress Optimization[J]. IEEE Trans Vis Comput Graph (S1077-2626), 2018, 24(5): 1814-1827.
- [30] Gajer P, Kobourov S G. GRIP: Graph Drawing with Intelligent Placement[C]// Proceedings of the 8th International Symposium on Graph Drawing. London: Spring, 2001: 222-228.
- [31] Hachul S, Junger M. Drawing Large Graphs with a Potential-Field-Based Multilevel Algorithm (Extended Abstract)[C]// Lecture Notes in Computer Science. Heidelberg: Springer, 2004: 285-295.
- [32] Nguyen A, Hong S H. k-core based multi-level graph visualization for scale-free networks[C]// Pacific Visualization Symposium. Seoul: IEEE, 2017: 21-25.
- [33] Gansner E R, Hu Y, North S C. A maxent-stress model for graph layout[J]. IEEE Transactions on Visualization and Computer Graphics (S1077-2626), 2013, 19(6): 927-940.
- [34] Kirmani S, Raghavan P. Scalable parallel graph partitioning[C]// Proceedings of the International Conference on High Performance Computing. Denver, Colorado: ACM, 2013: 1-10.
- [35] Robertson G G, Mackinlay J D, Card S K. Cone-Trees: Animated 3D Visualizations of Hierarchical Information[C]// Sigchi Conference on Human Factors in Computing Systems. New Orleans: ACM, 1991: 189-194.
- [36] Archambault D, Purchase H C. The “Map” in the mental map: Experimental results in dynamic graph drawing[J]. International Journal of Human-Computer Studies (S1071-5819), 2013, 71(11): 1044-1055.
- [37] Hayashi A, Matsubayashi T, Hoshide T I, et al. Initial positioning method for online and real time dynamic graph drawing for time varying data[C]// Proceedings of the 17<sup>th</sup> International Conference on Information Visualisation. Washington D C: IEEE Computer Society Press, 2013: 435-444.
- [38] Alper B, Bach B, Riche N H, et al. Weighted graph comparison techniques for brain connectivity analysis[C]//Sigchi Conference on Human Factors in Computing Systems. New York: ACM, 2013:483-492.
- [39] Behrisch M, Bach B, Riche N H, et al. Matrix Reordering Methods for Table and Network Visualization[J]. Computer Graphics Forum (S0167-7055), 2016, 35(3): 693-716.
- [40] Ghoniem M, Fekete J D, Castagliola P. On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis[J]. Information Visualization (S1473-8716), 2005, 4(2): 114-135.
- [41] Henry N, Fekete J D. MatrixExplorer: a Dual-Representation System to Explore Social Networks[J]. IEEE Transactions on Visualization and Computer Graphics (S1473-8716), 2006, 12(5): 677-684.
- [42] Henry N, Fekete J D. MatLink: Enhanced Matrix Visualization for Analyzing Social Networks[C]// IFIP Conference on Human-Computer Interaction. Heidelberg: Springer, 2007: 288-302.
- [43] Lex A, Streit M, Schulz H J, et al. StratomeX: Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization[J]. Computer Graphics Forum (S0167-7055), 2012, 31(3): 1175-1184.
- [44] Rufiange S, Melançon G. AniMatrix: A Matrix-Based Visualization of Software Evolution[C]// Software Visualization. Victoria: IEEE, 2014: 137-146.
- [45] Nobre C, Streit M, Lex A. Juniper: A tree+ table approach



- to multivariate graph visualization[J]. *IEEE Transactions on Visualization and Computer Graphics* (S1077-2626), 2019, 25(1): 544-554.
- [46] Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and their Compositionality[C]// *Proceedings of the 26<sup>th</sup> International Conference on Neural Information Processing Systems*. Nevada: Curran Associates Inc, 2013, 3111-3119.
- [47] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations[C]// *Proceedings of the 20<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2014: 701-710.
- [48] Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks[C]// *Proceedings of the 22<sup>nd</sup> ACM SIGKDD international conference on Knowledge discovery and data mining*. California: ACM, 2016: 1225-1234.
- [49] Tang J, Qu M, Wang M, et al. Line: Large-scale information network embedding[C]// *Proceedings of the 24th International Conference on World Wide Web*. Florence: International World Wide Web Conferences Steering Committee, 2015: 1067-1077.
- [50] Narayanan A, Chandramohan M, Chen L, et al. subgraph2vec: Learning Distributed Representations of Rooted Sub-graphs from Large Graphs[C]// *Proceedings of the 23<sup>rd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Canada: ACM, 2016: 385-394.
- [51] Xiaojun Xu, Chang Liu, Qian Feng, Heng Yin, Le Song, and Dawn Song. Neural network-based graph embedding for cross-platform binary code similarity detection[C]// *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. Dallas: ACM, 2017: 363-376.
- [52] Dong Y X, Chawla N V, Swami A. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks[C]// *Proceedings of the 23<sup>rd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Canada: ACM, 2017: 135-144.
- [53] Sun Y Z, Han J W. Mining Heterogeneous Information Networks: Principles and Methodologies[J]. *ACM SIGKDD explorations newsletter* (S1931-0145), 2012, 14(2): 439-473.
- [54] Fu T Y, Lee W C, Lei Z. HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning[C]// *ACM Conference on Information and Knowledge Management*. Singapore: ACM, 2017: 1797-1806.
- [55] Shi C, Hu B B, Zhao W X, et al. Heterogeneous Information Network Embedding for Recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering* (S2326-3865), 2018, 31(2): 357-370.
- [56] Wang X, Zhang Y D, Shi C. Hyperbolic Heterogeneous Information Network Embedding[C]// *33<sup>rd</sup> AAAI Conference on Artificial Intelligence*. Honolulu: AAAI, 2019: 5337-5344.
- [57] Van der M L, Hinton G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research* (S1532-4435), 2008, 9(11): 2579-2605.
- [58] Tang J, Liu J, Zhang M, et al. Visualizing Large-scale and High-dimensional Data[C]// *Proceedings of the 25<sup>th</sup> International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. Canada: ACM, 2016: 287-297.
- [59] Pienta R, Hohman F, Endert A, et al. VIGOR: Interactive Visual Exploration of Graph Query Results[J]. *IEEE Transactions on Visualization and Computer Graphics* (S1077-2626), 2018, 24(1): 215-225.
- [60] Yan K, Cui W W, Zhao T J. Frequent Pattern-based Graph Exploration[C]// *Proceedings of the 12<sup>th</sup> International Symposium on Visual Information Communication and Interaction*. Shanghai: ACM, 2019: 1-8.
- [61] Elseidy M, Abdelhamid E, Skiadopoulou S, et al. GRAMI: Frequent Subgraph and Pattern Mining in a Single Large Graph[J]. *Proceedings of the VLDB Endowment* (S2150-8097), 2014, 7(7): 517-528.
- [62] Chen W, Guo F Z, Han D M, et al. Structure-Based Suggestive Exploration: A New Approach for Effective Exploration of Large Networks[J]. *IEEE Transactions on Visualization and Computer Graphics* (S1473-8716), 2019, 25(1): 555-565.