2017届研究生硕士学位论文

分类号： 学校代码：10269

密 级： 学号：51151500102



**East China Normal University**

**硕士学位论文**

**MASTER’S DISSERTATION**

**论文题目：零售商店选址可视分析**

**院 系：计算机科学与软件工程学院**

**专 业：软件工程**

**研究方向：信息可视化**

**指导教师：王长波 教授**

**学位申请人：李柯林**

2017 年 4月

Dissertation for master degree in 2017 University code: 10269

Student ID: 51141500016

**East China Normal University**

**Title：Visual analysis of retailing store location selection**

**Department: Computer Science and Software Engineering**

**Major: Software Engineering**

**Research Direction: Information Visualization**

**Supervisor: Changbo Wang Professor**

**Candidate: Kelin Li**

**April, 2017**

**华东师范大学学位论文原创性声明**

郑重声明：本人呈交的学位论文《零售商店选址可视分析》，是在华东师范大学攻读**硕士**/博士（请勾选）学位期间，在导师的指导下进行的研究工作及取得的研究成果。除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

**作者签名： 日期：** 年 月 日

**华东师范大学学位论文著作权使用声明**

《零售商店选址可视分析》系本人在华东师范大学攻读学位期间在导师指导下完成的**硕士**/博士（请勾选）学位论文，本论文的著作权归本人所有。本人同意华东师范大学根据相关规定保留和使用此学位论文，并向主管部门和学校指定的相关机构送交学位论文的印刷版和电子版；允许学位论文进入华东师范大学图书馆及数据库被查阅、借阅；同意学校将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于（请勾选）

（ ）1.经华东师范大学相关部门审查核定的“内部”或“涉密”学位论文\*，于 年 月 日解密，解密后适用上述授权。

（ ）2.不保密，适用上述授权。

导师签名 本人签名

年 月 日

\* “涉密”学位论文应是已经华东师范大学学位评定委员会办公室或保密委员会审定过的学位论文（需附获批的《华东师范大学研究生申请学位论文“涉密”审批表》方为有效），未经上述部门审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权）。

**李柯林 硕士学位论文答辩委员会成员名单**

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 职称 | 单位 | 备注 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 摘要--900

大型零售商店在我国发展迅速，如何选择合适的新店位置，对企业获得更大的利润至关重要**。**在大数据时代到来的今天，零售业海量数据的产生，使得选址问题能够更加客观更加科学的分析。但是同样，由于数据量巨大以及跨学科研究的成本过高，没有一个明确的模式来对选址问题进行全面的研究。传统的经济学研究方法大多采用抽样调查以及专家评估的方式，已经无法满足如今越来越庞大的商业区选址的分析要求

现有的关于商圈的研究主要集中在商业领域和经济学领域，本文以大数据分析的角度入手，对人流数据和商业数据进行了深层次的研究，并通过可视化的方式从多角度进行分析。与基于统计抽样的位置推荐研究不同，我们通过挖掘销售数据，商圈数据和多维地理空间数据来设计视觉分析系统。 具体来说，我们优化了商圈吸引力模型来预测客户流大小，并提出了数据驱动的视觉比较模型来支持本文的分析。 此外，我们构建了一个交互式视觉分析系统，用于直观的进行选址研究

本文提出了不同人群行为的分析方法，基于数据分析的商圈吸引力模型的优化方法，选址分析方法以及设计了直观的可视化系统。首先为了准确的得到去往商圈购物的人数，我们基于多维地理空间数据和交通卡刷卡数据对人群进行分类，之后我们根据不同人群的行为特征对商圈进行等级划分，并对不同类型人群聚集行为进行了研究。。

本文提出了两种对图布局进行评估分析的方法。首先针对用户需要关注拓扑重要的节点的情形，结合视觉因素对图布局效果的影响，提出了一种基于视觉感知的图布局评估方法。具体地，基于用户实验分析影响节点的视觉重要性的因子并设计模型对节点的视觉重要性进行量化，通过对比节点的视觉重要性和拓扑结构重要性之间的差异来评估图布局的优劣，一方面借助热力图来直观地展示两者之间的差异，另一方面通过计算精确得到差异值，选择出更符合人类的视觉感知的图布局算法。

同时，本文还提出了客观的图布局整体质量评估方法。首先建立图布局主观质量评分的数据库，请用户通过主观实验对不同图布局效果进行评分；接下来对于每一个图布局，计算各个可能影响到图布局效果的指标；最后，我们把图布局中影响布局效果的指标作为自变量，用户对不同图布局的主观评分作为因变量，建立二者的回归模型，通过回归模型对一个图布局进行整体质量评分。

本文中的两种图布局评估方法具有不同的适用场景，通过文中提出的图布局评估方法选取出符合用户心理映像的图布局，使其深入地理解图中的特征信息。实验分析显示出我们的方法与实际情况切合度很好。

关键词：可视化，图布局，视觉感知，评估模型，拓扑结构

# ABSTRACT

With the arrival of information age, complex networks include social network, financial transaction network, biology network and so on, producing a large number of large-scale data every day. Data analysis and data mining for complex networks can explore a lot of meaningful information. In the complex network, the number of nodes is very large, the relationship between edges is also very complex, so traditional ways such as texts and forms can’t meet the user's requirements of understanding the structure of network. The visual layout of network presents the underlying structure of the network intuitively, so that people can understand the relational data more deeply. Therefore, the automatic visual layout method of complex network has been the major method of analyzing the network data.

There exists a large amount of network layout methods, and graph layout algorithms have different advantages and disadvantages from different aspects. The quality of graph layout greatly affects the user's understanding of graph structure. Some visualization means are mainly proposed from a technical perspective, while ignoring the psychological cognition of human. As a consequence, many visualization results can’t be understood and accepted by users. Based on human perception, the scientific evaluation of the intuition and effectiveness of different graph layouts can help the users select the layout that conforms to their psychological recognition, so that users can see through the internal relationship of the network data. So evaluating the graph layouts based on the characteristic of a graph and users’ psychological cognitive has significantly theoretical and application value.

This paper proposes two graph layout evaluation methods. Firstly, this paper proposes an evaluation method of graph layouts based on visual perception, and it is applicable to the situation when users focus on important nodes. It analyzes the nodes’ visual importance based on a user experiment and designs a model to quantify the nodes’ visual importance. Then it evaluates the pros and cons of graph layouts by comparing the topological importance and visual importance of nodes. A heatmap-based visualization is used to provide visual presentation for the difference between the topological importance and visual importance of nodes. Meanwhile, a metric is built to quantify the difference precisely. Finally, experiments are done under different scale of data sets to further analyze the characteristics of these graph layout methods. The layout with small differences between visual importance and topology importance is preferred.

At the same time, this paper proposes an objective overall quality assessment method for graph layout algorithms. Firstly, we build the subjective rating database of graph layouts. The subject experiment is designed to rate different graph layouts. Then, for each graph layout, we use the readability metrics of the layout as independent variables, the subjective score of users as the dependent variable, to establish the regression model. Through the regression model, we can get the overall quality score of a graph layout.

Two graph layout evaluation methods in this paper have different applicable scenarios. Selecting the graph layout that conforms to the user's psychological recognition helps users deeply understand the feature of the graph. The experiment shows that our method is in good agreement with the actual situation.

**Keywords:** *Visualization, Graph Layout, Perception Cognition, Evaluation Model, Topological Structure*

# 目录

[第一章 绪论 1](#_Toc479544919)

[1.1 研究背景及意义 1](#_Toc479544920)

[1.2 国内外研究现状 2](#_Toc479544921)

[1.3 研究目标和内容 5](#_Toc479544922)

[1.4 论文组织结构 5](#_Toc479544923)

[1.5 本章小结 6](#_Toc479544924)

[第二章 相关工作 7](#_Toc479544925)

[2.1 节点的拓扑结构重要性 7](#_Toc479544926)

[2.2 图布局相关工作 8](#_Toc479544927)

[2.2.1图布局算法 9](#_Toc479544928)

[2.2.2 图布局算法的评估 13](#_Toc479544929)

[2.3 视觉在可视化领域的应用 14](#_Toc479544930)

[2.4 图像质量评价 15](#_Toc479544931)

[2.5 回归分析 17](#_Toc479544932)

[2.5.1 线性回归 18](#_Toc479544933)

[2.5.2 支持向量回归 18](#_Toc479544934)

[2.6 本章小结 19](#_Toc479544935)

[第三章 基于视觉感知的图布局评估 21](#_Toc479544936)

[3.1 基本思路 21](#_Toc479544937)

[3.2 用户调查实验分析 23](#_Toc479544938)

[3.3 节点视觉重要性计算 24](#_Toc479544939)

[3.4 节点拓扑重要性与视觉重要性对比 27](#_Toc479544940)

[3.4.1基于热图的可视化 27](#_Toc479544941)

[3.4.2 差异量化分析 28](#_Toc479544942)

[3.5 本章小结 30](#_Toc479544943)

[第四章 基于回归的图布局质量评估 31](#_Toc479544944)

[4.1 图布局的指标特征 31](#_Toc479544945)

[4.2 图布局的主观质量评价 37](#_Toc479544946)

[4.3 回归分析 39](#_Toc479544947)

[4.4 本章小结 42](#_Toc479544948)

[第五章 实验分析 43](#_Toc479544949)

[5.1 基于视觉感知的图布局评估 43](#_Toc479544950)

[5.1.1 食物关系网络 43](#_Toc479544951)

[5.1.2 蛋白质关系网络 48](#_Toc479544952)

[5.1.3 综合对比 51](#_Toc479544953)

[5.2 基于回归的图布局整体质量评估 52](#_Toc479544954)

[5.2.1 指标评估 53](#_Toc479544955)

[5.2.2 拟合效果评估 54](#_Toc479544956)

[5.3 本章小结 56](#_Toc479544957)

[第六章 总结和展望 57](#_Toc479544958)

[6.1 总结 57](#_Toc479544959)

[6.2 未来的研究工作 57](#_Toc479544960)

[参考文献 59](#_Toc479544961)

[附录 64](#_Toc479544962)

[致 谢 72](#_Toc479544963)

[研究生期间科研成果 74](#_Toc479544964)

# 图清单

图2.1 节点中心性示意图 9

图2.2 一个图的三种布局 9

图2.3 使用FR模型的力引导布局 10

图2.4 Fruchterman-Reingold布局、ForceAtlas2布局、Yifanhu布局和Openord布局 11

图2.5 地图布局 12

图2.6 聚类布局 13

图2.7 一个两种美学标准冲突的例子 14

图2.8 视觉注意机制示例图 15

图2.9 图像质量评估的具体流程 17

图2.10 对回归分析分类的三个度量 18

图2.11 svm向高维空间的映射 19

图3.1 基于视觉感知的图布局评估流程图 22

图3.2 节点P的周边区域 24

图3.3 邻接节点的分布的对称性 26

图3.4 节点重要性热图示意图 28

图4.1 基于回归的图布局整体质量评估的流程图 32

图4.2 图布局最小角度指标示意图 33

图4.3 图布局边正交性指标示意图 34

图4.4 图布局中节点分布的条形图展示 35

图4.5 图布局中边长度一致性指标的示意图 36

图4.6 图布局主观评分与各指标的散点图 40

图4.7 模型选择流程图 41

图5.1 食物关系网示意图 43

图5.2 食物关系网Fruchterman-Reingold布局热图 44

图5.3 食物关系网ForceAtlas2布局热图 45

图5.4 食物关系网Openord布局热图 46

图5.5 食物关系网Yifanhu布局热图 47

图5.6 蛋白质关系网Fruchterman-Reingold布局热图 48

图5.7 蛋白质关系网ForceAtlas2布局热图 49

图5.8 蛋白质关系网Openord布局热图 50

图5.9 蛋白质关系网Yifanhu布局热图 51

图5.10 图布局质量评价算法的性能评价框架 52

图5.11 基于线性核函数的SVR预测结果 55

# 表清单

[表3.1 由不同影响因素设计的用户实验图 23](#_Toc479578639)

[表3.2 实验结果人数百分比 24](#_Toc479578640)

[表3.3拓扑重要性与视觉重要性差异量化值示例 29](#_Toc479578641)

[表4.1 节点在图布局中x轴、y轴方向的分布情况 35](#_Toc479578642)

[表4.2 主观实验中图结构的基本信息 37](#_Toc479578643)

[表4.3 图布局主观质量评价评分表 38](#_Toc479578644)

[表4.4 图布局的平均意见分数 38](#_Toc479578645)

[表4.5 SVR中不同的核函数 42](#_Toc479578646)

[表5.1 食物关系网络拓扑重要性与视觉重要性差异量化值 47](#_Toc479578647)

[表5.2蛋白质关系网络拓扑重要性与视觉重要性差异量化值 50](#_Toc479578648)

[表5.3 不同回归方法的性能指标对比 54](#_Toc479578649)

# 第一章 绪论

## 1.1 研究背景及意义

近年来，信息技术逐渐进入各行各业，产生了超越以往年代的海量数据。世界各地的移动设备、社交网络、电子商务网站、传感器等时时刻刻都在产生着各类各样的数据。如今，大数据已成为目前研究的热点，从海量的数据中获取有价值的信息是大数据研究中的一个热点问题。大数据中的“大”字，意味着数据数量与质量都远超普通的数据，然而在大数据技术中并不是数据量越大技术就越先进，而在于对这些海量的数据进行专业化处理，从中挖掘出有价值的信息。所以说，大数据并不在“大”，而在于“有价值”。数据的价值含量、挖掘成本远比数量更加重要。现如今越来越多的行业都积累了足够数量的数据，如何通过使用大数据技术利用这些数据的价值成为各行各业需要考虑的一个重要课题。

传统的数据处理方法大都是基于统计的分析。然而，统计分析的方法只适用于数据量比较小的应用场景，当数据量变得越来越大时，这种方法也就无法满足人们发掘数据价值的需求。近年来提出的数据挖掘、机器学习的方法虽然能够通过算法有效地从大量数据中找到隐藏其中的信息，但是信息的展示不够直观。对于普通用户和专家来讲，信息可视化是一种更加友好的展示数据信息本质特征的方法。因为一图胜千言，信息可视化技术能够直观生动地呈现大数据的特点，同时让用户和数据进行有效的互动。当繁复的数据以直观的可视化图形的形式展示在用户面前时，分析者往往能够直接地获取数据背后隐藏的信息。信息可视化经过几十年的发展，已经积累了大量的可视化方法与技术，为可视化的应用打下了坚实的基础。然而，事实上绝大多数可视化技术只被少数研究人员所使用，难以获取广泛的应用和认可。这是因为许多可视化手段单纯从技术角度进行革新，而忽略了人类的认知规律和心理映像，因此出现了很多可视化效果难以被用户理解和接受的现象。从人类的感知出发，对可视化手段的直观性和有效性进行科学的评价，从而选择、创造出符合用户心理映像的可视化方法，使其一眼看穿大数据，将是目前可视化领域面临的最大挑战[1]。

复杂网络包括社交网络、金融交易网络、生物网络等，它已成为最近几十年研究的热点话题。特别是近几年来，随着Facebook、微博、微信等国内外大型在线社交网络的不断出现，产生了越来越多的大规模网络数据，复杂网络越发引起了人们的关注。复杂网络中，节点的数目很多，边的关系也非常复杂，传统的表格等文字表现方式已经无法满足用户对网络结构数据理解、挖掘的需求。通过网络图可视化的方式，可以用图形的方式将隐藏在网络数据里的信息快速直观地加以展示，从而使人们直观地获取数据中的关系特征，更加深入地理解关系数据，所以复杂网络的自动可视化布局算法也是近几十年内研究的热门领域。现有的网络图可视化布局算法有很多，包括力引导布局、圆形布局、地图布局等几大类，每一类中又有许多布局算法[2]，然而各种图布局算法各有优劣，图布局的目标是直观、形象地展示出网络的本质特征。从人类的认知规律和心理映像出发，对不同的图布局算法进行评估，从众多图布局算法中选取适合的布局具有重要的现实意义。

## 1.2 国内外研究现状

近些年来，信息可视化受到越来越多的关注，Shneiderman根据信息的特征把信息可视化技术分为一维信息(1-dimensional)、二维信息(2-dimensional)、三维信息(3-dimensional)、多维信息(multi- dimensional)、层次信息(tree)、网络信息(network)、时序信息(temporal)可视化[3]。20年来研究人员针对以上可视化类型的信息提出了许多新的方法和技术，并在各个行业获得了广泛的应用[1]。

网络信息可视化是当前最为热门的数据可视化研究方向之一，已经有超过50年的研究历史，图布局的目的是为复杂网络设计自动化布局算法，图的可视化布局并不是唯一的。目前也有许多种网络可视化算法，在Visual Complexity[[1]](#footnote-1)中收录的网络可视化应用就有600多种。根据Shneiderman[3]提出的分类标准，网络图可视化主要有以下几种布局方法:基于力引导布局、基于地图布局、基于环形布局、基于相对空间布局、基于聚类布局等[4]。好的网络布局可以在有限的空间展示大量的数据，也可以增加用户理解网络结构的容易程度。事实上，已有实验研究表明图布局很大程度上影响了人们对数据的感知理解[5]。可视化的质量随布局算法不同而改变，所以对图布局算法的评估十分重要，符合人类的认知规律和心理映像的图布局易于理解，能够使用户深入理解图结构，从而吸引更多用户的使用。

针对图布局算法的评估主要分为以下两种：指标评估和可读性评估。所谓指标评估方法，大都使用的是美学指标，目前主要的美学指标有：边的弯曲程度[6]、整体布局的对称性[7]、边交叉的数量[8]、边长的均一性[9]、节点分布均匀程度[10]、节点密集与稀疏的程度[10]、空间利用效率[11]等。每个指标从不同的角度对图布局算法进行评估，不过美学标准之间可能会互相冲突，图布局中不可能同时实现所有的美学标准。实现其中一种美学标准可能会以消耗大部分其他标准为代价，以至于图布局的整体质量并不确定。近年来还出现了一些从其他角度提出的指标评估，Wang[12]设计了AmbiguityVis系统，考虑网络布局中出现的边交叉、点重叠等造成的歧义现象，以出现歧义性的多少为指标对各种可视化布局进行评估。

所谓图布局的可读性评估方法，主要是通过设计巧妙的、基于用户实验的主观评价，来更加真实地反映用户实际使用过程中图布局的效果。这种方式得到的评价结果可靠，无技术障碍，而且根据具体需求设计不同的实验，比如从图布局中完成指定任务、主观搜索相关信息的难易程度等，对图布局的不同方面的功能有比较明确的针对性和侧重。例如，Pohl[13]通过用户实验对力引导布局(force-directed layout)、正交布局(orthogonal layout)、层次性布局(hierarchical layout)进行比较与评估，他们设置了特定的任务，通过用户完成任务的时间以及正确率来判断布局的优劣，通过眼动仪形成的热图对实验结果进行分析和解释。Marriott [14]进行了一个对图布局进行重新绘制的实验，对比像对称性、节点的共线性、边排布的正交性这样的视觉感知因素对图布局的影响。他们要求参与实验的人员学习、记忆重现出之前看到的网络布局结构。然而，这些基于用户实验的图布局的评估需要邀请多个用户进行实验，在每次应用时需要重复进行。这种方式没有应用数学模型对其进行描述，主观实验的方式在可重复性、可操作性以及成本上都难以用于实践中大量的图布局设计与优化中。从应用角度来看需要消耗更多的人力物力，不易于实施，且难以实现实时的评价。

如果可以设计客观的图布局评价算法，其评价结果几乎与主观评价实验的结果完全一致，那么我们就完全可以在实践中用客观的图布局评价算法来代替主观实验的方式。客观算法的优点在于，其不需要消耗人力物力、实验结果具有可重复性、可信度较高、实时性好等。不过客观图布局评价算法的设计也不是空穴来风，设计好一个客观图布局的评价算法，我们也需要对该算法进行评价。如果通过测定发现客观图布局评价算法对特定数据集的评价与主观实验的评价结果一致性越高，该客观算法越好。

在本文中，我们设计了一种客观的图布局评价方法。我们首先把图布局的指标评估和图布局可读性的评估相结合，通过图布局的指标评估提取出图布局中的可量化特征，然后通过把这些特征作为因子，图布局可读性整体评分作为因变量进行回归分析，得到回归模型，通过回归模型获取对不同图布局算法的整体评分，整体评分结果就是客观算法对于该图布局的客观评价。这种方法综合考虑了多种特征对于图布局好坏的影响，对主观实验得到的图布局评分进行拟合。

在复杂网络的拓扑结构中，每个节点的重要程度是不同的，节点重要性是复杂网络研究中一个重要的概念，有许多不同的方法用来评估节点的重要性[15]。拓扑结构中节点重要性的研究具有现实意义，能够帮助用户分析出重要的节点，以加强用户对重要节点的关注。人类感知在可视化领域中起着重要的作用，好的视觉感知能够极大地提升信息展示的数量和质量。在网络可视化布局中，受视觉因素的影响，人类对不同节点进行感知的重要性顺序是不同的。当节点的视觉感知重要性顺序和其拓扑结构重要性顺序一致时，重要的节点就能够快速的被感知、发现，从而得到更好的可视化展示效果。我们提出了一种基于视觉的图布局评估方法，通过对图布局中节点视觉重要性和拓扑结构重要性的对比对不同图布局算法进行评估。

## 1.3 研究目标和内容

本篇论文主要研究对不同图布局算法进行评估的方法，从两个不同的角度出发提出了对图布局算法评估的方法。一种方法结合了视觉感知，适用于用户需要关注拓扑重要节点的情形。另一种方法结合了图布局的指标评估和基于用户实验的图布局的评估，提出了一种综合了各个指标的图布局评估方法，适用于图布局的整体质量评分。

本文的主要贡献有如下几点：

（1）结合人类视觉感知造成的用户对图布局中不同节点的不同感知效果，提出了节点的视觉重要性的定义，并通过用户实验得到影响节点视觉重要性的指标，最后提出了节点的视觉重要性的量化模型。

（2）提出了一种参考了视觉感知的图布局评估方法。通过热图和具体数值量化得到一个图布局中节点视觉重要性和拓扑结构重要性的差异，进一步对比不同图布局中差异的大小，从而对不同布局算法加以评估，选取出符合人类视觉感知的布局。

（3）设计了一种客观的图布局整体质量评价算法。首先建立图布局主观质量评分的数据库，请用户通过主观实验对不同图布局的整体质量进行评分；其次对于每一个图布局，计算图布局中各个可能影响到图布局效果的指标；最后，我们把不同图布局中影响布局效果的指标作为自变量，用户对不同图布局的主观评分作为因变量，建立二者的回归模型，通过回归模型对一个图布局进行整体质量评分。

## 1.4 论文组织结构

本论文的组织结构如下：

第一章主要介绍了论文的研究背景和意义，分析了图布局评估的重要意义、现状和存在问题，介绍了论文的研究目标、研究思路以及论文的组织结构。

第二章主要是对本文的相关工作进行介绍，包括节点拓扑结构重要性的概念、图布局算法及评估。介绍了视觉在可视化领域的应用、图像质量评价的主要研究方法以及通过线性回归和SVR解决回归问题的方法等。

第三章主要阐述了基于视觉的图布局评估方法，首先结合视觉感知中的理论提出了节点视觉重要性的概念；其次通过用户实验获取影响节点视觉重要性的因子，提出了节点视觉重要性的计算模型；最后详细介绍了通过对比节点拓扑重要性和视觉重要性从而评估图布局算法的方法，不同图布局之间节点视觉重要性和拓扑重要性的对比方法包括热图的对比以及具体数值的量化对比。

第四章讲述了基于回归的图布局整体质量评估，首先介绍了图布局中影响布局效果的指标及其计算方法；其次介绍了由用户对图布局进行主观评分的方法；最后描述了通过对图布局客观指标和主观质量评分进行回归分析，建立二者之间关系的方法。

第五章主要应用了本文提出的两种图布局评估方法进行实验并分析实验结果。第一部分采用了两个具体的实际数据集，对于每个数据集采用四种图布局算法并使用基于视觉的图布局评估方法对不同图布局进行评估；第二部分使用训练数据对基于回归的图布局主观质量评分结果进行验证，验证回归模型的准确性和稳定性。

第六章对论文内容进行总结，对未来的研究工作进行展望和规划。

## 1.5 本章小结

本章论述了本文的研究背景及研究意义，介绍了当前国内外的研究现状，阐述了图布局评估的研究意义、主要难题和解决方法，本文研究目标、研究思路，以及论文的组织结构。之后的章节会详细地论述本章所提出的内容。

# 第二章 相关工作

本章主要介绍图布局、视觉在可视化领域的应用、图像质量评价的相关工作研究，以及和本文研究相关的技术，主要包括节点拓扑结构重要性的概念和计算，回归分析的主要方法等。

## 2.1 节点的拓扑结构重要性

从数学角度研究网络时，只关心节点之间的关联，至于节点的位置、颜色、形状则不在考虑范围之内。通常把不依赖于节点和边的位置和形态所表现出来的网络结构叫作网络拓扑。由于网络结构的复杂性，复杂网络中每个节点的重要程度也不同。目前有许多针对节点拓扑结构重要性的研究，下文我们介绍几个常见的节点拓扑结构重要性度量指标。

一个节点的度中心性(Degree Centrality)，是指网络拓扑中与该节点直接相连的点的个数，它体现了节点的直接影响力，考察了该节点的直接社会关系。越大，节点的度中心性越大。

 （2.1）

接近度中心性(Closeness Centrality)是指该点与图中所有其他点的最短路径之和的倒数，反映了节点在网络中居于中心的程度，也体现了节点通过社会网络对其他节点的间接影响力，即考察该点的间接社会关系。位于网络中心的程度高的节点一般来说是比较重要的节点，该节点产生的消息能够以较少的时间和较小的代价迅速传遍整个网络，这样可以减少传播的时间和花销。节点接近度的经典定义是Sabidussi提出的[16]。假设表示图中从节点到节点的最短路径，则节点的接近度中心性定义为：

 （2.2）

由于节点到其他所有节点距离之和越小，节点位于网络中心的中心度越高，二者成反比，因此我们对节点到其他所有节点距离之和平均值取倒数得到，节点的接近度中心性()越大，节点越重要。

介数中心性(Betweenness Centrality)最早是Freeman[17]于1977 年在研究社会网络时提出的，从一个全新的角度反映了个体在社会中的地位。节点*p*的介数含义为网络的所有经过*p*的最短路径的数量。记间最短路径的集合为*Sij*，则节点*p*的介数中心性定义为：

 （2.3）

其中，表示经过节点p的最短路径的数量。*B(p)*越大，某节点在关系网络中最短路径上出现的次数越大，其他节点与其交流更密切，则其影响范围越大。使用介数可以用于寻找网络中“流量”大的重要节点。

特征向量中心性(Eigenvector Centrality)指标[18]考虑了与其相邻节点的影响力，是其相邻节点的中心化指标的线性叠加，它的值不仅与其相邻节点的数目有关，也取决于其邻接节点的重要性。为第*i*个节点的值，表示网络的邻接矩阵，有：

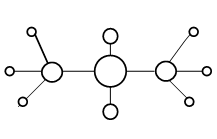
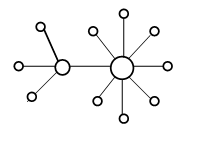
 （2.4）

求解以上方程，可能会有多个值符合以上方程，从中选取它的最大值，我们获取的相关向量中第*i*个分量即图中第*i*个节点的中心度指数。越大，节点的间接影响力越大，节点越重要。

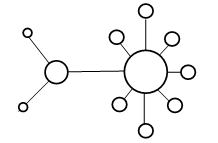
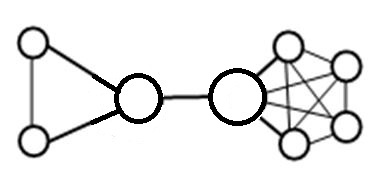
这里，我们分别通过以上四个指标计算节点的拓扑结构重要性，即度数中心性、接近度中心性、介数中心性和特征向量中心性，它们从不同角度展示了节点的重要程度。为了使人们对节点的拓扑结构重要性有更加直观的认识，我们做出节点中心性示意图如图2.1，节点越大表示其重要性越高。

## 2.2 图布局相关工作

可视化利用强大的人类视觉系统把复杂的认知过程转换为简单的视觉操作[19]，网络图的可视化基本可以分为4个步骤。首先要获取数据，对数据结构进行分析；然后对数据进行处理，将数据转化为适合网络图展示的结构；再选择适合该数据的可视化布局算法；最终生成可视化布局终端界面。选择合适的图布局算法是网络图可视化中至关重要的一步，下文我们对图布局算法和图布局算法的评估进行介绍。



a. 节点的度数中心性 b. 节点的接近度中心性



c. 节点的介数中心性 d. 节点的特征向量中心性

图2.1 节点中心性示意图

**2.2.1图布局算法**

图通常以点线连接图的方式可视化出来，然而，图的可视化布局并不是唯一的。如图2.2所示，它给定了一个图的三种画法，它们的唯一区别是布局[20]。从图中可以很明显看出最左边的图具有最好的可视化效果，因为它把图中边和点的结构特征清楚地展示了出来，另外两幅图则要花一些工夫才能看懂。从这个例子中可以看出不同图布局具有不同的展示效果，图布局很大程度上影响了人们对数据的感知理解。



图2.2 一个图的三种布局[20]

下面我们针对几个比较常用的网络图布局方法进行介绍。

（1）力导引布局(FDA，Force-Directed Algorithm)

力导引布局方法充分展示了网络的整体结构，其产生的网络布局结构美观，是一种广泛应用的网络结点布局技术。这种方法是由Eades在1984年提出的[21]，它把结点之间的边看做是固定长度的弹簧，若两点间距离过远，就会相互吸引，距离过近又会互相排斥。直到系统的总能量减少到最小值时会进入一个稳定的状态，运动停止，得到了力引导的布局。这种算法的复杂度是，复杂度比较高。在这种图布局算法提出之后，出现了很多基于该算法的改进，一些FDA算法以提高性能为目的[22]，另一些FDA算法则致力提高图布局的美观性和可读性。

而后又出现了以最初力引导算法为基础的改进算法，其中较著名的是FR算法，它分别从理论依据，美学标准和展现能力方面做出了改进[23]（如图2.3所示）。

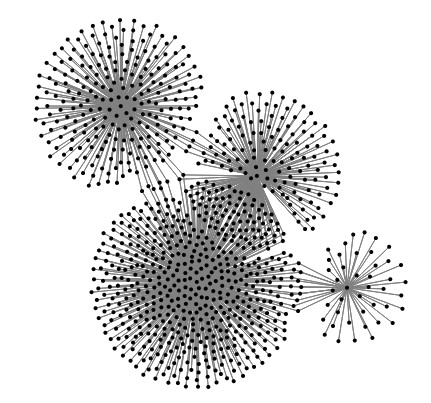


图2.3 使用FR模型的力引导布局

设显示区域的高为*H*，宽度为*W*，结点的位置为*pos*，位置偏移量为*dis*。定义*a*为显示区域：

 （2.4）

平衡距离：

 （2.5）

*|v|*代表图中结点的个数。

两点间的几何距离：

 （2.6）

两点之间的引力和斥力：

 （2.7）

 （2.8）

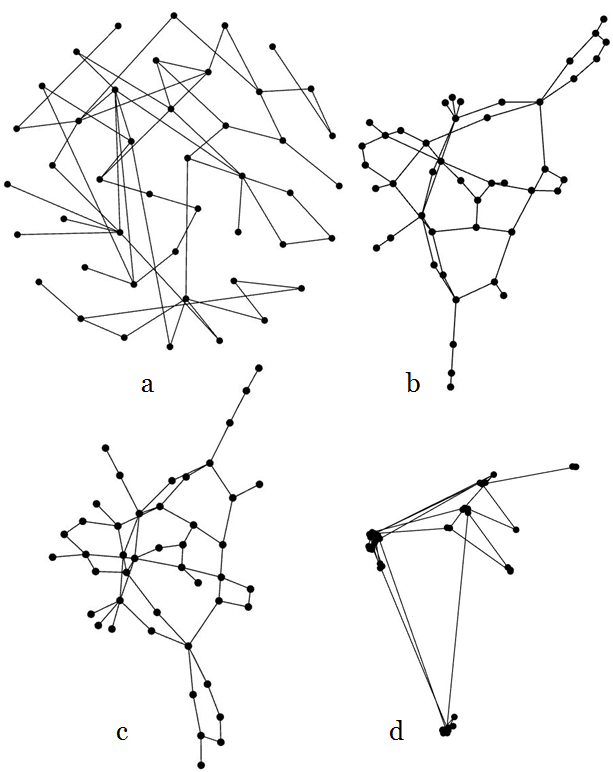


图2.4 Fruchterman-Reingold布局、ForceAtlas2布局、Yifanhu布局和Openord布局

FR算法实现时首先要计算结点之间的斥力和引力，最后综合引力、斥力以及结点间的最小和最大位移来确定结点间的距离。除了FR算法外，本文还使用到了ForceAtlas2算法[24]、Yifanhu算法[25]和Openord算法[26]三种力引导布局算法，它们具有不同的展示效果。图2.4为使用Gephi[[2]](#footnote-2)生成同一个图的四种力引导布局的示意图。

（2）地图布局

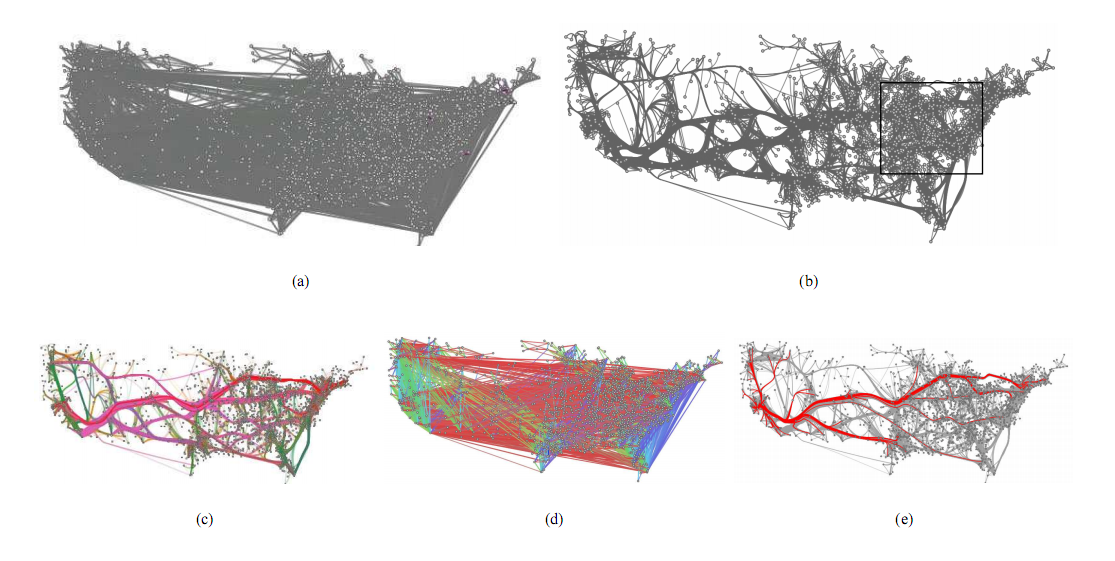


图2.5 地图布局[27]

对于地图布局来讲，每个节点的位置是固定的，在不同的地理坐标点上放置不同的结点，在有关联的结点之间绘制一条边，地图 (世界地图、大洲地图、国家地图、省或市地图等)是图布局的背景，这种布局方法能够给用户展示结点在地理位置上的布局和关联[27]（图2.5）。

地图布局的优点是用户容易理解，地图背景最大的特点是直观，可以直观地显示网络结点的位置迁移和两个地点之间的关联；缺点在于网络节点在背景图上位置由其地理位置确定，会出现节点重叠、边交叉的问题，当处理可视化节点较多或者边数量很大的大型网络结构时，展示效果比较混乱。根据该算法的改进算法一般针对点重叠、边交叉的问题加以改进，通过边绑定的方法实现更清晰直观的地图网络图布局[27]。

（3）聚类布局(Cluster)

所谓图聚类是指把图中相对关系紧密的结点及其相关的边分组形成一个可以用一个空间位置上聚集到一起的子图。子图内各结点具有较高的相似性，而不同的子图之间各结点的相似性较低。它根据节点的不同属性及节点相互间的连接关系，通过聚类对网络节点进行分组。图2.6展示了一个多层聚类布局，同一类别的节点在空间上同一层上[28]。

聚类布局能够帮助用户发现网络图中结点之间隐藏的关系，将关联比较紧的点在空间布局中放在一起，更为直观清晰，一般多与其他布局方法结合使用[29]。

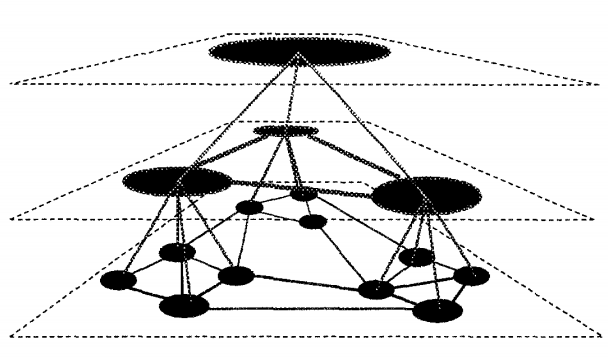


图2.6 聚类布局[28]

**2.2.2 图布局算法的评估**

根据美学标准提出来的针对图布局的评估是一种主流的评估方法[30]，然而美学标准之间可能会互相冲突，图布局中不可能同时实现所有的美学标准。实现其中一种美学标准可能会以消耗大部分其他标准为代价[20]，如图2.7所示，它展示了两种冲突的美学标准[31]，左边的图展示了最大的对称性，但是有更多的边交叉；右边的图有最少的边交叉，然而缺少对称性。冲突的美学标准会造成图布局的整体质量并不确定的现象。为了综合考虑各个美学标准对图布局质量造成的综合影响，Ware[32]提出了一种预估用户寻找图布局中两点之间最短路径的时间的方法，在实验中，用户被要求寻找图布局中两点之间的最短路径，他们完成任务的时间被记录了下来，接着Ware在用户完成任务的时间以及美学标准之间建立了线性关系，对寻找图布局中两点之间最短路径的时间加以预测。然而这种方法获取的是对于图布局中两点之间最小路径这一任务的评价，而不是对图布局质量的总体评估，本文的工作实现了对图布局效果的整体评估。

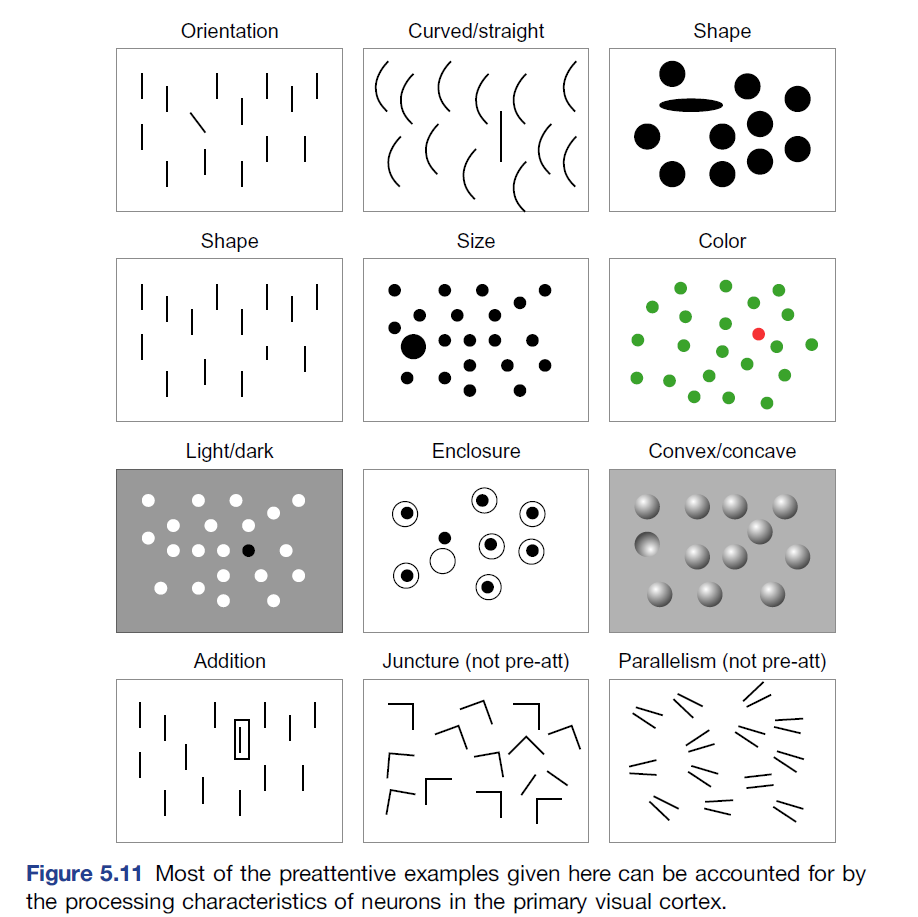


图2.7 一个两种美学标准冲突的例子[31]

## 2.3 视觉在可视化领域的应用

人眼视觉系统(Human Visual System, HVS)的信息处理机制是一个高度复杂的过程，为了研究其原理需要涉及多门学科的交叉领域，包括生物学、解剖学、神经生理学、心理学等多个领域，视觉感知直接影响到人们对信息的认知效果。视觉感知在信息可视化中起着重要的作用，它既可以用来评估可视化的效果，也可以对可视化进行优化。近几年来出现了许多以视觉感知理论为指导的可视化的研究，Colin等[32]提出了一种评估视觉消耗的方法并把它应用到寻找图中两点间最短路径上去，Lee等[33]提出了一种针对分类数据的感知驱动视觉优化的方法，主要研究颜色在可视化中对分类数据的作用，Albuquerque等[34]提出一种基于感知的可视化质量度量技术，用来选择高维数据集的可视化方式，对它们进行优劣排序。

在人们生活中，人的视觉系统总能快速定位显著目标区域，并对其进行识别、分析等任务，这种主动选择性的心理活动被称为视觉关注机制(Visual Attention)[35]。例如图2.8中，在人们看到其中的图片时，注意力很快会被图中的倾斜的线段、曲线中弯曲的线段以及圆中的椭圆等所吸引[36]。视觉关注与跟观察到内容的显著性有很大关联，这被称为是自底向上（数据驱动的）视觉关注机制。数据驱动的视觉关注模型关注颜色、亮度、方向、形状、大小等视觉特征[36]，特征显著的目标就是引起人类注意的目标。本文以数据驱动的视觉关注模型为指导，提出了图布局中节点的视觉重要性模型，这一部分内容会在后续章节中继续描述。

图2.8 视觉注意机制示例图[36]

## 2.4 图像质量评价

为了能够科学地设计客观图布局算法，解决对图布局算法整体评分的可量化的问题，本文作者查阅了许多相关领域的研究成果，发现在图像领域，研究者们早已遇到并且着手解决了类似的问题。因为在图像领域，图像的质量评价也是一个很重要的问题，而且最早的评价方法也都是基于主观实验。类似于图布局评价的主观实验方法，图像质量评价的主观实验方法有难以重复、需要过多的人力物力、无法实时给出结果等缺点。在图像质量评价领域的研究者很早就发现了这个问题，并进行了相关的研究。

在过往的几十年里，随着数字化时代的到来，图像技术飞速发展，图像已成为人们获取信息的重要手段，每天人类产生的图像数量也以很高的速度增长[37]。然而，图像在传输、处理到显示的一系列过程中，都有可能产生质量降低的情况，图像质量对获取的信息的准确度和充分度起着决定性的作用，图像质量评价技术意义重大。

图像质量评估算法可以分为主观图像质量评估算法和客观图像质量评估算法[38]。主观图像质量评估通过主观实验评估图像质量，而客观图像质量评估则通过对人眼视觉特性进行建模，设计出来对应的图像质量评估的算法。其中，根据对原始图像信息的依赖程度，客观图像质量评估算法可分为三类[39]：（1）全参考图像质量评估算法(Full Reference)，需要原始图像的所有信息；（2）半参考图像(Reduced Reference)质量评估算法，需要原始图像的特征信息；（3）无参考图像质量评估算法，不需要原始图像。

无参考的图像质量评价算法设计与本文的图布局客观评价算法类似，根据图像自身的特征对图像进行评估。我们可以参照无参考图像质量评价算法的设计流程(图2.9)，进行图布局评价算法的设计。

在设计无参考图像质量评价算法之前，需要建立一个主观实验结果的数据库，用于训练参数和回归分析。一部分主观实验数据用于训练，另一部分用于评估训练完成后的算法。在训练时，提取的图像特征有：梯度、锐利程度等空域特征，小波、DCT系数等频域特征以及基于图像自然统计的其他特征等。训练完成后，将客观算法在测试集上与主观实验的结果进行对比，主要的对比指标有PLCC，SROCC和RMSE[40]。其中，PLCC用于衡量客观算法预测的准确性；SROCC用于衡量客观算法预测结果与主观算法的结果的排序的一致程度；RMSE用于衡量客观算法与主观实验预测结果的一致性。



图2.9 图像质量评估的具体流程

## 2.5 回归分析

回归分析(regression analysis)是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法[41]。在设计图像评价算法时，常常会使用到回归分析算法来融合多种特征，因此在本文中，参照图像质量评价算法的设计，图布局的质量评价算法也涉及了回归分析的步骤。回归分析在各个领域有着广泛的应用，目前也存在许多回归分析的方法，按照涉及的变量的多少，自变量和因变量之间的关系类型，或因变量的多少，回归分析方法有着不同的分类（图2.10）。本文中主要使用了线性回归和支持向量回归两种回归方法。



图2.10 对回归分析分类的三个度量

**2.5.1 线性回归**

线性回归是统计学中回归分析的一种常用的方法，线性回归[42] (Linear Regression)通过对线性回归方程最小平方函数进行优化进行建模回归分析。我们假设问题是一个线性模型就可以描述的，那么通过建立线性模型，然后通过线性回归的方法拟合模型参数，就可以得到最后的模型。线性模型可以通过下面的公式表示：

 （2.9）

求解线性模型可以通过梯度下降法求解。梯度下降法的流程是，先计算当前的目标函数梯度，然后将梯度乘以步长对目标进行迭代。经过多次迭代可以得到最优解。

**2.5.2 支持向量回归**

支持向量机 (Support Vector Machine, SVM) [43]是基于统计学习理论的一种机器学习方法，在统计样本量较少的情况下也能获取好的统计效果。SVM把样本空间中的输入数据映射到高维特征空间中进行线性分析(图2.11)。支持向量回归(Support Vector Regression, SVR)是支持向量机在解决回归拟合问题上的应用。

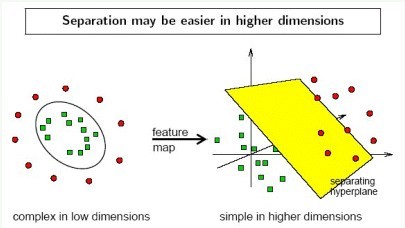


图2.11 svm向高维空间的映射

假设目标函数是，我们想要拟合出该曲线的参数。根据SVM的理论，我们可以得到如下的优化目标：



（2.10）

其中为松弛变量，是为了保证最大间隔分类。

将上述问题转化为对偶问题，然后用SMO算法进行求解就可以算出拟合后的结果。

## 2.6 本章小结

本章首先描述了图布局中节点拓扑结构重要性的定义；其次介绍了图布局这一可视化技术的相关工作以及本文中使用到的几种图布局算法；最后介绍了视觉在可视化领域的应用，图像质量评价的方法以及几种常用的回归分析方法，为后续的工作提供了理论基础。

# 第三章 基于视觉感知的图布局评估

本章阐述了一种基于视觉感知的图布局评估方法，首先，我们介绍了图布局中节点的视觉重要性的想法的引入，并根据用户实验提出了节点的视觉重要性计算模型，接着提出了采用热图和具体量化方法对比不同布局中节点拓扑结构重要性与视觉重要性的差异的方法，最后通过不同图布局之间差异的对比选取出最优的布局。

## 3.1 基本思路

人类感知在可视化领域中起着重要的作用，好的视觉感知能够极大地提升信息展示的数量和质量。在图可视化布局中，由于每个节点具有不同的特征，例如节点的度数不同，周边的节点和边的数目有差异等，基于上文中提到的视觉领域的数据驱动的视觉关注机制，特征显著的目标会先引起人们的注意。用户对不同节点感知的先后顺序是不同的，不同的节点具有不同的视觉重要性。我们定义特征显著、先引起用户注意力的节点为视觉重要性高的节点，反之，则为视觉重要性低的节点。

当用户首先看到某一个节点时，他会倾向认为这个节点更加重要，在实际情况下从拓扑结构去看不一定这样。同样，如果某个节点的视觉重要性较低，但事实上该节点具有较高的拓扑结构重要性，用户也可能倾向地认为这个节点的拓扑结构重要性比较低。在第二章中，我们介绍了节点拓扑结构重要性*T(p)*的计算，*T(p)*包括度中心性、接近度中心性、介数中心性、特征向量中心性四个指标，它们从不同方面展示了节点的拓扑结构重要性。我们希望节点的拓扑结构重要性与其视觉重要性相统一，这样拓扑结构重要性高的节点可以迅速被感知、发现。在网络可视化图中快速发掘网络中重要的节点，具有实际的意义。针对不同情况下的复杂网络，发掘重要节点，并采取相应的措施。例如在传染病网络中，可以快速找到病源，并对其治疗、隔离；在犯罪关系网中，分析出团伙的头目人员，有目标性的进行抓捕。当节点的视觉感知重要性顺序和其拓扑结构重要性顺序一致时，重要节点能够更快的被用户察觉，从而得到更好的可视化展示效果。因此，在图布局中，考虑到节点的拓扑结构重要性，把拓扑结构重要性高的节点放在更易于被用户观察和感知的位置，具有很高的实用价值。

基于以上想法，我们提出了基于视觉感知的图布局评估方法。它适用于用户需要关注图中的重要节点的情形，例如当用户需要关注介数大的节点时，我们应用本文提出的方法比较使用不同图布局时节点介数中心性和视觉重要性之间的差异，选取差异小的图布局方法，即为更符合人类视觉认知、更好的图布局方法。具体评价方法的流程图如图3.1所示。



图3.1 基于视觉感知的图布局评估流程图

## 3.2 用户调查实验分析

表3.1 由不同影响因素设计的用户实验图

|  |  |
| --- | --- |
| 影响因素 | 用户实验图 |
| 1.节点的度数 |  |
| 2.节点周边区域内点的个数 |  |
| 3.节点周边区域内边的个数 |  |
| 4.节点周边区域内边的交点数 |  |
| 5. 节点邻接节点分布的均匀性 |  |

实验通过一个发布在网上的调查问卷实现，实验参与者有20人，包括某高校软件工程专业的本科生和研究生以及社会工作者。参与者的年龄在20岁到50岁之间，参与者的平均年龄为26岁。实验开始首先向参与者介绍图是一个被边连接的节点的集合这一概念，接着要求他们分别选出表3.1的5幅图中更能吸引其注意力的节点是a还是b。最终实验结果统计如表3.2所示。

从实验结果中我们可以看出，节点的度数越高，视觉重要性越高；节点周边区域内的点的个数与节点的视觉重要性呈正相关；节点的视觉重要性随其周边边的数目增大而增大；节点的视觉重要性与其周边区域内边的交点数呈负相关；节点相连的节点分布越均匀，用户对其关注越多，节点的视觉重要性越高。

表3.2 实验结果人数百分比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 人数百分比 | 表3.1.1 | 表3.1.2 | 表3.1.3 | 表3.1.4 | 表3.1.5 |
| 中心节点a | 90% | 25% | 80% | 40% | 90% |
| 中心节点b | 10% | 75% | 20% | 60% | 10% |

## 3.3 节点视觉重要性计算

为了进一步对每个节点的视觉重要性加以量化，我们提出了节点的视觉重要性计算模型。根据用户调查的结果，我们得到了5个影响节点视觉重要性的因子，除此之外，已有眼动仪实验表明[44]图像的中心区域更能吸引用户的注意力，因此在节点距离图布局中心的距离也会影响节点的视觉重要性，越靠近中心区域，节点的视觉重要性越高。在本文中，我们使用下面几个指标：节点的度数*d*，节点周边区域的相邻节点数*n*，节点周边区域的边的数目*e*，节点周边交叉点*c*，节点相连节点分布的均匀性*ne*，以及节点到图布局中心的距离*cd*，对节点的视觉重要性*V*进行量化。

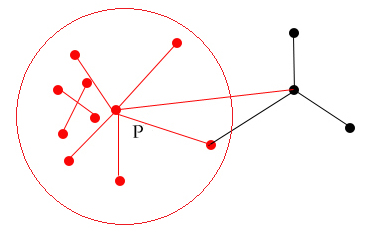


图3.2 节点*P*的周边区域

对于节点P来说，我们把节点*P*周边区域*Area*定义为以*P*为圆心*R*为半径的圆，图3.2所示的红色圆圈即为节点*P*的周边区域。*W*为图布局的宽度，*H*为图布局的高度，我们把图布局中节点所占半径的平均值定义为，因此可提出对节点周边区域的半径*R*的定义如下：

 （3.1）

其中*k*是参数，*k*过大或过小都可能会导致周边区域范围过小或过大，从而导致不同节点周边区域的各参数基本相同。经过实验，我们把*k*取值为2.5，可以根据用户的需求或者不同场景调节*k*的大小。

节点周边区域内的相邻节点数*n*指出现在节点*P*周边区域*Area*中的节点的数目，节点周边区域的边的数目*e*指出现在节点周边区域*Area*的边的数目，节点周边交叉点数目*c*指节点周边区域*Area*内出现的所有边交叉的数目，基于以上设定，各参数的定义如下：

 （3.2）

 （3.3）

 （3.4）

 （3.5）

其中，表示位于节点周边区域*Area*内的节点，表示位于节点周边区域*Area*内的边。

对于节点邻接节点的分布的对称性*ne*，我们通过以下方式进行量化，首先，对于与节点*P*邻接的所有节点，求出它们的质心，具体计算如下：

 （3.6）

接着，计算节点*P*和质心点*pm*之间的距离作为节点的均匀性指标，如图3.3所示，图布局*a*中点*P*周边节点分布不够均匀，值比较大；而图布局*b*中点*P*周边节点分布比较均匀，值为0。值越小说明与该节点相邻的节点分布更为均匀，该节点的视觉重要性越大。

 （3.7）

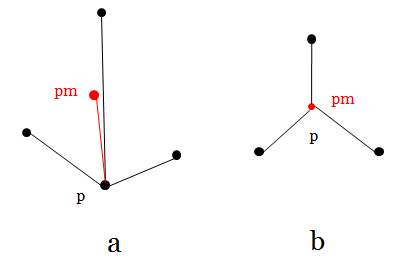


图3.3 邻接节点的分布的对称性

对于节点到图布局中心的距离*cdis*，首先我们定义图布局的中心节点为，则：

 （3.8）

因此，节点到图布局中心的距离*cdis*为:

 （3.9）

节点距离图布局中心的位置越近，*cd*越小，节点的视觉重要性越高。

我们综合考虑以上各因素对视觉重要性*V(p)*的影响，由于不同的节点视觉重要性指标具有不同的数值范围，因此我们先对每一个指标进行归一化处理，归一化可以把不同量级的数据归纳到同一个数量级中，以消除不同指标之间数量级不同的影响。本文中我们使用了min-max标准化（Min-Max Normalization)方法对节点的视觉重要性指标进行归一化。该方法也称为离差标准化，是对原始数据的线性变化，处理后的数据在[0，1]之间，具体公式如下：



（3.10）

其中max为样本数据的最大值，min为样本数据的最小值。

在对节点的各个视觉重要性指标进行归一化后，对于与节点的视觉重要性呈正相关的节点，直接相加，对于与节点视觉重要性呈负相关的因子，用1减去该指标后相加，最后获得节点p的视觉重要性。具体公式如下。

 （3.11）

对于图布局中的每一个节点，我们可以根据以上节点的视觉重要性量化模型，获取节点的视觉重要性。其中视觉重要性高的节点，在图布局中更容易被用户注意到，如果具有高的拓扑重要性的节点恰好具有高的视觉重要性，可以更快地被用户感知、发掘。

## 3.4 节点拓扑重要性与视觉重要性对比

为了观察节点的拓扑重要性和视觉重要性是否统一，这里我们将节点的视觉重要性，分别与节点的度中心性、接近度中心性、介数中心性、特征向量中心性进行对比，如果差异比较小，则说明该节点的视觉重要性与其拓扑结构重要性比较统一。

**3.4.1基于热图的可视化**

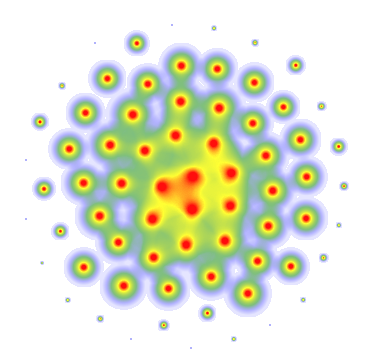
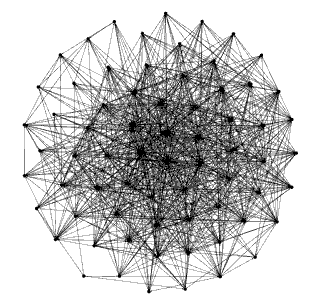
为了展示节点视觉重要性和节点拓扑结构重要性的差异，首先我们采用可视化的展示手段来观测二者大体的差异。我们分别把节点的度中心性、接近度中心性、介数中心性、特征向量中心性、视觉重要性作为节点的权重，并结合可视化布局中根据布局算法获取的节点位置，绘制出图布局的热度图。在热度图中，重要的节点颜色更深，会明显地显示出来。如果根据节点拓扑结构重要性得到的热图和由节点视觉重要性得到的热图比较相似，则其拓扑结构重要性与其视觉重要性更为统一。通过这种方法，用户可以从总体上获取节点拓扑结构重要性与其视觉重要性的差异比对。

由于节点的度中心性、接近度中心性、介数中心性、特征向量中心性这些值的范围较大，直接把它们作为权重在热度图中展示效果不够好，我们对节点的拓扑结构重要性先进行归一化（公式3.10），再乘以参数*f*，这种方法可以缩小数据的绝对数值，便于热图中的展示，我们对以上几个数据做出了如下变换：

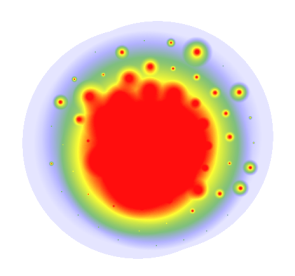
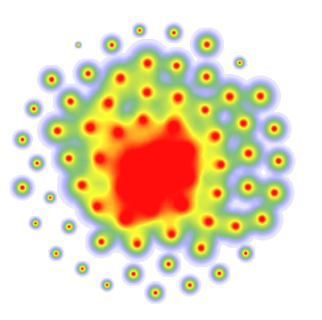
 （3.12）

其中，为归一化之后的结果，本文中取*f*值为60。

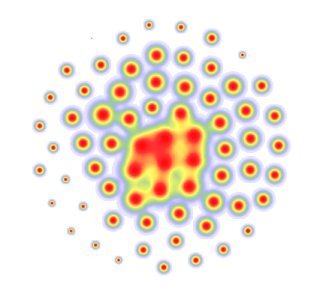
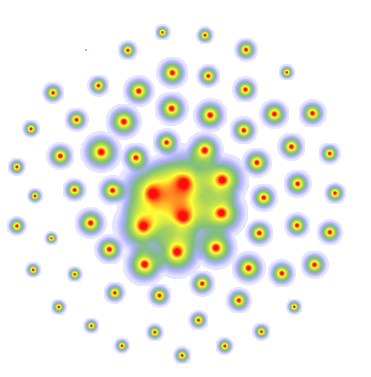
接下来使用变换之后的数据做出对应的热图展示。对于一种图布局算法，用户可以获取对应的五幅热图（如图3.4 b、c、d、e、f所示，图3.4.a为原始图布局）。通过分别将节点的度中心性热图、介数中心性热图、接近度中心性热图同节点的视觉重要性热图进行对比，用户可以从大体上观测到此图布局算法中节点视觉重要性和拓扑结构重要性分布的一致性，通过不同图布局中热图一致性的对比来获取图布局的大体评估。如果图布局中拓扑重要性热图与视觉重要性热图比较一致，则拓扑重要的节点能够得到更好的展示。



a. 图布局 b. 视觉重要性热图



c. 度中心性热图 d. 介数中心性热图



e. 接近度中心性热图 f. 特征向量中心性热图

图3.4 节点重要性热图示意图

**3.4.2 差异量化分析**

除了通过对比结点的视觉重要性热图和拓扑结构重要性（包括节点的度中心性、接近度中心性、介数中心性、特征向量中心性）热图来从大体上获取二者之间的差异外，我们还希望进一步从细节上获取不同图布局下二者差异的大小，从而通过对比选取差异小的图布局。因此我们通过对具体数值差异的量化来实现这一效果。具体步骤是对于每一种拓扑结构重要性（度中心性、接近度中心性、介数中心性、特征向量中心性），将每个节点的视觉重要性与其对应的拓扑结构重要性作差，再对图布局中所有节点的差异求平均值，最后计算出总体的差异值。

由于节点不同的重要性指标值之间的差异较大，我们通过归一化来消除数据指标的不一致、不可比性，以便于我们对不同指标进行分析和对比。首先对每个指标通过上文提到的归一化方法进行处理（公式3.10），再进行进一步的计算，对于图布局中计算节点拓扑重要性与视觉重要性差异具体公式为：

 （3.13）

其中，和分别代表和归一化后的值，对图G中每个节点P计算和的差值并取绝对值得到差异值，最后对所有节点的取平均值，获得整个图布局中和的总体差异值。越大，代表该图布局中节点之间拓扑重要性与视觉重要性差异值越大。对于每种图布局算法，我们可以分别获取度中心性、接近度中心性、介数中心性、特征向量中心性四个指标与节点视觉重要性的整体差异值，表3.3中展示了食物关系网络中采取Fruchterman-Reingold图布局算法和ForceAtlas2图布局算法时的各个拓扑重要性与视觉重要性的差异值的一个示例。不同布局算法中，拓扑重要性与视觉重要性的差异值越小，该算法展示重要节点的效果越好，例如表3.3中，对于度中心性，Fruchterman-Reingold布局展示效果比较好，对于接近度中心性，ForceAtlas2布局展示效果较好。通过对使用不同图布局时拓扑重要性与视觉重要性差异量化值的对比进行图布局的评估，可以挑选出更加符合用户视觉感知的图布局算法。

表3.3 拓扑重要性与视觉重要性差异量化值示例

|  |  |  |
| --- | --- | --- |
|  | Fruchterman-Reingold | ForceAtlas2 |
| 度 | 0.1495 | 0.2504 |
| 介数 | 0.1688 | 0.4602 |
| 接近度 | 0.3301 | 0.2304 |
| 特征向量 | 0.1154 | 0.1596 |

对于不同的布局算法，执行3.4.1、3.4.2中的两步对比对图布局进行评估。通过对比，节点视觉重要性和拓扑结构重要性相差小的布局，评估为优。

## 3.5 本章小结

本章主要介绍了基于视觉感知的图布局评估想法的引入，对图布局中节点视觉重要性影响因子进行的用户调研，图布局中节点视觉重要性模型以及节点的视觉重要性和拓扑结构重要性的对比方法，最终描述了对图布局算法进行评估的标准。

# 第四章 基于回归的图布局质量评估

在本章中，我们通过回归分析得到图布局的整体质量评估，具体做法是把图布局中影响布局效果的指标作为自变量，用户对不同图布局的主观评分作为因变量，目的是确认图布局的主观评分与影响布局效果指标之间的关系。这是一个多元回归分析的问题。在进行回归分析之前，我们先建立一个主观实验结果的数据库，用于训练参数和回归分析。一部分主观实验数据用于训练，另一部分用于评估训练完成后的算法。在训练时，提取的图布局特征有：图密度、边交叉、最小角度、边正交性、节点分布的均匀性、边长一致性。训练完成后，将客观算法在测试集上与主观实验的结果进行对比，如果客观算法在测试集上的结果与主观实验结果越接近则说明客观算法越有价值。与图像质量评价领域一样，我们采用PLCC、SROCC和RMSE三个指标来评估客观算法的好坏。整个评价方法的流程如图4.1所示。

## 4.1 图布局的指标特征

图布局算法评估中有许多公认的影响图布局效果的指标，它们可能会影响用户对图布局整体质量的评分。为了综合考虑各种因素对图布局的影响，我们把这些指标提取出来，作为特征，然后把它们组合起来，作为图布局的整体特征。在这里，我们参考Purchase提出的图布局算法的美学标准[7]，使用了边交叉、最小角度、边正交性、边长一致性这些指标，同时又使用了图密度，并提出了一种对节点分布均匀性进行度量的指标。

对于图*G*来讲，它有*n*个节点，*m*条边，图中的节点用*V*表示。

（1）图密度(graph density)

在数学中，一个稠密图是指一个边的数目接近可能出现的最大边数的图。相反的，如果一个图只有很少的几条边，会被称为稀疏图。图密度的大小决定一个图是稠密的还是稀疏的。对于无边的简单图来说，图密度的定义如下所示：

（4.1）



图4.1 基于回归的图布局整体质量评估的流程图

图的稠密程度可能会影响到图布局效果的整体评分，因此我们把它也作为一个影响图布局评分的特征。

（2）边交叉(edge crossing)

边交叉指标基于*c*，即在整个图布局中边交叉的数目。边交叉即指平面上两条边相交出现交点，整个图布局中边交叉出现得越少，可视化效果越好。为了获取边交叉指标，我们对边交叉数目进行归一化处理，将实际边交叉的数目除以图布局中所有可能出现的边交叉数目。我们首先计算图中所有可能出现的边交叉的数目，对于*m*条线段来讲，*m*条线段可能出现的最多交点的情况是任意两条线段都相交，因此*m*条线段可能出现的最多交点数为：

（4.2）

对图来讲，图中的节点不属于两条边的交点，两条相邻的边是不可能相交的，因此我们需要排除这种情况，不可能的边交叉的数目如下：

（4.3）

其中，指图中第*j*个节点。因此，图布局中所有可能出现的边交叉数目如下：

（4.4）

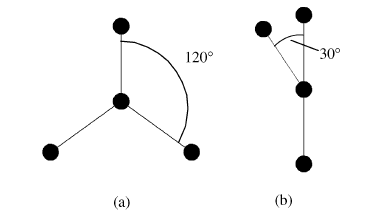
图布局的边交叉指标为：

（4.5）

的值在0到1之间，0代表最小的边交叉，1代表出现最多可能的边交叉。图布局的边交叉指标越大，图布局的可读性越差。

（3）最小角度(Minimum Angle)

图布局的最小角度指标基于，即第*i*个节点的最优最小角度。对于一个节点来说，当其邻接边在节点周围均匀分布时，可视化效果最好，如图4.2a所示，此时，节点邻接的边中，任意相邻两条边之间的角度都是相同的。



a b

图4.2 图布局最小角度指标示意图

节点的最优最小角度的计算如下：

（4.6）

其中，为节点的度数。因此对于图4.2b中的中心节点最优最小角度为。最小角度指标用来衡量实际情况与最优分布之间的差距，即节点边的偏离程度与其最佳偏离程度之间的差距。具体的计算如下：

（4.7）

是第*i*个节点中邻接边的实际最小角度。首先计算出图布局中每个节点实际最小角度与最优最小角度的差异指标值，再求出所有节点差异指标值的平均值，即该图布局的最小角度指标。值越小，图布局的可读性越好。

（4）边正交性(Edge Orthogonality)

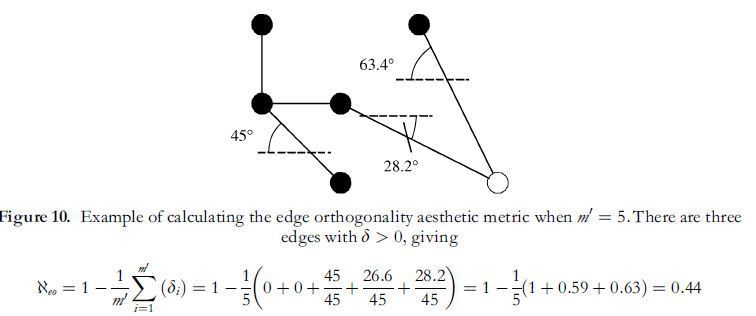


图4.3 图布局边正交性指标示意图

边正交性展示了图布局中边偏移正交线的程度，对于第*i*条边来说，它的边偏移因素代表了该边偏移正交线的程度。图4.3是对计算的一个示例，该图布局中三条边的偏移正交线的角度分别为，和。用来计算该边偏移水平线或垂直线的程度：

（4.8）

对于一个图布局来说，我们先计算每条边的偏移度，再求所有边的偏移度的平均值，计算如下：

（4.9）

越大，表明图布局中边偏移正交线越多，可视化效果越差。

（5） 节点分布的均匀性(Uniform node distribution)

在图布局中，节点是集中分布还是分散分布也对图布局的效果有着很大的影响，为了把这一特征抽象出来，我们分别统计节点在*x*轴和*y*轴方向上的分布。首先，我们分别取图布局中所有节点*x*坐标的最小值和最大值以及y坐标的最小值和最大值，把最小值和最大值之间均分为10段，分别统计落在每一段区间中的节点数目*Ni*，可以得到*x*轴方向和*y*轴方向展示图布局中节点分布均匀性的两组统计值。表4.1展示了统计值的一个示例。

表4.1 节点在图布局中x轴、y轴方向的分布情况

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **段1** | **段2** | **段3** | **段4** | **段5** | **段6** | **段7** | **段8** | **段9** | **段10** |
| **X轴** | 4 | 8 | 12 | 7 | 7 | 9 | 7 | 7 | 4 | 4 |
| **Y轴** | 1 | 4 | 10 | 9 | 9 | 8 | 6 | 6 | 8 | 8 |

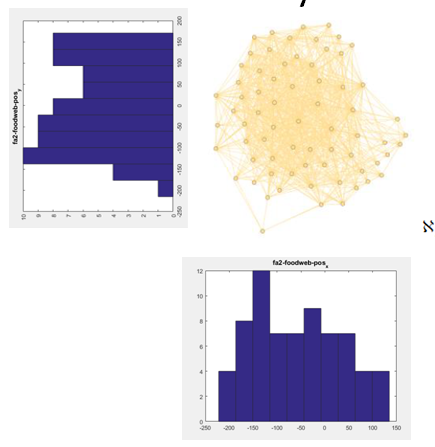


图4.4 图布局中节点分布的条形图展示

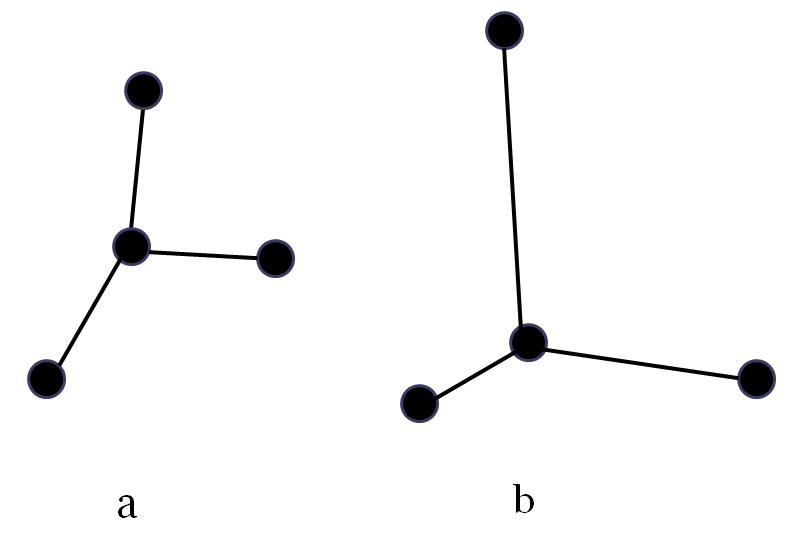
我们通过条形图对点的分布进行可视化展示，用户可以从条形图上观测到节点的分布是否均匀，如果条形图中每个矩形的高度相差不大，则说明节点在空间上分布比较均匀。图4.4展示了一个图布局中节点分布的条形图展示，其中，位于图布局下方的条形图展示了节点在*x*轴方向的分布，位于图布局左方的条形图展示了节点在*y*轴方向的分布。然而，条形图的分布只能获取视觉上的感知，为了进一步得到反映节点分布均匀性的量化的指标，作为数值便于下一步的使用。我们提出图布局中节点在*x*轴和*y*轴上分布的均匀性指标如下：

（4.10）

、为统计值的均值，为*x*轴，*y*轴上节点的空间分布统计值的标准差，、为数据的变异系数(Coefficient of Variation)，它可以消除平均数不同对多组数据离散程度比较的影响，用来反映数据的离散程度，即节点在*x*轴和*y*轴上分布的均匀性，、越小，则图布局中节点分布越均匀。

（6） 边长度的均一性(Uniform Edge Length)

一致的边的长度使得图布局更加容易使图布局中的形状均匀，比长短不一的边造成的复杂形状更加易于识别，因此可能会有更好的可视化效果。例如图4.5中，图布局a中边长更为一致，可读性比图布局b更好。

图4.5 图布局中边长度一致性指标的示意图

对于图布局中边长度的均一性指标，首先，我们计算出图布局中每条边的长度，然后对不同边长度的差异进行分析。我们使用以下计算方法对图布局中边长的均匀性进行量化：

（4.11）

其中，为图布局中所有边长的方差，为图布局中所有边长均值。越小，图布局中边长分布越均匀，图布局效果越好。

## 4.2 图布局的主观质量评价

我们参照图像质量的主观评价方法，综合大量观看者对图布局的评分得到图布局的最终质量分数。主要的评分过程可分为以下几步：（1）搭建评价环境，使用在线调查问卷的形式请用户评分；（2）请被测试用户浏览完同一个图结构的四种布局方法；（3）看完四种布局方法后，根据直观感受以及个人经验进行评分；（4）综合考虑所有用户的评分获取图布局的最终质量得分。我们常使用平均意见分数(Mean Opinion Score，MOS)作为图布局质量综合考虑的计量方式，即取所有观察者打分的平均值。

根据评分过程，我们首先需要搭建图布局的数据库。在本文中，我们建立了一个含有60幅（15\*4）参考图布局的数据库（见附录）。首先使用Gephi随机生成15个图结构，我们选择生成不同节点规模、边规模、图密度的图结构。15个图的结构信息如表4.2所示。接着对于每个图结构分别使用Fruchterman-Reingold算法、ForceAtlas2算法、Openord算法和Yifanhu算法4种图布局算法生成4幅图布局，例如对于图1来讲图1-1到图1-4分别为其通过Fruchterman-Reingold算法、ForceAtlas2算法、Openord算法和Yifanhu算法生成的图布局。接下来邀请用户对图布局进行评分。

表4.2 主观实验中图结构的基本信息

|  |  |  |  |
| --- | --- | --- | --- |
|  | **节点数目** | **边数目** | **图密度** |
| **图1** | 19 | 29 | 0.170 |
| **图2** | 10 | 18 | 0.400 |
| **图3** | 40 | 118 | 0.151 |
| **图4** | 120 | 368 | 0.052 |
| **图5** | 10 | 15 | 0.333 |
| **图6** | 10 | 33 | 0.733 |
| **图7** | 10 | 27 | 0.6 |
| **图8** | 20 | 39 | 0.205 |
| **图9** | 20 | 84 | 0.442 |
| **图10** | 20 | 112 | 0.589 |
| **图11** | 40 | 77 | 0.099 |
| **图12** | 40 | 87 | 0.112 |
| **图13** | 40 | 210 | 0.269 |
| **图14** | 80 | 314 | 0.099 |
| **图15** | 20 | 54 | 0.284 |

表4.3 图布局主观质量评价评分表

|  |  |  |
| --- | --- | --- |
| **级别** | **绝对度量评估** | **相对度量评估** |
| **5** | 最好 | 一批中最好的 |
| **4** | 较好 | 好于该批的平均水平 |
| **3** | 一般 | 该批的平均水平 |
| **2** | 较差 | 差于该批的平均水平 |
| **1** | 最差 | 一批中最差的 |

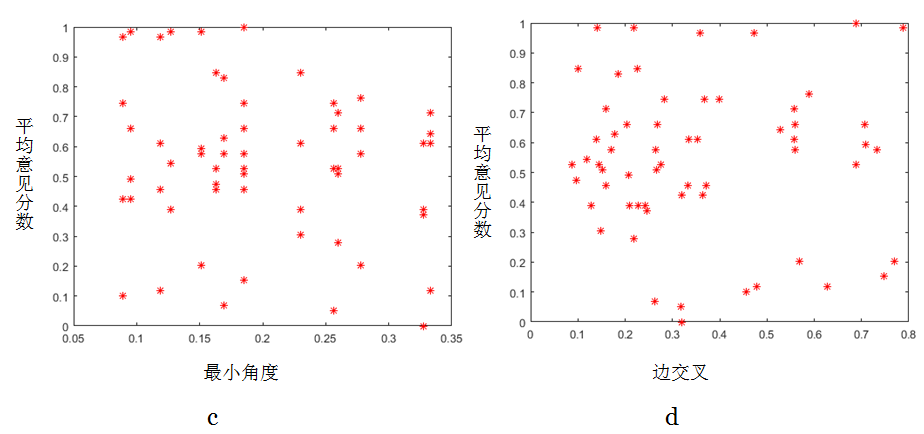
表4.4 图布局的平均意见分数

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 图1-1 | 图1-2 | 图1-3 | 图1-4 | 图2-1 | 图2-2 | 图2-3 | 图2-4 | 图3-1 | 图3-2 |
| **评分** | 3.64 | 3.05 | 2.00 | 3.00 | 4.05 | 3.77 | 2.55 | 3.55 | 4.23 | 3.55 |
|  | 图3-3 | 图3-4 | 图4-1 | 图4-2 | 图4-3 | 图4-4 | 图5-1 | 图5-2 | 图5-3 | 图5-4 |
| **评分** | 2.18 | 3.68 | 4.27 | 3.41 | 3.23 | 3.27 | 3.91 | 3.73 | 2.32 | 3.64 |
|  | 图6-1 | 图6-2 | 图6-3 | 图6-4 | 图7-1 | 图7-2 | 图7-3 | 图7-4 | 图8-1 | 图8-2 |
| **评分** | 4.64 | 3.55 | 2.55 | 3.59 | 4.68 | 3.77 | 2.41 | 3.41 | 4.00 | 3.77 |
|  | 图8-3 | 图8-4 | 图9-1 | 图9-2 | 图9-3 | 图9-4 | 图10-1 | 图10-2 | 图10-3 | 图10-4 |
| **评分** | 2.14 | 3.41 | 4.59 | 3.64 | 2.32 | 3.23 | 4.59 | 4.00 | 2.27 | 3.14 |
|  | 图11-1 | 图11-2 | 图11-3 | 图11-4 | 图12-1 | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 |
| **评分** | 3.91 | 3.41 | 2.05 | 3.36 | 4.27 | 3.64 | 3.05 | 2.82 | 4.64 | 3.77 |
|  | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 | 图12-2 |
| **评分** | 3.14 | 3.32 | 4.64 | 3.45 | 3.05 | 3.05 | 4.00 | 3.55 | 3.23 | 3.36 |

本篇论文中图布局的主观质量评分由22个观察者打分完成，包括某高校软件工程专业的本科生和研究生以及社会工作者。参与者的年龄在20岁到50岁之间，参与者的平均年龄为26岁。具体评分过程是通过一个发布在网上的调查问卷实现的。实验开始首先向参与者介绍图是一个被边连接的节点的集合这一概念，接着介绍评分时使用五级评分的质量尺度，评分从1-5，表示从不满意到满意，具体度量如表4.3所示。

接下来观察者被要求在观测完同一个图结构的四种布局方式后分别对四幅图进行评分，依次对60幅图布局打分，我们取所有观察者打分的平均值作为每个图布局的主观质量评分。最后得到的60幅图布局的主观评分统计结果如表4.4所示，通过这种方式，我们建立了图布局主观质量评分的数据库。在进行下一步的回归分析前，对于图布局的主观质量评分，我们首先使用公式3.10对其进行归一化，以消除不同评分数量级不同的影响。

## 4.3 回归分析



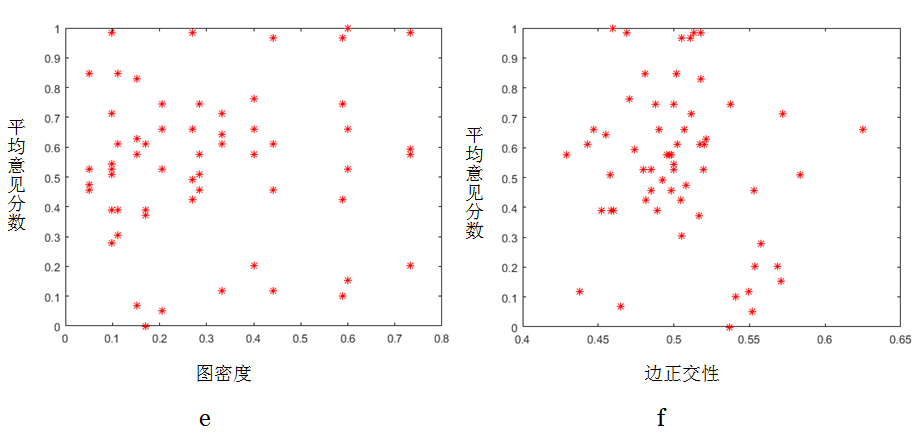
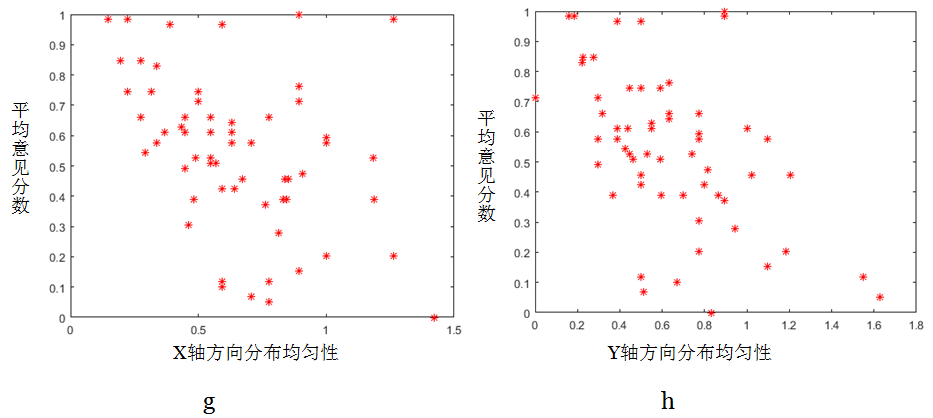
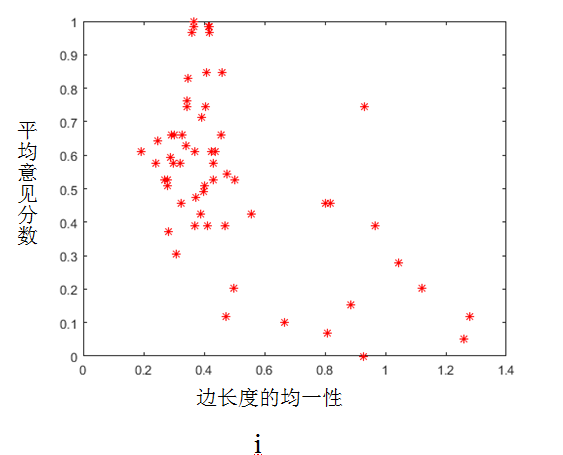


图4.6 图布局主观评分与各指标的散点图

在本文中，我们把图布局中影响布局效果的指标作为自变量，用户对不同图布局的主观评分作为因变量，目的是确认图布局的主观评分与影响布局效果指标之间的关系，这是一个回归分析的问题。

我们先通过散点图观察图布局主观评分和单个图布局效果指标之间的关系，如图4.6所示。通过散点图可以发现，图布局主观评分和任意一个图布局效果指标之间并不存在直接的线性关系，图布局的主观评分和多个图布局效果指标相关。因此我们要解决的是一个多元回归的问题，通过多个指标对图布局主观评分进行回归拟合。常用的回归分析模型有线性回归、主成分分析、支持向量回归等。

本文中我们比较了多种回归方法。用到了线性回归以及SVR解决这个回归问题，其中，线性回归通过最小二乘法进行拟合，SVR进行回归时的流程如图4.7所示。核函数的选择，SVR的参数调整以及模型的效果评估为支持向量回归的重要步骤，对回归的结果有着重要影响。下文中对这三个重要步骤的流程加以介绍：



图4.7 模型选择流程图

（1）核函数的选择

核函数的作用是实现训练集的输入维度到高维特征空间之间的映射，核函数的选择对SVR的性能有着很大的影响。常用的核函数有线性核函数(Linear kernel function)、高斯核函数(Gaussian kernel function)和多项式核函数(Polynomial kernel function)。

（2）支持向量回归的参数选择

给定一种核函数，输入空间到高维特征空间的映射会随着核参数的改变而变化，例如，对于多项式核函数来说，次数可以取2或3。不同的核参数也会影响到SVR模型的性能。我们使用不同的核参数，并选择出最优的参数。

（3）模型的效果评估

我们在（1）、（2）中选取不同的核函数和核参数，可以得到不同的回归模型，通过对不同回归模型的性能进行对比，选取最优的回归模型。

在解决本文的回归问题时，为了获取最优的回归结果，我们分别使用线性回归以及线性核函数、高斯核函数和多项式核函数进行支持向量回归(表4.5)，得到不同的回归模型。通过对采用不同回归方法时得到的回归模型效果的评估和对比，选取最优的回归模型。

表4.5 SVR中不同的核函数

|  |  |  |
| --- | --- | --- |
| 核函数 | 描述 | 公式 |
| 高斯核函数 | Gaussian核 |  |
| 线性核函数 | Linear核 |  |
| 多项式核函数 | Polynomial核 |  |

经过不同回归模型之间的对比，我们选取出来了其中最优的回归模型对图布局的整体质量进行评分。对于一个给定的图布局算法，通过计算图布局算法的各项指标值，应用回归模型计算得到此图布局算法的整体质量评分，从而通过不同图布局算法之间评分的对比，选取最优的图布局算法。

## 4.4 本章小结

本章介绍了基于回归的图布局整体质量评估的具体过程，主要包括图布局的指标特征的计算，图布局的主观质量评价方法，回归分析的具体方法。在下文中我们将通过具体的实验对回归模型的结果进行验证。

# 第五章 实验分析

本章节通过具体实验对本文中提出的两种图布局评估方法加以应用。对于基于视觉的图布局评估，我们采用具体数据集来应用评估具体的图布局算法。对于基于回归分析的图布局质量评价方法，我们通过对测试集中由回归模型得到的图布局质量评分同实际的图布局主观质量评分进行对比，选择合适的回归模型并对图布局主观质量评价方法的正确性进行验证。

## 5.1 基于视觉感知的图布局评估

本文选取了2个不同规模的网络结构数据集，依次使用Gephi中的Fruchterman-Reingold算法、ForceAtlas2算法、Openord算法和Yifanhu算法获取图布局，并通过对节点视觉重要性与其拓扑结构重要性的对比，对不同的布局算法进行评估，接下来我们对文中的两个用例进行介绍。

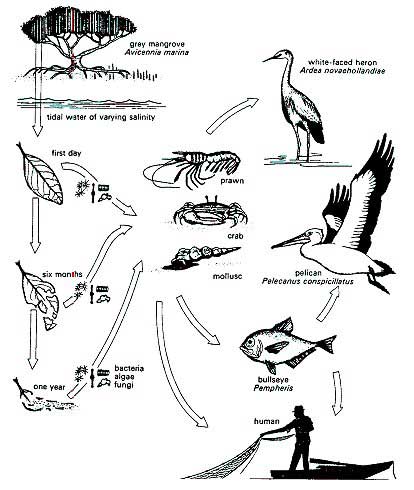
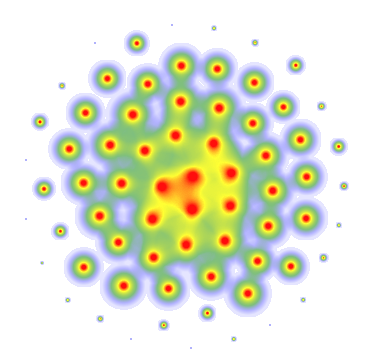
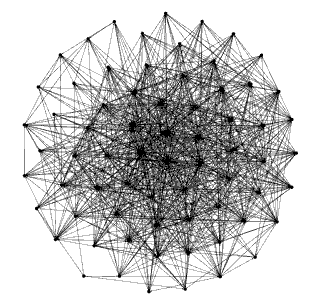
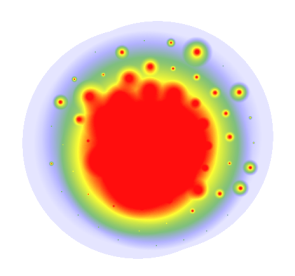
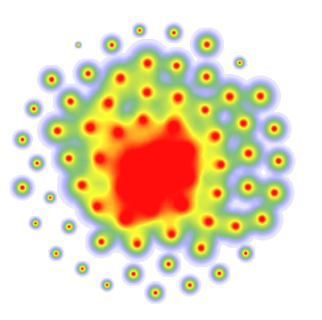
**5.1.1 食物关系网络**

图5.1 食物关系网示意图

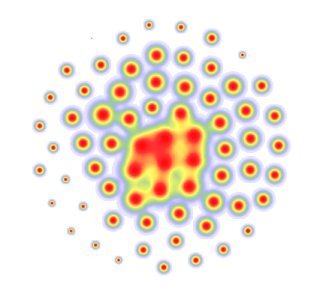
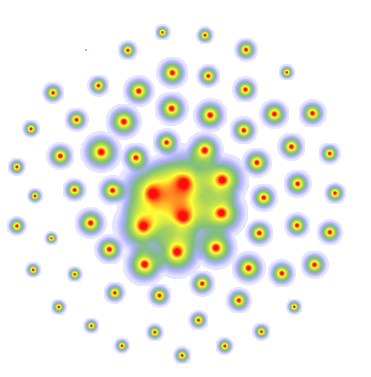
食物网也被称为食物链，生态系统中不同物种间的捕食关系是一个网络，图5.1展示了食物网的一个示例[[3]](#footnote-3)。在食物网中，物种之间会相互影响，一个物种的灭绝可能会影响到整个生态系统。例如，当被猎食者物种丧失之后，它的捕食者会有更大的灭绝的可能性；同样，捕食者的灭绝也会对被猎食的物种产生影响，因为捕食者的灭绝可能会增加猎物之间的竞争，使其中较弱的竞争者更容易遭受灭绝。而这些物种的损失很有可能进一步对其他猎物和捕食者物种造成影响，从而对整个食物网络造成重大影响。物种之间的相互关系即其在网络中的拓扑位置决定了可能产生最大影响的物种，影响力大的物种会对所在地区的整体物种分布产生重要的作用[45]。因此，发掘食物网络中的重要节点，对生态系统的研究意义重大。



a. 图布局 b. 视觉重要性热图



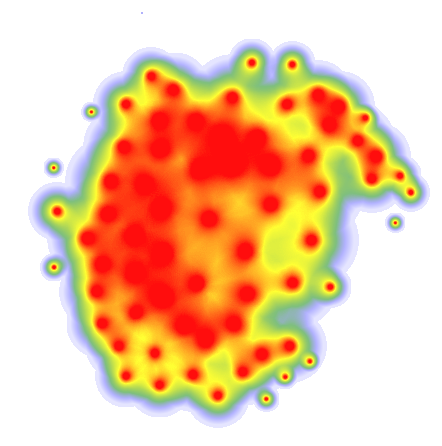
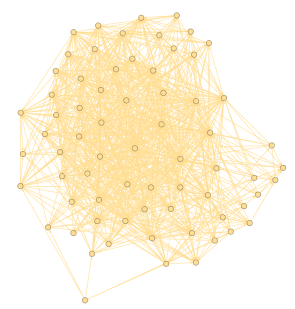
c. 度中心性热图 d. 介数中心性热图



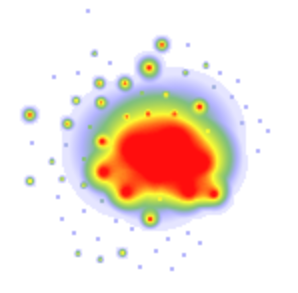
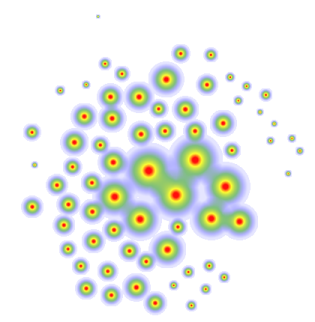
e. 接近度中心性热图 f. 特征向量中心性热图

图5.2 食物关系网Fruchterman-Reingold布局热图

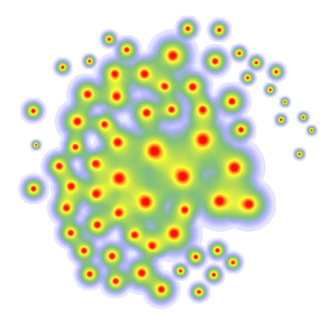
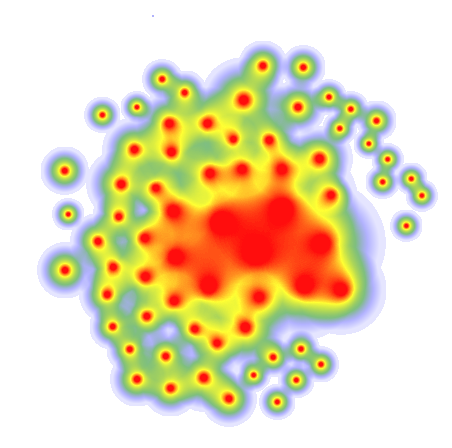
用例1采用的数据是一个食物关系网络(food web)，它来自于一个生态网络分析的网站，是对美国南佛罗里达洲的营养动力学分析，它展示了美国南佛罗里达罗里达洲不同物种的动植物间的捕食关系[46]。我们使用的是众多食物关系网中的一个沼泽地食物网络[47]，它含有69个节点916条边。我们分别使用Gephi中提供的Fruchterman-Reingold算法、ForceAtlas2算法、Openord算法和Yifanhu算法生成图布局。图5.2.a展示的是Fruchterman-Reingold算法生成的节点布局，接下来对该食物关系网计算图中每个节点的拓扑结构重要性（度中心性、接近度中心性、介数中心性、特征向量中心性）和视觉重要性，把它们作为热图绘制时节点的权重，由此可以得到各个热图（图5.2.b-5.2.f）。



a. 图布局 b. 视觉重要性热图

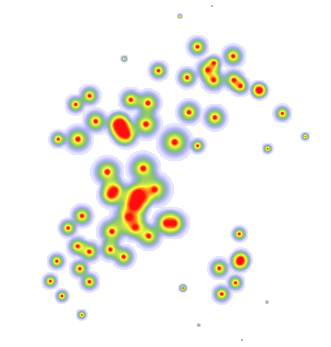
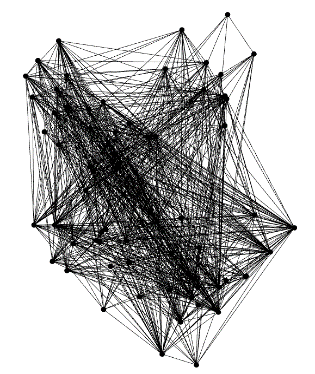


c. 度中心性热图 d. 介数中心性热图

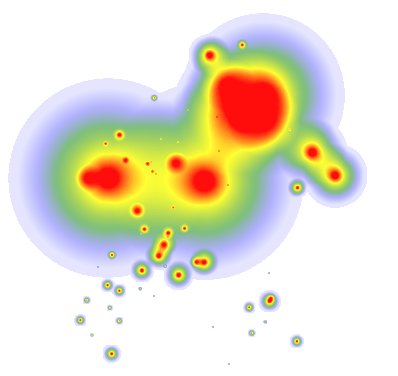
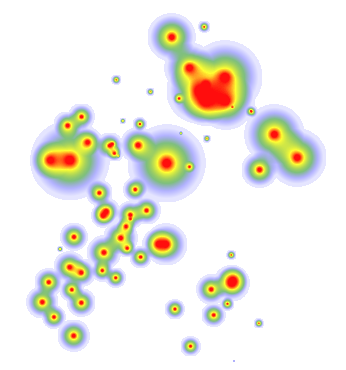


e. 接近度中心性热图 f. 特征向量中心性热图

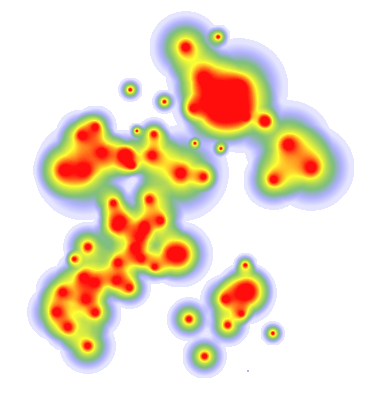
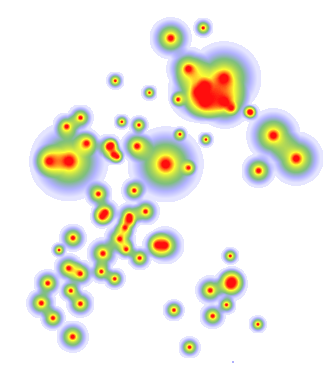
图5.3 食物关系网ForceAtlas2布局热图



a. 图布局 b. 视觉重要性热图



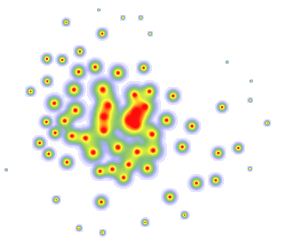
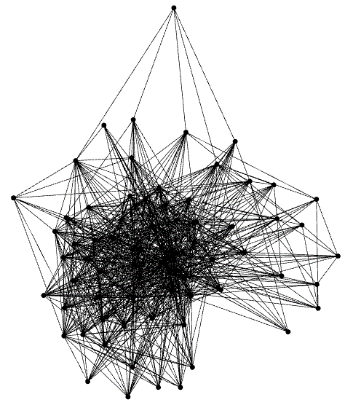
c. 度中心性热图 d. 介数中心性热图



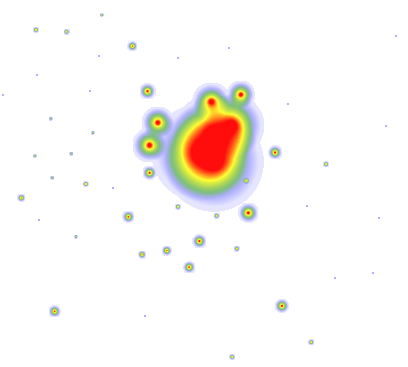
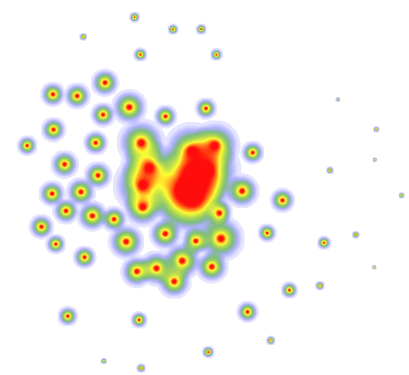
e. 接近度中心性热图 f. 特征向量中心性热图

图5.4 食物关系网Openord布局热图

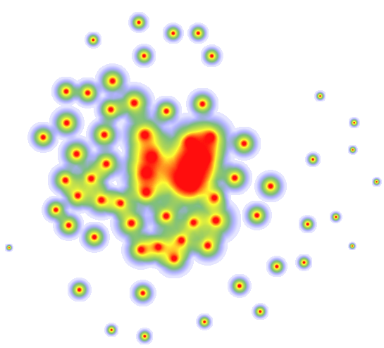
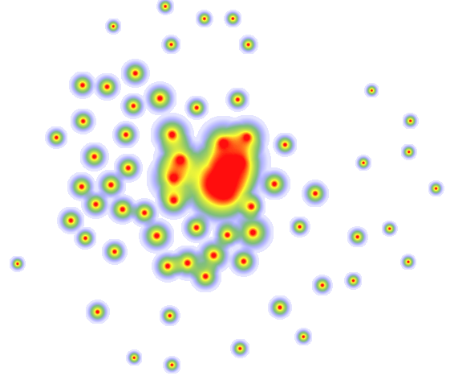
图5.3-5.5则分别展示了使用ForceAtlas2算法、Openord算法和Yifanhu算法时的布局效果和热图。通过以上各幅图的对比，我们对节点拓扑结构重要性与视觉重要性的分布有了大体的认识。从热图中的整体效果来看，视觉重要性热图与度中心性热图、特征向量中心性热图比较一致，而介数中心性热图和接近度中心性热图与视觉重要性热图的一致性稍微差一些。



a. 图布局 b. 视觉重要性热图



c. 度中心性热图 d. 介数中心性热图



e. 接近度中心性热图 f. 特征向量中心性热图

图5.5 食物关系网Yifanhu布局热图

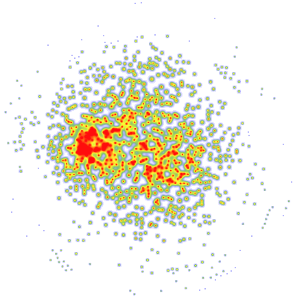
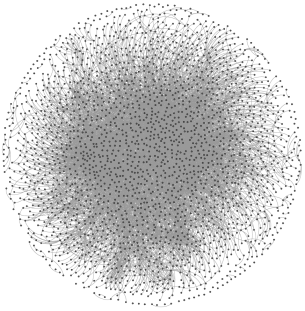
表5.1 食物关系网络拓扑重要性与视觉重要性差异量化值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Fruchterman-Reingold | ForceAtlas2 | Openord | Yifanhu |
| 度中心性 | 0.1495 | 0.2504 | 0.3395 | 0.1476 |
| 介数中心性 | 0.1688 | 0.4602 | 0.3121 | 0.1374 |
| 接近度中心性 | 0.3301 | 0.2304 | 0.5147 | 0.3718 |
| 特征向量中心性 | 0.1154 | 0.1596 | 0.2721 | 0.0808 |

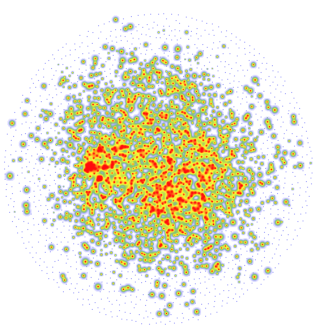
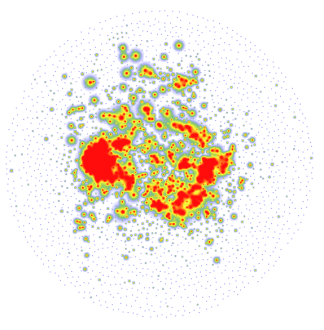
接下来，将节点的拓扑结构重要性与其视觉重要性之间的差异进一步量化展示，如表5.1所示。从表中来看，Yifanhu布局在展示节点的度中心性、介数中心性以及特征向量中心性时，要优于其他三种算法，ForceAtlas2布局则在对节点的接近度中心性进行呈现时具有更好的效果。因此，对于本案例中的食物网，要研究节点的度中心性、介数中心性以及特征向量中心性时使用Yifanhu布局展示效果较好，要关注节点的接近度中心性时使用ForceAtlas2布局较好。

**5.1.2 蛋白质关系网络**

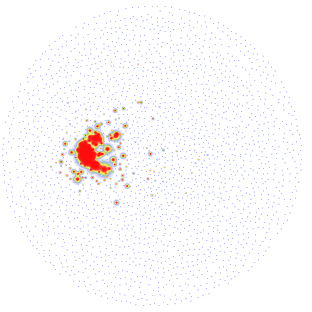
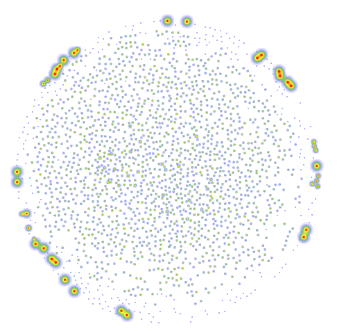
蛋白质之间通过相互作用关系组成一定的结构功能来发挥作用，现有的检测方法已经发现，不同的蛋白质之间有成千上万的关联关系，蛋白质之间相互作用的网络被称为蛋白质关系网络(Protein-Protein Interaction Network)。研究发现随机删除蛋白质关系网络中的几个节点并不太影响网络的整体结构，相反的，删除某几个高度连通的节点极大地改变了网络的结构，网络中少数重要的节点对整个网络的影响大于大部分不重要的节点，对蛋白质关系网络至关重要[48]。因此在蛋白质网络中发掘重要节点有着很高的实用价值。



a. 图布局 b. 视觉重要性热图

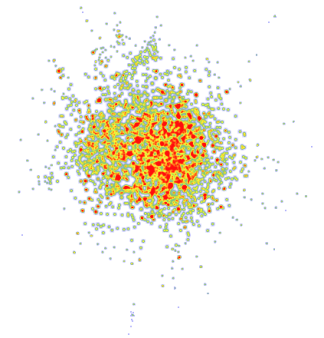


c. 度中心性热图 d. 介数中心性热图

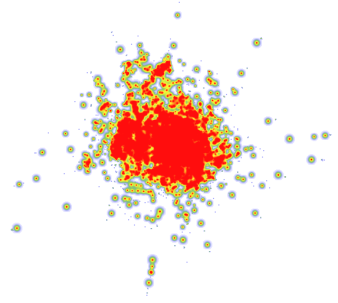
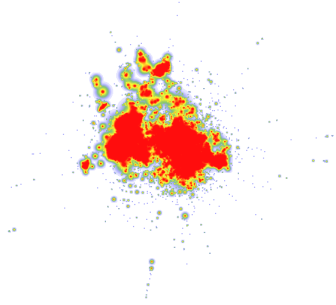


e. 接近度中心性热图 f. 特征向量中心性热图

图5.6 蛋白质关系网Fruchterman-Reingold布局热图



a. 图布局 b. 视觉重要性热图



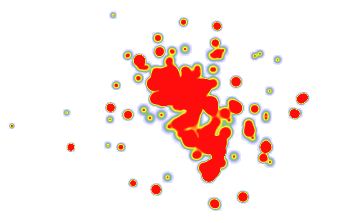
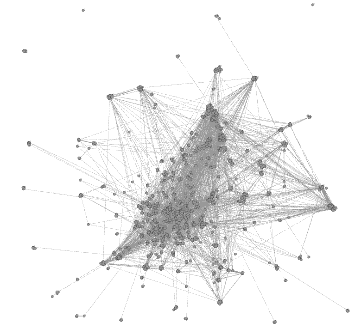
c. 度中心性热图 d. 介数中心性热图



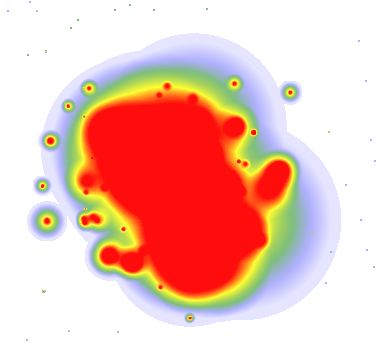
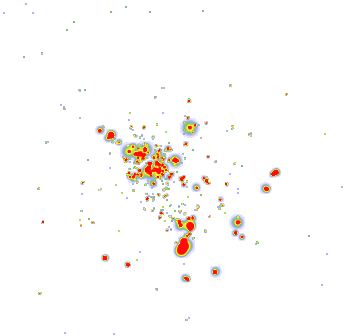
e. 接近度中心性热图 f. 特征向量中心性热图

图5.7 蛋白质关系网ForceAtlas2布局热图

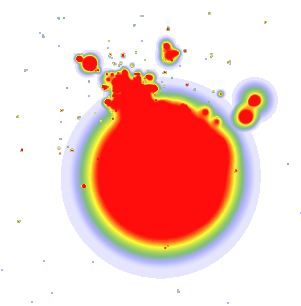
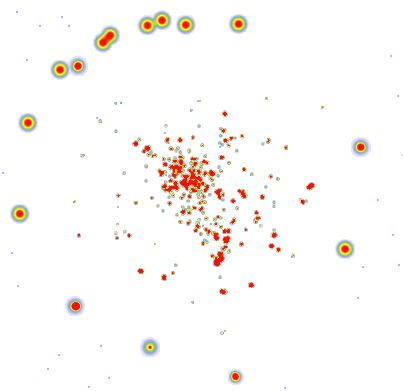
用例2中采用了一个具有2361个顶点，7182条边的蛋白质关系网络[49]。该网络分别通过Fruchterman-Reingold算法、ForceAtlas2算法、Openord算法和Yifanhu算法获取的效果图如图5.6、5.7、5.8、5.9所示。从热图来看，四种布局中视觉重要性热图和度中心性热图、介数中心性热图都比较一致，而ForceAtlas2布局中接近度中心性热图和视觉重要性热图比较一致。在该数据的差异表（表5.2）中，ForceAtlas2布局在展示节点的接近度中心性时效果更好，Fruchterman-Reingold布局则能更好的展示节点的度数中心性以及特征向量中心性，Openord布局中节点的介数中心性和视觉重要性更为一致。因此在本案例中，要研究节点的度数中心性以及特征向量中心性采用Fruchterman-Reingold布局比较好，关注节点的接近度中心性时宜采用ForceAtlas2布局，重视节点的介数中心性时应使用Openord布局。



a. 图布局 b. 视觉重要性热图



c. 度中心性热图 d. 介数中心性热图

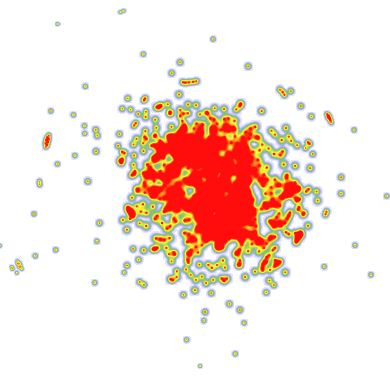
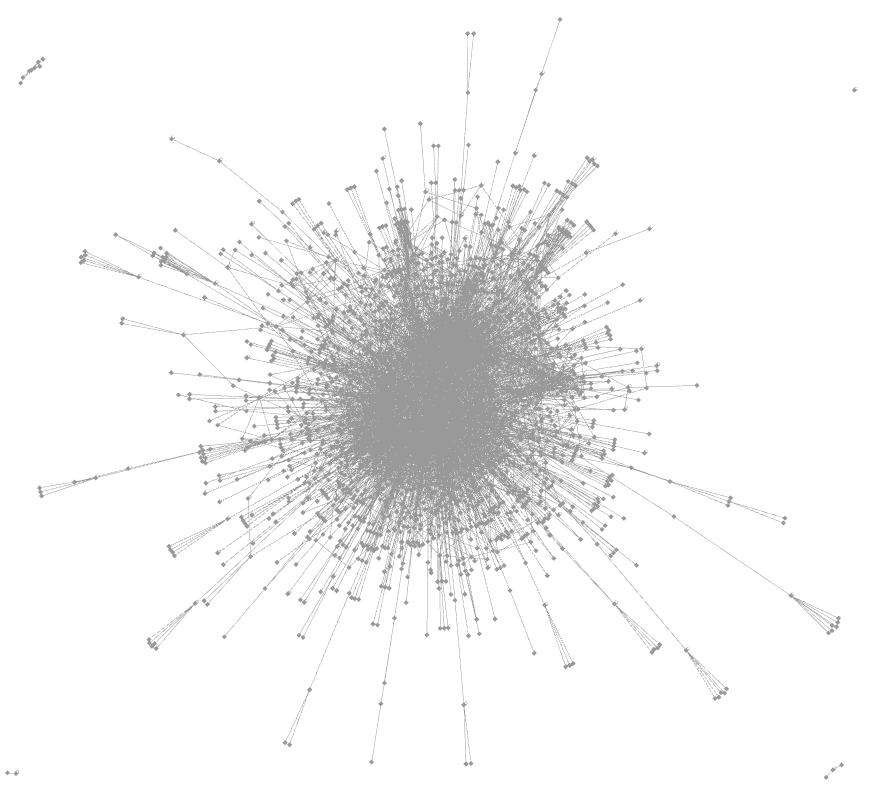


e. 接近度中心性热图 f. 特征向量中心性热图

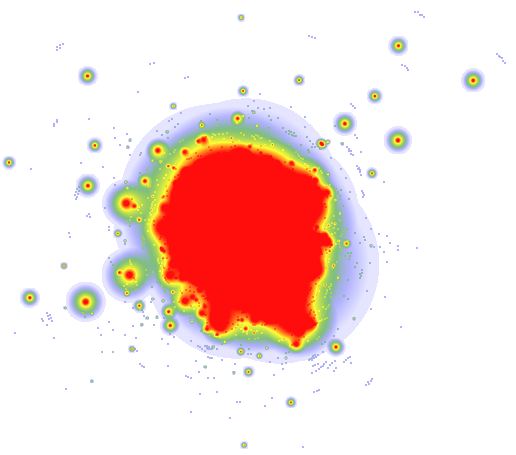
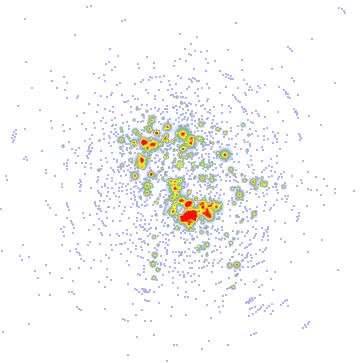
图5.8 蛋白质关系网Openord布局热图

表5.2蛋白质关系网络拓扑重要性与视觉重要性差异量化值

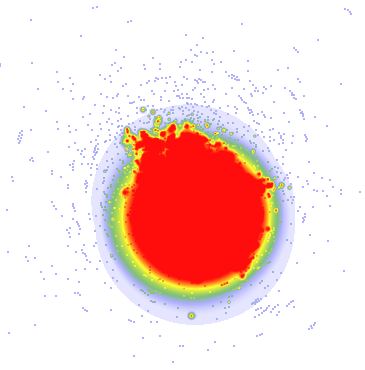
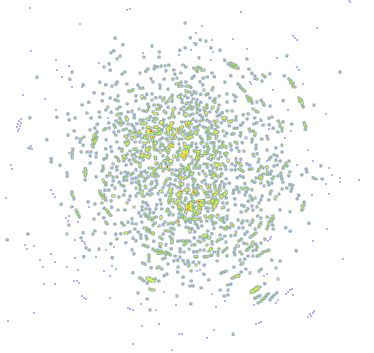
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | FR | ForceAtlas2 | Openord | Yifanhu |
| 度中心性 | 0.4180 | 0.6461 | 0.5905 | 0.5915 |
| 介数中心性 | 0.4952 | 0.7244 | 0.4895 | 0.4932 |
| 接近度中心性 | 0.3249 | 0.2472 | 0.6684 | 0.6698 |
| 特征向量中心性 | 0.4966 | 0.7272 | 0.6712 | 0.6726 |



a. 图布局 b. 视觉重要性热图



c. 度中心性热图 d. 介数中心性热图



e. 接近度中心性热图 f. 特征向量中心性热图

图5.9 蛋白质关系网Yifanhu布局热图

**5.1.3 综合对比**

综合观察分析案例1和案例2通过Fruchterman-Reingold算法、ForceAtlas2算法、Openord算法和Yifanhu算法进行可视化布局时各节点拓扑重要性与视觉重要性之间的差异，我们发觉ForceAtlas2布局中节点接近度中心性与视觉重要性比较接近，而节点的度中心性、介数中心性和特征向量中心性在不同的案例中适用的最优布局不同。

以上两个案例展示了在对具体数据集进行应用时，采用基于视觉的图布局评估方法评估、选择不同图布局的做法，对图布局的评估不限于上文中使用的四种布局。在实际应用中，首先要明确研究目的，关注点为哪一种拓扑重要性，接下来对于备选的图布局，对比其拓扑重要性与视觉重要性的差异，差异最小的图布局即为最符合人类视觉特性的图布局。

## 5.2 基于回归的图布局整体质量评估

图布局评价算法基于特征提取和机器学习，实现的时候我们先把图布局效果按内容随机分为两类，第一类是训练数据，包括整个图布局数据库67%的数据（40幅图），第二类是测试数据，包括另外33%的数据（20幅图）。图布局评价算法在第一类数据集上进行训练得到回归模型，在第二类数据集上测试，分析模型给出的客观图布局质量评价与实际主观图布局质量评价之间的差异，进行效果的评估，具体过程如图5.10所示。



图5.10 图布局质量评价算法的性能评价框架

本文中我们分别使用线性回归和支持向量回归，对于支持向量回归使用线性核函数、高斯核函数和多项式核函数进行拟合，得到不同的回归模型。利用这些回归模型对测试数据中图布局的指标进行回归，得到预测的图布局总体评分。为了判断回归模型的效果，我们使用不同方法对比预测的图布局的总体评分与主观实验中实际图布局的总体评分，对客观质量评价方法的性能进行评估。

**5.2.1 指标评估**

为了判断回归模型的效果，我们使用不同的指标将预测的图布局的总体评分与主观实验中实际图布局的总体评分进行比对。首先，我们定义为用户对图布局的主观质量评价，为通过本文中的整体质量评估模型得到的图布局质量评价值，主要使用了以下几种指标：

（1）SROCC(Spearman’s rank ordered correlation coefficient)

SROCC反映2组变量之间联系的密切程度，即Spearman秩相关系数:

（5.1）

其中，和分别为和在各自数列中排序的位置。SROCC取值在区间[-1,1]之间，其绝对值越接近1，表明预测单调性越高，它用于衡量算法预测的单调性(Monotonicity)。

（2）RMSE (Root mean squared error)

RMSE表示客观评分对主观评分的估计值主观评分MOS之间的误差，均方根误差比较算法评价值与用户对图布局质量主观评分之间的绝对误差，来衡量算法预测的准确性(Accuracy)。由于质量评分的取值范围有所不同，需要在计算指标之前进行归一化处理。误差值越小，说明该客观评价方法越好。

（5.2）

（3）PLCC(Pearson product-moment correlation coefficient)



（5.3）

Pearson线性相关系数能够反应模型的预测准确度。取值在区间[-1,1]之间，其绝对值越接近1，表明预测准确度越高。

根据以上三个指标的对比，可得出客观质量评价算法与图布局主观质量评分的一致性程度，从而分析得出客观算法的性能。其中，皮尔逊线性相关性系数(Pearson Linear Correlation Coefficient, PLCC)，用于衡量客观算法预测的准确性；斯皮尔曼排序相关性系数(Spearman Rank order Correlation Coefficient, SROCC)，用于衡量客观算法预测结果的单调性；均方根误差(Root Mean Square Error, RMSE) ，用于衡量客观算法与主观实验预测结果的一致性。前两个指标越大（越接近1），说明客观算法效果越好；后一个指标越小（越接近0），说明客观算法效果越好。

我们首先进行线性回归对图布局质量评分进行拟合，接下来使用线性核函数、高斯核函数和多项式核函数进行支持向量回归，对预测的图布局的总体评分与主观实验中实际图布局的总体评分进行比对。具体的指标值对比如下：

表5.3 不同回归方法的性能指标对比

|  |  |  |  |
| --- | --- | --- | --- |
| 回归方法 | PLCC | SROCC | RMSE |
| 线性回归 | 0.5791 | 0.4461 | 0.1612 |
| ***线性核函数*** | ***0.9061*** | ***0.7852*** | ***0.0837*** |
| 高斯核函数 | 0.8192 | 0.6843 | 0.1134 |
| 多项式核函数（次数为2） | 0.6853 | 0.6187 | 0.1440 |
| 多项式核函数（次数为3） | 0.5354 | 0.3632 | 0.1670 |

从表5.3中可以看到，基于线性核函数的回归算法得到的三个指标都比其他几种方法得到的指标要好。使用线性核函数时PLCC的值为0.9061，SROCC为0.7852，RMSE为0.0837，回归的结果比较好。因此我们选取使用线性核函数的支持向量回归模型得到图布局质量评分。

**5.2.2 拟合效果评估**

接下来我们通过散点图来观测由回归模型得到的图布局质量评分与用户的实际主观评分之间的关系。图5.11显示的是基于线性核函数的支持向量回归算法客观评价的打分情况。横坐标表示客观评价打分，纵坐标表示主观评价打分，每一个点表示一个数据样本。图中的红色直线为图中所有点拟合出的一条直线，图中的蓝色虚线为*y=x*。从图中可以看出样本点拟合出来的直线和*y=x*比较接近，也就是说，主观评价算法与客观评价算法的结果是基本一致的，客观算法在一定程度上可以替代主观算法对于图布局的结果的评价，说明基于线性核函数的SVR算法对图布局整体质量评分的拟合效果较好，在本数据集上主观、客观方法得到的结果基本一致。

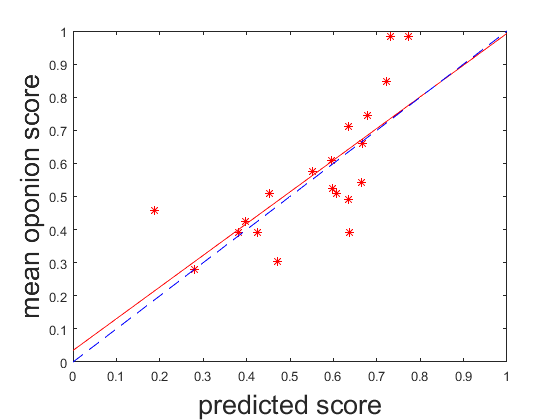


图5.11 基于线性核函数的SVR预测结果

因此，基于回归算法在主观评分标注的训练集上训练出的模型，我们可以预测图布局的评分。具体做法是先计算出图布局中影响布局效果的指标值（提取特征），再通过回归模型进行训练（训练，回归分析），最后用训练好的模型进行预测（测试），对于不同的图布局可以得到不一样的客观评分。该评分就表示图布局的效果好坏，评分高的图布局整体布局效果更好。这样，我们就通过客观评价算法实现了对图布局主观评分结果的拟合。

## 5.3 本章小结

本章中我们主要进行了两种图布局评价算法的用户实验。首先，把基于视觉的图布局评估方法应用到两个数据集中去，并评估出其在不同情况下适用的最优图布局算法；其次，对于基于回归的图布局质量评估算法，我们使用不同的回归模型，获取图布局的客观质量评价，并通过指标对比图布局客观质量评价算法与图布局主观质量评分的一致性程度，从而分析得出客观算法的性能并选择出最优的回归模型加以应用。

# 第六章 总结和展望

## 6.1 总结

图布局算法的评估是一项有很大实际意义的研究工作，有助于使用者选取最优布局，更好地理解图的结构信息。对于图来讲，不同图具有不同的节点以及关联关系，不同研究者也具有不同的研究目的，因此从众多图布局算法中选择一种合适的图布局算法涉及很多因素。本文从两个不同的角度出发，提出了两种对图布局算法进行评估的方法：（1）提出了一种基于视觉的图布局评估方法，首先提出了图布局中单个节点的视觉重要性概念，通过实验确立了图布局中影响节点视觉重要性的因素，提出了节点的视觉重要性计算模型，对于特定的数据集和研究目的，通过节点的视觉重要性和拓扑结构重要性之间的对比选择出更符合视觉特性的布局算法；（2）提出了一种基于回归的图布局整体质量评估方法，首先通过实验建立了图布局主观质量评分数据库，接下来综合了图布局中各个可能影响到图布局效果的指标，把它们作为因子，建立图布局指标与图布局评分的回归模型，从而通过模型对图布局进行质量评分。

以上两种方法基于两个不同的全新的角度，实现了对图布局算法的评价，两种评价方法具有不同的适合的应用场景。对于基于视觉的图布局算法的评估方法，它更加适用于用户有明确的研究目的或需求对图布局中重要节点进行发掘的情形，通过本文提出的布局评价算法，找出在视觉上能够更好地展示出重要节点的图布局算法。而基于回归的图布局整体质量评价则展示了图布局的整体评分，它是综合了各种图布局指标之后得到的图布局的评分，适用于需要对图布局进行整体评价的情况。

## 6.2 未来的研究工作

我们的工作还有一些不全面的地方，例如在计算节点的视觉重要性时，我们没有考虑与节点相连边的布局结构对其视觉重要性的影响，如果节点所在边布局对称性好，其视觉重要性可能会更高；还有本文中提到的布局方法中，节点的颜色相同，如果节点具有不同的颜色，其视觉重要性也会发生相应的变化。在接下来的工作中，我们计划考虑更多因素对节点视觉重要性的影响，进一步完善节点的视觉重要性标准，我们也会进一步采用更多的数据集和布局算法加以实验。在基于回归的图布局整体质量评价中，由于人力的限制，我们只建立了有60个图布局的图布局主观评分数据库，下步工作中，我们会进一步对现有的图布局评分数据库进行扩充，一方面增加图布局的数量，另一方面邀请更多的用户对图布局进行评分，以获取与实际情况更为符合的评分数据库。同时采用更多的图布局可读性指标对图布局评分进行拟合，以获取更为准确的图布局评分模型。

**参考文献**

1. 任磊, 杜一, 马帅,等. 大数据可视分析综述[J]. 软件学报, 2014(9):1909-1936.
2. 孙扬, 蒋远翔, 赵翔, 等. 网络可视化研究综述[J]. 计算机科学, 2010 (2): 12-18.
3. Kim S C, Seo K K, Kim I K, et al. Readings in information visualization: using vision to think[M]. DBLP, 1999: 1-712.
4. Shneiderman B, Aris A. Network Visualization by Semantic Substrates[J]. IEEE Transactions on Visualization & Computer Graphics, 2006, 12(5):733.
5. Purchase H C, Carrington D, Allder J A. Empirical Evaluation of Aesthetics-based Graph Layout[J]. Empirical Software Engineering, 2002, 7(3):233-255.
6. Tamassia R. On embedding a graph in the grid with the minimum number of bends[J]. 1987, 16(3):421-444.
7. Biedl T C, Marks J, Ryall K, et al. Graph Multidrawing: Finding Nice Drawings Without Defining Nice[C]// International Symposium on Graph Drawing. Springer-Verlag, 1998:347-355.
8. Purchase H C. Which Aesthetic Has the Greatest Effect on Human Understanding[M]// Graph Drawing. Springer Berlin Heidelberg, 1997:248-261.
9. Davidson R. Drawing graphs nicely using simulated annealing[J]. Acm Transactions on Graphics, 1996, 15(4):301-331.
10. Tamassia R, Battista G D, Batini C. Automatic graph drawing and readability of diagrams[J]. IEEE Transactions on Systems Man & Cybernetics, 1988, 18(1):61-79.
11. Bennett C, Ryall J, Spalteholz L, et al. The aesthetics of graph visualization[C]// Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging. Eurographics Association, 2007:57-64.
12. Wang Y, Shen Q, Archambault D, et al. AmbiguityVis: Visualization of Ambiguity in Graph Layouts[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(1): 359-368.
13. Pohl M, Schmitt M, Diehl S. Comparing the readability of graph layouts using eyetracking and task-oriented analysis[C]// Computational Aesthetics 2009: Eurographics Workshop on Computational Aesthetics, Victoria, British Columbia, Canada. DBLP, 2009:49-56.
14. Marriott K, Purchase H, Wybrow M, et al. Memorability of Visual Features in Network Diagrams[J]. IEEE Transactions on Visualization & Computer Graphics, 2012, 18(12):2477-2485.
15. 赫南, 李德毅, 淦文燕, 等. 复杂网络中重要性节点发掘综述[J]. 计算机科学, 2007, 34(12): 1-5.
16. Sabidussi G. The Centrality Index of a Graph[J]. Psychometrika, 1966, 31(4):581-603.
17. Freeman L C. A set of measures of centrality based on betweenness.[J]. Sociometry, 1977, 40(1):35-41.
18. Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification[J]. The Journal of Mathematical Sociology, 1972, 2(1):113-120.
19. Tory M, Möller T. Human factors in visualization research.[J]. Visualization & Computer Graphics IEEE Transactions on, 2004, 10(1):72.
20. Huang W. Evaluating Overall Quality of Graph Visualizations Indirectly and Directly[M]// Handbook of Human Centric Visualization. Springer New York, 2014:373-390.
21. Eades P A. A Heuristic for Graph Drawing[J]. Congressus Numerantium, 1984, 42(42):149-160.
22. Harel D, Koren Y. A fast multi-scale method for drawing large graphs[J]. Journal of Graph Algorithms & Applications, 2015, 1984(3):183-196.
23. Fruchterman T M J, Reingold E M. Graph drawing by force directed placement[J]. Software Practice & Experience, 21.
24. Jacomy M, Venturini T, Heymann S, et al. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software.[J]. Plos One, 2014, 9(6):e98679.
25. Hu Y. Efficient, High-Quality Force-Directed Graph Drawing[J]. Mathematica Journal, 2005, 10(1): 37-71.
26. Martin S. OpenOrd: an open-source toolbox for large graph layout[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2011, 7868(3):786806.
27. Cui W, Zhou H, Qu H, et al. Geometry-Based Edge Clustering for Graph Visualization[J]. IEEE Transactions on Visualization & Computer Graphics, 2008, 14(6):1277.
28. Eades P, Feng Q W. Multilevel visualization of clustered graphs[J]. International Journal of Research & Reviews in Computer Science, 1996, 1190(1):101-112.
29. Noack A. An Energy Model for Visual Graph Clustering[J]. Proceedings of International Symposium on Graph Drawing Lncs, 2003, 2912:425-436.
30. Battista G D, Eades P, Tamassia R, et al. Graph Drawing: Algorithms for the Visualization of Graphs[J]. Upper Saddle River Nj, 1998.
31. Huang W, Huang M L, Lin C C. Evaluating overall quality of graph visualizations based on aesthetics aggregation[J]. Information Sciences, 2015, 330:444-454.
32. Ware C, Purchase H C, Colpoys L, et al. Cognitive measures of graph aesthetics[J]. Information Visualization, 2002, 1(2):103-110.
33. Lee S, Sips M, Seidel H P. Perceptually driven visibility optimization for categorical data visualization[J]. Visualization & Computer Graphics IEEE Transactions on, 2013, 19(10):1746.
34. Albuquerque G, Eisemann M, Magnor M. Perception-based visual quality measures[C]// IEEE Conference on Visual Analytics Science and Technology, Vast 2011, Providence, Rhode Island, Usa, October. DBLP, 2011:13-20.
35. Itti L, Koch C. Computational modelling of visual attention[J]. Nature Reviews Neuroscience, 2001, 2(3):194.
36. Ware C. Information visualization: perception for design[M]. Morgan Kaufmann Publishers Inc. 2000.
37. Bovik A C. Automatic Prediction of Perceptual Image and Video Quality[J]. Proceedings of the IEEE, 2013, 101(9):2008-2024.
38. Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4):600-12.
39. 吴金建. 基于人类视觉系统的图像信息感知和图像质量评价[D]. 西安：西安电子科技大学, 2014.
40. FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF VIDEO QUALITY ASSESSMENT COMLS E,，S Chisholm，Q, Rapporteur,2008.
41. 方开泰, 全辉, 陈庆云. 实用回归分析[M]. 北京: 科学出版杜, 1988.
42. Neter J, Kutner M H, Nachtsheim C J, et al. Applied linear statistical models /[J]. Journal of the American Statistical Association, 1996, 39(3):19–32.
43. Cortes C, Vapnik V. Support-Vector Networks[J]. Machine Learning, 1995, 20(3):273-297.
44. Osberger W, Maeder A J. Automatic Identification of Perceptually Important Regions in an Image[C]// Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. IEEE, 1998:701-704 vol.1.
45. Bauer B, Jordán F, Podani J. Node centrality indices in food webs: Rank orders versus distributions[J]. Ecological Complexity, 2010, 7(4): 471-477.
46. Melián C J, Bascompte J. Food web cohesion[J]. Ecology, 2004, 85(2): 352-358.
47. Heymans J J, Ulanowicz R E, Bondavalli C. Network analysis of the South Florida Everglades graminoid marshes and comparison with nearby cypress ecosystems[J]. Ecological Modelling, 2002, 149(1): 5-23.
48. Jeong H, Mason S P, Barabási A L, et al. Lethality and centrality in protein networks[J]. Nature, 2001, 411(6833): 41-42.
49. Vladimir Batagelj and Andrej Mrvar (2006): Pajek datasets.<URL: <http://vlado.fmf.uni-lj.si/pub/networks/data/>>.

**附录**

附录展示了图布局数据库中的60个图布局，图1到图15为15个不同规模的图结构，每个图结构中1到4分别为图的Fruchterman-Reingold布局、ForceAtlas2布局、Openord布局和Yifanhu布局。

|  |  |  |
| --- | --- | --- |
| fr | fa2 | |
| **图1-1** | **图1-2** | |
| yifanhu | openord | |
| **图1-3** | **图1-4** | |
| fr | fa2 | |
| **图2-1** | **图2-2** | |
| openord | yifanhu | |
| **图2-3** | **图2-4** | |
| fr | | fa2 |
| **图3-1** | | **图3-2** |
| openord | | yifanhu |
| **图3-3** | | **图3-4** |
| fr | | fa2 |
| **图4-1** | | **图4-2** |
| openord | | yifanhu |
| **图4-3** | | **图4-4** |

|  |  |
| --- | --- |
| frr | fa2 |
| **图5-1** | **图5-2** |
| openord | yifanhu |
| **图5-3** | **图5-4** |
| fr | fa2 |
| **图6-1** | **图6-2** |
| openord | yifanhu |
| **图6-3** | **图6-4** |

|  |  |
| --- | --- |
| fr | fa2 |
| **图7-1** | **图7-2** |
| openord | yifanhu |
| **图7-3** | **图7-4** |
| fr | fa2 |
| **图8-1** | **图8-2** |
| openord | yifunhu |
| **图8-3** | **图8-4** |

|  |  |
| --- | --- |
| fr | fa2 |
| **图9-1** | **图9-2** |
| openord | yifanhu |
| **图9-3** | **图9-4** |
| fr | fa2 |
| **图10-1** | **图10-2** |
| openford | yifanhu |
| **图10-3** | **图10-4** |

|  |  |  |
| --- | --- | --- |
| fr | fa2 | |
| **图11-1** | **图11-2** | |
| openord | yifanhu | |
| **图11-3** | **图11-4** | |
| fr | fa2 | |
| **图12-1** | **图12-2** | |
| openord | yifanhu | |
| **图12-3** | **图12-4** | |
| fr | | fa2 |
| **图13-1** | | **图13-2** |
| openord | | yifanhu |
| **图13-3** | | **图13-4** |
| fr | | fa2 |
| **图14-1** | | **图14-2** |
| openord | | yifanhu |
| **图14-3** | | **图14-4** |
| fr | fa2 | |
| **图15-1** | **图15-2** | |
| openord | yifanhu | |
| **图15-3** | **图15-4** | |

# 致 谢

光阴荏苒，岁月如歌，转眼间我的硕士研究生生活即将结束，毕业论文的写作也已接近尾声。回想起丽娃河畔、理科大楼内三年的学习生活，真的是浮想联翩，感慨万千！

首先，感谢母校华师大给我提供了优质的学习环境，给了我更高、更新的起点，让我有了丰富的知识储备，使我更加从容自信地面对人生的挑战！

其次，感谢我的导师王长波教授，以及张康、殷红等老师，特别是对我支持和帮助最多的王长波老师！在我三年的学习生活中他给予我无微不至的学业指导和生活关照。在科研方面，老师结合我的兴趣帮我有针对性地制定学习计划和研究计划，认真教导我怎么做研究，每周定期和我一起交流，帮我理清学习思路、排除学习障碍。在生活方面，老师时常帮助我解决生活中的疑惑和困难，同时又组织实验室同学一起交流，增进了师生及同学之间的友谊。在我写作论文时，无论是开题还是写作过程中，老师都给予了认真指导，同时给我的论文提出了许多建设性的意见，在这个过程中，我学会了如何去发现问题、分析问题和解决问题。在跟随老师学习的这段时间里，老师渊博的知识、严谨的治学态度和忘我的工作精神一直深深地感染着我，大大激发了我的学习及生活热情，使我受益终生。在此，我衷心地感谢老师，谢谢您对我的谆谆教诲和悉心照顾！

再次，感谢刘玉华、叶鹏、孟宇、孔凡龙、张成海、张霞、王超、李晨等师兄师姐们；感谢感谢王秋丹、孙凯、田浩、张泰滺、阮骥鸣以及我所有的师弟师妹们！三年的时间，你们一直在帮助我、陪伴我，为我排除学习上的困扰，增添生活中的乐趣，让我在实验室享受到了家庭的温暖，感受到了兄弟姐妹般的浓厚情谊，真诚地感谢你们的陪伴！

然后，我要特别感谢我的父母以及一直关心我的亲人们，感谢您们多年的陪伴和教导，感谢您们带给我的温暖和关怀，在前进的道路上您们无私奉献、无怨无悔地在生活上为我遮风挡雨，学业上为我指明方向！感谢您们教会了我做人做事，感谢您们培养我长大成才。

今后的学习、工作、生活中我会不断努力，奋发向上，再接再厉，再创辉煌！充分利用自己所学的专业知识，回馈祖国，回馈社会，为祖国的繁荣富强做贡献，为母校华师大增光添彩！

最后衷心感谢各位在百忙之中抽出时间为我审阅论文并提出宝贵建议的评审老师和专家们。

# 研究生期间科研成果

[1] Li J, Liu Y, Wang C. Evaluation of graph layout methods based on visual perception[C]// Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing. ACM, 2016: 90.

[2] Cavelan A, Li J, Robert Y, et al. When Amdahl Meets Young/Daly[C]// IEEE International Conference on CLUSTER Computing. IEEE, 2016:203-212.

[3] 李佳凡.基于视觉感知的图布局评估方法：中国，201610959140.0.

1. http://www.visualcomplexity.com/vc/index.cfm [↑](#footnote-ref-1)
2. https://gephi.org/ [↑](#footnote-ref-2)
3. http://www.fao.org/forestry/mangrove/3648/en/ [↑](#footnote-ref-3)