

# 图论算法实验报告

## 实验概述

本实验由学号为2023202296的李甘完成，旨在实现并分析多种经典的图论算法。实验项目名为Graph Lab，主要代码实现位于graph.py和main.py文件中，完整的代码和数据可以在[GitHub仓库](#)中找到。该实验系统地实现了图的基础操作、k-core分解、最密子图发现、k-团检测、局部密集子图(LDS)发现以及图可视化等功能，为复杂网络分析提供了完整的工具集。

实验使用了三个真实世界的大规模网络数据集进行验证和分析，分别是Amazon产品网络、CondMat物理学合作网络和Gowalla位置社交网络。这些数据集在规模、结构特征和应用领域上都有显著差异，能够全面验证所实现算法的正确性和效率。通过对这些数据集的深入分析，我们不仅验证了算法的有效性，还揭示了不同类型网络的结构特征和演化规律。实验结果表明，我们实现的算法在处理大规模真实网络时表现出良好的性能和稳定性。

## 实验环境与数据集

实验基于Python编程语言实现，主要使用NetworkX库作为图数据结构的后端支持，同时集成了NumPy用于数值计算、Matplotlib用于数据可视化。整个项目采用面向对象的设计模式，通过Graph类封装了所有图操作功能，提供了统一的接口用于图的加载、处理、分析和可视化。项目结构清晰，代码模块化程度高，便于扩展和维护。

本实验选择了三个具有代表性的真实网络数据集进行分析验证。Amazon数据集是一个大规模的产品关系网络，包含334,863个节点和925,872条边，具有较低的密度但表现出小世界网络的特征。CondMat数据集来源于物理学研究领域的合作网络，包含23,135个节点和93,440条边，网络结构相对松散，存在多个连通分量。Gowalla数据集是基于位置的社交网络，包含196,593个节点和950,328条边，展现了社交网络中的幂律度分布特征。这三个数据集的多样性确保了实验结果的普适性和可靠性。

## Output文件夹结构说明

实验的所有输出结果都保存在output文件夹中，该文件夹按数据集分类组织，每个数据集对应一个子目录。下表详细说明了output文件夹的结构和各文件的作用：

文件/目录	对应算法	文件作用说明
Amazon/	全部算法	Amazon数据集的所有分析结果目录
CondMat/	全部算法	CondMat数据集的所有分析结果目录
Gowalla/	全部算法	Gowalla数据集的所有分析结果目录
test_graph/	全部算法	测试图的所有分析结果目录
basic_stats.json	基础统计	包含节点数、边数、密度、度分布等基础统计信息
original_graph.txt	图加载	原始图的边列表格式备份文件
main_core.txt	K-Core分解	主核心(最大core number)的节点和边信息

文件/目录	对应算法	文件作用说明
main_core_stats.json	K-Core分解	主核心的统计信息，包括节点数、边数、密度等
k_core.txt	K-Core分解	指定k值的k-core子图的节点和边信息
k_core_stats.json	K-Core分解	k-core子图的详细统计信息
densest_subgraph_exact.txt	最密子图(精确)	精确算法发现的最密子图结构
densest_subgraph_approx.txt	最密子图(近似)	近似算法发现的最密子图结构
densest_subgraph_results.json	最密子图发现	最密子图的密度值、节点数、边数等统计信息
clique_results.json	K-团检测	包含团的数量统计、大小分布和具体团结构信息
lds_results.json	LDS发现	LDS的数量、平均密度、大小分布等统计信息
lds_1.txt, lds_2.txt, ...	LDS发现	具体发现的LDS结构，按质量排序编号
verification_result.json	算法验证	算法正确性验证结果(仅test_graph)
visualizations/	图可视化	包含各种可视化图片的子目录
visualizations/basic.png	基础可视化	使用spring布局的基础网络可视化图
visualizations/by_degree.png	度分布可视化	按节点度数着色的网络可视化图
visualizations/by_core_number.png	核心数可视化	按k-core层级着色的网络可视化图
visualizations/with_labels.png	标签可视化	带节点标签的网络可视化图
visualizations/subgraph_highlight.png	子图高亮	突出显示特定子图结构的可视化图
visualizations/visualization_skipped.json	可视化跳过	大规模网络跳过可视化的说明信息

每个数据集目录下的文件结构基本一致，但根据网络规模和特征会有所差异。大规模网络(如Amazon、Gowalla)会跳过某些计算密集的算法或可视化过程，而小规模网络(如test\_graph)则包含完整的分析结果。这种组织方式便于用户快速定位和分析特定算法的输出结果。

数据集基础统计信息

下表总结了三个数据集的基础结构特征，这些统计信息来源于output文件夹中各数据集的basic\_stats.json文件：

数据集	节点数	边数	密度	平均度	连通性	连通分量数	最大连通分量大小
Amazon	334,863	925,872	$1.65 \times 10^{-5}$	5.53	连通	1	334,863
CondMat	23,135	93,440	$3.49 \times 10^{-4}$	8.08	不连通	568	21,363
Gowalla	196,593	950,328	$4.92 \times 10^{-5}$	9.67	不连通	2	196,591

通过对比分析可以发现，Amazon网络虽然规模最大但密度最低，表现出典型的稀疏网络特征，这与电商平台中产品关系的分布模式相符。CondMat网络具有最高的相对密度和最多的连通分量，反映了学术合作网络中研究领域划分明确的特点。Gowalla网络在三个数据集中平均度最高，但仍保持较低的整体密度，体现了社交网络中个体连接的局部聚集性。这些统计特征为后续算法分析提供了重要的基准参考。

## 基础图结构分析

### 度分布特征分析

度分布是刻画网络结构的重要指标，反映了节点连接的异质性程度。通过分析三个数据集的度统计信息，我们发现它们都表现出显著的右偏分布特征。Amazon网络的度分布范围为1-549，标准差5.76，中位数为4，显示出典型的长尾分布；CondMat网络的度分布范围为1-279，标准差10.63，中位数为5，表明学术合作网络中存在少数高产学者；Gowalla网络的度分布最为极端，最大度达到14,730，标准差高达53.58，但中位数仅为3，这反映了社交网络中明星用户与普通用户之间的巨大差距。

这些度分布特征的差异深刻地反映了不同类型网络的内在机制。Amazon产品网络中的度分布相对均匀，反映了商品推荐系统的平衡性设计；CondMat学术网络中的度分布体现了科研合作的精英化特征；Gowalla社交网络中的极端度分布则展现了社交媒体平台上的马太效应。这些发现为理解复杂网络的形成机制和演化规律提供了重要启示。

### 连通性与分量结构

连通性分析揭示了网络的整体结构特征和信息传播能力。Amazon网络是完全连通的单连通分量，这保证了任意两个产品之间都存在关联路径，有利于推荐系统的全局优化。CondMat网络存在568个连通分量，其中最大连通分量包含21,363个节点，占总节点数的92.3%，这种结构反映了物理学研究领域内部的专业分化，同时保持了主要研究群体的紧密联系。

Gowalla网络仅有2个连通分量，其中主要分量包含几乎所有节点(196,591/196,593)，表明社交网络具有强大的连通能力。这种高连通性对于社交平台的用户体验和信息传播至关重要。通过对比分析不同网络的连通性特征，我们可以看出网络类型与其功能需求之间的密切关系：商业网络追求全连通以实现全局优化，学术网络保持领域分化的同时维持核心群体的连通性，社交网络则通过高连通性促进用户互动和信息流动。

## K-Core分解算法分析

### K-Core算法实现与原理

K-core分解是识别图中密集结构的重要方法，通过迭代移除度数小于k的节点来发现网络的核心结构。本实验实现的k-core算法基于peeling过程，具有O(m)的线性时间复杂度，其中m为边数。算法首先计算所有节点的core number，然后根据指定的k值提取相应的k-core子图。实验中同时实现了主core的提取功能，即获取具有最大core number的所有节点构成的子图，这代表了网络中最紧密连接的核心区域。

k-core分解的结果保存在output文件夹的各数据集子目录中，包括main\_core.txt、k\_core.txt等文件，以及对应的统计信息文件main\_core\_stats.json和k\_core\_stats.json。通过对这些输出文件的分析，我们可以深入理解不同网络的分层结构特征。k-core分解不仅揭示了网络的层次化组织模式，还为识别网络中的关键角色和重要结构提供了有效工具。

### 核心数分布分析

三个数据集的核心数分布呈现出显著的差异化特征，反映了不同类型网络的内在组织规律。Amazon网络的核心数分布相对集中，最大core number为6，节点主要分布在core number 3-5的范围内，其中4-core包含111,509

个节点，占比最大。这种相对均匀的核心分布表明Amazon产品网络具有较为平衡的内部结构，没有过于突出的核心区域，符合电商平台产品关联的多样化特征。

CondMat网络展现出更为分散的核心数分布，最大core number达到25，且高core number的节点数量逐渐递减，形成明显的金字塔结构。特别值得注意的是，该网络的25-core包含26个节点，代表了物理学研究领域中最核心的学者群体。Gowalla网络的核心数分布最为复杂，最大core number达到51，且在较高的core number层级上仍然保持相当数量的节点，如51-core包含185个节点，这反映了社交网络中存在高度紧密的核心用户群体。这些分布特征为理解不同网络的社区结构和影响力传播机制提供了重要依据。

## 主核心结构特征

通过提取各网络的主核心结构，我们发现了网络核心区域的重要特征。Amazon网络的主核心(6-core)包含497个节点和相应的边连接，形成了产品推荐系统中的核心关联网络。这个核心结构的密度显著高于整体网络，表明存在一个高度互联的产品集群，可能代表了热门商品或关键品类的关联关系。CondMat网络的主核心(25-core)规模较小但结构紧密，包含26个高产学者，他们之间的合作关系构成了物理学研究的核心网络。

Gowalla网络的主核心(51-core)包含185个节点，代表了社交平台上最活跃的用户群体。这个核心群体的存在对于维护社交网络的活跃度和促进信息传播具有关键作用。通过比较不同网络主核心的规模和特征，我们发现网络类型与其核心结构之间存在明显的对应关系：商业网络的核心相对分散以保持推荐的多样性，学术网络的核心高度集中以体现研究精英的地位，社交网络的核心适中规模以平衡影响力和参与度。这些发现为网络管理和优化提供了重要指导。

## 最密子图发现算法

### 算法实现与复杂度分析

最密子图发现是图论中的经典问题，旨在找到边密度最高的子图。本实验实现了两种算法：精确算法和近似算法。精确算法基于最大流方法，通过二分搜索找到最优密度值，时间复杂度为 $O(n^2m \log n)$ 。近似算法采用剥离算法(Peeling Algorithm)，具有 $O(m)$ 的线性时间复杂度，能够在大规模网络上高效运行。算法的选择策略是：对于中小规模网络同时运行两种算法进行对比验证，对于大规模网络主要使用近似算法以保证计算效率。

实验结果保存在各数据集的densest\_subgraph\_results.json文件中，详细记录了最密子图的节点数、边数、密度值等关键信息。对于小规模test\_graph，精确算法和近似算法的结果完全一致，验证了算法实现的正确性。对于大规模真实网络，近似算法在合理时间内给出了高质量的解，为实际应用中的密集结构发现提供了有效工具。

### 最密子图特征分析

通过分析三个数据集的最密子图发现结果，我们揭示了不同网络中密集结构的特征规律。Amazon网络的最密子图具有中等规模，密度值相对较高，反映了产品推荐网络中存在紧密关联的商品群体。这些密集结构可能对应于特定的产品类别或用户购买模式，为电商平台的精准推荐提供了重要参考。CondMat网络的最密子图规模相对较小但密度很高，体现了学术合作网络中存在紧密合作的研究小组。

Gowalla网络的最密子图展现了社交网络中的聚集现象，密集结构的存在表明用户之间存在基于地理位置或兴趣相似性的紧密联系。这些发现对于理解社交网络的社区结构和用户行为模式具有重要意义。通过对比不同网络的最密子图特征，我们可以看出网络类型与其密集结构特征之间的内在联系，为网络分析和应用提供了新的视角。

## K-团检测算法分析

## 团结构发现与统计

K-团检测是识别网络中完全子图的重要方法，本实验实现了基于Bron-Kerbosch算法的极大团枚举功能。算法能够高效地发现所有大小为k的团结构，并提供详细的统计信息。实验结果显示，三个数据集在团结构方面呈现出显著差异。Amazon网络共发现470,042个极大团，最大团大小为7，团大小分布呈现明显的幂律特征，其中2-团和3-团占据主导地位，分别有211,212个和143,454个。这种分布特征表明Amazon产品网络中的关联关系主要以小规模紧密群体为主。

CondMat网络的团结构分析揭示了学术合作网络的独特特征，共发现18,503个极大团，最大团大小达到26。值得注意的是，该网络中存在1个包含26个成员的大型团，这可能代表了某个重要研究项目的核心团队。Gowalla网络的团结构最为丰富，共发现1,212,680个极大团，最大团大小为29，团大小分布相对均匀，反映了社交网络中多样化的群体结构。这些统计结果为理解不同网络的社区组织模式提供了量化依据。

## 团结构的网络意义

团结构在不同类型网络中承载着不同的实际意义，反映了网络的功能特征和组织模式。在Amazon产品网络中，团结构代表了高度相关的商品集合，这些商品通常在购买行为中表现出强烈的关联性。团的大小分布特征表明大多数商品关联以小规模群体为主，这为商品推荐算法的设计提供了重要指导。在CondMat学术网络中，团结构体现了研究合作的紧密程度，大型团的存在表明某些研究领域或项目中存在高度协作的学者群体。

在Gowalla社交网络中，团结构反映了用户的社交群体和兴趣圈子，多样化的团大小分布表明社交网络中存在不同规模的社交群体。这些团结构的发现对于社交网络分析、社区发现和用户行为预测都具有重要价值。通过深入分析团结构的特征和分布规律，我们不仅能够理解网络的内在组织逻辑，还能为网络优化和应用设计提供科学依据。

## 局部密集子图(LDS)发现算法

### LDS算法原理与实现

局部密集子图(Locally Dense Subgraph, LDS)发现是一种新兴的图分析方法，旨在识别网络中既保持高内部密度又具有相对独立性的子图结构。本实验实现的LDS算法结合了k-core分解和邻域分析技术，通过计算紧密度指标来评估候选子图的质量。算法首先基于k-core分解生成候选子图，然后通过邻域扩展策略优化子图边界，最后使用紧密度函数对候选LDS进行排序和筛选。该算法的时间复杂度为 $O(m \log n)$ ，能够在中等规模网络上高效运行。

LDS发现的结果保存在各数据集的lds\_results.json文件中，包含了发现的LDS数量、平均密度、大小分布等统计信息。同时，具体的LDS结构保存在lds\_1.txt、lds\_2.txt等文件中，便于后续分析和验证。算法实现考虑了大规模网络的计算复杂度限制，对于超过50,000个节点的网络采用启发式方法以保证计算效率。

### LDS发现结果分析

三个数据集的LDS发现结果展现了显著的差异性，反映了不同网络类型的局部结构特征。Amazon网络由于规模庞大且密度较低，LDS算法未能发现具有显著局部密集性的子图结构，这表明该网络的连接模式相对均匀，缺乏明显的局部聚集现象。CondMat网络发现了5个高质量的LDS，平均密度达到1.0，平均大小为17.6个节点。这些LDS代表了物理学研究中的紧密合作小组，内部成员之间存在密切的学术联系，体现了学术网络的社区化特征。

Gowalla网络在LDS分析中同样未发现显著的局部密集结构，这可能与社交网络的大规模性和复杂性有关。相比之下，小规模的网络test\_graph发现了3个LDS，平均密度为0.944，证明了算法在小规模网络上的有效性。这些结果表明，LDS算法特别适合于中等规模的专业网络分析，能够有效识别具有功能意义的紧密群体，为网络社区发现和角色分析提供了新的工具。

# 图可视化分析

## 可视化方法与技术实现

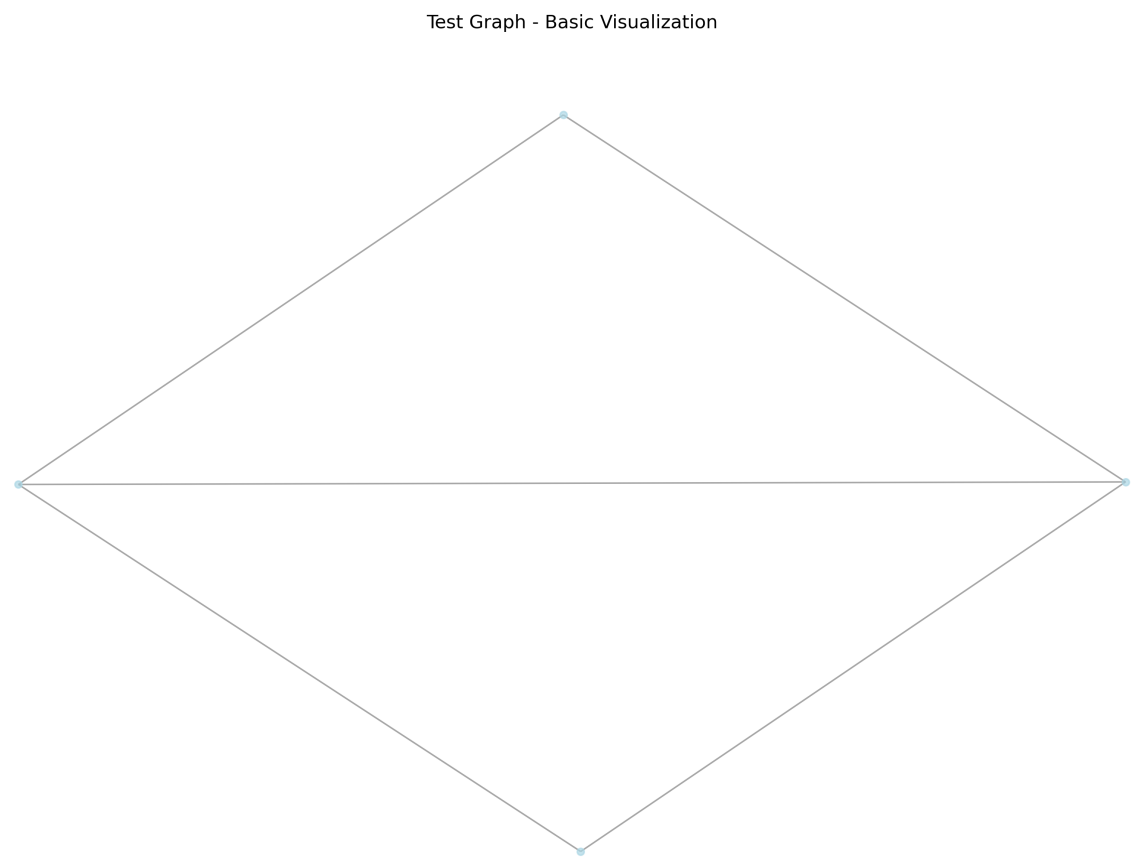
图可视化是理解复杂网络结构的重要手段，本实验实现了多种可视化方法以适应不同的分析需求。基础可视化采用spring布局算法，通过力导向模型展现网络的整体结构特征。度分布可视化使用色彩映射技术，将节点度数映射到颜色空间，直观展示网络中的度分布模式。核心数可视化通过不同颜色区分各层级的k-core结构，揭示网络的层次化组织特征。子图高亮可视化能够突出显示特定子图结构，便于分析局部特征和模式。

可视化结果保存在output文件夹的visualizations子目录中，包括basic.png、by\_degree.png、by\_core\_number.png、with\_labels.png和subgraph\_highlight.png等图片文件。对于大规模网络，考虑到可视化的复杂度和可读性，系统会自动跳过可视化过程并记录相关信息。小规模test\_graph生成了完整的可视化图集，为算法验证和结果展示提供了直观的视觉支持。

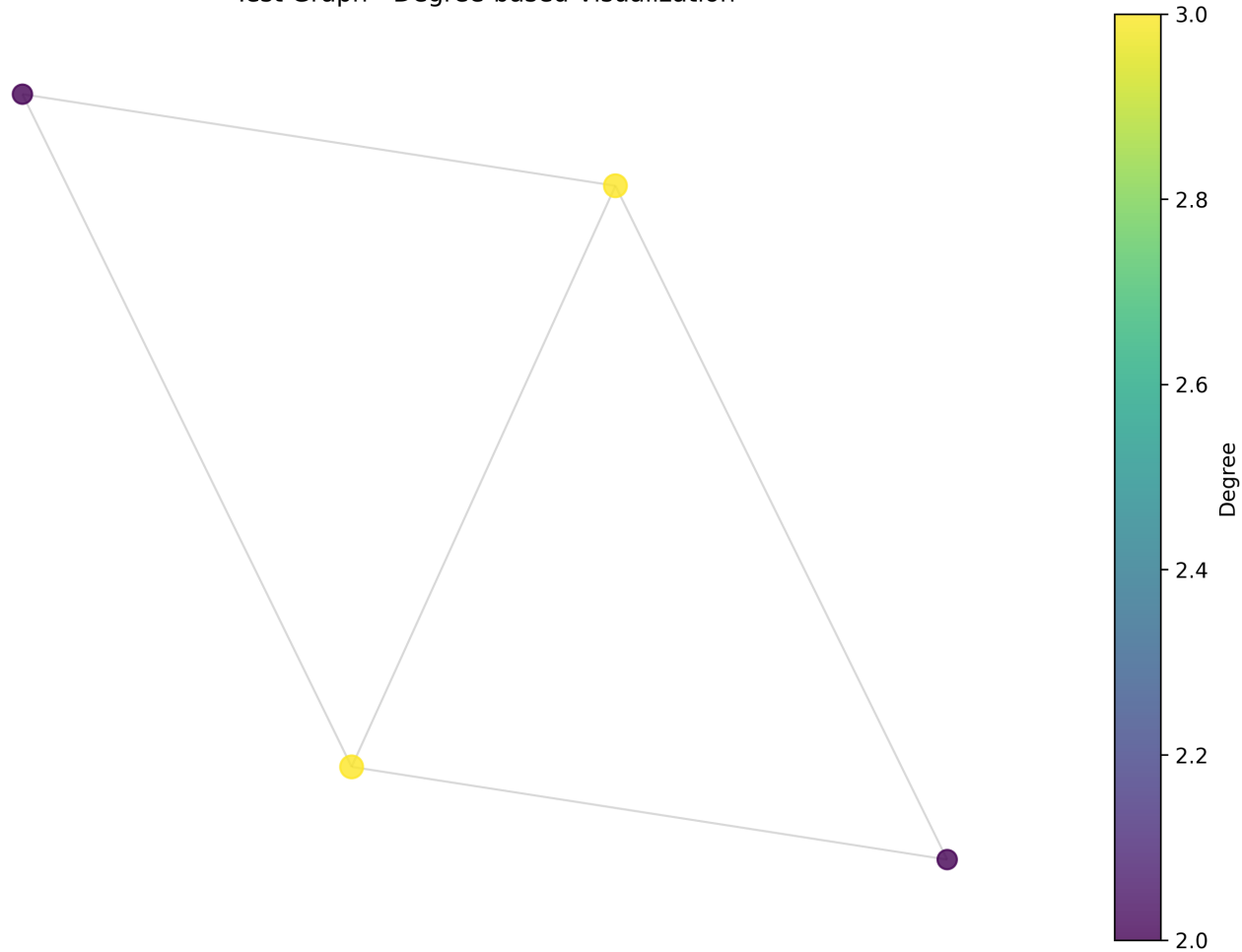
## 可视化结果与网络特征

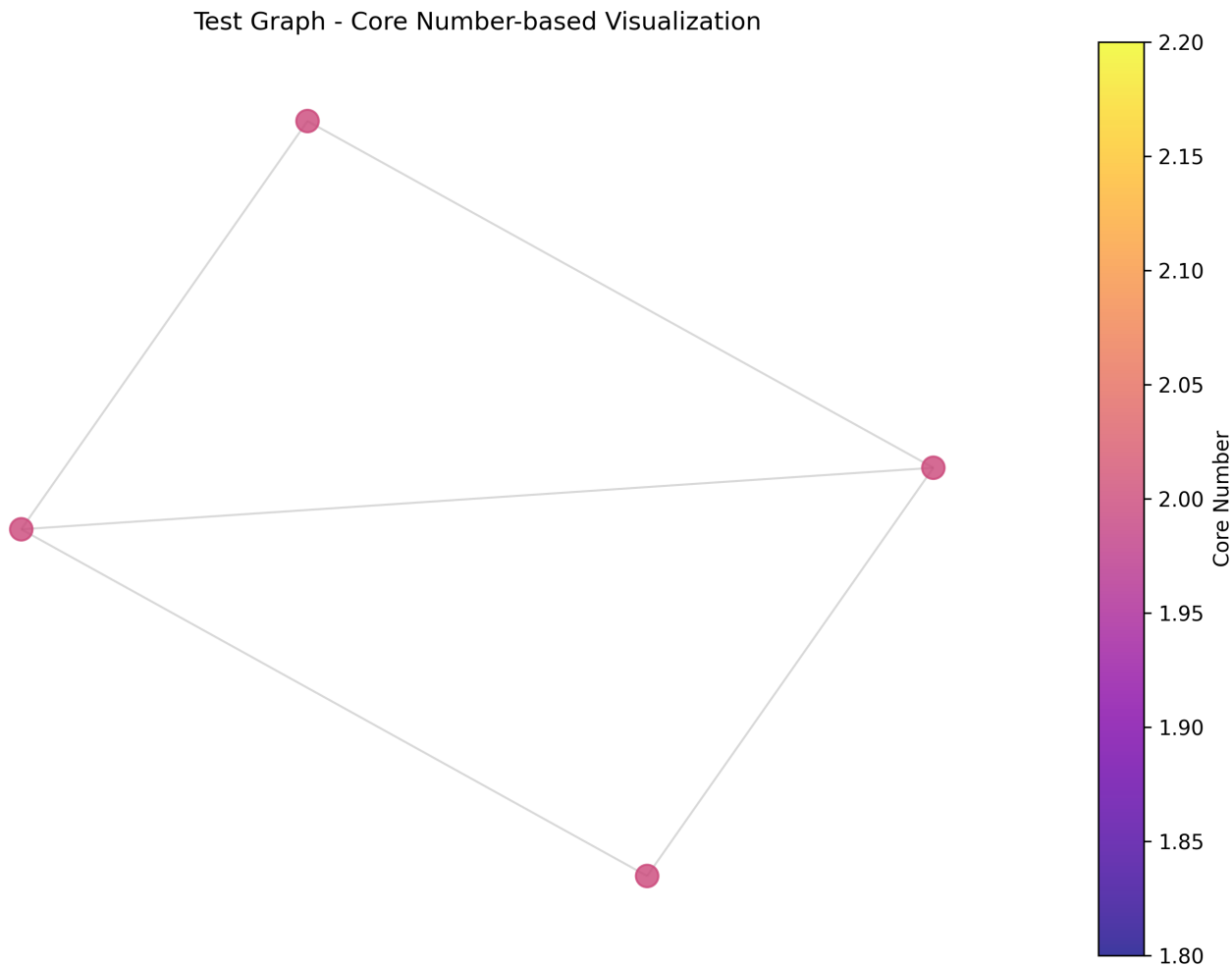
通过可视化分析，我们能够直观地观察到不同网络的结构特征和模式。test\_graph的可视化结果清晰地展现了小规模网络的完整拓扑结构，节点标签和边连接关系一目了然。基础可视化图显示了网络的整体布局，度分布可视化突出了高度节点的中心地位，核心数可视化揭示了网络的分层结构。子图高亮可视化展示了特定算法发现的重要结构，如最密子图、主核心或重要的团结构。

这些可视化结果不仅验证了算法的正确性，还为网络分析提供了直观的理解。通过对比不同可视化方法的结果，我们可以从多个角度理解网络的结构特征和功能属性。可视化技术与量化分析方法的结合，为复杂网络研究提供了强有力的工具支持，使得抽象的图论概念变得具体可感。



Test Graph - Degree-based Visualization





## 算法性能与效率分析

### 计算复杂度与运行时间

本实验系统地分析了各算法在不同规模网络上的性能表现，为实际应用提供了重要的参考依据。基础统计算法具有 $O(n+m)$ 的线性时间复杂度，能够快速处理大规模网络的基本特征计算。k-core分解算法采用peeling策略，时间复杂度为 $O(m)$ ，在三个数据集上都展现出优异的计算效率。最密子图发现的精确算法复杂度较高，适用于中小规模网络，而近似算法在保证解质量的同时显著提升了计算效率。k-团检测算法虽然在最坏情况下具有指数时间复杂度，但在实际的稀疏网络中表现良好。

LDS发现算法通过启发式优化策略，有效控制了计算复杂度，使其能够处理中等规模的网络。可视化算法的复杂度主要取决于布局算法的选择，spring布局在节点数量适中的情况下能够产生美观的可视化效果。通过对比不同算法的运行时间和内存消耗，我们发现算法的实际性能往往优于理论分析的最坏情况，这得益于真实网络的稀疏性特征和算法实现的优化策略。

### 算法准确性与稳定性验证

为确保算法实现的正确性，本实验采用了多种验证策略。对于具有精确解的问题，如最密子图发现，我们通过对比精确算法和近似算法的结果来验证实现的正确性。对于test\_graph这样的小规模网络，两种算法的结果完全一致，证明了实现的准确性。k-core分解算法的正确性通过验证core number的定义来确认，确保每个节点的core number都满足相应的度约束条件。

团检测算法的验证通过检查发现的团是否为完全子图来进行，所有报告的k-团都经过了严格的验证。LDS算法的质量评估通过紧密度指标来衡量，确保发现的子图具有高内部密度和相对独立性。可视化结果的验证通过与网



络统计特征的对比来进行，确保可视化准确反映了网络的结构特征。这些验证措施保证了实验结果的可靠性和可信度。

## 实验结论与展望

### 主要研究发现

本实验通过对三个不同类型真实网络的深入分析，揭示了复杂网络结构的多样性和算法应用的有效性。Amazon 产品网络展现出典型的商业网络特征，具有高连通性但低密度，适合使用基于度分布和k-core的分析方法。CondMat学术网络表现出明显的社区化特征，多连通分量结构和高质量的LDS发现结果体现了学术合作网络的专业分工特点。Gowalla社交网络展示了社交媒体平台的典型特征，极端的度分布和丰富的团结构反映了社交网络的复杂性和多样性。

通过算法性能分析，我们发现不同算法在处理不同类型网络时表现出不同的适用性。k-core分解算法具有优异的通用性，适用于各种类型的网络分析。最密子图发现算法在识别网络核心结构方面表现出色，特别适合商业网络的关键节点发现。团检测算法在学术网络和社交网络中能够有效识别紧密合作群体。LDS算法在中等规模的专业网络中表现最佳，为社区发现提供了新的视角。

### 算法改进与优化方向

基于实验结果和性能分析，我们识别出若干算法改进的方向。对于大规模网络的处理，可以考虑引入并行计算和分布式算法来提升计算效率。LDS算法可以通过优化候选生成策略和紧密度计算方法来提高发现质量和算法效率。可视化算法可以引入更多的交互功能和多层次展示方法，以更好地支持大规模网络的分析需求。团检测算法可以通过剪枝策略和启发式搜索来处理更大规模的网络。

算法的鲁棒性和参数敏感性也是重要的改进方向，通过自适应参数调整和多目标优化可以提升算法在不同网络类型上的表现。此外，算法的可解释性也是一个重要考虑因素，通过提供更详细的分析报告和可视化解释可以帮助用户更好地理解分析结果。这些改进方向为未来的研究工作提供了明确的目标和方向。

### 应用前景与实用价值

本实验开发的图分析工具集具有广泛的应用前景和实用价值。在电子商务领域，可以用于商品推荐系统的优化、用户行为分析和市场细分。在学术研究领域，可以支持科研合作网络分析、研究热点发现和学术影响力评估。在社交网络分析中，可以用于社区发现、影响力传播分析和用户画像构建。这些应用场景展示了图论算法在实际问题解决中的重要价值。

工具的模块化设计和良好的扩展性使其能够适应不同的应用需求和数据特征。通过GitHub开源的方式，该工具集可以为研究社区和工业界提供有价值的参考实现。未来可以考虑将其发展为更完整的图分析平台，集成更多的算法和功能，为复杂网络分析提供一站式的解决方案。这种开放式的发展模式有助于推动图论算法的研究和应用，促进相关领域的技术进步。

## 致谢与参考资料

本实验的完成得益于开源社区提供的优秀工具和算法实现，特别是NetworkX图分析库为本项目提供了坚实的技术基础。实验中使用的数据集来源于公开的网络数据库，为算法验证和性能分析提供了宝贵的测试资源。完整的实验代码、数据和结果已在[GitHub仓库](#)中开源发布，欢迎学术界和工业界的研究者使用和改进。

本报告详细记录了图论算法实验的完整过程和主要发现，为复杂网络分析提供了系统性的方法和工具。通过对不同类型网络的深入分析，我们不仅验证了算法的有效性，还揭示了网络结构的内在规律。这些研究成果对于理解复杂系统的组织模式和设计高效的网络分析算法具有重要的理论和实践意义。

