

# 基于图数据库的自定义图分析算法评测

## 1. 评测背景

知识图谱本质是基于图的语义网络，而图数据库又是以图模式存储管理数据，因此图数据库用于存储知识图谱数据具有得天独厚的优势。

通过在知识图谱上进行查询、分析、推理是实现知识图谱应用的关键核心技术，而反映到图数据库上则是基于图数据库上的图查询与图分析。BFS和DFS是图数据库最核心的算法之一，基于此之上，众多学者针对不同需求设计了衍一系列的图查询与图分析算法，如Jaccard相似度算法、Louvain算法、直径估计算法等，且设计了不同变种算法。但遗憾的是，传统的图数据库查询语言并不完全内置了这些算法，如SPARQL，Gremlin等。

因此，通过本次测评将这些图查询和图分析经典算法进行实现并在图数据库中验证，对于检验算法有效性和实用性具有重要意义。

## 2. 任务描述

本任务属于链接数据、知识融合和知识图谱存储管理。即指定一个特定的图查询或分析算法，用户通过实验平台的原子函数和可视化自定义函数编写模块，实现该算法，并借助实验平台验证算法的准确性和效率。以下为指定要求实现的图查询与图分析算法：

- (1) 路径查询类算法：直径估计算法
- (2) 社区发现类算法：Louvain算法
- (3) 重要性分析类算法：介度中心度算法
- (4) 关联性分析类算法：Jaccard相似度算法、度数关联度算法

本评测任务所使用的知识图谱由LDBC SNB Datagen生成的模拟社交网络数据，测试和验证数据集的影响因子分别为SF1、SF10。

## 3. 输入输出

### > 输入

输入为一个特定查询或分析需求，并指定查询算法类型。

### > 输出

输出执行结果，以及程序执行时间等信息。

### > 输入样例

Q1：估算有向图中的直径数（通过直径估计算法计算）。

Q2：对图的社区进行识别，返回各社区成员ID（通过Louvain算法计算）。

Q3：查询节点sn:per00000000000000000933的介度中心度值（通过介度中心度算法计算）

Q4：查询与节点sn:per00000000000000000933相似度最高的5个点（通过Jaccard相似度算法计算）

Q5：计算无向图（子图）的度数关联度值（通过度数关联度算法计算）

### > 输出样例

Q1：输出图的直径数

"8"

Q2：输出划分的各社区信息，每组为社区成员的节点id值

[

[0,1],

[2,4,5],

[3]

]

Q3：返回节点sn:per00000000000000000933的介度中心度值(保留4位小数)

"0.3668"

Q4：返回与节点sn:per00000000000000000933相似度最高的5个点

[2,15,18,133,23]

Q5：计算图（子图）的度数关联度值(保留4位小数)

"-0.1567"

### > 说明

为帮助参赛选手把更多的精力用于算法的实现，将提供竞赛平台为用户提供可视化自定义算法编写和编译功能，并且可通过可视化的操作方式来创建和测试算法。

## 4. 评价指标

本任务的评价指标包括准确率、耗时等两个指标按一定的权重比例进行计分，具体记分标准如下：

- 1) 每个算法为20分，算法结果不准确为0分；
- 2) 在算法准确的基础上，按照执行时间进行排序，耗时最短的得20分，随着执行时间变大，依次分数减少，最小不少于5分；
- 3) 各算法得分累加则为最终得分。

## 5. 数据集

### > 数据来源

本次测评的数据来自于LDBC SNB Datagen生成的模拟社交网络数据，测试和验证数据集的影响因子分别为SF1、SF10。

### > 测试数据

在任务发布后，我们将发布影响因子SF1的测试数据集，以及各个算法执行的准确答案，参赛选手可以在竞赛平台上编写算法并进行测试验证，竞赛平台将实时展示各参赛队伍当前最好成绩榜单。

### > 验证数据

在任务截止提交后，我们会发布影响因子SF10的数据集，比赛平台将重新载入新的数据，并对各个参赛队伍最终提交的算法进行验证，以该算法执行结果和耗时计算得分。

## 6. 任务提交

本次评测将采取刷榜方式，各任务验证集发布后，允许参赛队伍多次向平台提交测试（每天提交不超过100次），参赛队伍的排名在测试结果出来后实时更新。

**最终提交文件要求：**每一个参赛队需提交的材料如下。

1. 相关代码及说明
2. 方法描述文档（非评测论文，评测论文撰写要求见CCKS 2024官网）
3. 在比赛平台上提交最终代码（逾期未提交将默认提交）
4. 以上两个文件需在任务提交截止日期前发送至邮箱  
liwenjiehn@pku.edu.cn。邮件的标题为：“CCKS-gAnalysis-参赛队名称”，例如“CCKS- gAnalysis-火箭队”。
5. 代码及其文档需打包成一个文件（tar, zip, gzip, rar等均可），用code.xxx命名。
6. 本次评测将依托gStore竞赛平台（<http://contest.gstore.cn>）展开，请有意向的参赛队伍关注平台上的竞赛列表。

## 7. 时间安排

评测任务发布：4月20日

报名时间：5月15日 - 8月1日（通过在线表格填报）

<https://docs.qq.com/form/page/DY0FpZUFEEGx5WmVp>

提交阶段：5月15日 - 8月9日

（每天可以重复提交，榜单根据成绩实时刷新）

复测阶段：8月10日 - 8月14日

排名通知：8月15日

评测论文提交：9月1日（暂定，具体时间请参考CCKS2024官网）

CCKS会议日期（评测报告及颁奖）：9月19日 - 22日

## 8. 评审规则

1. 参赛选手需要提交“参赛队名，队长信息（姓名，邮箱，联系电话），参赛单位名称”等信息，报名方式稍后在评测网站发布。
2. 报名截止到测试数据集发布，在测试数据集发布之后，未报名的选手/队伍不能再报名或提交。
3. 每支队伍需指定一名队长，队伍名称不超过15个字符，队伍成员不超过4人。
4. 每名选手只能参加一支队伍，一旦发现某选手以注册多个账号的方式参加多支队伍，将取消所有相关队伍的参赛资格。
5. 允许使用开源代码或工具，但不允许使用任何未公开发布或需要授权的代码或工具。允许使用外部数据，但该数据必须是公开的，并在提交最终结果时一并提交（如数据过大，需提供下载地址）。
6. 鼓励使用通用的、创新的算法、模型解决问题，不允许基于测试集或KB用规则Case By Case解决问题，若发现主办方有权取消参赛资格。
7. 参赛选手最终需要提交可运行的代码和方法描述文档，若在排行榜上的结果无法复现，将取消参赛资格。
8. 欢迎国内外在校生及社会在职人士参加。比赛组织方成员不可参赛。
9. 参加技术创新奖评选的队伍，可以任选一经典图分析算法进行实现，提交

源代码和方法描述文档，发送到liwenjiehn@pku.edu.cn邮箱，组织方经过复测评审后确定获奖队伍。技术创新奖与其他奖项不冲突，可以同时参与。

## 9. 奖励规则

第一名10000,

第二名5000,

第三名2500,

技术创新奖2500

同时排名前三队伍将获授精美参赛奖牌、证书。

## 10. 组织者

邹磊       （北京大学王选计算机研究所）

李文杰     （湖南第一师范学院）