**Neo4j实习说明**

**一、数据集**

**案例数据集stormofswords.csv**：权游中不同角色间的接触数据，数据量352条。

**实习数据集wordnet-valid.csv/wordnet-test.csv**：不同词汇间的关系数据，使用synset\_id代表单词，数据量各5000条。具体id对应的词汇及其释义可参照wordnet-mlj12-definitions.txt。

***Tips：***所有数据集均已上传至大数据平台，使用命令LOAD CSV WITH HEADERS FROM "file:/xxx.csv" AS row可直接导入。

**二、案例参考——Neo4j案例.ipynb**

数据模型(:Character {name})-[:INTERACTS {weight}]->(:Character {name})

主要包括三部分：①导入原始数据；②人物网络分析（人物数量、最短路径、关键节点、节点中心度）；③使用python-igraph（PageRank、社区发现算法）。

***Tips：***实习任务所有任务点的实现均包含在参考案例中，请大家根据实习数据集进行相应的修改实现。

**三、实习任务**

选取wordnet-valid.csv/wordnet-test.csv任一实习数据集进行知识图谱查询分析。

**任务点如下：**

1. 导入数据，输出数据量。同时做唯一限制性约束，确保词汇节点name唯一。
2. 可视化显示部分数据。
3. 统计知识图谱中词汇数量。
4. 统计每个词汇相关的其他词汇数目，输出最小值、最大值、平均值、标准差。
5. 查询网络直径及其长度。
6. 任选两个词汇节点，分别查询任意最短路径和全部最短路径。
7. 查询网络中的关键节点，通过可视化验证。
8. 分析词汇节点中心度（度中心性、介数中心性、紧度中心性）。

***Tips：***鼓励大家在程序分析外附加文字描述，同时可以在实习数据集上尝试其他相关的知识图谱网络分析（Neo4j GDS库包含很多图科学算法）。