



## 课程作业 M - 地理定位网络科学

COMPSCI4077/COMPSCI5107/COMPSCI5078

乔蒙-M-何塞

(i)

- 我们将为您提供一个数据集（在数据文件夹团队中）。编写 python 代码，将推文整理成 1km x 1km 的网格。绘制图表和/或数字，分析数据的分布。

我们收集数据时使用的坐标系为伦敦 = [-0.563, 51.261318, 0.28036, 51.686031]。

- [25]

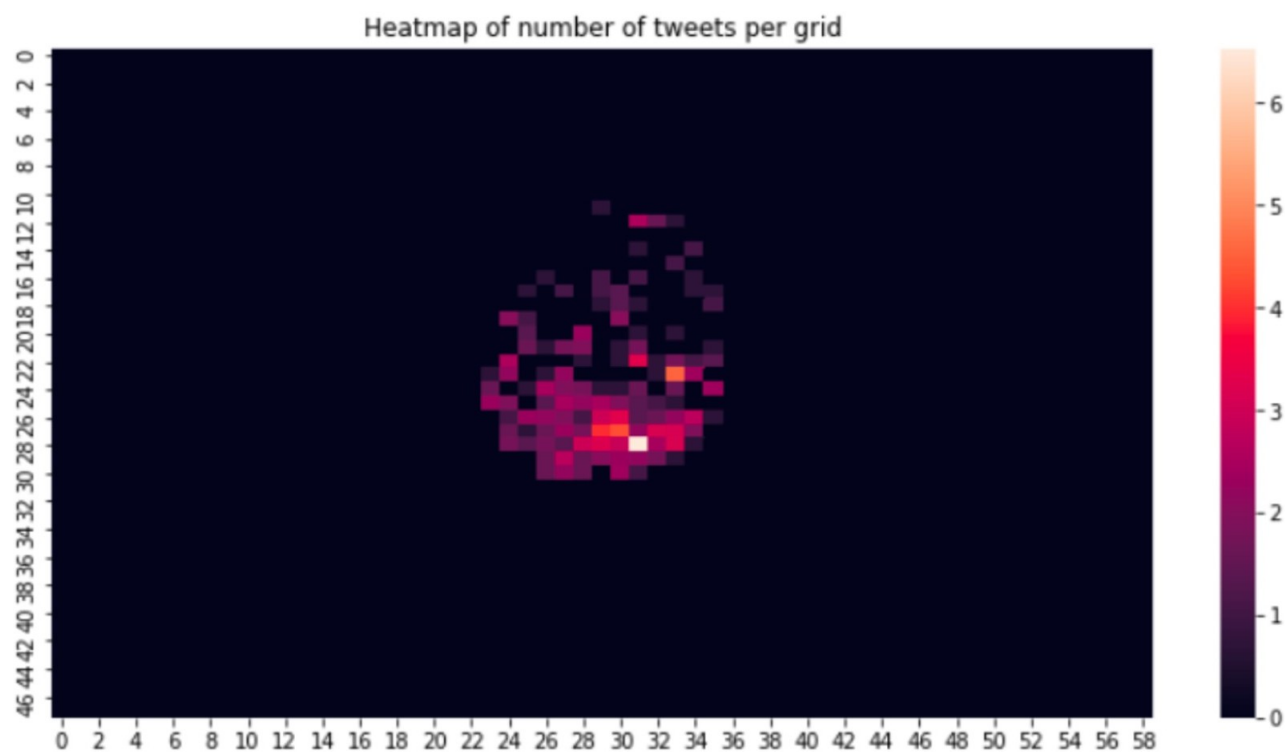
# 解决方案

- 用伪代码描述算法
  - (5 分)
- 提供数据统计（推文总数、单元格中的推文数量以及分布情况等），并解释统计结果--这意味着什么？
  - (5 分)
- 为网格数据的可视化提供图表和/或数字
  - (10 分 - 编码和说明 5 分；显示输出 5 分)

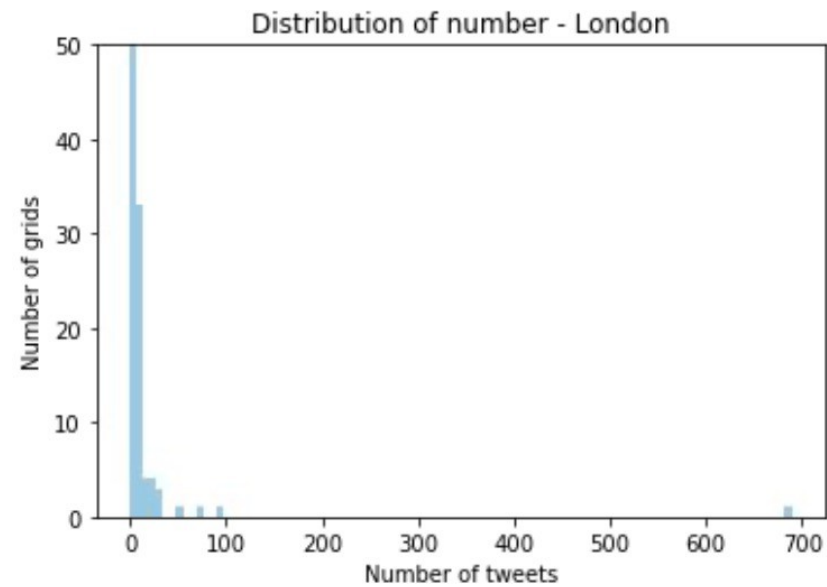
- 描述你对数据（以及由此产生的可视化效果）的看法/解释--你可能想强调使用该数据的任何潜在地理定位技术的问题。

- (5 分)

# 热图--如何解读？



# 数据分布？ 如何解释



(ii)

2. 您将得到一组高质量、低质量和背景推文。基于该数据集开发新闻价值评分方法。通过经验调整阈值来修改新闻价值，并讨论结果。

• [30]

# 解决方案

- 在报告中
- 关于新闻价值评分的幻灯片
- 直接实施
  - 解释您的新闻价值计算方法，并提供算法/伪代码
    - (15 分)
- 进行数据分析并对各种阈值进行分析；数据分析可包括使用或不使用停顿词、调整阈值等。
- 批判性讨论
  - 评分方法的质量
  - 您的高质量、低质量和背景质量数据集



- (15 分)

(iii)

(i) 使用上述新闻价值评分技术分析给出的地理标记数据集 (i)，并讨论结果

• [25]

# 解决方案

- 选取数据集中的每一条推文
- 为其新闻价值打分
- 删除的那个没有新闻价值？
- 保留有新闻价值的
  - 将您的评分方法应用于任务 1 中的数据 -

- 调查得分较低和得分较高的推文；找到合适的阈值将它们区分开来，并删除新闻价值得分较低的推文
- 用您能收集到的任何辅助信息证明所使用的阈值是合理的
- (10分)

- 提供数据统计（推文总数、具有一定新闻价值的推文数量、分布情况等。）
- (5 分)

在有新闻价值的数据上应用您创建的可视化方法；绘制数字/图表，并将其与 (1) 中的结果进行比较。有什么不同？

- 重绘热图和直方图
  - 10 分

# [开放式任务] - 10 分

- 举例说明并讨论因推文或消息来源的性质而导致的地理定位问题
- 这里的想法是展示你的技能；将你的知识应用到可能是现实生活中的任务中去
- 探索新闻价值得分和变化
- 我可能在课堂上提到过其中的一些内容
  - 可能不在幻灯片中
- 本部分用于奖励举措

# 报告 - 10 分

- a. 结构和格式
- b. 阐述观点
  - a. 得出的新见解
- c. 创造性地完成任务
  - a. 不仅仅是照搬斯利奥德斯的要点