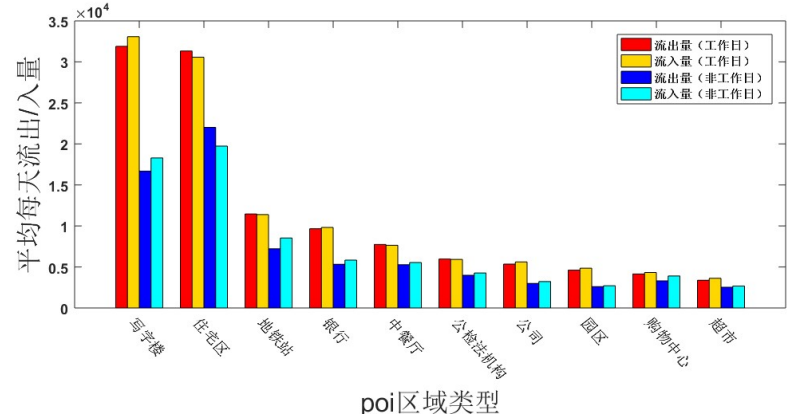
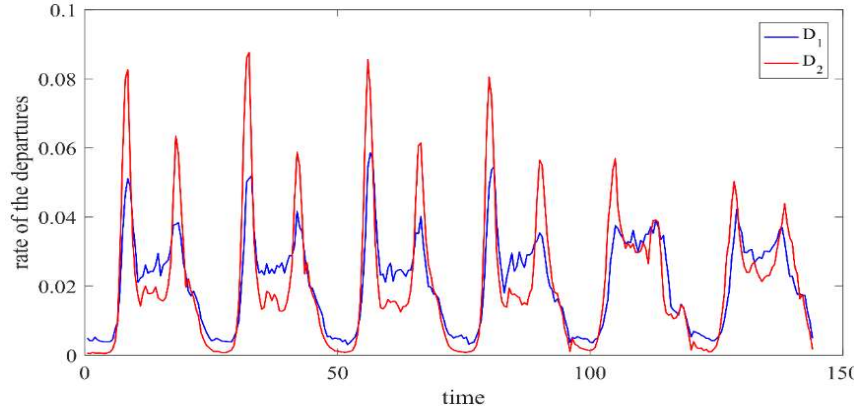
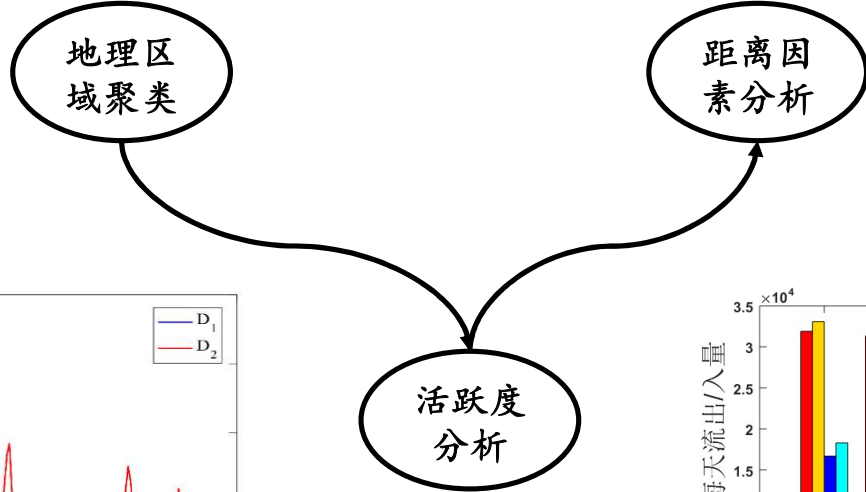
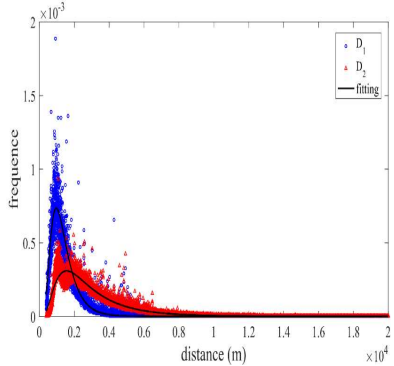
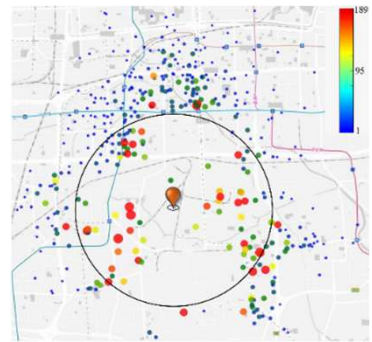
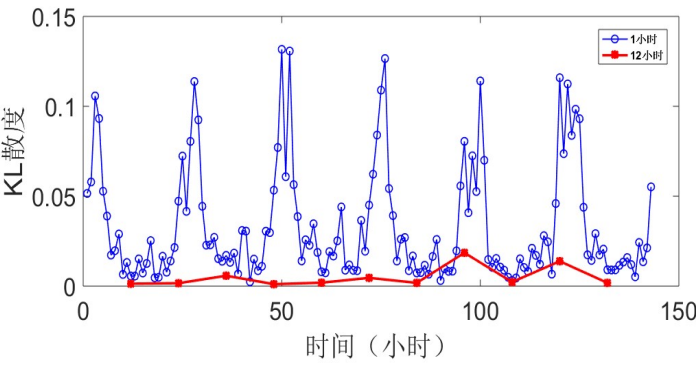
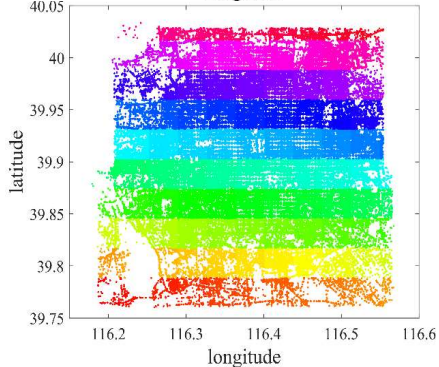
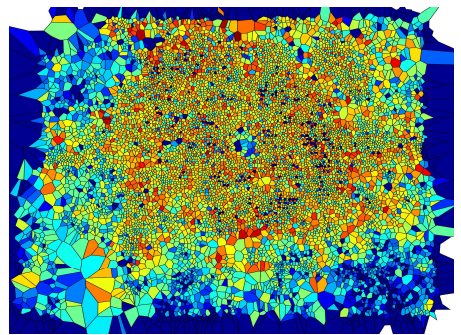
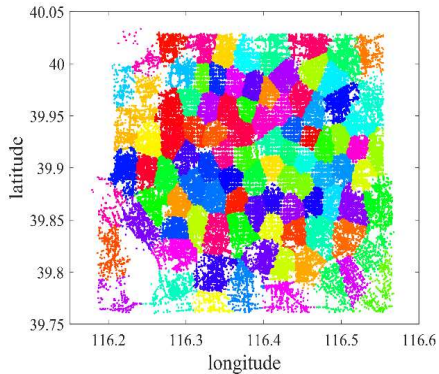


基于城市兴趣点聚类的无桩共享单车时空流动模型

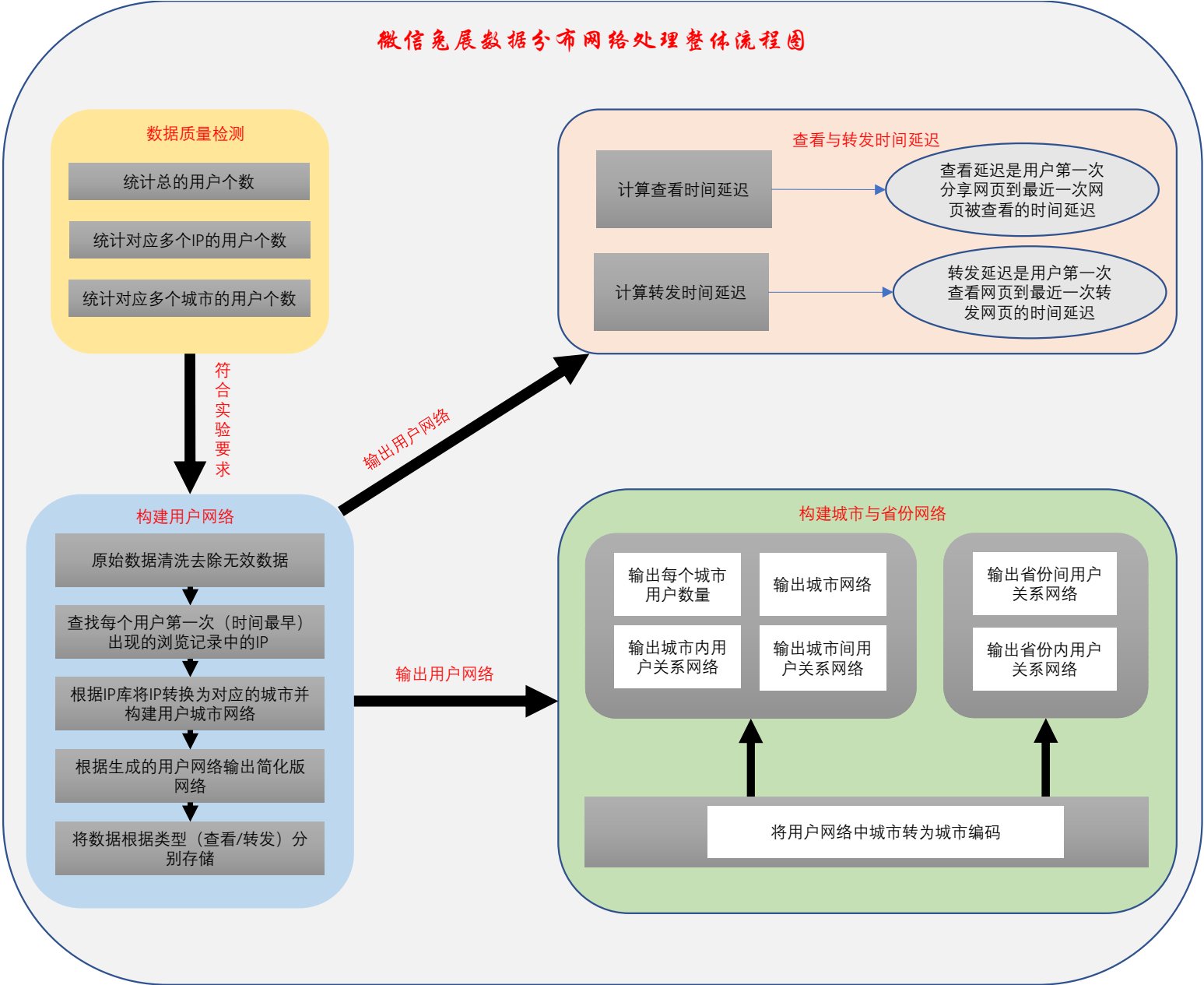
- ofo: 平均52w辆, 平均骑行1.5km, 10min
- mobike: 平均41w辆, 平均骑行3km, 30min

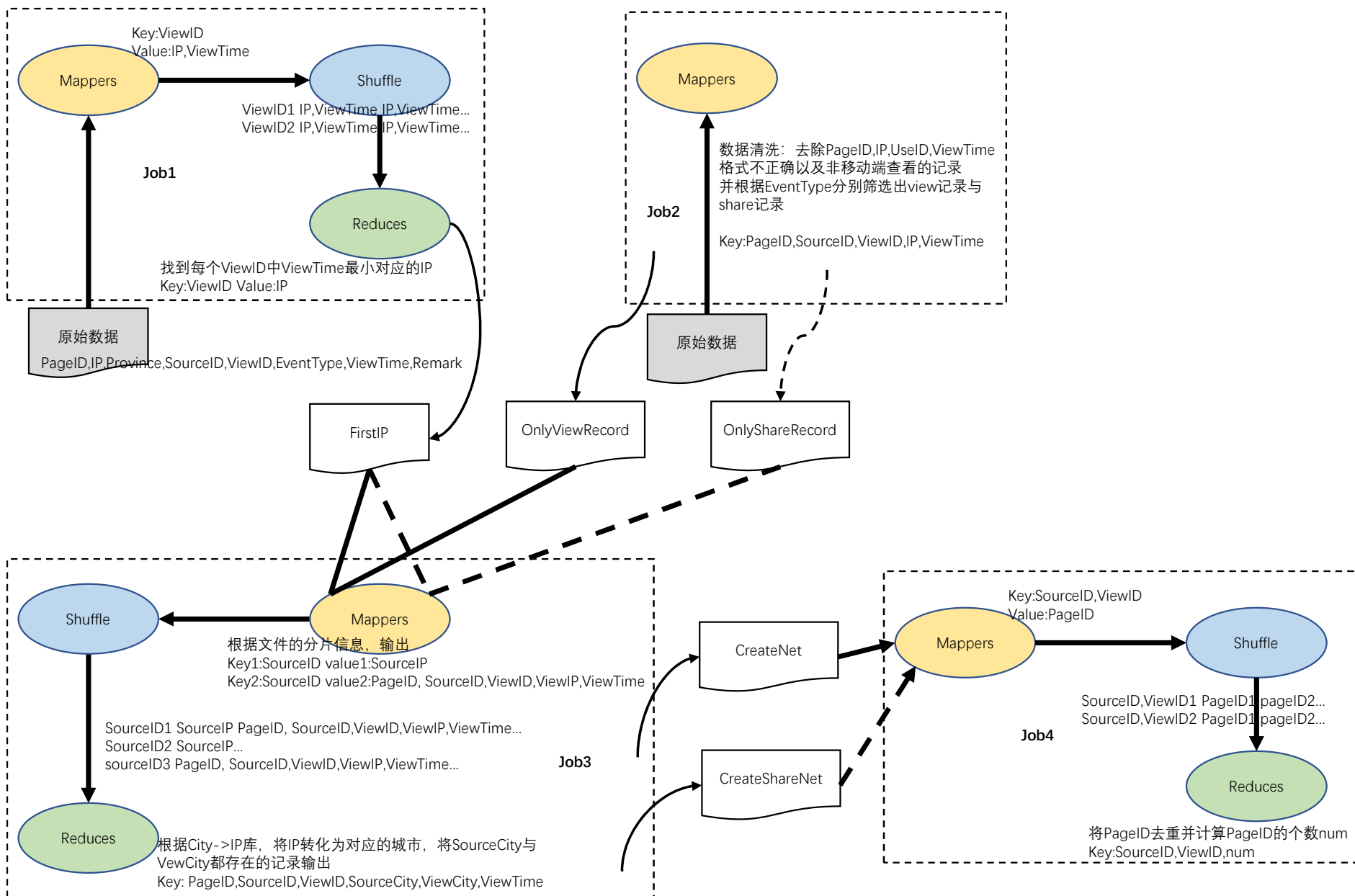


基于MapReduce的微信数据处理流程



微信免展数据分布网络处理整体流程图





数据清洗

查找IP

微信数据集：每条记录包含博文ID、用户ID（源用户，查看用户）、查看用户所在省份、查看用户IP、用户浏览或转发博文时间等信息

去除博文ID、用户ID、查看用户IP格式错误以及非移动端查看的记录

查看用户是否转发博文？

只查看

转发

输出数据：每条记录包含博文ID、用户ID、查看用户IP以及时间信息

输出数据：每条记录包含博文ID、用户ID、查看用户IP以及时间信息

查找源用户第一次（时间最早）浏览记录中对应的IP，即源用户IP

输出数据：每条记录包含源用户ID、源用户IP

将两个文件根据源用户ID进行连接并根据ip库将ip转换为对应城市

查看

转发

输出数据：每条记录包含博文ID、用户ID、用户城市以及时间信息

输出数据：每条记录包含博文ID、用户ID、用户城市以及时间信息

提取用户关系网络

地理网络特征分析

用户关系网络：每条记录包含博文ID、用户ID、用户城市以及时间信息

对每个城市进行编码，将用户关系网络中的用户城市转成对应的城市编码

每个城市对应的用户个数：每个记录包含查看用户城市编码、查看用户个数

构建城市内/间的用户关系网络：每个记录包含用户城市编码、用户ID

构建省份内/间的用户关系网络：每个记录包含用户省份编码、用户ID

城市内用户查看/被查看行为分布：每个记录包含用户ID、用户被本城市、其他城市查看/被查看数量

城市关系网络：每个记录包含用户城市编码、查看用户数量、转发用户数量

用户网络时间特征分析

用户关系网络：每条记录包含博文ID、用户ID、用户城市以及时间信息

要求同一个pageID同一个用户只记录一次行为，根据时间找最小查看或者转发时间的记录，并解析出时间对应的小时（0-23），星期（0-6）

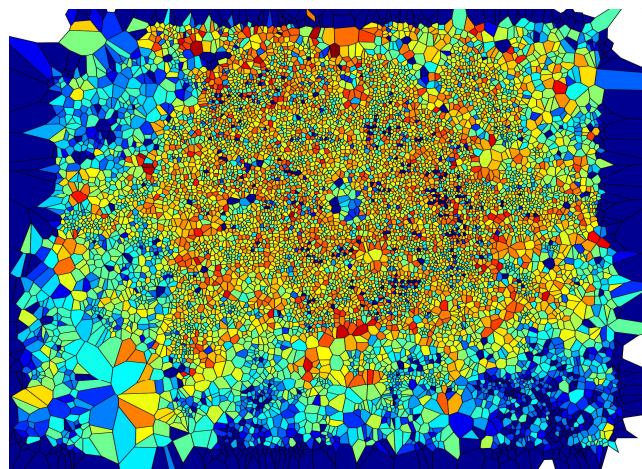
转发延迟：用户第一次查看网页到最近一次转发网页的时间延迟

查看延迟：用户第一次分享网页到最近一次网页被查看的时间延迟

输出数据：每个记录包含用户ID、查看时间、转发时间、转发延迟

输出数据：每个记录包含用户ID、查看时间、转发时间、查看延迟

共享单车一日流动停放仿真设计与分析



时间分布模型

OD转移矩阵

区域聚类

特征分析

模型构建

仿真系统设计

模型验证

