



# Research Internship Summary



**Reporter: Tan Kaizhen (Undergraduate student)**  
**Instructor: Huan Weihua (PhD student)**  
**Huang Wei (Professor)**



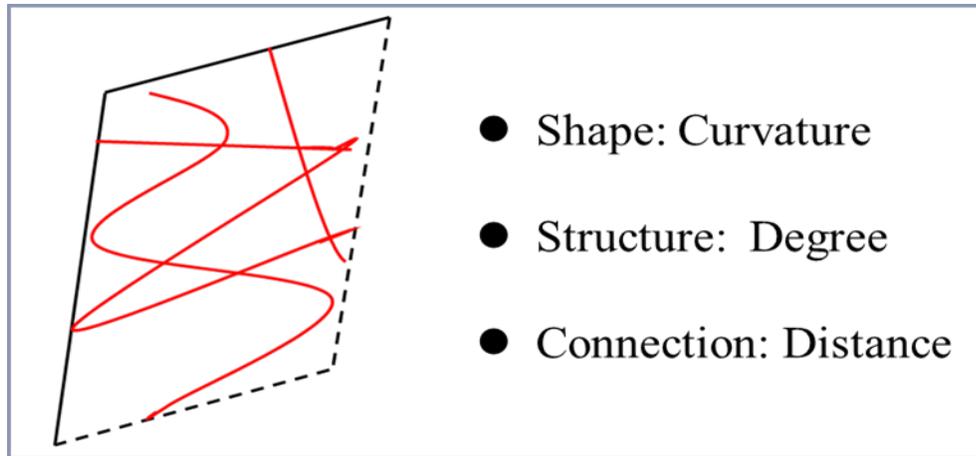
## Main Topic: Spatio-temporal Interaction Mechanism of Human Activities and Traffic Congestion Propagation

1、利用时空轨迹-POI语义数据，对交通拥堵传播模式进行聚类，目的是为了研究拥堵传播这一动态过程的特征。



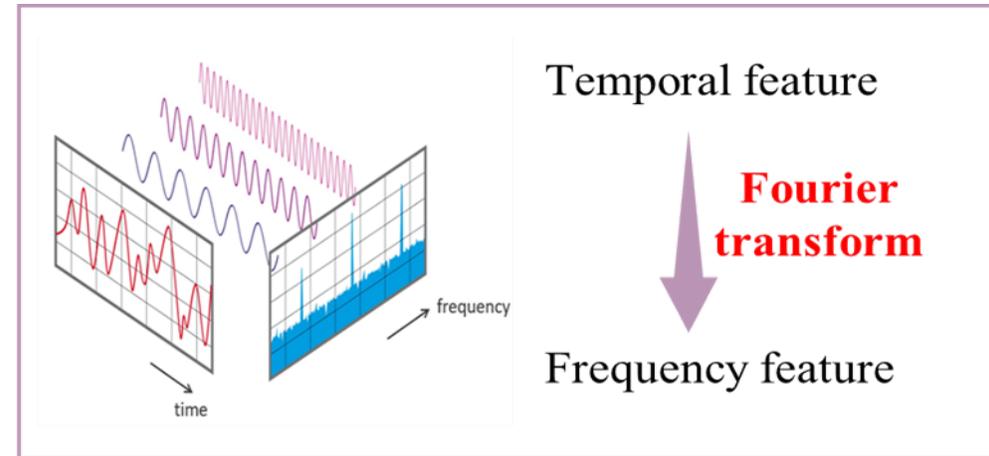
2、利用时空轨迹数据，对动态拥堵网络图进行社区挖掘，找出有显著差异的相邻时间切片，主要做算法层面的创新；再结合POI语义数据，发现一些造成差异的关联性。

## Embedded graphs construction



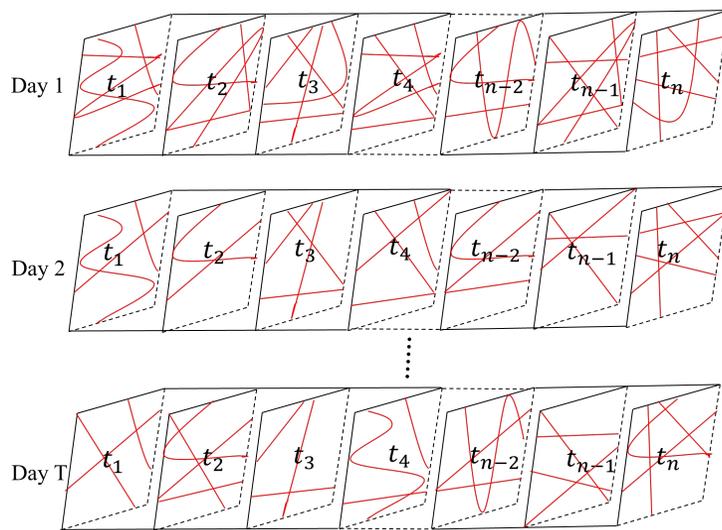
Prior knowledge  
(Topological features)

+

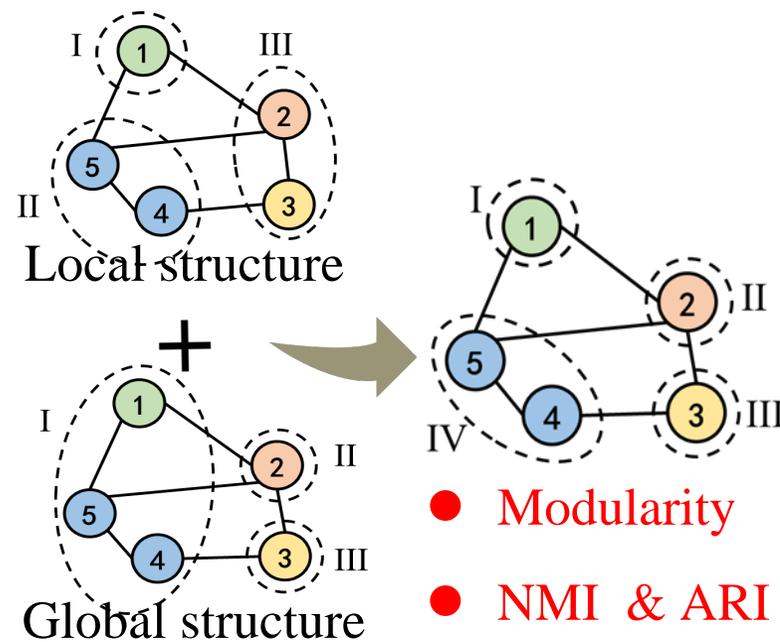


Data-driven knowledge  
(TTI → Frequency features)

## Adaptive modularity optimization



Dynamic congestion networks



Communities detection & Evaluation

## 2、Validation and Analysis

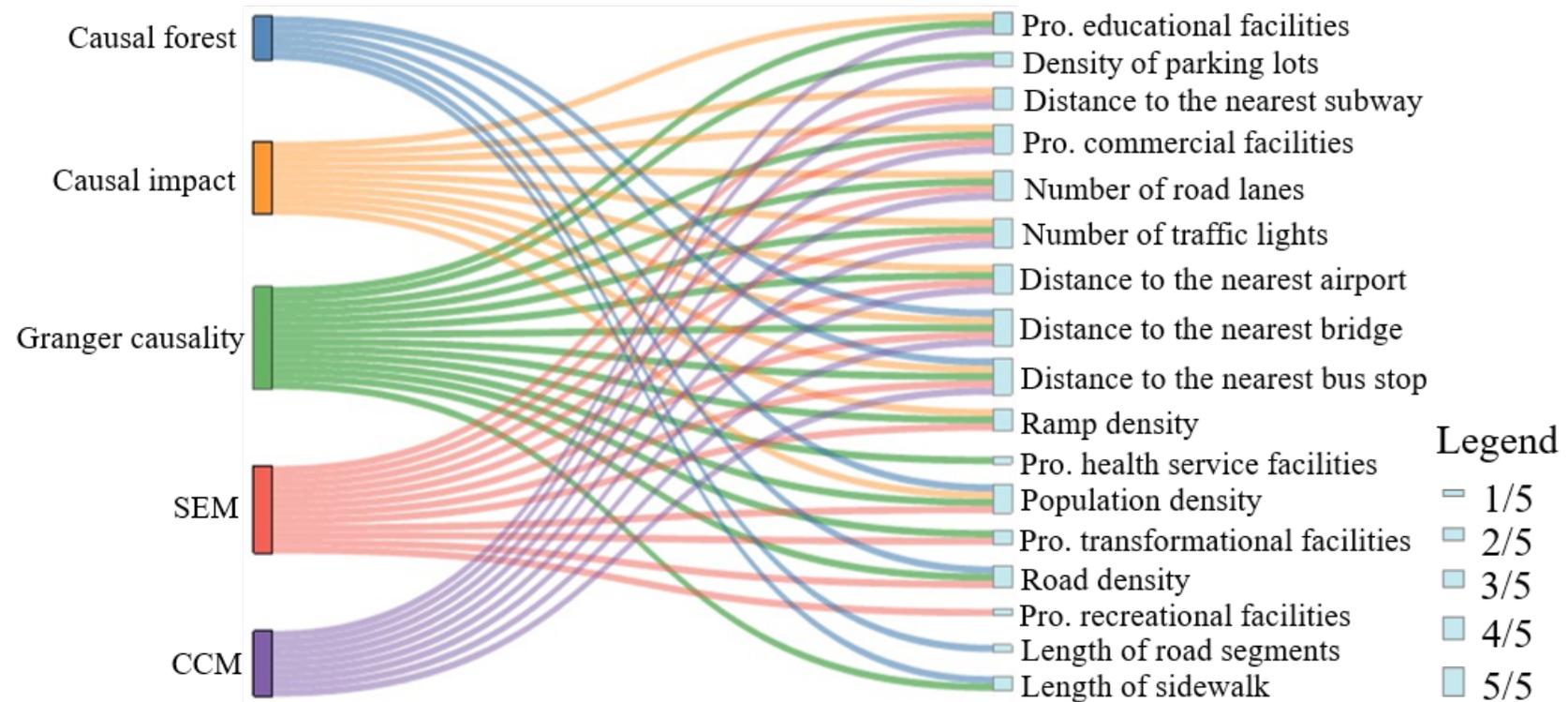
### Benchmark dataset

- Four types of dynamic benchmark graphs
- 15000 vertices

### Real world dataset

- Average speed dataset in New York City  
(Dec 2, 2018 - Dec 31, 2018)
- Average speed dataset in Shanghai  
(April 1, 2015 – April 30, 2015)

Visualizing the interaction between detected communities and the 11 causal factors



1. 提出了道路的空间拓扑特征的概念，融合了距离、结构 (node degree)、形状 (curvature) 等特征值。
2. 对于道路tti的时间序列数据，添加了傅里叶变换，将时间域特征转化为频域特征，消除了数据采集的时间分辨率对于结果的影响。
3. 运用了一种改进后的社区挖掘方法识别动态网络图中发生的结构变化，追踪拥堵传播 (Adaptive modularity-based method for tracking congestion propagation in dynamic traffic networks)，并与传统的基于模块度和LDA (主题模型) 的社区挖掘方法进行对比。

静态网络：每一时刻的节点数目不变，结构不变。

动态网络：节点数目会变化，结构会变化。

现有研究大多考虑的是局部结构的演化。

本研究同时考虑了网络图的局部特征和全局特征 (整体结构差异)。

使用NMI和ARI作为前后时间切片差异的衡量指标，使用模块度作为社区挖掘效果好坏的衡量指标

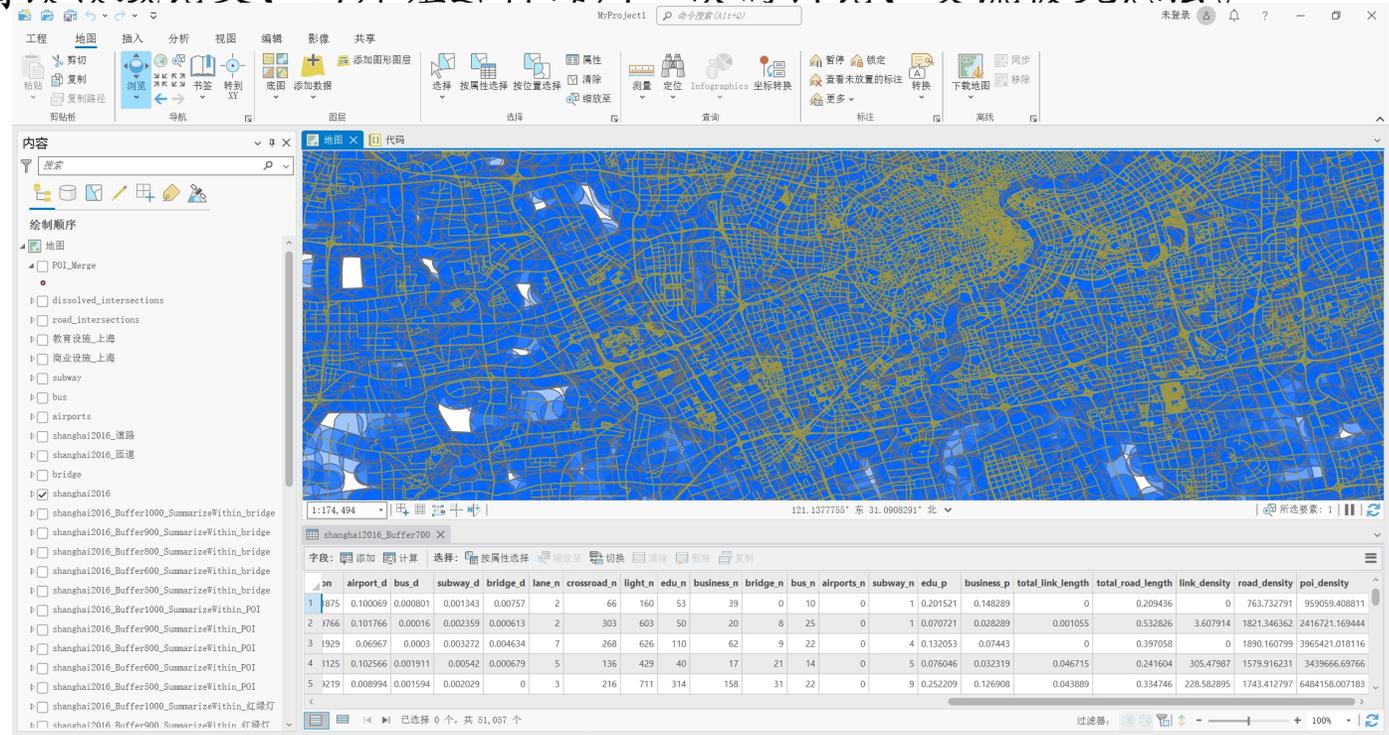
。

# 4、My Contribution

1. Map matching, 计算每个时间切片中每条道路上的平均速度和tti指数, 得到拥堵网络图。
2. 计算拥堵道路的时间相似度, 得到特征矩阵, 方便师姐进行后续处理。
3. 提出了道路的拓扑结构相似性的概念, 提出聚类算法并不适用于作为动态过程的拥堵传播这一观点。
4. 处理了所有POI数据, 创建道路缓冲区, 将11个拥堵权重因子添加到映射; 验证缓冲区的最佳大小。

阅读领域论文, 与师姐进行每周一次的讨论, 交流彼此想法。

名称	修改日期	类型	大小
Temporal_similarity_0401_1.csv	2024/5/28 21:04	Microsoft Excel 逗...	3,353 KB
test6.py	2024/5/28 21:03	JetBrains PyCharm ...	2 KB
test3.py	2024/5/28 18:21	JetBrains PyCharm ...	3 KB
test4.py	2024/5/28 18:20	JetBrains PyCharm ...	3 KB
Temporal_similarity_0401_001.csv	2024/5/28 17:56	Microsoft Excel 逗...	3,151 KB
Temporal_similarity_0401_000001.csv	2024/5/28 17:47	Microsoft Excel 逗...	3,187 KB
新建 文本文档.txt	2024/5/28 17:16	文本文档	0 KB
calculate_similarity_matrix.py	2024/5/28 16:58	JetBrains PyCharm ...	3 KB
Temporal_similarity_0401_0001.csv	2024/5/28 15:38	Microsoft Excel 逗...	3,151 KB
Temporal_similarity_0401_0000.csv	2024/5/28 15:07	Microsoft Excel 逗...	3,172 KB
888.py	2024/5/27 23:38	JetBrains PyCharm ...	2 KB
TSI_subgraph_0401.csv	2024/5/27 22:36	Microsoft Excel 逗...	3,352 KB
CDD_similarity_0401_0.csv	2024/5/27 22:36	Microsoft Excel 逗...	3,178 KB
map_matching.py	2024/5/15 23:42	JetBrains PyCharm ...	3 KB
Index_Shanghai2016.zip	2024/5/13 14:59	压缩(zipped)文件夹	6,176 KB
average_speeds.csv	2024/5/12 0:16	Microsoft Excel 逗...	76,223 KB
要求.txt	2024/5/12 0:05	文本文档	7 KB
求平均速度.py	2024/5/12 0:01	JetBrains PyCharm ...	2 KB
2.py	2024/5/11 20:36	JetBrains PyCharm ...	3 KB
1.py	2024/5/11 20:35	JetBrains PyCharm ...	2 KB
Index_Shanghai2016.dbf	2024/5/9 16:15	DBF 文件	52,006 KB
Index_Shanghai2016.shp	2024/5/9 16:15	SHP 文件	10,088 KB
Index_Shanghai2016.shx	2024/5/9 16:15	SHX 文件	399 KB
Index_Shanghai2016.cpg	2024/5/9 16:15	CPG 文件	1 KB
Index_Shanghai2016.prj	2024/5/9 16:15	PRJ 文件	1 KB



## 挑战：

- 1、Map matching时，如何区分高架/快速路和地面上的GPS轨迹点？
- 2、得到道路拥堵时间切片图以后，怎么将拥堵通过道路传播这一动态的过程表现出来？

## 解决方法：

1、领域内经典问题，至今没有什么特别好的解决方案。在行驶过程中，高架路和地面道路的轨迹路径和轨迹速率的变化频率不同。高架路上的路径通常较长且变化少，而地面道路则更为复杂；高架路上的速率是比较平稳和连续的，而地面道路上则会显示较多的停顿或变速。

可是，我们研究的是交通拥堵，而高架上也是有可能发生拥堵的，如果把所有的停顿变速视为在地面道路上发生的事件，是否会降低检测到在高架上发生拥堵的概率？

2、社区发生变化—>拥堵传播过程。在每一时刻，我们都需要考虑前一刻的拥堵图结构。对每一个子图使用改进后的社团挖掘算法，并得到相邻时刻子图之间的差异性（用AMI, ARI来衡量），进而找出有显著差异的拥堵路段图（相邻时间切片），找出发生异常的时刻。然后再根据已创建的，融合建成环境数据POI数据的buffer静态图，找到有显著差异的拥堵路段图与buffer静态图的映射，发现一些关联性，分析形成原因。

科研训练——天马行空的思维与现实的引力

思维模式的对比：可行性、可量化和跳脱框架

参数设定的严谨：理论如何支撑实际

数据的重要性：数据质量、采集难度、获取难度

祛魅——看山还是山

真实意义的消解：工作与个人提升

科学系统的搭建：点点创新之积累