Measuring the vibrancy of Austin neighborhoods using taxi data with PageRank algorithm

Yalin Yang
Doctoral student
Geographic information science

- Research objective

- Study Area and dataset

- Methods

- Results and conclusion

- Future Work

- Reference

Research objective

Measuring the vibrancy of Austin neighborhoods using taxi data with PageRank algorithm

- Vibrancy

  - Building a more vibrant city is one important goal of urban development (Montgomery, 1998).

- Estimation

Study Area and Dataset

Taxi dataset

	⊙ completed_on ⌄	# distance_travelled ⌄	# end_location_lat ⌄	# end_location_long ⌄	⊕ end_location_location ⌄
1	2016-06-04T04:35:56	285.0	30.27	-97.75	POINT(-97.75 30.27)
2	2016-06-04T04:51:15	1029.0	30.27	-97.74	POINT(-97.74 30.27)
3	2016-06-04T05:27:32	8459.0	38.68	-121.04	POINT(-121.04 38.68)
4	2016-06-04T06:51:49	443.0	38.68	-121.04	POINT(-121.04 38.68)
5	2016-06-04T08:17:57	568.0	38.68	-121.04	POINT(-121.04 38.68)
6	2016-06-04T15:13:38	4051.0	30.27	-97.74	POINT(-97.74 30.27)
7	2016-06-04T15:26:07	790.0	30.27	-97.75	POINT(-97.75 30.27)
8	2016-06-05T03:50:57	2171.0	30.27	-97.75	POINT(-97.75 30.27)
9	2016-06-05T04:33:20	10260.0	30.27	-97.75	POINT(-97.75 30.27)
10	2016-06-05T07:12:48	5294.0	30.24	-97.78	POINT(-97.78 30.24)

RideAustin_Weather.csv   >_ Query   ⬇ Download
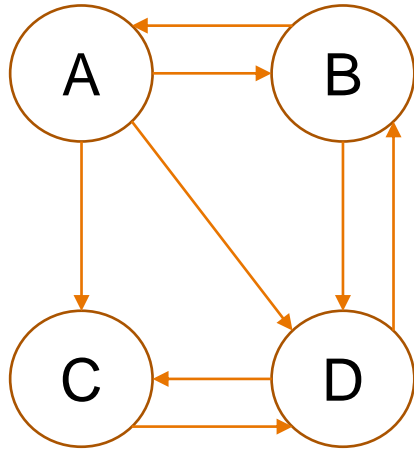
- 481137 records
- 11/2016 – 2/2017

Austin boundary shapefile

Methods

PageRank algorithm

$$Lr = r$$

Start with $r = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, L = \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 1 & 0 \end{bmatrix}$
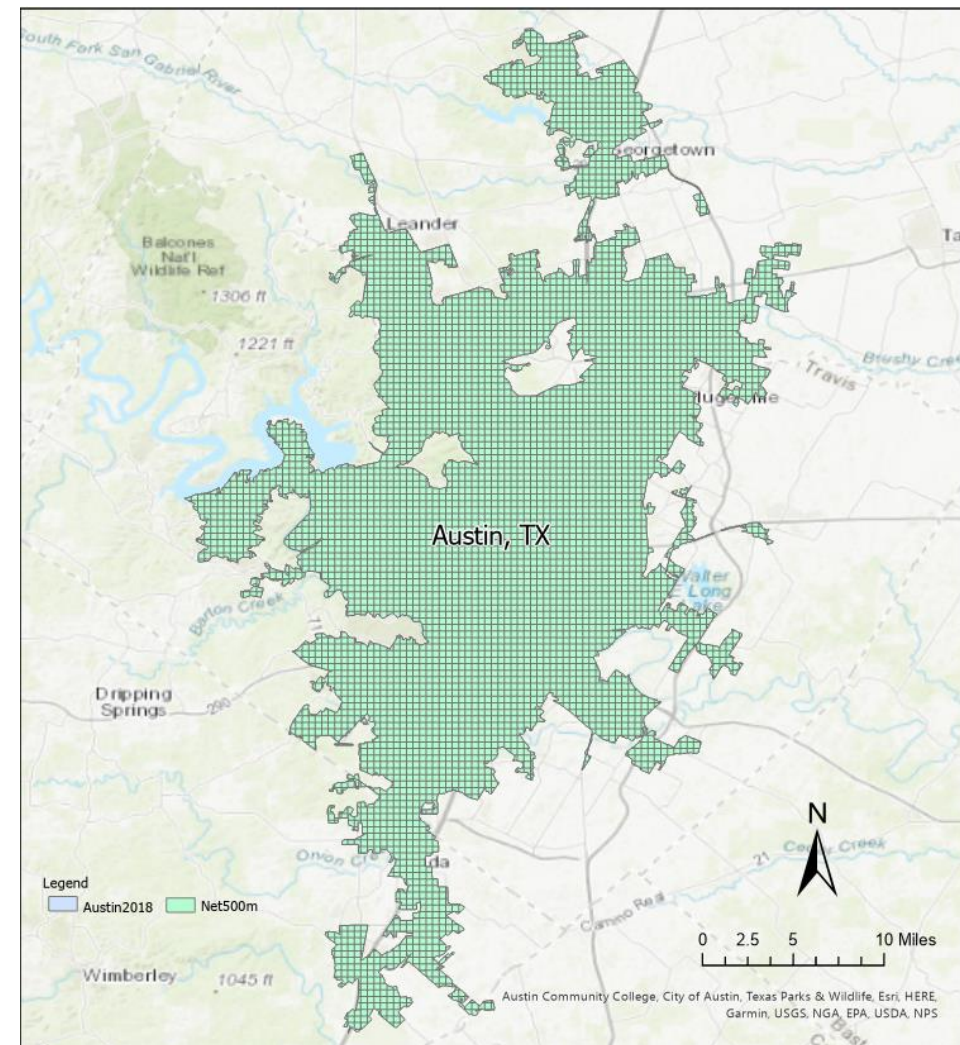
Add damping factor

$$r = d(L \cdot r) + \frac{1-d}{n}$$

$$r = \begin{bmatrix} r_A \\ r_B \\ r_C \\ r_D \end{bmatrix}, L = \begin{bmatrix} L_{A \to A} & L_{B \to A} & L_{C \to A} & L_{D \to A} \\ L_{A \to B} & L_{B \to B} & L_{C \to B} & L_{D \to B} \\ L_{A \to C} & L_{B \to C} & L_{C \to C} & L_{D \to C} \\ L_{A \to D} & L_{B \to D} & L_{C \to D} & L_{D \to D} \end{bmatrix}$$

Methods

Austin fishnet map

Workflow
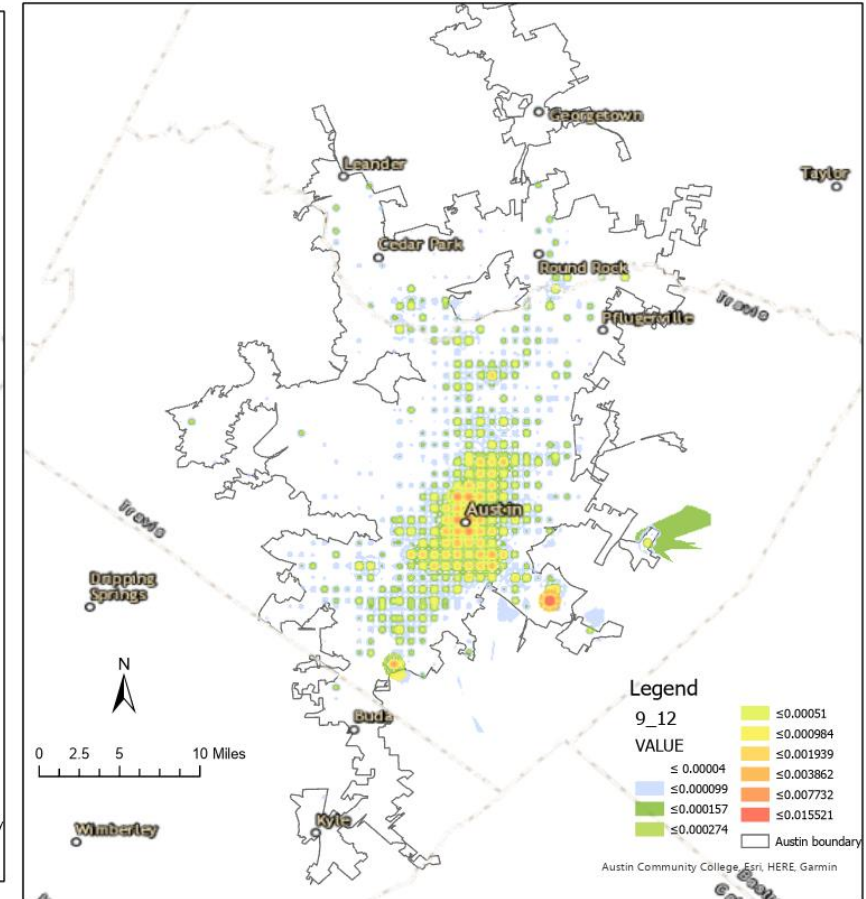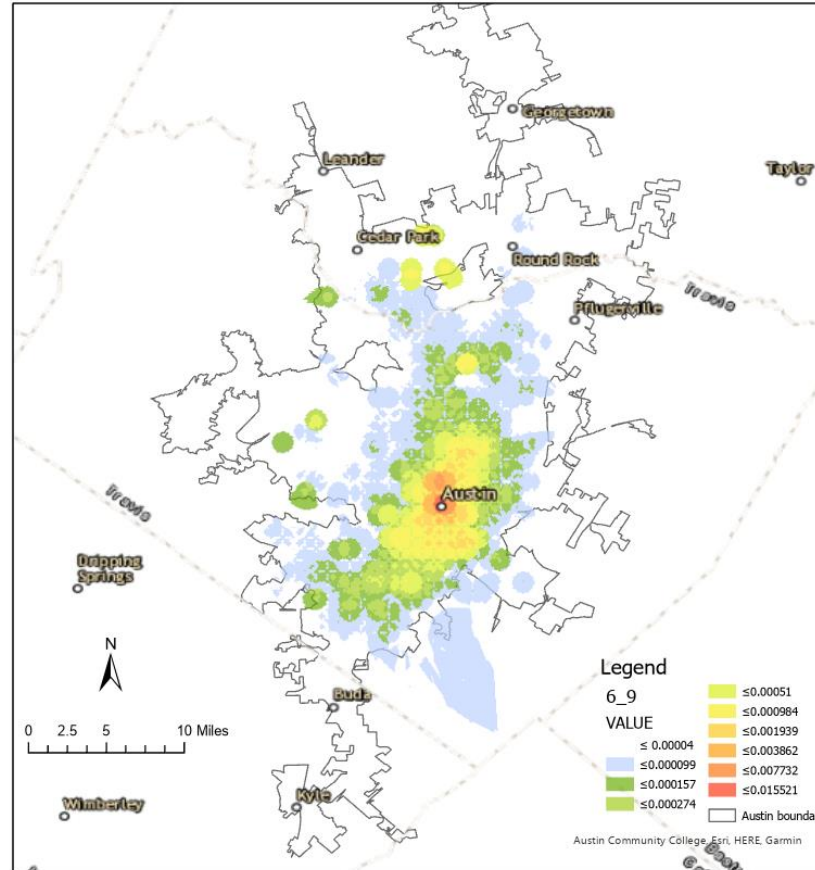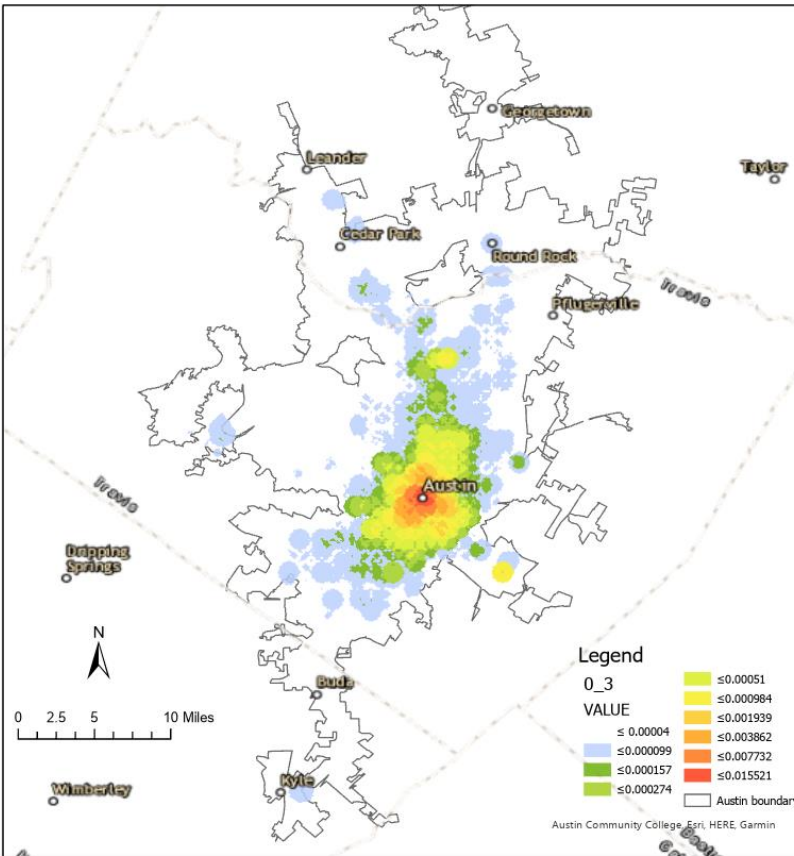
- Generate grid (500*500)

- Filter taxi data by different time period

- Use PageRank algorithm to get rank of each grid

- Interpret results with POI

Results (legend are shared)

Hourly vibrancy in Austin : 0:00 - 3:00

Hourly vibrancy in Austin : 6:00 - 9:00

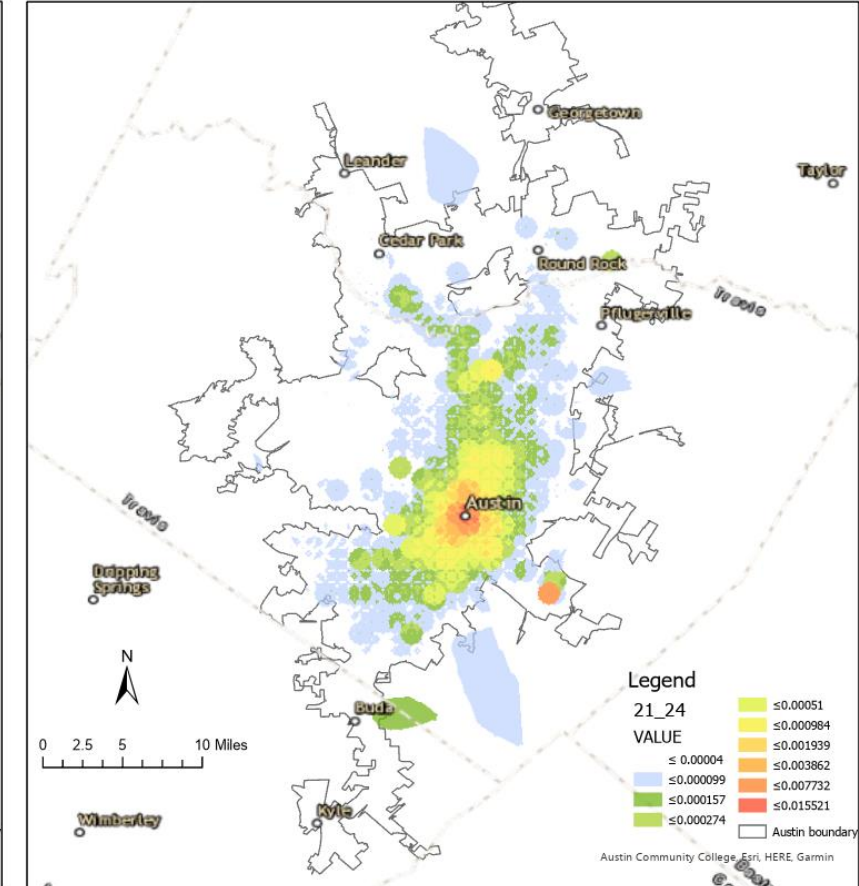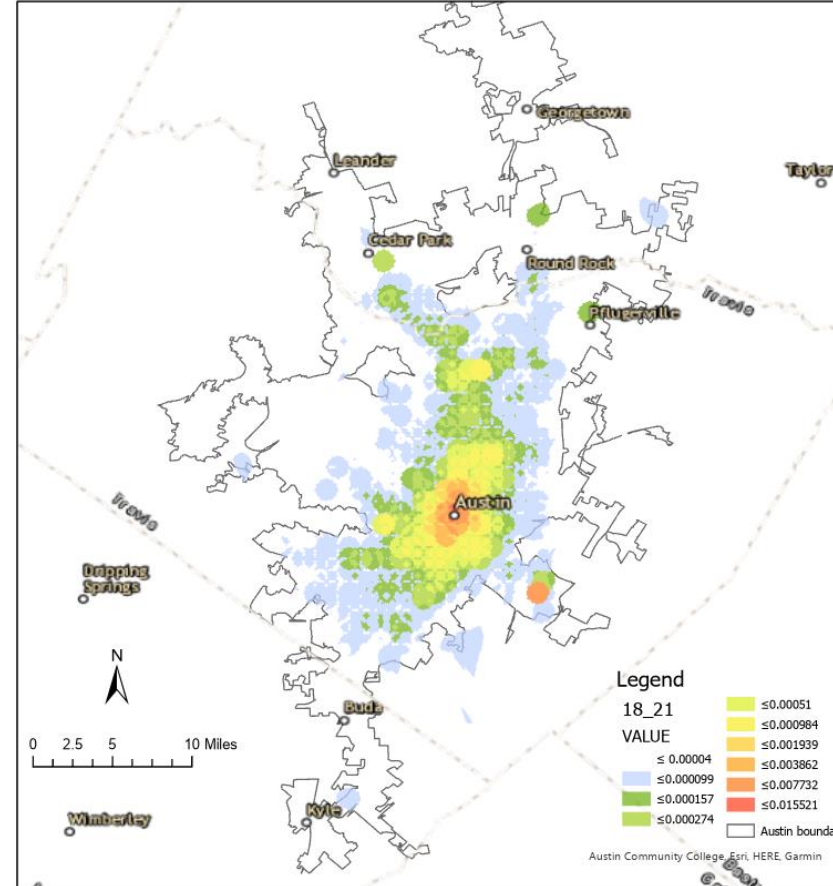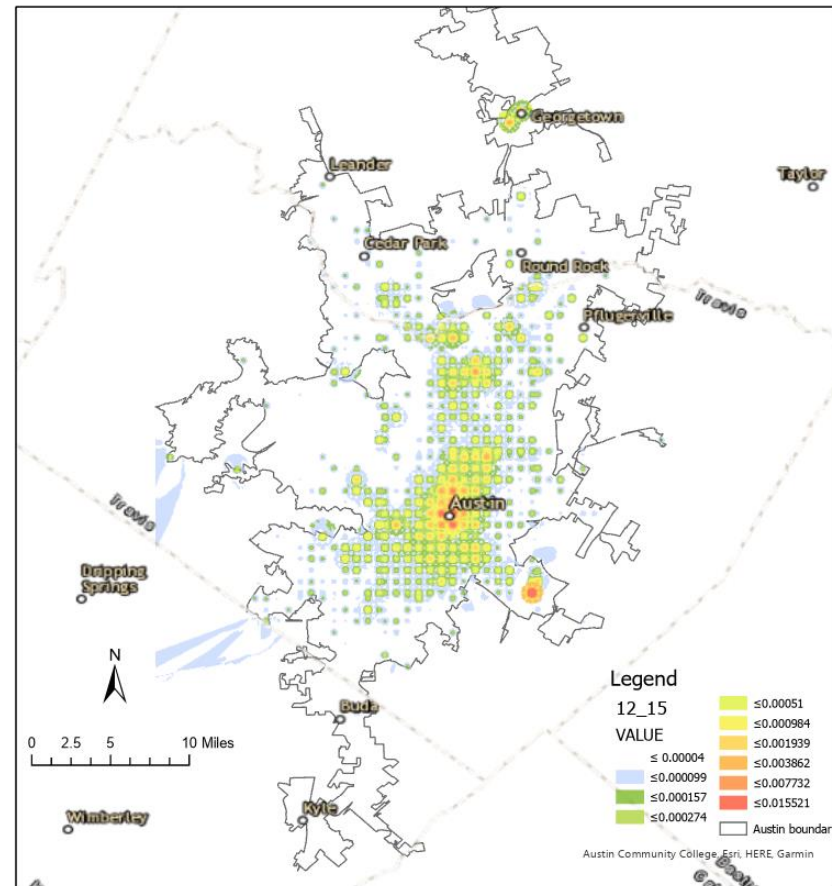Hourly vibrancy in Austin : 9:00 - 12:00

Results (legend are shared)

Hourly vibrancy in Austin : 12:00 - 15:00

Hourly vibrancy in Austin : 18:00 - 21:00

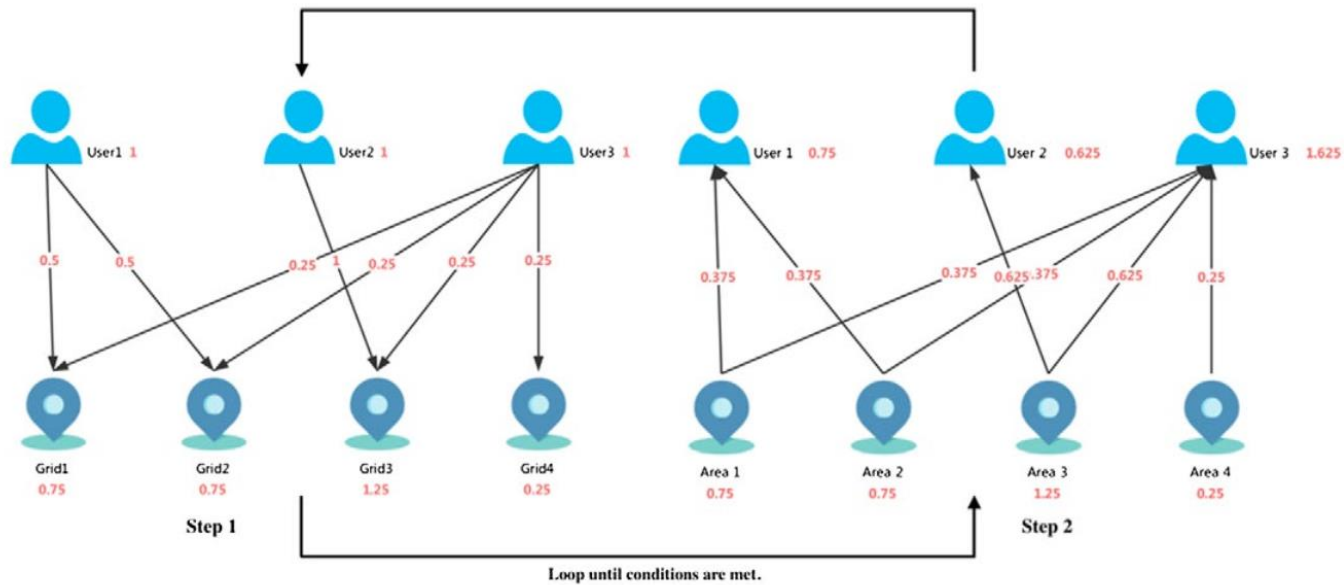Hourly vibrancy in Austin : 21:00 - 24:00

Conclusion

- Obvious temporal and spatial pattern are discovered (feasible)

- Drawbacks

  - Only one factor within consideration

  - Pre-biased data are used

```python
In [14]: for index in range(len(time_range)-1):
             tv =  [filter_by_hour(data,time_range[index],time_range[index+1]) for data in completed_time]
             tdf = df[tv]
             print(f'{time_range[index]} : {tdf.shape}')
             fn = os.path.join('../Data/Table','RideAustin_' + str(time_range[index]) +'_' + str(time_range[index+1]) +'.csv')
             tdf.to_csv(fn)
0 : (95296, 8)
3 : (100151, 8)
6 : (77979, 8)
9 : (18060, 8)
12 : (25794, 8)
15 : (45121, 8)
18 : (49425, 8)
21 : (60570, 8)
```
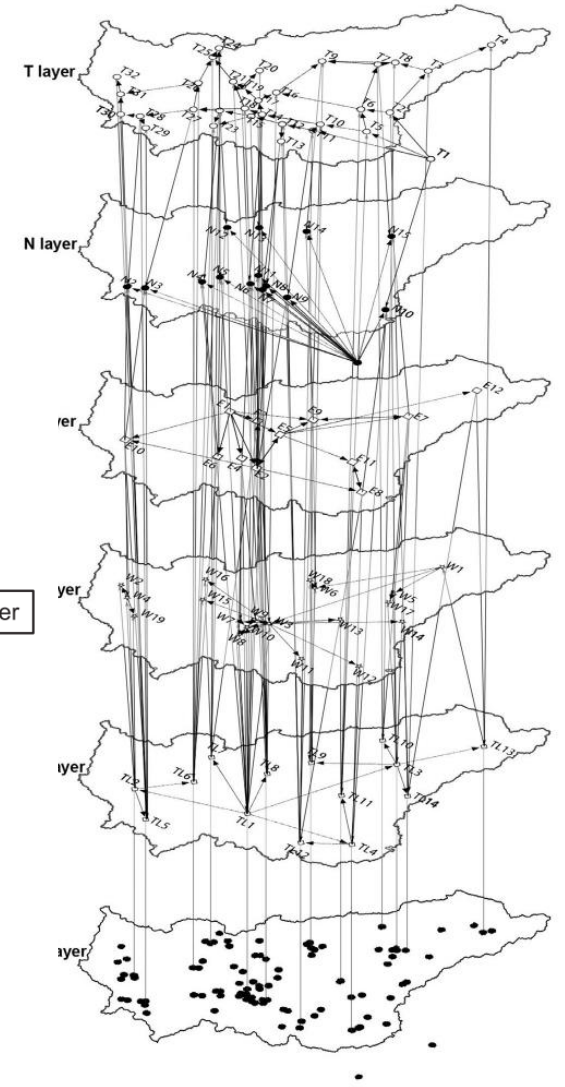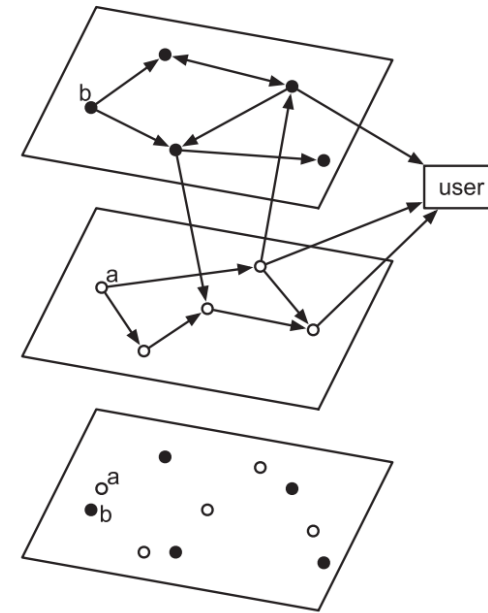
Future Work

- Using trajectory data to implement PR algorithm
  - Add weight factor based on distance and number of nodes

- Build multilayer network model
  - Add multiple related data

Reference

- Jia, Chen, et al. "Measuring the vibrancy of urban neighborhoods using mobile phone data with an improved PageRank algorithm." *Transactions in GIS* 23.2 (2019): 241-258.

- Zhao, Chen, Nan Li, and Dongping Fang. "Criticality assessment of urban interdependent lifeline systems using a biased PageRank algorithm and a multilayer weighted directed network model." *International Journal of Critical Infrastructure Protection* 22 (2018): 100-112.