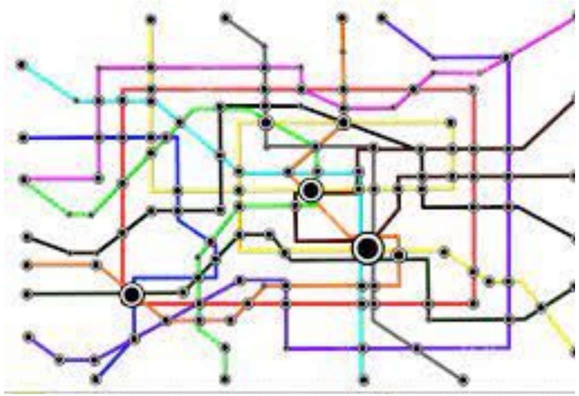
Network Analysis on Four Different City Subways

Wheaton Wang, Kyle Yu, Brandon Bai.

Overview

- Subway network is composed of stations(Node) and routes(Edge).
- Designing a subway network is complex and costly.

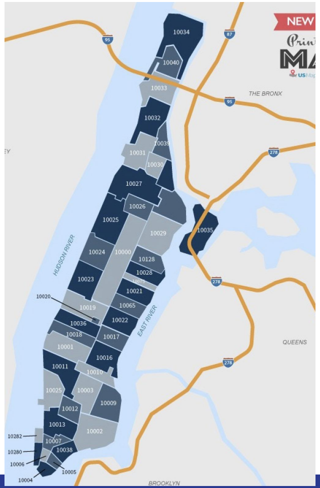


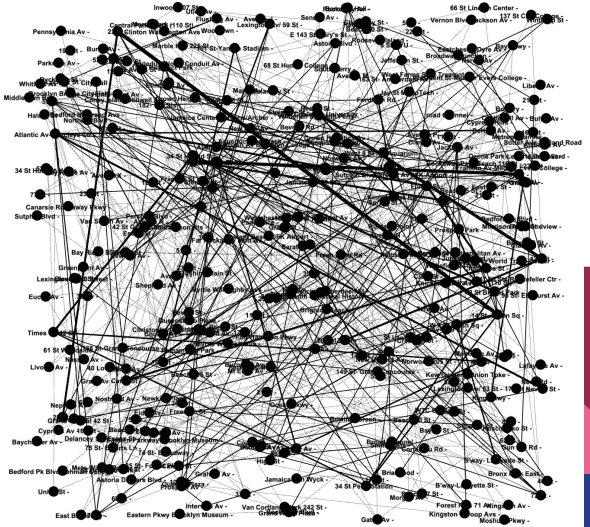
Goals

- By analyzing existing subway networks, we want to…
  - Milestone 1: Find important factors affecting the site selection of subway stations.
  - Milestone 2 & 3: Be able to analyze some famous cities' subways and decide if they are successful subways (e.g. performance under attack).
  - Aim to solve the problem to potentially save money for city subway system.

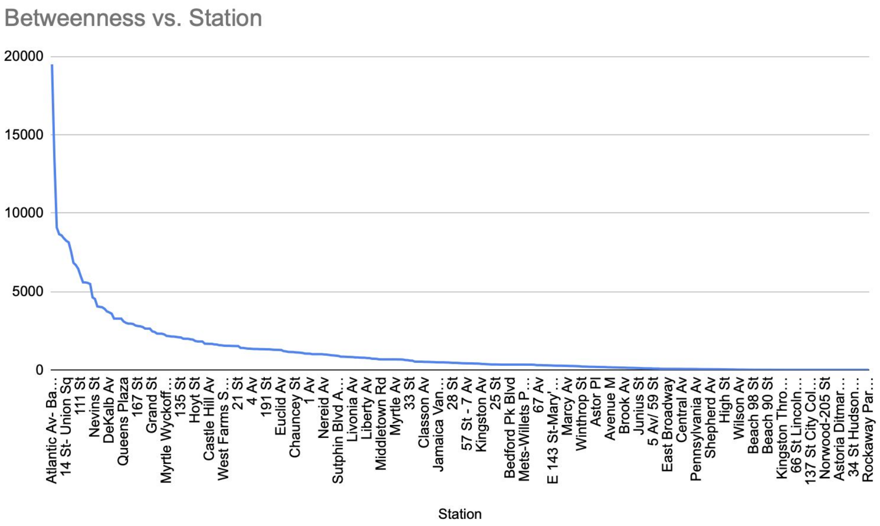
Milestone 1 Recall

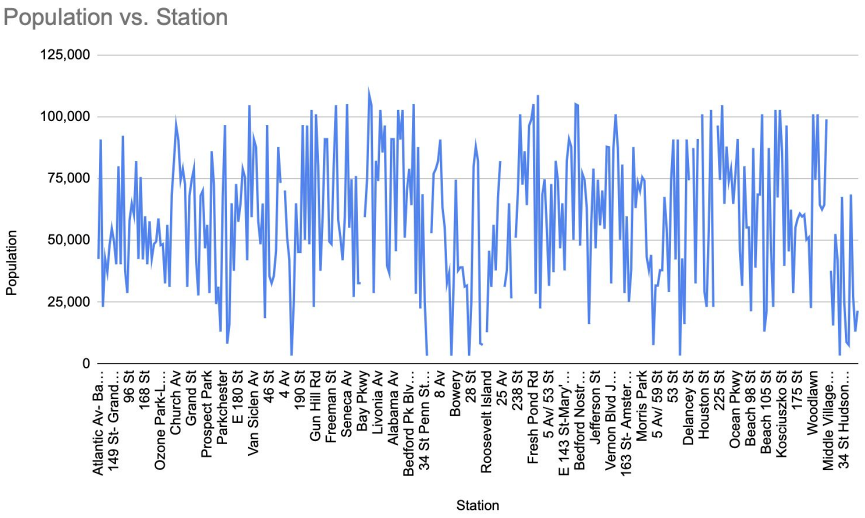
- For Milestone 1, we focused on the **New York City** Subway Network.
- Aimed to find the relationship between betweenness centrality and population distribution (whether betweenness centrality affects the site selection of subway stations)





Milestone 1 Recall (Continued)





Main tasks

- Analyze the **robustness** of a city subway network under different network attack protocols for both Milestone 2 and Milestone 3.
- See how physical accidents or abrupt increasing ridership affect overall ability.

Prior Work

- W. Ellens & R. E. Kooij [1] and Feng, Li, Bao et al. [2] suggest Robustness is a measurement to decide whether a network can continue performing well when facing failures or attacks.
- Z. Jianhua et al. [3] gives the example methods to analyze the subway network robustness (in Shanghai) by efficiency, betweenness centrality, network size, largest connected cluster, and functionality loss using three attacking protocols - largest degree, highest betweenness, and random attacks.
- L.Tian, A.Bashan, D. Shi & Y. Liu [4] introduces articulation point removal (AP) approach to analyze transportation network robustness.
- F. Morone & H. A. Makse [5] states a method (Collective Influence) to measure the importance of a node by aggregating degrees over all neighbors within a ball of size, and F. Morone, B. Min, L. Bo, R. Mari & H. A. Makse [6] adapts the CI algorithm into simplified version and in linear time.

Approaches - Data Collection & Network Generation

- Created network graphs for 4 cities, New York (USA), Shanghai (China), London (England), and Chicago (USA).

New York

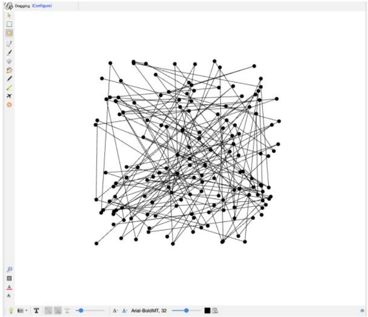
Shanghai

London

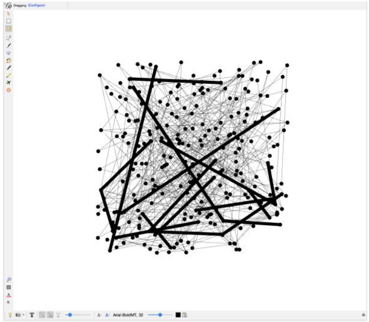
Chicago

Approaches - Data Collection & Network Generation

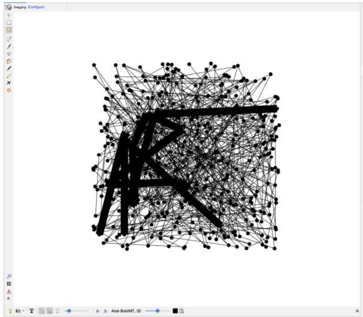
- Data collection is challenging. More cities mean more manual process.
  - New York: collected from Kaggle
  - Shanghai: collected on Shanghai Metro Website
  - London: collected on Transportation For London website
  - Chicago: collected on Chicago Data Portal
- Python script is used to convert raw Json data to network graphs.



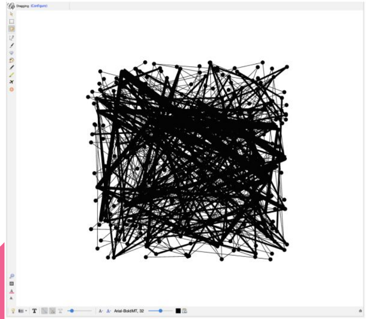
(a) Chicago Network



(b) London Network



(c) Shanghai Network



(d) NYC Network

Approaches - Robustness Analysis

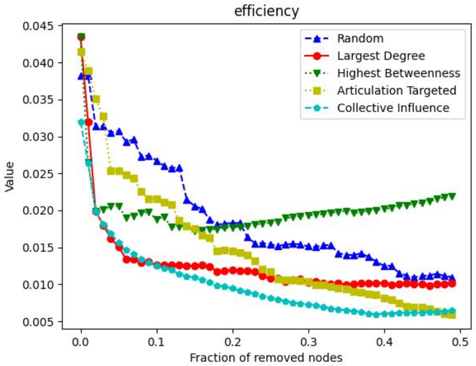
- Utilized 5 network attack protocols and 6 network properties to analyze the **robustness** of 4 cities.
    - Highest Betweenness centrality attack
    - Highest degree attack
    - Random attack
    - Articulation Point attack
    - Collective Influence attack
- Observed network efficiency, average betweenness centrality of nodes, average betweenness centrality of edges, number of edges, largest connected cluster, functionality loss of the network.

Results - Efficiency Analysis

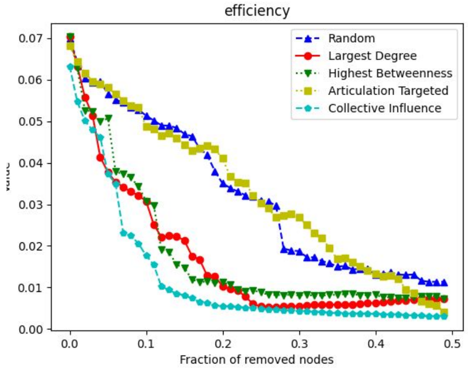
The network efficiency is given as follows:

$$E = \frac{2}{N(N-1)} \sum \frac{1}{d_{ij}}$$

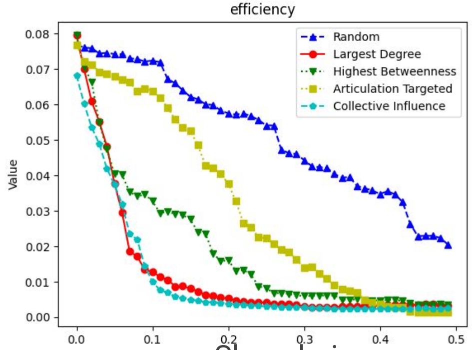
- Collective Influence attack have the largest influence
- When p approaches 0.5, all attacking approaches will have similar results.
- Articulation point targeted will damage most in the end.



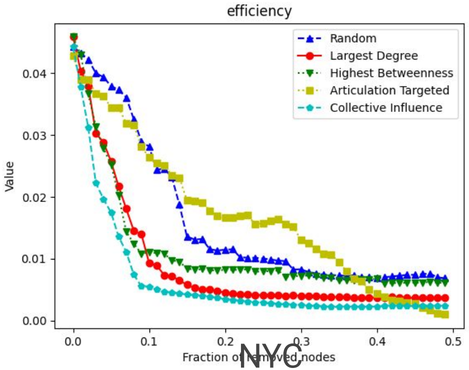
Chicago



London



Shanghai



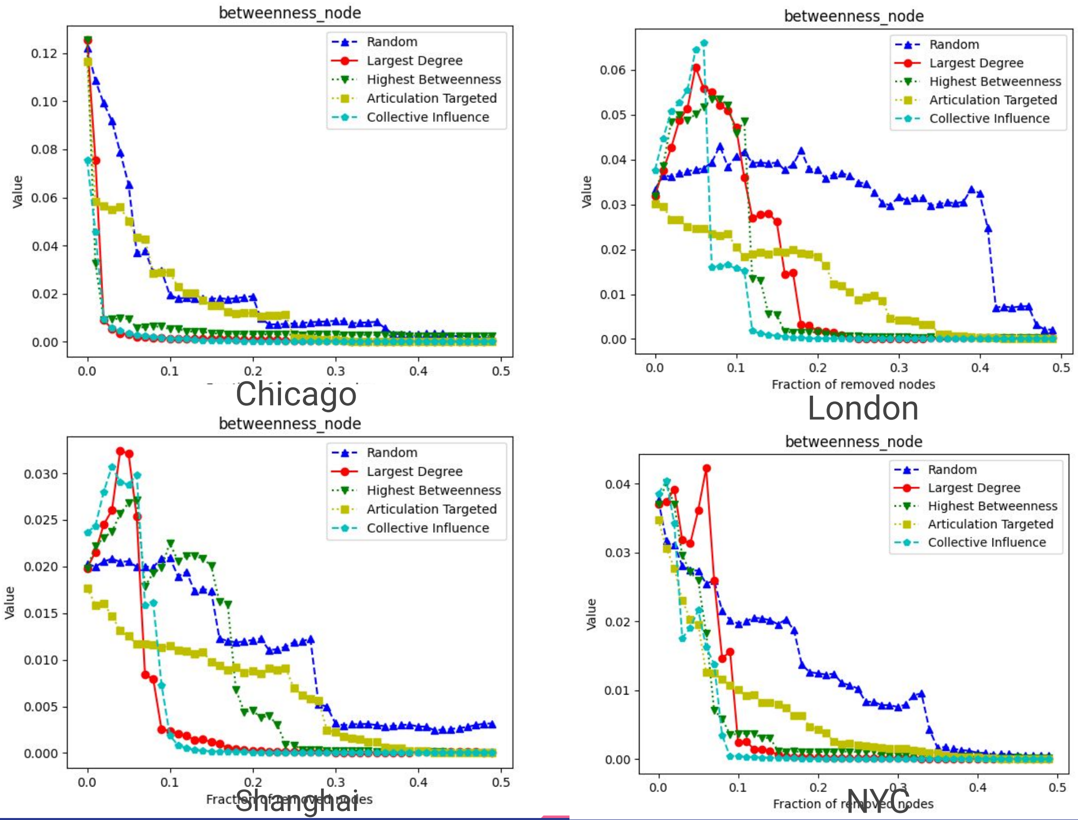
NYC

Results - Average Betweenness Centrality of Nodes

Average Betweenness Centrality

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- Sudden drop exists
- Chicago Subway Network different from other subway networks

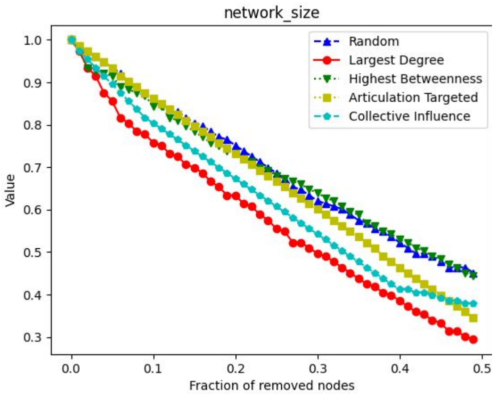


Results - Network Size

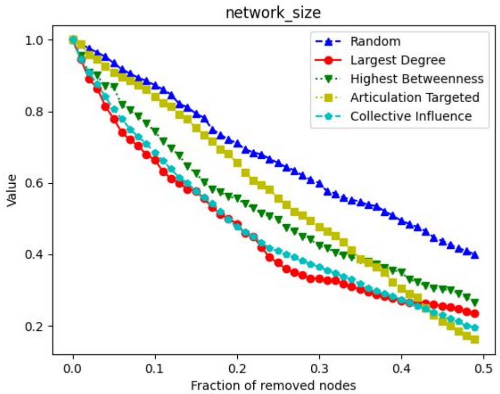
Network size

$$A_{normalized} = \frac{N}{N_0}$$

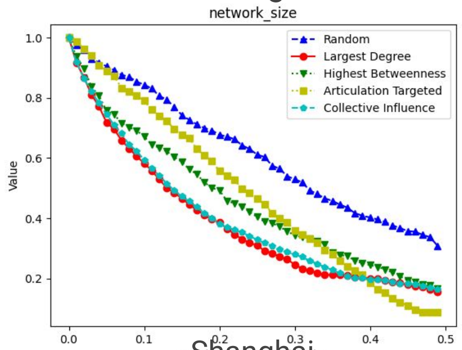
- Drop linearly
- More robust when random attacks
- More fragile when largest degree or collective influence attacks



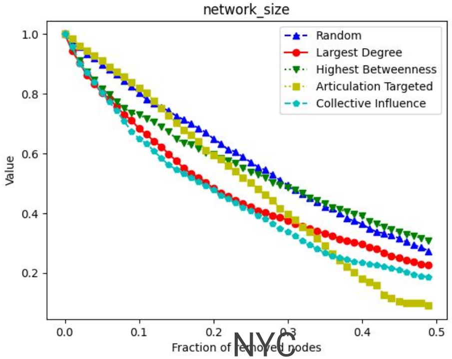
Chicago



London



Shanghai



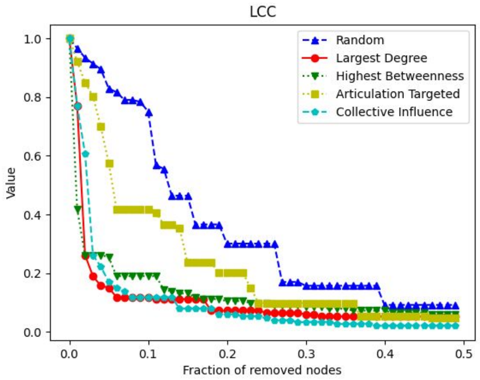
NYC

Results - Largest Connected Cluster

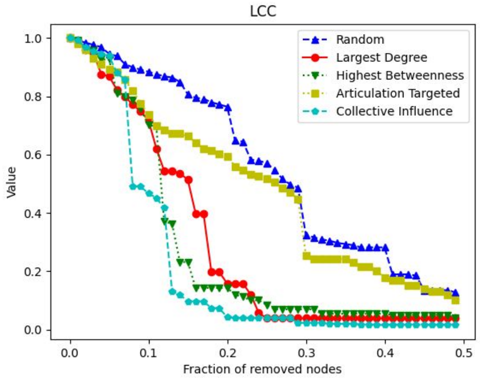
Largest Connected Cluster

$$LCC = \frac{S}{S_0}$$

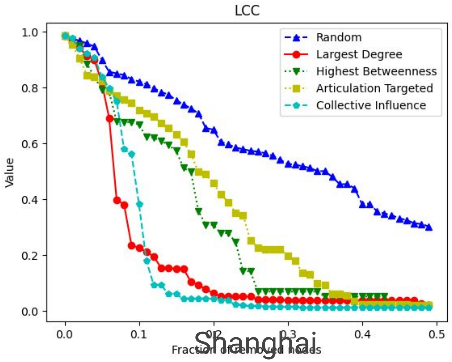
- Random attacks and Articulation Targeted have least damage
- Other attacking protocols have similar results.



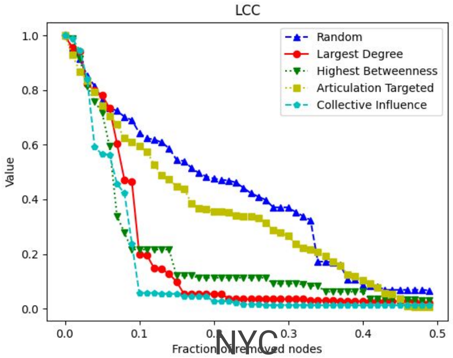
Chicago



London



Shanghai



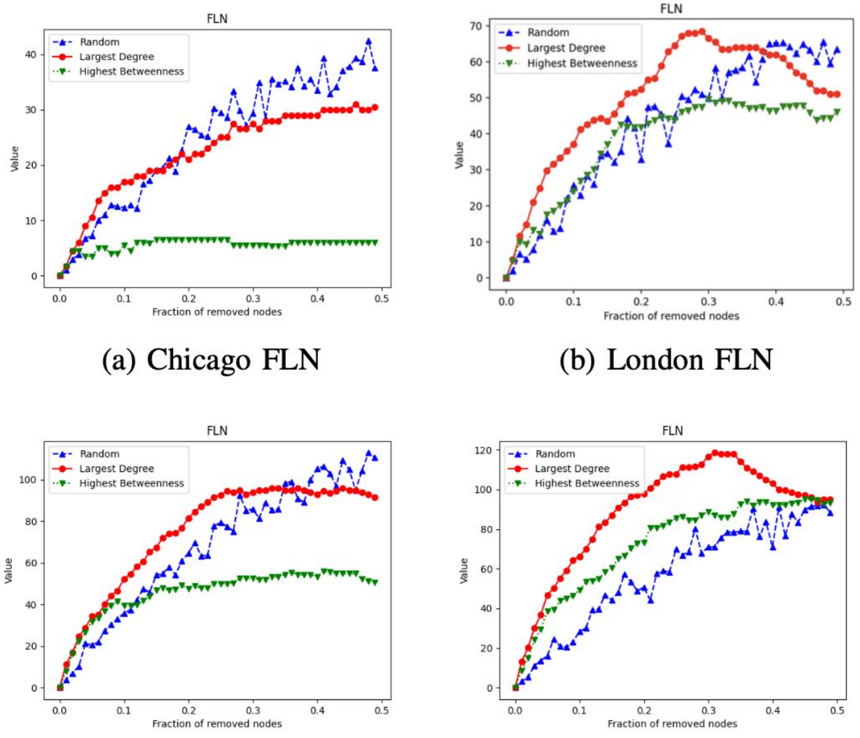
NYC

Results - Functionality Loss of the Network

$$F_l(v_j) = F_{l-1}(v_j) - \frac{1}{d_{ij}^2 k_j} F_{l-1}(v_j)$$

$$FLN = \sum_{j=1, j \neq i}^{N} FL(v_j)$$

- Different city has distinct results based on the city subway shapes and properties.



(a) Chicago FLN (b) London FLN

(c) Shanghai FLN (d) NYC FLN

Conclusions

- The 4 cities are relatively robust to random attack protocols but are fragile to other attacks.
- Different cities are variedly affected under the same attack protocol due to shape of the network graphs.
- **Collective Influence** attack and **Highest Degree** attack have the largest overall effects; **Articulations Targeted** attack have more effects when there are less nodes. The most destructive strategy is to use CI/HD + AT

Limitation & Possible Solutions

- 5 network attack protocols may be not enough. In real world, the 5 may not represent realistic situations.
  - Solution: Research more attack protocols in the future, such as min-sum attack, coreHD, and more.
- Scalability limits our application to a small scope. Always needs manual process to collect subway datasets.
  - Solution: May apply web crawlers to collect data more efficiently.

Contributions

- All:
  - data collecting, graph building, network analysis, and presentation preparation, write report
- Wheaton Wang:
  - collected London's subway data set and built the network
  - implemented methods to simulated Articulation Point attack and Collective Influence attack
- Kyle Yu:
  - collected Shanghai's subway data set and built the network
  - implemented the methods to generate network efficiency graph and size graphs, LCC graphs
- Brandon Bai:
  - collected Chicago's subway data set and built the network
  - implemented the methods to generate normalized average betweenness and FLN graphs

Lessons Learned

- Data collection is the most difficult. Always need manual process and patience.
- Horizontal comparison and visualization of different data is very important.

References

[1] W. Ellens and R. E. Kooij, “Graph measures and network robustness,” arXiv.org, 07-Nov-2013. [Online]. Available: https://arxiv.org/abs/1311.5064v1. [Accessed: 16-Nov-2022].

[2] Feng, J., Li, X., Mao, B., Xu, Q., & Bai, Y. (2017). Weighted complex network analysis of the Beijing subway system: Train and passenger flows. Physica A: Statistical Mechanics and its Applications, 474, 213–223. doi:10.1016/j.physa.2017.01.085

[3] Zhang, Jianhua et al. “Networked Analysis of the Shanghai Subway Network, in China.” Physica A 390.23-24 (2011): 4562–4570. Web.

[4] Tian, L., Bashan, A., Shi, D.-N., & Liu, Y.-Y. (2017). Articulation points in complex networks. Nature Communications, 8(1), 14223. doi:10.1038/ncomms14223

[5] Morone, F., & Makse, H. A. (2015). Influence maximization in complex networks through optimal percolation. Nature, 524(7563), 65–68. doi:10.1038/nature14604

[6] Morone, F., Min, B., Bo, L., Mari, R., & Makse, H. A. (2016). Collective Influence Algorithm to find influencers via optimal percolation in massively large social media. *Scientific Reports*, *6*(1), 30062. doi:10.1038/srep30062

[7] https://new.mta.info/maps/subway-line-maps

[8] https://www.newyork-demographics.com/zip_codes_by_population

Thank you!