Complex Network Analysis on Bike-Sharing Systems

Bike-sharing systems (BSS) are emerging transportation systems arising from the sharing economy. Due to the data accessibility and important statistical features, those systems have gained increasing scientific interest. To better understand user behaviors in BSS and their relations to system performance, researchers have developed various data-driven approaches. The objective of this problem is to analyze BSS network topology and predict user behaviors in support of system design and operation.

Problem Statement:

Bike-sharing system: Divvy bike system in Chicago

Task 1. BSS network analysis

1) First, based on the data file, *July_trip_raw_data_2016.csv*, and the following definition, establish the trip network: the trip network is directed, where the stations are defined as nodes, and the links represent the transit from one station to another station. The weight on a link is the total number of trips that happened from one station to another station in a month. For example, in June, 14 users are riding from station A to station B and ten users riding from station B to station A, then the weight of link (A → B) is 14, and the weight of link (B → A) is 10.
2) Second, transform the weighted network to a binary network with a link weight threshold of X. That means you need to remove the links that have a weight smaller than X, and only keep the links with a weight greater than X. You need to determine the value of X by yourself and justify your decision.
3) Third, visualize the directed binary trip network and explore its network statistics, such as degree distribution, network density, average cluster coefficient, etc.
4) You can use any software tool to visualize and analyze the network, e.g., Gephi, R, or NetworkX package.

Task 2. BSS network prediction

1) Follow 1) – 2) in Task 1 to process the data file, *July_trip_raw_data_2017.csv*, and create the real directed binary trip network of July 2017.
2) Develop a predictive model on your own discretion to predict the directed binary BSS network of July 2017 using the station information and the directed binary BSS network of July 2016 created in Task 1.
3) Evaluate the performance of the proposed model by comparing the predicted networks against the real BSS network of July 2017.

Data Sources:

1. 2016 July trip raw data of Divvy bike system, filename: *July_trip_raw_data_2016.csv*
2. 2017 July trip raw data of Divvy bike system, filename: *July_trip_raw_data_2017.csv*
   (In these two files, there are only two columns, one is the start station ID, and another one is the end station ID. Note that these are the trip records of the entire month. Do not delete the repeated items; they represent one trip that repeatedly happened at different time points, and you need them to calculate the link weight)
3. 2016 station information of Divvy bike system, filename: *Divvy_Station_2016.csv*
4. 2017 station information of Divvy bike system, filename: *Divvy_Station_2017.csv*

Format of Deliverables:

1. Please write a report of no more than three pages to summarize your key findings.
2. Please prepare a PPT presentation to introduce your approaches, findings, and conclusions. You will present the slides during your second-round interview.