Overview of the data structure:

* ride\_id: Unique identifier for each ride
* rideable\_type: The type of bike used for the ride (classic bike, electric bike, etc.)
* started\_at: Start date and time of the ride
* ended\_at: End date and time of the ride
* start\_station\_name: Name of the starting station
* start\_station\_id: Unique identifier for the starting station
* end\_station\_name: Name of the ending station
* end\_station\_id: Unique identifier for the ending station
* start\_lat, start\_lng: Latitude and longitude of the start station
* end\_lat, end\_lng: Latitude and longitude of the end station
* member\_casual: Type of user (member or casual)

**A. Community Structure and Network Resilience**

**Objective**: This analysis aims to understand how the intrinsic community structures within the bike-sharing network—clusters of stations that exhibit more frequent interactions among themselves than with other parts of the network—affect its overall resilience.

**Approach**:

1. **Network Construction**: Create a graph where nodes represent the bike stations and edges represent trips between these stations. The weight of an edge could reflect the volume of trips, indicating the strength of the connection between stations.
2. **Community Detection**: Use community detection algorithms (such as modularity optimization, Girvan-Newman algorithm, or Louvain method) to identify clusters or communities within the network. These communities are characterized by higher intra-community connectivity compared to inter-community connections.
3. **Resilience Analysis**: Evaluate the network's resilience by simulating the removal of nodes (stations) or edges (trips) from the network. This can be done in two ways:
   * **Random Failures**: Randomly remove nodes or edges and assess the impact on network connectivity and the size of the largest connected component.
   * **Targeted Attacks**: Remove nodes or edges based on certain criteria (e.g., highest degree, betweenness centrality) and observe the effects on network resilience.
4. **Community Role**: Analyze how the removal of stations within a single community versus across different communities affects the overall network structure. Investigate whether certain communities serve as critical bridges or bottlenecks for network connectivity.

**Expected Insights**:

* Identify key communities within the bike-sharing network that are crucial for maintaining system-wide connectivity.
* Understand the role of community structures in absorbing disruptions and maintaining service continuity.
* Develop strategies for strengthening weak links between communities to enhance network resilience.

# Task 2 Station Utilization and Suitability Analysis

Aim:

* Evaluate how well the bike-sharing stations are distributed across the network
* Evaluate station usage rate
* Identify potential gaps

**Metrics to Construct**

1. **~~Usage Rate Metrics~~**
   * **~~Total Number of Trips~~**~~: The total number of trips starting and ending at each station.~~
   * **~~Average Trip Duration~~**~~: The average duration of trips originating from and concluding at each station.~~
   * **~~Peak Usage Times~~**~~: The most common times of day and days of the week each station is used.~~
2. **~~Connectivity Metrics~~**
   * **~~Degree Centrality~~**~~: The number of direct connections each station has to others, indicating its importance in the network.~~
   * **~~Betweenness Centrality~~**~~: A measure of how often a station acts as a bridge along the shortest path between two other stations, highlighting stations that facilitate long-distance trips within the network.~~
3. **Accessibility Metrics**
   * **Average Distance to Nearest Neighbors**: The average distance (physical or in terms of trip duration) from each station to its nearest neighbor stations, which can help identify potentially isolated or poorly served areas.
   * **Cluster Coefficient**: A measure of how well-connected a station's neighbors are to each other, indicating local cluster or community structure.
4. **Utilization Efficiency Metrics**
   * **Station Balance**: The ratio of incoming to outgoing trips, identifying stations that consistently experience bike shortages or surpluses.

**How to Measure Using the Station Network**

1. **Analyze Spatial Distribution**: Use GIS (Geographic Information Systems) tools or mapping libraries in Python (e.g., **geopandas**, **matplotlib**, **folium**) to visualize the spatial distribution of these metrics across the network. This can help identify geographical patterns, such as high-demand areas or regions lacking sufficient station coverage.
2. **Identify Under and Over-utilized Stations**: By comparing usage rate metrics and connectivity metrics, identify stations that are under-utilized (low trip numbers, low centrality) or over-utilized (high trip numbers, potential for congestion).
3. **Recommendations for Optimization**: Based on your findings, recommend actions such as adding new stations, redistributing bikes, or adjusting the placement of existing stations to improve the overall efficiency and accessibility of the network.

4. Simulation of Adding Bikes

Simulate adding bikes to the network by adjusting the available bike count at the identified stations. This can be approached in several ways, depending on the complexity of the simulation you're willing to undertake:

Simple Adjustment: Increase the bike count at shortage stations by a fixed number or percentage and decrease correspondingly at surplus stations, then recalculate the station balance metric.

Dynamic Model: Use a more sophisticated model that takes into account the time of day, day of the week, and possibly external factors like weather. This model would dynamically adjust the bike availability at different times, simulating how actual usage might change with more (or fewer) bikes.

5. Impact Assessment

Assess the impact of these adjustments by evaluating:

Change in Station Balance: How does the station balance metric change with the adjusted bike counts? A move towards a ratio closer to 1:1 would indicate an improvement.

Usage Changes: Evaluate how the overall trip volumes change for the affected stations. An increase in trips starting at previously bike-short stations would indicate a positive impact.

Network Effects: Consider the broader network effects, such as changes in trip patterns or impacts on adjacent stations.

6. Analysis of Results

Analyze the simulation results to determine if adding bikes as simulated effectively addresses the identified issues. Consider:

Effectiveness: Did the additional bikes reduce instances of shortages or surpluses? By how much?

Operational Feasibility: Is the required increase or decrease in bikes at specific stations operationally feasible?

Unintended Consequences: Are there any negative impacts on other parts of the network, such as increased congestion at nearby stations?