CNotes - not formated

> 💡 definition of centrality measure and define each centrality measure along with how we incorporated/used each centrality measure for TfL network and what does that mean for the whole network

Centrality Measures: An Overview

Centrality measures are metrics used in network analysis to identify the most important or influential nodes within a graph. In the context of the TfL network, these measures help us understand which stations play crucial roles in the overall structure and flow of the transportation system.

1. Degree Centrality

Definition: Degree centrality measures the number of direct connections a node has to other nodes in the network.

Mathematical Formula:
C_D(v) = deg(v) / (n - 1)

Where:

- C_D(v) is the degree centrality of node
- deg(v) is the degree of node v
- n is the total number of nodes in the network

Code Implementation:

```
degree_centrality = nx.degree_centrality(G)
```

Implication for TfL: Stations with high degree centrality are those with many direct connections to other stations. These are likely to be major interchange stations or hubs where multiple lines intersect. They are crucial for facilitating transfers and managing passenger flow across the network.

1. Betweenness Centrality

Definition: Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes in the network.

Mathematical Formula:
$C_B(v) = \sum(s \neq v \neq t) (\sigma_{st}(v) / \sigma_{st})$

Where:

- $C_B(v)$ is the betweenness centrality of node v
- σ_{st} is the total number of shortest paths from node s to node t
- $\sigma_{st}(v)$ is the number of those paths that pass through v

Code Implementation:

```
betweenness_centrality = nx.betweenness_centrality(G)
```

Implication for TfL: Stations with high betweenness centrality act as critical junctions in the network. They are important for maintaining efficient travel routes and could be potential bottlenecks if disrupted. These stations might not necessarily have many direct connections but are crucial for connecting different parts of the network.

1. Closeness Centrality

Definition: Closeness centrality measures how close a node is to all other nodes in the network.

Mathematical Formula:
$C_C(v) = (n - 1) / \sum(u \neq v) d(v,u)$

Where:

- $C_C(v)$ is the closeness centrality of node v
- n is the total number of nodes in the network
- d(v,u) is the shortest path distance between nodes v and u

Code Implementation:

```
closeness_centrality = nx.closeness_centrality(G)
```

Implication for TfL: Stations with high closeness centrality are, on average, the most accessible from all other stations in the network. These stations are strategically positioned for quick access

to many parts of the network, making them ideal for emergency response or network management.

1. Eigenvector Centrality

Definition: Eigenvector centrality measures the influence of a node in the network based on the centrality of its neighbors.

Mathematical Formula:
$A x = \lambda x$

Where:

- A is the adjacency matrix of the graph
- λ is the largest eigenvalue
- x is the corresponding eigenvector

Code Implementation:

```
eigenvector_centrality = nx.eigenvector_centrality(G)
```

Implication for TfL: Stations with high eigenvector centrality are connected to other important stations. This measure helps identify not just stations with many connections, but those connected to other well-connected stations. These are likely to be key stations in the busiest or most central areas of the network.

1. PageRank

Definition: PageRank is a variant of eigenvector centrality that considers the direction and weight of links.

Mathematical Formula:
$PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$

Where:

- PR(A) is the PageRank of node A
- d is a damping factor (typically 0.85)
- PR(Ti) is the PageRank of nodes linking to A
- C(Ti) is the number of outbound links from node Ti

Code Implementation:

```
pagerank = nx.pagerank(G)
```

Implication for TfL: While the TfL network is undirected, PageRank can still provide valuable insights. Stations with high PageRank are likely to be important transfer points or destinations, considering both their connections and the importance of the stations they're connected to.

In our analysis of the TfL network, we've applied these centrality measures to identify key stations from different perspectives. By visualizing the top 5 stations for each measure, we can understand the network's structure from multiple angles:

- Identify major hubs and interchange points
- Locate critical junctions for network flow
- Find strategically positioned stations for quick access
- Recognize stations in the most central or busiest areas
- Determine overall important stations considering both connections and neighboring importance

This multi-faceted approach provides a comprehensive understanding of the TfL network's structure and the relative importance of different stations, which can be crucial for network planning, optimization, and management.

Top 5 nodes by Degree Centrality:

[('bakerstreet', 0.019021739130434784), ('kingscrossstpancras', 0.019021739130434784), ('stratford', 0.019021739130434784), ('waterloo', 0.016304347826086956), ('oxfordcircus', 0.016304347826086956)]

Top 5 nodes by Betweenness Centrality:

[('bank', 0.2398771409848292), ('waterloo', 0.2254827592539357), ('kingscrossstpancras', 0.2233532023252586), ('greenpark', 0.21981973398200746), ('bakerstreet', 0.21915043621261734)]

Top 5 nodes by Closeness Centrality:

[('greenpark', 0.11295273173726213), ('westminster', 0.11004784688995216), ('bondstreet', 0.10994920824619062), ('kingscrossstpancras', 0.10988354732756046), ('oxfordcircus', 0.10952380952380952)]

Top 5 nodes by Eigenvector Centrality:

[('oxfordcircus', 0.3968157978956346), ('greenpark', 0.39036243116617064), ('piccadillycircus', 0.30975629560797274), ('bondstreet', 0.2835513647087155), ('westminster', 0.22314485167327802)]

Top 5 nodes by PageRank:

[('stratford', 0.0068488344763078775), ('kingscrossstpancras', 0.0063662557647093), ('bakerstreet', 0.006109027217569682), ('paddington', 0.005900841200001562), ('canningtown', 0.005836020535399946)

Insights on TfL Network Centrality Measures

1. Degree Centrality
   The top stations by degree centrality (Baker Street, King's Cross St. Pancras, Stratford, Waterloo, and Oxford Circus) represent major interchange hubs in the network. These stations serve multiple lines, facilitating transfers and connecting different parts of London. Their high degree centrality indicates their crucial role in network connectivity and passenger distribution.
2. Betweenness Centrality
   Bank, Waterloo, King's Cross St. Pancras, Green Park, and Baker Street emerge as key

stations for betweenness centrality. These stations act as critical junctions, lying on many shortest paths between other stations. Bank's top position suggests its vital role in connecting different parts of the network, particularly in the City of London. The presence of both central London stations (like Green Park) and major interchanges highlights the importance of both geographical location and network structure in determining a station's betweenness centrality.

3. Closeness Centrality
   Green Park, Westminster, Bond Street, King's Cross St. Pancras, and Oxford Circus show high closeness centrality. These stations are strategically located for quick access to many parts of the network. Their central positions in London's geography and the network topology make them ideal for rapid transit across the city. The prominence of West End stations (Green Park, Westminster, Bond Street, Oxford Circus) underscores the centrality of this area in London's public transport system.

4. Eigenvector Centrality
   Oxford Circus, Green Park, Piccadilly Circus, Bond Street, and Westminster score highly on eigenvector centrality. This measure highlights the West End's significance in the network, as these stations are not only well-connected themselves but are also linked to other important stations. The concentration of high eigenvector centrality in this area reflects its role as a key destination and transit hub in London.

5. PageRank
   Stratford, King's Cross St. Pancras, Baker Street, Paddington, and Canning Town top the PageRank measure. This diverse set of stations includes both central London hubs and important outlying stations. Stratford's top position likely reflects its role as a major interchange in East London, connecting multiple lines and services. The inclusion of Canning Town highlights the importance of stations that might not be central but play crucial roles in their local network segments.

Summary and Network Implications

1. Multi-faceted Importance: Stations like King's Cross St. Pancras and Baker Street appear in multiple top-5 lists, underscoring their multifaceted importance to the network. They serve as major interchanges, critical junctions, and well-connected hubs.

2. Geographical Distribution: The centrality measures reveal a balance between central London stations (e.g., Oxford Circus, Green Park) and strategically located outer stations (e.g.,

Stratford, Canning Town). This highlights the network's structure in serving both the dense city center and the broader metropolitan area.

3. West End Significance: The high rankings of West End stations across multiple measures (particularly in closeness and eigenvector centrality) emphasize this area's critical role in London's transport network, likely reflecting its importance as a commercial, cultural, and tourist center.
4. Network Resilience: Stations with high betweenness centrality (like Bank and Waterloo) are crucial for network flow. Their prominence suggests that disruptions at these stations could have significant impacts on the entire network's efficiency.
5. Strategic Importance of East London: The high rankings of Stratford and Canning Town, particularly in PageRank, indicate the growing importance of East London in the network, possibly reflecting urban development and expansion in this area.
6. Interchange Hubs: The recurrence of stations like King's Cross St. Pancras across different measures underscores the vital role of major interchange hubs in knitting together different parts of the network and facilitating efficient travel across London.

In conclusion, these centrality measures provide a nuanced view of the TfL network's structure. They highlight not only the importance of traditional central London hubs but also the strategic value of well-connected stations in outer areas. This analysis can inform decisions on capacity management, service frequency, and infrastructure development, ensuring that the network continues to effectively serve London's diverse transportation needs.