[redacted]

the Girvan-Newman Method

To develop an automated algorithm for determining the right set of communities using the Girvan-Newman method, it is essential to define a clear stopping criterion that ensures the identified communities are both meaningful and well-structured. The Girvan-Newman algorithm is a hierarchical, divisive method that progressively removes edges with the highest betweenness centrality to reveal community structures within a graph. The challenge lies in determining the optimal point at which to stop the algorithm to yield the most accurate community detection.

Steps in the Algorithm

1. Initialization

The process begins by considering the entire graph G as a single community. All nodes are connected, and the aim is to iteratively identify communities by selectively removing edges.

2. Edge Betweenness Centrality Calculation

For each edge in the graph, calculate its betweenness centrality. Betweenness centrality measures the extent to which an edge lies on the shortest paths between pairs of nodes. Edges with high betweenness centrality are typically the connectors between different communities. Mathematically, the betweenness centrality of an edge e is given by:

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

where σ_{st} is the total number of shortest paths between nodes s and t, and $\sigma_{st}(e)$ is the number of those paths that pass through edge e.

3. Edge Removal

Identify and remove the edge with the highest betweenness centrality. This step is crucial because it helps in isolating communities by cutting the most significant links between them.

4. Community Detection

After each edge removal, the graph may split into multiple connected components, each of which is considered a potential community. The structure of these communities is refined with each iteration.

5. Modularity Calculation

Modularity Q is a key metric used to assess the quality of the community structure at each stage. It compares the density of edges within communities to the density of edges between communities. The formula for modularity is:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

where:

- A_{ij} is the adjacency matrix element representing the presence (1) or absence (0) of an edge between nodes i and j.
- k_i and k_j are the degrees of nodes i and j.
- m is the total number of edges in the graph.
- $\delta(c_i, c_j)$ is the Kronecker delta function, which equals 1 if nodes i and j belong to the same community, and 0 otherwise.

6. Stopping Criterion

The effectiveness of the Girvan-Newman algorithm hinges on defining an appropriate stopping criterion. The most widely accepted stopping criterion is based on maximizing the modularity Q. The algorithm should continue removing edges and recalculating modularity after each step until the modularity score reaches its peak. The peak modularity indicates that the division of the graph into communities is optimal—where intra-community edge density is maximized, and inter-community edge density is minimized. The stopping criterion can be mathematically expressed as:

$$Q_{\text{current}} < Q_{\text{previous}}$$

where Q_{current} is the modularity after the latest edge removal, and Q_{previous} is the modularity from the previous step. Once this condition is met, the algorithm halts.

7. Final Community Structure

The community structure at the point of maximum modularity is considered the optimal division of the network. This final step ensures that the communities identified are both cohesive and distinct, providing a clear and meaningful segmentation of the graph.