**Different approaches to evaluating communities in social networks.**

**What is Community Detection?**

* Community detection is the task of identifying groups of nodes in a network where **nodes within the same group are highly connected**, and **nodes between groups are sparsely connected**.
* Example:  
  In a social media platform like Facebook, a *community* might be a friend group, college batch, or a workplace team.
* After detecting such groups, we need to **evaluate** whether the grouping makes sense. That’s where the **evaluation approaches** come in.

**I. Internal Evaluation Methods (Structure-Based Evaluation)**

These methods rely solely on the **topology or structure** of the graph. They do **not** use any external labels or ground-truth data.

**1. Modularity**

* **Definition**: Modularity measures how well the network is divided into communities by comparing the **density of links inside communities** vs. **links between communities**.
* **Formula**:  
    
  (You don’t need to memorize this — just understand the meaning.)
* **Key Idea**: A higher modularity value (close to 1) indicates strong community structure.
* **Example**: In a university, students from the same department interact more with each other than with students from other departments. This shows high modularity.
* **Used In**: Louvain algorithm, FastGreedy algorithm.
* **Interpretation**:
  + High modularity (0.4 to 0.7) = good separation
  + Low modularity (~0) = random groupings

**2. Conductance**

* **Definition**: Conductance measures how easily a community can be "cut off" from the rest of the network. It looks at **the number of external connections** vs **internal connections**.
* **Formula**:  
  Conductance = (Edges leaving the community) / (Total degree of nodes in the community)
* **Goal**: Lower conductance means better separation.
* **Example**: If you create a study group and most discussions stay inside the group with very few members talking to outsiders, conductance is low — which is good.
* **Use Case**: Great for identifying well-separated communities in large graphs.

**3. Density**

* **Definition**: Measures how many connections exist within a community **compared to the total number of possible connections**.
* **Formula**:  
  Density = Actual edges / Maximum possible edges in the group
* **Example**: In a project team of 5 people where everyone works closely with everyone, the community has high density.
* **Use Case**: Helps measure tightness of small communities.

**4. Clustering Coefficient**

* **Definition**: Measures how much the neighbors of a node are connected to each other.
* **Local Version**: Looks at individual nodes.
* **Global Version**: Average over all nodes in a community.
* **Example**: If Alice is friends with Bob and Charlie, and Bob and Charlie are also friends, Alice has a high clustering coefficient.
* **Interpretation**: High coefficient → likely presence of tightly-knit subgroups.

**5. Average Path Length within Communities**

* **Definition**: Measures the average number of steps required to connect any two nodes within the same community.
* **Lower values** indicate better internal connectivity.
* **Example**: In a good office team, you can reach any colleague within 1-2 email forwards or messages.

**II. External Evaluation Methods (Ground-Truth-Based Evaluation)**

These methods evaluate detected communities by comparing them with **known real groupings or labels**.

**1. Normalized Mutual Information (NMI)**

* **Definition**: NMI measures the amount of information shared between the detected and actual communities.
* **Range**: 0 (no match) to 1 (perfect match).
* **Example**: Suppose you know students' departments and detect communities using an algorithm. If NMI = 0.9, then the detected groups closely match actual departments.
* **Use Case**: When labeled data is available.

**2. Adjusted Rand Index (ARI)**

* **Definition**: ARI calculates the similarity between the clustering results and the actual grouping by **adjusting for random chance**.
* **Advantage**: Penalizes random groupings.
* **Example**: If your detection puts most marketing team members together, ARI will reflect that accuracy, and also penalize the mistakes.

**3. Purity**

* **Definition**: Measures how "pure" each detected group is based on known labels. Each group is assigned to the most common actual class in it.
* **Easy to compute**, but may give high scores for small fragmented groups.
* **Example**: If your detected group mostly contains CS students, purity is high. But if you make one-person groups, purity becomes 1, which is misleading.

**4. Precision, Recall, and F1-Score**

* **Precision**: Fraction of node pairs that are correctly grouped out of all pairs predicted to be together.
* **Recall**: Fraction of true same-group pairs that were correctly detected.
* **F1 Score**: Harmonic mean of precision and recall.
* **Use Case**: Works well with partially labeled datasets.

**III. Functional Evaluation Methods (Application-Based Evaluation)**

These methods check how useful the detected communities are in **practical scenarios**, such as marketing, recommendation, or influence spread.

**1. Recommendation Accuracy**

* **Definition**: Evaluates whether the communities help in giving accurate and relevant recommendations.
* **Example**: In Amazon, customers are grouped based on purchase behavior. Better communities lead to more accurate product recommendations.
* **Use Case**: E-commerce, content platforms, streaming apps.

**2. Information Spread (Simulation Tests)**

* **Definition**: Tests how well the community structure supports or limits the spread of information, messages, or viruses.
* **Example**: In Twitter, a meme might go viral within a strong fan community faster than a weak one.
* **Use Case**: Social media campaigns, public health announcements.

**3. Behavioral Similarity**

* **Definition**: Evaluates if members of the same community behave similarly (e.g., similar likes, posts, or purchases).
* **Example**: In a Spotify group, members who like similar songs form a behaviorally consistent community.
* **Use Case**: Ad targeting, behavioral analytics, personalization engines.

**Final Tips for Writing in Exam**

* **Always explain the purpose first** (Why evaluate?)
* **Include a definition, formula (optional), and example** for each method
* **Use practical examples** to show real-world usefulness
* **Mention use-cases and limitations briefly**