

# CS246 Exam 2025 — Question 7: Community Detection (9 points)

---

## Part 1 (3 points) — Conductance vs. Edge Cut

**Question:** Why is conductance a better metric than edge cut for evaluating cluster quality?

**Answer:** Edge cut only counts edges crossing the boundary, ignoring cluster size. This allows degenerate solutions — e.g., isolating a single node with 1 connection gives cut = 1, which is minimal but meaningless.

Conductance normalizes the cut by the cluster's **volume** (sum of degrees):

$$\phi(A) = \frac{\text{edges crossing boundary}}{\min(\text{vol}(A), 2m - \text{vol}(A))}$$

This penalizes tiny clusters (small volume inflates conductance) and ensures good communities are balanced with both sides internally well-connected. Edge cut measures *absolute* separation; conductance measures *relative* separation.

---

## Part 2 (3 points) — Louvain Algorithm Convergence

**Question:** Is Louvain guaranteed to converge? Is it guaranteed to find maximum modularity?

**Converges? Yes.** Each node move only happens if it **strictly increases** modularity ( $\Delta Q > 0$ ). Since modularity is bounded ( $Q \in [-1, 1]$ ) and strictly increases with each move, the algorithm cannot cycle and must terminate.

**Global maximum? No.** Louvain is **greedy** — it only considers single-node moves and always accepts improvements. It converges to a **local maximum**, not necessarily the global one. Finding the global optimum might require simultaneously moving multiple nodes, which the greedy strategy cannot discover.

---

## Part 3 (3 points) — Can NN-Descent Replace PPR or Louvain?

**Answer: No.** They solve fundamentally different problems at different stages of the same pipeline:

Algorithm	Role	Input → Output
NN-Descent	Graph construction	Data points → K-NN graph
PPR / Louvain	Community detection	Graph → Partition into communities

---

NN-Descent builds an approximate K-nearest neighbor graph from raw data. It has no mechanism for defining communities — it doesn't compute conductance, modularity, or any partition quality metric. The correct pipeline is:

**Raw data → NN-Descent (build graph) → Louvain/PPR (detect communities)**

NN-Descent is a *prerequisite* for community detection when starting from raw data, not a substitute.