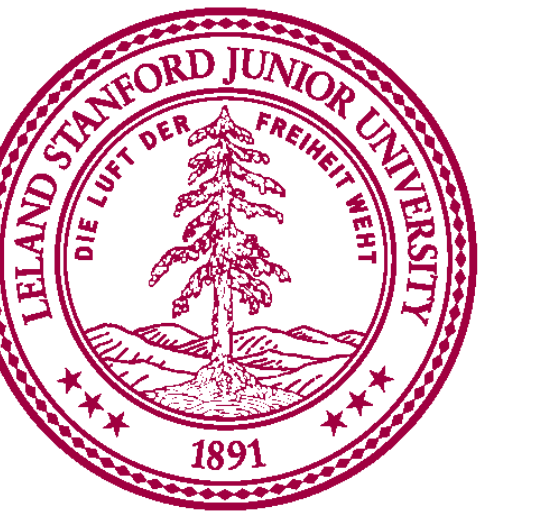


Model-based Approach to Detecting Densely Overlapping Communities in Networks



Jaewon Yang and Jure Leskovec
Stanford University

Problem Definition

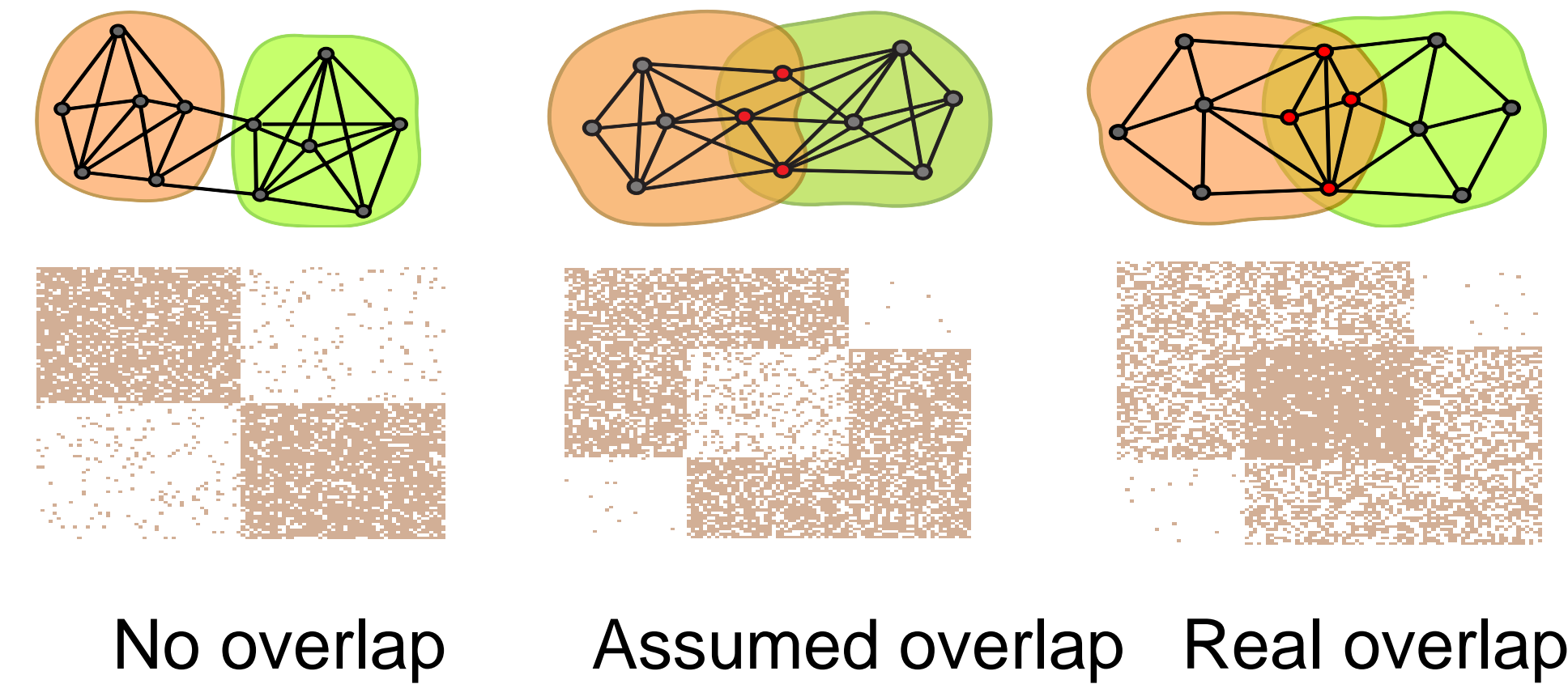
- Given a network $G(V, E)$, find communities C_i .
 - Community:** A cluster of nodes that are densely linked one another.
 - Web pages about the same topic are densely linked.
 - Proteins of the same role interact frequently.
- Goal:** Develop a model that detects overlapping communities.
 - A node can belong to multiple communities.

Previous Approach VS Ours

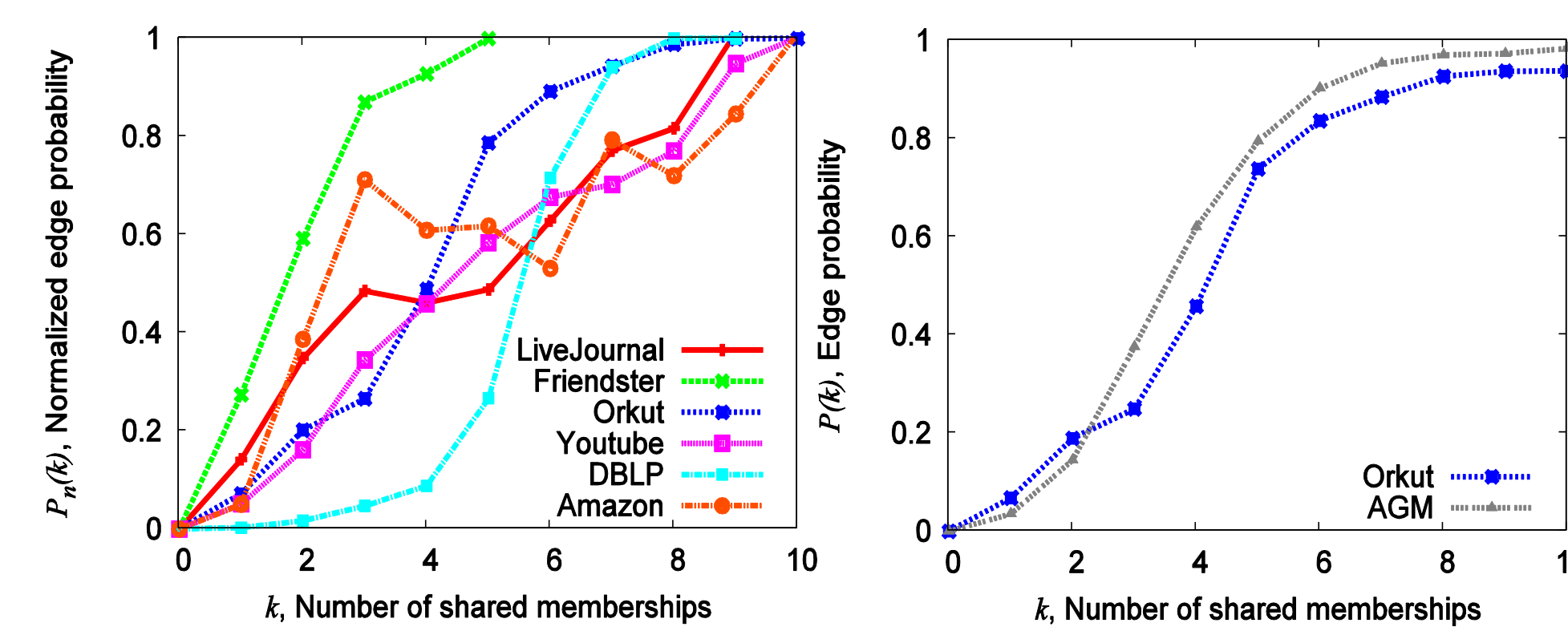
- Previous approach:
 - Detect communities from a given network.
 - Interpret communities based on "Ground-truth".
 - E.g., people working in the same area
- Our approach:**
 - Observe a network with **explicitly labeled ground-truth communities**.
 - Develop a **model** explaining the relationship between the network and the ground-truth communities.
 - Detect communities from a given network without ground-truth **using the model**.
- Examples of explicitly defined ground-truth:
 - Researchers publishing to the same conference,
 - Social network users belonging to the topic-based group.

Community Overlaps

- Current assumption: nodes in the overlap are **sparse** linked.



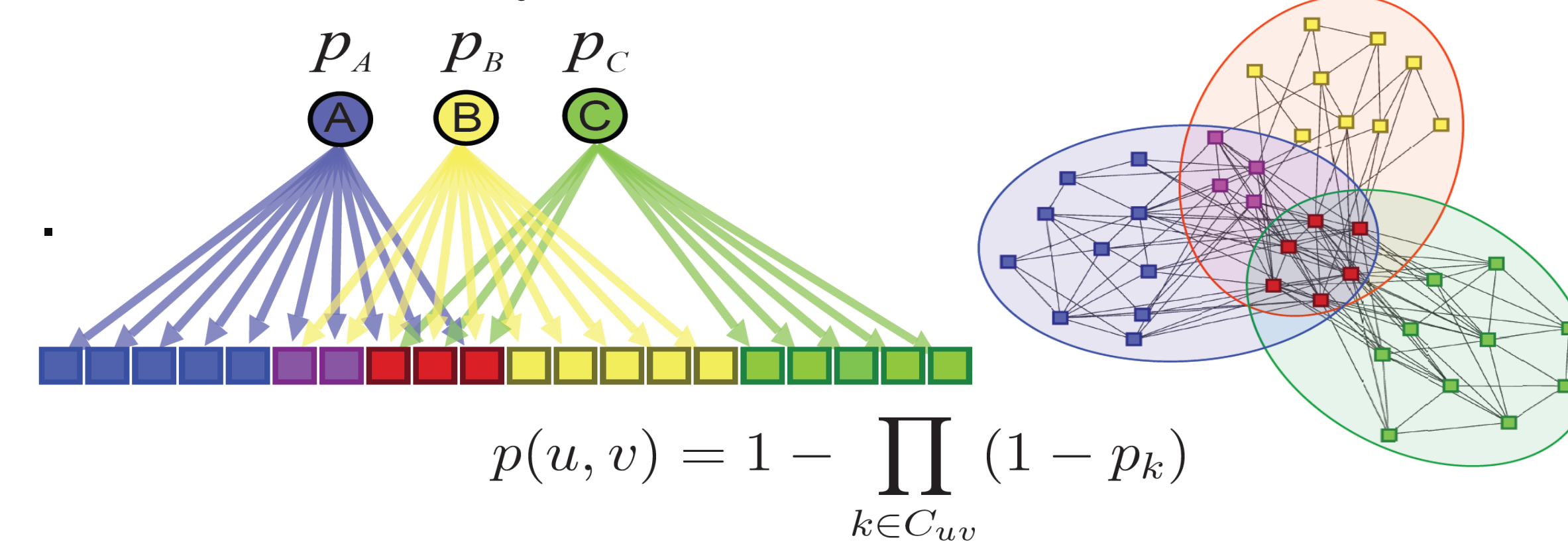
- Ground-truth communities form **denser** overlaps.



Affiliation Graph Model (AGM)

- Given community memberships, AGM generates a network.

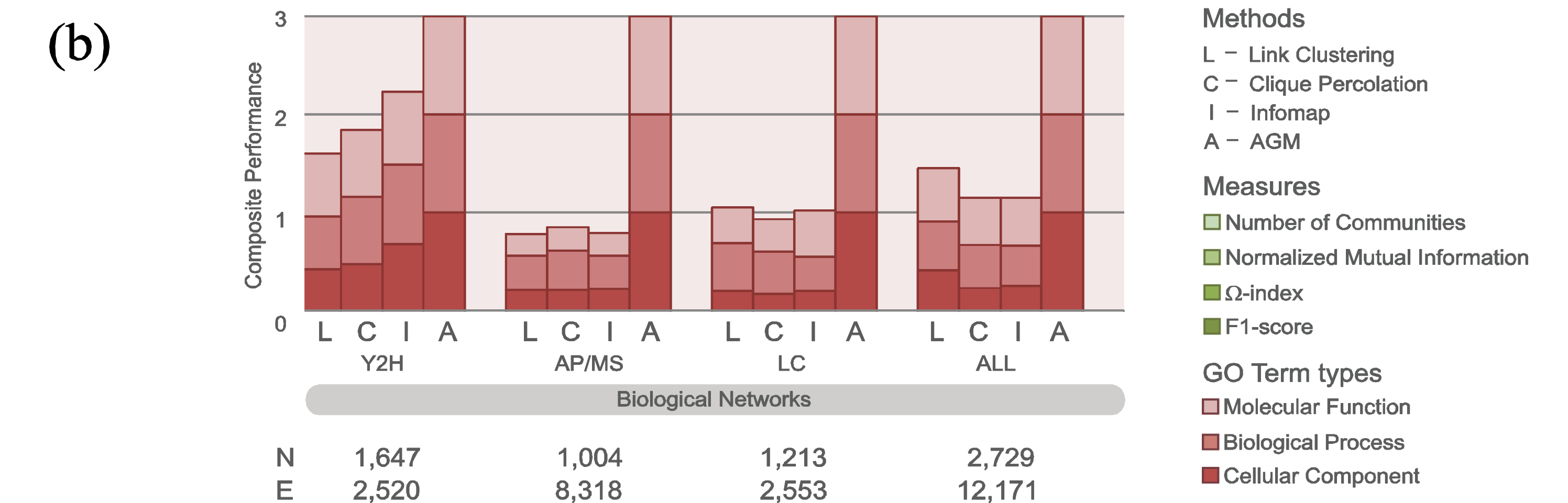
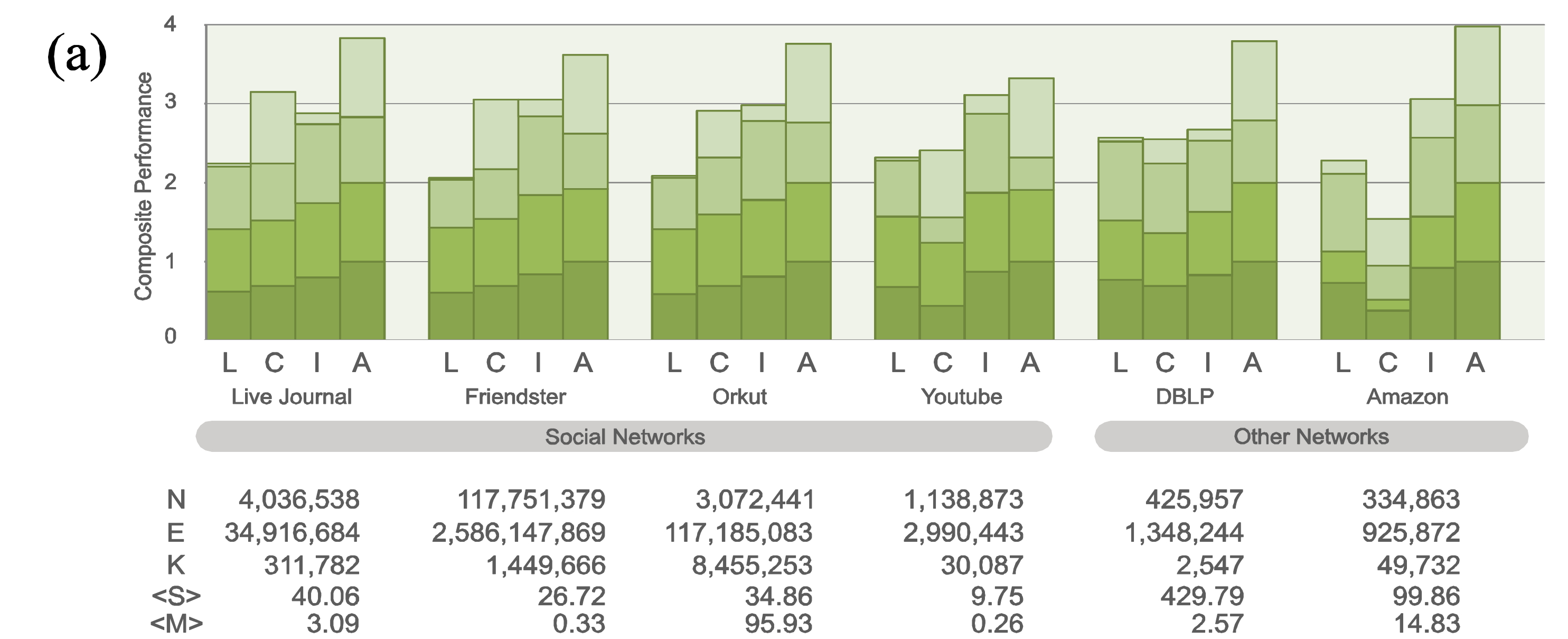
- Connect members of each community c with probability p_c .



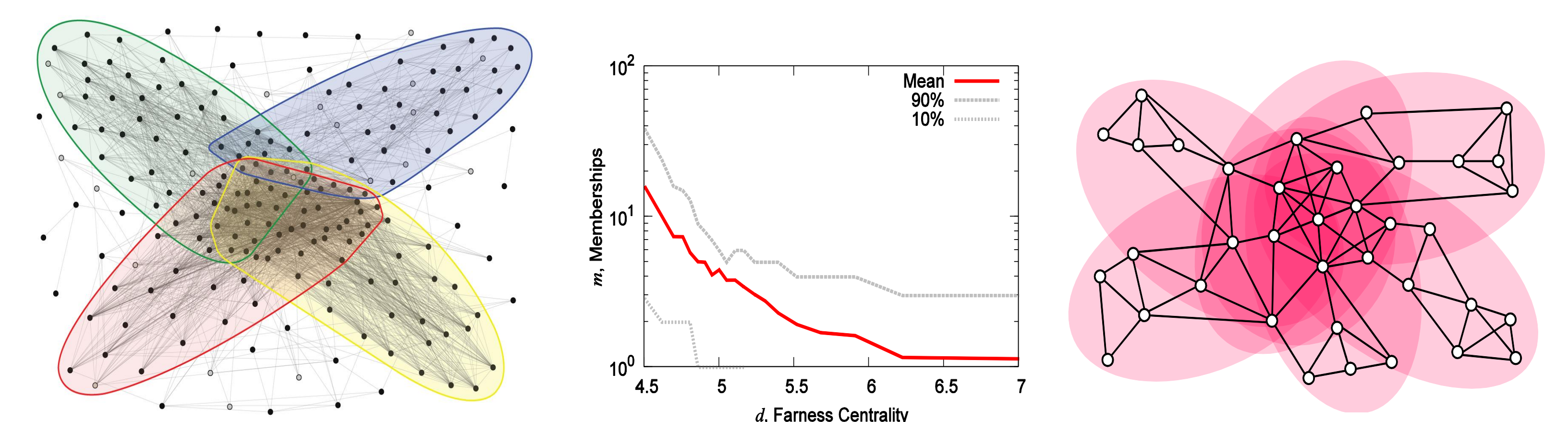
- Community detection by AGM:
 - Given a network, search **the most probable community memberships and probabilities p_c** .

Evaluation

- Networks with ground-truth communities: Measure "accuracy" of identifying ground-truth communities.
- Biological networks: Measure the **quality** of detected communities by meta-data.



Connection to Core-Periphery Structure



- Nodes in dense **overlaps** form dense **cores**.