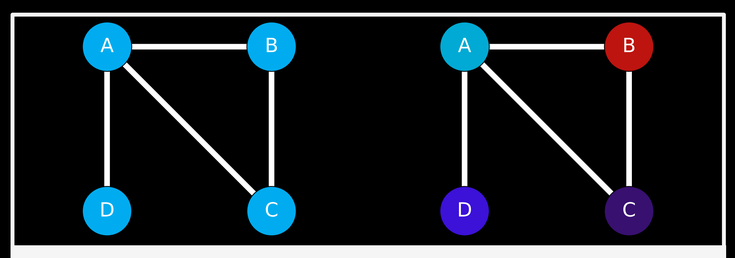
Robustness against Frustration in Community Detection

SRC 2025: Lucas Lopes Felipe, Ph.D. at the Federal University of Rio de Janeiro, Brazil - lucaslopes.me/sigmetrics2025

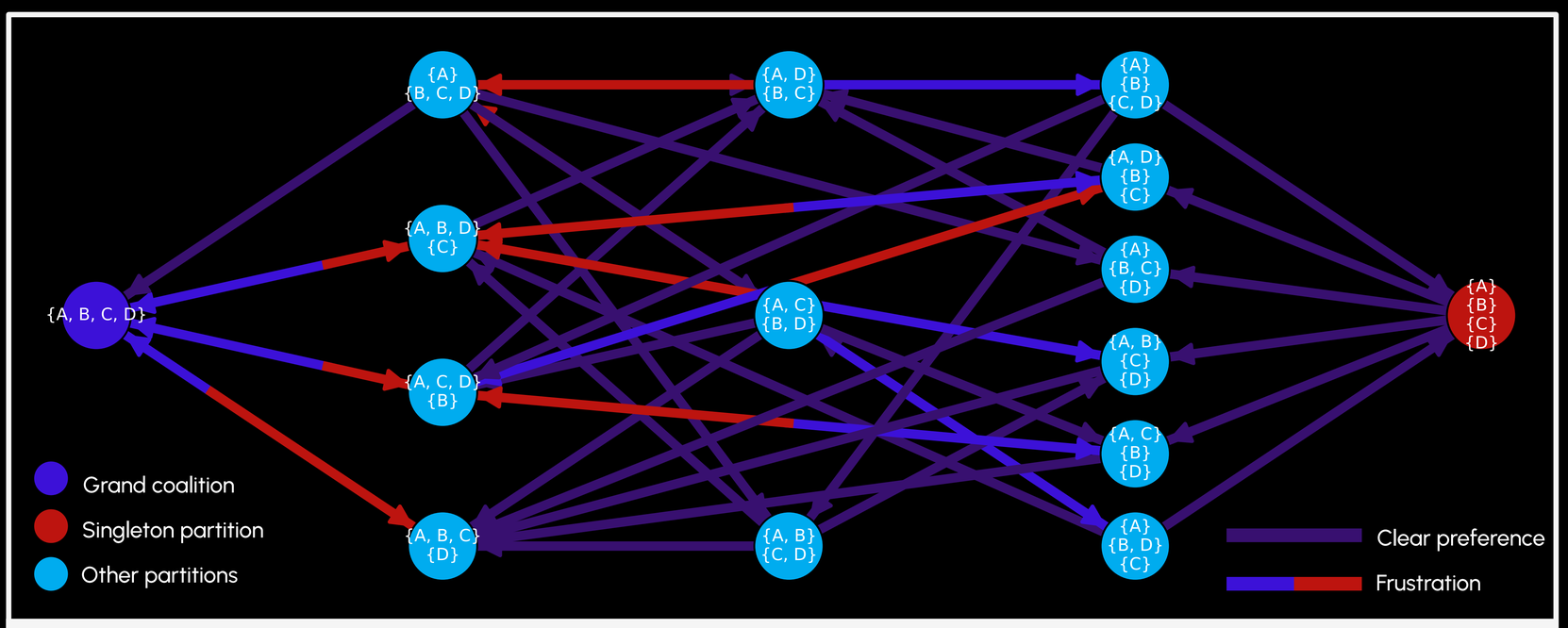
Clustering groups similar items; community detection is network clustering



Partition Segment the graph's vertices into disjoint sets (communities).
Left: The grand coalition forms a single community containing all vertices.
Right: The singleton partition places each vertex in its own community.

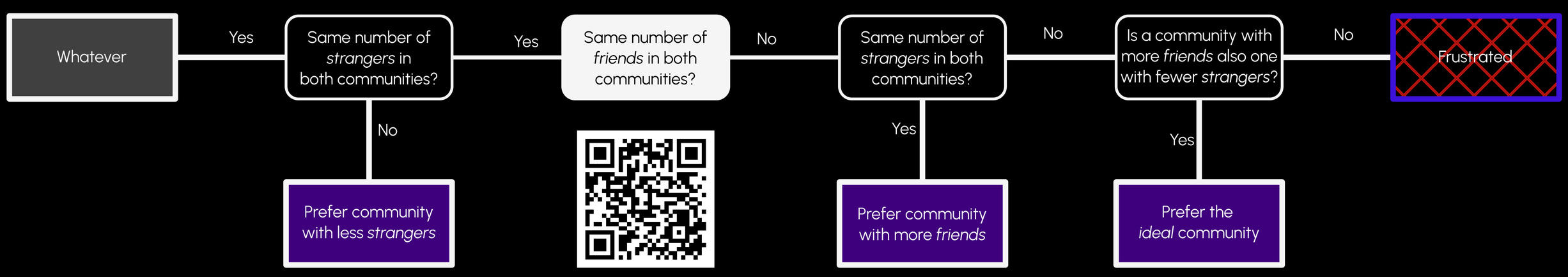
$$\gamma^\star = \frac{\mathrm{F}_i^{\mathcal{A}} - \mathrm{F}_i^{\mathcal{B}}}{(\mathrm{F}_i^{\mathcal{A}} - \mathrm{F}_i^{\mathcal{B}}) + (\mathrm{S}_i^{\mathcal{A}} - \mathrm{S}_i^{\mathcal{B}})}$$

Indifference Threshold creates equilibrium between two communities when a node (i) has more friends (neighbors) in community (A) and fewer strangers (non-neighbors) in community (B).



Meta-graph of Partitions To grasp the vast number of possible ways a graph can be partitioned, we can create a meta-graph. In this meta-graph, each vertex represents a unique partition. Each edge shows that two partitions differ by the movement of just one node. We use unidirectional edges when one partition is clearly preferred over the other. Bidirectional edges are used when the moving node is "frustrated," meaning it has more friends in one partition but fewer strangers in the other, unable to satisfy both the "maximize friends" and "minimize strangers" criteria simultaneously.

1) Equilibrium selection: How to select a partition among candidate solutions?
2) Equilibrium convergence: How many moves until find a candidate solution?



$$\begin{cases} \text{Internal Link Reward} &= (1-\gamma) \\ \text{Internal Non-Link Penalty} &= (-\gamma) \end{cases}$$

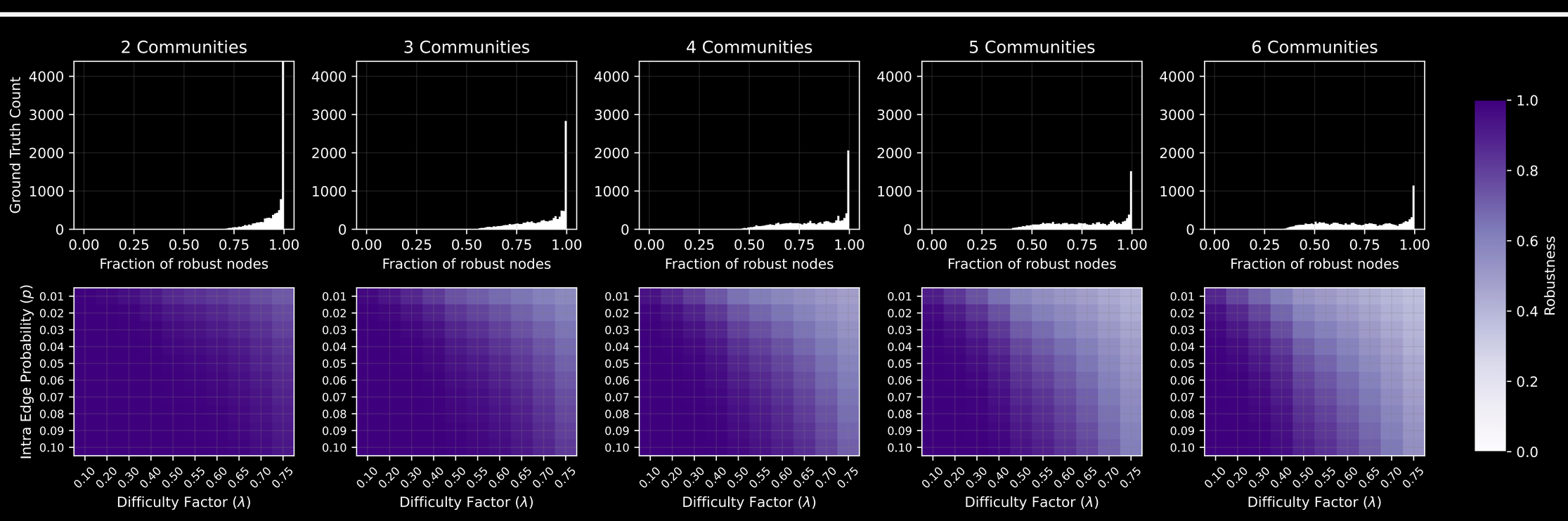
Reward & Penalty If we set a reward of (1-γ) for internal links (friends) and a penalty of (-γ) for internal non-links (strangers) we can formulate a single measure. This measure perfectly balances the goals of maximizing friends and minimizing strangers.

$$\Phi^\gamma(\pi) = (1-\gamma)\left(\sum_{k=1}^{K} m_k\right) - \gamma\left(\sum_{k=1}^{K}\binom{n_k}{2} - m_k\right)$$

Partition Quality Using the reward/penalty for friends and strangers, we can calculate a partition's quality by summing internal links and non-links across all communities. This model is known as the Constant Potts Model (CPM).

$$\gamma = \frac{b}{c}$$

Resolution If we assume the resolution is a rational number, say b/c then any move that improves the potential will increase it by at least 1/c. Since the total possible increase is bounded by by $2V^2$, the algorithm must converge in pseudo-polynomial time $O(cV^2)$.



Robustness of Ground Truth Partitions Using synthetic networks generated with the Symmetric Assortative Planted Partition Model (SAPPM)—a special case of the Stochastic Block Model (SBM) where communities have an equal number of nodes and the probability of an nternal link (p) is always greater than or equal to the probability of external links (q)—we can calculate the robustness of these "ground truth" partitions. This helps us verify if there's a correlation between robustness and accurate partitions. We've observed that for clearly defined partitions (either due to a. a high internal link probability (p) or a low "difficulty factor" (λ=q/p)), the ground truth partition is indeed robust. This suggests a positive correlation between accuracy and robustness.

CPM converges to robust partitions in pseudo-polynomial time!