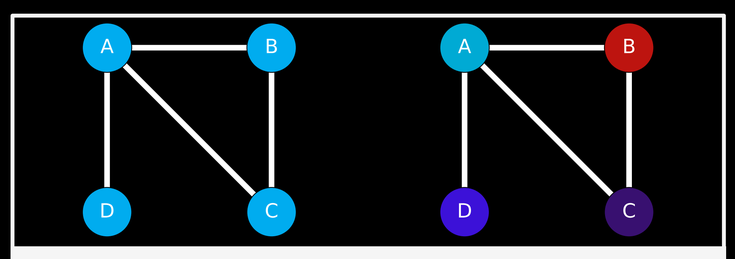
Robustness against Frustration in Community Detection

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

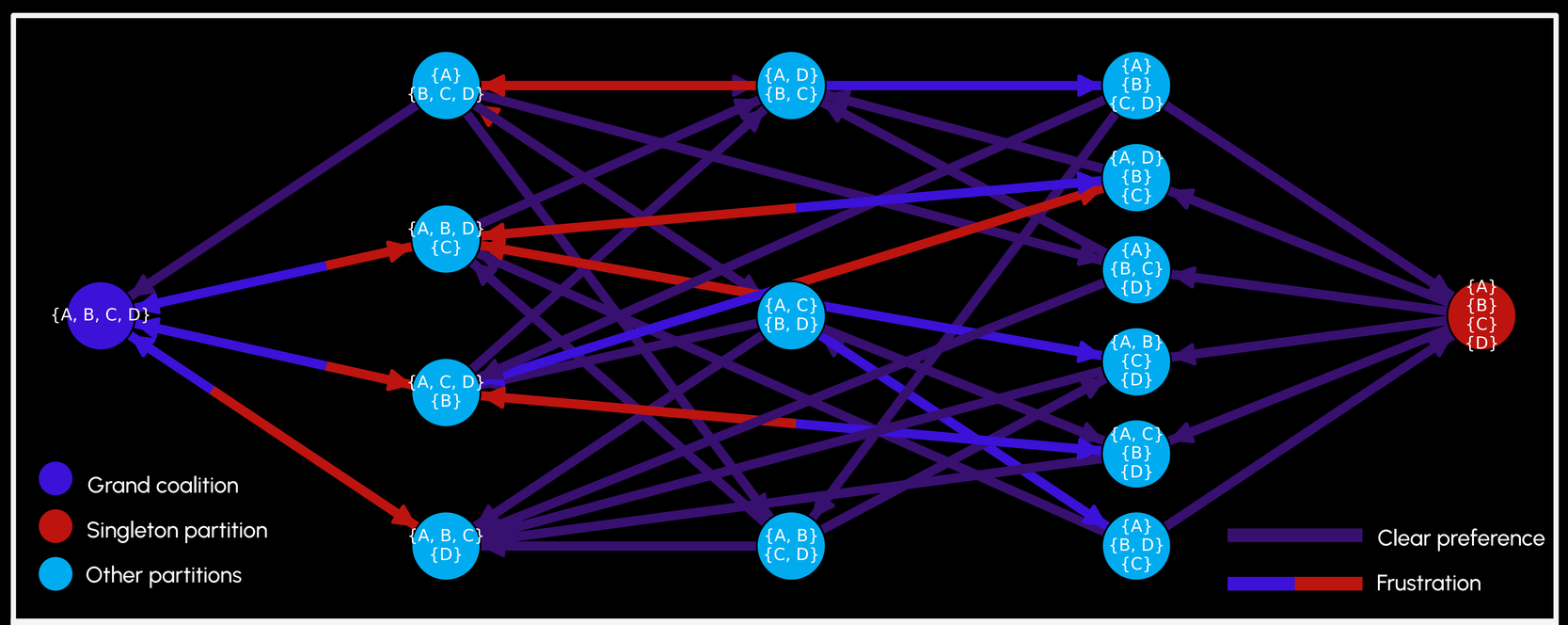
Clustering groups similar items; community detection is network clustering



Partition Segment the vertices of a graph into disjoint sets (communities)
(Left) The grand coalition is a single community containing all vertices of the graph
(Right) The singleton partition partitions each vertex in its own community

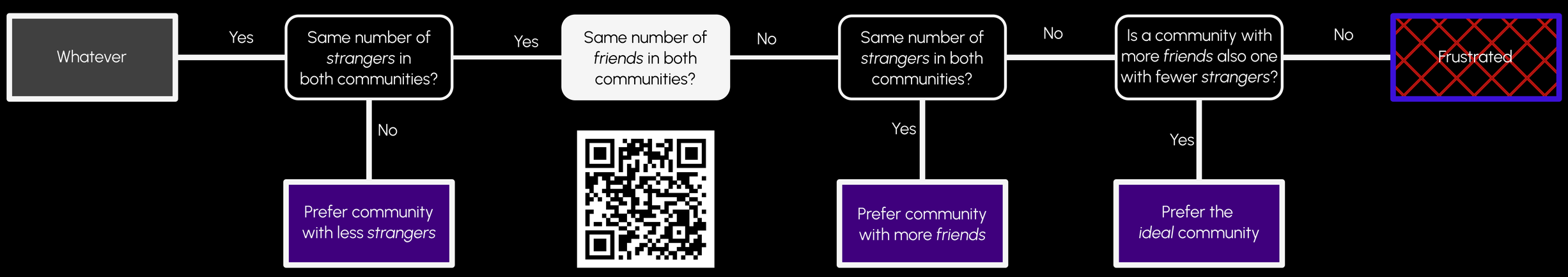
$$\gamma^\star = \frac{\mathrm{F}_i^{\mathcal{A}} - \mathrm{F}_i^{\mathcal{B}}}{(\mathrm{F}_i^{\mathcal{A}} - \mathrm{F}_i^{\mathcal{B}}) + (\mathrm{S}_i^{\mathcal{A}} - \mathrm{S}_i^{\mathcal{B}})}$$

Indifference Threshold When a node (i) has more Friends (neighbors) in a community (A) and fewer Stranger (non-neighbors) in the other communtiy (B), the indifference threshold balance the two communities to become equivalent.



Meta-graph of Partitions To conceptualize the vast space of possible partitions that a graph have, we can derive the meta-graph, where each vertex represent a possible partition, and each edge indicates that the two partitions differentiate by only one node move. We consider unidirectional edges for the case where there is a clear preference of a partition over the other, and bidirectional edges in case where the moving node is frustred, meaning that it has more friends in one partition, but fewer stranger in the other, unable to satisfy both criteria (max. friends & min. strangers) simultaniasly.

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



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

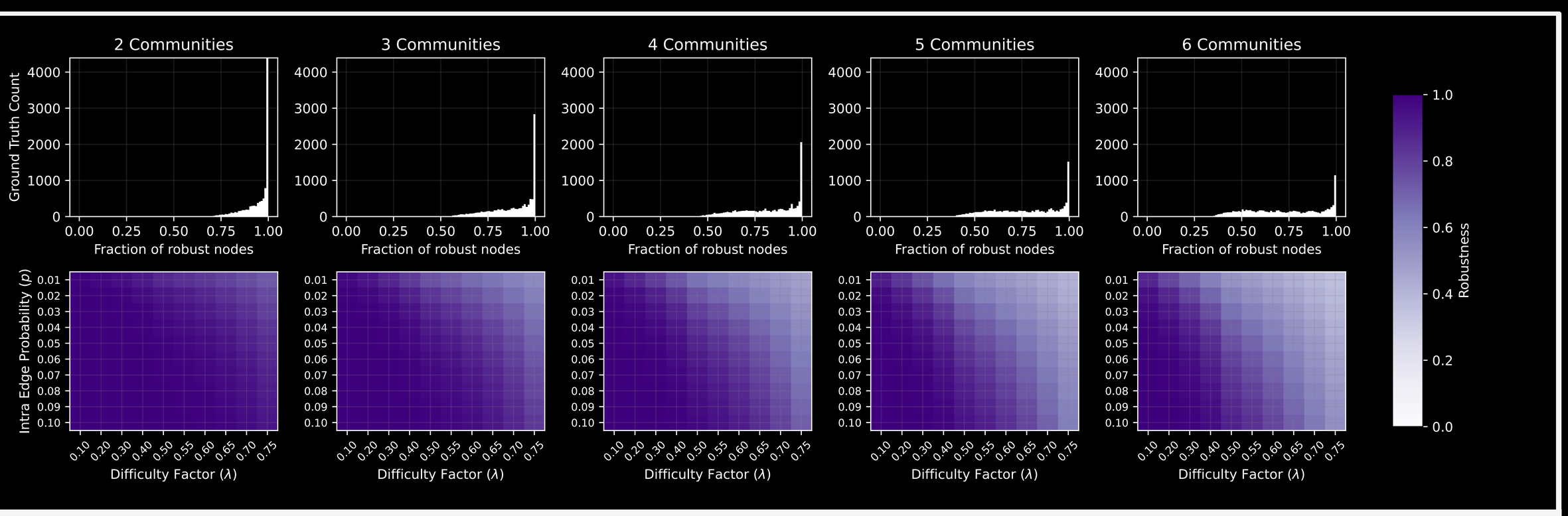
Reward & Penalty If we set a reward of (1-γ) for Internal Links (Friends) and a penalty of (-γ) for Internal Non-Links (Strangers) we can formulate a single measure that balances the preference for maximizing friends and minimizing strangers.

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

Partition Quality From the reward and penalty for friends and strangers we can derive the quality of a partition by computing the internal links and internal non-links among all the communities of the partition. This model is known as the Constant Potts Model (CPM).

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

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



Robustness of Ground Truth Partitions From synthetic networks generated with the Symmetric Assortative Planted Partition Model (SAPPM), a special case of the Stochastic Block Model (SBM) where communities have equal size in nodes and the probability of an internal link (p). is always greater or equal than the probability of external links (q), we can calculate the robustness of these ground truth partitions in order to verify if there is a correlation between robustness and accurate partitions. We can observe that for clear definied partitions (either because the internal link probability is high (p), or the difficulty factor (λ=q/p) is low) the ground truth partition is indeed robust. This suggest a positive correlation between accuracy and robustness.

CPM converges in pseudo-polynomial time to robust partitions!