BEYOND NODE CLASSIFICATION: GROUP DETECTION THROUGH LEARNED GRAPH COARSENING

Anonymous authors
Paper under double-blind review

ABSTRACT

Money-laundering schemes, coordinated manipulation on social platforms, and fraud rings are inherently group-level: the object of interest is a coordinated set of nodes whose members may appear unremarkable in isolation. Yet graph-learning approaches often reduce such problems to node-level prediction. In this paper, we address group detection directly, treating the group itself as the prediction unit. We formulate the task as learned graph coarsening: from training groups, we learn a target subspace under which unknown groups emerge as supernodes. We characterize the target-subspace properties required for coarsening to detect groups and translate them into a learning objective. We parameterize the target with a GNN, enabling sparse computation and inductive transfer to unseen graphs. We establish bounds relating these target-subspace properties to group-recovery error. To convert the learned target into group predictions, we adapt Ward clustering, introduce a merge score linked to coarsening error, and derive sufficient conditions for exact group recovery. Experiments demonstrate high-precision detection across synthetic, financial, and large community graphs, with transfer to unseen graphs.

1 INTRODUCTION

Node classification, which assigns a label to each node from its features and graph context, is a fundamental graph-learning task (Kipf & Welling, 2017; Hamilton et al., 2017). It naturally applies when the prediction target is an individual entity, as in assigning functions to proteins in biological networks. GNNs are a leading approach to node classification, incorporating relational information by propagating and aggregating features over the graph (Kipf & Welling, 2017; Wu et al., 2019). Their success has also encouraged other problems to be reduced to node-level prediction, even when the behavior of interest is not a property of any single node.

In many applications, that behavior belongs to a group of nodes acting together: it emerges from the interactions among its members and is not apparent from any of them in isolation. Examples include coordinated abuse on social platforms, collusive behavior in online marketplaces, and money laundering in financial networks. In anti-money laundering (AML), a scheme is carried out by a coordinated set of accounts—a *gang*—whose members appear unremarkable individually, while their collective transaction pattern is suspicious (Weber et al., 2019; Altman et al., 2023; Song et al., 2024). GNN approaches to financial crime nevertheless classify individual accounts or transactions (Weber et al., 2019; Pareja et al., 2020; Liu et al., 2021). Even with perfect node labels, the output is a list of flagged accounts, and recovering the gangs from it remains a separate problem.

The relevant question becomes "which nodes, acting together, constitute a target group?" rather than "does this node have the target property?" We refer to this task as *group detection*. A target group may contain only a few nodes and take varied forms: cliques, cycles, stars, or irregular motifs. Such groups are also rare: most nodes belong to none of them. Community detection, by contrast, partitions the entire graph into comparatively large communities.

In this work, we address group detection directly, taking the group rather than the node as the unit of prediction. Given training groups, our objective is to infer the number, size, and membership of previously unknown groups. We formulate this as a *graph-coarsening* problem. Graph coarsening partitions nodes into connected blocks represented by supernodes, traditionally to reduce graph size while preserving spectral properties (Loukas & Vandergheynst, 2018; Loukas, 2019). The behavior of spectral coarsening depends on its target subspace $\mathcal{R}$, which specifies the information to preserve

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and thus which nodes are merged. Standard methods treat $\mathcal{R}$ as a fixed, generic low-frequency target. We instead learn $\mathcal{R}$ from training groups so that coarsening acts as a detector, with unknown target groups emerging directly as supernodes.

We characterize the properties required of an effective target subspace: it must preserve each group as a whole while suppressing internal distinctions that would cause the coarsener to keep its members separate. Building on this analysis, we introduce a GNN based on learnable Chebyshev graph filters that learns such a target subspace directly from training groups. The learned filters transfer to unseen graphs, where they construct graph-specific target subspaces that guide the detection of new groups. Thus, GNN-based learning determines *what kind of group* to seek, while graph coarsening determines *where its members are*. Our main contributions are:

- **Group detection through learned graph coarsening.** We adapt graph coarsening to group detection by learning its target subspace from training groups, enabling unknown groups to emerge as individual supernodes. To the best of our knowledge, this is the first work to use learned graph coarsening for group detection.
- **Target-subspace characterization.** We characterize what a target subspace must preserve and suppress for coarsening to identify groups as individual supernodes. These properties guide our learning objective, and we derive bounds relating them to group preservation and contraction.
- **Inductive and scalable target-subspace learning.** We develop a GNN based on Chebyshev graph filters that learns from training groups a transferable rule for constructing such targets. Sparse graph operations make this construction scalable to large graphs, while the learned filters transfer to unseen graphs and produce graph-specific targets.
- **Target-aware coarsening and exact recovery.** We adapt Ward clustering for group detection by using the learned target to define a merge score linked to coarsening approximation error, and derive sufficient conditions for exact group recovery.
- **Empirical validation and transfer.** We evaluate the proposed approach on synthetic graphs with controlled planted motifs, financial transaction networks, and large SNAP community graphs. The results show that the learned filters transfer to unseen graphs, achieving high-precision group detection across diverse group structures and graph scales.

2 Related Work

Graph coarsening and condensation. Graph coarsening aggregates nodes into supernodes while preserving selected graph properties (Loukas & Vandergheynst, 2018; Loukas, 2019). Graph condensation learns small synthetic graphs that preserve GNN performance, primarily for node-level tasks (Jin et al., 2022). We instead learn the coarsening target from training groups, turning coarsening itself into a group detector.

Subgraph classification and discovery. Subgraph classifiers label candidate subgraphs supplied as input (Alsentzer et al., 2020; Wang & Zhang, 2021; Bellei et al., 2024). RevTrack (Song et al., 2024), the closest method to ours, also searches for suspicious subgraphs, but requires candidate sender and receiver sets and represents candidates through their endpoints, abstracting internal paths. Our method requires no candidate endpoints and infers membership from internal group structure.

Subgraph mining and community detection. Subgraph mining seeks recurring structural patterns. SPMiner (Ying et al., 2024) searches for frequent motifs, while MotiFiesta (Oliver et al., 2022) learns related motif instances. Unlike these methods, we learn from training groups to detect new groups without requiring them to occur frequently. Community detection instead seeks cohesive graph structure: SEAL (Zhang et al., 2020) grows communities from seed nodes, CLARE (Wu et al., 2022) localizes and refines their boundaries, and ProCom (Wu et al., 2024) learns node-assignment patterns from example communities. Our target groups may be small, structurally varied, and not densely connected; we learn a coarsening target under which such groups emerge as supernodes.

3 Preliminaries and Paper Roadmap

Graph and target groups. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{W})$ be a connected, undirected weighted graph with node set $\mathcal{V} = \{1, \ldots, N\}$, $N = |\mathcal{V}|$, and nonnegative weighted adjacency matrix $\boldsymbol{W}$. We add self-loops through $\widetilde{\boldsymbol{W}} = \boldsymbol{W} + \boldsymbol{I}$, and define $\widetilde{\boldsymbol{D}} = \mathrm{diag}(\widetilde{\boldsymbol{W}}\mathbf{1})$, with $\widetilde{d}_i$ denoting its i-th diagonal

entry and $\mathbf{1} \in \mathbb{R}^N$ the all-ones vector. The normalized adjacency and symmetric normalized Laplacian are $\widehat{\boldsymbol{A}} = \widetilde{\boldsymbol{D}}^{-1/2}\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{D}}^{-1/2}$ and $\boldsymbol{L} = \boldsymbol{I} - \widehat{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. The eigenvalues of $\boldsymbol{L}$ satisfy $0 = \lambda_0 \le \cdots \le \lambda_{N-1} = \lambda_{\max} < 2$.

We address the problem of detecting groups in a graph. Let $\mathcal{S}_\star := \{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ denote the collection of pairwise-disjoint target groups to be detected. Each target group $\mathcal{S} \in \mathcal{S}_\star$ is a nonempty connected proper subset of $\mathcal{V}$, typically small relative to the graph and characterized by a collective pattern of interest; in AML, it represents a gang of coordinated accounts. Define its internal edge set as $\mathcal{E}_{\text{int}}(\mathcal{S}) := \{\{i,j\} \in \mathcal{E} : i,j \in \mathcal{S}\}$ and its boundary edge set as $\partial\mathcal{S} := \{\{i,j\} \in \mathcal{E} : i \in \mathcal{S},\ j \notin \mathcal{S}\}$, with each undirected edge counted once. For $i \in \mathcal{S}$, let $d_\partial(i) := \sum_{j \notin \mathcal{S}} W_{ij}$ be its total edge weight to nodes outside $\mathcal{S}$. Conversely, for $j \notin \mathcal{S}$, let $b_{\mathcal{S}}(j) := \sum_{i \in \mathcal{S}} W_{ji}$ be its total edge weight to nodes inside $\mathcal{S}$. The cut, augmented volume, and conductance of $\mathcal{S}$ are respectively defined as

$$\text{cut}(\mathcal{S}) = \sum_{i \in \mathcal{S}} d_\partial(i) = \sum_{j \notin \mathcal{S}} b_{\mathcal{S}}(j), \qquad \text{vol}(\mathcal{S}) = \sum_{i \in \mathcal{S}} \widetilde{d}_i, \qquad \Phi(\mathcal{S}) = \frac{\text{cut}(\mathcal{S})}{\text{vol}(\mathcal{S})}\,. \tag{1}$$

Small $\Phi(\mathcal{S})$ indicates that $\mathcal{S}$ is well separated from the surrounding graph.

Let $r_{\mathcal{S}} := \text{vol}(\mathcal{S})/\,\text{vol}(\mathcal{V})$, and $\mathbf{1}_{\mathcal{S}} \in \{0,1\}^N$ be the binary indicator of $\mathcal{S}$, with $(\mathbf{1}_{\mathcal{S}})_i = 1$ if $i \in \mathcal{S}$ and zero otherwise. Define the normalized degree-weighted indicator and its centered counterpart by

$$\boldsymbol{v}_{\mathcal{S}} := \frac{\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}_{\mathcal{S}}}{\sqrt{\text{vol}(\mathcal{S})}}, \qquad \bar{\boldsymbol{v}}_{\mathcal{S}} := \frac{\widetilde{\boldsymbol{D}}^{1/2}(\mathbf{1}_{\mathcal{S}} - r_{\mathcal{S}}\mathbf{1})}{\sqrt{\text{vol}(\mathcal{S})(1 - r_{\mathcal{S}})}}, \qquad \|\boldsymbol{v}_{\mathcal{S}}\|_2 = \|\bar{\boldsymbol{v}}_{\mathcal{S}}\|_2 = 1\,. \tag{2}$$

The signal $\boldsymbol{v}_{\mathcal{S}}$ is supported on $\mathcal{S}$, whereas $\bar{\boldsymbol{v}}_{\mathcal{S}}$ contrasts $\mathcal{S}$ with its complement. Both distinguish the complete group from the rest of the graph.

Graph coarsening. Let $\mathcal{R} \subseteq \mathbb{R}^N$ denote the target subspace supplied to the coarsener. Guided by $\mathcal{R}$, the coarsener partitions $\mathcal{V}$ into disjoint connected blocks $\mathcal{P} = \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$, each represented by a supernode. Let $\boldsymbol{v}_{\mathcal{B}_r}$ be the normalized degree-weighted indicator of block $\mathcal{B}_r$, as in Eq. 2, collected in $\boldsymbol{V}_{\mathcal{P}} := [\boldsymbol{v}_{\mathcal{B}_1}, \ldots, \boldsymbol{v}_{\mathcal{B}_n}] \in \mathbb{R}^{N \times n}$. Since the blocks are disjoint, $\boldsymbol{V}_{\mathcal{P}}^\top \boldsymbol{V}_{\mathcal{P}} = \boldsymbol{I}$. Coarsening a signal $\boldsymbol{x}$ gives the coarse coordinates $\boldsymbol{V}_{\mathcal{P}}^\top \boldsymbol{x}$, while lifting them back gives $\widetilde{\boldsymbol{x}} = \boldsymbol{\Pi}_{\mathcal{P}}\boldsymbol{x}$, $\boldsymbol{\Pi}_{\mathcal{P}} := \boldsymbol{V}_{\mathcal{P}}\boldsymbol{V}_{\mathcal{P}}^\top$. Following Loukas (2019), coarsening quality is characterized by a constant $\varepsilon_{\mathcal{P}} \ge 0$ satisfying $\|\boldsymbol{x} - \boldsymbol{\Pi}_{\mathcal{P}}\boldsymbol{x}\| \le \varepsilon_{\mathcal{P}}\|\boldsymbol{x}\|$, $\boldsymbol{x} \in \mathcal{R}$, for an appropriate graph-signal norm. Thus, $\mathcal{R}$ specifies the information to preserve, whereas $\boldsymbol{\Pi}_{\mathcal{P}}$ is determined by the returned partition. $\boldsymbol{\Pi}_{\mathcal{P}}$ is the Euclidean orthogonal projector performing degree-weighted averaging within each block. It preserves the zero-frequency mode, $\boldsymbol{\Pi}_{\mathcal{P}}\boldsymbol{u}_0 = \boldsymbol{u}_0$, and can be applied in linear time (Loukas, 2019, Prop. 7).

Paper roadmap. Our framework has three parts: what the target subspace must satisfy, how to learn such a subspace, and how to coarsen with it. Sec. 4 characterizes the properties a target subspace must satisfy for a coarsening to return a group as a single supernode, and expresses them through two quantities, capture and confusability. Sec. 5 turns these properties into a training objective and learns the target subspace from training groups with a GNN. Sec. 6 coarsens a graph under the learned target, and Sec. 7 evaluates the resulting detector. Fig. 1 shows the framework end-to-end.

4 Graph Coarsening as a Group Detector: Properties of $\mathcal{R}$

We formulate group detection as graph coarsening, designing its target subspace so that unknown groups emerge as supernodes. Formally, $\mathcal{S}$ is detected when $\mathcal{S} \in \mathcal{P}$: its nodes form one supernode without absorbing outside nodes. The target subspace $\mathcal{R}$ specifies the graph-signal behavior to preserve and thereby guides the partition. To induce this block, $\mathcal{R}$ must retain the direction distinguishing $\mathcal{S}$ as a whole from its surroundings, which helps prevent over-merging, while suppressing directions that distinguish its members and thus oppose their contraction.

Screened geometry and coarsening similarity. To make these requirements precise, we need to quantify which group-related directions are retained by $\mathcal{R}$ and how accurately the coarsening preserves them. We treat the group indicators in Eq. 2 as graph signals and equip the signal space with a graph-aware geometry. The Laplacian energy $\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}$ is a natural choice, but it vanishes on the zero-frequency mode and is only a seminorm, so its orthogonal projections need not be unique. We

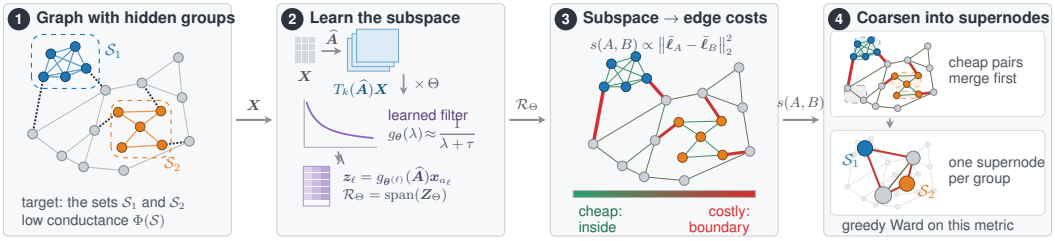


Figure 1: Group detection by learned graph coarsening. Chebyshev filters build a target subspace (2) that becomes an edge cost (3); greedy coarsening then returns each group as one supernode (4).

therefore introduce the shifted Laplacian $\boldsymbol{M}_\tau = \boldsymbol{L} + \tau \boldsymbol{I}$, $\tau \geq 0$, and refer to the geometry it induces as *screened*; the added term contributes Euclidean signal energy to the Laplacian geometry. For graph signals $\boldsymbol{x}$ and $\boldsymbol{y}$, let $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{M}_\tau} := \boldsymbol{x}^\top \boldsymbol{M}_\tau \boldsymbol{y}$ and $\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau} := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\boldsymbol{M}_\tau}}$. We write $\perp_{\boldsymbol{M}_\tau}$ for orthogonality in this geometry. For $\boldsymbol{X} \in \mathbb{R}^{N \times p}$ and $\boldsymbol{Y} \in \mathbb{R}^{N \times q}$, their screened cross-Gram matrix is $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\boldsymbol{M}_\tau} := \boldsymbol{X}^\top \boldsymbol{M}_\tau \boldsymbol{Y} \in \mathbb{R}^{p \times q}$; its (a, b)-th entry is $\langle \boldsymbol{x}_a, \boldsymbol{y}_b \rangle_{\boldsymbol{M}_\tau}$. For screened projections, we use $\tau > 0$, for which $\boldsymbol{M}_\tau$ is positive definite and the projections are unique. At $\tau = 0$, the geometry reduces to the classical Laplacian seminorm.

With this geometry in place, we assess how faithfully the resulting partition preserves the target subspace $\mathcal{R}$. Given an approximation tolerance $\varepsilon_{\mathcal{P}}$, we say that a coarsening is $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-*similar* if

$$\|\boldsymbol{x} - \boldsymbol{\Pi}_{\mathcal{P}} \boldsymbol{x}\|_{\boldsymbol{M}_\tau} \leq \varepsilon_{\mathcal{P}} \|\boldsymbol{x}\|_{\boldsymbol{M}_\tau} \qquad \text{for every} \qquad \boldsymbol{x} \in \mathcal{R}\,. \tag{3}$$

For $\tau = 0$, this is the restricted spectral approximation of Loukas (2019). For $\tau > 0$, the screened norm additionally penalizes Euclidean reconstruction error within supernodes.

Screened orthogonal projector. For an r-dimensional subspace $\mathcal{R} \subseteq \mathbb{R}^N$, let $\boldsymbol{U}_{\mathcal{R}}^\tau \in \mathbb{R}^{N \times r}$ be an $\boldsymbol{M}_\tau$-orthonormal basis satisfying $\mathrm{range}(\boldsymbol{U}_{\mathcal{R}}^\tau) = \mathcal{R}$, and $\langle\!\langle \boldsymbol{U}_{\mathcal{R}}^\tau, \boldsymbol{U}_{\mathcal{R}}^\tau \rangle\!\rangle_{\boldsymbol{M}_\tau} = \boldsymbol{I}_r$. The $\boldsymbol{M}_\tau$-orthogonal projector onto $\mathcal{R}$ is

$$\boldsymbol{Q}_{\mathcal{R}}^\tau := \boldsymbol{U}_{\mathcal{R}}^\tau (\boldsymbol{U}_{\mathcal{R}}^\tau)^\top \boldsymbol{M}_\tau. \tag{4}$$

Thus, $\boldsymbol{Q}_{\mathcal{R}}^\tau \boldsymbol{x} \in \mathcal{R}$ and $\boldsymbol{x} - \boldsymbol{Q}_{\mathcal{R}}^\tau \boldsymbol{x} \perp_{\boldsymbol{M}_\tau} \mathcal{R}$. If $\mathcal{R} = \mathrm{range}(\boldsymbol{Z})$ for a full-column-rank matrix $\boldsymbol{Z}$, one may construct the basis as $\boldsymbol{U}_{\mathcal{R}}^\tau = \boldsymbol{Z}(\boldsymbol{G}_{\mathcal{R}}^\tau)^{-1/2}$, where $\boldsymbol{G}_{\mathcal{R}}^\tau = \langle\!\langle \boldsymbol{Z}, \boldsymbol{Z} \rangle\!\rangle_{\boldsymbol{M}_\tau}$ is the Gram matrix of $\boldsymbol{Z}$.

Screened block distortion. The similarity criterion in Eq. 3 uses the screened geometry $\boldsymbol{M}_\tau$, whereas coarsening reconstructs signals with the Euclidean block-averaging projector $\boldsymbol{\Pi}_{\mathcal{P}}$, which need not be orthogonal in the screened geometry. Our analysis must therefore account for this geometric mismatch. Let $\mathcal{A}_{\mathcal{P}} := \mathrm{range}(\boldsymbol{V}_{\mathcal{P}})$ be the block-constant subspace. Applying Eq. 4 with $\boldsymbol{Z} = \boldsymbol{V}_{\mathcal{P}}$, its $\boldsymbol{M}_\tau$-orthogonal projector is $\boldsymbol{Q}_{\mathcal{P}}^\tau := \boldsymbol{V}_{\mathcal{P}} \big(\boldsymbol{V}_{\mathcal{P}}^\top \boldsymbol{M}_\tau \boldsymbol{V}_{\mathcal{P}}\big)^{-1} \boldsymbol{V}_{\mathcal{P}}^\top \boldsymbol{M}_\tau$. Since $\boldsymbol{Q}_{\mathcal{P}}^\tau$ gives the closest block-constant approximation in the screened norm, we quantify its discrepancy from $\boldsymbol{\Pi}_{\mathcal{P}}$ by the *screened block-distortion factor*

$$\mu_{\mathcal{P}}^\tau := \sup_{\substack{\boldsymbol{x} \in \mathbb{R}^N \\ (\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{P}}^\tau)\boldsymbol{x} \neq \boldsymbol{0}}} \frac{\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}{\|(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{P}}^\tau)\boldsymbol{x}\|_{\boldsymbol{M}_\tau}} = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}{\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}\,. \tag{5}$$

Thus, $\mu_{\mathcal{P}}^\tau = 1$ when the two projectors coincide, while larger values indicate greater mismatch. For every nontrivial coarsening, $1 \leq \mu_{\mathcal{P}}^\tau \leq \bar{\mu}^\tau$ with $\bar{\mu}^\tau := \frac{2\tau + \lambda_1 + \lambda_{\max}}{2\sqrt{(\tau + \lambda_1)(\tau + \lambda_{\max})}}$ (see App A.1).

Requirements on $\mathcal{R}$. Coarsening reconstructs a signal by averaging it within each block. It therefore preserves directions that are constant within a block and suppresses directions that vary across its nodes. A suitable target should represent $\mathcal{S}$ as a whole while retaining as little variation among its members as possible. We quantify these complementary properties by *capture* and *confusability* (see Fig. 2), defined next.

4.1 Capture: Preserving the Complete Group

For a target subspace to preserve $\mathcal{S}$ as a complete group, it must retain the direction that contrasts $\mathcal{S}$ with its complement. Define the screened-normalized group indicator by

$$\widehat{\boldsymbol{v}}_{\mathcal{S}} := \frac{\bar{\boldsymbol{v}}_{\mathcal{S}}}{\sqrt{\Phi(\mathcal{S})/(1 - r_{\mathcal{S}}) + \tau}}, \qquad \|\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} = 1\,. \tag{6}$$

Definition 1. The capture of $\mathcal{S}$ by $\mathcal{R}$ is

$$C^\tau_{\mathcal{R}}(\mathcal{S}) := \|\boldsymbol{Q}^\tau_{\mathcal{R}}\widehat{\boldsymbol{v}}_{\mathcal{S}}\|^2_{\boldsymbol{M}_\tau} \in [0,1]\,. \tag{7}$$

Thus, $C^\tau_{\mathcal{R}}(\mathcal{S})$ is the fraction of the complete-group direction retained by the target, and it equals one exactly when $\widehat{\boldsymbol{v}}_{\mathcal{S}} \in \mathcal{R}$.

Connection to Coarsening. To measure whether the coarse blocks cross the boundary of $\mathcal{S}$, define its *group-indicator distortion* $\delta^\tau_{\mathcal{P}}(\mathcal{S}) := \left\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\widehat{\boldsymbol{v}}_{\mathcal{S}}\right\|_{\boldsymbol{M}_\tau}$. Zero distortion means that $\mathcal{S}$ is captured by a union of coarse blocks. However, it does not rule out fragmentation: $\mathcal{S}$ may still be split across several such blocks.

Proposition 2. *Let $\mathcal{P}$ be a degree-weighted block coarsening that is $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similar. Then*

$$\delta^\tau_{\mathcal{P}}(\mathcal{S}) \le \varepsilon_{\mathcal{P}}\sqrt{C^\tau_{\mathcal{R}}(\mathcal{S})} + \mu^\tau_{\mathcal{P}}\sqrt{1 - C^\tau_{\mathcal{R}}(\mathcal{S})}\,. \tag{8}$$

Moreover, if $\mathcal{S}$ is contained in a coarse block $\mathcal{B}$, then

$$\rho_{\mathcal{B}}(\mathcal{S}) \ge 1 - \left(\frac{\Phi(\mathcal{S}) + \tau(1 - r_{\mathcal{S}})}{\tau + \lambda_1}\right)\left(\delta^\tau_{\mathcal{P}}(\mathcal{S})\right)^2, \tag{9}$$

where $\rho_{\mathcal{B}}(\mathcal{S}) = \frac{\mathrm{vol}(\mathcal{S})}{\mathrm{vol}(\mathcal{B})}$ is the purity of $\mathcal{B}$. In particular, $\delta^\tau_{\mathcal{P}}(\mathcal{S}) = 0$ and $\mathcal{S} \subseteq \mathcal{B}$ imply $\mathcal{B} = \mathcal{S}$, and hence exact detection.

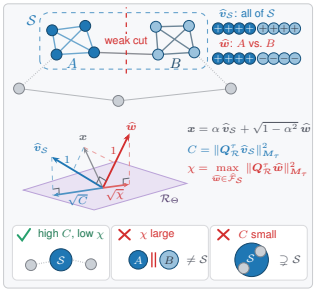


Figure 2: Capture vs. confusability.

Eq. 8 separates the error of the capture component, controlled by $\varepsilon_{\mathcal{P}}$, from that of the residual component, scaled by $\mu^\tau_{\mathcal{P}}$. Thus, high capture drives the distortion toward $\varepsilon_{\mathcal{P}}$, while Eq. 9 limits the outside volume absorbed by a block containing $\mathcal{S}$. The proof of Prop. 2 is in App. A.2.

4.2 Confusability: Excluding Internal Divisions

To contract all nodes of $\mathcal{S}$ into one supernode, $\mathcal{R}$ should suppress directions that distinguish its members. If $\mathcal{R}$ retains such a direction, contracting $\mathcal{S}$ eliminates it under block averaging and incurs reconstruction error, so the coarsener may instead split $\mathcal{S}$ across several supernodes. We formalize these internal differences through the following subspace. (Proofs in App. A.3)

Definition 3. The *internal fluctuation space* of $\mathcal{S}$ is

$$\mathcal{F}_{\mathcal{S}} = \left\{\boldsymbol{w} \in \mathbb{R}^N : \mathrm{supp}(\boldsymbol{w}) \subseteq \mathcal{S},\ \langle \boldsymbol{w}, \widehat{\boldsymbol{v}}_{\mathcal{S}}\rangle_2 = 0\right\}, \tag{10}$$

where $\mathrm{supp}(\boldsymbol{w}) := \{i \in \mathcal{V} : w_i \neq 0\}$. For each $\mathbf{0} \neq \boldsymbol{w} \in \mathcal{F}_{\mathcal{S}}$, define $\widehat{\boldsymbol{w}} := \boldsymbol{w}/\|\boldsymbol{w}\|_{\boldsymbol{M}_\tau}$. The corresponding unit sphere is $\widehat{\mathcal{F}}^\tau_{\mathcal{S}} := \{\widehat{\boldsymbol{w}} : \mathbf{0} \neq \boldsymbol{w} \in \mathcal{F}_{\mathcal{S}}\}$. When τ is clear, we write simply $\widehat{\mathcal{F}}_{\mathcal{S}}$.

Each internal fluctuation has support contained in $\mathcal{S}$ and is orthogonal to its complete-group direction, so it describes variation among the members of $\mathcal{S}$.

Definition 4 (Confusability)**.** The confusability of $\mathcal{S}$ with respect to $\mathcal{R}$ is

$$\chi^\tau_{\mathcal{R}}(\mathcal{S}) = \max_{\widehat{\boldsymbol{w}}\in\widehat{\mathcal{F}}_{\mathcal{S}}} \|\boldsymbol{Q}^\tau_{\mathcal{R}}\widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} \in [0,1]\,. \tag{11}$$

Confusability is the largest fraction of the screened energy of an internal fluctuation retained by $\mathcal{R}$. Low confusability means that $\mathcal{R}$ largely suppresses every internal distinction, making contraction compatible with preserving the target. When $\chi^\tau_{\mathcal{R}}(\mathcal{S}) \approx 1$, $\mathcal{R}$ nearly preserves at least one internal-fluctuation direction, opposing contraction. The following result quantifies this obstruction.

Theorem 5. *Let $\mathcal{P}$ be the partition returned by a block coarsening that is $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similar and contracts $\mathcal{S}$ into one supernode. Write $\chi = \chi^\tau_{\mathcal{R}}(\mathcal{S})$. If $\chi > 0$, then*

$$\varepsilon_{\mathcal{P}} \ge \left[\sqrt{\chi} - \sqrt{(\mu^\tau_{\mathcal{P}})^2 - 1}\,\sqrt{1-\chi}\right]_+. \tag{12}$$

As $\chi \to 1$, the lower bound approaches $\varepsilon_{\mathcal{P}} \ge 1$, making accurate target preservation incompatible with contracting the group.

Screened dual. We next construct a target subspace that, for a given group, completely suppresses its internal fluctuations. For $\tau > 0$, let $\boldsymbol{z}^\star := \boldsymbol{M}_\tau^{-1}\boldsymbol{z}$ denote the *screened dual* of $\boldsymbol{z}$.

Proposition 6. *Let $\bar{\boldsymbol{v}}_{\mathcal{S}}^\star := \boldsymbol{M}_\tau^{-1}\bar{\boldsymbol{v}}_{\mathcal{S}}$ be the screened dual of $\bar{\boldsymbol{v}}_{\mathcal{S}}$. Then, with $\mathcal{R}_\chi^\star(\mathcal{S}) := \operatorname{span}\{\bar{\boldsymbol{v}}_{\mathcal{S}}^\star\}$,*

$$\mathcal{R}_\chi^\star(\mathcal{S}) \in \arg\min_{\mathcal{R}} \chi_{\mathcal{R}}^\tau(\mathcal{S}) \qquad \textit{and} \qquad \chi_{\mathcal{R}_\chi^\star}^\tau(\mathcal{S}) = 0\,. \tag{13}$$

Thus, $\mathcal{R}_\chi^\star(\mathcal{S})$ suppresses all internal fluctuations of $\mathcal{S}$. It provides a zero-confusability reference for the learned targets introduced in Sec. 5.

5 Learning Transferable Coarsening Target Subspaces

Sec. 4 showed that a suitable target must retain complete-group directions and suppress internal fluctuations. A fixed low-frequency eigenspace $\mathcal{R}_r = \operatorname{span}\{\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{r-1}\}$ is a natural unsupervised target, but one value of r may not suit all groups. It also requires per-graph eigendecomposition and ignores node features (App. B). To overcome these limitations, our central proposal is to learn from training groups a rule for constructing graph-specific coarsening targets. The learned rule exploits node features, transfers inductively to unseen graphs, and is computationally inexpensive.

GNN target. We realize this idea with GNN-based learnable filters. Their spectral responses are learned to retain complete-group directions while suppressing internal fluctuations. Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d] \in \mathbb{R}^{N\times d}$ contain the input channels, using node features when available and random channels otherwise. We use H degree-K filters $g_{\boldsymbol{\theta}_h}(\widehat{\boldsymbol{A}}) := \sum_{k=0}^K \theta_{h,k} T_k(\widehat{\boldsymbol{A}})$, where T_k is the kth Chebyshev polynomial and $\boldsymbol{\theta}_h = (\theta_{h,0}, \ldots, \theta_{h,K})^\top$ contains the coefficients of filter h. Collecting them in $\Theta = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_H] \in \mathbb{R}^{(K+1)\times H}$, we apply each filter to all d input channels and define

$$\boldsymbol{Z}_\Theta := [g_{\boldsymbol{\theta}_1}(\widehat{\boldsymbol{A}})\boldsymbol{X}, \ldots, g_{\boldsymbol{\theta}_H}(\widehat{\boldsymbol{A}})\boldsymbol{X}] \in \mathbb{R}^{N\times q}, \qquad q := Hd, \qquad \mathcal{R}_\Theta := \operatorname{range}(\boldsymbol{Z}_\Theta)\,. \tag{14}$$

The ith row of $\boldsymbol{Z}_\Theta$ is a q-dimensional embedding of node i. Each filter learns a propagation response shared across the input channels, while different filters can retain different structural patterns. For any graph g, substituting $\widehat{\boldsymbol{A}}^{(g)}$ and $\boldsymbol{X}^{(g)}$ in Eq. 14 produces its graph-specific target $\mathcal{R}_\Theta^{(g)}$, so the learned filters transfer to unseen graphs. Degree-K Chebyshev filters combine information up to K hops and require only sparse graph operations, at cost $O(HKd|\mathcal{E}|)$ per graph and without eigendecomposition (Defferrard et al., 2016). Further details of the Chebyshev parameterization and its implementation are given in App. C.1.

Computing capture and confusability. To relate the target requirements to the GNN parameters, we express capture and confusability of Sec. 4 directly through the learned filter outputs. Let $\boldsymbol{G}_\Theta := \boldsymbol{Z}_\Theta^\top \boldsymbol{M}_\tau \boldsymbol{Z}_\Theta$ and define the *screened response* $\boldsymbol{\ell}_\Theta := \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau \boldsymbol{Z}_\Theta (\boldsymbol{G}_\Theta^\dagger)^{1/2} \in \mathbb{R}^{N\times q}$, where $\boldsymbol{G}_\Theta^\dagger$ is the pseudoinverse of $\boldsymbol{G}_\Theta$. The row $\boldsymbol{\ell}_{\Theta,v,:}$ is node v's screened response. For $\mathcal{B} \subseteq \mathcal{V}$, let $\bar{\boldsymbol{\ell}}_{\mathcal{B}} := \frac{1}{\operatorname{vol}(\mathcal{B})}\sum_{v\in\mathcal{B}} \widetilde{d}_v \boldsymbol{\ell}_{\Theta,v,:}^\top$ be its degree-weighted mean response. Define the within-group response scatter matrix $\boldsymbol{\Sigma}_{\mathcal{B}}(\boldsymbol{\ell}_\Theta) := \sum_{v\in\mathcal{B}} \widetilde{d}_v (\boldsymbol{\ell}_{\Theta,v,:}^\top - \bar{\boldsymbol{\ell}}_{\mathcal{B}})(\boldsymbol{\ell}_{\Theta,v,:}^\top - \bar{\boldsymbol{\ell}}_{\mathcal{B}})^\top$. Capture and confusability satisfy

$$C_{\mathcal{R}_\Theta}^\tau(\mathcal{S}) = \frac{\operatorname{vol}(\mathcal{S})}{\Phi(\mathcal{S}) + \tau(1 - r_{\mathcal{S}})} \left\|\bar{\boldsymbol{\ell}}_{\mathcal{S}} - \bar{\boldsymbol{\ell}}_{\mathcal{V}}\right\|_2^2, \qquad \chi_{\mathcal{R}_\Theta}^\tau(\mathcal{S}) \le \frac{1}{\tau + \lambda_1}\lambda_{\max}\big(\boldsymbol{\Sigma}_{\mathcal{S}}(\boldsymbol{\ell}_\Theta)\big)\,. \tag{15}$$

Thus, capture is the normalized energy of the contrast between the group's mean response and the global mean response, while confusability is controlled by within-group response variation. In particular, $\chi_{\mathcal{R}_\Theta}^\tau(\mathcal{S}) = 0$ if and only if $\boldsymbol{\ell}_{\Theta,v,:}^\top = \bar{\boldsymbol{\ell}}_{\mathcal{S}}$ for every $v \in \mathcal{S}$. Extending the screened-dual construction of Sec. 4 to all target groups, let $\bar{\boldsymbol{V}}_\star = [\bar{\boldsymbol{v}}_{\mathcal{S}_1}, \ldots, \bar{\boldsymbol{v}}_{\mathcal{S}_m}]$, $\boldsymbol{Z}^\star = \boldsymbol{M}_\tau^{-1}\bar{\boldsymbol{V}}_\star$, and $\mathcal{R}^\star = \operatorname{range}(\boldsymbol{Z}^\star)$. Then $\chi_{\mathcal{R}^\star}^\tau(\mathcal{S}_j) = 0$ for every j. Proofs and further analysis are given in App C.2.

5.1 Learning objective for multiple-group detection

The construction above makes the capture and confusability of any group functions of Θ. Because a graph may contain many groups whose number and memberships are unknown at inference, one graph-specific subspace must support their simultaneous detection. We therefore learn a single filter bank jointly from multiple training groups.

Learning from internal and boundary edges. Eq. 15 motivates agreement within groups and separation across their boundaries. We measure the response difference between adjacent nodes by $d^2_\Theta(u,v) := \|\boldsymbol{\ell}_{\Theta,u,:} - \boldsymbol{\ell}_{\Theta,v,:}\|_2^2$. Let $w_{\mathrm{int}}(\mathcal{S}) := \sum_{(u,v)\in\mathcal{E}_{\mathrm{int}}(\mathcal{S})} W_{uv}$ and $w_\partial(\mathcal{S}) := \sum_{(u,v)\in\partial\mathcal{S}} W_{uv}$. The corresponding weighted mean contrasts are $\mathrm{int}_\mathcal{S}(\Theta) := w_{\mathrm{int}}(\mathcal{S})^{-1} \sum_{(u,v)\in\mathcal{E}_{\mathrm{int}}(\mathcal{S})} W_{uv} d^2_\Theta(u,v)$ and $\mathrm{bnd}_\mathcal{S}(\Theta) := w_\partial(\mathcal{S})^{-1} \sum_{(u,v)\in\partial\mathcal{S}} W_{uv} d^2_\Theta(u,v)$. For training groups $\mathcal{S}_1,\dots,\mathcal{S}_m$ with positive internal and boundary weight, we formalize this trade-off through the learning objective

$$\mathcal{J}_{\mathrm{edge}}(\Theta) := \frac{1}{m}\sum_{j=1}^{m} \left[\mathrm{bnd}_{\mathcal{S}_j}(\Theta) - \gamma\,\mathrm{int}_{\mathcal{S}_j}(\Theta)\right], \qquad \gamma \geq 0\,, \tag{16}$$

where γ controls the trade-off between boundary separation and within-group response agreement. The analysis in Sec. 6 connects these patterns to Ward scores: within-group agreement favors inexpensive internal merges, whereas boundary separation favors expensive cross-boundary merges.

Although Eq. 16 admits direct gradient-based optimization, our experiments use a faster pooled quadratic surrogate shared across the training graphs. The surrogate preserves the boundary-reward and internal-penalty structure of $\mathcal{J}_{\mathrm{edge}}$ and reduces training to a single generalized eigenproblem. Its construction and closed-form solution are given in App. C.3. At inference, the learned filters are fixed and construct a graph-specific target on an unseen graph, after which the coarsener predicts groups without labels from that graph.

6 Adapting Ward Clustering for Group Detection

Secs. 4–5 characterize and learn the target subspace $\mathcal{R}_\Theta$. We now adapt Ward clustering to use this subspace for group detection. Standard Ward is a hierarchical method that starts from singleton nodes and repeatedly merges the block pair producing the smallest increase in within-cluster distortion. Our adaptation restricts merges to adjacent blocks and instead scores how much information from $\mathcal{R}_\Theta$ each merge discards in the screened geometry. Thus, every supernode represents a connected block. Repeated merging produces a hierarchy, and we return the coarsest partition satisfying the prescribed approximation tolerance. Capture and confusability yield a score-gap guarantee for exact recovery.

For the current partition $\mathcal{P}_n = \{B_1,\dots,B_n\}$, the realized block-averaging error $\varepsilon_{\mathcal{P}_n}$ is the smallest tolerance in Eq. 3 for $\boldsymbol{\Pi}_{\mathcal{P}_n}$. The screened block projector gives the best block-constant approximation in the screened norm. We define its worst relative error on $\mathcal{R}_\Theta$ as $\varepsilon^Q_{\mathcal{P}_n} := \max_{\boldsymbol{x}\in\mathcal{R}_\Theta\setminus\{\mathbf{0}\}} \|(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}_n})\boldsymbol{x}\|_{M_\tau}/\|\boldsymbol{x}\|_{M_\tau}$. The following result relates the two errors.

***Proposition* 7.** *For every partition $\mathcal{P}_n$, $\varepsilon^Q_{\mathcal{P}_n} \leq \varepsilon_{\mathcal{P}_n} \leq \mu^\tau_{\mathcal{P}_n}\varepsilon^Q_{\mathcal{P}_n}$, where $\mu^\tau_{\mathcal{P}_n}$ is defined in Eq. 5.*

Since $\mu^\tau_{\mathcal{P}_n} \leq \bar{\mu}_\tau$ (see App. A.1), Prop. 7 gives $\varepsilon_{\mathcal{P}_n} \leq \bar{\mu}_\tau\varepsilon^Q_{\mathcal{P}_n}$. Thus, minimizing $\varepsilon^Q_{\mathcal{P}_n}$ minimizes a uniform upper bound on the realized error. We use it as our surrogate. Optimization over all partitions is combinatorial, motivating the greedy score below.

Ward score. At each step, we define a merge score that measures how much target information a candidate merge discards. For adjacent blocks $\mathcal{B}_i,\mathcal{B}_j \in \mathcal{P}_n$, let $\mathcal{P}_{n-1} = \mathcal{P}_n \oplus (\mathcal{B}_i,\mathcal{B}_j)$ and define the block-difference direction $\boldsymbol{h}_{\mathcal{B}_i,\mathcal{B}_j} := \boldsymbol{v}_{\mathcal{B}_i}/\sqrt{\mathrm{vol}(\mathcal{B}_i)} - \boldsymbol{v}_{\mathcal{B}_j}/\sqrt{\mathrm{vol}(\mathcal{B}_j)}$, normalized as $\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j} := \boldsymbol{h}_{\mathcal{B}_i,\mathcal{B}_j}/\|\boldsymbol{h}_{\mathcal{B}_i,\mathcal{B}_j}\|_{M_\tau}$. If $\mathcal{B}_i,\mathcal{B}_j \subseteq \mathcal{S}$, then $\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j} \in \widehat{\mathcal{F}}^\tau_\mathcal{S}$. The component of this contrast eliminated by the merge, normalized in the screened geometry, is $\widehat{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j} := (\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}_{n-1}})\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}/\|(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}_{n-1}})\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}\|_{M_\tau}$. We define the Ward score as

$$s_{\mathcal{P}_n}(\mathcal{B}_i,\mathcal{B}_j) := \|\boldsymbol{Q}^\tau_{\mathcal{R}_\Theta}\widehat{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}\|^2_{M_\tau}. \tag{17}$$

The score is therefore the fraction of the eliminated direction retained by $\mathcal{R}_\Theta$. The coarsener selects the minimum-score adjacent pair; implementation details and the algorithm are given in App. D.1.

***Proposition* 8.** *For any merge sequence from singletons, with $\dim(\mathcal{R}_\Theta) \leq q$, $(\varepsilon^Q_{\mathcal{P}_n})^2 \leq \sum_{k=n+1}^{N} s_{\mathcal{P}_k}(\mathcal{B}_i^{(k)},\mathcal{B}_j^{(k)}) \leq q(\varepsilon^Q_{\mathcal{P}_n})^2$, where $(\mathcal{B}_i^{(k)},\mathcal{B}_j^{(k)})$ is the pair merged at the k-block partition.*

Thus, the cumulative merge score controls the surrogate within a factor q, motivating the greedy selection of the minimum-score pair.

6.1 EXACT GROUP RECOVERY

We call $\mathcal{P}_n$ $\mathcal{S}$-clean if every block is contained in $\mathcal{S}$ or disjoint from it. Within an $\mathcal{S}$-clean partition, an adjacent pair $\mathcal{B}_i, \mathcal{B}_j$ is *internal* if both blocks are contained in $\mathcal{S}$ and *boundary* if exactly one is contained in $\mathcal{S}$. Denote their scores by $s_{\mathrm{int}}(\mathcal{B}_i, \mathcal{B}_j; \mathcal{S})$ and $s_{\partial}(\mathcal{B}_i, \mathcal{B}_j; \mathcal{S})$, respectively. Exact recovery requires the lowest internal score to remain below every boundary score until $\mathcal{S}$ forms one block.

Internal geometry. To connect these merge scores to the learned response, we first consider the raw score $s^0(\mathcal{B}_i, \mathcal{B}_j) := \|\boldsymbol{Q}^{\tau}_{\mathcal{R}_\Theta} \bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}\|^2_{\boldsymbol{M}_\tau}$. It replaces the eliminated direction $\widehat{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}$ in Eq. 17 with the original block contrast $\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}$. Prop. 9 relates the raw and actual Ward scores.

***Proposition* 9.** *Define $\eta_n(\mathcal{B}_i, \mathcal{B}_j) := |\langle \bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}, \widehat{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j} \rangle_{\boldsymbol{M}_\tau}|^2 \in (0, 1]$ as the squared alignment between the raw and deflated contrasts. Writing $s = s^0(\mathcal{B}_i, \mathcal{B}_j)$ and $\eta = \eta_n(\mathcal{B}_i, \mathcal{B}_j)$,*

$$\left[\sqrt{\eta s} - \sqrt{(1-\eta)(1-s)}\right]^2_+ \le s_{\mathcal{P}_n}(\mathcal{B}_i, \mathcal{B}_j) \le 1 - \left[\sqrt{\eta(1-s)} - \sqrt{(1-\eta)s}\right]^2_+ . \tag{18}$$

Moreover, $\eta_n(\mathcal{B}_i, \mathcal{B}_j) \ge \left(\mu^{\tau}_{\mathcal{P}_{n-1}}\right)^{-2}$; see App. D.

Thus, $\eta_n(\mathcal{B}_i, \mathcal{B}_j)$ controls how faithfully the raw score represents the actual Ward score. When $\eta_n(\mathcal{B}_i, \mathcal{B}_j) = 1$, deflation leaves the contrast unchanged and $s_{\mathcal{P}_n}(\mathcal{B}_i, \mathcal{B}_j) = s^0(\mathcal{B}_i, \mathcal{B}_j)$.

The raw score can be written as $s^0(\mathcal{B}_i, \mathcal{B}_j) = \|\bar{\boldsymbol{\ell}}_{\mathcal{B}_i} - \bar{\boldsymbol{\ell}}_{\mathcal{B}_j}\|_2^2 / \|\boldsymbol{h}_{\mathcal{B}_i,\mathcal{B}_j}\|^2_{\boldsymbol{M}_\tau}$. Thus, when $\eta_n(\mathcal{B}_i, \mathcal{B}_j)$ is sufficiently large, Prop. 9 shows that block-mean response agreement lowers the upper bound on s_{int}, while boundary response separation raises the lower bound on s_{∂}. The learning objective in Sec. 5 promotes both, helping preserve the score gap after deflation.

***Lemma* 10.** *If $\mathcal{B}_i, \mathcal{B}_j \subseteq \mathcal{S}$, then $\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j} \in \widehat{\mathcal{F}}^{\tau}_{\mathcal{S}}$ and hence $s^0(\mathcal{B}_i, \mathcal{B}_j) \le \chi^{\tau}_{\mathcal{R}_\Theta}(\mathcal{S})$.*

Boundary geometry. For $e = (i, j) \in \partial\mathcal{S}$, define the screened effective resistance and boundary alignment by $R_e := (\boldsymbol{e}_i - \boldsymbol{e}_j)^{\top}\left(\widetilde{\boldsymbol{D}}^{1/2}\boldsymbol{M}_\tau\widetilde{\boldsymbol{D}}^{1/2}\right)^{-1}(\boldsymbol{e}_i - \boldsymbol{e}_j)$, and $\zeta_e := \left(\mathrm{vol}(\mathcal{S})(1 - r_{\mathcal{S}})\left(\frac{\Phi(\mathcal{S})}{1-r_{\mathcal{S}}} + \tau\right) R_e\right)^{-\frac{1}{2}}$. Here, R_e measures screened connectivity across e: it is small when its endpoints are joined by strong or numerous paths and large near a weak bottleneck. The quantity $\zeta_e \in (0, 1]$ measures the edge's alignment with the complete-group direction.

***Theorem* 11.** *For a connected group $\mathcal{S}$, define $\kappa_{\mathcal{S}} := \min_{e\in\partial\mathcal{S}} \zeta_e^2$, $\eta_\star := \bar{\mu}_\tau^{-2}$. Let $C_{\mathcal{S}} := C^{\tau}_{\mathcal{R}_\Theta}(\mathcal{S})$, $\chi_{\mathcal{S}} := \chi^{\tau}_{\mathcal{R}_\Theta}(\mathcal{S})$. At every $\mathcal{S}$-clean partition, internal and boundary candidates satisfy*

$$s_{\mathrm{int}}(\mathcal{B}_i, \mathcal{B}_j; \mathcal{S}) \le 1 - \left[\sqrt{\eta_\star(1-\chi_{\mathcal{S}})} - \sqrt{(1-\eta_\star)\chi_{\mathcal{S}}}\right]^2_+ =: U_{\mathcal{S}}, \tag{19}$$

$$s_{\partial}(\mathcal{B}_i, \mathcal{B}_j; \mathcal{S}) \ge \left[\sqrt{\kappa_{\mathcal{S}} C_{\mathcal{S}}} - \sqrt{(1-\kappa_{\mathcal{S}})(1-C_{\mathcal{S}})}\right]^2_+ =: L_{\mathcal{S}}. \tag{20}$$

If $U_{\mathcal{S}} < L_{\mathcal{S}}$, the greedy Ward hierarchy forms $\mathcal{S}$ as a supernode before any merge crosses its boundary.

Low confusability lowers the internal upper bound, while high capture raises the boundary lower bound. Hence, $U_{\mathcal{S}} < L_{\mathcal{S}}$ ensures that an internal merge is always preferred to a boundary merge until $\mathcal{S}$ becomes a supernode. Proofs and our group detection algorithm are given in App. D.

7 EXPERIMENTS

We evaluate group recovery by our method across synthetic, financial, and product co-purchase networks. Synthetic graphs permit controlled variation of group size, internal density, and background connectivity, while Elliptic++ tests temporal transfer on Bitcoin wallet graphs (Elmougy & Liu, 2023). We further evaluate scalability and community recovery on the Amazon co-purchase graph, whose ground-truth communities correspond to product categories (Yang & Leskovec, 2012). Dataset construction and evaluation protocols are detailed in App. E.1.

Evaluation metrics. Coarsening produces a hierarchy of partitions $\mathcal{P}_N, \mathcal{P}_{N-1}, \ldots$, where $\mathcal{P}_n$ contains n supernodes. A *coarsening cut* is a partition $\mathcal{P}_n$ selected by the stopping rule. For group

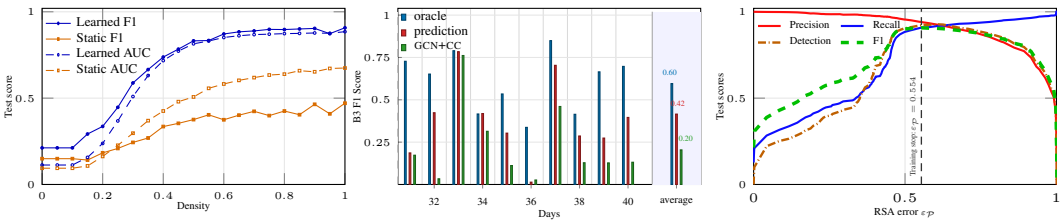


Figure 3: Group recovery across datasets. (left) Synthetic macro F1 (solid) and PR-AUC (dashed): blue is the learned target and green is the rank-matched static target. (center) Elliptic++ B^3 F1 by day: oracle coarsening (blue), coarsening + node prediction (red), and GNN+components (green). Days 31–40 are held out. (right) Held-out com-Amazon communities along the complete learned hierarchy, plotted against its RSA error $\varepsilon_{\mathcal{P}}$: group F1 (green), detection rate (brown), group precision (red), and group recall (blue). The dashed vertical line marks the cut selected on training communities.

$\mathcal{S}$, let $\mathcal{B}_n^\star(\mathcal{S}) = \arg\max_{\mathcal{B}\in\mathcal{P}_n} |\mathcal{S}\cap\mathcal{B}|$ be its best-matching supernode at the selected cut. The *group precision* and *group recall* at that cut are $p_n(\mathcal{S}) = |\mathcal{S}\cap\mathcal{B}_n^\star(\mathcal{S})|/|\mathcal{B}_n^\star(\mathcal{S})|$, and $r_n(\mathcal{S}) = |\mathcal{S}\cap\mathcal{B}_n^\star(\mathcal{S})|/|\mathcal{S}|$. Their harmonic mean is the *group F1 score*, $F_{1,\mathcal{P}_n}(\mathcal{S}) = \frac{2p_n(\mathcal{S})r_n(\mathcal{S})}{p_n(\mathcal{S})+r_n(\mathcal{S})}$. A group is detected when both $p_n(\mathcal{S})$ and $r_n(\mathcal{S})$ exceed 0.5; reported metrics are averaged over target groups.

For label-assisted group detection, we additionally report B^3 F1 (Amigó et al., 2009). B^3 F1 gives partial credit for approximately correct group membership rather than requiring exact recovery. Finally, PR-AUC summarizes the precision–recall trajectory over the full hierarchy and does not depend on the stopping threshold. Further metric details appear in App. E.1.

Synthetic. Each synthetic graph contains 12 planted random groups of size 7–20 with a prescribed within-group edge density. To study the effect of group conductance, we vary the within-group edge density; denser groups generally have lower conductance. To isolate the benefit of learning the coarsening geometry, we compare the learned subspace with a rank-matched static spectral subspace using the same deflated Ward coarsener. Fig. 3 (left) reports macro group F1 at a training-selected cut and PR-AUC over the full hierarchy, averaged over 15 held-out test graphs. The learned subspace performs better throughout: at density 0.40, its F1 and PR-AUC are 0.738 and 0.717, versus 0.335 and 0.425 for the static target; at density 1.00, they reach 0.907 and 0.885, versus 0.471 and 0.674.

Elliptic++. Elliptic++ is a Bitcoin fraud-detection dataset organized into daily wallet-transfer graphs, where nodes are wallets, edges are transactions, and wallets are labeled licit, illicit, or unknown. We define a malicious group as a connected component of at least two known-illicit wallets. To test inductive temporal generalization, models are fitted on days 26–30 and evaluated on held-out days 31–40. Detecting a malicious group requires both consolidating its wallets into one supernode and correctly flagging that supernode as illicit. We therefore train a node label predictor on the concatenated node features and learned screened responses $[X, \ell_{\boldsymbol{\Theta}}]$, labeling a supernode illicit when a majority of its members are predicted illicit. As a node-classification baseline, *GNN+components* classifies wallets with a GNN and returns the connected components induced by those predicted illicit. We also report *Oracle coarsening*, which replaces predicted labels with ground truth, providing an upper bound on the classification stage and isolating the group-recovery quality of the coarsening.

Fig. 3(center) reports B^3 F1 for each held-out day and average over the test period. Coarsening with label prediction achieves an average F1 of 0.416, compared with 0.204 for GNN+components, and performs better on nine of ten test days. Oracle coarsening reaches 0.596, indicating that classification accounts for a substantial part of the remaining error.

Amazon co-purchase network. com-Amazon is a product co-purchase network with $N = 334{,}863$ nodes and $E = 925{,}872$ edges. We apply our method and report, in Fig. 3 (right), macro group precision, recall, and F1 for test communities along the hierarchy as a function of the RSA error $\varepsilon_{\mathcal{P}}$. At $\varepsilon_{\mathcal{P}} = 0$, the partition $\mathcal{P}_N$ consists of singletons, giving $p_N(\mathcal{S}) = 1$ and $r_N(\mathcal{S}) = 1/|\mathcal{S}| \approx 0$; at the final partition $\mathcal{P}_1$, all nodes belong to one supernode, giving $r_1(\mathcal{S}) = 1$ and $p_1(\mathcal{S}) = |\mathcal{S}|/N \approx 0$. F1 is therefore maximized at an intermediate cut: at $\varepsilon_{\mathcal{P}} = 0.554$, our method attains precision 0.941, recall 0.907, and macro group F1 0.907, detecting 699 of 759 held-out communities (92.1%). We compare against *Louvain* (Blondel et al., 2008), a modularity-based community-detection method, and obtain precision 0.950, recall 0.832, and group F1 0.861, detecting 637 communities (83.9%).

Conclusion We introduced group detection through learned graph coarsening, where transferable Chebyshev filters construct targets that make unseen groups emerge as supernodes. Our analysis connects target capture and confusability to coarsening error and exact recovery. Across synthetic, Elliptic++, and com-Amazon graphs, the method improves over static spectral, node-classification, and Louvain baselines. Future work will consider overlapping or evolving groups.

REPRODUCIBILITY STATEMENT

The appendices provide complete proofs, experimental protocols, metric definitions, uncertainty procedures, and result tables. Code and resolved experimental configurations are available in the anonymous supplementary repository: `https://anonymous.4open.science/r/Group-Detection-via-Graph-Coarsening-E0E0/`.

AI USE STATEMENT

Generative AI assisted with coding, verification, literature organization, and editing. The authors reviewed all resulting derivations, analyses, citations, and prose and take full responsibility for the paper.

REFERENCES

Emily Alsentzer, Samuel Finlayson, Michelle Li, and Marinka Zitnik. Subgraph neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Erik Altman, Jovan Blanuša, Luc von Niederhäusern, Béni Egressy, Andreea Anghel, and Kubilay Atasu. Realistic synthetic financial transactions for anti-money laundering models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Enrique Amigó, Julio Gonzalo, Javier Artiles, and Felisa Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, 12(4):461–486, 2009. doi: 10.1007/s10791-008-9066-8.

Claudio Bellei, Muhua Xu, Ross Phillips, Tom Robinson, Mark Weber, Tim Kaler, Charles E. Leiserson, Arvind, and Jie Chen. The shape of money laundering: Subgraph representation learning on the blockchain with the Elliptic2 dataset. In *KDD Workshop on Machine Learning in Finance (KDD-W)*, 2024.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008 (10):P10008, 2008.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

Youssef Elmougy and Ling Liu. Demystifying fraudulent transactions and illicit nodes in the bitcoin network for financial forensics. In *ACM SIGKDD conference on knowledge discovery and data mining (KDD)*, 2023.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Wei Jin, Lingxiao Zhao, Shichang Zhang, Yozen Liu, Jiliang Tang, and Neil Shah. Graph condensation for graph neural networks. In *International Conference on Learning Representations (ICLR)*, 2022.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Yang Liu, Xiang Ao, Zidi Qin, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Pick and choose: A GNN-based imbalanced learning approach for fraud detection. In *Web Conference*, 2021.

Andreas Loukas. Graph reduction with spectral and cut guarantees. *Journal of Machine Learning Research*, 20(116):1–42, 2019.

Andreas Loukas and Pierre Vandergheynst. Spectrally approximating large graphs with smaller graphs. In *International Conference on Machine Learning (ICML)*, 2018.

Carlos G. Oliver, Dexiong Chen, Vincent Mallet, Pericles Philippopoulos, and Karsten M. Borgwardt. Approximate network motif mining via graph learning. *arXiv preprint arXiv:2206.01008*, 2022.

Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B. Schardl, and Charles E. Leiserson. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In *AAAI Conference on Artificial Intelligence*, 2020.

Kiwhan Song, Mohamed Ali Dhraief, Muhua Xu, Locke Cai, Xuhao Chen, Arvind Mithal, and Jie Chen. Identifying money laundering subgraphs on the blockchain. In *ACM International Conference on AI in Finance (ICAIF)*, 2024.

Xiyuan Wang and Muhan Zhang. GLASS: GNN with labeling tricks for subgraph representation learning. In *International Conference on Learning Representations (ICLR)*, 2021.

Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I. Weidele, Claudio Bellei, Tom Robinson, and Charles E. Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*, 2019.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In *International Conference on Machine Learning (ICML)*, 2019.

Xixi Wu, Yun Xiong, Yao Zhang, Yizhu Jiao, Caihua Shan, Yiheng Sun, Yangyong Zhu, and Philip S. Yu. CLARE: A semi-supervised community detection algorithm. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2022.

Xixi Wu, Kaiyu Xiong, Yun Xiong, Xiaoxin He, Yao Zhang, Yizhu Jiao, and Jiawei Zhang. ProCom: A few-shot targeted community detection algorithm. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024.

Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. In *ACM SIGKDD Workshop on Mining Data Semantics*, 2012.

Rex Ying, Tianyu Fu, Andrew Wang, Jiaxuan You, Yu Wang, and Jure Leskovec. Representation learning for frequent subgraph mining. *arXiv preprint arXiv:2402.14367*, 2024.

Yao Zhang, Yun Xiong, Yun Ye, Tengfei Liu, Weiqiang Wang, Yangyong Zhu, and Philip S. Yu. SEAL: Learning heuristics for community detection with generative adversarial networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2020.

A PROOFS

A.1 PROOF OF $\mu_{\mathcal{P}}^{\tau}$

Screened block distortion. Let $\mathcal{P}$ be a partition, let $\mathbf{\Pi}_{\mathcal{P}}$ be its Euclidean block-averaging projector, and let $\boldsymbol{Q}_{\mathcal{P}}^{\tau}$ be the screened-orthogonal projector onto the same block-constant subspace. For $\tau > 0$, define

$$\mu_{\mathcal{P}}^{\tau} := \sup_{\substack{\boldsymbol{x}\in\mathbb{R}^N \\ (\boldsymbol{I}-\boldsymbol{Q}_{\mathcal{P}}^{\tau})\boldsymbol{x}\neq\mathbf{0}}} \frac{\|(\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{P}})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}{\|(\boldsymbol{I}-\boldsymbol{Q}_{\mathcal{P}}^{\tau})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}. \tag{21}$$

This factor compares the error of Euclidean block averaging with the smallest screened error attainable in the block-constant subspace.

***Proposition* 12.** *Suppose the graph is connected and $\mathcal{P}$ is a nontrivial degree-weighted block coarsening. Then*

$$\mu_{\mathcal{P}}^{\tau} = \max_{\boldsymbol{x}\neq\mathbf{0}} \frac{\|(\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{P}})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}{\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}, \tag{22}$$

and, writing λ_1 for the first nonzero eigenvalue of $\boldsymbol{L}$,

$$1 \le \mu_{\mathcal{P}}^{\tau} \le\ \bar{\mu}_\tau, \qquad \bar{\mu}_\tau := \frac{2\tau+\lambda_1+\lambda_{\max}}{2\sqrt{(\tau+\lambda_1)(\tau+\lambda_{\max})}}. \tag{23}$$

Proof. Write $T := \boldsymbol{I} - \mathbf{\Pi}_{\mathcal{P}}$. The two projectors $\mathbf{\Pi}_{\mathcal{P}}$ and $\boldsymbol{Q}_{\mathcal{P}}^{\tau}$ have the same range, namely the block-constant subspace. Hence $T\boldsymbol{Q}_{\mathcal{P}}^{\tau} = \mathbf{0}$. For any $\boldsymbol{x}$, set $\boldsymbol{y} = (\boldsymbol{I}-\boldsymbol{Q}_{\mathcal{P}}^{\tau})\boldsymbol{x}$. Then $T\boldsymbol{x} = T\boldsymbol{y}$. Moreover, screened orthogonality gives $\|\boldsymbol{y}\|_{\boldsymbol{M}_\tau} \le \|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}$. It follows that the maximum of $\|T\boldsymbol{x}\|_{\boldsymbol{M}_\tau}/\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}$ can be attained with $\boldsymbol{x} \in \mathrm{range}(\boldsymbol{I}-\boldsymbol{Q}_{\mathcal{P}}^{\tau})$. For such a vector, the denominator in Eq. 21 is $\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}$. This proves Eq. 22.

Because $\mathcal{P}$ is nontrivial, the range of T contains a nonzero vector $\boldsymbol{x}$. For this vector $T\boldsymbol{x} = \boldsymbol{x}$, so Eq. 22 gives $\mu_{\mathcal{P}}^{\tau} \ge 1$.

We now prove the upper bound. Let U have Euclidean-orthonormal columns spanning $\mathrm{range}(T)$. Since T is a Euclidean orthogonal projector, $T = UU^{\top}$. The zero-frequency eigenvector belongs to the block-constant subspace: indeed,

$$\boldsymbol{u}_0 = \sum_{\mathcal{B}\in\mathcal{P}} \sqrt{\frac{\mathrm{vol}(\mathcal{B})}{\mathrm{vol}(\mathcal{V})}}\,\boldsymbol{v}_{\mathcal{B}}.$$

Consequently, $U^{\top}\boldsymbol{u}_0 = \mathbf{0}$. All columns of U therefore lie in $\boldsymbol{u}_0^{\perp}$, where the eigenvalues of $\boldsymbol{M}_\tau$ lie in

$$[\alpha,\beta], \qquad \alpha := \tau+\lambda_1, \qquad \beta := \tau+\lambda_{\max}.$$

Define the positive-definite matrices

$$A := U^{\top}\boldsymbol{M}_\tau U, \qquad B := U^{\top}\boldsymbol{M}_\tau^{-1}U.$$

From Eq. 22, the change of variables $\boldsymbol{y} = \boldsymbol{M}_\tau^{1/2}\boldsymbol{x}$ gives

$$\begin{aligned}(\mu_{\mathcal{P}}^{\tau})^2 &= \max_{\boldsymbol{y}\neq\mathbf{0}} \frac{\boldsymbol{y}^{\top}\boldsymbol{M}_\tau^{-1/2}UAU^{\top}\boldsymbol{M}_\tau^{-1/2}\boldsymbol{y}}{\boldsymbol{y}^{\top}\boldsymbol{y}} \\ &= \lambda_{\max}\Big(A^{1/2}BA^{1/2}\Big).\end{aligned}$$

For the second equality, the nonzero eigenvalues of $\boldsymbol{M}_\tau^{-1/2}UAU^{\top}\boldsymbol{M}_\tau^{-1/2}$ equal those of $A^{1/2}BA^{1/2}$.

For every $t \in [\alpha,\beta]$,

$$t + \frac{\alpha\beta}{t} \le \alpha+\beta,$$

because $(t-\alpha)(t-\beta) \le 0$. Applying this inequality to the eigenvalues of $\boldsymbol{M}_\tau$ on $\boldsymbol{u}_0^{\perp}$ and then multiplying by $U^{\top}$ and U gives

$$A + \alpha\beta B \preceq (\alpha+\beta)\boldsymbol{I}.$$

Thus

$$A^{1/2}BA^{1/2} \preceq \frac{(\alpha+\beta)A - A^2}{\alpha\beta} \preceq \frac{(\alpha+\beta)^2}{4\alpha\beta}\boldsymbol{I}.$$

The last inequality follows from $\left(A - \frac{\alpha+\beta}{2}\boldsymbol{I}\right)^2 \succeq \mathbf{0}$. Taking the largest eigenvalue and its square root yields

$$\mu_{\mathcal{P}}^{\tau} \le \frac{\alpha+\beta}{2\sqrt{\alpha\beta}} = \frac{2\tau + \lambda_1 + \lambda_{\max}}{2\sqrt{(\tau+\lambda_1)(\tau+\lambda_{\max})}}.$$

□

A.2 Proofs for Section 4.1 (capture)

Proposition. *Let $\mathcal{P}$ be a degree-weighted block coarsening that is $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similar. Then*

$$\delta_{\mathcal{P}}^{\tau}(\mathcal{S}) \le \varepsilon_{\mathcal{P}}\sqrt{C_{\mathcal{R}}^{\tau}(\mathcal{S})} + \mu_{\mathcal{P}}^{\tau}\sqrt{1 - C_{\mathcal{R}}^{\tau}(\mathcal{S})}. \tag{24}$$

Moreover, if $\mathcal{S}$ is contained in a coarse block $\mathcal{B}$, then

$$\rho_{\mathcal{B}}(\mathcal{S}) \ge 1 - \left(\frac{\Phi(\mathcal{S}) + \tau(1 - r_{\mathcal{S}})}{\tau + \lambda_1}\right)\left(\delta_{\mathcal{P}}^{\tau}(\mathcal{S})\right)^2, \tag{25}$$

where $\rho_B(\mathcal{S}) = \frac{\mathrm{vol}(\mathcal{S})}{\mathrm{vol}(\mathcal{B})}$ is the purity of $\mathcal{B}$. In particular, $\delta_{\mathcal{P}}^{\tau}(\mathcal{S}) = 0$ and $\mathcal{S} \subseteq \mathcal{B}$ imply $\mathcal{B} = \mathcal{S}$, and hence exact detection.

Proof. For $\tau > 0$, decompose the screened-normalized group indicator into its projection onto $\mathcal{R}$ and its screened-orthogonal residual:

$$\widehat{\boldsymbol{v}}_{\mathcal{S}} = \boldsymbol{Q}_{\mathcal{R}}^{\tau}\widehat{\boldsymbol{v}}_{\mathcal{S}} + (\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{R}}^{\tau})\widehat{\boldsymbol{v}}_{\mathcal{S}}.$$

Applying $\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}}$ and then the triangle inequality gives

$$\begin{aligned}\delta_{\mathcal{P}}^{\tau}(\mathcal{S}) &= \|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} \\ &\le \|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\boldsymbol{Q}_{\mathcal{R}}^{\tau}\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} + \|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{R}}^{\tau})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}.\end{aligned} \tag{26}$$

The vector $\boldsymbol{Q}_{\mathcal{R}}^{\tau}\widehat{\boldsymbol{v}}_{\mathcal{S}}$ belongs to $\mathcal{R}$. Hence, $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similarity, as defined in Eq. 3, bounds the first term by

$$\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\boldsymbol{Q}_{\mathcal{R}}^{\tau}\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} \le \varepsilon_{\mathcal{P}}\|\boldsymbol{Q}_{\mathcal{R}}^{\tau}\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} = \varepsilon_{\mathcal{P}}\sqrt{C_{\mathcal{R}}^{\tau}(\mathcal{S})}.$$

For the second term, the definition of the screened block-distortion factor gives

$$\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{R}}^{\tau})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} \le \mu_{\mathcal{P}}^{\tau}\|(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{R}}^{\tau})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}.$$

Because $\boldsymbol{Q}_{\mathcal{R}}^{\tau}$ is an $\boldsymbol{M}_\tau$-orthogonal projector and $\|\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau} = 1$, the screened Pythagorean identity yields

$$\begin{aligned}\|(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{R}}^{\tau})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}^2 &= \|\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}^2 - \|\boldsymbol{Q}_{\mathcal{R}}^{\tau}\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}^2 \\ &= 1 - C_{\mathcal{R}}^{\tau}(\mathcal{S}).\end{aligned}$$

Substitution into Eq. 26 proves Eq. 8.

For the purity bound, suppose that $\mathcal{S} \subseteq \mathcal{B}$ for a block $\mathcal{B} \in \mathcal{P}$. Recall that

$$\boldsymbol{u}_0 = \frac{\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}}{\sqrt{\mathrm{vol}(\mathcal{V})}} \quad \text{and} \quad \boldsymbol{v}_{\mathcal{S}} = \frac{\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}_{\mathcal{S}}}{\sqrt{\mathrm{vol}(\mathcal{S})}}.$$

Using $r_{\mathcal{S}} = \mathrm{vol}(\mathcal{S})/\mathrm{vol}(\mathcal{V})$, the centered group indicator in Eq. 2 can be written as

$$\bar{\boldsymbol{v}}_{\mathcal{S}} = \frac{\boldsymbol{v}_{\mathcal{S}} - \sqrt{r_{\mathcal{S}}}\,\boldsymbol{u}_0}{\sqrt{1 - r_{\mathcal{S}}}}.$$

Its squared screened norm is $\|\bar{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}^2 = \Phi(\mathcal{S})/(1-r_{\mathcal{S}}) + \tau$. Therefore,

$$\widehat{\boldsymbol{v}}_{\mathcal{S}} = \frac{\boldsymbol{v}_{\mathcal{S}} - \sqrt{r_{\mathcal{S}}}\,\boldsymbol{u}_0}{\sqrt{\Phi(\mathcal{S}) + \tau(1-r_{\mathcal{S}})}}. \tag{27}$$

The block projector preserves the zero-frequency mode: $\boldsymbol{\Pi}_{\mathcal{P}}\boldsymbol{u}_0 = \boldsymbol{u}_0$. Applying $\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}}$ to Eq. 27 thus gives

$$(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\widehat{\boldsymbol{v}}_{\mathcal{S}} = \frac{(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\boldsymbol{v}_{\mathcal{S}}}{\sqrt{\Phi(\mathcal{S}) + \tau(1-r_{\mathcal{S}})}}. \tag{28}$$

Since $\mathcal{S}$ is contained in $\mathcal{B}$, $\boldsymbol{v}_{\mathcal{S}}$ is orthogonal to the indicators of every block other than $\mathcal{B}$. Its projection onto the block-constant subspace is consequently

$$\boldsymbol{\Pi}_{\mathcal{P}}\boldsymbol{v}_{\mathcal{S}} = \langle \boldsymbol{v}_{\mathcal{B}}, \boldsymbol{v}_{\mathcal{S}} \rangle_2 \, \boldsymbol{v}_{\mathcal{B}}.$$

The overlap is

$$\langle \boldsymbol{v}_{\mathcal{B}}, \boldsymbol{v}_{\mathcal{S}} \rangle_2 = \frac{\sum_{i\in\mathcal{S}} \widetilde{d}_i}{\sqrt{\mathrm{vol}(\mathcal{B})\,\mathrm{vol}(\mathcal{S})}} = \sqrt{\frac{\mathrm{vol}(\mathcal{S})}{\mathrm{vol}(\mathcal{B})}}.$$

As $\boldsymbol{\Pi}_{\mathcal{P}}$ is a Euclidean orthogonal projector and $\|\boldsymbol{v}_{\mathcal{S}}\|_2 = 1$, it follows that

$$\begin{aligned}\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\boldsymbol{v}_{\mathcal{S}}\|_2^2 &= \|\boldsymbol{v}_{\mathcal{S}}\|_2^2 - \|\boldsymbol{\Pi}_{\mathcal{P}}\boldsymbol{v}_{\mathcal{S}}\|_2^2 \\ &= 1 - \frac{\mathrm{vol}(\mathcal{S})}{\mathrm{vol}(\mathcal{B})} = 1 - \rho_{\mathcal{B}}(\mathcal{S}).\end{aligned} \tag{29}$$

Finally, because $(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\widehat{\boldsymbol{v}}_{\mathcal{S}} \perp \boldsymbol{u}_0$, we have $\boldsymbol{M}_\tau = \boldsymbol{L} + \tau\boldsymbol{I} \succeq (\tau + \lambda_1)\boldsymbol{I}$. Combining this inequality with Eq. 28 and Eq. 29 yields

$$\begin{aligned}\big(\delta_{\mathcal{P}}^{\tau}(\mathcal{S})\big)^2 &= \|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}^2 \\ &\geq (\tau + \lambda_1)\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}})\widehat{\boldsymbol{v}}_{\mathcal{S}}\|_2^2 \\ &= \frac{\tau + \lambda_1}{\Phi(\mathcal{S}) + \tau(1-r_{\mathcal{S}})}\big(1 - \rho_{\mathcal{B}}(\mathcal{S})\big).\end{aligned}$$

Rearranging proves Eq. 9.

If $\delta_{\mathcal{P}}^{\tau}(\mathcal{S}) = 0$, then Eq. 29 implies $\rho_{\mathcal{B}}(\mathcal{S}) = 1$. Every augmented degree is positive, so a proper containment $\mathcal{S} \subsetneq \mathcal{B}$ would give $\mathrm{vol}(\mathcal{S}) < \mathrm{vol}(\mathcal{B})$. Thus $\mathcal{B} = \mathcal{S}$, establishing exact detection. □

A.3 Proofs for Section 4.2 (Confusability)

***Lemma* 13.** *If a degree-weighted block coarsening $\boldsymbol{\Pi}_{\mathcal{P}}$ contracts $\mathcal{S}$ into one supernode, then*

$$\boldsymbol{\Pi}_{\mathcal{P}}\widehat{\boldsymbol{w}} = \boldsymbol{0} \qquad \textit{for every } \widehat{\boldsymbol{w}} \in \widehat{\mathcal{F}}_{\mathcal{S}}^{\tau}.$$

Proof. Let $\mathcal{B} \in \mathcal{P}$ be the block containing $\mathcal{S}$, and let $\widehat{\boldsymbol{w}} \in \widehat{\mathcal{F}}_{\mathcal{S}}^{\tau}$. By definition, $\widehat{\boldsymbol{w}}$ is supported on $\mathcal{S}$ and is Euclidean-orthogonal to $\widehat{\boldsymbol{v}}_{\mathcal{S}}$.

On $\mathcal{S}$, the entries of $\widehat{\boldsymbol{v}}_{\mathcal{S}}$ are a common nonzero multiple of $\sqrt{\widetilde{d}_i}$. Indeed, by Eq. 2 and Eq. 6, for $i \in \mathcal{S}$,

$$(\widehat{\boldsymbol{v}}_{\mathcal{S}})_i = \frac{(1-r_{\mathcal{S}})\sqrt{\widetilde{d}_i}}{\sqrt{\mathrm{vol}(\mathcal{S})(1-r_{\mathcal{S}})}\sqrt{\Phi(\mathcal{S})/(1-r_{\mathcal{S}}) + \tau}}.$$

Orthogonality to $\widehat{\boldsymbol{v}}_{\mathcal{S}}$ therefore implies

$$\sum_{i\in\mathcal{S}} \sqrt{\widetilde{d}_i}\,\widehat{w}_i = 0.$$

Since $\widehat{\boldsymbol{w}}$ is supported on $\mathcal{S} \subseteq \mathcal{B}$, its inner product with the indicator of $\mathcal{B}$ is

$$\langle \boldsymbol{v}_{\mathcal{B}}, \widehat{\boldsymbol{w}} \rangle_2 = \frac{1}{\sqrt{\mathrm{vol}(\mathcal{B})}} \sum_{i\in\mathcal{S}} \sqrt{\widetilde{d}_i}\,\widehat{w}_i = 0.$$

Its inner product with every other block indicator also vanishes, because those blocks do not intersect its support. Thus $\widehat{\boldsymbol{w}}$ is orthogonal to every column of $\boldsymbol{V}_{\mathcal{P}}$. Using $\boldsymbol{\Pi}_{\mathcal{P}} = \boldsymbol{V}_{\mathcal{P}}\boldsymbol{V}_{\mathcal{P}}^{\top}$ gives $\boldsymbol{\Pi}_{\mathcal{P}}\widehat{\boldsymbol{w}} = \boldsymbol{0}$. □

Consequently, if $\mathbf{0} \neq \widehat{\boldsymbol{w}} \in \mathcal{R} \cap \widehat{\mathcal{F}}_{\mathcal{S}}$, contracting $\mathcal{S}$ gives $\|\widehat{\boldsymbol{w}} - \boldsymbol{\Pi}_{\mathcal{P}}\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau} = 1$. Hence, an $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similar coarsening with $\varepsilon_{\mathcal{P}} < 1$ cannot both preserve this direction and contract $\mathcal{S}$. Thus, a suitable target should capture the complete-group direction while having little overlap with the internal fluctuation space. We use confusability to quantify this overlap.

Theorem. *Let $\tau > 0$, and let $\mathcal{P}$ be the partition of a degree-weighted block coarsening that is $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similar and contracts the proper group $\mathcal{S}$ into one supernode. Write $\chi = \chi^{\tau}_{\mathcal{R}}(\mathcal{S})$. If $\chi > 0$, then*

$$\varepsilon_{\mathcal{P}} \geq \left[\sqrt{\chi} - \sqrt{(\mu^{\tau}_{\mathcal{P}})^2 - 1}\,\sqrt{1-\chi}\right]_+ . \tag{30}$$

Proof. Choose $\widehat{\boldsymbol{w}} \in \widehat{\mathcal{F}}^{\tau}_{\mathcal{S}}$ attaining the maximum in the definition of confusability:

$$\|\boldsymbol{Q}^{\tau}_{\mathcal{R}}\widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} = \chi.$$

The maximum is attained because $\widehat{\mathcal{F}}^{\tau}_{\mathcal{S}}$ is compact. Since $\chi > 0$, the vector

$$\boldsymbol{y} := \frac{\boldsymbol{Q}^{\tau}_{\mathcal{R}}\widehat{\boldsymbol{w}}}{\sqrt{\chi}}$$

is well defined, belongs to $\mathcal{R}$, and has $\|\boldsymbol{y}\|_{\boldsymbol{M}_\tau} = 1$. Screened orthogonality of $\boldsymbol{Q}^{\tau}_{\mathcal{R}}$ also gives

$$\langle \widehat{\boldsymbol{w}}, \boldsymbol{y}\rangle_{\boldsymbol{M}_\tau} = \frac{\langle \widehat{\boldsymbol{w}}, \boldsymbol{Q}^{\tau}_{\mathcal{R}}\widehat{\boldsymbol{w}}\rangle_{\boldsymbol{M}_\tau}}{\sqrt{\chi}} = \sqrt{\chi}. \tag{31}$$

Set $T := \boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}}$. For $\tau > 0$, define the adjoint of T in the screened inner product by

$$T^{*_\tau} := \boldsymbol{M}_\tau^{-1} T^{\top} \boldsymbol{M}_\tau.$$

This is the screened analogue of a matrix transpose: for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$,

$$\langle \boldsymbol{x}, T\boldsymbol{y}\rangle_{\boldsymbol{M}_\tau} = \langle T^{*_\tau}\boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{M}_\tau}. \tag{32}$$

We first bound $T^{*_\tau}\widehat{\boldsymbol{w}}$. The characterization of a vector norm by inner products, followed by Eq. 32, gives

$$\begin{aligned}
\|T^{*_\tau}\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau} &= \max_{\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}=1} \langle T^{*_\tau}\widehat{\boldsymbol{w}}, \boldsymbol{x}\rangle_{\boldsymbol{M}_\tau} \\
&= \max_{\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}=1} \langle \widehat{\boldsymbol{w}}, T\boldsymbol{x}\rangle_{\boldsymbol{M}_\tau} \\
&\leq \max_{\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}=1} \|\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau}\,\|T\boldsymbol{x}\|_{\boldsymbol{M}_\tau} \\
&\leq \mu^{\tau}_{\mathcal{P}}.
\end{aligned}$$

The first inequality is Cauchy–Schwarz. The last uses $\|\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau} = 1$ and the definition of $\mu^{\tau}_{\mathcal{P}}$ in Eq. 5.

Next, Lemma 13 gives $\boldsymbol{\Pi}_{\mathcal{P}}\widehat{\boldsymbol{w}} = \mathbf{0}$ and hence $T\widehat{\boldsymbol{w}} = \widehat{\boldsymbol{w}}$. Applying Eq. 32 with both arguments equal to $\widehat{\boldsymbol{w}}$ yields

$$\langle T^{*_\tau}\widehat{\boldsymbol{w}}, \widehat{\boldsymbol{w}}\rangle_{\boldsymbol{M}_\tau} = \langle \widehat{\boldsymbol{w}}, T\widehat{\boldsymbol{w}}\rangle_{\boldsymbol{M}_\tau} = \|\widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} = 1.$$

Therefore,

$$\langle T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}, \widehat{\boldsymbol{w}}\rangle_{\boldsymbol{M}_\tau} = 1 - 1 = 0.$$

In other words, the difference $T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}$ is screened-orthogonal to $\widehat{\boldsymbol{w}}$. Pythagoras now gives

$$\begin{aligned}
\|T^{*_\tau}\widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} &= \|\widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} + \|T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} \\
&= 1 + \|T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau}.
\end{aligned} \tag{33}$$

Since the left-hand side is at most $(\mu^{\tau}_{\mathcal{P}})^2$, we conclude that

$$\|T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}\|^2_{\boldsymbol{M}_\tau} \leq (\mu^{\tau}_{\mathcal{P}})^2 - 1. \tag{34}$$

By Eq. 31, $\langle \widehat{\boldsymbol{w}}, \boldsymbol{y}\rangle_{\boldsymbol{M}_\tau} = \sqrt{\chi}$. Since $\|\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau} = 1$, subtracting the component of $\boldsymbol{y}$ along $\widehat{\boldsymbol{w}}$ gives

$$\left\langle \widehat{\boldsymbol{w}}, \boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}}\right\rangle_{\boldsymbol{M}_\tau} = \sqrt{\chi} - \sqrt{\chi} = 0.$$

Thus $\boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}}$ is screened-orthogonal to $\widehat{\boldsymbol{w}}$. Applying Pythagoras to $\boldsymbol{y} = \sqrt{\chi}\,\widehat{\boldsymbol{w}} + (\boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}})$ and using $\|\boldsymbol{y}\|_{\boldsymbol{M}_\tau} = 1$ gives

$$\|\boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau}^2 = \|\boldsymbol{y}\|_{\boldsymbol{M}_\tau}^2 - \chi = 1 - \chi. \tag{35}$$

We established above that $T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}$ is also screened-orthogonal to $\widehat{\boldsymbol{w}}$. Expand both vectors in the following inner product:

$$\begin{aligned}\langle T^{*_\tau}\widehat{\boldsymbol{w}}, \boldsymbol{y}\rangle_{\boldsymbol{M}_\tau} &= \big\langle \widehat{\boldsymbol{w}} + (T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}), \sqrt{\chi}\,\widehat{\boldsymbol{w}} + (\boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}})\big\rangle_{\boldsymbol{M}_\tau} \\ &= \sqrt{\chi} + \big\langle T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}, \boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}}\big\rangle_{\boldsymbol{M}_\tau}.\end{aligned}$$

The other two cross terms vanish by the orthogonality just established, and $\langle \widehat{\boldsymbol{w}}, \widehat{\boldsymbol{w}}\rangle_{\boldsymbol{M}_\tau} = 1$. Cauchy–Schwarz bounds the remaining cross term from below:

$$\begin{aligned}\big\langle T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}, \boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}}\big\rangle_{\boldsymbol{M}_\tau} &\geq -\|T^{*_\tau}\widehat{\boldsymbol{w}} - \widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau}\|\boldsymbol{y} - \sqrt{\chi}\,\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau} \\ &\geq -\sqrt{(\mu_{\mathcal{P}}^\tau)^2 - 1}\,\sqrt{1-\chi},\end{aligned}$$

where the last line uses Eq. 34 and Eq. 35. Hence

$$\langle T^{*_\tau}\widehat{\boldsymbol{w}}, \boldsymbol{y}\rangle_{\boldsymbol{M}_\tau} \geq \sqrt{\chi} - \sqrt{(\mu_{\mathcal{P}}^\tau)^2 - 1}\sqrt{1-\chi}. \tag{36}$$

To relate this inner product to coarsening error, first apply Cauchy–Schwarz and $\|\widehat{\boldsymbol{w}}\|_{\boldsymbol{M}_\tau} = 1$:

$$\langle \widehat{\boldsymbol{w}}, T\boldsymbol{y}\rangle_{\boldsymbol{M}_\tau} \leq \|T\boldsymbol{y}\|_{\boldsymbol{M}_\tau}.$$

The definition of the adjoint makes the left-hand side equal to $\langle T^{*_\tau}\widehat{\boldsymbol{w}}, \boldsymbol{y}\rangle_{\boldsymbol{M}_\tau}$. Moreover, $\boldsymbol{y} \in \mathcal{R}$ and $\|\boldsymbol{y}\|_{\boldsymbol{M}_\tau} = 1$, so $(\mathcal{R}, \varepsilon_{\mathcal{P}}, \tau)$-similarity gives $\|T\boldsymbol{y}\|_{\boldsymbol{M}_\tau} \leq \varepsilon_{\mathcal{P}}$. Combining these facts with Eq. 36 yields

$$\varepsilon_{\mathcal{P}} \geq \sqrt{\chi} - \sqrt{(\mu_{\mathcal{P}}^\tau)^2 - 1}\sqrt{1-\chi}.$$

Finally, $\varepsilon_{\mathcal{P}} \geq 0$, so the same inequality holds with the right-hand side replaced by its positive part. □

Proposition. *Let $\tau > 0$ and let $\mathcal{S}$ be a proper group with $|\mathcal{S}| \geq 2$. Define the screened dual of $\bar{\boldsymbol{v}}_{\mathcal{S}}$ by*

$$\bar{\boldsymbol{v}}_{\mathcal{S}}^\star := \boldsymbol{M}_\tau^{-1}\bar{\boldsymbol{v}}_{\mathcal{S}}, \qquad \mathcal{R}_\chi^\star(\mathcal{S}) := \operatorname{span}\{\bar{\boldsymbol{v}}_{\mathcal{S}}^\star\}.$$

Then

$$\mathcal{R}_\chi^\star(\mathcal{S}) \in \arg\min_{\mathcal{R}} \chi_{\mathcal{R}}^\tau(\mathcal{S}), \qquad \chi_{\mathcal{R}_\chi^\star}^\tau(\mathcal{S}) = 0. \tag{37}$$

Proof. The vectors $\bar{\boldsymbol{v}}_{\mathcal{S}}$ and $\widehat{\boldsymbol{v}}_{\mathcal{S}}$ differ by a nonzero scalar. Thus, for every $\widehat{\boldsymbol{w}} \in \widehat{\mathcal{F}}_{\mathcal{S}}^\tau$, the defining Euclidean orthogonality of the internal fluctuation space gives

$$\widehat{\boldsymbol{w}}^\top \bar{\boldsymbol{v}}_{\mathcal{S}} = 0.$$

By the definition of the screened dual, $\boldsymbol{M}_\tau \bar{\boldsymbol{v}}_{\mathcal{S}}^\star = \bar{\boldsymbol{v}}_{\mathcal{S}}$. Consequently,

$$\begin{aligned}\langle \widehat{\boldsymbol{w}}, \bar{\boldsymbol{v}}_{\mathcal{S}}^\star\rangle_{\boldsymbol{M}_\tau} &= \widehat{\boldsymbol{w}}^\top \boldsymbol{M}_\tau \bar{\boldsymbol{v}}_{\mathcal{S}}^\star \\ &= \widehat{\boldsymbol{w}}^\top \bar{\boldsymbol{v}}_{\mathcal{S}} = 0.\end{aligned}$$

Hence every internal fluctuation is screened-orthogonal to $\mathcal{R}_\chi^\star(\mathcal{S})$. Its screened orthogonal projection onto this subspace is therefore zero:

$$\boldsymbol{Q}_{\mathcal{R}_\chi^\star(\mathcal{S})}^\tau \widehat{\boldsymbol{w}} = \boldsymbol{0} \qquad \text{for every } \widehat{\boldsymbol{w}} \in \widehat{\mathcal{F}}_{\mathcal{S}}^\tau.$$

Taking the maximum in the definition of confusability yields

$$\chi_{\mathcal{R}_\chi^\star}^\tau(\mathcal{S}) = \max_{\widehat{\boldsymbol{w}} \in \widehat{\mathcal{F}}_{\mathcal{S}}^\tau} \Big\| \boldsymbol{Q}_{\mathcal{R}_\chi^\star(\mathcal{S})}^\tau \widehat{\boldsymbol{w}} \Big\|_{\boldsymbol{M}_\tau}^2 = 0.$$

Confusability is nonnegative for every subspace $\mathcal{R}$, so zero is its minimum possible value. Therefore $\mathcal{R}_\chi^\star(\mathcal{S})$ is a minimizer. □

B STATIC EIGENSPACES: STRUCTURAL CONDITIONS FOR GROUP RECOVERY

We specialize the target-subspace analysis to

$$\mathcal{R}_K = \operatorname{span}\{\boldsymbol{u}_0, \dots, \boldsymbol{u}_{K-1}\}, \qquad 2 \le K < N.$$

Throughout, the graph is connected, $\tau > 0$, and all degrees and volumes include the self-loops used in the main text. Eigenvalues are indexed from zero, so λ_{K-1} is the largest retained eigenvalue and λ_K is the first omitted eigenvalue. At a repeated cutoff eigenvalue, the selected eigenvectors specify the subspace; exact quantities can depend on that selection. We write

$$C_K(\mathcal{S}) = C^{\tau}_{\mathcal{R}_K}(\mathcal{S}), \qquad \chi_K(\mathcal{S}) = \chi^{\tau}_{\mathcal{R}_K}(\mathcal{S}), \qquad a_{\mathcal{S}} = \frac{\Phi(\mathcal{S})}{1 - r_{\mathcal{S}}}.$$

The quantity $a_{\mathcal{S}}$ is the Laplacian energy of the centered unit indicator $\bar{\boldsymbol{v}}_{\mathcal{S}}$. A useful static target must retain this complete-group contrast while excluding modes that resolve internal divisions. We first quantify these two requirements, then give guarantees for the actual screened, deflated Ward score.

B.1 CAPTURE, CONDUCTANCE, AND BOUNDARY IRREGULARITY

Let $p_j = \boldsymbol{u}_j^{\top} \widehat{\boldsymbol{v}}_{\mathcal{S}}$. Because the indicator is centered,

$$p_0 = 0, \qquad \sum_{j=1}^{N-1} p_j^2 = \frac{1}{a_{\mathcal{S}} + \tau}, \qquad \sum_{j=1}^{N-1} \lambda_j p_j^2 = \frac{a_{\mathcal{S}}}{a_{\mathcal{S}} + \tau}.$$

The spectral projector commutes with $\boldsymbol{M}_\tau$, and therefore

$$\boldsymbol{Q}^{\tau}_{\mathcal{R}_K} = \boldsymbol{U}_K \boldsymbol{U}_K^{\top}, \qquad C_K(\mathcal{S}) = \sum_{j=1}^{K-1} (\lambda_j + \tau) p_j^2. \tag{38}$$

Thus, the projector itself is independent of τ, but the fraction of screened energy it captures generally is not. To describe boundary irregularity, put $\mathcal{T} = \mathcal{V} \setminus \mathcal{S}$ and define the normalized boundary degrees

$$\rho_i = \frac{d_{\partial}(i)}{\widetilde{d}_i} \quad (i \in \mathcal{S}), \qquad q_j = \frac{b_{\mathcal{S}}(j)}{\widetilde{d}_j} \quad (j \in \mathcal{T}).$$

Their degree-weighted means are $\Phi(\mathcal{S})$ and $\Phi(\mathcal{T})$, respectively. Define

$$\sigma^2_{\partial,\mathcal{S}} := \frac{\sum_{i \in \mathcal{S}} \widetilde{d}_i (\rho_i - \Phi(\mathcal{S}))^2 + \sum_{j \in \mathcal{T}} \widetilde{d}_j (q_j - \Phi(\mathcal{T}))^2}{\operatorname{vol}(\mathcal{S})(1 - r_{\mathcal{S}})}. \tag{39}$$

This statistic includes all complement nodes, including those with $b_{\mathcal{S}}(j) = 0$. It measures irregularity of the boundary relative to the full two-block partition, rather than only among boundary endpoints.

***Proposition* 14** (Structural capture bounds)**.** *The boundary statistic satisfies the exact identities*

$$\sigma^2_{\partial,\mathcal{S}} = \|(\boldsymbol{L} - a_{\mathcal{S}} \boldsymbol{I}) \bar{\boldsymbol{v}}_{\mathcal{S}}\|_2^2 = (a_{\mathcal{S}} + \tau) \sum_{j=1}^{N-1} (\lambda_j - a_{\mathcal{S}})^2 p_j^2. \tag{40}$$

Moreover,

$$1 - C_K(\mathcal{S}) \le \frac{a_{\mathcal{S}}(\lambda_K + \tau)}{\lambda_K (a_{\mathcal{S}} + \tau)}. \tag{41}$$

If $a_{\mathcal{S}} < \lambda_K$, then also

$$1 - C_K(\mathcal{S}) \le \frac{(\lambda_K + \tau)\sigma^2_{\partial,\mathcal{S}}}{(a_{\mathcal{S}} + \tau)(\lambda_K - a_{\mathcal{S}})^2}. \tag{42}$$

Proof. Set $b = \sqrt{\operatorname{vol}(\mathcal{S})(1 - r_{\mathcal{S}})}$. On $\mathcal{S}$, the residual $(\boldsymbol{L} - a_{\mathcal{S}} \boldsymbol{I}) \bar{\boldsymbol{v}}_{\mathcal{S}}$ equals $\sqrt{\widetilde{d}_i}(\rho_i - \Phi(\mathcal{S}))/b$. On $\mathcal{T}$, it equals $-\sqrt{\widetilde{d}_j}(q_j - \Phi(\mathcal{T}))/b$. Squaring proves the first identity; expansion in the eigenbasis proves the second. Equation Eq. 38 gives

$$1 - C_K = \sum_{j \ge K} (\lambda_j + \tau) p_j^2.$$

For $j \geq K$, $\lambda_j + \tau \leq (1 + \tau/\lambda_K)\lambda_j$, which proves Eq. 41. For the second bound, the function $t \mapsto (t+\tau)/(t-a_{\mathcal{S}})^2$ is decreasing for $t > a_{\mathcal{S}}$. Apply this inequality termwise to the same tail and use Eq. 40. □

Low conductance places the mean indicator frequency below the cutoff. Small boundary variance additionally concentrates its spectral mass near that mean. In particular, if $\sigma_{\partial,\mathcal{S}} = 0$, the centered indicator is an exact eigenvector with eigenvalue $a_{\mathcal{S}}$; it is fully captured whenever $a_{\mathcal{S}} < \lambda_K$.

B.2 Confusability and internal connectivity

Assume $|\mathcal{S}| \geq 2$. Define the internal Poincare constant using the original augmented degrees,

$$\gamma_{\mathcal{S}} := \min_{\substack{f \neq 0 \\ \sum_{i\in\mathcal{S}} \widetilde{d}_i f_i = 0}} \frac{\sum_{\{i,j\}\in\mathcal{E}_{\mathrm{int}}(\mathcal{S})} W_{ij}(f_i - f_j)^2}{\sum_{i\in\mathcal{S}} \widetilde{d}_i f_i^2}. \tag{43}$$

It is positive when the induced group is connected. This is an internal spectral gap with the degree measure of the full graph; it is not the normalized-Laplacian gap obtained by renormalizing to internal degrees. Large $\gamma_{\mathcal{S}}$ means that every nonconstant internal signal pays substantial edge energy. Also set $\rho_{\min,\mathcal{S}} = \min_{i\in\mathcal{S}} \rho_i$.

***Proposition* 15** (Static confusability)**.** *The exact specialization is*

$$\chi_K(\mathcal{S}) = \max_{\widehat{\boldsymbol{w}}\in\widehat{\mathcal{F}}_{\mathcal{S}}^{\tau}} \sum_{j=1}^{K-1} (\lambda_j + \tau)\big(\boldsymbol{u}_j^{\top}\widehat{\boldsymbol{w}}\big)^2. \tag{44}$$

Moreover, it satisfies

$$\chi_K(\mathcal{S}) \leq X_{K,\mathcal{S}} := \min\left\{1, \frac{\lambda_{K-1}+\tau}{\gamma_{\mathcal{S}} + \rho_{\min,\mathcal{S}} + \tau} \min\left\{1, \frac{\lambda_{K-1}}{\gamma_{\mathcal{S}}}\right\}\right\}. \tag{45}$$

In particular, when $\lambda_{K-1} < \gamma_{\mathcal{S}}$,

$$\chi_K(\mathcal{S}) \leq \frac{\lambda_{K-1}(\lambda_{K-1}+\tau)}{\gamma_{\mathcal{S}}(\gamma_{\mathcal{S}}+\tau)}. \tag{46}$$

Proof. Every $\boldsymbol{w} \in \mathcal{F}_{\mathcal{S}}$ is Euclidean orthogonal to $\boldsymbol{u}_0$, giving Eq. 44. Let $\boldsymbol{P}_F$ denote the Euclidean projector onto $\mathcal{F}_{\mathcal{S}}$ and $\boldsymbol{P}_K = \boldsymbol{U}_K\boldsymbol{U}_K^{\top}$. For any $\boldsymbol{y} \in \mathcal{R}_K$, the Poincare inequality gives

$$\gamma_{\mathcal{S}}\|\boldsymbol{P}_F\boldsymbol{y}\|_2^2 \leq \sum_{\{i,j\}\in\mathcal{E}_{\mathrm{int}}(\mathcal{S})} W_{ij}\left(\frac{y_i}{\sqrt{\widetilde{d}_i}} - \frac{y_j}{\sqrt{\widetilde{d}_j}}\right)^2 \leq \boldsymbol{y}^{\top}\boldsymbol{L}\boldsymbol{y} \leq \lambda_{K-1}\|\boldsymbol{y}\|_2^2.$$

Transposing the product of orthogonal projectors therefore yields $\|\boldsymbol{P}_K\boldsymbol{w}\|_2^2 \leq \min\{1, \lambda_{K-1}/\gamma_{\mathcal{S}}\}\|\boldsymbol{w}\|_2^2$. For a fluctuation supported on $\mathcal{S}$, boundary edges contribute $\sum_{i\in\mathcal{S}} \rho_i w_i^2$, so

$$\boldsymbol{w}^{\top}\boldsymbol{M}_{\tau}\boldsymbol{w} \geq (\gamma_{\mathcal{S}} + \rho_{\min,\mathcal{S}} + \tau)\|\boldsymbol{w}\|_2^2.$$

Bound the numerator of Eq. 44 by $(\lambda_{K-1}+\tau)\|\boldsymbol{P}_K\boldsymbol{w}\|_2^2$ to conclude. □

For a nested choice of eigenvectors, both C_K and χ_K are nondecreasing in K. Increasing the cutoff improves complete-group capture but can also retain internal divisions. At $K = N$, both equal one for a nonsingleton group, and every Ward merge has score one. At $K = 1$, all merge scores vanish because every partition preserves $\boldsymbol{u}_0$. Thus, neither endpoint provides a discriminatory merge score.

B.3 From static eigenmaps to Ward decisions

The screened response associated with $\mathcal{R}_K$ has coordinates

$$\ell_{K,j}(i) = \sqrt{\lambda_j + \tau}\,\frac{u_j(i)}{\sqrt{\widetilde{d}_i}}, \qquad 0 \le j < K. \tag{47}$$

Write

$$\psi_j(i) = \frac{u_j(i)}{\sqrt{\widetilde{d}_i}}, \qquad \bar{\psi}_{j,A} = \frac{1}{\mathrm{vol}(A)} \sum_{i \in A} \widetilde{d}_i \psi_j(i).$$

For adjacent blocks A and B, let $\bar{\boldsymbol{h}}_{A,B}$ denote their $\boldsymbol{M}_\tau$-normalized block contrast. The corresponding raw Ward score is

$$s_K^0(A,B) = \frac{\sum_{j=1}^{K-1} (\lambda_j + \tau)\big(\bar{\psi}_{j,A} - \bar{\psi}_{j,B}\big)^2}{\|\boldsymbol{h}_{A,B}\|^2_{\boldsymbol{M}_\tau}}. \tag{48}$$

Its denominator is determined by the two blocks and their boundary:

$$\begin{aligned}\|\boldsymbol{h}_{A,B}\|^2_{\boldsymbol{M}_\tau} = \tau\left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)}\right) + \frac{\mathrm{cut}(A)}{\mathrm{vol}(A)^2} + \frac{\mathrm{cut}(B)}{\mathrm{vol}(B)^2} \\ + \frac{2W(A,B)}{\mathrm{vol}(A)\,\mathrm{vol}(B)}, \qquad W(A,B) = \sum_{i \in A, j \in B} W_{ij}.\end{aligned} \tag{49}$$

Equation Eq. 48 gives the raw score a direct interpretation: it is the normalized separation between the degree-weighted block means in the retained eigenmap. Two blocks are inexpensive to merge when the retained eigenvectors give them similar mean responses. In particular, if $A, B \subseteq \mathcal{S}$, then their normalized contrast is an internal fluctuation, and hence

$$s_K^0(A,B) \le \chi_K(\mathcal{S}) \le X_{K,\mathcal{S}}. \tag{50}$$

Thus, the confusability bound in Proposition 15 is also a uniform upper bound on every raw merge internal to $\mathcal{S}$.

The algorithm uses the deflated rather than the raw contrast. Given the current partition $\mathcal{P}$ and the post-merge partition $\mathcal{P}' = \mathcal{P} \oplus (A,B)$, define

$$\widehat{\boldsymbol{h}}_{A,B} = \frac{(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}'})\bar{\boldsymbol{h}}_{A,B}}{\|(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}'})\bar{\boldsymbol{h}}_{A,B}\|_{\boldsymbol{M}_\tau}}.$$

The actual score is

$$s_{\mathcal{P},K}(A,B) = \sum_{j=1}^{K-1} (\lambda_j + \tau)\big(\boldsymbol{u}_j^\top \widehat{\boldsymbol{h}}_{A,B}\big)^2. \tag{51}$$

The constant eigenvector contributes nothing because it belongs to every post-merge block space.

To quantify the effect of deflation, let

$$\eta_{\mathcal{P}}(A,B) = \left|\left\langle \bar{\boldsymbol{h}}_{A,B}, \widehat{\boldsymbol{h}}_{A,B} \right\rangle_{\boldsymbol{M}_\tau}\right|^2. \tag{52}$$

The raw-to-deflated comparison in Proposition 9 shows that the raw score remains informative whenever this alignment is bounded away from zero. In particular, using $\eta_{\mathcal{P}}(A,B) \ge \eta_\star := \bar{\mu}_\tau^{-2}$ and Eq. 50, every internal candidate at an $\mathcal{S}$-clean partition satisfies

$$s_{\mathcal{P},K}(A,B) \le U_{K,\mathcal{S}} := 1 - \left[\sqrt{\eta_\star(1 - X_{K,\mathcal{S}})} - \sqrt{(1-\eta_\star)X_{K,\mathcal{S}}}\right]_+^2. \tag{53}$$

The conclusion is operational: internal connectivity controls $X_{K,\mathcal{S}}$, which controls raw internal scores, while $\eta_\star$ quantifies the worst loss caused by deflation. Therefore, choosing the retained frequencies below the group's internal variation scale keeps all merges needed to complete the group uniformly inexpensive in the actual algorithm—not only in the raw eigenmap.

C Learning the Subspace

C.1 Chebyshev Filter Bank and Target Construction

Sparse filter evaluation. The eigenvalues of $\widehat{\boldsymbol{A}} = \boldsymbol{I} - \boldsymbol{L}$ lie in $(-1, 1]$, so Chebyshev polynomials can be applied to $\widehat{\boldsymbol{A}}$ without rescaling. They satisfy

$$T_0(t) = 1, \qquad T_1(t) = t, \qquad T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t).$$

Consequently, all propagated feature matrices needed by Eq. 14 can be computed through

$$\boldsymbol{Y}_0 = \boldsymbol{X}, \qquad \boldsymbol{Y}_1 = \widehat{\boldsymbol{A}}\boldsymbol{X}, \qquad \boldsymbol{Y}_{k+1} = 2\widehat{\boldsymbol{A}}\boldsymbol{Y}_k - \boldsymbol{Y}_{k-1}. \tag{54}$$

Filter h then produces

$$g_{\boldsymbol{\theta}_h}(\widehat{\boldsymbol{A}})\boldsymbol{X} = \sum_{k=0}^{K} \theta_{h,k}\boldsymbol{Y}_k.$$

The recurrence requires sparse matrix–feature products; neither the eigenvectors nor dense powers of $\widehat{\boldsymbol{A}}$ are constructed. Since $\boldsymbol{Y}_k$ is a degree-k polynomial of $\widehat{\boldsymbol{A}}$ applied to $\boldsymbol{X}$, its row at vertex v depends only on input features within k hops of v. Thus each output row depends on at most a K-hop neighborhood.

Spectral action. Writing $\widehat{\boldsymbol{A}} = \boldsymbol{U}(\boldsymbol{I} - \boldsymbol{\Lambda})\boldsymbol{U}^\top$ gives

$$g_{\boldsymbol{\theta}_h}(\widehat{\boldsymbol{A}})\boldsymbol{X} = \boldsymbol{U}\ \mathrm{diag}\big(g_{\boldsymbol{\theta}_h}(1-\lambda_0), \ldots, g_{\boldsymbol{\theta}_h}(1-\lambda_{N-1})\big)\,\boldsymbol{U}^\top\boldsymbol{X}. \tag{55}$$

The learned coefficients therefore determine a spectral response shared across the d input channels. Different filters can assign different gains to the same graph frequencies.

***Lemma* 16** (Polynomial-basis equivalence)**.** *For every degree K and input matrix $\boldsymbol{X}$,*

$$\mathrm{span}\{T_k(\widehat{\boldsymbol{A}})\boldsymbol{x}_a : 0 \le k \le K,\ 1 \le a \le d\} = \mathrm{span}\{\widehat{\boldsymbol{A}}^k\boldsymbol{x}_a : 0 \le k \le K,\ 1 \le a \le d\}.$$

For every Chebyshev filter bank there is a monomial filter bank producing exactly the same $\boldsymbol{Z}_\Theta$, and conversely.

Proof. The polynomials $T_0, \ldots, T_K$ form a basis of the polynomials of degree at most K: T_k has degree k and a nonzero leading coefficient. Their change-of-basis matrix with $1, t, \ldots, t^K$ is therefore triangular and invertible. Applying either polynomial basis to $\widehat{\boldsymbol{A}}$ and then to each $\boldsymbol{x}_a$ preserves this change of basis. Applying the corresponding coefficient transformation separately to each filter gives identical filtered output matrices. □

The lemma concerns the set of subspaces attainable by the model. A numerical coefficient vector generally changes its filter when moved between the two bases without the corresponding change-of-basis transformation. Chebyshev polynomials satisfy $|T_k(t)| \le 1$ on $[-1, 1]$, which helps control the magnitude of individual basis functions; the conditioning of a particular filtered-feature Gram matrix still depends on the graph, features, and coefficients.

Constructing the screened target. The bank output has $q = Hd$ columns, although its rank can be smaller than q. With

$$\boldsymbol{G}_\Theta := \boldsymbol{Z}_\Theta^\top \boldsymbol{M}_\tau \boldsymbol{Z}_\Theta,$$

the screened-orthogonal projector onto the constructed target is

$$\boldsymbol{Q}^\tau_{\mathcal{R}_\Theta} = \boldsymbol{Z}_\Theta \boldsymbol{G}_\Theta^\dagger \boldsymbol{Z}_\Theta^\top \boldsymbol{M}_\tau. \tag{56}$$

The corresponding Gram-normalized screened response used in the capture, confusability, and merge-score formulas is

$$\boldsymbol{\ell}_\Theta = \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau \boldsymbol{Z}_\Theta (\boldsymbol{G}_\Theta^\dagger)^{1/2}. \tag{57}$$

The pseudoinverse accommodates dependent or zero output columns without changing the subspace $\mathcal{R}_\Theta = \mathrm{range}(\boldsymbol{Z}_\Theta)$.

Cost and transfer. With a sparse graph, evaluating Eq. 54 costs $O(K(|\mathcal{E}| + N)d)$. Combining its outputs for all H filters costs $O(H(K+1)Nd)$; the recurrence can be streamed while retaining the H output accumulators. These costs describe construction of $\boldsymbol{Z}_\Theta$. If the screened response is needed, forming $\boldsymbol{M}_\tau \boldsymbol{Z}_\Theta$ adds $O((|\mathcal{E}| + N)q)$, forming $\boldsymbol{G}_\Theta$ adds $O(Nq^2)$, and a dense decomposition of this $q \times q$ Gram matrix adds $O(q^3)$.

For a new graph, the same learned Θ is applied to its own normalized adjacency and input matrix. This constructs a graph-specific $\mathcal{R}_\Theta$ without carrying eigenvectors or node identities from the training graphs. The input channel count d must remain fixed; when random channels replace node features, their distribution plays the role of the shared input specification.

C.2 Capture and confusability through the screened response

***Theorem* 17** (Capture and confusability through the screened response)**.** *Let $\tau > 0$, let $\mathcal{S}$ be a proper group with $|\mathcal{S}| \geq 2$, and let $\mathcal{R} = \mathrm{range}(\boldsymbol{Z})$. Define*

$$\boldsymbol{G} := \boldsymbol{Z}^\top \boldsymbol{M}_\tau \boldsymbol{Z}, \qquad \boldsymbol{\ell}_\mathcal{R} := \widetilde{\boldsymbol{D}}^{-1/2} \boldsymbol{M}_\tau \boldsymbol{Z} (\boldsymbol{G}^\dagger)^{1/2}.$$

For a nonempty set A, write

$$\bar{\boldsymbol{\ell}}_A := \frac{1}{\mathrm{vol}(A)} \sum_{v \in A} \widetilde{d}_v \boldsymbol{\ell}_{\mathcal{R},v,:}^\top,$$

and define its within-group response covariance by

$$\boldsymbol{\Sigma}_\mathcal{S}(\boldsymbol{\ell}_\mathcal{R}) := \sum_{v \in \mathcal{S}} \widetilde{d}_v (\boldsymbol{\ell}_{\mathcal{R},v,:}^\top - \bar{\boldsymbol{\ell}}_\mathcal{S})(\boldsymbol{\ell}_{\mathcal{R},v,:}^\top - \bar{\boldsymbol{\ell}}_\mathcal{S})^\top.$$

Then capture has the exact response representation

$$C_\mathcal{R}^\tau(\mathcal{S}) = \frac{\mathrm{vol}(\mathcal{S})}{\Phi(\mathcal{S}) + \tau(1 - r_\mathcal{S})} \|\bar{\boldsymbol{\ell}}_\mathcal{S} - \bar{\boldsymbol{\ell}}_\mathcal{V}\|_2^2. \tag{58}$$

Confusability has the exact variational representation

$$\chi_\mathcal{R}^\tau(\mathcal{S}) = \max_{\substack{\boldsymbol{f} \neq \boldsymbol{0},\ \mathrm{supp}(\boldsymbol{f}) \subseteq \mathcal{S} \\ \sum_{v \in \mathcal{S}} \widetilde{d}_v f_v = 0}} \frac{\left\| \sum_{v \in \mathcal{S}} \widetilde{d}_v f_v (\boldsymbol{\ell}_{\mathcal{R},v,:}^\top - \bar{\boldsymbol{\ell}}_\mathcal{S}) \right\|_2^2}{\|\widetilde{\boldsymbol{D}}^{1/2} \boldsymbol{f}\|_{\boldsymbol{M}_\tau}^2}. \tag{59}$$

In particular, because the graph is connected,

$$\chi_\mathcal{R}^\tau(\mathcal{S}) \leq \min\left\{1,\ \frac{\lambda_{\max}(\boldsymbol{\Sigma}_\mathcal{S}(\boldsymbol{\ell}_\mathcal{R}))}{\tau + \lambda_1}\right\} \leq \frac{\lambda_{\max}(\boldsymbol{\Sigma}_\mathcal{S}(\boldsymbol{\ell}_\mathcal{R}))}{\tau + \lambda_1}. \tag{60}$$

Moreover, $\chi_\mathcal{R}^\tau(\mathcal{S}) = 0$ if and only if $\boldsymbol{\ell}_{\mathcal{R},v,:}$ is identical for all $v \in \mathcal{S}$.

For the static eigenspace $\mathcal{R}_K = \mathrm{span}(\boldsymbol{U}_K)$, let $\boldsymbol{\Omega}_K = \mathrm{diag}(\lambda_0 + \tau, \ldots, \lambda_{K-1} + \tau)$. Its screened response and capture specialize to

$$\boldsymbol{\ell}_K = \widetilde{\boldsymbol{D}}^{-1/2} \boldsymbol{U}_K \boldsymbol{\Omega}_K^{1/2}, \tag{61}$$

$$C_{\mathcal{R}_K}^\tau(\mathcal{S}) = \frac{\sum_{k=1}^{K-1} (\lambda_k + \tau)(\boldsymbol{u}_k^\top \boldsymbol{v}_\mathcal{S})^2}{\Phi(\mathcal{S}) + \tau(1 - r_\mathcal{S})}. \tag{62}$$

If $\boldsymbol{F}_\mathcal{S}$ has Euclidean-orthonormal columns spanning $\mathcal{F}_\mathcal{S}$, its confusability is exactly

$$\chi_{\mathcal{R}_K}^\tau(\mathcal{S}) = \lambda_{\max}\Big((\boldsymbol{F}_\mathcal{S}^\top \boldsymbol{M}_\tau \boldsymbol{F}_\mathcal{S})^{-1/2} \boldsymbol{F}_\mathcal{S}^\top \boldsymbol{U}_K \boldsymbol{\Omega}_K \boldsymbol{U}_K^\top \boldsymbol{F}_\mathcal{S} (\boldsymbol{F}_\mathcal{S}^\top \boldsymbol{M}_\tau \boldsymbol{F}_\mathcal{S})^{-1/2} \Big). \tag{63}$$

Proof. The screened projector onto $\mathcal{R}$ is $\boldsymbol{Q}_\mathcal{R}^\tau = \boldsymbol{Z} \boldsymbol{G}^\dagger \boldsymbol{Z}^\top \boldsymbol{M}_\tau$. Hence, for any $\boldsymbol{x}$,

$$\|\boldsymbol{Q}_\mathcal{R}^\tau \boldsymbol{x}\|_{\boldsymbol{M}_\tau}^2 = \|(\boldsymbol{G}^\dagger)^{1/2} \boldsymbol{Z}^\top \boldsymbol{M}_\tau \boldsymbol{x}\|_2^2. \tag{64}$$

For capture, use

$$\bar{\boldsymbol{v}}_\mathcal{S} = \frac{\widetilde{\boldsymbol{D}}^{1/2}(\boldsymbol{1}_\mathcal{S} - r_\mathcal{S} \boldsymbol{1})}{\sqrt{\mathrm{vol}(\mathcal{S})(1 - r_\mathcal{S})}}.$$

The definition of $\boldsymbol{\ell}_{\mathcal{R}}$ gives

$$(\boldsymbol{G}^{\dagger})^{1/2}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\tau}\bar{\boldsymbol{v}}_{\mathcal{S}} = \frac{\sum_{v\in\mathcal{S}}\widetilde{d}_v\boldsymbol{\ell}_{\mathcal{R},v,:}^{\top} - r_{\mathcal{S}}\sum_{v\in\mathcal{V}}\widetilde{d}_v\boldsymbol{\ell}_{\mathcal{R},v,:}^{\top}}{\sqrt{\mathrm{vol}(\mathcal{S})(1-r_{\mathcal{S}})}} = \sqrt{\frac{\mathrm{vol}(\mathcal{S})}{1-r_{\mathcal{S}}}}(\bar{\boldsymbol{\ell}}_{\mathcal{S}} - \bar{\boldsymbol{\ell}}_{\mathcal{V}}).$$

Since

$$\|\bar{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}^2 = \frac{\Phi(\mathcal{S})}{1-r_{\mathcal{S}}} + \tau = \frac{\Phi(\mathcal{S}) + \tau(1-r_{\mathcal{S}})}{1-r_{\mathcal{S}}},$$

substituting $\widehat{\boldsymbol{v}}_{\mathcal{S}} = \bar{\boldsymbol{v}}_{\mathcal{S}}/\|\bar{\boldsymbol{v}}_{\mathcal{S}}\|_{\boldsymbol{M}_\tau}$ into Eq. 64 proves Eq. 58.

For confusability, every $\boldsymbol{w} \in \mathcal{F}_{\mathcal{S}}$ can be written as $\boldsymbol{w} = \widetilde{\boldsymbol{D}}^{1/2}\boldsymbol{f}$, where $\boldsymbol{f}$ is supported on $\mathcal{S}$. The orthogonality defining $\mathcal{F}_{\mathcal{S}}$ is equivalent to $\sum_{v\in\mathcal{S}}\widetilde{d}_v f_v = 0$. Applying Eq. 64 to $\boldsymbol{w}$ therefore gives

$$\|\boldsymbol{Q}_{\mathcal{R}}^{\tau}\boldsymbol{w}\|_{\boldsymbol{M}_\tau}^2 = \left\|\sum_{v\in\mathcal{S}}\widetilde{d}_v f_v \boldsymbol{\ell}_{\mathcal{R},v,:}^{\top}\right\|_2^2 = \left\|\sum_{v\in\mathcal{S}}\widetilde{d}_v f_v (\boldsymbol{\ell}_{\mathcal{R},v,:}^{\top} - \bar{\boldsymbol{\ell}}_{\mathcal{S}})\right\|_2^2.$$

Dividing by $\|\boldsymbol{w}\|_{\boldsymbol{M}_\tau}^2$ and maximizing over nonzero internal fluctuations proves Eq. 59.

For the covariance bound, form a matrix whose row indexed by $v \in \mathcal{S}$ is $\sqrt{\widetilde{d}_v}(\boldsymbol{\ell}_{\mathcal{R},v,:} - \bar{\boldsymbol{\ell}}_{\mathcal{S}}^{\top})$. Its column Gram matrix is $\boldsymbol{\Sigma}_{\mathcal{S}}(\boldsymbol{\ell}_{\mathcal{R}})$. Therefore the numerator in Eq. 59 is at most

$$\lambda_{\max}(\boldsymbol{\Sigma}_{\mathcal{S}}(\boldsymbol{\ell}_{\mathcal{R}}))\sum_{v\in\mathcal{S}}\widetilde{d}_v f_v^2 = \lambda_{\max}(\boldsymbol{\Sigma}_{\mathcal{S}}(\boldsymbol{\ell}_{\mathcal{R}}))\|\boldsymbol{w}\|_2^2.$$

An internal fluctuation is orthogonal to $\boldsymbol{u}_0$: its support and defining zero-mean condition give $\boldsymbol{u}_0^{\top}\boldsymbol{w} = 0$. Thus

$$\|\boldsymbol{w}\|_{\boldsymbol{M}_\tau}^2 \geq (\tau + \lambda_1)\|\boldsymbol{w}\|_2^2.$$

Together with $\chi_{\mathcal{R}}^{\tau}(\mathcal{S}) \leq 1$, this proves Eq. 60.

If the response is constant on $\mathcal{S}$, the numerator in Eq. 59 vanishes for every admissible $\boldsymbol{f}$, so $\chi = 0$. Conversely, if two vertices $i, j \in \mathcal{S}$ have different response rows, choose $f_i = 1/\widetilde{d}_i$, $f_j = -1/\widetilde{d}_j$, and set all other entries to zero. This is admissible and makes the numerator nonzero. Hence $\chi > 0$.

Finally, an $\boldsymbol{M}_\tau$-orthonormal basis of $\mathcal{R}_K$ is $\boldsymbol{U}_K\boldsymbol{\Omega}_K^{-1/2}$. Substituting it into the response definition proves Eq. 61. The centered indicator has coefficient zero on $\boldsymbol{u}_0$ and coefficient $(\boldsymbol{u}_k^{\top}\boldsymbol{v}_{\mathcal{S}})/\sqrt{1-r_{\mathcal{S}}}$ on $\boldsymbol{u}_k$ for $k \geq 1$. Its screened projection therefore gives Eq. 62. For $\boldsymbol{w} = \boldsymbol{F}_{\mathcal{S}}\boldsymbol{a}$, spectral truncation gives

$$\|\boldsymbol{Q}_{\mathcal{R}_K}^{\tau}\boldsymbol{w}\|_{\boldsymbol{M}_\tau}^2 = \boldsymbol{a}^{\top}\boldsymbol{F}_{\mathcal{S}}^{\top}\boldsymbol{U}_K\boldsymbol{\Omega}_K\boldsymbol{U}_K^{\top}\boldsymbol{F}_{\mathcal{S}}\boldsymbol{a},$$

whereas $\|\boldsymbol{w}\|_{\boldsymbol{M}_\tau}^2 = \boldsymbol{a}^{\top}\boldsymbol{F}_{\mathcal{S}}^{\top}\boldsymbol{M}_\tau\boldsymbol{F}_{\mathcal{S}}\boldsymbol{a}$. Maximizing this generalized Rayleigh quotient proves Eq. 63. □

***Proposition* 18** (Screened-dual target and its response)**.** *Let $\mathcal{S}_1, \ldots, \mathcal{S}_m$ be pairwise disjoint proper groups. Define*

$$\bar{\boldsymbol{V}}_{\star} := [\bar{\boldsymbol{v}}_{\mathcal{S}_1}, \ldots, \bar{\boldsymbol{v}}_{\mathcal{S}_m}], \qquad \boldsymbol{Z}^{\star} := \boldsymbol{M}_\tau^{-1}\bar{\boldsymbol{V}}_{\star}, \qquad \mathcal{R}^{\star} := \mathrm{range}(\boldsymbol{Z}^{\star}).$$

Set

$$\boldsymbol{G}^{\star} := \bar{\boldsymbol{V}}_{\star}^{\top}\boldsymbol{M}_\tau^{-1}\bar{\boldsymbol{V}}_{\star}, \qquad \boldsymbol{H}^{\star} := \bar{\boldsymbol{V}}_{\star}^{\top}\bar{\boldsymbol{V}}_{\star}.$$

The unwhitened screened response is

$$\widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau\boldsymbol{Z}^{\star} = \widetilde{\boldsymbol{D}}^{-1/2}\bar{\boldsymbol{V}}_{\star},$$

and the Gram-normalized response is

$$\boldsymbol{\ell}^\star = \widetilde{\boldsymbol{D}}^{-1/2}\bar{\boldsymbol{V}}_\star\big((\boldsymbol{G}^\star)^\dagger\big)^{1/2}. \tag{65}$$

In particular, every row of $\boldsymbol{\ell}^\star$ within a fixed group $\mathcal{S}_j$ is identical. Consequently,

$$\chi^\tau_{\mathcal{R}^\star}(\mathcal{S}_j) = 0 \qquad (j = 1, \dots, m). \tag{66}$$

Its capture of $\mathcal{S}_j$ is

$$C^\tau_{\mathcal{R}^\star}(\mathcal{S}_j) = \frac{[\boldsymbol{H}^\star(\boldsymbol{G}^\star)^\dagger \boldsymbol{H}^\star]_{jj}}{\Phi(\mathcal{S}_j)/(1 - r_{\mathcal{S}_j}) + \tau}. \tag{67}$$

Proof. The identity $\boldsymbol{M}_\tau \boldsymbol{Z}^\star = \bar{\boldsymbol{V}}_\star$ immediately gives both response formulas. More explicitly, before Gram normalization, the entry at vertex v in column j is

$$\Big(\widetilde{\boldsymbol{D}}^{-1/2}\bar{\boldsymbol{v}}_{\mathcal{S}_j}\Big)_v = \frac{\mathbf{1}_{\mathcal{S}_j}(v) - r_{\mathcal{S}_j}}{\sqrt{\mathrm{vol}(\mathcal{S}_j)(1 - r_{\mathcal{S}_j})}}.$$

Because the groups are disjoint, this entry is constant as v ranges over any one group $\mathcal{S}_i$. Right multiplication by the fixed matrix $\big((\boldsymbol{G}^\star)^\dagger\big)^{1/2}$ preserves that property. Theorem 17 therefore gives Eq. 66.

For capture, apply Eq. 64 with $\boldsymbol{Z} = \boldsymbol{Z}^\star$ and $\widehat{\boldsymbol{v}}_{\mathcal{S}_j} = \bar{\boldsymbol{v}}_{\mathcal{S}_j}/\|\bar{\boldsymbol{v}}_{\mathcal{S}_j}\|_{\boldsymbol{M}_\tau}$. Since $(\boldsymbol{Z}^\star)^\top \boldsymbol{M}_\tau = \bar{\boldsymbol{V}}_\star^\top$, its numerator is

$$\bar{\boldsymbol{v}}_{\mathcal{S}_j}^\top \bar{\boldsymbol{V}}_\star (\boldsymbol{G}^\star)^\dagger \bar{\boldsymbol{V}}_\star^\top \bar{\boldsymbol{v}}_{\mathcal{S}_j} = [\boldsymbol{H}^\star(\boldsymbol{G}^\star)^\dagger \boldsymbol{H}^\star]_{jj}.$$

Dividing by $\|\bar{\boldsymbol{v}}_{\mathcal{S}_j}\|^2_{\boldsymbol{M}_\tau} = \Phi(\mathcal{S}_j)/(1 - r_{\mathcal{S}_j}) + \tau$ proves Eq. 67. □

C.3 Closed-form optimization of the edge surrogate

The exact edge objective in the main text is defined through the jointly Gram-normalized response of the complete multi-channel embedding. This normalization makes the objective a nonlinear trace quotient of the filter coefficients and, in general, prevents a closed-form solution. Our goal in this section is therefore not to claim a closed-form optimizer of the exact multi-channel objective. Instead, we introduce a channel-pooled surrogate that preserves its supervised edge contrasts, explain precisely how the two objectives differ, and show that the surrogate reduces to a generalized eigenvalue problem with an exact global solution.

Throughout this section, we fix one graph and suppress graph indices. Let $\mathcal{S}_1, \dots, \mathcal{S}_m$ be its training groups, each with positive internal and boundary weight.

C.3.1 Exact objective and the source of nonlinearity

The objective from the main text is

$$\mathcal{J}_{\mathrm{edge}}(\Theta) = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{w_\partial(\mathcal{S}_j)}\sum_{(u,v)\in\partial\mathcal{S}_j} W_{uv} d^2_\Theta(u,v) - \frac{\gamma}{w_{\mathrm{int}}(\mathcal{S}_j)}\sum_{(u,v)\in\mathcal{E}_{\mathrm{int}}(\mathcal{S}_j)} W_{uv} d^2_\Theta(u,v)\right], \tag{68}$$

where

$$d^2_\Theta(u,v) := \|\boldsymbol{\ell}_{\Theta,u,:} - \boldsymbol{\ell}_{\Theta,v,:}\|_2^2. \tag{69}$$

The first term rewards response separation across group boundaries, while the second penalizes response variation within a group.

For an edge (u, v), define the degree-normalized contrast direction

$$\boldsymbol{q}_{uv} := \frac{\boldsymbol{e}_u}{\sqrt{\widetilde{d}_u}} - \frac{\boldsymbol{e}_v}{\sqrt{\widetilde{d}_v}}. \tag{70}$$

Recall that the complete embedding, its screened Gram matrix, and its normalized screened response are

$$\boldsymbol{Z}_\Theta, \qquad \boldsymbol{G}_\Theta := \boldsymbol{Z}_\Theta^\top \boldsymbol{M}_\tau \boldsymbol{Z}_\Theta, \qquad \boldsymbol{\ell}_\Theta := \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau \boldsymbol{Z}_\Theta (\boldsymbol{G}_\Theta^\dagger)^{1/2}. \tag{71}$$

The difference between response rows u and v is therefore

$$\boldsymbol{\ell}_{\Theta,u,:} - \boldsymbol{\ell}_{\Theta,v,:} = \boldsymbol{q}_{uv}^\top \boldsymbol{M}_\tau \boldsymbol{Z}_\Theta (\boldsymbol{G}_\Theta^\dagger)^{1/2}, \tag{72}$$

and hence

$$d_\Theta^2(u,v) = \boldsymbol{q}_{uv}^\top \boldsymbol{M}_\tau \boldsymbol{Z}_\Theta \boldsymbol{G}_\Theta^\dagger \boldsymbol{Z}_\Theta^\top \boldsymbol{M}_\tau \boldsymbol{q}_{uv}. \tag{73}$$

Collect the boundary rewards and internal penalties in the symmetric signed edge-contrast matrix

$$\boldsymbol{C}_\gamma := \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{w_\partial(\mathcal{S}_j)}\sum_{(u,v)\in\partial\mathcal{S}_j} W_{uv}\boldsymbol{q}_{uv}\boldsymbol{q}_{uv}^\top - \frac{\gamma}{w_{\mathrm{int}}(\mathcal{S}_j)}\sum_{(u,v)\in\mathcal{E}_{\mathrm{int}}(\mathcal{S}_j)} W_{uv}\boldsymbol{q}_{uv}\boldsymbol{q}_{uv}^\top\right]. \tag{74}$$

Substituting Eq. 73 into Eq. 68 and using cyclic invariance of the trace gives the exact matrix representation

$$\mathcal{J}_{\mathrm{edge}}(\Theta) = \mathrm{tr}\Big[\boldsymbol{G}_\Theta^\dagger \boldsymbol{Z}_\Theta^\top \boldsymbol{M}_\tau \boldsymbol{C}_\gamma \boldsymbol{M}_\tau \boldsymbol{Z}_\Theta\Big]. \tag{75}$$

To see why this expression is not generally quadratic in Θ, write $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d]$ and define the Chebyshev dictionary of input channel c by

$$\boldsymbol{\Phi}_c := \big[\boldsymbol{x}_c, T_1(\widehat{\boldsymbol{A}})\boldsymbol{x}_c, \ldots, T_K(\widehat{\boldsymbol{A}})\boldsymbol{x}_c\big] \in \mathbb{R}^{N\times(K+1)}. \tag{76}$$

For $\Theta = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_H]$, the output of head h on channel c is

$$g_{\boldsymbol{\theta}_h}(\widehat{\boldsymbol{A}})\boldsymbol{x}_c = \boldsymbol{\Phi}_c\boldsymbol{\theta}_h. \tag{77}$$

Up to a permutation of columns, the complete embedding is therefore

$$\boldsymbol{Z}_\Theta = \big[\boldsymbol{\Phi}_1\Theta, \ldots, \boldsymbol{\Phi}_d\Theta\big]. \tag{78}$$

Consequently, $\boldsymbol{G}_\Theta$ contains all cross-channel blocks

$$\Theta^\top \boldsymbol{\Phi}_c^\top \boldsymbol{M}_\tau \boldsymbol{\Phi}_{c'}\Theta, \qquad c, c' = 1, \ldots, d. \tag{79}$$

These blocks are coupled through the pseudoinverse $\boldsymbol{G}_\Theta^\dagger$ in Eq. 75. Thus, even though $\boldsymbol{Z}_\Theta$ is linear in Θ, the exact objective is not a quadratic form in Θ.

C.3.2 The surrogate response and its relation to the exact response

To obtain a tractable objective, we retain the same filtered signals and the same boundary and internal edge comparisons, but normalize the shared coefficient directions through their average screened energy across input channels. For each channel, define the raw screened response

$$\widetilde{\boldsymbol{\ell}}_{\Theta,c} := \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau\boldsymbol{\Phi}_c\Theta \in \mathbb{R}^{N\times H}. \tag{80}$$

We concatenate these responses and scale by $1/\sqrt{d}$:

$$\boldsymbol{\ell}_\Theta^{\mathrm{sur}} := \frac{1}{\sqrt{d}}\big[\widetilde{\boldsymbol{\ell}}_{\Theta,1}, \ldots, \widetilde{\boldsymbol{\ell}}_{\Theta,d}\big] \in \mathbb{R}^{N\times dH}. \tag{81}$$

The associated surrogate edge distance is

$$\begin{aligned}\big(d_\Theta^{\mathrm{sur}}(u,v)\big)^2 &:= \Big\|\boldsymbol{\ell}_{\Theta,u,:}^{\mathrm{sur}} - \boldsymbol{\ell}_{\Theta,v,:}^{\mathrm{sur}}\Big\|_2^2 \\ &= \frac{1}{d}\sum_{c=1}^{d}\Big\|\widetilde{\boldsymbol{\ell}}_{\Theta,c,u,:} - \widetilde{\boldsymbol{\ell}}_{\Theta,c,v,:}\Big\|_2^2 \\ &= \frac{1}{d}\sum_{c=1}^{d}\sum_{h=1}^{H}\Big(\boldsymbol{q}_{uv}^\top\boldsymbol{M}_\tau\boldsymbol{\Phi}_c\boldsymbol{\theta}_h\Big)^2.\end{aligned} \tag{82}$$

We define the surrogate by replacing the exact distance in Eq. 68 with this channel-pooled distance:

$$\begin{aligned}\mathcal{J}_{\mathrm{sur}}(\Theta) := \frac{1}{m}\sum_{j=1}^{m}\Bigg[&\frac{1}{w_\partial(\mathcal{S}_j)}\sum_{(u,v)\in\partial\mathcal{S}_j} W_{uv}\big(d_\Theta^{\mathrm{sur}}(u,v)\big)^2 \\ &- \frac{\gamma}{w_{\mathrm{int}}(\mathcal{S}_j)}\sum_{(u,v)\in\mathcal{E}_{\mathrm{int}}(\mathcal{S}_j)} W_{uv}\big(d_\Theta^{\mathrm{sur}}(u,v)\big)^2\Bigg].\end{aligned} \tag{83}$$

The exact and surrogate objectives therefore have identical boundary-reward-minus-internal-penalty forms. Their difference is the response used to measure an edge:

$$\begin{aligned}
&\text{exact:} && \boldsymbol{\ell}_\Theta = \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau \boldsymbol{Z}_\Theta (\boldsymbol{G}_\Theta^\dagger)^{1/2}, && d_\Theta^2(u,v) = \|\boldsymbol{\ell}_{\Theta,u,:} - \boldsymbol{\ell}_{\Theta,v,:}\|_2^2,\\
&\text{surrogate:} && \boldsymbol{\ell}_\Theta^{\mathrm{sur}} = \frac{1}{\sqrt{d}}\big[\widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau \boldsymbol{\Phi}_1\Theta, \ldots, \widetilde{\boldsymbol{D}}^{-1/2}\boldsymbol{M}_\tau \boldsymbol{\Phi}_d\Theta\big], && \big(d_\Theta^{\mathrm{sur}}(u,v)\big)^2 = \|\boldsymbol{\ell}_{\Theta,u,:}^{\mathrm{sur}} - \boldsymbol{\ell}_{\Theta,v,:}^{\mathrm{sur}}\|_2^2.
\end{aligned} \tag{84}$$

Why this is a useful surrogate. The substitution above is limited to the normalization step. It retains the following properties of the exact objective.

1. It uses the same Chebyshev-filtered signals $\boldsymbol{\Phi}_c\boldsymbol{\theta}_h$, the same training groups, the same weighted boundary and internal edges, and the same trade-off parameter γ.
2. It preserves the intended local geometry: increasing a boundary response difference increases the objective, whereas increasing an internal response difference decreases it. Thus it trains the same type of block-level response contrasts that determine the raw Ward scores discussed in the main text, while using the pooled normalization above.
3. It averages squared response differences across channels. Consequently, it is invariant to an orthogonal change of basis in the input channels; the value does not depend on an arbitrary rotation of the feature coordinates.
4. Its scale is controlled by the pooled screened-energy constraint derived below. The constraint also prevents all heads from selecting the same coefficient direction.
5. For a single input channel, the surrogate becomes an exact reparameterization of the rank-H objective after screened-energy normalization, as shown below.

The surrogate is not claimed to be a universal upper or lower bound on $\mathcal{J}_{\mathrm{edge}}$ when $d > 1$. Its benefit is that it preserves the supervised edge signal while removing the joint Gram normalization that prevents closed-form optimization.

C.3.3 Derivation of the quadratic trace objective

We now derive the coefficient-space form of $\mathcal{J}_{\mathrm{sur}}$. For channel c and head h, the difference of the raw screened response across (u, v) is

$$\widetilde{\boldsymbol{\ell}}_{\Theta,c,u,h} - \widetilde{\boldsymbol{\ell}}_{\Theta,c,v,h} = \boldsymbol{q}_{uv}^\top \boldsymbol{M}_\tau \boldsymbol{\Phi}_c \boldsymbol{\theta}_h. \tag{85}$$

Using the definition of $\boldsymbol{C}_\gamma$ in Eq. 74, the signed edge contrast of this single channel-head pair is

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}\Bigg[\frac{1}{w_\partial(\mathcal{S}_j)}\sum_{(u,v)\in\partial\mathcal{S}_j} W_{uv}\left(\boldsymbol{q}_{uv}^\top \boldsymbol{M}_\tau \boldsymbol{\Phi}_c\boldsymbol{\theta}_h\right)^2 \\
&\qquad\qquad - \frac{\gamma}{w_{\mathrm{int}}(\mathcal{S}_j)}\sum_{(u,v)\in\mathcal{E}_{\mathrm{int}}(\mathcal{S}_j)} W_{uv}\left(\boldsymbol{q}_{uv}^\top \boldsymbol{M}_\tau \boldsymbol{\Phi}_c\boldsymbol{\theta}_h\right)^2\Bigg] \\
&= \boldsymbol{\theta}_h^\top \boldsymbol{\Phi}_c^\top \boldsymbol{M}_\tau \boldsymbol{C}_\gamma \boldsymbol{M}_\tau \boldsymbol{\Phi}_c \boldsymbol{\theta}_h.
\end{aligned} \tag{86}$$

This motivates the channel-specific contrast and energy matrices

$$\boldsymbol{A}_c := \boldsymbol{\Phi}_c^\top \boldsymbol{M}_\tau \boldsymbol{C}_\gamma \boldsymbol{M}_\tau \boldsymbol{\Phi}_c, \qquad \boldsymbol{S}_c := \boldsymbol{\Phi}_c^\top \boldsymbol{M}_\tau \boldsymbol{\Phi}_c. \tag{87}$$

They have the direct interpretations

$$\boldsymbol{\theta}_h^\top \boldsymbol{A}_c \boldsymbol{\theta}_h = \text{signed edge contrast of channel } c \text{ and head } h, \qquad \boldsymbol{\theta}_h^\top \boldsymbol{S}_c \boldsymbol{\theta}_h = \|\boldsymbol{\Phi}_c\boldsymbol{\theta}_h\|_{\boldsymbol{M}_\tau}^2. \tag{88}$$

Average these matrices over channels:

$$\bar{\boldsymbol{A}} := \frac{1}{d}\sum_{c=1}^{d}\boldsymbol{A}_c, \qquad \bar{\boldsymbol{S}} := \frac{1}{d}\sum_{c=1}^{d}\boldsymbol{S}_c. \tag{89}$$

Expanding the surrogate distance in Eq. 83, and then summing over channels and heads, gives

$$\begin{aligned}\mathcal{J}_{\mathrm{sur}}(\Theta) &= \frac{1}{d}\sum_{c=1}^{d}\sum_{h=1}^{H}\boldsymbol{\theta}_h^\top \boldsymbol{A}_c\boldsymbol{\theta}_h \\ &= \sum_{h=1}^{H}\boldsymbol{\theta}_h^\top \bar{\boldsymbol{A}}\boldsymbol{\theta}_h \\ &= \boxed{\mathrm{tr}(\Theta^\top \bar{\boldsymbol{A}}\Theta)}.\end{aligned} \tag{90}$$

This is the desired quadratic representation of the surrogate objective.

The quadratic objective is not scale invariant: if it has a positive direction, rescaling that direction makes its value arbitrarily large. We therefore normalize the filters by their average screened energy. From Eq. 87,

$$\begin{aligned}\frac{1}{d}\sum_{c=1}^{d}(\boldsymbol{\Phi}_c\Theta)^\top \boldsymbol{M}_\tau(\boldsymbol{\Phi}_c\Theta) &= \frac{1}{d}\sum_{c=1}^{d}\Theta^\top \boldsymbol{S}_c\Theta \\ &= \Theta^\top \bar{\boldsymbol{S}}\Theta.\end{aligned} \tag{91}$$

Requiring this matrix to equal $\boldsymbol{I}_H$ gives every head unit average screened energy and makes different heads orthogonal in the pooled energy geometry. The resulting optimization problem is

$$\boxed{\begin{array}{ll}\underset{\Theta\in\mathbb{R}^{(K+1)\times H}}{\text{maximize}} & \mathcal{J}_{\mathrm{sur}}(\Theta) = \mathrm{tr}(\Theta^\top \bar{\boldsymbol{A}}\Theta) \\ \text{subject to} & \Theta^\top \bar{\boldsymbol{S}}\Theta = \boldsymbol{I}_H.\end{array}} \tag{92}$$

For completeness, when $d = 1$ we have $\boldsymbol{Z}_\Theta = \boldsymbol{\Phi}_1\Theta$ and therefore

$$\mathcal{J}_{\mathrm{edge}}(\Theta) = \mathrm{tr}\Big[(\Theta^\top \boldsymbol{S}_1\Theta)^\dagger \Theta^\top \boldsymbol{A}_1\Theta\Big]. \tag{93}$$

Under the normalization $\Theta^\top \boldsymbol{S}_1\Theta = \boldsymbol{I}_H$, this reduces to $\mathrm{tr}(\Theta^\top \boldsymbol{A}_1\Theta)$. Thus, for one channel, Eq. 92 is an exact reparameterization rather than a surrogate approximation.

***Theorem* 19** (Closed-form solution of the pooled surrogate)**.** *Let $p := \mathrm{rank}(\bar{\boldsymbol{S}})$ and $1 \le H \le p$. Write the positive-support eigendecomposition of $\bar{\boldsymbol{S}}$ as*

$$\bar{\boldsymbol{S}} = \boldsymbol{U}_+\boldsymbol{D}_+\boldsymbol{U}_+^\top, \qquad \boldsymbol{D}_+ \succ \boldsymbol{0}, \qquad \boldsymbol{W} := \boldsymbol{U}_+\boldsymbol{D}_+^{-1/2}. \tag{94}$$

Let

$$\boldsymbol{T} := \boldsymbol{W}^\top \bar{\boldsymbol{A}}\boldsymbol{W}, \qquad \boldsymbol{T}\boldsymbol{v}_h = \lambda_h\boldsymbol{v}_h, \qquad \lambda_1 \ge \cdots \ge \lambda_p, \tag{95}$$

where $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are Euclidean-orthonormal. Then

$$\Theta^\star = \boldsymbol{W}[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_H] \tag{96}$$

is a global maximizer of Eq. 92, and

$$\max_{\Theta^\top \bar{\boldsymbol{S}}\Theta = \boldsymbol{I}_H} \mathcal{J}_{\mathrm{sur}}(\Theta) = \sum_{h=1}^{H}\lambda_h. \tag{97}$$

Equivalently, its columns satisfy

$$\bar{\boldsymbol{A}}\boldsymbol{\theta}_h^\star = \lambda_h\bar{\boldsymbol{S}}\boldsymbol{\theta}_h^\star, \qquad (\boldsymbol{\theta}_h^\star)^\top \bar{\boldsymbol{S}}\boldsymbol{\theta}_{h'}^\star = \delta_{hh'}. \tag{98}$$

If the number of heads may be chosen subject to an upper bound H, one optimal choice retains the eigenvectors associated with the positive eigenvalues among $\lambda_1, \ldots, \lambda_H$. If $H < p$ and $\lambda_H > \lambda_{H+1}$, the optimal coefficient subspace is unique, although its basis is determined only up to an orthogonal transformation.

Proof. We first handle a possibly singular $\bar{\boldsymbol{S}}$. Let $\boldsymbol{a} \in \ker(\bar{\boldsymbol{S}})$. By Eq. 89,

$$0 = \boldsymbol{a}^\top \bar{\boldsymbol{S}} \boldsymbol{a} = \frac{1}{d} \sum_{c=1}^{d} \|\boldsymbol{\Phi}_c \boldsymbol{a}\|_{\boldsymbol{M}_\tau}^2. \tag{99}$$

Every term is nonnegative and $\boldsymbol{M}_\tau \succ \boldsymbol{0}$, so $\boldsymbol{\Phi}_c \boldsymbol{a} = \boldsymbol{0}$ for every channel c. It follows from Eq. 87 that

$$\boldsymbol{A}_c \boldsymbol{a} = \boldsymbol{\Phi}_c^\top \boldsymbol{M}_\tau \boldsymbol{C}_\gamma \boldsymbol{M}_\tau \boldsymbol{\Phi}_c \boldsymbol{a} = \boldsymbol{0}, \tag{100}$$

and hence

$$\ker(\bar{\boldsymbol{S}}) \subseteq \ker(\bar{\boldsymbol{A}}). \tag{101}$$

Thus, components of Θ in $\ker(\bar{\boldsymbol{S}})$ affect neither the constraint nor the objective. We may restrict to the canonical representatives satisfying $\Theta \in \mathrm{range}(\bar{\boldsymbol{S}})$.

The matrix $\boldsymbol{W}$ in Eq. 94 spans this range and satisfies

$$\boldsymbol{W}^\top \bar{\boldsymbol{S}} \boldsymbol{W} = \boldsymbol{D}_+^{-1/2} \boldsymbol{U}_+^\top (\boldsymbol{U}_+ \boldsymbol{D}_+ \boldsymbol{U}_+^\top) \boldsymbol{U}_+ \boldsymbol{D}_+^{-1/2} = \boldsymbol{I}_p. \tag{102}$$

Every canonical coefficient bank can be written uniquely as $\Theta = \boldsymbol{W}\boldsymbol{V}$, where $\boldsymbol{V} \in \mathbb{R}^{p \times H}$. Under this change of variables,

$$\Theta^\top \bar{\boldsymbol{S}} \Theta = \boldsymbol{I}_H \quad \Longleftrightarrow \quad \boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_H, \tag{103}$$

and

$$\mathcal{J}_{\mathrm{sur}}(\Theta) = \mathrm{tr}\left(\boldsymbol{V}^\top \boldsymbol{W}^\top \bar{\boldsymbol{A}} \boldsymbol{W} \boldsymbol{V}\right) = \mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{T} \boldsymbol{V}). \tag{104}$$

Using the eigendecomposition $\boldsymbol{T} = \sum_{i=1}^{p} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$, any feasible $\boldsymbol{V}$ satisfies

$$\mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{T} \boldsymbol{V}) = \sum_{i=1}^{p} \lambda_i \|\boldsymbol{V}^\top \boldsymbol{v}_i\|_2^2. \tag{105}$$

Set $\alpha_i := \|\boldsymbol{V}^\top \boldsymbol{v}_i\|_2^2$. Since $\boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_H$, these weights satisfy

$$0 \le \alpha_i \le 1, \qquad \sum_{i=1}^{p} \alpha_i = \mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{V}) = H. \tag{106}$$

Because the eigenvalues are in decreasing order,

$$\sum_{i=1}^{p} \lambda_i \alpha_i \le \sum_{h=1}^{H} \lambda_h. \tag{107}$$

Equality is attained by $\boldsymbol{V}^\star = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_H]$. Therefore $\Theta^\star = \boldsymbol{W}\boldsymbol{V}^\star$ is globally optimal and has the value in Eq. 97. This is the Ky Fan maximum principle specialized to the present problem.

To obtain the generalized eigenvalue form, note that $\bar{\boldsymbol{A}}$ is symmetric and Eq. 101 implies $\mathrm{range}(\bar{\boldsymbol{A}}) \subseteq \mathrm{range}(\bar{\boldsymbol{S}})$. Hence $\boldsymbol{U}_+ \boldsymbol{U}_+^\top \bar{\boldsymbol{A}} = \bar{\boldsymbol{A}}$. Starting from

$$\boldsymbol{D}_+^{-1/2} \boldsymbol{U}_+^\top \bar{\boldsymbol{A}} \boldsymbol{U}_+ \boldsymbol{D}_+^{-1/2} \boldsymbol{v}_h = \lambda_h \boldsymbol{v}_h \tag{108}$$

and multiplying by $\boldsymbol{U}_+ \boldsymbol{D}_+^{1/2}$ gives

$$\begin{aligned} \bar{\boldsymbol{A}} \boldsymbol{U}_+ \boldsymbol{D}_+^{-1/2} \boldsymbol{v}_h &= \lambda_h \boldsymbol{U}_+ \boldsymbol{D}_+^{1/2} \boldsymbol{v}_h \\ &= \lambda_h \bar{\boldsymbol{S}} \boldsymbol{U}_+ \boldsymbol{D}_+^{-1/2} \boldsymbol{v}_h. \end{aligned} \tag{109}$$

Since $\boldsymbol{\theta}_h^\star = \boldsymbol{W}\boldsymbol{v}_h = \boldsymbol{U}_+ \boldsymbol{D}_+^{-1/2} \boldsymbol{v}_h$, this proves the generalized eigenvalue relation in Eq. 98; its normalization follows from Eq. 102.

For a fixed rank r, the same argument gives the optimal value $\sum_{h=1}^{r} \lambda_h$. If r may vary up to H, including a positive eigenvalue increases the objective, including a zero eigenvalue leaves it unchanged, and including a negative eigenvalue decreases it. Thus, one optimal variable-rank choice retains precisely the positive modes up to the head budget. Finally, the strict gap $\lambda_H > \lambda_{H+1}$ makes the leading H-dimensional eigenspace of $\boldsymbol{T}$ unique; applying the injective map $\boldsymbol{W}$ proves the corresponding uniqueness of the coefficient subspace. □

Algorithm 1 Closed-form learning of the channel-pooled Chebyshev filter bank

Require: Weighted adjacency $\boldsymbol{W}$, augmented degree matrix $\widetilde{\boldsymbol{D}}$, normalized adjacency $\widehat{\boldsymbol{A}}$, features $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_d]$, training groups $\{\mathcal{S}_j\}_{j=1}^m$, filter degree K, head budget H, screening $\tau > 0$, and internal penalty $\gamma \geq 0$

Ensure: Filter coefficients $\Theta^\star$ and the training-graph target $\mathcal{R}_{\Theta^\star}$

Form $\boldsymbol{M}_\tau = \boldsymbol{I} - \widehat{\boldsymbol{A}} + \tau \boldsymbol{I}$
Compute $\boldsymbol{T}_0 = \boldsymbol{X}$, $\boldsymbol{T}_1 = \widehat{\boldsymbol{A}}\boldsymbol{X}$, and $\boldsymbol{T}_{k+1} = 2\widehat{\boldsymbol{A}}\boldsymbol{T}_k - \boldsymbol{T}_{k-1}$ for $k = 1, \dots, K-1$
for $c = 1, \dots, d$ **do**
    Form the channel dictionary $\boldsymbol{\Phi}_c = [\boldsymbol{T}_0(:, c), \dots, \boldsymbol{T}_K(:, c)]$
    Compute the contrast matrix $\boldsymbol{A}_c$ and energy matrix $\boldsymbol{S}_c$ using Eq. 87
end for
Pool the channel matrices: $\bar{\boldsymbol{A}} = d^{-1} \sum_{c=1}^d \boldsymbol{A}_c$ and $\bar{\boldsymbol{S}} = d^{-1} \sum_{c=1}^d \boldsymbol{S}_c$
Compute the positive-support decomposition $\bar{\boldsymbol{S}} = \boldsymbol{U}_+ \boldsymbol{D}_+ \boldsymbol{U}_+^\top$ and set $\boldsymbol{W}_{\mathrm{pool}} = \boldsymbol{U}_+ \boldsymbol{D}_+^{-1/2}$
Solve $\boldsymbol{W}_{\mathrm{pool}}^\top \bar{\boldsymbol{A}} \boldsymbol{W}_{\mathrm{pool}} \boldsymbol{v}_h = \lambda_h \boldsymbol{v}_h$ with $\lambda_1 \geq \cdots \geq \lambda_p$ and $\boldsymbol{v}_h^\top \boldsymbol{v}_{h'} = \delta_{hh'}$
Set $r = \min\{H, \#\{h : \lambda_h > 0\}\}$
if $r = 0$ **then**
    return no positive-gain filter bank
end if
Set $\Theta^\star = \boldsymbol{W}_{\mathrm{pool}}[\boldsymbol{v}_1, \dots, \boldsymbol{v}_r]$
Form $\boldsymbol{Z}_{\Theta^\star} = [\boldsymbol{\Phi}_1 \Theta^\star, \dots, \boldsymbol{\Phi}_d \Theta^\star]$
return $\Theta^\star$ and $\mathcal{R}_{\Theta^\star} = \mathrm{range}(\boldsymbol{Z}_{\Theta^\star})$

D Proofs of Adapting Ward Clustering for Group Detection

This section state the entire framework algorithm and proves the results in Section 6. We retain the notation of the main text throughout. In particular, all orthogonality, projections, and norms below are with respect to the screened inner product induced by $\boldsymbol{M}_\tau$, and the projector onto the block-constant space of a partition $\mathcal{P}$ is $\boldsymbol{Q}_{\mathcal{P}}^\tau$.

D.1 Algorithm

The stopping budget B_{stop} may be selected on training groups and transferred unchanged, or after the same graph-size normalization used in the experiments. If a merge-score threshold is preferred, replace the cumulative-budget test by $s^\star \geq t_{\mathrm{stop}}$; this is the stopping rule used by the score-gap recovery certificate.

D.2 Proof of Proposition 7

Proposition. *For every partition $\mathcal{P}_n$,*

$$\varepsilon_{\mathcal{P}_n}^Q \leq \varepsilon_{\mathcal{P}_n} \leq \mu_{\mathcal{P}_n}^\tau \varepsilon_{\mathcal{P}_n}^Q , \tag{110}$$

where $\mu_{\mathcal{P}_n}^\tau$ is defined in Eq. 5.

Proof. Fix a partition $\mathcal{P}_n$ and a vector $\boldsymbol{x}$. Add and subtract $\boldsymbol{Q}_{\mathcal{P}_n}^\tau \boldsymbol{x}$ to obtain

$$(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{P}_n}^\tau)\boldsymbol{x} + (\boldsymbol{Q}_{\mathcal{P}_n}^\tau - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{x}. \tag{111}$$

The first term is $\boldsymbol{M}_\tau$-orthogonal to the block-constant space $\mathcal{A}_{\mathcal{P}_n}$. The second term belongs to $\mathcal{A}_{\mathcal{P}_n}$ because both $\boldsymbol{Q}_{\mathcal{P}_n}^\tau \boldsymbol{x}$ and $\boldsymbol{\Pi}_{\mathcal{P}_n} \boldsymbol{x}$ belong to that space. Therefore, the two terms in Eq. 111 are $\boldsymbol{M}_\tau$-orthogonal, and the Pythagorean identity gives

$$\begin{aligned} \|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}^2 &= \|(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{P}_n}^\tau)\boldsymbol{x}\|_{\boldsymbol{M}_\tau}^2 + \|(\boldsymbol{Q}_{\mathcal{P}_n}^\tau - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}^2 \\ &\geq \|(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{P}_n}^\tau)\boldsymbol{x}\|_{\boldsymbol{M}_\tau}^2. \end{aligned} \tag{112}$$

Dividing by $\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}^2$ and maximizing over nonzero $\boldsymbol{x} \in \mathcal{R}_\Theta$ proves $\varepsilon_{\mathcal{P}_n}^Q \leq \varepsilon_{\mathcal{P}_n}$.

Algorithm 2 Group detection on an unseen graph by screened Ward coarsening

Require: Weighted adjacency $\boldsymbol{W}$, augmented degree matrix $\widetilde{\boldsymbol{D}}$, normalized adjacency $\widehat{\boldsymbol{A}}$, features $\boldsymbol{X}$, learned coefficients $\Theta^\star \in \mathbb{R}^{(K+1)\times r}$, filter degree K, screening $\tau > 0$, and cumulative-score budget B_{stop}

Ensure: A hierarchy $\mathcal{H}$ and its selected partition $\widehat{\mathcal{P}}$

Form $\boldsymbol{M}_\tau = \boldsymbol{I} - \widehat{\boldsymbol{A}} + \tau \boldsymbol{I}$
Compute $\boldsymbol{T}_0 = \boldsymbol{X}$, $\boldsymbol{T}_1 = \widehat{\boldsymbol{A}}\boldsymbol{X}$, and $\boldsymbol{T}_{k+1} = 2\widehat{\boldsymbol{A}}\boldsymbol{T}_k - \boldsymbol{T}_{k-1}$ for $k = 1, \dots, K-1$
for $h = 1, \dots, r$ **do**
    Form the head output $\boldsymbol{Z}_h = \sum_{k=0}^{K} \theta^\star_{k,h} \boldsymbol{T}_k$
end for
Set $\boldsymbol{Z}_{\Theta^\star} = [\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_r]$ and $\boldsymbol{G} = \boldsymbol{Z}_{\Theta^\star}^\top \boldsymbol{M}_\tau \boldsymbol{Z}_{\Theta^\star}$
Set the learned target projector $\boldsymbol{Q}^\tau_{\mathcal{R}_{\Theta^\star}} = \boldsymbol{Z}_{\Theta^\star} \boldsymbol{G}^\dagger \boldsymbol{Z}_{\Theta^\star}^\top \boldsymbol{M}_\tau$
Initialize the singleton partition $\mathcal{P} \leftarrow \{\{v\} : v \in \mathcal{V}\}$, hierarchy $\mathcal{H} \leftarrow [\mathcal{P}]$, and cumulative score $b \leftarrow 0$
while $|\mathcal{P}| > 1$ **do**
    for each adjacent block pair $A, B \in \mathcal{P}$ **do**
        Compute the screened Ward score $s_{\mathcal{P}}(A, B)$ using Eq. 17
    end for
    Choose $(A^\star, B^\star) \in \arg\min_{A \sim B} s_{\mathcal{P}}(A, B)$ and set $s^\star = s_{\mathcal{P}}(A^\star, B^\star)$
    if $b + s^\star > B_{\mathrm{stop}}$ **then**
        break
    end if
    Merge $A^\star$ and $B^\star$: $\mathcal{P} \leftarrow \mathcal{P} \oplus (A^\star, B^\star)$
    Update $b \leftarrow b + s^\star$ and append $\mathcal{P}$ to $\mathcal{H}$
end while
Set $\widehat{\mathcal{P}} = \mathcal{P}$
return $\mathcal{H}$ and $\widehat{\mathcal{P}}$

For the reverse comparison, both projectors have range $\mathcal{A}_{\mathcal{P}_n}$, so $(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{Q}^\tau_{\mathcal{P}_n} = \boldsymbol{0}$. Consequently,

$$(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_n})(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}_n})\boldsymbol{x}.$$

Apply the operator characterization of $\mu^\tau_{\mathcal{P}_n}$ in Eq. 5 to the vector $(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}_n})\boldsymbol{x}$:

$$\|(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_n})\boldsymbol{x}\|_{\boldsymbol{M}_\tau} \le \mu^\tau_{\mathcal{P}_n} \|(\boldsymbol{I} - \boldsymbol{Q}^\tau_{\mathcal{P}_n})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}.$$

After division by $\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}$ and maximization over $\mathcal{R}_\Theta \setminus \{\boldsymbol{0}\}$, this yields $\varepsilon_{\mathcal{P}_n} \le \mu^\tau_{\mathcal{P}_n} \varepsilon^Q_{\mathcal{P}_n}$. Together with the first inequality, the proposition follows. □

D.3 Proof of Proposition 8

The next lemma makes precise why $\widehat{\boldsymbol{h}}_{A,B}$ is the unique direction discarded when A and B are merged.

***Lemma* 20** (Rank-one change of the block projector)**.** *Let $A, B \in \mathcal{P}_n$ be distinct blocks and let $\mathcal{P}_{n-1} = \mathcal{P}_n \oplus (A, B)$. Then*

$$\mathcal{A}_{\mathcal{P}_n} = \mathcal{A}_{\mathcal{P}_{n-1}} \oplus_{\boldsymbol{M}_\tau} \mathrm{span}\{\widehat{\boldsymbol{h}}_{A,B}\}, \qquad \boldsymbol{Q}^\tau_{\mathcal{P}_n} - \boldsymbol{Q}^\tau_{\mathcal{P}_{n-1}} = \widehat{\boldsymbol{h}}_{A,B}\widehat{\boldsymbol{h}}_{A,B}^\top \boldsymbol{M}_\tau. \tag{113}$$

In particular, the denominator in the definition of $\widehat{\boldsymbol{h}}_{A,B}$ is nonzero.

Proof. Before the merge, $\boldsymbol{v}_A$ and $\boldsymbol{v}_B$ are separate basis directions of $\mathcal{A}_{\mathcal{P}_n}$. After the merge, they are replaced by

$$\boldsymbol{v}_{A\cup B} = \frac{\sqrt{\mathrm{vol}(A)}\,\boldsymbol{v}_A + \sqrt{\mathrm{vol}(B)}\,\boldsymbol{v}_B}{\sqrt{\mathrm{vol}(A) + \mathrm{vol}(B)}}.$$

All other block indicators remain unchanged. Hence $\mathcal{A}_{\mathcal{P}_{n-1}} \subset \mathcal{A}_{\mathcal{P}_n}$ and $\dim(\mathcal{A}_{\mathcal{P}_n}) - \dim(\mathcal{A}_{\mathcal{P}_{n-1}}) = 1$.

The raw contrast $\boldsymbol{h}_{A,B}$ belongs to $\mathcal{A}_{\mathcal{P}_n}$. It is not in $\mathcal{A}_{\mathcal{P}_{n-1}}$: on A and B its degree-normalized values have opposite signs, whereas every vector in the post-merge space has one common degree-normalized value on $A \cup B$. Thus $(\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_{n-1}})\boldsymbol{h}_{A,B} \neq \boldsymbol{0}$. By construction, this residual lies in $\mathcal{A}_{\mathcal{P}_n}$ and is $\boldsymbol{M}_\tau$-orthogonal to $\mathcal{A}_{\mathcal{P}_{n-1}}$. After normalization it is exactly $\widehat{\boldsymbol{h}}_{A,B}$. Because the orthogonal complement is one-dimensional, this proves the direct-sum statement.

The $\boldsymbol{M}_\tau$-orthogonal projector onto $\operatorname{span}\{\widehat{\boldsymbol{h}}_{A,B}\}$ is $\widehat{\boldsymbol{h}}_{A,B}\widehat{\boldsymbol{h}}_{A,B}^{\top}\boldsymbol{M}_\tau$, since $\|\widehat{\boldsymbol{h}}_{A,B}\|_{\boldsymbol{M}_\tau} = 1$. The projector onto an orthogonal direct sum is the sum of the two component projectors. Hence

$$\boldsymbol{Q}^{\tau}_{\mathcal{P}_n} = \boldsymbol{Q}^{\tau}_{\mathcal{P}_{n-1}} + \widehat{\boldsymbol{h}}_{A,B}\widehat{\boldsymbol{h}}_{A,B}^{\top}\boldsymbol{M}_\tau,$$

which is the second identity in Eq. 113. □

Proposition. *For any merge sequence from singletons, with* $\dim(\mathcal{R}_\Theta) \le q$*,*

$$(\varepsilon^{Q}_{\mathcal{P}_n})^2 \le \sum_{k=n+1}^{N} s_{\mathcal{P}_k}(\mathcal{B}_i^{(k)}, \mathcal{B}_j^{(k)}) \le q(\varepsilon^{Q}_{\mathcal{P}_n})^2, \tag{114}$$

where $(\mathcal{B}_i^{(k)}, \mathcal{B}_j^{(k)})$ *is the pair merged at the* k*-block partition.*

Proof. Recall that

$$\varepsilon^{Q}_{\mathcal{P}_n} := \max_{\boldsymbol{x}\in\mathcal{R}_\Theta\setminus\{\boldsymbol{0}\}} \frac{\|(\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_n})\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}{\|\boldsymbol{x}\|_{\boldsymbol{M}_\tau}}.$$

Since

$$\boldsymbol{Z}_\Theta(\boldsymbol{G}_\Theta^{\dagger})^{1/2}$$

spans $\mathcal{R}_\Theta$ and is $\boldsymbol{M}_\tau$-orthonormal on the positive spectral support of $\boldsymbol{G}_\Theta$, every $\boldsymbol{x} \in \mathcal{R}_\Theta$ can be written as

$$\boldsymbol{x} = \boldsymbol{Z}_\Theta(\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{c},$$

with $\boldsymbol{c}$ restricted to that support, and

$$\|\boldsymbol{x}\|^2_{\boldsymbol{M}_\tau} = \|\boldsymbol{c}\|_2^2.$$

Moreover, since $\boldsymbol{Q}^{\tau}_{\mathcal{P}_n}$ is an $\boldsymbol{M}_\tau$-orthogonal projector,

$$\begin{aligned}\|(\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_n})\boldsymbol{x}\|^2_{\boldsymbol{M}_\tau} &= \boldsymbol{x}^{\top}\boldsymbol{M}_\tau(\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_n})\boldsymbol{x} \\ &= \boldsymbol{c}^{\top}(\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau(\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_n})\boldsymbol{Z}_\Theta(\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{c}. \end{aligned}\tag{115}$$

Define

$$\boldsymbol{E}_{\mathcal{P}_n} := (\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau(\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_n})\boldsymbol{Z}_\Theta(\boldsymbol{G}_\Theta^{\dagger})^{1/2}. \tag{116}$$

Then

$$(\varepsilon^{Q}_{\mathcal{P}_n})^2 = \max_{\boldsymbol{c}\neq\boldsymbol{0}} \frac{\boldsymbol{c}^{\top}\boldsymbol{E}_{\mathcal{P}_n}\boldsymbol{c}}{\|\boldsymbol{c}\|_2^2}.$$

Since $\boldsymbol{E}_{\mathcal{P}_n}$ is positive semidefinite, the Rayleigh quotient gives

$$(\varepsilon^{Q}_{\mathcal{P}_n})^2 = \lambda_{\max}(\boldsymbol{E}_{\mathcal{P}_n}). \tag{117}$$

Now consider the merge performed at the k-block partition $\mathcal{P}_k$, and write its pair as (A_k, B_k). Applying Lemma 20 gives

$$\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_{k-1}} = \boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_k} + \widehat{\boldsymbol{h}}_{A_k,B_k}\widehat{\boldsymbol{h}}_{A_k,B_k}^{\top}\boldsymbol{M}_\tau.$$

Substitution in Eq. 116 yields the rank-one update

$$\boldsymbol{E}_{\mathcal{P}_{k-1}} = \boldsymbol{E}_{\mathcal{P}_k} + \boldsymbol{a}_k\boldsymbol{a}_k^{\top}, \qquad \boldsymbol{a}_k := (\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau\widehat{\boldsymbol{h}}_{A_k,B_k}. \tag{118}$$

The squared norm of the update vector is precisely the merge score:

$$\begin{aligned}\|\boldsymbol{a}_k\|_2^2 &= \widehat{\boldsymbol{h}}_{A_k,B_k}^{\top}\boldsymbol{M}_\tau\boldsymbol{Z}_\Theta\boldsymbol{G}_\Theta^{\dagger}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau\widehat{\boldsymbol{h}}_{A_k,B_k} \\ &= \|\boldsymbol{Q}^{\tau}_{\mathcal{R}_\Theta}\widehat{\boldsymbol{h}}_{A_k,B_k}\|^2_{\boldsymbol{M}_\tau} = s_{\mathcal{P}_k}(A_k, B_k). \end{aligned}\tag{119}$$

At the singleton partition, $\boldsymbol{Q}^{\tau}_{\mathcal{P}_N} = \boldsymbol{I}$, so $\boldsymbol{E}_{\mathcal{P}_N} = \boldsymbol{0}$. Telescoping Eq. 118 from N blocks down to n blocks gives

$$\boldsymbol{E}_{\mathcal{P}_n} = \sum_{k=n+1}^{N} \boldsymbol{a}_k \boldsymbol{a}_k^{\top}, \qquad \operatorname{tr}(\boldsymbol{E}_{\mathcal{P}_n}) = \sum_{k=n+1}^{N} s_{\mathcal{P}_k}(A_k, B_k).$$

Finally, $\boldsymbol{E}_{\mathcal{P}_n}$ is positive semidefinite and has rank at most $\dim(\mathcal{R}_\Theta) \le q$. Thus

$$\lambda_{\max}(\boldsymbol{E}_{\mathcal{P}_n}) \le \operatorname{tr}(\boldsymbol{E}_{\mathcal{P}_n}) \le q\lambda_{\max}(\boldsymbol{E}_{\mathcal{P}_n}).$$

Combining this inequality with Eq. 117 and Eq. 119 proves Eq. 114. □

D.4 Proof of Proposition 9

We first record the elementary geometric fact used in the proof of Proposition 9.

***Lemma* 21** (Projection energy of two aligned directions)**.** *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be unit vectors in an inner-product space, and let $\boldsymbol{Q}$ be an orthogonal projector. If*

$$s = \|\boldsymbol{Q}\boldsymbol{x}\|^2, \qquad t = \|\boldsymbol{Q}\boldsymbol{y}\|^2, \qquad \eta = |\langle \boldsymbol{x}, \boldsymbol{y}\rangle|^2,$$

then

$$\Big[\sqrt{\eta s} - \sqrt{(1-\eta)(1-s)}\Big]_+^2 \le t \le 1 - \Big[\sqrt{\eta(1-s)} - \sqrt{(1-\eta)s}\Big]_+^2. \tag{120}$$

Proof. Decompose both vectors into their components in the range of $\boldsymbol{Q}$ and its orthogonal complement:

$$\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{x}, \qquad \boldsymbol{y} = \boldsymbol{Q}\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{y}.$$

The two subspaces are orthogonal. Hence, by the triangle inequality and Cauchy–Schwarz,

$$\begin{aligned} |\langle \boldsymbol{x}, \boldsymbol{y}\rangle| &\le |\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle| + |\langle (\boldsymbol{I}-\boldsymbol{Q})\boldsymbol{x}, (\boldsymbol{I}-\boldsymbol{Q})\boldsymbol{y}\rangle| \\ &\le \sqrt{st} + \sqrt{(1-s)(1-t)}. \end{aligned}$$

Set $\alpha = \arccos\sqrt{s}$, $\beta = \arccos\sqrt{t}$, and $\vartheta = \arccos\sqrt{\eta}$; all three angles lie in $[0, \pi/2]$. The preceding inequality becomes $\cos\vartheta \le \cos(\alpha - \beta)$ and therefore $|\alpha - \beta| \le \vartheta$. Equivalently,

$$\max\{0, \alpha - \vartheta\} \le \beta \le \min\{\pi/2, \alpha + \vartheta\}.$$

Applying the decreasing function $\cos^2$ on $[0, \pi/2]$ and expanding $\cos(\alpha \pm \vartheta)$ gives the two inequalities in Eq. 120. The positive parts account for the cases in which an endpoint is clipped to 0 or $\pi/2$. □

Proposition. *Define $\eta_n(\mathcal{B}_i, \mathcal{B}_j) := |\langle \bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}, \widehat{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j}\rangle_{\boldsymbol{M}_\tau}|^2 \in (0, 1]$ as the squared alignment between the raw and deflated contrasts. Writing $s = s^0(\mathcal{B}_i, \mathcal{B}_j)$ and $\eta = \eta_n(\mathcal{B}_i, \mathcal{B}_j)$,*

$$\Big[\sqrt{\eta s} - \sqrt{(1-\eta)(1-s)}\Big]_+^2 \le s_{\mathcal{P}_n}(\mathcal{B}_i, \mathcal{B}_j) \le 1 - \Big[\sqrt{\eta(1-s)} - \sqrt{(1-\eta)s}\Big]_+^2. \tag{121}$$

Moreover, $\eta_n(\mathcal{B}_i, \mathcal{B}_j) \ge \big(\mu^{\tau}_{\mathcal{P}_{n-1}}\big)^{-2}$.

Proof. Let

$$\boldsymbol{r}_{A,B} := (\boldsymbol{I} - \boldsymbol{Q}^{\tau}_{\mathcal{P}_{n-1}})\bar{\boldsymbol{h}}_{A,B}.$$

Lemma 20 shows that $\boldsymbol{r}_{A,B} \neq \boldsymbol{0}$ and $\widehat{\boldsymbol{h}}_{A,B} = \boldsymbol{r}_{A,B}/\|\boldsymbol{r}_{A,B}\|_{\boldsymbol{M}_\tau}$. Since $\boldsymbol{r}_{A,B}$ is orthogonal to $\boldsymbol{Q}^{\tau}_{\mathcal{P}_{n-1}}\bar{\boldsymbol{h}}_{A,B}$,

$$\begin{aligned} \langle \bar{\boldsymbol{h}}_{A,B}, \widehat{\boldsymbol{h}}_{A,B}\rangle_{\boldsymbol{M}_\tau} &= \|\boldsymbol{r}_{A,B}\|_{\boldsymbol{M}_\tau}, \\ \eta_n(A, B) &= \|\boldsymbol{r}_{A,B}\|^2_{\boldsymbol{M}_\tau} \in (0, 1]. \end{aligned}$$

Apply Lemma 21 in the $\boldsymbol{M}_\tau$ inner product with

$$\boldsymbol{x} = \bar{\boldsymbol{h}}_{A,B}, \qquad \boldsymbol{y} = \widehat{\boldsymbol{h}}_{A,B}, \qquad \boldsymbol{Q} = \boldsymbol{Q}^{\tau}_{\mathcal{R}_\Theta}.$$

Then $s = \|\boldsymbol{Q}^{\tau}_{\mathcal{R}_\Theta}\bar{\boldsymbol{h}}_{A,B}\|^2_{\boldsymbol{M}_\tau} = s^0(A, B)$ and $t = \|\boldsymbol{Q}^{\tau}_{\mathcal{R}_\Theta}\widehat{\boldsymbol{h}}_{A,B}\|^2_{\boldsymbol{M}_\tau} = s_{\mathcal{P}_n}(A, B)$. The two inequalities in Eq. 18 now follow directly.

It remains to lower-bound the alignment. The raw contrast is Euclidean orthogonal to the post-merge block space: its contributions on A and B cancel against $\boldsymbol{v}_{A\cup B}$, and it has disjoint support from every other post-merge block. Therefore

$$(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_{n-1}})\bar{\boldsymbol{h}}_{A,B} = \bar{\boldsymbol{h}}_{A,B}.$$

Moreover, $\bar{\boldsymbol{h}}_{A,B} - \boldsymbol{r}_{A,B} = \boldsymbol{Q}^{\tau}_{\mathcal{P}_{n-1}}\bar{\boldsymbol{h}}_{A,B}$ belongs to the post-merge block space, so

$$(\boldsymbol{I} - \boldsymbol{\Pi}_{\mathcal{P}_{n-1}})\boldsymbol{r}_{A,B} = \bar{\boldsymbol{h}}_{A,B}.$$

The definition of $\mu^{\tau}_{\mathcal{P}_{n-1}}$ in Eq. 5 consequently gives

$$1 = \|\bar{\boldsymbol{h}}_{A,B}\|_{\boldsymbol{M}_\tau} \le \mu^{\tau}_{\mathcal{P}_{n-1}}\|\boldsymbol{r}_{A,B}\|_{\boldsymbol{M}_\tau} = \mu^{\tau}_{\mathcal{P}_{n-1}}\sqrt{\eta_n(A,B)}.$$

Rearranging proves $\eta_n(A,B) \ge (\mu^{\tau}_{\mathcal{P}_{n-1}})^{-2}$. □

D.5 Proof of Lemma 10

Lemma. *If $\mathcal{B}_i, \mathcal{B}_j \subseteq \mathcal{S}$, then $\bar{\boldsymbol{h}}_{\mathcal{B}_i,\mathcal{B}_j} \in \widehat{\mathcal{F}}^{\tau}_{\mathcal{S}}$ and hence $s^0(\mathcal{B}_i,\mathcal{B}_j) \le \chi^{\tau}_{\mathcal{R}_\Theta}(\mathcal{S})$.*

Proof. Suppose $A, B \subseteq \mathcal{S}$. The vector $\boldsymbol{h}_{A,B}$ is supported on $\mathcal{S}$. On $\mathcal{S}$, the degree-normalized group direction $\widehat{\boldsymbol{v}}_{\mathcal{S}}$ is proportional to $\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}_{\mathcal{S}}$. On the other hand,

$$\begin{aligned}\boldsymbol{h}_{A,B}^{\top}\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}_{\mathcal{S}} &= \frac{1}{\mathrm{vol}(A)}\sum_{v\in A}\widetilde{d}_v - \frac{1}{\mathrm{vol}(B)}\sum_{v\in B}\widetilde{d}_v \\ &= 1 - 1 = 0.\end{aligned}$$

Thus $\boldsymbol{h}_{A,B} \in \mathcal{F}_{\mathcal{S}}$, and normalizing it in the screened norm gives $\bar{\boldsymbol{h}}_{A,B} \in \widehat{\mathcal{F}}^{\tau}_{\mathcal{S}}$. The definition of confusability therefore implies

$$s^0(A,B) = \|\boldsymbol{Q}^{\tau}_{\mathcal{R}_\Theta}\bar{\boldsymbol{h}}_{A,B}\|^2_{\boldsymbol{M}_\tau} \le \chi^{\tau}_{\mathcal{R}_\Theta}(\mathcal{S}).$$

□

We next derive the response formula used to interpret this raw score. From the screened projector formula,

$$\begin{aligned}s^0(A,B) &= \frac{\boldsymbol{h}_{A,B}^{\top}\boldsymbol{M}_\tau\boldsymbol{Z}_\Theta\boldsymbol{G}_\Theta^{\dagger}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau\boldsymbol{h}_{A,B}}{\|\boldsymbol{h}_{A,B}\|^2_{\boldsymbol{M}_\tau}} \\ &= \frac{\|(\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau\boldsymbol{h}_{A,B}\|_2^2}{\|\boldsymbol{h}_{A,B}\|^2_{\boldsymbol{M}_\tau}}.\end{aligned} \tag{122}$$

By the definition of $\boldsymbol{\ell}_\Theta$ in the main text,

$$\boldsymbol{M}_\tau\boldsymbol{Z}_\Theta(\boldsymbol{G}_\Theta^{\dagger})^{1/2} = \widetilde{\boldsymbol{D}}^{1/2}\boldsymbol{\ell}_\Theta.$$

Also,

$$\boldsymbol{h}_{A,B} = \frac{\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}_A}{\mathrm{vol}(A)} - \frac{\widetilde{\boldsymbol{D}}^{1/2}\mathbf{1}_B}{\mathrm{vol}(B)}.$$

Substituting these identities into the numerator of Eq. 122 gives

$$\begin{aligned}(\boldsymbol{G}_\Theta^{\dagger})^{1/2}\boldsymbol{Z}_\Theta^{\top}\boldsymbol{M}_\tau\boldsymbol{h}_{A,B} &= \frac{1}{\mathrm{vol}(A)}\sum_{v\in A}\widetilde{d}_v\boldsymbol{\ell}_{\Theta,v,:}^{\top} - \frac{1}{\mathrm{vol}(B)}\sum_{v\in B}\widetilde{d}_v\boldsymbol{\ell}_{\Theta,v,:}^{\top} \\ &= \bar{\boldsymbol{\ell}}_A - \bar{\boldsymbol{\ell}}_B.\end{aligned}$$

Hence

$$\boxed{s^0(A,B) = \frac{\|\bar{\boldsymbol{\ell}}_A - \bar{\boldsymbol{\ell}}_B\|_2^2}{\|\boldsymbol{h}_{A,B}\|^2_{\boldsymbol{M}_\tau}}.} \tag{123}$$

This identity explains the role of the learned response: similar degree-weighted mean responses make an internal raw score small, and Proposition 9 transfers that conclusion to the actual deflated Ward score.

D.6 Proof of Theorem 11

The internal bound above is controlled by confusability. The boundary bound is controlled by capture together with the alignment constant $\kappa_{\mathcal{S}}$. The following lemma supplies the missing geometric step.

Lemma 22 (Alignment of a boundary merge)**.** *Let $\mathcal{P}_n$ be $\mathcal{S}$-clean and let $A \subseteq \mathcal{S}$ and $B \subseteq \mathcal{S}^c$ be adjacent blocks. For every original boundary edge $e = (u, v)$ with $u \in A$ and $v \in B$,*

$$|\langle \widehat{\boldsymbol{h}}_{A,B}, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau}|^2 \geq \zeta_e^2 \geq \kappa_{\mathcal{S}}. \tag{124}$$

Proof. For this proof, consider the screened Riesz representative of the degree-normalized endpoint difference,

$$\boldsymbol{t}_e := \boldsymbol{M}_\tau^{-1} \widetilde{\boldsymbol{D}}^{-1/2} (\boldsymbol{e}_u - \boldsymbol{e}_v).$$

It satisfies, for every $\boldsymbol{x}$,

$$\langle \boldsymbol{t}_e, \boldsymbol{x} \rangle_{\boldsymbol{M}_\tau} = \frac{x_u}{\sqrt{\widetilde{d}_u}} - \frac{x_v}{\sqrt{\widetilde{d}_v}}, \qquad \|\boldsymbol{t}_e\|_{\boldsymbol{M}_\tau}^2 = R_e. \tag{125}$$

After merging A and B, every vector in $\mathcal{A}_{\mathcal{P}_{n-1}}$ has the same degree-normalized value at u and v. The first identity in Eq. 125 therefore shows that $\boldsymbol{t}_e$ is $\boldsymbol{M}_\tau$-orthogonal to $\mathcal{A}_{\mathcal{P}_{n-1}}$.

Because the current partition is $\mathcal{S}$-clean, $\widehat{\boldsymbol{v}}_{\mathcal{S}} \in \mathcal{A}_{\mathcal{P}_n}$. Lemma 20 then implies

$$\begin{aligned} \langle \boldsymbol{t}_e, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau} &= \langle \boldsymbol{Q}_{\mathcal{P}_n}^\tau \boldsymbol{t}_e, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau} \\ &= \langle \boldsymbol{t}_e, \widehat{\boldsymbol{h}}_{A,B} \rangle_{\boldsymbol{M}_\tau} \langle \widehat{\boldsymbol{h}}_{A,B}, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau}. \end{aligned}$$

Direct substitution of the normalized group direction into Eq. 125, together with the definition of ζ_e in the main text, gives

$$|\langle \boldsymbol{t}_e, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau}| = \zeta_e \sqrt{R_e}.$$

Meanwhile, Cauchy–Schwarz and $\|\widehat{\boldsymbol{h}}_{A,B}\|_{\boldsymbol{M}_\tau} = 1$ give

$$|\langle \boldsymbol{t}_e, \widehat{\boldsymbol{h}}_{A,B} \rangle_{\boldsymbol{M}_\tau}| \leq \|\boldsymbol{t}_e\|_{\boldsymbol{M}_\tau} = \sqrt{R_e}.$$

Combining the last three displays and cancelling $\sqrt{R_e} > 0$ yields

$$|\langle \widehat{\boldsymbol{h}}_{A,B}, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau}| \geq \zeta_e.$$

Squaring proves the first inequality in Eq. 124; the second follows from $\kappa_{\mathcal{S}} = \min_{e \in \partial \mathcal{S}} \zeta_e^2$. □

Theorem. *For a connected group $\mathcal{S}$, define $\kappa_{\mathcal{S}} := \min_{e \in \partial \mathcal{S}} \zeta_e^2$, $\eta_\star := \bar{\mu}_\tau^{-2}$. Let $C_{\mathcal{S}} := C_{\mathcal{R}_\Theta}^\tau(\mathcal{S})$, $\chi_{\mathcal{S}} := \chi_{\mathcal{R}_\Theta}^\tau(\mathcal{S})$. At every $\mathcal{S}$-clean partition, internal and boundary candidates satisfy*

$$s_{\mathrm{int}}(\mathcal{B}_i, \mathcal{B}_j; \mathcal{S}) \leq 1 - \left[\sqrt{\eta_\star(1-\chi_{\mathcal{S}})} - \sqrt{(1-\eta_\star)\chi_{\mathcal{S}}}\right]_+^2 =: U_{\mathcal{S}}, \tag{126}$$

$$s_\partial(\mathcal{B}_i, \mathcal{B}_j; \mathcal{S}) \geq \left[\sqrt{\kappa_{\mathcal{S}} C_{\mathcal{S}}} - \sqrt{(1-\kappa_{\mathcal{S}})(1-C_{\mathcal{S}})}\right]_+^2 =: L_{\mathcal{S}}. \tag{127}$$

If $U_{\mathcal{S}} < L_{\mathcal{S}}$, the greedy Ward hierarchy forms $\mathcal{S}$ as a supernode before any merge crosses its boundary.

Proof. Write $C_{\mathcal{S}} = C_{\mathcal{R}_\Theta}^\tau(\mathcal{S})$ and $\chi_{\mathcal{S}} = \chi_{\mathcal{R}_\Theta}^\tau(\mathcal{S})$, as in the theorem.

First consider an internal candidate $A, B \subseteq \mathcal{S}$. By Lemma 10,

$$s^0(A, B) \leq \chi_{\mathcal{S}}.$$

Proposition 9 and the uniform bound on the screened block-distortion factor give

$$\eta_n(A, B) \geq (\mu_{\mathcal{P}_{n-1}}^\tau)^{-2} \geq \eta_\star.$$

For the upper inequality in Eq. 18, define

$$F(s, \eta) := 1 - \left[\sqrt{\eta(1-s)} - \sqrt{(1-\eta)s}\right]_+^2.$$

The angular representation

$$F(s,\eta) = \sin^2\left(\min\left\{\frac{\pi}{2}, \arcsin\sqrt{s} + \arccos\sqrt{\eta}\right\}\right)$$

shows that F is nondecreasing in s and nonincreasing in η. Therefore

$$\begin{aligned} s_{\mathrm{int}}(A,B;\mathcal{S}) &\le F(s^0(A,B), \eta_n(A,B)) \\ &\le F(\chi_{\mathcal{S}}, \eta_\star) = U_{\mathcal{S}}. \end{aligned}$$

Now consider a boundary candidate $A \subseteq \mathcal{S}$, $B \subseteq \mathcal{S}^c$. Lemma 22 gives

$$|\langle \widehat{\boldsymbol{h}}_{A,B}, \widehat{\boldsymbol{v}}_{\mathcal{S}} \rangle_{\boldsymbol{M}_\tau}|^2 \ge \kappa_{\mathcal{S}}.$$

Apply Lemma 21 with $\boldsymbol{x} = \widehat{\boldsymbol{v}}_{\mathcal{S}}$, $\boldsymbol{y} = \widehat{\boldsymbol{h}}_{A,B}$, and $\boldsymbol{Q} = \boldsymbol{Q}^\tau_{\mathcal{R}_\Theta}$. Since

$$\|\boldsymbol{Q}^\tau_{\mathcal{R}_\Theta} \widehat{\boldsymbol{v}}_{\mathcal{S}}\|^2_{\boldsymbol{M}_\tau} = C_{\mathcal{S}},$$

the lower inequality of that lemma yields

$$\begin{aligned} s_\partial(A,B;\mathcal{S}) &\ge \left[\sqrt{\kappa_{\mathcal{S}} C_{\mathcal{S}}} - \sqrt{(1-\kappa_{\mathcal{S}})(1-C_{\mathcal{S}})}\right]_+^2 \\ &= L_{\mathcal{S}}. \end{aligned}$$

It remains to show that the score gap forces formation of $\mathcal{S}$. The singleton partition is $\mathcal{S}$-clean. Suppose inductively that the current partition is $\mathcal{S}$-clean and that $\mathcal{S}$ is still represented by at least two blocks. Because the subgraph induced by $\mathcal{S}$ is connected, at least one original edge joins two distinct current blocks inside $\mathcal{S}$. Hence an internal adjacent candidate exists, and its score is at most $U_{\mathcal{S}}$. Every boundary candidate has score at least $L_{\mathcal{S}}$. Under the strict condition $U_{\mathcal{S}} < L_{\mathcal{S}}$, a minimum-score merge therefore cannot cross the boundary of $\mathcal{S}$.

An internal merge preserves $\mathcal{S}$-cleanliness and reduces by one the number of blocks covering $\mathcal{S}$; a merge wholly outside $\mathcal{S}$ also preserves cleanliness. Since only finitely many outside merges are possible, continued coarsening must eventually select enough internal merges to make $\mathcal{S}$ one block, and no boundary merge can occur before that event. This proves the theorem. If $\mathcal{S}$ is a singleton, it is already a block at initialization. □

E Additional results

E.1 Experimental datasets and evaluation details

Synthetic graphs. For each random seed, we generate three fitting graphs and five independently generated test graphs, each with $N = 3{,}000$ nodes. Twelve disjoint groups are planted by randomly permuting the nodes and drawing each group size uniformly from $\{7, \dots, 20\}$. The background is formed by sampling $N\bar{d}/2$ random endpoint pairs with target mean degree $\bar{d} = 4$; self-loops are discarded and repeated edges are collapsed. For a group of size s and nominal density ρ, we first add a randomly ordered spanning path to ensure connectivity and then sample additional within-group edges until the group contains $\max\{s-1, \mathrm{round}(\rho\binom{s}{2})\}$ edges. Consequently, small values of ρ are bounded below by the spanning-path density, and background edges may further increase the realized density. We therefore record the realized, rather than nominal, within-group density after graph generation. Each node receives 32 independent Gaussian feature channels, standardized columnwise within its graph. The sweep uses $\rho \in \{0, 0.05, \dots, 1\}$ and seeds $\{1, 2, 3\}$. For a fixed seed, group memberships, background edges, and node features are shared across densities, so only the planted within-group edges change. Within each fitting graph, a seeded split assigns half of the planted groups to representation learning and cut selection; all groups in the five separately generated test graphs remain unseen. Reported values first aggregate the five test graphs within each seed and then average over the three seeds, giving 15 held-out test graphs per density.

Elliptic++. Elliptic++ contains Bitcoin wallet addresses, address–address transaction records, wallet features, and illicit, licit, or unknown labels over 49 temporal steps (Elmougy & Liu, 2023). We use the wallet-address graph and construct one graph for each step from 26 through 40. For each step, we retain one feature record per active address, keeping the final record when an address is duplicated, and merge the corresponding class label; missing labels are treated as unknown. The address graph is induced on these wallets, symmetrized, converted to an unweighted graph, and stripped of self-loops and duplicate edges. A target illicit group is a connected component of the subgraph induced by known-illicit wallets, with components of size one discarded. These groups are therefore label-derived proxies for coordinated illicit activity, rather than independently verified criminal organizations. Days 26–30 form the fitting period and days 31–40 are untouched temporal test graphs. Before normalization, we remove identifiers, class and time-step fields, and the four absolute block-height attributes `first_block_appeared_in`, `last_block_appeared_in`, `first_sent_block`, and `first_received_block`, preventing calendar position from directly revealing the temporal split. Non-finite values are replaced by zero; zero-variance columns are removed; and the remaining wallet features are standardized using means and standard deviations fitted only on days 26–30. Each wallet-feature vector is then normalized to unit ℓ_2 norm and concatenated with 64 column-standardized Gaussian range-finder channels. Known-label wallets on the fitting days are split stratifiably into 80% training and 20% validation subsets. The reported run uses a stratified half of the training subset, corresponding to 4,095 wallets or 40% of all known-label fitting-period wallets, to fit the node-label predictors; the validation subset is retained for early stopping and decision-threshold selection. Unknown-label wallets remain in the graphs but are excluded from classifier losses and evaluation metrics.

Amazon co-purchase network. The Amazon co-purchase graph represents products as nodes and frequently co-purchased product pairs as undirected edges. The processed graph contains $N = 334{,}863$ nodes and $E = 925{,}872$ edges, together with the 5,000 top-quality product-category communities distributed by SNAP (Yang & Leskovec, 2012). We remove self-loops, collapse repeated edges, and reindex the original product identifiers to contiguous node indices. Community members absent from the graph are discarded, and only communities with at least two retained products are considered. The supplied file contains many exactly repeated node sets: only 1,517 of its 5,000 entries are distinct, with one set repeated 48 times. To prevent identical communities from appearing in both splits, we retain the first occurrence of each distinct node set before performing a seeded random 50/50 split, giving 758 training and 759 test communities. The retained communities range from 3 to 328 products, with a median size of 10. They remain genuinely overlapping: approximately 40.3% of the communities intersect another retained community, and 37.6% of their member products belong to more than one community. Since Amazon supplies no node attributes, every product receives 32 independent Gaussian range-finder channels, standardized columnwise. These features contain no category labels and allow the polynomial filters to probe graph structure. The representation and stopping cut are learned from the 758 training communities, while all reported recovery metrics are computed on the 759 held-out communities. This is therefore a held-out-community evaluation on one graph, rather than transfer to an unseen graph.

Evaluation and aggregation. At each level of a coarsening hierarchy, every target group $\mathcal{S}$ is compared with the supernode having the largest overlap with it; ties are resolved in favor of the smaller supernode. Group precision, recall, and F1 are computed from this best match as defined in the main text, with F1 set to zero when the precision–recall denominator is zero. Macro precision, recall, and F1 are averaged separately over target groups; in particular, macro F1 is the mean of the per-group F1 scores and not the harmonic mean of macro precision and recall. A group is counted as detected when both its precision and recall exceed 0.51. PR-AUC is obtained by sorting the macro precision–recall points over the complete hierarchy and applying trapezoidal integration, so it does not depend on a single selected cut. Synthetic results are first pooled over the five test graphs belonging to each seed and then averaged across seeds; Amazon metrics are evaluated on the single held-out community split.

For the label-assisted Elliptic++ experiment, we additionally report B^3 precision, recall, and F1 (Amigó et al., 2009). Before scoring, each predicted group is projected onto its known-label members, and predicted groups with fewer than two such members are treated as unevaluable. For each known wallet, B^3 precision measures the fraction of its predicted group belonging to its true group, while recall measures the fraction of its true group contained in the prediction. These values are averaged

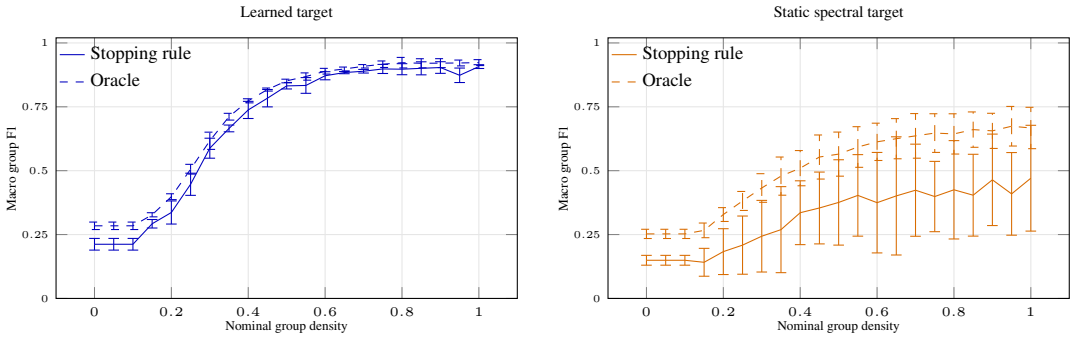


Figure 4: Macro group F1 at the training-selected stopping cut (solid) and the post-hoc test-label oracle cut (dashed). Left: learned target. Right: rank-matched static spectral target. Error bars indicate one standard deviation across the three seed-level test-graph averages.

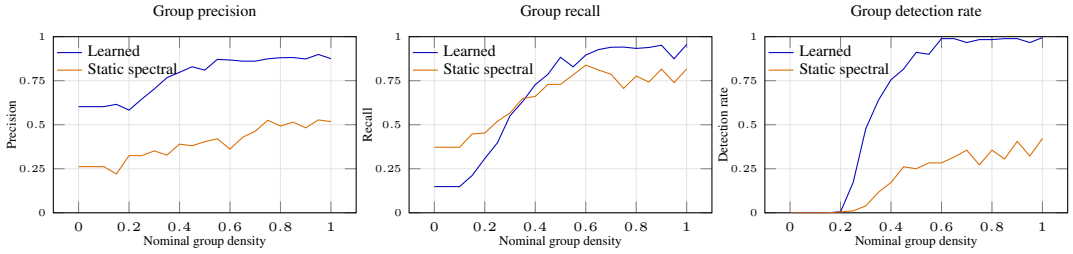


Figure 5: Stopping-rule performance on the held-out graphs. Left: macro group precision. Center: macro group recall. Right: group detection rate, where a group is detected when both its precision and recall exceed 0.51. Blue denotes the learned target and orange denotes the rank-matched static spectral target.

over wallets and combined by their harmonic mean, thereby providing partial credit for approximately recovered groups. Aggregate temporal scores pool the underlying numerators and denominators across days rather than averaging the daily F1 values. When B^3 is reported for the overlapping Amazon communities, we use the multiplicity-corrected extension for overlapping ground truth (Amigó et al., 2009). Test-label oracle cuts are reported only as diagnostic upper bounds; deployable cuts are selected using fitting data.

E.2 Additional density-sweep results

Experimental protocol. We evaluate 21 nominal within-group densities using three random seeds. For each seed, we generate three training graphs and five independently generated test graphs. Every graph contains 3,000 nodes, 12 disjoint planted groups of size 7–20, a background mean degree of four, and 16-dimensional node features. The learned target uses four degree-32 Chebyshev-filter heads, resulting in an effective target rank of 64. The static spectral target uses the 64 lowest-frequency eigenvectors of the symmetric normalized Laplacian, so the two representations have equal rank. Both targets are evaluated using the same deflated Ward coarsening procedure with screening parameter $\tau = 0.5$.

The stopping rule selects a normalized cumulative merge-score threshold using the training groups and applies it unchanged to the held-out graphs. In contrast, the oracle cut maximizes macro group F1 using the test-group labels and is reported only as a post-hoc upper bound. PR-AUC integrates the macro precision–recall trajectory over the complete hierarchy and is therefore independent of either cut. For each seed, metrics are first averaged over its five test graphs; the reported means and standard deviations are then computed over the three seed-level averages.

Effect of cut selection. Figure 4 compares the training-selected and oracle cuts. Oracle selection improves both targets, but it does not eliminate the learned target's advantage. At density 0.50, the learned target obtains F1 scores of 0.832 and 0.852 under the stopping and oracle cuts, respectively, whereas the static target obtains 0.376 and 0.565. The larger oracle improvement for the static target indicates that cut transfer contributes to its lower performance. Nevertheless, the remaining oracle gap shows that cut selection alone does not explain the difference between the representations.

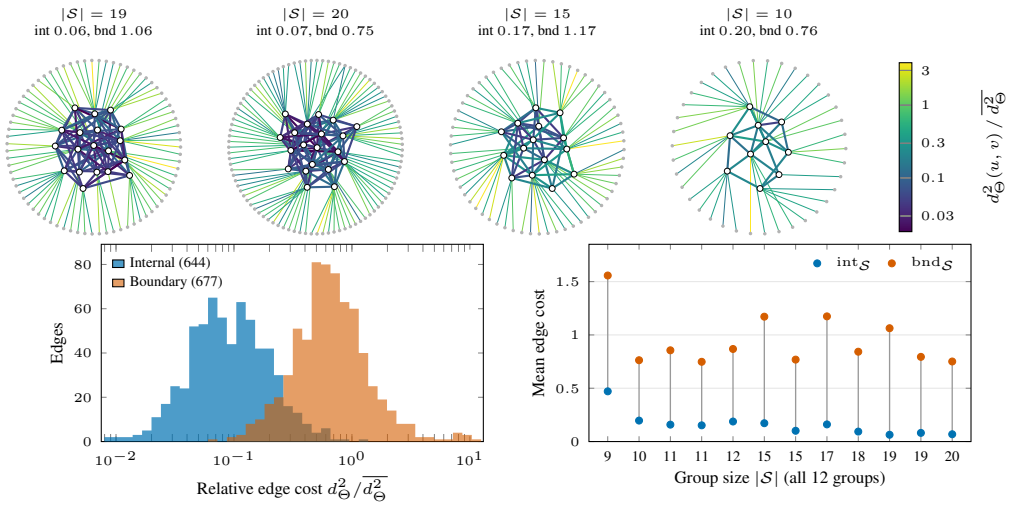


Figure 6: Learned response contrasts on a held-out synthetic graph. *Top:* Four example planted groups. Thick edges connect members of the group, while thin edges cross its boundary; color represents the relative response contrast $d^2_\Theta(u,v)/\overline{d^2_\Theta}$. *Bottom left:* Pooled distributions over all internal and boundary edges. *Bottom right:* Mean internal and boundary costs for each of the 12 planted groups, ordered by size. Costs are normalized by their graph-wide mean. The filter bank was learned from separate training graphs.

Precision, recall, and detection. Figure 5 separates the stopping-rule results into precision (left), recall (center), and detection rate (right). The learned target has higher mean precision at all 21 densities and higher mean recall at 13 densities. At density 0.25, the static target has higher recall (0.519 versus 0.397) but substantially lower precision (0.324 versus 0.645), resulting in F1 scores of 0.209 and 0.447 for the static and learned targets, respectively. At density 0.50, the learned target leads in both precision (0.810 versus 0.404) and recall (0.883 versus 0.729). Its detection rate is correspondingly 0.911, compared with 0.250 for the static target, and increases to 0.994 versus 0.422 at density 1.00. Thus, at moderate and high densities, the learned target's advantage cannot be explained solely by a different precision–recall operating point.

E.3 Internal and boundary costs under the learned target

The learning objective is designed to produce similar screened responses within a target group and distinct responses across its boundary. We examine whether this separation transfers to a held-out synthetic graph and whether it produces the merge-cost ordering required by the coarsener.

Experimental setting. We use the first held-out graph at nominal planted-group density 0.5 from the experiment in App. E.2. The graph contains $N = 3000$ nodes and 12 planted groups of size 7–20, with 16-dimensional Gaussian node features. The learned target has four degree-32 Chebyshev heads and effective rank $q = 64$; its parameters are fitted only on separate training graphs. For each edge (u, v), we measure the squared difference between the normalized screened responses at its endpoints, $d^2_\Theta(u,v)$, and divide it by the mean cost over all graph edges. The held-out graph contains 644 within-group edges, 677 group-boundary edges, and 19 edges connecting two planted groups.

Separation on unseen groups. Figure 6 (top) illustrates the learned costs for four example groups. Their internal edges generally receive low costs, whereas edges leaving the groups receive visibly larger costs. This pattern also holds across the complete graph: the pooled distributions in Fig. 6 (bottom left) have median relative costs 0.088 for internal edges and 0.645 for boundary edges. The corresponding edge-level AUC is 0.971, meaning that a randomly selected boundary edge is assigned a larger cost than a randomly selected internal edge with probability 0.971.

The separation is consistent across individual groups. As shown in Fig. 6 (bottom right), every planted group has a larger mean boundary cost than mean internal cost. Mean internal costs range from 0.06 to 0.47, whereas mean boundary costs range from 0.75 to 1.56; consequently, even the largest group-level internal mean remains below the smallest boundary mean. The boundary-to-internal mean-cost ratio ranges from 3.3 to 16.7, with a median of 7.1.

Connection to greedy coarsening. The raw Ward merge score additionally normalizes the response contrast by the screened block-contrast norm. On singleton edges, it remains strongly aligned with the relative response cost: their Spearman correlation is 0.94, and the raw score separates boundary from internal edges with AUC 0.919. Thus, the response separation visible in Fig. 6 is largely preserved in the scores that determine the early merge order.

Strict separation of every internal and boundary edge is neither observed nor required. The block-level conditions in Sec. 6.1 allow local edge overlap provided that internal merges remain preferable along the relevant coarsening sequence. Indeed, the transferred stopping rule detects all 12 planted groups on this graph and obtains macro group F1 0.911. Group recall ranges from 0.67 to 1.00 and precision from 0.65 to 0.95, with means 0.89 and 0.79, respectively. These recovery results confirm that the learned cost contrast is sufficient to support accurate group formation despite a small overlap between the edge-level distributions.

Comparison with the static target. The rank-matched static eigenspace also contains structural information, but its edge ordering is less consistent. Its pooled relative-cost AUC is 0.852, compared with 0.971 for the learned target, and its median per-group AUC is 0.854, compared with 0.970. After Ward normalization, the corresponding AUCs are 0.783 and 0.919. Although the static target has a larger median ratio of group-level mean boundary to internal cost, that ratio is driven by a small number of large boundary values and does not reflect the substantial overlap between the full distributions. The learned target therefore provides a more reliable ordering of candidate merges across the graph, consistent with its stronger group-recovery results in App. E.2.

E.4 Elliptic++ temporal-transfer results

Data and temporal split. Elliptic++ contains Bitcoin wallet-address graphs over 49 temporal steps. We define a target gang as a connected component containing at least two known-illicit wallets. Unknown labels constitute 50–90% of the wallets on the evaluated days; these wallets remain in the graph but are excluded from classifier losses and evaluation metrics. Days 26–30 form the fitting period, and days 31–40 are reserved for out-of-period testing. Within the fitting period, known-label wallets are split stratifiably into 80% training and 20% validation subsets. The classifiers use a stratified half of the training subset, corresponding to 4,095 wallets or 40% of all known-label wallets on days 26–30. Absolute block-height attributes are removed to prevent calendar position from revealing the temporal split. The remaining wallet features are standardized using statistics fitted only on days 26–30 and then normalized by row. Additional construction and preprocessing details appear in App. E.1.

Methods and stopping rule. The learned Chebyshev filter bank is fitted using the closed form surrogate objective. Each daily graph is then processed by the same deflated Ward coarsener. coarsening+prediction trains a class-weighted ridge classifier using the wallet features X concatenated with the learned screened responses ℓ. A Ward supernode is declared illicit when a strict majority of its known-label members are predicted illicit. GNN+components instead applies a GCN node classifier and returns the connected components induced by wallets predicted illicit, without coarsening. Oracle coarsening uses the same Ward hierarchy but replaces predicted node labels with ground truth, providing an upper-bound diagnostic for the effect of node classification. Each Ward variant uses a single cumulative merge-score threshold shared across all days.

Evaluation metrics. B^3 precision and recall are computed over known-label wallets and give partial credit when only part of a gang is recovered. Matched-group evaluation is stricter: a predicted group and a target gang are matched only when more than 51% of the target lies in the prediction and more than 51% of the prediction's known members belong to that target. The resulting one-to-one matching defines group precision, group recall, and matched-group F1. Figure 3(b) reports the daily and pooled B^3 F1 scores, while Fig. 7 reports the corresponding matched-group F1 scores. Aggregate results and matched-gang counts are summarized in Table 1.

Temporal transfer. Table 1 shows that both learned classifiers deteriorate outside the fitting period, but the decline is substantially smaller for coarsening+prediction. Its B^3 F1 decreases from 0.607 to 0.416, a reduction of 0.191, whereas GNN+components falls from 0.650 to 0.204, a reduction of 0.446. The daily results in Fig. 3(center) show that coarsening+prediction exceeds GNN+components

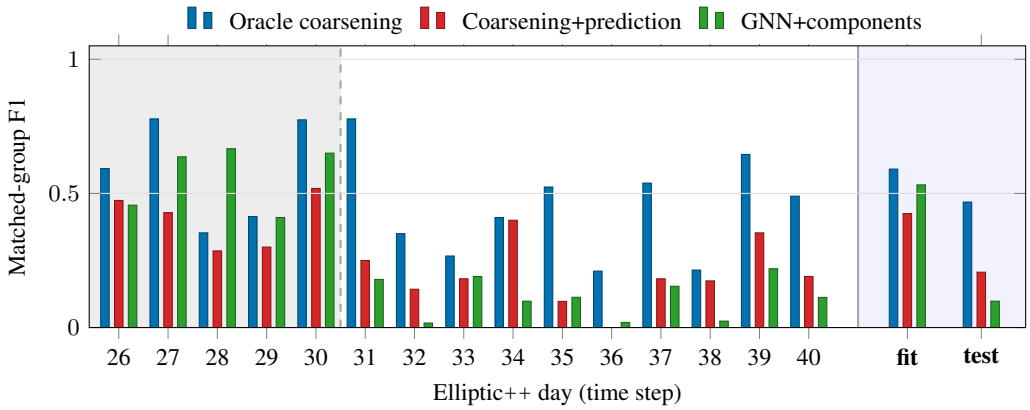


Figure 7: Matched-group F1 across Elliptic++ time steps. Days 26–30 (gray) form the fitting period, and days 31–40 are held-out temporal tests. The two groups of bars at right show results pooled over the fitting and test periods, respectively. Oracle coarsening uses true node labels, whereas coarsening+ prediction and GNN+components use predicted labels.

	Fitting period (days 26–30)			Test period (days 31–40)		
Method	B^3 F1	Group F1	Matched	B^3 F1	Group F1	Matched
Oracle coarsening	0.675	0.591	44	0.596	0.468	76
coarsening+prediction	0.607	0.425	24	0.416	0.206	29
GNN+components	0.650	0.532	46	0.204	0.098	43

Table 1: Aggregate Elliptic++ results over the fitting and held-out test periods. "Matched" is the number of target gangs satisfying the two-sided 0.51 overlap criterion, out of 70 fitting-period and 172 test-period gangs. Fitting-period group metrics are partly in-sample because they include wallets used to train the node classifiers.

in B^3 F1 on nine of the ten test days. Under the stricter criterion in Fig. 7, it leads on seven test days. The opposite aggregate ordering during the fitting period should be interpreted cautiously because those scores are partly in-sample.

Classification and coarsening error. As shown in Table 1, replacing ridge predictions with ground-truth node labels raises test B^3 F1 from 0.416 to 0.596. This 0.180 difference shows that node classification is an important source of error. Because Oracle coarsening and coarsening+prediction use separately selected stopping thresholds, however, the difference is a diagnostic rather than an exact additive decomposition. Even Oracle coarsening remains below perfect recovery, indicating residual error from the coarsening hierarchy and from using one stopping threshold across temporally different graphs.

The difference between approximate membership and near-whole recovery is also substantial. Table 1 reports test B^3 F1 versus matched-group F1 of 0.596 versus 0.468 for Oracle coarsening, 0.416 versus 0.206 for coarsening+prediction, and 0.204 versus 0.098 for GNN+components. Thus, all three methods often place a meaningful fraction of a gang together without satisfying the stricter two-sided matching criterion.

Output selectivity and difficult days. Although GNN+components matches 43 test gangs compared with 29 for coarsening+prediction in Table 1, it produces 3,950 test groups rather than 197. Its matched-group precision is consequently only 0.061, compared with 0.266 for coarsening+prediction. The coarsened method therefore returns fewer but substantially more selective candidate groups.

The daily results also reveal pronounced temporal variation. In Fig. 3(b), coarsening+prediction reaches its highest test B^3 F1 on day 33 (0.784) and its lowest on day 36 (0.014). Figure 7 similarly shows no near-whole gang recovery for this method on day 36. Oracle coarsening also reaches its lowest daily B^3 F1 on day 36 (0.337), indicating that this time step is difficult even when node-classification error is removed.

Table 2: Dataset statistics and learned-coarsening configurations. Communities are split equally between training and testing. All methods use degree-32 Chebyshev filters fitted in closed form, followed by deflated Ward coarsening. The Amazon communities are deduplicated before splitting. For email-Eu-core, the reported results average five independent splits.

Dataset	Nodes	Edges	Communities	Test	Splits	Features	Heads	τ
com-Amazon	334,863	925,872	1,517	759	1	Random, 32	4	0.5
com-DBLP	317,080	1,049,866	5,000	2,500	1	Structural, 9	16	16
email-Eu-core	1,005	16,064	40	20	5	Structural, 9	16	32

E.5 Additional community-recovery results

Datasets. We evaluate community recovery on three networks with different scales and forms of ground truth. The Amazon co-purchase graph represents products as nodes and co-purchases as edges; its communities are product categories. The DBLP coauthorship graph represents authors connected by coauthorship, with publication venues providing overlapping communities. Both datasets include the 5,000 top-quality communities distributed by SNAP (Yang & Leskovec, 2012). The email-Eu-core graph represents email communication between members of a European research institution, with organizational departments used as disjoint ground-truth communities.

The Amazon community file contains many exact duplicate node sets: only 1,517 of its 5,000 entries are distinct, and one set is repeated 48 times. Indeed, a direct random split places 1,937 of the 2,500 test entries in exact correspondence with a training entry. We therefore retain only the first occurrence of each distinct node set before splitting, leaving 758 training and 759 test communities. The retained communities have median size 10 and maximum size 328. They remain genuinely overlapping: approximately 40% share nodes with another community, and 38% of their member nodes belong to more than one community. DBLP uses a seeded 2,500/2,500 split of its supplied communities. For email-Eu-core, we report averages over five seeded splits, each containing 20 training and 20 test departments.

Amazon provides no node attributes, so we use 32 independent standardized Gaussian channels as a structural range finder. For DBLP and email-Eu-core, we use nine standardized, label-free structural features: degree, log-degree, triangle count, clustering coefficient, mean neighbor degree, core number, and the two-, three-, and four-step random-walk return probabilities. Table 2 summarizes the datasets and model configurations.

Evaluation protocol. For each dataset, the learned representation is fitted using only the training communities. Deflated Ward then constructs a complete hierarchy, beginning with singleton nodes and continuing to one supernode. A single operating point is selected by maximizing macro group F1 on the training communities, and all reported scores are computed on held-out communities. This protocol evaluates both the quality of the learned hierarchy and the ability of the stopping rule to transfer without test labels.

We compare with Louvain (Blondel et al., 2008), using the NetworkX implementation with seed zero. Louvain produces a multilevel dendrogram, and its operating level is selected using the same training-F1 rule. At each cut, a target community is compared with the predicted group containing the largest number of its nodes. It is counted as detected when both its group precision and recall exceed 0.5. Macro group F1 gives equal weight to every target community, whereas B^3 F1 evaluates membership at the node level. For overlapping Amazon and DBLP ground truth, B^3 uses the multiplicity-corrected extension (Amigó et al., 2009).

Comparison with Louvain. Table 3 shows that the learned coarsening is strongest on Amazon. It achieves nearly the same precision as Louvain (0.941 versus 0.950), while increasing recall from 0.832 to 0.907. Consequently, macro group F1 rises from 0.861 to 0.907, and B^3 F1 from 0.686 to 0.774. The learned cut detects 699 of the 759 held-out communities, 62 more than Louvain. It also achieves this with 20,496 output groups rather than 53,145, showing that the improvement is not obtained by fragmenting the graph into a finer partition.

On DBLP, the two methods have nearly identical macro group F1: 0.685 for learned coarsening and 0.694 for Louvain. Their operating characteristics are complementary. Learned coarsening has substantially higher precision (0.862 versus 0.756), while Louvain favors recall (0.732 versus 0.627).

Table 3: Held-out community recovery at the training-selected global cut. Precision, recall, and group F1 are macro averages over test communities. "Detected" reports communities satisfying the two-sided 0.51 criterion, and "Output groups" is the number of blocks at the selected partition. Bold denotes the better value within each dataset. Email-Eu-core counts are summed over five splits; its remaining metrics are averaged across splits (group-F1 standard deviation: 0.015 for learned coarsening and 0.049 for Louvain).

Dataset	Method	Precision	Recall	Group F1	Detected	B^3 F1	Output groups
com-Amazon	Ours	0.941	**0.907**	**0.907**	**699/759 (92.1%)**	**0.774**	20,496
	Louvain	**0.950**	0.832	0.861	637/759 (83.9%)	0.686	53,145
com-DBLP	Ours	**0.862**	0.627	0.685	1,410/2,500 (56.4%)	0.266	113,756
	Louvain	0.756	**0.732**	**0.694**	**1,499/2,500 (60.0%)**	**0.314**	61,523
email-Eu-core	Ours	**0.594**	0.457	**0.439**	**27/100 (27.0%)**	0.345	166
	Louvain	0.320	**0.792**	0.378	26/100 (26.0%)	**0.563**	38

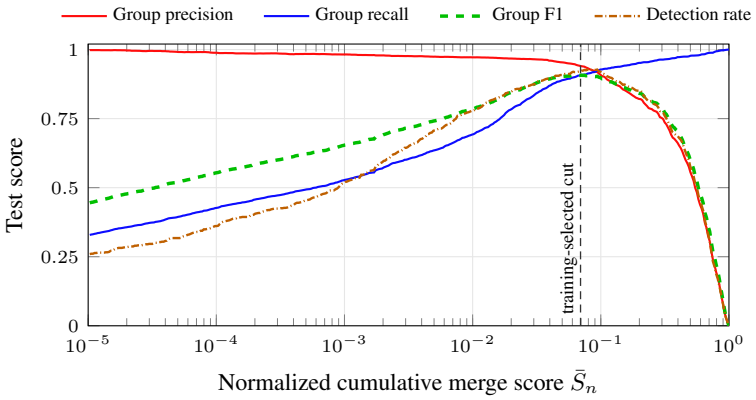


Figure 8: Held-out Amazon performance along the learned coarsening hierarchy, indexed by the normalized cumulative merge score $\bar{S}_n = \sum_{k \le n} s_k / \sum_k s_k$ on a logarithmic scale. Group F1 (green), detection rate (brown), group precision (red), and group recall (blue) are evaluated on the 759 test communities. The vertical line marks the cut selected using training communities ($\bar{S}_n = 0.070$, 20,496 supernodes). Values below 10^{-5} are omitted from the displayed horizontal range.

The small 0.009 group-F1 difference therefore reflects a precision–recall tradeoff rather than a failure to recover useful communities.

The learned coarsening obtains the higher macro group F1 on email-Eu-core (0.439 versus 0.378), along with substantially higher precision (0.594 versus 0.320) and a slightly larger number of detected departments (27 versus 26 across the five splits). Louvain's much coarser partition contains only 38 groups, close to the 40 departments, and consequently favors recall and node-weighted B^3 F1. In contrast, the learned cut returns smaller, more selective candidates and performs better when each target department is weighted equally through macro group F1.

Amazon hierarchy and stopping behavior. Figure 3(right) reports Amazon precision, recall, and F1 against the restricted spectral-approximation error, while Fig. 8 displays the same hierarchy against its normalized cumulative merge score. The latter does not require evaluating a spectral error and shows how recovery evolves at every merge level. Test F1 increases steadily over several orders of magnitude and reaches its maximum near the training-selected cut. A post-hoc test-label oracle changes the number of detected communities only from 699 to 701, demonstrating that the training-selected stopping rule transfers accurately to the held-out communities.

The remaining misses at this single cut are concentrated among the largest categories. Detection rates are 98% for communities of size 2–10, 96% for sizes 11–20, 85% for sizes 21–50, and 51% for sizes 51–100. The selected cut therefore performs especially well at the scale containing most Amazon categories. More importantly, all 759 held-out communities become detectable at some level of the complete learned hierarchy, compared with 755 for the five-level Louvain dendrogram. Thus, the learned representation preserves recoverable versions of every held-out category; the few misses at the global cut reflect the requirement that one resolution serve communities of widely different sizes.

Information retained across the hierarchy. The DBLP results show a similar distinction between hierarchy quality and single-cut performance. Although Louvain has a small advantage at the selected

global cut, 73.6% of the held-out DBLP communities are detectable somewhere in the learned hierarchy, compared with 64.9% across the Louvain dendrogram. Moreover, selecting the learned cut with test labels would detect 1,418 communities rather than 1,410 under the training-selected cut. The stopping rule therefore transfers almost optimally; the principal limitation is the use of one global resolution for communities spanning a wide range of sizes and connectivities.

Across the three benchmarks, learned coarsening attains the higher macro group F1 on Amazon and email-Eu-core and comes within 0.009 of Louvain on DBLP, while providing markedly higher precision on DBLP and email-Eu-core. Beyond the selected partition, its complete hierarchy retains more recoverable community structure than the much shallower Louvain dendrogram on Amazon and DBLP. These results show that the learned coarsening is not merely a substitute for a modularity partition: it provides a fine-grained, transferable hierarchy from which operating points can be selected according to the desired balance between purity, coverage, and resolution.