Bioinformatics, 2026, pp. 1–7

doi:
Proceedings paper

PROCEEDINGS PAPER

SoftBlock: Transferable Soft Blocks for Overlapping Protein Complex Recovery under a Frozen Protocol

Haziq Jeelani[1,*] and Fayeq Jeelani Syed[2]

[1]Claremont Graduate University, United States and [2]Indiana University, United States; faysyed@iu.edu
[*]Corresponding author. haziq.jeelani@cgu.edu

Abstract

Motivation: Protein complexes overlap, and protein–protein interaction networks vary substantially across experimental sources. This often pushes complex-recovery pipelines toward graph-specific tuning, making transfer and cross-dataset comparison brittle.
Results: We present SoftBlock, a coarse-to-fine pipeline that learns soft memberships once on a reference interaction network (STRING) and transfers them as overlapping blocks to other networks. Within each block we run Markov clustering, then consolidate candidates with graph-only scoring and diversity-aware top-N selection. Under a strict frozen protocol (tune on STRING, freeze elsewhere), we evaluate on seven human interaction networks against CORUM and Complex Portal gold standards. SoftBlock performs consistently across graphs, outperforming strong overlap baselines and remaining competitive with PCGAN under matched conditions.
Availability and Implementation: Code is available at `https://github.com/haziqjeelani/paper_repos/tree/main/2026/SoftBlock`.
Contact: haziq.jeelani@cgu.edu
Supplementary Information: Supplementary data are available online.

Key words: protein complexes, PPI networks, overlapping communities, transferability, frozen protocol

Introduction

Protein complexes are fundamental functional units, and protein–protein interaction (PPI) networks are a common substrate for recovering them. The challenge is not only overlap—proteins participate in multiple complexes—but also dataset shift: PPIs are noisy and incomplete, and assay-specific biases (e.g., affinity purification–mass spectrometry (AP–MS), yeast two-hybrid (Y2H), or curated compilations) change graph structure. Methods tuned to one benchmark can therefore drift when moved to a new PPI.

Evaluation is also sensitive to operating point. Best-match scores can improve simply by emitting more clusters, so comparisons can become hard to interpret. Fixed-cap operating points (top-N predictions) and frozen protocols—tune once and avoid per-graph retuning—help keep comparisons defensible.

SoftBlock is built around a simple hypothesis: learn a coarse, transferable prior once on a rich reference PPI, then reuse it to constrain overlap-aware complex recovery on other graphs. Concretely, we compute SoftBlock soft memberships on STRING, transfer them as overlapping blocks on each target PPI, run MCL within blocks to enumerate candidates, and then deduplicate, rerank, and select a fixed-size, non-redundant set of predictions.

Contributions.

- A transferable soft-membership prior: SoftBlock memberships computed once on a reference PPI are reused as overlapping blocks across target PPIs.
- A practical overlap-aware coarse-to-fine pipeline: block-local solving (MCL (Enright, 2002; van Dongen, 2000)), global deduplication, graph-only reranking, and a diversity-aware top-N selector for an explicit operating point.
- A strict frozen-protocol evaluation: tune on STRING and freeze elsewhere across seven PPIs and two gold standards, using fixed-cap operating points and OS metrics (Brohée and van Helden, 2006), plus a matched-protocol comparison to PCGAN (Pan et al., 2023).

Related Work

Overlap complex detection.
Early complex discovery methods typically look for dense subgraphs or locally cohesive regions, including MCODE (Bader and Hogue, 2003) and ClusterONE (Nepusz et al., 2012). Overlap-aware community methods represent overlap more explicitly (e.g., OSLOM2 (Lancichinetti et al., 2011), SLPA (Xie et al., 2011), link communities (Ahn et al., 2010), BigCLAM (Yang and Leskovec, 2012)). Their performance

can be sensitive to operating point and graph-specific hyperparameters.

Learning-based complex prediction.

More recent approaches bring supervision or generative modeling into the pipeline (e.g., Super.Complex (Palukuri and Marcotte, 2021), reinforcement learning (Palukuri et al., 2023), and PCGAN (Pan et al., 2023)). These methods can be strong on specific graphs and protocols, but comparisons across papers are hard without matching the graph construction, the gold set snapshot, and the evaluation operating point.

Evaluation and operating-point control.

Complex prediction is commonly evaluated with best-match scores and overprediction-aware metrics such as Sn/PPV/Accuracy and MMR (Brohée and van Helden, 2006). Because the number of predicted clusters drives the precision/recall trade-off, fixed-cap operating points (top-N clusters) are a practical way to keep comparisons interpretable.

Methods

Overview.

SoftBlock turns a reference PPI with rich evidence (STRING) and a target PPI $G = (V, E)$ into a set of overlapping complex predictions in three stages: (i) compute a transferable soft-membership prior on the reference graph, (ii) use that prior to define overlapping blocks on the target graph and run a local solver inside each block to generate candidates, and (iii) consolidate and select a fixed operating point using graph-only scores and overlap suppression. Figure 1 summarizes the pipeline.

SoftBlock Soft Memberships as Transferable Block Priors

On the reference PPI $G_{\text{ref}} = (V_{\text{ref}}, E_{\text{ref}})$, SoftBlock produces a row-stochastic soft membership matrix $R \in [0,1]^{|V_{\text{ref}}| \times K}$. We compute Louvain communities (Blondel et al., 2008) on G_{ref} (resolution = 1.0) and take the K largest as prototype node sets $\{C_k\}_{k=1}^{K}$. A 1-layer GCN encoder (Kipf and Welling, 2017) trained with Deep Graph Infomax (Veličković et al., 2019) (TSVD-64 node features; 512-d embeddings; full-batch Adam, lr $= 10^{-3}$, weight decay $= 5 \times 10^{-3}$, 300 epochs with patience 50; corruption permutes node-feature rows with adjacency fixed) outputs ℓ_2-normalized embeddings z_v; prototype vectors are centroids $\mu_k = |C_k|^{-1} \sum_{v \in C_k} z_v$, and memberships use a temperature-softmax over dot products: $R_{vk} = \text{softmax}_k(\tau\, z_v^\top \mu_k)$. We use $\tau = 30$ and one soft k-means refinement step (initialized by the Louvain prototypes), both selected on STRING and then frozen.

To transfer this prior to a target graph $G = (V, E)$, we restrict R to the node overlap $V \cap V_{\text{ref}}$ and interpret each column k as a coarse block membership function. We optionally apply a frozen calibration by sharpening/flattening each row: for a power $a > 0$, we compute $\tilde{R}_{vk} \propto R_{vk}^a$ and renormalize so $\sum_k \tilde{R}_{vk} = 1$.

Overlapping Coarse Blocks

We convert soft memberships into overlapping blocks by assigning each node v to multiple modules. In the default top-k rule, we assign v to the indices of its largest k membership values (optionally thresholded by $\tilde{R}_{vk} \geq p_{\min}$), yielding blocks $\{B_1, \ldots, B_K\}$ with substantial overlap. Overlap at the block level enables downstream recovery of overlapping complexes.

Coarse-to-Fine Local Solving Inside Blocks

For each block B_k, we extract the induced subgraph $G[B_k]$ and run a local complex finder. Our default solver is MCL (Enright, 2002; van Dongen, 2000), which is widely used for PPI clustering and works well as a dense-subgraph enumerator inside blocks. To isolate the contribution of the transfer prior (rather than MCL alone), we also run wrapper variants that swap the local solver inside each block (e.g., SLPA (Xie et al., 2011) or a ClusterONE-like greedy growth (Nepusz et al., 2012)) while keeping the rest of the protocol fixed. We then union candidates across blocks and deduplicate near-duplicates with a Jaccard-based merge.

Multiscale & Multi-K Union

To increase robustness without per-graph tuning, we union candidate clusters across multiple prototype counts $K \in \{6, 8, 16\}$ by training R on STRING separately for each K and merging the resulting candidate pools (optionally also across a small set of membership calibrations). All such choices are selected once on STRING and frozen for all transfer PPIs.

Graph-Only Reranking and Diversity-Aware Top-N Selection

From the candidate pool, we select a defensible operating point. We score each candidate cluster using only the target graph, e.g., (weighted) density and cohesiveness. For a cluster C with $n = |C|$ and internal weight sum $W_{\text{in}}(C)$, weighted density is $W_{\text{in}}(C)/\binom{n}{2}$; a boundary-aware cohesiveness score is $2W_{\text{in}}(C)/(2W_{\text{in}}(C) + W_{\text{cut}}(C))$, where $W_{\text{cut}}(C)$ sums weights of edges leaving C. We optionally train a lightweight linear reranker on STRING using only graph-derived cluster features and gold supervision; however, unless explicitly stated, we use a graph-only scorer to keep the evaluation focused on transfer and operating-point control.

To avoid spending the top-N budget on redundant micro-clusters, we apply a simple diversity filter after sorting by score: we greedily keep a candidate only if its Jaccard overlap with all previously selected clusters is below a threshold (default 0.5). This overlap-suppression heuristic is intentionally simple, but it makes fixed-cap comparisons substantially more stable.

Hybrid Candidate Augmentation (Link Communities)

Because link communities (Ahn et al., 2010) can be strong on STRING (and can also contribute candidates outside $V \cap V_{\text{ref}}$), we evaluate a hybrid recipe that unions SoftBlock candidates with link-communities candidates and reranks the combined pool. To avoid per-dataset branching, the frozen `hybrid_auto` gate is deterministic: we include link-communities candidates iff the target graph is unweighted (keeping clusters with $|C| \geq 4$) or weighted with confidence-like weights in $[0,1]$ (keeping clusters with $|C| \geq 3$); otherwise we disable link communities and use blocks-only candidates.

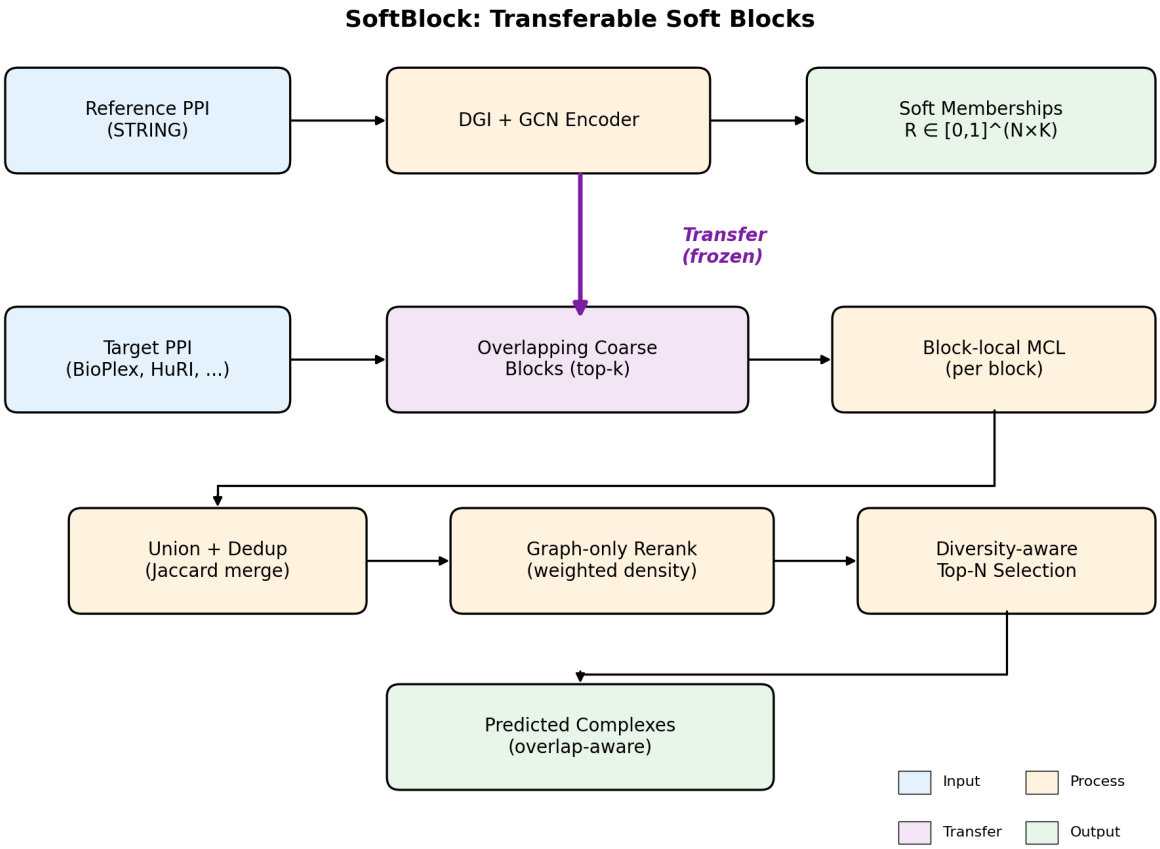

Fig. 1. SoftBlock pipeline: transferable soft blocks from a reference PPI (STRING) enable overlap-aware coarse-to-fine complex recovery on diverse target PPIs under a frozen protocol.

Table 1. Human PPI graphs and gold standards. Gold complexes are intersected to each graph node set (minimum complex size = 3).

Graph	Nodes	Edges	CORUM	CP
STRING	15 882	236 712	1317	1198
BioPlex	13 923	118 144	1164	961
HuRI	8 109	51 686	537	616
BioGRID	27 590	1 002 631	1355	1157
ComPPI	15 277	170 728	1310	1158
IntAct	17 733	527 860	1315	1211
hu.MAP2	7 824	19 631	848	730

Experimental Setup

Graphs and Gold Standards

We evaluate on seven human PPIs: STRING (Szklarczyk et al., 2018), BioPlex (Huttlin et al., 2015), HuRI (Luck et al., 2020), BioGRID (Oughtred et al., 2020), IntAct (del Toro et al., 2021), ComPPI (Veres et al., 2014), and hu.MAP2 (Drew et al., 2021). Gold standards are CORUM (Giurgiu et al., 2018) and Complex Portal (CP) (Meldal et al., 2014), intersected to each graph's node universe with a minimum complex size of 3. Table 1 summarizes graph statistics.

Metrics

We report (i) mean best-match F1 against CORUM and Complex Portal, and (ii) overprediction-aware OS metrics (Sn/PPV/Accuracy and greedy maximum matching ratio (MMR)) (Brohée and van Helden, 2006). When we report a single number "averaged over CORUM and CP," we compute the metric separately for each gold standard on the same predictions and then take the arithmetic mean (we do not merge the gold sets).

Best-match F1.

For a gold complex G and a predicted cluster C, let precision be $|G \cap C|/|C|$ and recall be $|G \cap C|/|G|$. Best-match F1 assigns each gold complex the maximum F1 over all predicted clusters with at least a small overlap (we use $|G \cap C| \geq 2$), and reports the mean over gold complexes.

OS metrics.

We use the overlap score $\mathrm{OS}(G, C) = |G \cap C|^2/(|G||C|)$. Sensitivity Sn is the mean over gold complexes of the best overlap score ($\max_C \mathrm{OS}(G, C)$), and positive predictive value PPV is the mean over predicted clusters of the best overlap score ($\max_G \mathrm{OS}(G, C)$). Accuracy is $\sqrt{\mathrm{Sn} \cdot \mathrm{PPV}}$. Greedy maximum matching ratio (MMR) approximates the maximum-weight matching ratio on the bipartite graph between gold complexes and predicted clusters.

Table 2. Frozen protocol performance (uncapped candidate pools). We report #PC and mean best-match F1 on CORUM (C) and Complex Portal (CP). Best Baseline is selected by mean best-match F1 averaged over C/CP; it is link communities on all graphs except HuRI (ClusterONE).

	SoftBlock			Best Baseline		
Graph	#PC	$F1_C$	$F1_{CP}$	#PC	$F1_C$	$F1_{CP}$
BioGRID	4560	.301	.371	7940	.182	.174
BioPlex	8000	.386	.475	7838	.340	.425
ComPPI	8000	.392	.418	7984	.355	.367
HuRI	4733	.236	.295	1530	.096	.141
IntAct	8000	.221	.270	7974	.186	.240
hu.MAP2	2641	.473	.592	1227	.389	.485
STRING	8000	.513	.553	7983	.493	.508

Frozen Protocol

All hyperparameters (block assignment, solver parameters, reranking score, and the cap N) are tuned on STRING only. The resulting configuration is then applied without modification to all other PPIs. Unless otherwise stated, SoftBlock uses top-k block assignment ($k = 4$, $p_{\min} = 0$), membership calibration $a = 1.5$, multi-K union (separate reference memberships for $K \in \{6, 8, 16\}$), and MCL inflation 4.0, followed by global deduplication (Jaccard 0.85; union-merge near-duplicates) and graph-only weighted-density reranking. For the large-pool table we keep up to 8000 clusters; for fixed-cap comparisons we use $N = 2000$ with overlap suppression (greedy NMS, max Jaccard < 0.5). We rerun across three random seeds (42/43/44) and report mean ± std and significance checks in the Supplementary Material.

Baselines

We compare to strong baselines under the same frozen protocol, including MCODE and ClusterONE (Bader and Hogue, 2003; Nepusz et al., 2012) and overlap-aware methods (OSLOM2, SLPA, link communities, BigCLAM) (Lancichinetti et al., 2011; Xie et al., 2011; Ahn et al., 2010; Yang and Leskovec, 2012). For each baseline, we select its hyperparameters on STRING only and reuse the same setting unchanged across all transfer PPIs. We also evaluate learning-based predictors on hu.MAP2 (Super.Complex and an RL predictor) (Palukuri and Marcotte, 2021; Palukuri et al., 2023) under the same evaluator.

Results

Main Cross-PPI Results Under a Frozen Protocol

Uncapped candidate pools.

Best-match scores can improve simply by emitting more candidate complexes. We therefore report the uncapped prediction count (#PC) in Table 2; this should be interpreted together with the fixed-cap comparisons (Table 3).

Fixed-cap comparisons.

Table 3 uses a shared target cap ($N = 2000$) with overlap suppression; #PC reports the realized output size, so some baselines are cap- or NMS-limited and output fewer than N predictions. HuRI is the most extreme case: the best baseline yields 417 clusters; at a rate-matched SoftBlock operating point ($N = 417$) we obtain Acc = .035, MMR = .018, and F1 = .043 (mean over seeds 42/43/44). Figure 2 summarizes the gap to the best baseline across graphs.

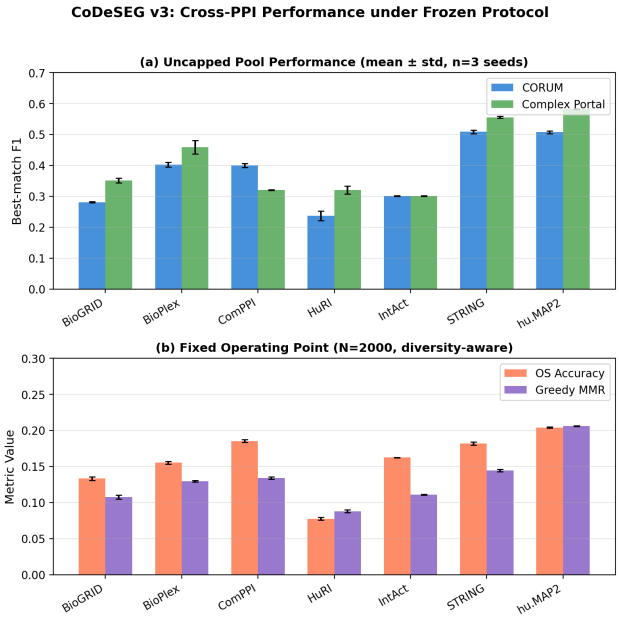

Fig. 2. SoftBlock cross-PPI performance under the frozen protocol. (a) Best-match F1 against CORUM and Complex Portal gold standards. (b) Operating-point metrics at $N = 2000$ with diversity-aware selection.

Operating-Point Control with Diversity-Aware Selection

Why diversity-aware selection matters.

Fixed-cap selection by score alone can spend the budget on many near-duplicate micro-clusters, especially on sparse PPIs and under wrapper solvers. The overlap-suppression step (max Jaccard < 0.5) is a simple way to make the operating point less brittle without touching upstream candidate generation.

Operating-Point Sensitivity

Sweeping the cap and overlap threshold.

The frozen protocol fixes both the cap N and the overlap-suppression threshold (max Jaccard) on STRING. To check that our cross-PPI comparisons are not an artifact of a particular operating point, we sweep N (with max Jaccard fixed at 0.5) and sweep max Jaccard (with N fixed at 2000) on two contrasting graphs: BioGRID (dense, weighted) and HuRI (sparse, unweighted). Figure 3 shows the sweep curves; Table 4 summarizes ranks across sweep settings. SoftBlock is stable on BioGRID (top-ranked across settings) and remains competitive

Table 3. Frozen operating point (target cap $N = 2000$, diversity-aware selection). #PC is the realized output count after overlap suppression; we report averaged OS Accuracy, MMR, and best-match F1 across CORUM and Complex Portal. Best Baseline (selected by mean best-match F1 averaged over CORUM/CP) is link communities on all graphs.

	SoftBlock				Best Baseline			
Graph	#PC	Acc	MMR	F1	#PC	Acc	MMR	F1
BioGRID	2000	.150	.125	.299	2000	.135	.078	.151
BioPlex	2000	.161	.121	.322	2000	.157	.114	.303
ComPPI	2000	.186	.139	.281	1976	.168	.123	.253
HuRI	2000	.109	.086	.229	417	.028	.009	.017
IntAct	1905	.136	.089	.176	1681	.119	.069	.132
hu.MAP2	1385	.210	.146	.492	836	.196	.101	.426
STRING	2000	.207	.163	.444	2000	.192	.150	.416

Table 4. Sensitivity of the frozen operating point to the cap N and overlap-suppression threshold. We summarize the rank of SoftBlock among all evaluated methods under OS Accuracy and greedy MMR across the sweep settings (lower rank is better).

Graph	Sweep	#settings	Top1 Acc	Med rank	Worst	Top1 MMR	Worst
BioGRID	cap sweep	4	4	1.0	1	4	1
BioGRID	Jaccard sweep	4	3	1.0	2	4	1
HuRI	cap sweep	4	0	2.0	2	1	2
HuRI	Jaccard sweep	4	0	2.0	2	1	2

Table 5. Node overlap with the reference graph (STRING) and uncapped frozen-protocol performance. Overlap is $|V \cap V_{\text{ref}}|/|V|$ where V_{ref} are STRING nodes. Performance is shown for SoftBlock (`hybrid_auto`) using mean best-match F1 and mean greedy MMR (averaged over CORUM and Complex Portal).

Graph	Nodes	Overlap	Mean F1	Mean MMR
BioGRID	27590	53.2%	0.336	0.172
BioPlex	13923	79.0%	0.431	0.239
IntAct	17733	82.7%	0.246	0.158
HuRI	8109	83.6%	0.265	0.117
hu.MAP2	7824	84.6%	0.532	0.200
ComPPI	15277	86.8%	0.405	0.251
STRING	15882	100.0%	0.533	0.292

on HuRI (within the top two by OS Accuracy; top-two by MMR).

Transfer vs Node Overlap with STRING

SoftBlock transfers a block prior defined on $V \cap V_{\text{ref}}$, so it is natural to ask how performance changes as that overlap shrinks. Table 5 reports the overlap fraction for each graph along with uncapped mean best-match F1 and mean greedy MMR (averaged over CORUM and Complex Portal). Figure 4 visualizes the same relationship and compares SoftBlock to the strongest baseline on each graph.

Runtime and Scalability

Additional Gold: hu.MAP2 Complexes

To complement curated gold standards (CORUM, Complex Portal), we also treat the hu.MAP2 complex catalog as an additional gold standard on the hu.MAP2 graph. We compare to a published RL predictor and to SLPA, and we report a wrapper variant that runs SLPA inside transferred SoftBlock blocks (Table 7).

PCGAN-Protocol Head-to-Head on CPIN-H

We compare to PCGAN (Pan et al., 2023) under their published protocol on CPIN-H (Table 8). At a matched #PC, SoftBlock yields a slightly higher F-measure and MMR, trading lower recall for higher precision.

Split sensitivity.

Table 8 follows PCGAN's published split, which is induced by CORUM file order. To assess sensitivity, we reshuffle the CORUM list 100 times and recompute the split. Across shuffles, PCGAN varies substantially; the mean ΔF-measure (SoftBlock − PCGAN) is 0.060 [0.049, 0.070] (95% CI).

Discussion

What we can (and cannot) claim.

We avoid broad "SOTA" claims across mismatched datasets and protocols. The claim we are comfortable making is narrower: under the frozen protocol (tune on STRING, freeze elsewhere), SoftBlock is consistently competitive with (and often better than) the strong overlap baselines we evaluated, and it remains competitive with PCGAN when evaluated under matched conditions.

Runtime and scalability.

SoftBlock runs a local solver (MCL) inside many overlapping blocks, so runtime is dominated by this block-local stage. Even on the largest graph (BioGRID, ∼1M edges), end-to-end hybrid_auto inference remains tractable under the frozen configuration (Table 6). Additional diagnostics are in the Supplementary Material.

Biological plausibility beyond benchmarks.

Gold standards are incomplete and can miss compositional variants of known assemblies. To complement CORUM/Complex Portal matching, we provide GO-slim enrichment case studies for high-scoring predictions on STRING that are unmatched under our overlap-score criterion (max OS < 0.2). Representative examples show coherent cellular-component enrichment (see Supplementary).

Limitations.

First, a transferable prior anchored on STRING may underperform on target graphs with limited node overlap

Table 6. Runtime and memory footprint (peak RSS) for SoftBlock (`hybrid_auto`) and the strongest baseline per graph. We report wall-clock time in minutes for the main pipeline stages and the total, along with peak resident memory (GiB). Timings exclude CORUM/ComplexPortal evaluation. Stage abbreviations: MCL = block-local Markov Clustering; LC = link communities (when used); Ens = multi-K union; Hyb = hybrid union/rerank.

Graph	Nodes	Edges	MCL	LC	Ens	Hyb	Total	Peak GiB	Best baseline (min / GiB)
BioGRID	27590	1002631	2.25	0.00	0.13	0.08	2.45	2.76	Link communities (0.63 / 0.86)
IntAct	17733	527860	1.42	0.24	0.09	0.24	1.99	1.91	Link communities (0.24 / 0.78)
STRING	15882	236712	0.86	0.20	0.22	0.26	1.54	0.95	Link communities (0.20 / 0.74)
ComPPI	15277	170728	1.03	0.17	0.17	0.36	1.73	2.10	Link communities (0.17 / 0.73)
BioPlex	13923	118144	0.53	0.12	0.10	0.16	0.90	0.78	Link communities (0.12 / 0.72)
HuRI	8109	51686	0.30	0.06	0.06	0.08	0.51	0.75	ClusterONE (0.22 / 0.71)
hu.MAP2	7824	19631	0.32	0.06	0.04	0.05	0.47	0.74	Link communities (0.06 / 0.71)

Table 7. hu.MAP2 complexes as gold standard on hu.MAP2 graph. All capped methods use $N = 2000$ with diversity-aware selection.

Method	#PC	F1	Acc	MMR
Published RL predictor (capped)	2000	.483	.377	.201
SLPA baseline	1387	.553	.528	.227
SoftBlock + wrapper SLPA	2000	.567	.502	.277

Table 8. PCGAN-protocol head-to-head on CPIN-H (human): Rate-match Recall/Precision/F-measure and MMR at matched #PC = 1942.

Method	#PC	Rec	Prec	F	MMR
SoftBlock + graph-only	1942	.241	.257	.249	.063
PCGAN (published)	1942	.288	.213	.245	.058

or markedly different interaction semantics. Our block prior is defined on $V \cap V_{\text{ref}}$, so recovering complexes involving proteins absent from STRING remains limited unless additional candidate sources (e.g., link communities) are enabled; Figure 4 and Table 5 summarize this dependence. Second, comparisons across papers remain challenging due to differences in graph construction, gold-set versions, and metric definitions; we therefore emphasize like-for-like head-to-heads.

Reproducibility

Code and tracked artifacts are available at: `https://github.com/haziqjeelani/paper_repos/tree/main/2026/SoftBlock`. The main frozen-protocol tables can be reproduced with the scripts under `paper_v3/src/`; the default entry point is `make -C paper_v3 operating-point SEED=42`. Additional analyses are provided in the Supplementary Material.

Conclusion

SoftBlock shows that a transferable soft-membership prior, paired with a simple local solver and explicit operating-point control, can deliver robust overlapping complex recovery across diverse human PPIs under a strict frozen protocol.

Data Availability

Code to reproduce the main tables and figures is available at `https://github.com/haziqjeelani/paper_repos/tree/main/2026/SoftBlock`. The PPI networks and gold standards used in this work are publicly available from the cited sources; the repository includes scripts to download/fetch and preprocess inputs, along with pinned CORUM/Complex Portal snapshots used for the paper benchmark under `paper_v3/data/`.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no competing interests.

Acknowledgements

None.

References

Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307): 761–764, 2010. doi: 10.1038/nature09182.

G. D. Bader and C. W. V. Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4(1), 2003. doi: 10.1186/1471-2105-4-2.

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008, 2008. doi: 10.1088/1742-5468/2008/10/p10008.

S. Brohée and J. van Helden. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*, 7(1), 2006. doi: 10.1186/1471-2105-7-488.

N. del Toro, A. Shrivastava, E. Ragueneau, B. Meldal, C. Combe, E. Barrera, L. Perfetto, K. How, P. Ratan, G. Shirodkar, O. Lu, B. Mészáros, X. Watkins, S. Pundir, L. Licata, M. Iannuccelli, M. Pellegrini, M. J. Martin, S. Panni, M. Duesbury, S. D. Vallet, J. Rappsilber, S. Ricard-Blum, G. Cesareni, L. Salwinski, S. Orchard, P. Porras, K. Panneerselvam, and H. Hermjakob. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Research*, 50(D1):D648–D653, 2021. doi: 10.1093/nar/gkab1006.

K. Drew, J. B. Wallingford, and M. E. M. Bhatt. hu.MAP 2.0: integration of over 15,000 proteomic experiments builds a global compendium of human multiprotein assemblies. *Molecular Systems Biology*, 17(5), 2021. doi: 10.15252/msb.202010016.

A. J. Enright. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30(7):1575–1584, 2002. doi: 10.1093/nar/30.7.1575.

M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, and A. Ruepp. CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Research*, 47(D1):D559–D563, 2018. doi: 10.1093/nar/gky973.

E. L. Huttlin, L. Ting, R. J. Bruckner, F. Gebreab, M. P. Gygi, J. Szpyt, S. Tam, G. Zarraga, G. Colby, K. Baltier, R. Dong, V. Guarani, L. P. Vaites, A. Ordureau, R. Rad, B. K. Erickson, M. Wühr, J. Chick, B. Zhai, D. Kolippakkam, J. Mintseris, R. A. Obar, T. Harris, S. Artavanis-Tsakonas, M. E. Sowa, P. De Camilli, J. A. Paulo, J. W. Harper, and S. P. Gygi. The BioPlex network: A systematic exploration of the human interactome. *Cell*, 162(2):425–440, 2015. doi: 10.1016/j.cell.2015.06.043.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*, 2017. arXiv:1609.02907.

A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato. Finding statistically significant communities in networks. *PLoS ONE*, 6(4):e18961, 2011. doi: 10.1371/journal.pone.0018961.

K. Luck, D.-K. Kim, L. Lambourne, K. Spirohn, B. E. Begg, W. Bian, R. Brignall, T. Cafarelli, F. J. Campos-Laborie, B. Charloteaux, D. Choi, A. G. Coté, M. Daley, S. Deimling, A. Desbuleux, A. Dricot, M. Gebbia, M. F. Hardy, N. Kishore, J. J. Knapp, I. A. Kovács, I. Lemmens, M. W. Mee, J. C. Mellor, C. Pollis, C. Pons, A. D. Richardson, S. Schlabach, B. Teeking, A. Yadav, M. Babor, D. Balcha, O. Basha, C. Bowman-Colin, S.-F. Chin, S. G. Choi, C. Colabella, G. Coppin, C. D'Amata, D. De Ridder, S. De Rouck, M. Duran-Frigola, H. Ennajdaoui, F. Goebels, L. Goehring, A. Gopal, G. Haddad, E. Hatchi, M. Helmy, Y. Jacob, Y. Kassa, S. Landini, R. Li, N. van Lieshout, A. MacWilliams, D. Markey, J. N. Paulson, S. Rangarajan, J. Rasla, A. Rayhan, T. Rolland, A. San-Miguel, Y. Shen, D. Sheykhkarimli, G. M. Sheynkman, E. Simonovsky, M. Taşan, A. Tejeda, V. Tropepe, J.-C. Twizere, Y. Wang, R. J. Weatheritt, J. Weile, Y. Xia, X. Yang, E. Yeger-Lotem, Q. Zhong, P. Aloy, G. D. Bader, J. De Las Rivas, S. Gaudet, T. Hao, J. Rak, J. Tavernier, D. E. Hill, M. Vidal, F. P. Roth, and M. A. Calderwood. A reference map of the human binary protein interactome. *Nature*, 580(7803):402–408, 2020. doi: 10.1038/s41586-020-2188-x.

B. H. M. Meldal, O. Forner-Martinez, M. C. Costanzo, J. Dana, J. Demeter, M. Dumousseau, S. S. Dwight, A. Gaulton, L. Licata, A. N. Melidoni, S. Ricard-Blum, B. Roechert, M. S. Skyzypek, M. Tiwari, S. Velankar, E. D. Wong, H. Hermjakob, and S. Orchard. The complex portal—an encyclopaedia of macromolecular complexes. *Nucleic Acids Research*, 43(D1):D479–D484, 2014. doi: 10.1093/nar/gku975.

T. Nepusz, H. Yu, and A. Paccanaro. Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods*, 9(5):471–472, 2012. doi: 10.1038/nmeth.1938.

R. Oughtred, J. Rust, C. Chang, B.-J. Breitkreutz, C. Stark, A. Willems, L. Boucher, G. Leung, N. Kolas, F. Zhang, S. Dolma, J. Coulombe-Huntington, A. Chatr-aryamontri, K. Dolinski, and M. Tyers. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science*, 30(1):187–200, 2020. doi: 10.1002/pro.3978.

M. V. Palukuri and E. M. Marcotte. Super.complex: A supervised machine learning pipeline for molecular complex detection in protein-interaction networks. *PLOS ONE*, 16(12):e0262056, 2021. doi: 10.1371/journal.pone.0262056.

M. V. Palukuri, R. S. Patil, and E. M. Marcotte. Molecular complex detection in protein interaction networks through reinforcement learning. *BMC Bioinformatics*, 24(1), 2023. doi: 10.1186/s12859-023-05425-7.

Y. Pan, Y. Wang, J. Guan, and S. Zhou. PCGAN: a generative approach for protein complex identification from protein interaction networks. *Bioinformatics*, 39(8), 2023. doi: 10.1093/bioinformatics/btad473.

D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. von Mering. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1):D607–D613, 2018. doi: 10.1093/nar/gky1131.

S. van Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, 2000. `https://micans.org/mcl/`.

P. Veličković, W. Fedus, W. L. Hamilton, P. Lio, Y. Bengio, and R. D. Hjelm. Deep graph infomax. *International Conference on Learning Representations*, 2019. arXiv:1809.10341.

D. V. Veres, D. M. Gyurkó, B. Thaler, K. Z. Szalay, D. Fazekas, T. Korcsmáros, and P. Csermely. ComPPI: a cellular compartment-specific database for protein–protein interaction network analysis. *Nucleic Acids Research*, 43(D1):D485–D493, 2014. doi: 10.1093/nar/gku1007.

J. Xie, B. K. Szymanski, and X. Liu. SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 344–349. IEEE, 2011. doi: 10.1109/icdmw.2011.154.

J. Yang and J. Leskovec. Community-affiliation graph model for overlapping network community detection. In *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012. doi: 10.1109/icdm.2012.139.

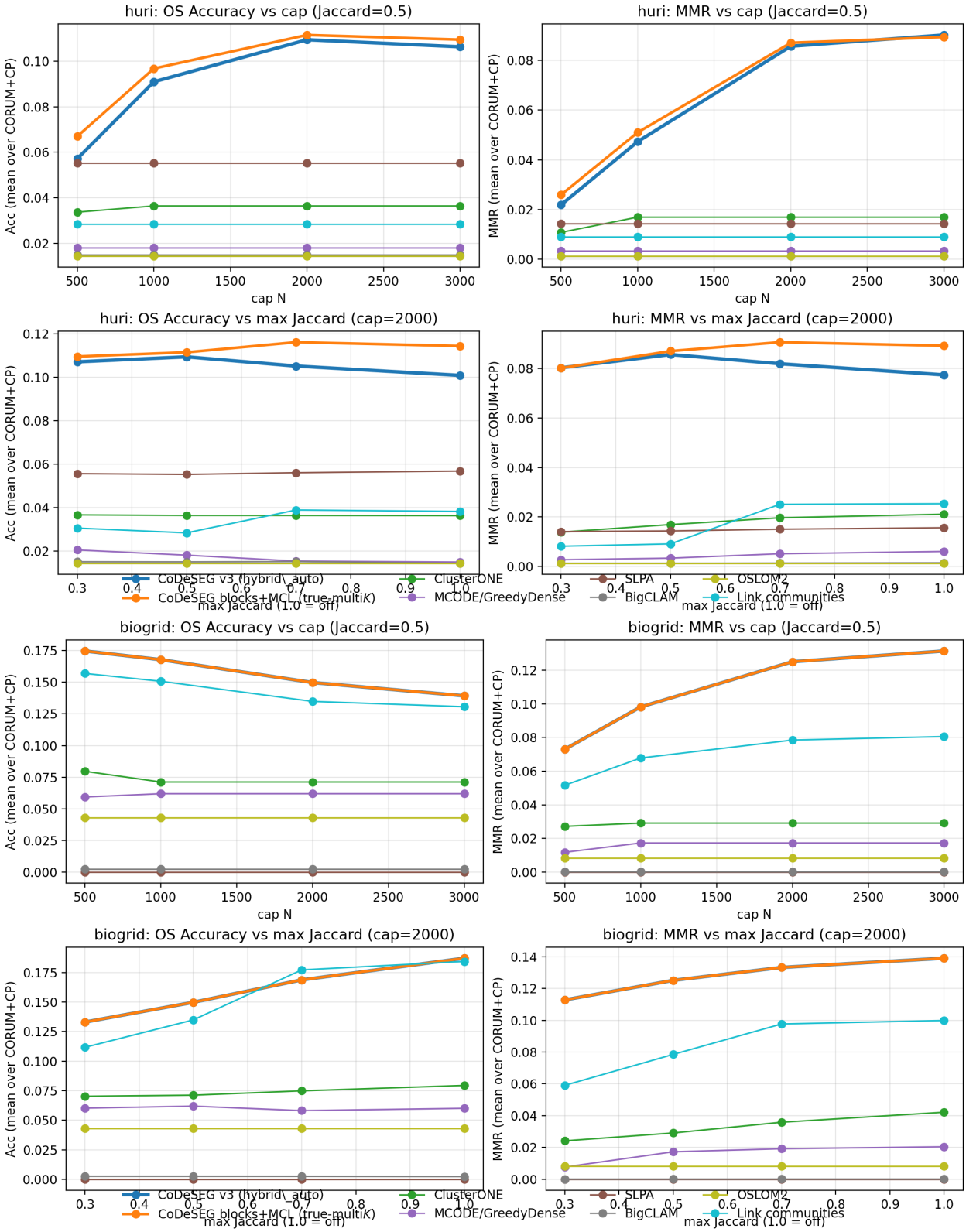

Fig. 3. Sensitivity of fixed-cap comparisons to the cap N and overlap suppression (max Jaccard). We sweep N at fixed max Jaccard = 0.5 and sweep max Jaccard at fixed $N = 2000$ on BioGRID (dense/weighted) and HuRI (sparse/unweighted). Metrics are OS Accuracy and greedy MMR, averaged over CORUM and Complex Portal.

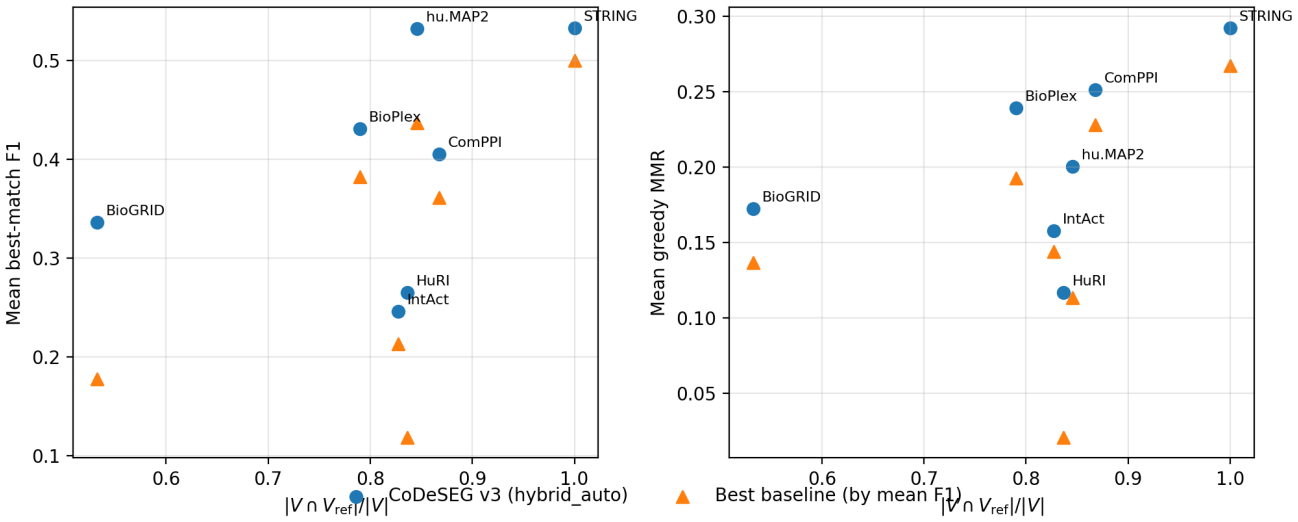

Fig. 4. Transfer performance vs. node overlap with STRING ($|V \cap V_{\text{ref}}|/|V|$). Blue circles show SoftBlock (`hybrid_auto`) and orange triangles show the best baseline per graph (by mean best-match F1). Metrics are mean best-match F1 and mean greedy MMR, averaged over CORUM and Complex Portal.