A Supplementary material

We implemented SSL in Python 3.6 and ran experiments on an 8-core Intel Core i7-8565U machine with 16GB RAM. Code and data are available at https://github.com/AU-DIS/SSL.

A.1 Hyperparameter setting

We calibrated SSL using grid search on the hyperparameters $\texttt{maxiter}_{\texttt{in}}$, $\texttt{maxiter}_{\texttt{out}}$, $a_{\texttt{tol}}$, and α. Table 1 shows the ranges of tested hyperparameters. We observe no significant difference vs. hyperparameters across different datasets, vindicating the robustness of our method. Surprisingly, we observe the same robustness for the regularization parameter μ. Similarly, we run a grid search on the number of sampled graphs k, the number of edges per sample t, and the threshold θ.

Parameter	Value/range	Description
$\texttt{maxiter}_{\texttt{in}}$	500–1000	number of inner iterations
$\texttt{maxiter}_{\texttt{out}}$	3–5	number of outer iterations
$a_{\texttt{tol}}$	10^{-5}	loss tolerance
α	0.02	gradient step size
k	30	number of sampled graphs
t	0.8	percentage of edges to sample
θ	0.2	fraction of solutions for the threshold strategy

Table 1: SSL hyperparameters and default values.

Figure 6 illustrates the impact of the parameter θ and the percentage of sampled edges t on the Threshold and Neighborhood strategy for different conductance values. Both methods achieve the best accuracy and stabilize around $t = 0.8$ percent of sampled edges and $\theta = 0.2$.

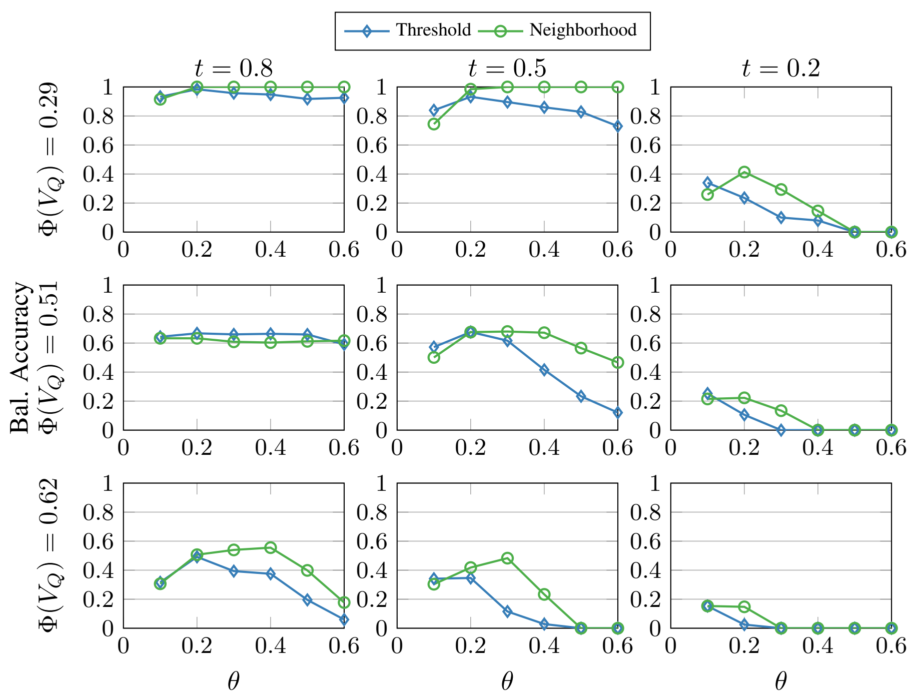


Figure 6: Balanced accuracy vs. threshold θ for $t \in \{0.8, 0.5, 0.2\}$ percentage of sampled edges and three conductance values $\Phi(V_Q)$ with Threshold and Neighborhood bagging-like strategies on $k = 30$ graphs sampled from the Football dataset.

A.2 Dataset description

Table 2 and Table 3 presents the characteristics of the datasets used in the experimental evaluations in terms of the number of nodes V, edges E, network type, and parameters.

Dataset	$\lvert V\rvert$	$\lvert E\rvert$	**Graph type**
Football [34]	115	613	Contact
Malaria [35]	306	9042	Biological
HighSchool [36]	327	5 818	Proximity

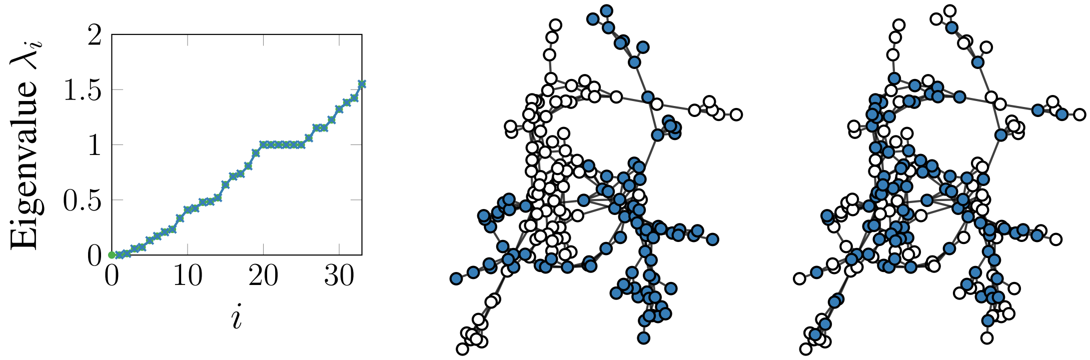
Table 2: Real graphs in our evaluation: number of nodes $|V|$, number of vertices $|E|$, graph type.

Dataset	**Source**	$\lvert V\rvert$	$\lvert E\rvert$	**Parameters**
Barabási–Albert	[37]	200	5907	$m_{new} = 50$
Erdős–Rényi	[38]	200	8185	$p_{new} = 0.5$

Table 3: Synthetic graphs in our evaluation: number of nodes $|V|$, number of vertices $|E|$, number of edges to attach from a new node to existing nodes m_{new}, edge creation probability p_{new}.

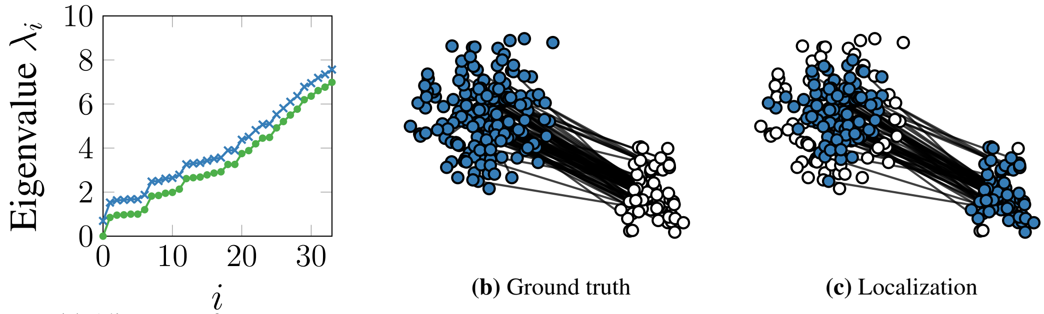
A.3 Challenging Cases

In this section, we present examples of problem instances that are challenging for SSL. In Figure 7, we observe a case where the spectrum of the query graph and that of the detected subgraph are well aligned, yet the localized subgraph deviates substantially from the ground truth. Likewise, in Figure 8, while SSL finds a spectrum correlated to the one of the query, it returns a subgraph comprising nodes that are only connected by one edge. In both cases, the challenge arises from the sensitivity of the spectrum at weakly connected parts of the graph. Changing the Laplacian in such parts by adding $\mathbf{v}$ has a larger impact on the spectrum than changing the Laplacian in a well-connected neighborhood. These types of graphs lure the optimization process into a local optimum, as the optimizer has a large incentive to separate these weakly connected parts of the graph.



(a) Alignment of spectra **(b)** Ground truth **(c)** Localization

Figure 7: Alignment of the spectrum $\boldsymbol{\lambda}_Q$ of Q and the corresponding part of the spectrum $\boldsymbol{\lambda}(\mathbf{L} - \mathbf{E} + \operatorname{diag}(\mathbf{v}))$ of G after convergence, ground truth V_S (blue) and $V \backslash V_S$ (white), and corresponding localization by SSL; while the spectra are perfectly aligned, the detected subgraph is not the ground truth. The depicted graph is a protein-protein interaction network from the D&D dataset [39].



(a) Alignment of spectra

(b) Ground truth

(c) Localization

Figure 8: Alignment of the spectrum $\boldsymbol{\lambda}_Q$ of Q and the corresponding part of the spectrum $\boldsymbol{\lambda}(\mathbf{L} - \mathbf{E} + \mathrm{diag}(\mathbf{v}))$ of G after convergence, ground truth V_S (blue) and $V \backslash V_S$ (white), and corresponding localization by SSL; while spectra are well aligned, the detected subgraph is not the ground truth, yet matches nodes with degree 0 or 1. The depicted graph is **arenas** [40].

A.4 SSL pseudocode

Algorithm 1 shows the psedocode of SSL, as presented in Section 3.

Algorithm 1 SSL

Input: $\mathbf{A}$ adjacency matrix of the full graph; $\boldsymbol{\lambda}_Q$ spectrum of the query subgraph.

Params: μ regularization coefficient; $a_{\texttt{tol}}$ loss tolerance; α gradient step size; $\texttt{maxiter}_{\texttt{in}}$ maximum number of inner iterations; $\texttt{maxiter}_{\texttt{out}}$ maximum number of outer iterations

Output: Vector $\mathbf{v}$, threshold τ

$\mathbf{L} \leftarrow \mathbf{D} - \mathbf{A}$

loss $\leftarrow \infty$

$c \leftarrow 2\sqrt{n - n_q}\,\max(\boldsymbol{\lambda}_Q)$

$\mathbf{v}_0 \leftarrow \frac{c}{|V|}\mathbf{1}$

$\mathbf{E}_0 \leftarrow \mathbf{0}$

while $q \leq \texttt{maxiter}_{\texttt{out}}$ **and** $\texttt{loss} \geq a_{\texttt{tol}}$ **do**

// Compute $\mathbf{v}_{q+1}$ by iterating (8) $\texttt{maxiter}_{\texttt{in}}$

$\quad \mathbf{v}_{q+1} \leftarrow \arg\min_{\mathbf{v}:\|\mathbf{v}\|=c} f(\mathbf{v}, \mathbf{E}_q)$

// Update $\mathbf{E}_{q+1}$ via (9)

$\quad \mathbf{E}_{q+1} \leftarrow \texttt{E\_from\_v}(\mathbf{v}_{q+1})$

// Update the threshold τ

$\quad \tau \leftarrow \texttt{k\_means\_1d}(\mathbf{v}_{q+1})$

$\quad \texttt{loss} \leftarrow f(\mathbf{v}_\tau, \mathbf{E}_\tau)$

$\quad q \leftarrow q + 1$

return $\mathbf{v}_q, \tau$