Novel Exploration via Orthogonality

Andreas Theophilou
University of Bath
United Kingdom
ajt80@bath.ac.uk

Özgür Şimşek
University of Bath
United Kingdom
o.simsek@bath.ac.uk

Abstract

Efficient exploration remains one of the key open problems in reinforcement learning. Discovering novel states or transitions requires policies that efficiently direct the agent away from regions of the state space that are already well explored. We introduce *Novel Exploration via Orthogonality (NEO)*, an approach that automatically uncovers not only which regions of the environment are novel but also how to reach them by leveraging Laplacian representations. NEO uses the eigenvectors of a modified graph Laplacian to induce gradient flows from states that are frequently visited (less novel) to states that are seldom visited (more novel). We show that NEO's modified Laplacian yields eigenvectors whose extreme values align with the most novel regions of the state space. We provide bounds for the eigenvalues of the modified Laplacian; and we show that the smoothest eigenvectors with real eigenvalues below certain thresholds provide guaranteed gradients to novel states for both undirected and directed graphs. In an empirical evaluation in online, incremental settings, NEO outperformed related state-of-the-art approaches, including eigen-options and cover options, in a large collection of undirected and directed domains with varying structures.

1 Introduction

Temporal abstraction has been widely studied as an approach to efficient exploration, which remains one of the central research areas in reinforcement learning. State of the art methods include eigen-options(13; 14) and cover time options(9). In symmetric settings, both of these methods can produce policies that push the agent toward the far reaches of the state space. However, such regions may already be heavily visited, and policies that push the agent naively to far away reaches of the environment are not always effective explorers.

An alternative approach is count-based exploration(2; 28; 24; 20). These methods address novelty through pseudo-rewards, for example, by giving the agent a novelty bonus proportional to $1/n$ for having visited a state-action pair n times. Some of these methods come with theoretical guarantees. However, these methods do not provide explicit gradient guidance to the agent. Rather than providing explicit policies for exploration, they rely on primitive actions and standard reinforcement update rules, such as Q-value learning, to propagate their intrinsic rewards, which can introduce a lag in exploration until the agent's value estimates have sufficiently propogated.

A natural alternative is to construct policies that explicitly drive the agent toward the most novel states (1). With access to an oracle solver for shortest paths in the transition graph, in symmetric settings, such policies can guarantee reaching a maximally novel state within a given horizon. However, this approach focuses exploration very narrowly on a single target , rather than diversifying across multiple novel regions.

We introduce *Novel Exploration via Orthogonality (NEO)*, a method that maintains the strengths of prior approaches whilst addressing their weaknesses. As with related spectral methods, NEO uses the

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

Laplacian eigenvector's smooth and orthogonal property to encourage exploration towards distinct regions of the state space. Unlike prior methods, NEO introduces gradient guarantees and focuses the exploration towards regions which are novel.

Importantly, whereas previous Laplacian approaches assume an undirected transition graph, many real-world environments are inherently directed. Recent work has proposed using the polar decomposition of a transition matrix to recover real eigenvalues to obtain eigen-options in the directed setting (5); however, this approach uses a symmetric, Hermitian matrix, which cannot provide directed gradient flows beyond the symmetric graph obtained in the construction. Moreover, eigen-options in directed settings by polar decomposition is used only as a baseline; it has not been shown in evaluations to perform desirably (5). By careful construction, we obtain eigenvectors with gradient guarantees in strongly connected directed graphs for any eigenvector with an associated real eigenvalue below a given value. Conceptually, as we create a modified Laplacian and take its eigenvectors, we effectively convert novelty to energy where the most novel states dominate the contributions to the inner product $\langle u, v\rangle = \sum_i u_i v_i$ and are the highest energy states of the smoothest Laplacian eigenvectors with real eigenvalues below a novelty-based threshold.

We provide experimental results in a wide variety of environments, both directed and undirected, that illustrate the approach and its effectiveness. In online, incremental settings, NEO outperformed related state-of-the-art approaches, including eigen-options and cover options, in a large collection of undirected and directed domains with varying structures.

2 Preliminaries

A finite Markov decision process (MDP) is a 5-tuple $\langle S, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma\rangle$, where S is a set of states, $\mathcal{A}$ a set of actions, $\mathcal{T}(s, a, s') \in [0, 1]$ the probability of transitioning to state s' upon taking action a in state s, $\mathcal{R}(s, a, s')$ the expected reward for that transition, and $\gamma \in [0, 1]$ the discount factor. At decision stage t, $t \geq 0$, the agent is in state $s_t \in S$, selects action $a_t \in \mathcal{A}$, and transitions to state s_{t+1}, receiving reward r_{t+1}. A *policy* $\pi(s, a)$ gives the probability of selecting action a when in state s. The value function $V^\pi(s) = \sum_{a\in\mathcal{A}} \pi(s, a) \sum_{s'\in\mathcal{S}} \mathcal{T}(s, a, s')\Big[\mathcal{R}(s, a, s') + \gamma\, V^\pi(s')\Big]$ is the expected discounted sum of future rewards when following policy π from state s. An optimal policy is one that maximizes $V^\pi(s)$ for every state $s \in S$. We refer to the actions of an MDP as *primitive actions* and represent *temporally-extended actions* using the options framework (23; 18). An *option* o is a triple $\langle I_o, \pi_o, \beta_o\rangle$, where $I_o \subseteq S$ is the set of states in which the option can be initiated, π_o is the policy followed during option execution, and $\beta_o : S \to [0, 1]$ is the *termination function*, expressing the probability that the option terminates at a given state.

Associated with an MDP $\langle S, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma\rangle$, we define a directed graph $G = (V, E)$, where $V = S$ is the set of vertices and E is the set of edges, with $(s, s') \in E$ iff $\exists a \in \mathcal{A}$ such that $\mathcal{T}(s, a, s') > 0$. Given this graph, we define *adjacency matrix* A, where $A_{ij} = 1$ if $(i, j) \in E$ and 0 otherwise, and *random walk transition matrix* P, where $P_{ij} = \frac{A_{ij}}{\deg^+(i)}$, where deg^+ denotes the out-degree of a node, $deg^+(i) = \sum_j A_{ij}$.

The *Rayleigh quotient* of matrix $M \in \mathbb{C}^{n\times n}$ at vector $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ is the scalar $R(M; \mathbf{x}) = \frac{\mathbf{x}^* M \mathbf{x}}{\mathbf{x}^* \mathbf{x}}$, where $\mathbf{x}^*$ is the conjugate transpose of $\mathbf{x}$. Let λ denote the eigenvalues and f the eigenvectors. If M is Hermitian ($M = M^*$), then $R(M; \mathbf{x}) \in \mathbb{R}$ for all $\mathbf{x} \neq \mathbf{0}$, and $\lambda_{\min}(M) = \min_{\mathbf{x}\neq 0} R(M; \mathbf{x})$, $\lambda_{\max}(M) = \max_{\mathbf{x}\neq 0} R(M; \mathbf{x})$, with the extrema attained exactly at the corresponding eigenvectors f of M. For a complex number, we use $\Re$ to refer to the real part and $\Im$ to refer to the imaginary part.

3 Source Laplacian

In this section, we define the *source Laplacian* L_ζ by adding non-negative diagonal weights to the *random-walk Laplacian* $L^{rw} = I - P$, thereby concentrating energy on selected nodes.

Let $G = (V, E)$ be a directed, strongly connected graph of n nodes, with row-stochastic transition matrix $P = (P_{ij}) : P_{ij} \geq 0,\ \sum_{j=1}^n P_{ij} = 1\ (\forall\, i)$. Let I denote the identity matrix. Introduce nonnegative weights $\zeta = (\zeta_i)_{i=1}^n$ and define $n \times n$ matrix Γ, with $\Gamma_{ii} = \zeta_i$ and all other entries zero. Define $L_\zeta = I + \Gamma - P$. Adding Γ increases the diagonal entries and, in eigenvectors of L_ζ, pulls

the eigenvector value at weighted nodes toward zero, raising the relative prominence of less weighted nodes where energy gets concentrated. A node i with $\zeta_i = 0$ is a *source node*. A node i with $\zeta_i > 0$ is a *sink node*.

The Rayleigh quotient of source Laplacian L_ζ for any $x \in \mathbb{C}^n \setminus \{0\}$ is:

$$\mathcal{R}(x) = \frac{x^* L_\zeta x}{x^* x} = \frac{x^*(I+\Gamma)x - x^* P x}{x^* x} = \frac{\sum_{i=1}^n (1+\zeta_i)|x_i|^2 - \sum_{i,j=1}^n \overline{x_i}\, P_{ij}\, x_j}{\sum_{i=1}^n |x_i|^2}, \tag{1}$$

where $\overline{x_i}$ is the complex conjugate of x_i. If the graph is symmetric, the Rayleigh quotient can be expressed as:

$$\mathcal{R}(x) = \frac{x^\top L_\zeta\, x}{x^\top x} = \frac{\frac{1}{2}\sum_{i,j} P_{ij}(x_i - x_j)^2 \;+\; \sum_i \zeta_i\, x_i^2}{\sum_i x_i^2}\,. \tag{2}$$

This is the standard Rayleigh quotient for a symmetric matrix $L^{rw} = I - P$ with the additional term $\sum_i \zeta_i\, x_i^2$, which is the added squared smoothness error contributions $\zeta_i(x_i - 0)^2$ between the node values x_i and the zeros, weighted by ζ_i.

Below, we present three theorems. In theorem 3.1, we prove a lower bound for the eigenvalues of the source Laplacian and an upper bound for its smoothest eigenvalue. In theorem 3.2, we show that, for certain eigenvalues, the maximum of the associated eigenvector is attained at a source node. In theorem 3.3, for the source Laplacian with Γ weighted by visitation values, we prove that, certain eigenvectors (those with associated eigenvalues lower than the maximum visitation value) give us guaranteed gradients (and thus paths) from nodes with higher visitation values to nodes with lower visitation values.

Theorem 3.1 (Bounds for the eigenvalues λ of the source Laplacian $L_\zeta = (I+\Gamma) - P$)**.** Suppose k nodes are sources ($\zeta_i = 0$) and $n-k$ are sinks, each with uniform weight $\mu > 0$ (i.e., $\zeta_i = \mu$ on sinks). Then, $0 \le R(\mathbf{1}) < \mu$, and the upper bound for the smoothest eigenvalue is μ. If all sources are replaced with weight α, where $\mu > \alpha \ge 0$, then $\alpha < R(\mathbf{1}) < \mu$, and the lower bound for the smoothest eigenvalue is α . If all $\zeta_i = \mu$, then $R(\mathbf{1}) = \mu$, and the smoothest eigenvalue is μ.

Proof. $R(\mathbf{1}) = \frac{1}{n}\sum_{i=1}^n \zeta_i$. In the *standard case*, with k sources and $n-k$ sinks, k entries are zero and $n-k$ entries equal μ, and $\sum_i \zeta_i = (n-k)\mu$. Consequently, $R(\mathbf{1}) = (n-k)\mu/n = (1-k/n)\mu$, which implies $0 \le R(\mathbf{1}) < \mu$. In the α *weighted variant*, $\sum_i \zeta_i = k\alpha + (n-k)\mu$, so $R(\mathbf{1}) = \frac{k\alpha+(n-k)\mu}{n}$, a combination of α and μ, hence $\alpha < R(\mathbf{1}) < \mu$. In the *uniform case*, $\sum_i \zeta_i = n\mu$, giving $R(\mathbf{1}) = \mu$. □

Theorem 3.2 (Maximality (in magnitude) of the eigenvector at sources)**.** Given source Laplacian $L_\zeta = (I+\Gamma) - P$, assume $L_\zeta f = \lambda f$ with real eigenvalue $\lambda < \min_{i:\zeta_i>0} \zeta_i$, that is, the eigenvalue is less than the minimum non-zero weight in ζ. Then the element of eigenvector f with the highest absolute value, $|f_i|$, must be a source node ($\zeta_i = 0$).

Proof. From $L_\zeta f = \lambda f$ and $L_\zeta = I + \Gamma - P$, we have $(I + \Gamma - P)f = \lambda f$. Then for the ith row (corresponding to node i), after rearranging, we get $(1 + \zeta_i - \lambda)f_i = \sum_j P_{ij} f_j$, which yields:

$$f_i = \frac{1}{1+\zeta_i - \lambda} \sum_j P_{ij} f_j \tag{3}$$

Taking absolute values at both sides of the equation, we obtain $|f_i| = \frac{1}{|1+\zeta_i-\lambda|}\left|\sum_j P_{ij} f_j\right|$. If the i^{th} node is a sink node, then $\zeta_i > \lambda$, and $|1 + \zeta_i - \lambda| > 1$, and so

$$|f_i| \le \frac{1}{1+\zeta_i - \lambda} \sum_j P_{ij}|f_j| < \max_j |f_j|, \tag{4}$$

implying that, for a sink node i, $|f_i|$ cannot be a maximum entry of f in magnitude. Therefore, maximality must be obtained at a source node. □

Theorem 3.3 (Novelty path via visit counts)**.** Let G be a strongly connected graph, and $N(i)$ a visitation value of node i. Define $L_\zeta = (I + \Gamma) - P$ with $\Gamma_{ii} = N(i) \in \mathbb{R}_{\geq 0}$. For any eigenvalue λ and associated eigenvector f (right eigenvector f if L_ζ is asymmetric) where $N(i) > \Re(\lambda)$, there exists a directed path $v_0 \to v_1 \to \cdots \to v_k$ such that $|f_{v_0}| < |f_{v_1}| < \cdots < |f_{v_k}|$ and $N(v_k) < \Re(\lambda)$.

Proof. Following the same reasoning for the i^{th} entry of $L_\zeta f = \lambda f$ as in Equation 3, we have $(1 + N(i) - \lambda) f_i = \sum_j P_{ij} f_j$. Therefore,

$$f_i = \frac{1}{1 + N(i) - \lambda} \sum_j P_{ij} f_j. \tag{5}$$

If $N(i) > \Re(\lambda)$, then

$$\begin{aligned} \big|1 + N(i) - \lambda\big| &= \big|(1 + N(i) - \Re(\lambda)) - \sqrt{-1}\,\Im(\lambda)\big| \\ &= \sqrt{(1 + N(i) - \Re(\lambda))^2 + \Im(\lambda)^2} \\ &\geq 1 + N(i) - \Re(\lambda) > 1. \end{aligned}$$

Therefore,

$$|f_i| = \frac{1}{\big|1 + N(i) - \lambda\big|} \Big| \sum_j P_{ij} f_j \Big| < \sum_j P_{ij} |f_j|, \tag{6}$$

implying $\sum_j P_{ij}|f_j| > |f_i|$. Thus some neighbor j has $|f_j| > |f_i|$. Iterating yields a strictly increasing chain in $|f|$, terminating at v_k with no neighbor with a larger magnitude, forcing $\big|1 + N(i) - \lambda\big| \leq 1$ and $N(v_k) < \Re(\lambda)$. □

Remark 3.1. Since theorem 3.3 uses only the magnitudes $|f_v|$, any complex phase of f is irrelevant to the strict-inequality relations among $|f_{v_i}|$, $i = 0, ..., k$.

In the undirected setting, L_ζ is Hermitian and all its eigenvalues and eigenvectors are real. Consequently, one may simply take the eigenvectors corresponding to the z smallest eigenvalues, knowing that each of these smoothest eigenvectors will define an orthogonal gradient field that carries mass from states with lower visitation values to states with higher visitation values, as long as the corresponding eigenvalue λ is less than the maximum visitation value. When G is directed, L_ζ becomes non-Hermitian, and complex eigenvalues (and eigenvectors) can appear. Nevertheless, any eigenvalue with $\Re(\lambda)$ below the maximum visitation value still induces a guaranteed gradient from nodes with higher visitation values to nodes with lower visitation values. In the appendix, we present a detailed exploration of theorem 3.3 when the eigenvalues and eigenvectors are real valued.

4 Novel Exploration with Orthogonality (NEO)

We now use the smoothest eigenvectors of the source Laplacian L_ζ to define multi-step exploration policies in the form of options. We call our approach *Novel Exploration with Orthogonality*, or NEO, and outline it in pseudocode in Algorithm 1. [1]

At each iteration of the algorithm, we start by following the agent's policy for H steps, recording every transition so as to assemble (or update) an empirical graph with transition matrix P and to accumulate raw visitation counts $N(s)$ for each state s encountered during the roll out. Raw counts can vary wildly between instantiations; we therefore transform them via a scaling function $F\colon \mathbb{R} \to \mathbb{R}$, using two parameters: $\delta > 0, k > 0$. Specifically,

$$F\big(N(s)\big) = \delta \left(\frac{N(s)}{\max_{i \in S} N(i)}\right)^{1/k}, \tag{7}$$

where the denominator $\max_{i \in S} N(i)$ ensures that the maximum scaled count prior to being multiplied by δ is less than or equal to 1, the exponent $1/k$ compresses their dynamic range, and the multiplication with δ further shrinks or increases them. In the appendix, we present results obtained by varying δ

[1] The code is available at `https://github.com/AndreasTheo/NEO`

and k. For all results presented in the main paper, we use $\delta = 0.5$ and $k = 64$. Using the scaled visitation counts $\zeta_i = F\big(N(i)\big)$, $i \in S$, we define an $n \times n$ matrix Γ, with $\Gamma_{ii} = \zeta_i$ and all other entries set to zero, and another $n \times n$ matrix $L_\zeta = I + \Gamma - P$, the source Laplacian.

We then compute the Z smoothest eigenvectors of L_ζ, those associated with its Z smallest real eigenvalues. For every eigenvalue $\lambda < \delta$, the associated eigenvector f defines a gradient that naturally points away from states that are heavily visited toward those that are less explored.

We instantiate one option per eigenvector. We use f^o to denote the eigenvector that corresponds to option o. This option can be initiated in any state. The option policy moves the agent to the neighboring state with the largest magnitude of the corresponding eigenvector, $|f^o|$. The option terminates with probability 1 when the agent reaches a local peak of $|f^o|$ (a state where the magnitude of the eigenvector exceeds or equals that of all its immediate neighbors), with probability zero elsewhere.

Algorithm 1 NEO

Input: update horizon H, option count Z, scaling function F.
$O = \{\}$ // set of options
$S = \{\}$ // set of states
loop
  Execute agent policy for H steps
  Update state set S, graph G, transition matrix P, and visitation counts$N(.)$
  $L_\zeta = I + \Gamma - P$, where $\Gamma_{ii} = F(N(i))$
  Compute eigenpairs $\{(\lambda_o,\ f_o)\}_{i=1}^{Z}$ of L_ζ, sorted by ascending λ_o
  for $i = 1$ to Z **do**
    $I_o = S$ // initiation set
    $\beta_o(s) = 1$ iff $\forall s' : (s \to s')$, $|f^o_{s'}| \leq |f^o_s|$ // termination condition
    $\pi_o(s)$: choose action that leads to next state s' that maximises $|f^o_{s'}|$ // option policy
    $o = \langle I_o, \pi_o, \beta_o \rangle$
    $O \leftarrow O \cup o$
  end for
end for=0

In Figure 1, we illustrate the eigenvectors of the source Laplacian in some directed and undirected domains. Plots (a) and (b) illustrate how the eigenvectors of the source Laplacian capture directed connectivity. In plot (a), we show a 20-node directed cycle, with one source ($\zeta = 0$; the node shown in the darkest shade of green) and all other nodes as equal sinks ($\zeta = 0.1$), yielding a smooth eigenvector that increases monotonically around the circle. In plot (b), we show a directed four-room grid with one-way doorways. Setting the top left corner as the source ($\zeta = 0$) and all other states as sinks ($\zeta = 0.1$) produces an eigenvector whose gradient reflects directed distance from sinks to the source.

Plots (c) and (d) show how we derive weights for Γ from visitation counts in directed four rooms and in a maze environment, respectively. In directed four rooms, the agent performed a random walk of 100 steps, starting at the top left corner; this was repeated for 100 trials. In the maze, the agent performed a random walk of 10000 steps, starting at the bottom left corner; this was repeated for 500 trials. In both domains, visitation counts were scaled by their maximum, raised to the power of $\frac{1}{4}$ to compress their dynamic range, and multiplied by $\delta = 0.5$. In directed four-rooms, one state was never visited, whereas all states in the maze received some visitation (minimum $F(N(.)) \approx 0.475$).

Using the scaled visitation counts $F(N(.))$ to add to the diagonals of Γ, plots (e)–(h) display the first four eigenvectors of the source Laplacian with novelty weighting for Γ in directed four rooms; plots (i)–(l) show the same in in the maze. We measured the cosine similarity between eigenvectors in directed four rooms (the eigenvectors displayed in plots (e)-(h)); the maximum value observed was 0.001. Orthogonality leads to distinct novel regions being high in magnitude across the eigenvector values. Notice that the smallest eigenvalue in plots (e) and (i) align with the theoretical lower and upper bounds set by the minimum diagonal value of Γ. We have a minimum $F(N(.))$ of 0 in plot (c) and 0.475 in plot (d) with a maximum $F(N(.))$ of 0.5 in both, resulting in eigenvalues of ≈ 0.136 in plot (e) and ≈ 0.48 in plot (i).

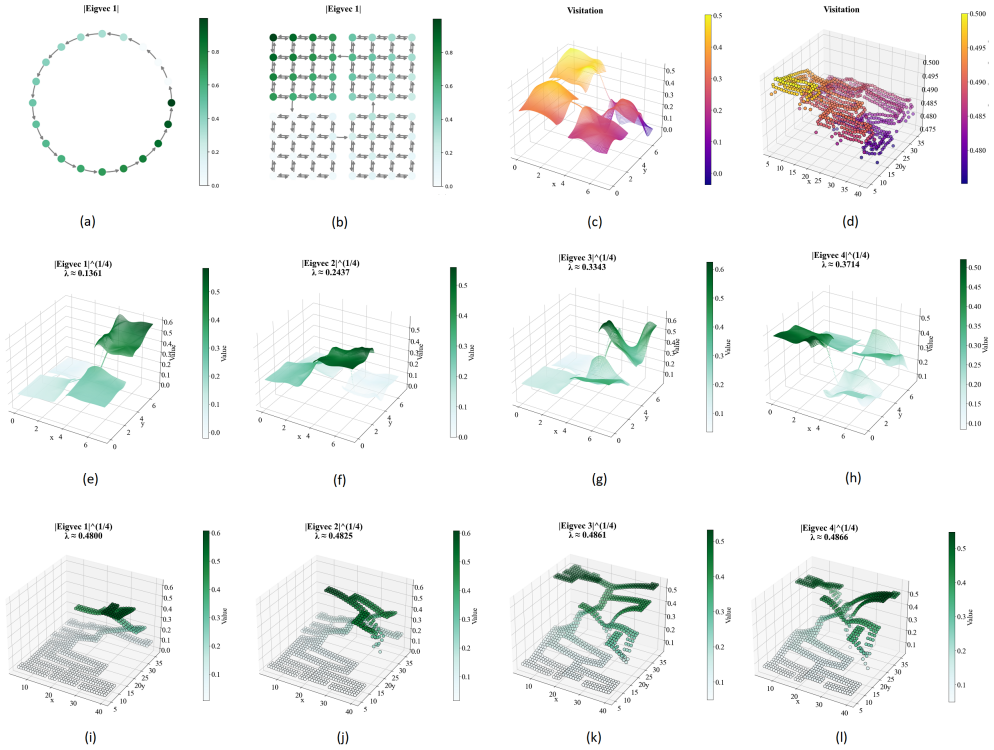

Figure 1: Smoothest eigenvector of L_ζ in (a) directed-cycle, (b) directed four rooms. Visitation-based sink weights in (c) directed four-rooms, (d) large maze. The first four eigenvectors of the source Laplacian L_ζ in (e–h) directed four rooms, and (i–l) large maze.

5 Related Work

Steinberger (3) demonstrated that the first eigenvector of an undirected graph Laplacian provides an approximate shortest-path solution via a single sink node, inspiring the spectral method proposed here. Earlier related works include eigen-options (13; 10; 14), which use eigenvectors of a normalized Laplacian for options policies, as well as proto value functions (15; 16), which use the same eigenvector construction for representation learning and efficient value estimation. Cover options (9) build on eigen-options, applying solely the fielder vector recursively. Diffusion options (5) use eigenvectors of the Laplacian to find diffusely separated option termination states; however, this method is not currently applicable in the online setting. The successor representation (6) has been shown to be closely related to the graph Laplacian (19). Another set of closely related graph-based methods include sub-goal discovery via betweenness (22), identifying bottleneck states (21), relative novelty (20), and modularity maximization (7), which show favorable performance of graph-based clustering methods for hierarchical options.

Recent approaches to approximate the eigenvector of the Laplacian (27; 25; 8) and commute times (4) via these eigenvectors provide a foundation for future work in extending Laplacian-based option discovery methods to the large and continuous setting. Also related are methods for extending novelty counts to the deep reinforcement learning setting (24; 28).

6 Empirical Evaluation

We use Q-learning with step size $\alpha = 0.4$, discount rate $\gamma = 0.99$, an ϵ-greedy policy with $\epsilon = 0.1$, augmented with a growing library of temporally extended actions in the form of options. The initial

Q values are set to 0 for primitive actions and -0.00001 for options, ensuring that the agent policy initially favors primitives unless exploration explicitly invokes an option. When the ϵ-greedy policy takes an exploration step, with probability $init_P$ the agent chooses a random option, otherwise a random primitive action. We fix $init_P$ at 0.1 for reward based evaluations in Figure 3; we evaluate the impact of the $init_P$ parameter in Figure 4. An option can be initiated in any state that was part of the graph used when the option was constructed (in other states, there would be no constructed option policy). Each transition triggers the usual one-step Q-learning update $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha\big[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\big]$. Additionally, at the end of an episode, the entire episode's transitions are replayed in reverse order, with the same Q-learning update to accelerate reward propagation in our sparse-reward task. Option values are learned using SMDP Q-learning backups: $Q(s,o) \leftarrow (1-\alpha)\, Q(s,o) + \alpha \left[\sum_{j=0}^{\tau-1} \gamma^j\, r_{t+j+1} + \gamma^\tau \max_{a'} Q\big(s_{t+\tau}, a'\big) \right]$, where s is the state in which option o was initiated, τ is option duration until termination, and r_{t+j+1} are the rewards accrued while executing option o.

In all domains, there is a single goal state. Upon reaching the goal state, the agent receives a reward of $+100$ and the episode terminates. All other transitions give a reward of 0. We use an adaptive episode horizon: 500 steps for domains under 1000 states, 1000 steps for domains with 501–3000 states, and otherwise s horizon that is equal to the number of states in the domain. For each evaluation agent instance, a goal location is randomly chosen and the farthest state from the goal is set to be the start state for all runs for the agent instance. For each method, we run 20 agent instances; we use a random seed equal to run number before selecting goal and start states ensuring all start and goal states are the same across compared agents.

Options are discovered online every H decision stages, with H set to five times the number of nodes in the domain's full transition graph. We generate four new options per update, capping the total stored options at 64 and discarding the oldest four options when capacity is exceeded and replacing them with the newest four options. Each option policy is defined by ascending a representation until reaching a local maximum: for eigen-options, the representation is an eigenvector of the graph Laplacian; for cover-options, it is the Fiedler vector over graph nodes; for the shortest-path-novelty(1) options, it is the shortest path distance to the most novel node in the agent's working graph; and for NEO options, it is the magnitude of the eigenvectors described previously in theorem 3.3. In directed domains, the polar-decomposition variant of eigen-options, NEO options, and the shortest-path-novelty options are applied. We cannot use cover options in directed domains because they rely on the construction of Fiedler vectors.

Every 2000 steps, which we define as an *epoch*, we freeze exploration and run an independent evaluation episode where no exploration actions are taken ($\epsilon = 0$). We plot the evaluation results with a 2σ error width. As a baseline, we also include a primitive Q-learner which receives an intrinsic novelty bonus $\beta/\sqrt{n(s,a)}$, with β = 0.01.

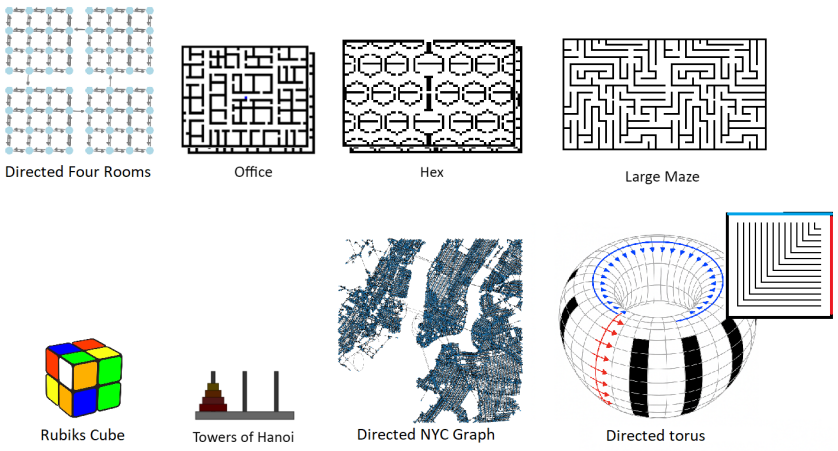

Figure 2: A visual depiction of the eight domains used in the empirical analysis.

In Figure 2, we present a visual depiction of the domains used in our analysis. Those under 1000 states include `directed four rooms`, a gridworld with four rooms connected via one-way hallway states, `Rubik's cube end game`, the undirected subgraph of the 2×2 Rubik's cube containing the states that are within 3 moves of the solved state, and the classic `towers of Hanoi` puzzle. At roughly 2500 states is the `office`, an undirected environment challenging to previous spectral-based exploration methods (7). Larger still are `large maze` and `hex`, each with approximately 5500 states. In addition to these undirected versions, we created and experimented on directed versions of large maze and hex by adding a single directed edge between the two most distant states. Our largest domains are a directed `New York City (NYC)` street graph consisting of 10000 nodes and the directed torus domain. In directed torus, the red walls on the right side wrap in a one-way direction (from right to left) to the walls on the left side, and the blue walls at the top wrap in a one-way direction (top to bottom) to the walls at the bottom. This domain was constructed as a more challenging alternative to a standard torus, allowing us to assess how agents navigate when traversing any of a large number of directed edges which can preclude finding returning paths.

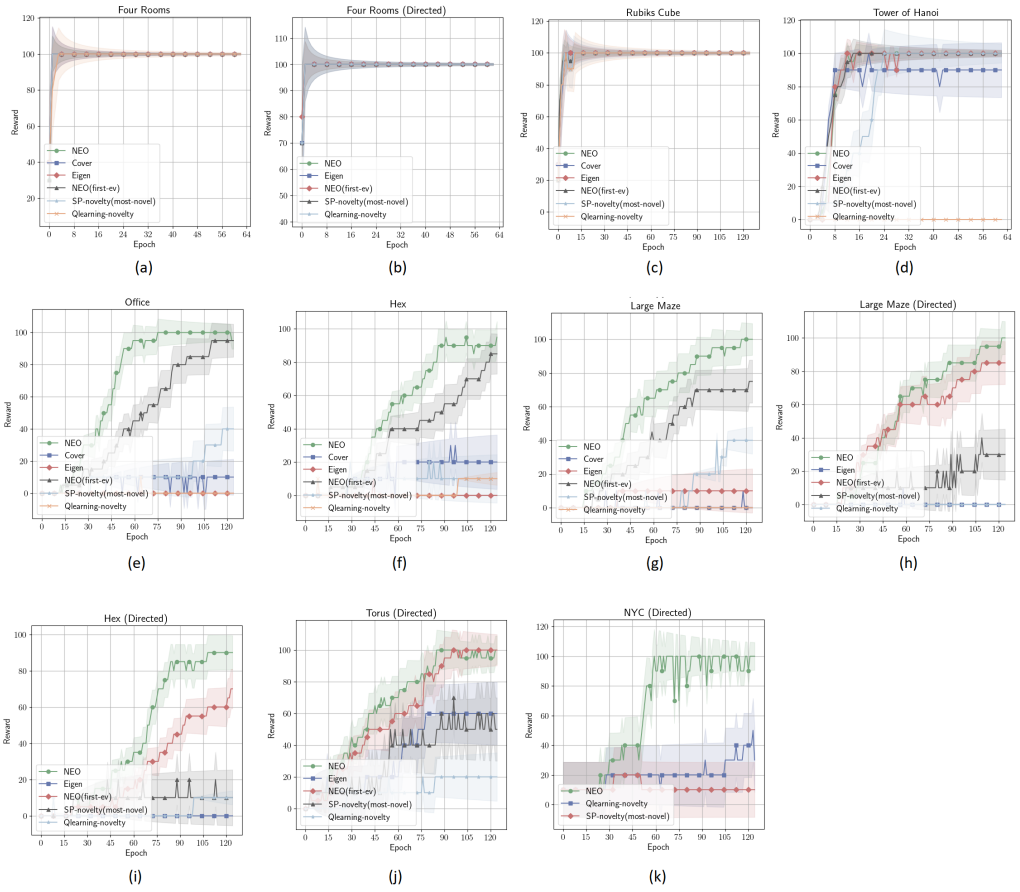

Figure 3: Comparative sparse reward empirical evaluation of learning agents with different option discovery methods across a set of domains.

The learning curves are shown in Figure 3. Plots (a) – (d) show the learning curves for our simplest benchmark domains: four rooms (undirected and directed), Rubik's cube end game, and the towers of Hanoi. All option-based agents reach the maximum reward relatively quickly, with the exception of the shortest-path-novelty method, which exhibits slower learning on Towers of Hanoi.

Plots (e) – (g) compare performance on three undirected domains: office, hex, and large maze, where the proposed method NEO shows a clear advantage over the alternative approaches. By the end of training, NEO agents achieve the maximum reward of 100 on average, while the other methods fail

to achieve higher than 40. We also include a variant of NEO that uses only the first (i.e., smoothest) eigenvector, yielding a single option per update interval, which outperforms the related methods. However, it is evident from the evaluations that leveraging multiple eigenvectors provides additional learning benefits beyond this single eigenvector baseline.

In plots (h) – (k), we present results in directed domains. As in the undirected setting, both NEO variants outperform the alternative methods across all directed tasks. The agent's state-transition graphs remained weakly connected throughout most of learning, and we obtained no complex eigenvalues in the construction of any option policy for NEO, with all 4 smoothest eigenvalues computed every H update iterations being real valued. We hypothesize that the presence of sink states stabilizes the eigenvalues and makes it more likely that they are real valued by influencing the stationary distribution, though we can not yet provide formal evidence for this claim. For the NYC directed domain, eigen-options could not be computed because the solver would crash. We suspect this issue arises from the interaction between the polar decomposition method and initially weakly connected graphs, which may impair numerical stability during option computation.

In Fig. 4, we present the number of unique nodes visited by each method, with various values of the option initialization probability $init_P$. Lower initialization probabilities consistently improve coverage for all methods. However, only NEO approaches full coverage in hex and large maze. These are the largest symmetric domains where it is possible to compare all methods. In these domains, NEO achieves a higher unique-node count during evaluations, roughly twice that of the competing approaches.

Overall, these evaluations demonstrate that NEO achieves or surpasses state-of-the-art performance in both undirected and directed environments, in terms of exploration and, by extension, learning.

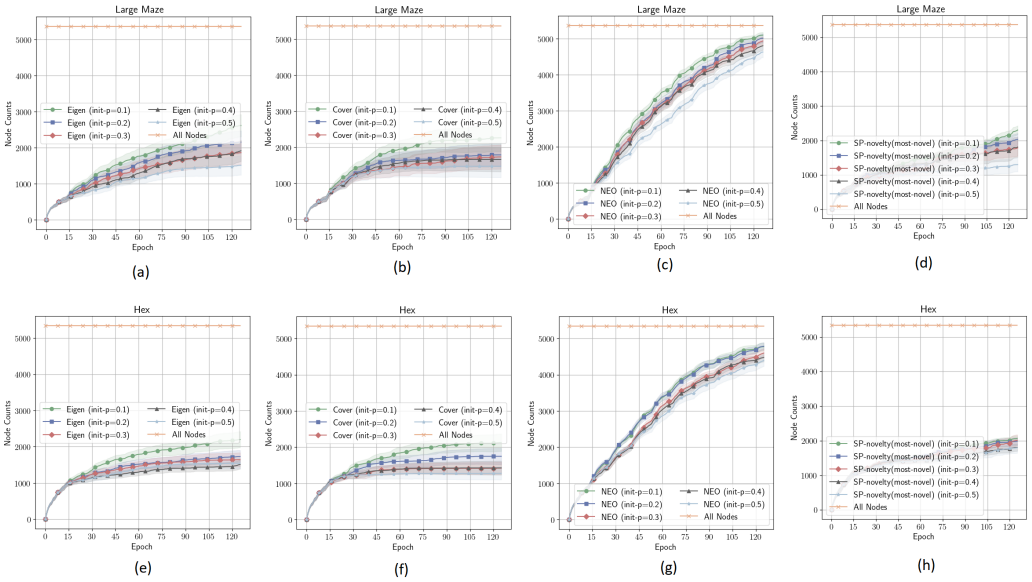

Figure 4: Comparative initialization probability and node count evaluations of learning agents with different option discovery methods across a set of domains.

One boundary of the Fiedler vector tends to sit in a well-visited region. Eigen-options and cover options include the Fiedler vector among their option sets. To understand the performance gap between these methods and NEO, we analyze how the *signed* option pair, corresponding to eigenvectors f and $-f$, can undo progress by steering the agent back to states that are relatively well explored. For this analysis, we use the two largest symmetric domains where Fiedler vectors can be computed, hex and large maze.

Consider an agent that begins near a corner of the full graph and builds the graph incrementally. That corner typically becomes the region that is the most connected and the most frequently visited.

Placing one boundary (say b_1, an extremum of the Fiedler vector f) in this region tends to be part of the first non-trivial solution of the Rayleigh quotient $\mathcal{R}(f) = \frac{\sum_{(u,v)\in E}\left(f_u - f_v\right)^2}{f^\top f}$ because the dense local connectivity allows f to taper off gradually, keeping each edge difference $(f_u - f_v)$ and thus the numerator small. If the second boundary were nearby, neighbouring nodes would carry opposite extreme values, the numerator would spike, and the quotient would rise. Minimising the quotient therefore pushes the second boundary to a geometrically far away state, such as another corner of the graph, rarely visited.

After 250K steps, we freeze the exploration graph and visitation counts $N(\cdot)$, compute eigen-options, and extract the Fiedler vector f. Let b_1 and b_2 be the two boundary states (the extrema of f), and define $b_{\text{low}} := \arg\min\{N(b_1), N(b_2)\}$, and $b_{\text{high}} := \arg\max\{N(b_1), N(b_2)\}$. We initialize the agent at b_{low} and execute the Fiedler-vector option that follows the increasing f-gradient, moving toward the opposite boundary until the gradient stalls at a local or global maximum. Our aim is to assess whether one of the pair tends to undo progress. Table 1 shows the visitation counts $N(.)$ at Fiedler vector boundaries b_{low} and b_{high}, the difference, and the visitation counts at the option termination state. For both domains, one of the signed options lifts the agent from relatively novel states to frequently visited states (e.g., mean $68.5 \rightarrow 822.6$, median $14.5 \rightarrow 745.0$ in hex). Because the two options are chosen with the same probability, the option that drives the agent back to the high-visitation boundary can directly undo the progress made by its counterpart that moves toward the low-visitation boundary.

Table 1: Visitation counts (in 20 runs).

	$N(b_{\text{low}})$		$N(b_{\text{high}})$		$N(b_{\text{high}}) - N(b_{\text{low}})$		$N(\text{termination})$	
Domain	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
Large Maze	129.8	39.0	1355.1	1070.5	1225.3	952.5	985.6	638.0
Hex	68.5	14.5	958.9	503.5	890.5	467.0	822.6	745.0

7 Conclusion and Future Work

We presented a principled approach for generating exploration options via a proposed source Laplacian, leveraging theoretical insights and spectral graph properties to guide exploration to novel parts of the state space. The proposed method is shown to drive agents toward novel regions, improving learning performance across a range of challenging domains compared to the existing state-of-the-art. The approach is shown to be applicable to both undirected and directed environments, demonstrating versatility.

Although the theoretical and empirical results presented in this paper center on using state novelty as the driving signal, the proposed framework is general and is not restricted to the use of novelty alone. In fact, any real-valued state functions could be plugged in. Future work can, for instance, replace or augment novelty with expected reward, reward prediction error, or competence signals. We discuss the future extension to symmetric Hilbert spaces in the appendix.

A current limitation is obtaining eigenvectors of the source Laplacian L_ζ in Hilbert space as done in the symmetric case of the random walk Laplacian (8). This limitation currently exists for all asymmetric Laplacians, with potentially complex eigenvalues, complex eigenvectors, and non-Euclidean inner products. Another limitation is that exploration policies are currently being learned for a particular environment, one at a time. Future work can explore how these policies can be learnt in a general way, to be used effectively in other parts of the environment or even in new environments, drawing inspiration from approaches to temporal abstraction that focus on generalisation (32).

Acknowledgements

We would like to thank the members of the Bath Reinforcement Learning Laboratory for their comments and useful suggestions.

References

[1] D. Precup, M. Stolle. Learning Options in Reinforcement Learning *Lecture Notes in Computer Science, vol 2371. Springer, Berlin, Heidelberg.*, 2002. https://www.cs.cmu.edu/~mstoll/pubs/stolle2002learning.pdf.

[2] A. Strehl, M. Littman. An analysis of model-based Interval Estimation for Markov Decision Processes, 2008. https://www.sciencedirect.com/science/article/pii/S0022000008000767.

[3] S. Steinerberger. A Spectral Approach to the Shortest Path Problem *https://arxiv.org/abs/2004.01163*, 2020. https://arxiv.org/abs/2004.01163.

[4] Kaixin Wang and Kuangqi Zhou and Jiashi Feng and Bryan Hooi and Xinchao Wang Reachability-Aware Laplacian Representation in Reinforcement Learning *arXiv preprint arXiv:210.13153*, 2022. https://arxiv.org/abs/2210.13153.

[5] A. Bar, R. Talmon, and R. Meir. Option discovery in the absence of rewards with manifold analysis. *arXiv preprint arXiv:2003.05878*, 2020. https://arxiv.org/abs/2003.05878.

[6] P. Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4):613–624, 1993.

[7] J. B. Evans and Ö. Şimşek. Creating multi-level skill hierarchies in reinforcement learning. *Advances in Neural Information Processing Systems*, 2023.

[8] D. Gomez, M. Bowling, and M. C. Machado. Proper laplacian representation learning. *arXiv preprint arXiv:2310.10833*, 2024. https://arxiv.org/abs/2310.10833.

[9] Y. Jinnai, J. W. Park, D. Abel, and G. D. Konidaris. Discovering options for exploration by minimizing cover time. In *International Conference on Machine Learning (ICML)*, 2019.

[10] M. Klissarov and M. C. Machado. Deep Laplacian-based options for temporally-extended exploration. In *International Conference on Machine Learning (ICML)*, 2023.

[11] A. Levy, G. Konidaris, R. Platt, and K. Saenko. Learning multi-level hierarchies with hindsight. In *International Conference on Learning Representations (ICLR)*, 2019.

[12] M. C. Machado, M. G. Bellemare, and M. Bowling. A Laplacian framework for option discovery in reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2017.

[13] M. C. Machado, C. Rosenbaum, X. Guo, M. Liu, G. Tesauro, and M. Campbell. Eigenoption discovery through the deep successor representation. In *International Conference on Learning Representations (ICLR)*, 2018.

[14] M. C. Machado, A. Barreto, D. Precup, and M. Bowling. Temporal abstraction in reinforcement learning with the successor representation. *Journal of Machine Learning Research*, 24(80):1–69, 2023.

[15] S. Mahadevan. Proto-value functions: Developmental reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2005.

[16] S. Mahadevan and M. Maggioni. Proto-value functions: A Laplacian framework for learning representation and control in Markov decision processes. *Journal of Machine Learning Research*, 8(Oct):2169–2231, 2007.

[17] M. Petrik. An analysis of Laplacian methods for value function approximation in MDPs. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[18] D. Precup. *Temporal Abstraction in Reinforcement Learning*. PhD thesis, University of Massachusetts Amherst, 2000.

[19] R. Ramesh, M. Tomar, and B. Ravindran. Successor options: An option discovery framework for reinforcement learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3304–3310. AAAI Press, 2019.

[20] Ö. Şimşek and A. G. Barto. Using relative novelty to identify useful temporal abstractions in reinforcement learning. In *Proceedings of the 21st International Conference on Machine Learning (ICML '04)*, pages 95–102. ACM, 2004.

[21] Ö. Şimşek, A. P. Wolfe, and A. G. Barto. Identifying useful subgoals in reinforcement learning by local graph partitioning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pages 816–823. ACM, 2005.

[22] Ö. Şimşek and A. G. Barto. Skill characterization based on betweenness. In *Advances in Neural Information Processing Systems 21 (NeurIPS '09)*, pages 1497–1504. Curran Associates, Inc., 2009.

[23] R. S. Sutton, D. Precup, and S. Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999.

[24] H. Tang, R. Houthooft, D. Foote, A. Stooke, O. X. Chen, Y. Duan, J. Schulman, F. DeTurck, and P. Abbeel. #exploration: A study of count-based exploration for deep reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 2753–2762, 2017.

[25] K. Wang, K. Zhou, Q. Zhang, J. Shao, B. Hooi, and J. Feng. Towards better laplacian representation in reinforcement learning with generalized graph drawing. In *International Conference on Machine Learning (ICML)*, 2021.

[26] K. Wang, K. Zhou, J. Feng, B. Hooi, and X. Wang. Reachability-aware laplacian representation in reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2023.

[27] Y. Wu, G. Tucker, and O. Nachum. The Laplacian in RL: Learning representations with efficient approximations. In *International Conference on Learning Representations (ICLR)*, 2019.

[28] K. Yang, J. Tao, J. Lyu, and X. Li. Exploration and anti-exploration with distributional random network distillation. *arXiv preprint arXiv:2401.09750*, 2024. `https://arxiv.org/abs/2401.09750`.

[29] F. Zhang. The Schur Complement and Its Applications. *Numerical Methods and Algorithms*, vol. 4. Springer, 2005. `https://doi.org/10.1007/b105056`.

[30] M. K. G. Kruse, J. M. Conroy, and H. G. Miller. The Rayleigh Quotient. *arXiv preprint arXiv:1112.0292*, 2011. `https://arxiv.org/abs/1112.0292`.

[31] R. S. Varga. Geršgorin and His Circles. *Springer Series in Computational Mathematics*, vol. 36. Springer, 2004. `https://doi.org/10.1007/978-3-642-17798-9`.

[32] Thomas P Cannon and Ö. Şimşek Accelerating Task Generalisation with Multi-Level Skill Hierarchies. *The Thirteenth International Conference on Learning Representations, 2025.*

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: Yes

   Justification: In the paper we do provide evidence that the proposed method (NEO) outperforms related state of the art methods, that we get gradient guarantees and bounds for our method.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: Yes

   Justification: We discuss the limitation in the supplementary material, particularly on the problem of obtaining directed eigenvectors in large continuous spaces.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: Yes

Justification: We provide the assumptions and complete, correct proofs, further details of related proofs are given are in the supplementary material.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: Yes

   Justification: We provide the details needed to re-implement the proposed method (NEO) and do the comparative evaluations of the main results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: Yes

Justification: We provide the source code and data to reproduce the main experiments.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: Yes

   Justification: We provide all of the parameters to reproduce the results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: Yes

   Justification: We use a 2-sigma error width within the learning plots and state this within the main paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: NA

   Justification: We provide the detail on the hardware used to run experiments in supplementary material.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: Yes

   Justification: We have reviewed the code of ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: NA

    Justification: We do not see negativity of societal impacts from our work.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: NA

    Justification: We do not see how the paper can pose a risk.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: Yes

    Justification:

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: NA

    Justification: The paper does not release new assets beyond the source code for experiments.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: NA

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: NA

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: NA

    Justification: LLMs are not involved in the core method development of this research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

8 Appendix

Hardware

All results are obtained using a 9900k CPU, 16GB of ram, a 256 SSD hard drive and a 2080Ti NVIDIA GPU. Results are obtained in hours on a local computer.

Theorem: Novelty Path via Visit Counts (extended)

Here we give an extended theorem and proof of **Theorem 3.3** in the main paper.

Setup: Let $G = (V, E)$ be a finite, directed, strongly connected graph with n nodes. Let

$$P = (P_{ij})_{i,j=1}^{n}, \quad P_{ij} \geq 0, \quad \sum_{j=1}^{n} P_{ij} = 1$$

be its row-stochastic adjacency matrix, and let

$$\Gamma = \operatorname{diag}\big(N(1), \ldots, N(n)\big), \quad N(i) \in_{\geq 0} .$$

Define the source Laplacian

$$L_\zeta \;=\; I + \Gamma - P.$$

Fix an eigenpair (λ, f), $L_\zeta\, f = \lambda f$, and choose a start-node i with $f_i \neq 0$. For each $v \in V$ set

$$z_v \;=\; 1 + N(v) - \lambda,$$

so that componentwise

$$(1 + N(v) - \lambda)\, f_v = \sum_{j=1}^{n} P_{vj}\, f_j, \quad f_v = \frac{1}{z_v} \sum_{j=1}^{n} P_{vj}\, f_j.$$

Standard inequalities: For any nonzero $z \in \mathbb{R}$ or $\mathbb{C}$ and any scalars a_j:

$$\Big|\tfrac{1}{z} \sum_j a_j\Big| = \frac{1}{|z|}\Big|\sum_j a_j\Big|, \qquad \Big|\sum_j a_j\Big| \leq \sum_j |a_j|.$$

Applied at i with $a_j = P_{ij} f_j$, this yields

$$|f_i| = \Big|\tfrac{1}{z_i} \sum_j P_{ij} f_j\Big| = \frac{1}{|z_i|}\Big|\sum_j P_{ij} f_j\Big| \leq \frac{1}{|z_i|} \sum_j P_{ij} |f_j|.$$

Theorem 8.1 (Novelty Path via Visit Counts (extended))**.** Under the above setup, assume the start-node i satisfies the novelty threshold:

- Case 1: $\lambda \in \mathbb{R},\ f \in \mathbb{R}^n,\ N(i) > \lambda$.
- Case 2: $\lambda \in \mathbb{R},\ f \in \mathbb{C}^n,\ N(i) > \lambda$.
- Case 3: $\lambda \in \mathbb{C},\ f \in \mathbb{C}^n,\ N(i) > \mathrm{Re}(\lambda)$.

Then there exists a directed path

$$i = v_0 \ \to\ v_1 \ \to\ \cdots \ \to\ v_k$$

along which

$$|f_{v_0}| < |f_{v_1}| < \cdots < |f_{v_k}|,$$

and which terminates at v_k satisfying:

$$\begin{cases} N(v_k) < \lambda, & \text{in Cases 1–2,} \\ N(v_k) \le \mathrm{Re}(\lambda), & \text{in Case 3.} \end{cases}$$

Proof. We structure the proof in three phases:

I. Strict increase at the start node. First we verify in each case that $|z_i| > 1$, where

$$z_i \ = \ 1 + N(i) - \lambda.$$

Case 1 and 2: Here $z_i = 1 + N(i) - \lambda \in \mathbb{R}$ is a real number. The assumption

$$N(i) > \lambda$$

implies

$$z_i = 1 + N(i) - \lambda > 1 \quad \Longrightarrow \quad |z_i| > 1.$$

Case 3: $\lambda \in \mathbb{C},\ f \in \mathbb{C}^n$**.** Now $z_i = 1 + N(i) - \lambda \in \mathbb{C}$. Its real part is

$$\Re(z_i) = 1 + N(i) - \Re(\lambda).$$

Since

$$N(i) > \Re(\lambda),$$

we have $\Re(z_i) > 1$. By $|z| \ge \Re(z)$ for any complex z,

$$|z_i| \ \ge \ \Re(z_i) \ > \ 1.$$

In all three cases we conclude $|z_i| > 1$. From the componentwise relation,

$$f_i = \frac{1}{z_i} \sum_j P_{ij} f_j.$$

Taking modulus and applying the division and triangle inequalities,

$$|f_i| = \frac{1}{|z_i|} \Big| \sum_j P_{ij} f_j \Big| \le \frac{1}{|z_i|} \sum_j P_{ij} |f_j| < \sum_j P_{ij} |f_j|.$$

Thus $\sum_j P_{ij}|f_j| > |f_i|$, and since $\sum_j P_{ij} = 1$, there exists at least one neighbor v_1 with

$$|f_{v_1}| \ > \ |f_i|.$$

II. Iteration and no cycles. Inductively, while $|z_{v_k}| > 1$, choose v_{k+1} among the out-neighbors of v_k satisfying $|f_{v_{k+1}}| > |f_{v_k}|$. This constructs a strictly increasing real sequence $|f_{v_0}| < |f_{v_1}| < \cdots$. In the finite graph G a strictly increasing real sequence cannot revisit any node, so it must terminate at some v_k with $|z_{v_k}| \le 1$.

III. Termination and Novelty Bound

At the terminal node v_k, no out-neighbor has larger modulus, We now translate $|z_{v_k}| \leq 1$ back into a statement about the novelty count $N(v_k)$:

- **Cases 1–2 ($\lambda \in \mathbb{R}$).** Here
$$z_{v_k} = 1 + N(v_k) - \lambda \quad \in \mathbb{R}.$$
Thus $|z_{v_k}| \leq 1$ is equivalent to
$$-1 \;\leq\; 1 + N(v_k) - \lambda \;\leq\; 1 \quad\Longrightarrow\quad N(v_k) \;\leq\; \lambda.$$
Strong connectivity guarantees at least one strictly positive $P_{v_k j}$, so in fact $N(v_k) < \lambda$.
- **Case 3 ($\lambda \in \mathbb{C}$).** Now $z_{v_k} \in \mathbb{C}$ with $\Re(z_{v_k}) = 1 + N(v_k) - \Re(\lambda)$. From $\Re(z_{v_k}) \leq |z_{v_k}| \leq 1$ we get
$$1 + N(v_k) - \Re(\lambda) \;\leq\; 1 \quad\Longrightarrow\quad N(v_k) \;\leq\; \Re(\lambda).$$

This completes the termination argument and establishes the desired novelty bound at v_k.

9 Tabulated Evaluation Details

9.1 Hyperparameter Settings

Parameter	Value / Frequency
Max Steps per Evaluation	250 000
Evaluation Frequency (epochs)	Every 2 000 steps
Option Initialization Probability	{0.1, 0.2, 0.3, 0.4, 0.5}
Option Discovery Frequency	$5\times$ (#Nodes)
Learning Rate	0.4
Discount Factor	0.99
Epsilon (ε-greedy)	0.1
Instances Per Agent	20

Table 2: Key hyperparameters used across all evaluations.

9.2 Methods Under Evaluation

Method	Description
Q-learning-novelty	Q-learning with novelty intrinsic reward ($\beta/p_N(s)$), with $\beta = 0.01$.
Cover options	To obtain cover options we use the fielder vector representations of normalized Laplacians L_n.
SP-novelty (most-novel)	The option policy follows the shortest path distance to the most novel state (if one exists).
NEO	(Proposed Method)

Table 3: Comparison of all option-discovery methods evaluated.

9.3 Evaluation Metrics

Metric	Description
Sparse Reward Task	One goal state yields 100; others 0. Success = any goal visit (measures discovery).
Count-Based Empirical Evaluation	No learning updates; all rewards zero. Count unique states & edges over 250 000 steps (measures coverage).

Table 4: Metrics for assessing exploration performance.

9.4 Environment Graph Statistics

Environment	#Nodes	#Edges
Four Rooms	104	168
Office	2 558	3 849
Double Large Maze	5 363	9 010
Towers of Hanoi	729	1 092
Hex	5 334	8 346
Torus	5 336	19 028
NYC	10 000	22 354
Rubiks	1 051	1 380

Table 5: Graph statistics for each test environment.

9.5 Novelty scaling function

In the main paper, we transform novelty counts via a scaling function $F : \mathbb{R} \to \mathbb{R}$,

$$F\big(N(s)\big) \;=\; \delta \left(\frac{N(s)}{\max_t N(t)}\right)^{1/k},$$

In figure 5 we show a sweep δ and k, where we represent δ by d in the plots. We sweep the parameters in the directed Hex domain and obtain node counts in the same manner as performed in the main paper (with no environment reward).

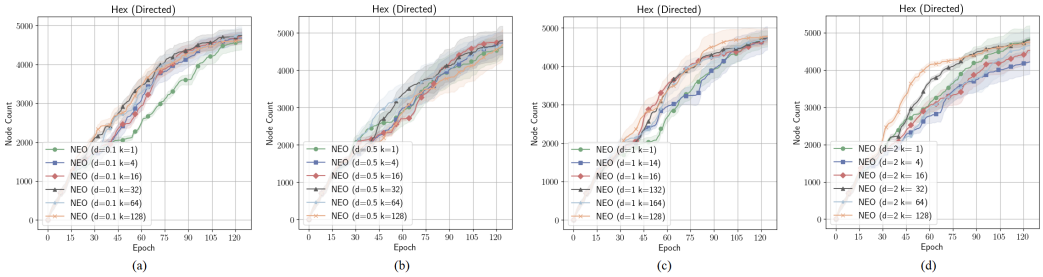

Figure 5: Hyperparameter sweeps of δ (d) and k in the directed Hex domain.

Symmetric Continuous Domains via Rayleigh + MSE

As mentioned in the main paper: "If the graph is symmetric for our source Laplacian then we can express the Rayleigh quiotient(30) as:

$$\mathcal{R}(x) = \frac{x^\top L_\zeta\, x}{x^\top x} = \frac{\frac{1}{2}\sum_{i,j} P_{ij}(x_i - x_j)^2 \;+\; \sum_i \zeta_i\, x_i^2}{\sum_i x_i^2}\,.$$

Which is the standard Rayleigh quiotient for a symmetric matrix $L^{rw} = I - P$ with the additional term $\sum_i \zeta_i\, x_i^2$ which is the added squared smoothness error contributions $\zeta_i(x_i - 0)^2$ between the values x_i and the pulls to zero weighted by ζ_i."

Therefore, extending the proposed method (NEO) to symmetric continuous domains is in principle, simply a matter of extending currently laplacian representation approaches such as (8) by adding a mean square error loss to move sink nodes towards zero based on sink strengths which we can obtain by pseudo-count methods such as distribution random network distillation(28).

10 Bounding $|\operatorname{Im} \lambda(\zeta)|$

We proceed in two stages. First, we show that when ζ is sufficiently large, the spectrum of our source Laplacian splits into two distinct families, type 1 and type 2 eigenvalues, each occupying a different region and characterized by different spectral radii. Second, we form the Schur complement (29) of a carefully chosen submatrix of the source Laplacian. By working with an expanded version of this complement, we derive a bound on the imaginary parts of the type 2 eigenvalues. This is achieved by quantifying how the real and imaginary components of the corresponding eigenvectors interact with the Schur complement (29).

10.1 Source Laplacian $L(\zeta) = (I + \Gamma) - P$

Fix $n \geq 2$. Let $T \subset \{1, \ldots, n\}$ be a non-empty set, $t := |T| \geq 1$. Its complement $S := \{1, \ldots, n\} \setminus T$ is a set, size $m := n - t$. For each $\zeta > 0$ define

$$\Gamma(\zeta) := \operatorname{diag}(\underbrace{\zeta, \ldots, \zeta}_{m \text{ indices } S}, \underbrace{0, \ldots, 0}_{t \text{ indices } T}).$$

Fix a real matrix $P = (p_{ij}) \in \mathbb{R}^{n\times n}, P_{ij} \geq 0, \quad \sum_{j=1}^{n} P_{ij} = 1 \quad (\forall\, i)$ and set

$$L(\zeta) := I_n + \Gamma(\zeta) - P.$$

After permuting indices to order S first,

$$L(\zeta) = \begin{bmatrix} A(\zeta) & B \\ C & D \end{bmatrix},$$

$$\begin{aligned} A(\zeta) &:= (\zeta+1)I_m - P_{SS}, \quad & B &:= -P_{ST}, \\ C &:= -P_{TS}, & D &:= I_t - P_{TT}. \end{aligned}$$

Only $A(\zeta)$ depends on ζ; B, C, D are ζ-independent.

10.2 Type-I eigenvalues stay outside $|\lambda| < \zeta$

The Gershgorin circle theorem(31) can be used to bound the spectrum of square matrices, we use it here to bound the spectrum of $A(\zeta)$ eigenvalues. [Gershgorin discs for $A(\zeta)$] Let $R := \|P_{SS}\|_{2\infty}$ (max row-sum) and $p_{\max} := \max_{1\leq r\leq m} |p_{rr}|$. If $\zeta > 2R + 2p_{\max}$ then $\operatorname{spec} A(\zeta) \subset \{z \in \mathbb{C} : |z| > \zeta/2\}$.

Proof. Row r of $A(\zeta)$ has centre $c_r = \zeta + 1 - p_{rr}$ and radius $R_r \leq R$. Hence $|z| \geq |c_r| - R_r \geq \zeta + 1 - p_{\max} - R > \zeta/2$. □

Definition 10.1. *Type-I eigenvalues* = eigenvalues of $A(\zeta)$.

Remark 10.1. By Lemma 10.2 no type-I eigenvalue can satisfy $|\lambda| < \zeta$ for large ζ; only type-II eigenvalues (defined later) can enter that disc.

10.3 Schur complement with $E = A(\zeta) - \lambda I_m$

We start from the eigenproblem $Mx = \lambda\, x$, which rearranges to $(M - \lambda I)\, x = 0$. Requiring $\det\big(M - \lambda I\big) = 0$ yields the admissible values of λ, and for each such λ we can solve $(M - \lambda I)x = 0$ to find the corresponding eigenvectors x. Writing our block matrix as $M - \lambda I$ also lets us form its Schur complement, so we can explicitly track how the spectrum moves with λ.

Given the block form we established in 10.1 when we take $L(\zeta) - \lambda I_n$ we have:

$$L(\zeta) - \lambda I_n = \begin{bmatrix} E & F \\ G & H \end{bmatrix}, \quad E := A(\zeta) - \lambda I_m,\ F := B,\ G := C,\ H := D - \lambda I_t.$$

therefore the Schur complement of $A - \lambda I$ is given by:

$$S_\zeta(\lambda) := D - \lambda I_t - C\big(A(\zeta) - \lambda I_m\big)^{-1} B. \tag{8}$$

note that by the Schur complement we also have:

$$\det\big(L(\zeta) - \lambda I_n\big) = \det\big(A(\zeta) - \lambda I_m\big)\ \det S_\zeta(\lambda), \tag{9}$$

We pick $E = A - \lambda I$ for the inverse block as it's the only k-dependent block, and we will show that it's inverse includes $1/(\zeta + 1)$ factors which we will use as the main factors in our upper bounds for the imaginary parts of the eigenvalues.

Definition 10.2. *Type-II eigenvalues* = roots of $\det S_\zeta(\lambda) = 0$.

10.4 Expanding $(A(\zeta) - \lambda I_m)^{-1}$

In order to obtain an expansion of $S_\zeta(\lambda)$ needed for our bounds, we have to obtain a Neumann expansion of $(A(\zeta) - \lambda I_m)^{-1}$.$[Neumann - seriesresolvent] Assume \|P_{SS}\|_2 < \zeta$, and $|\lambda| < \zeta$. Define

$$R(\lambda, \zeta) \ := \ \frac{P_{SS} + \lambda I_m}{\zeta + 1}, \qquad \rho_\zeta := \|R(\lambda, \zeta)\|_2 < 1.$$

Then

$$A(\zeta) - \lambda I_m = (\zeta+1)I_m - \big(P_{SS} + \lambda I_m\big) = (\zeta+1)\Big(I_m - R(\lambda,\zeta)\Big), \tag{10}$$

$$\begin{aligned}(A(\zeta) - \lambda I_m)^{-1} &= \frac{1}{\zeta+1}\Big(I_m + R(\lambda,\zeta) + R(\lambda,\zeta)^2 + \cdots\Big) \\ &= \frac{1}{\zeta+1}\Big[I_m + \frac{P_{SS}+\lambda I_m}{\zeta+1} + R_2(\lambda,\zeta)\Big],\end{aligned} \tag{11}$$

$$\|R_2(\lambda,\zeta)\|_2 \le \sum_{j=2}^{\infty} \|R(\lambda,\zeta)\|_2^j = \frac{\rho_\zeta^2}{1-\rho_\zeta}. \tag{12}$$

Proof. From

$$A(\zeta) - \lambda I_m \;=\; {}_m + P_{SS} - \lambda I_m = (\zeta+1)I_m - (P_{SS} + \lambda I_m),$$

factor out $(\zeta+1)$ to get (10). Since $\rho_\zeta < 1$, the Neumann series

$$(I_m - R)^{-1} = I_m + R + R^2 + R^3 + \cdots$$

converges in operator norm. Hence

$$(A(\zeta) - \lambda I_m)^{-1} = \frac{1}{\zeta+1}\,(I_m - R)^{-1} = \frac{1}{\zeta+1}\sum_{j=0}^{\infty} R^j.$$

Separating the $j = 0, 1$ terms,

$$\sum_{j=0}^{\infty} R^j = I_m + R + \sum_{j=2}^{\infty} R^j = I_m + \frac{P_{SS}+\lambda I_m}{\zeta+1} + R_2(\lambda,\zeta),$$

which yields (11). Finally, the tail $R_2(\lambda,\zeta) = \sum_{j=2}^{\infty} R^j$ satisfies

$$\|R_2(\lambda,\zeta)\|_2 \le \sum_{j=2}^{\infty} \|R\|_2^j = \frac{\rho_\zeta^2}{1-\rho_\zeta},$$

proving (12). □

11 Terms of $S_\zeta(\lambda)$

Theorem 11.1 (Schur expansion)**.** under the previous assumptions, with $\|P_{SS}\|_2 < \zeta$, and $|\lambda| < \zeta$. the following hold:

1. *:*

$$C(A - \lambda I_m)^{-1}B = \frac{CB}{\zeta+1} + \frac{C\big(P_{SS}+\lambda I_m\big)B}{(\zeta+1)^2} + \frac{C\,R_2(\lambda,\zeta)\,B}{\zeta+1}.$$

2. *:*

$$S_\zeta(\lambda) = (D - \lambda I_t) - \frac{CB}{\zeta+1} \;+\; E_\zeta(\lambda),$$

where

$$E_\zeta(\lambda) = -\frac{C\big(P_{SS}+\lambda I_m\big)B}{(\zeta+1)^2} - \frac{C\,R_2(\lambda,\zeta)\,B}{\zeta+1}. = -\frac{C\,P_{SS}\,B}{(\zeta+1)^2} - \frac{\lambda}{(\zeta+1)^2}\,CB - \frac{C\,R_2(\lambda,\zeta)B}{\zeta+1}$$

Proof. Substitute the Neumann expansion of $(A - \lambda I_m)^{-1}$ into $C(A - \lambda I_m)^{-1}B$:

$$C(A - \lambda I_m)^{-1}B = C\Big[\tfrac{1}{\zeta+1}I_m + \tfrac{1}{(\zeta+1)^2}(P_{SS}+\lambda I_m) + \tfrac{1}{\zeta+1}R_2(\lambda,\zeta)\Big]B,$$

which gives the stated expansion for part 1. Starting with our definition of $S_\zeta(\lambda)$ inserting the result of $C(A - \lambda I_m)^{-1}B$ yields

$$S_\zeta(\lambda) = D - \lambda I_t - \frac{CB}{\zeta+1} - \frac{C(P_{SS}+\lambda I_m)B}{(\zeta+1)^2} - \frac{C\,R_2(\lambda,\zeta)\,B}{\zeta+1},$$

and grouping the last two terms into $E_\zeta(\lambda)$ gives the expansion in part 2. □

Imaginary–part bound for Type-II eigenvalues

Theorem 11.2 (Explicit imaginary-part bound)**.** Let $\zeta > \|P_{SS}\|_2$ and suppose $\lambda(\zeta) \in \operatorname{spec} L(\zeta)$ with $|\lambda(\zeta)| < \zeta$. Let $y \neq 0$ be the corresponding Type-II eigenvector of the Schur complement $S_\zeta(\lambda)$, normalized so $\|y\|_2 = 1$. Split

$$D = D_H + D_A, \qquad D_H = \tfrac{1}{2}\,(D + D^\top), \quad D_A = \tfrac{1}{2}\,(D - D^\top),$$

and define

$$\alpha(\zeta) \;=\; y^* D_A y \in \mathbb{R}, \qquad \beta(\zeta) = |\alpha(\zeta)|.$$

Then

$$\big|\Im\lambda(\zeta)\big| \;\le\; \beta(\zeta) \;+\; \frac{\|B\|_2\,\|C\|_2}{\zeta+1} \;+\; \frac{\|B\|_2\,\|C\|_2\big(\|P_{SS}\|_2 + \zeta\big)}{(\zeta+1)^2}.$$

Proof. From $S_\zeta(\lambda)y = 0$ and $\|y\| = 1$ we get

$$0 \;=\; y^*\big(D - \lambda I\big)y \;-\; \frac{1}{\zeta+1}\,y^*CBy \;+\; y^*E_\zeta(\lambda)\,y, \tag{$P1$}$$

with

$$E_\zeta(\lambda) = -\,\frac{C\,(P_{SS} + \lambda I)\,B}{(\zeta+1)^2} - \frac{C\,R_2(\lambda,\zeta)\,B}{\zeta+1}, \quad \|R_2(\lambda,\zeta)\|_2 \le \frac{(\|P_{SS}\|_2 + \zeta)^2}{(\zeta+1)^2}. \tag{$P2$}$$

Taking imaginary parts and using $y^*Dy = y^*D_Hy + \alpha(\zeta)$ gives

$$0 = -\,\Im\lambda(\zeta) + \alpha(\zeta) - \frac{1}{\zeta+1}\,\Im\big(y^*CBy\big) + \Im\big(y^*E_\zeta(\lambda)y\big),$$

so

$$\big|\Im\lambda(\zeta)\big| \;\le\; |\alpha(\zeta)| \;+\; \frac{|y^*CBy|}{\zeta+1} \;+\; \big|y^*E_\zeta(\lambda)y\big|. \tag{$P4$}$$

Now finally we can bound each scalar explicitly,

$$\begin{aligned}
|y^*CB\,y| &= |(By)^*(Cy)| \le \|By\|_2\|Cy\|_2 \le \|B\|_2\,\|C\|_2,\\
|y^*E_\zeta(\lambda)y| &\le \|C\|_2\|B\|_2\Big(\frac{\|P_{SS}\|_2 + |\lambda|}{(\zeta+1)^2} + \frac{\|R_2(\lambda,\zeta)\|_2}{\zeta+1}\Big)\\
&\le \|B\|_2\,\|C\|_2\Big(\frac{\|P_{SS}\|_2 + \zeta}{(\zeta+1)^2} + \frac{(\|P_{SS}\|_2 + \zeta)^2}{(\zeta+1)^3}\Big)\\
&\le \frac{\|B\|_2\,\|C\|_2\big(\|P_{SS}\|_2 + \zeta\big)}{(\zeta+1)^2} \quad (\text{for } \zeta \ge \|P_{SS}\|_2).
\end{aligned}$$

and the result follows. □

Remark 11.1. The term $\beta(\zeta) = |y^*D_Ay|$ is the baseline imaginary-part contribution from the skew-symmetric part of D, vanishing exactly when D is symmetric. The remaining two terms decay like $(\zeta+1)^{-1}$ and $(\zeta+1)^{-2}$, respectively, reflecting the heavy-diagonal damping. Therefore, if $D = D^\top$ then $D_A = 0 \Rightarrow \beta(\zeta) = 0$ and the bound simplifies to

$$|\operatorname{Im}\lambda(\zeta)| \le \frac{\|B\|_2\,\|C\|_2}{\zeta+1} + \frac{\|B\|_2\,\|C\|_2\big(\|P_{SS}\|_2 + k\big)}{(\zeta+1)^2}.$$

The entire imaginary part is throttled by the heavy diagonal.

11.1 Eigenvalue bound observations

In figure 6 we generated nine plots showing the real versus imaginary parts of the eigenvalues of our source Laplacian with a sink strength ζ. According to our theory, the sink strength ζ imposes an upper bound on the imaginary parts of the type-2 eigenvalues. To test this, we work on the 104-node Directed Four Rooms domain, in which four nodes are designated as sources with sink-strength zero and the remaining 100 nodes as sinks with strength ζ). For each of the nine chosen values of ζ, we

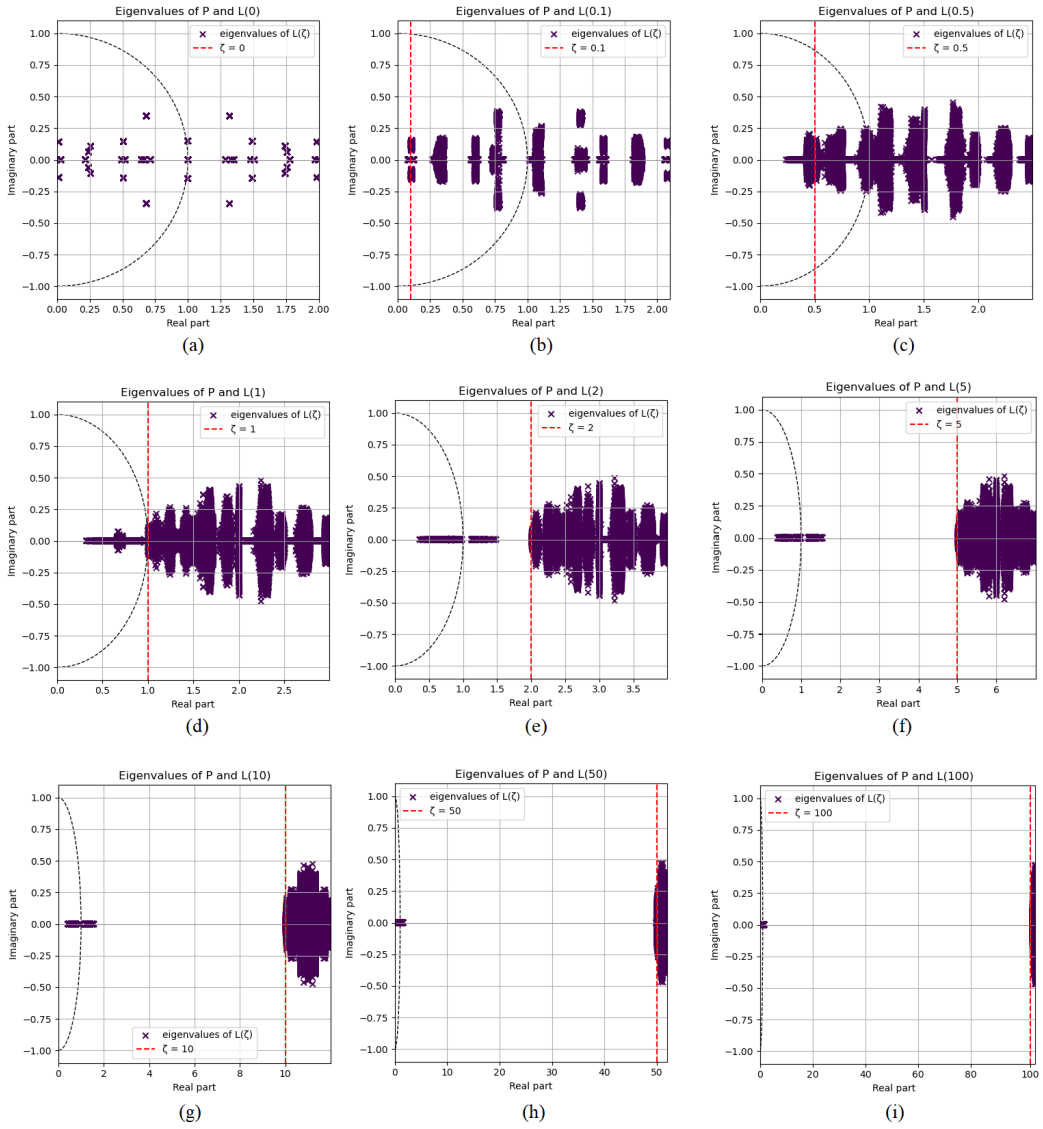

Figure 6: Eigenvalue plots where we show the real and imaginary parts for source Laplacian under different sink strengths

repeat the following 5,000 times: randomly select four source nodes, assign sink-strengths on the remaining 100 nodes, build the directed Laplacian, compute its spectrum, and collect all eigenvalues.

As ζ increases, the two classes of eigenvalues type 1 (those farther from zero) and type 2 (those clustering near zero) separate into distinct regions. In particular, the imaginary parts of the type 2 eigenvalues shrink toward zero as ζ grows. When $\zeta = 0$ (in plot A), our Laplacian reduces to the standard random-walk form where there is no distinction between type 1 and type 2 eigenvalues with both having significant imaginary parts. By contrast, at the largest ζ (plot I), only the type 1 eigenvalues that exceed exhibit significant nonzero imaginary parts.